Extended Depth-of-Field Imaging and Ranging in Microscopy

Paul Zammit

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

October 2016

Abstract

Conventional 3D imaging techniques such as laser scanning, focus-stacking and confocal microscopy either require scanning in all or a subset of the spatial dimensions, or else are limited by their depth of field (DOF). Scanning increases the acquisition time, therefore techniques which rely on it cannot be used to image moving scenes. In order to acquire both the intensity of the scene and its depth, extending the DOF without scanning is therefore necessary. This is traditionally achieved by stopping the system down (reducing the $f_{\#}$). This, however, has the highly undesirable effect of lowering both the throughput and the lateral resolution of the system. In microscopy in particular, both these parameters are critical, therefore there is scope in breaking this trade-off. The objective of this work, therefore, is to develop a practical and simple 3D imaging technique which is capable of acquiring both the irradiance of the scene and its depth in a single snapshot over an extended DOF without incurring a reduction in optical throughput and lateral resolution. To this end, a new imaging technique, referred to as complementary Kernel Matching (CKM), is proposed in this thesis.

To extend the DOF, in CKM a hybrid imaging technique known as wavefront coding (WC) has been used. WC permits the DOF to be extended by an order of magnitude typically without reducing the efficiency and the resolution of the system. Moreover, WC only requires the introduction of a phase mask in the aperture of the system, hence it also has the benefit of simplicity and practicality. Unfortunately, in practice, WC systems are found to suffer from post-recovery artefacts and distortion, which substantially degrade the quality of the acquired image. To date, this long-standing problem has found no solution and is probably the cause for the lack of exploitation of this imaging technique by the industry.

In CKM, use of a largely ignored phenomenon associated with WC was made to measure the depth of the sample. This is the lateral translation of the scene in proportion to its depth. Furthermore, once the depth of the scene is known, the ensuing artefacts and distortion due to the introduction of the WC element can be compensated for. As a result, a high quality intensity image of the scene and its depth profile (referred to in stereo vision parlance as a depth map) is obtained over a DOF which is typically an order of magnitude larger than that of an equivalent clear-aperture system. This implies that, besides being a 3D imaging technique, CKM is also a solution to one of the longest standing problem in WC itself. By means of WC, therefore, the

DOF was extended without scanning and without reducing the throughput and the optical resolution, allowing both an intensity image of the scene to be acquired and its depth map. In addition, CKM is inherently monocular, therefore it does not suffer from occlusion, which is a major problem affecting triangulation-based 3D imaging techniques such as the popular stereo vision. One therefore concludes that CKM fulfils the objectives set for this project.

In this work, various ways of implementing CKM were explored and compared; and the theory associated with them was developed. An experimental prototype was then built and the technique was demonstrated experimentally in microscopy. The results show that CKM eliminates WC artefacts and thus gives high quality images of the scene over an extended DOF. A DOF of $\sim 20\mu m$ was achieved on a $40\times, 0.5$NA system experimentally, however this can be increased if required. The experimental depth reconstructions of real samples (such as pollen grains and a silicon die) imaged in various modalities (reflection, transmission and fluorescence) were comparable to those given by a focus-stack. However, as with all other passive techniques, the performance of CKM depends on the texture and features in the scene itself. On a binary systematic scene consisting of regularly spaced dots with a linear depth gradient, an RMS error of $\pm 0.15\mu m$ was obtained from an image signal-to-noise ratio of $60dB$.

Finally, owing to its simplicity and large DOF, there is scope in investigating the possibility of using the same CKM setup for 3D point localisation applications such as super resolution. An initial investigation was therefore conducted by localising sub-resolution fluorescent beads. On a $40\times, 0.5$NA system, a mean precision of $148nm$ in depth and $< 30nm$ in the lateral dimensions was observed experimentally from $4,000$ photons per localisation over a DOF of $26\mu m$. From these experimental values, a mean localisation precision of $< 34nm$ in depth and $< 13nm$ in the lateral dimensions from $2,000$ photons per localisation over a DOF of $3\mu m$ is expected on a more typical $100\times, 1.4$NA system. This compares favourably to the competition, therefore we conclude that there is scope in investigating this technique for 3D point localisation applications further.

Table of contents

List of tables

List of figures

Acknowledgements

I would like to acknowledge the contribution and thank the following:

My supervisor, Prof. A. Harvey, for giving me the opportunity to work with his research group. His encouragement and guidance were indispensable and very instructive.

Dr. G. Carles who proposed the conjugate-masks CKM while I was independently developing the biplane method on point-source scenes (see Chapter 2) and for the myriad insightful technical discussions we had during the course of my studies. His contribution is greatly appreciated.

Finally, I would like to thank all members of the imaging concepts group (ICG) at the University of Glasgow who were very welcoming and helpful throughout my stay. Working with them was a pleasure. Thank you.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Paul Zammit
October 2016

Abbreviations

Acronyms / Abbreviations

ADU Analogue-To-Digital Unit

APSF Amplitude Point Spread Function

CKM Complementary Kernel Matching

CPM Cubic Phase Mask

CRLB Cramér-Rao Lower Bound

DFD Depth From Defocus

DH-PSF Double Helix Point Spread Function

DOF Depth Of Field

FFT Fast Fourier Transform

FOV Field Of View

FWHM Full-Width Half-Maximum

LSF Least-Square Fit

MLE Maximum Likelihood Estimation

MTF Modulation Transfer Function

MUM Multifocal Plane Microscopy

NA Numerical Aperture

NCC Normalised Cross Correlation

NSR Noise-To-Signal Ratio

OCT Optical Coherence Tomography

OPD Optical Path Difference

OTF Optical Transfer Function

PIV Particle Image Velocimetry

PSF Point Spread Function

PTF Phase Transfer Function

PWL Piece-wise Linear

SAD Sum Of Absolute Differences

SB-PSF Self-Bending Point Spread Function

SLM Spatial Light Modulator

SNR Signal-To-Noise Ratio

SP-PSF Saddle-Point Point Spread Function

SPIM Selective Plane Illumination Microscopy

SSD Sum Of Squared Differences

WC Wavefront Coding

Publications

Journal:

- P. Zammit, A. R. Harvey, and G. Carles, "Extended depth-of-field imaging and ranging in a snapshot," Optica, vol. 1, no. 4, pp. 209–216, 2014.
- Chapter 4 (main results of this study) is adapted to the following publication in preparation: P. Zammit, G. Carles, A. R. Harvey, "Snapshot 3D imaging with an extended depth of field." To be submitted to Nature Photonics.
- Chapter 2 (Section 2.1 and Section 2.5) is adapted to the following publication in preparation: P. Zammit, G. Carles, A. R. Harvey, "Augmenting the information content of the cubic point spread function." To be submitted to Optics Express.

Patent application:

- G. Carles, P. Zammit, A. R. Harvey, "Method and apparatus for evaluating defocus in an image of a scene" (patent application number: GB2015/051743).

Conference:

- P. Zammit, A. R. Harvey, and G. Carles, "Passive 3D imaging in a single snapshot with an extended depth of field," in Classical Optics 2014, p. CTh2C.5, OSA, 2014.
- A. R. Harvey, N. Bustin, J. Downing, G. Carles, G. Muyo, A. Wood, and P. Zammit, "Computational imaging for simpler optics," in International Optical Design Conference 2014, p. ITu3A–1, OSA, 2014.
- P. Zammit, A. R. Harvey, and G. Carles, "Single snapshot, extended depth of field, 3D-imaging for incoherent imaging applications," Photon14, 2014, London, UK.
- P. Zammit, A. R. Harvey, and G. Carles, "Practical single snapshot 3D imaging method with an extended depth of field," in Imaging and Applied Optics 2015, p. CT2E.2, OSA, 2015.

- P. Zammit, G. Carles, and A. R. Harvey, "3D imaging and ranging in a snapshot," Proc. SPIE, vol. 9630, pp. 963004–963004–9, 2015. [**invited**]
- P. Zammit, G. Carles, and A. R. Harvey, "Three-dimensional imaging and ranging in a snapshot with an extended depth-of-field," in Imaging and Applied Optics 2016, p. CW2D.1, OSA, 2016.

Book articles:

- AR Harvey, P. Zammit, G. Carles, G. Muyo, and S. Mezouari, "Wavefront-Coded, Hybrid Imaging for the Alleviation of Optical Aberrations," in Materials Science and Materials Engineering (MATS), 2016. [**invited**]

Chapter 1

Introductory Material

In this chapter, the organisation of this thesis is first presented, followed by an explicit statement of the motivation and objectives of this study. A review of the pertinent literature is then presented in order to establish the baseline for this work. This is then followed by a brief introduction to wavefront coding and the related theory, since this is essential to the understanding of the material in the rest of this thesis. An overview of the solution proposed in this work is discussed next, in order to put the reader in perspective. Lastly, the contributions of this study are summarised and the chapter is concluded.

1.1 Thesis organisation

The work conducted here was split into the following phases:

- Problem identification
- Solution identification
- Modelling and testing the proposed solution
- Implementation and experimental verification

The layout of this thesis was therefore designed to reflect this process. In this chapter, the first two points (problem and solution identification) are dealt with. These are supported by the relevant literature, therefore the literature review has been included in this chapter. In addition, background material of direct relevance to this work is covered in appropriate detail.

In Chapter 2, the theory relevant to the identified solution is presented in detail. The chapter starts by modelling the cubic point spread function (PSF) in space, which is at the heart of this work. We then proceed to the mathematical formulation of the proposed solution itself. Various ways of implementing it are then investigated and

compared. Lastly, an algorithm which can be used to implement the proposed solution is presented for two scenarios: extended scenes, which are the primary objective of this work and spares point-source scenes. This chapter is essential to the understanding of the remainder of this thesis.

The modelling and testing phase is presented in Chapter 3. Numerical simulations were used to confirm the theoretical inferences and predictions discussed in Chapter 2. A proof-of-concept experiment intended as an investigation of the feasibility of the solution is also presented.

In Chapter 4, the actual implementation of the proposed solution is presented. In this chapter the proposed solution is tested experimentally on extended scenes. Furthermore, an initial investigation of the suitability of the proposed solution for point localisation was also performed and is also included in this chapter.

Finally, the conclusions of this work are collected in Chapter 5, together with salient future work and improvements which are yet to be undertaken.

1.2 Objectives and motivation

As will be discussed in more detail in the following sections, mainstream 3D imaging techniques (*i.e.* techniques capable of acquiring both the irradiance and depth information of a sample) involve either reducing the numerical aperture (NA) of the optical system (such as stereo imaging and light field cameras/microscopes) or otherwise scanning -*i.e.* multiple acquisitions- (such as confocal imaging or Z-stacking alternatively known as focus stacking). However, reducing the NA has the obvious disadvantage of lowering the spatial resolution and the optical throughput, whereas scanning entails longer acquisition times in addition to the use of precision mechanical positioning equipment. The increase in acquisition time can make such techniques unsuitable to image moving scenes or samples such as live samples in microscopy or inspecting items on a conveyor belt. On the other hand, precision mechanical positioning equipment substantially increases the cost of the system and its complexity. Tackling these shortcomings is the motivation behind this work. The objective of this work can therefore be summarised as: the development of a practical and simple imaging technique which is capable of acquiring both the irradiance of the scene and its depth in a single snapshot without incurring a reduction in optical throughput and lateral resolution.

A key milestone in achieving this objective is to extend the depth of field (DOF) of the optical system without reducing its NA and without mechanical movement. A solution to this problem was proposed by Dowski *et. al.* in their seminal 1995 paper [1]. Wavefront coding (WC) requires only a phase mask to be placed in the aperture stop of the system to extend the DOF by a significant amount (typically by an order of magnitude). Despite this, the potential of WC has found little exploitation outside the

research community. This is primarily due to post-recovery artefacts which significantly degrade the quality of the recovered images [2]. This problem remains unsolved to date, therefore successful employment of this technology first requires a solution to this problem. As stated in [2], in order to eliminate the artefacts in question, knowledge of the depth of the sample is required. Given that the principal objective of this work is to develop a 3D imaging technique and hence the depth of the sample will be known, then, the artefact problem is inherently solved.

The objectives of this work can therefore be listed as follows:

- Identify a means of acquiring 3D information about the sample over an extended DOF without reducing NA and without mechanical movement in a single-snapshot. As stated previously, the 3D information in question consists of: (1) the irradiance of the sample and (2) its depth at each lateral location which is referred to as a "depth map" in stereo vision parlance. The technique proposed here for this purpose was dubbed Complementary Kernel Matching (CKM) and is based on WC.
- Model and test the technique identified above.
- Build a setup to demonstrate the technique (CKM) experimentally. For this purpose, microscopy was chosen as the application area, therefore CKM was implemented around a microscope.
- Demonstrate the removal of artefacts experimentally, hence achieving an image quality superior to that given by WC.
- Demonstrate 3D imaging of arbitrary extended samples experimentally.
- Finally, it is hypothesised that the same technique can be modified for applications which involved point localisation such as particle image velocimetry (PIV) and super-resolution. The advantage of this technique in these applications is expected to be a relatively large localisation depth range. A secondary objective of this work is therefore an initial appreciation of the potential of this technology for such purpose.

The original contributions of this work are:

- Complementary Kernel Matching, which is a novel technique based on WC with the unique capability of acquiring 3D images of an extended sample over an extended DOF in a single snapshot (without scanning or a reduction in NA).
- A solution to the long-standing problem of post-recovery artefacts in WC.

1.3 Literature review

The literature review conducted here is intended to establish the baseline for this work in the matter of 3D imaging technologies and techniques. It should be noted, however, that due to the broad spectrum of techniques available and their variants, it is virtually impossible to cover them all [3]. As a result, only the most prominent were considered and only those which are broadly applicable to microscopy. Techniques such as time of flight [4, 3], which is generally used in macroscopic imaging, were therefore not considered. Furthermore, here we are mostly interested in the capability and practical aspects of the techniques rather than the theory behind them; therefore, the theory will not be covered in depth. Appropriate references are, however, provided for further reading.

1.3.1 Laser triangulation

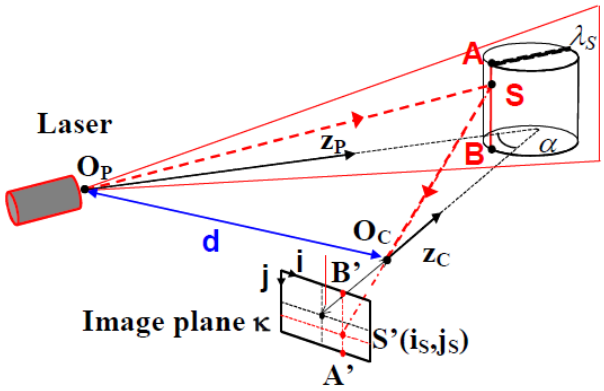


Fig. 1.1 Laser triangulation. Image taken from [3].

Laser triangulation is a widely-known principle which has found employment in a broad range of both microscopic [5] and macroscopic applications, such as 3D mapping of heritage sites and reverse engineering [3]. In addition, this principle lies at the heart of other important 3D imaging methods such as structured illumination and stereo vision, therefore it is of particular relevance to this work. Fig. 1.1 shows a representation of a typical laser scanning configuration. The baseline d measured from the exit pupil of the laser (O_p) to the entrance pupil of the detector (O_c) defines the angle between the optical axis of the illumination (Z_p) and that of the detector (z_c) denoted in the figure by α. This angle, in turn, stipulates the resolution and precision of the depth measurements.

For point scanning, the laser source shines a thin collimated beam onto the surface of the target which is then imaged onto the detector. Assuming that the beam impinges on the target, at point S in 3D space and that its image on the detector occurs at S', then the position of point S in 3D space can be calculated by triangulating the line of sight $\overline{S', O_c}$ and the known direction of the laser beam. To form a 3D image of a

surface scanning is therefore necessary; hence, this method suffers from a prolonged acquisition time. An improvement on this is what is known as slit scanning [6]. In slit scanning, instead of scanning a point in two dimensions, a light sheet produced by expanding a laser beam using a cylindrical lens is used instead. Scanning in this case is necessary along one dimension only, therefore the acquisition time is reduced significantly. Triangulation is then performed between the lines of sight $\overline{S'\left(\overline{i_s}, \overline{j_s}\right), O_c}$ to the plane of the light sheet λ_s. With reference to Fig. 1.1, $S'\left(\overline{i_s}, \overline{j_s}\right)$ denotes the set of points along line $\overline{A', B'}$ which are the image of each point S on the target illuminated by the light sheet along the illuminated line $\overline{A, B}$.

Besides a prolonged acquisition time, laser triangulation has other disadvantages. These include laser and/or camera occlusion, safety considerations due to laser radiation, susceptibility to the reflectivity of target in addition to laser speckle. The problem of occlusion can be dealt with using a technique known as synchronised scanning [7]. This, however, significantly increases the complexity of the system and hence the cost. The dependence on the reflectivity of the surface and the effect of laser speckle can be mitigated using a dual aperture mask [8]. This permits the position of the image on the detector to be determined differentially, therefore these effects cancel out. On the other hand, laser triangulators achieve high accuracies, and are insensitive to illumination conditions, as well as to the texture of the sample itself. This contributes to their wide appeal, despite the fact that they are rather complex and expensive.

1.3.2 Structured light (also known as pattern projection)

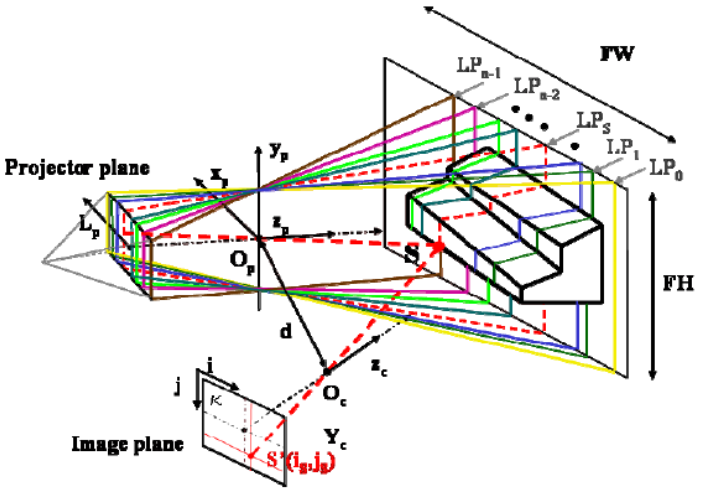


Fig. 1.2 Structured light 3D imaging techinque. Image taken from [3].

The structured light method operates on the same triangulation principle discussed above. In this case, a 2D pattern is projected onto the target illuminating various regions of it simultaneously as shown in Fig. 1.2 where a striped pattern has been assumed. This can be viewed as slit scanning (described above) with multiple slits illuminating the target simultaneously. Range can therefore be determined by triangulating the line

of sight of each point imaged on the detector to the correct illuminating light sheet. For instance, the position in 3D space of point S on the target which is illuminated by light sheet LP_s can be found by triangulating the line of sight $\overline{S'\left(i_s, j_s\right), O_c}$ with light sheet LP_s. Of key importance therefore is the identification of the correct illuminating light sheet corresponding to each point imaged on the detector. To this end, various projection strategies have been proposed in the literature, including the projection of dot patterns, grid lines, multi-colour patterns and binary or gray coded patterns [3]. In the latter, multiple vertical striped patterns are projected onto the target in succession and an image is acquired for each. The illuminating light sheet corresponding to each pixel of the detector can then be identified from the intensity pattern of the pixel itself in each acquired frame. This is one of the most popular methods used to solve this ambiguity. Finally, some patterns intended to maximise the resolution of the measurements have also been proposed in the literature, such as superimposed sinusoidal patterns with different frequencies in [9].

One of the main advantages of this technique over laser triangulation is the fact that incoherent illumination is used to project the pattern. This implies that rather than a dedicated laser scanner, a simple off-the-shelf projector can be used, which results in an appreciable cost reduction. Other advantages of this technique are its high accuracy, the fact that it does not rely on the texture of the target and its speed of acquisition compared to laser scanning. In fact, although this technique is not a single-snapshot technique, it can achieve real-time operation and has also been used to image moving targets. On the other hand, as with laser scanning, this technique is still susceptible to the reflectivity of the target's surface and to occlusion, although it is inherently immune to speckle-related problems. Finally, compared to passive systems such as stereo vision, this technique can be considered to be complicated.

1.3.3 Stereo vision

Stereo vision is another technique which relies on the triangulation principle explained above, although in this case, no active illumination is used. Instead, two or more cameras are positioned at an angle to each other; therefore, each camera captures an image of the scene from a different perspective. The acquired images are then transformed from camera coordinates to world coordinates by means of a camera model which takes into account projection and other distortion. Correspondence is then established by matching the features or the texture of the scene itself in the acquired images. Note that there are various ways of measuring correspondence, however these will be covered in Section 2.6 due to their relevance to the technique proposed here. Once correspondence has been established, triangulation is performed and hence range information is obtained.

The main advantages of stereo vision over the other triangulation methods lies in its simplicity and low cost hardware. Moreover, stereo vision is a single-snapshot technique, whereas the others require scanning to acquire the entire field of view (FOV). On the other hand, the fact that this technique relies on the features and the texture of the scene itself to establish correspondence implies that the accuracy and quality (density) of the acquired range information depends on the scene being imaged. As a result, stereo vision tends to display moderate to low range accuracy. Furthermore, as with all other triangulation techniques, stereo vision is prone to occlusion which limits the width of the baseline, hence its resolution and precision. Despite its disadvantages, the simplicity of this technique makes it very appealing. In fact, besides the substantial research endeavour in this field, especially in the computer science field, stereo vision has found employment in many applications including microscopy, robotic or computer vision and collision avoidance.

1.3.4 Interferometry

Interferometric techniques operate by sensing path length difference between a reference beam and a test beam. This is usually accomplished using classical configurations such as the well known Mach-Zehnder and Twyman-Green interferometers. These techniques achieve very high accuracy (a fraction of the wavelength) on a small depth range where the unambiguous range is equal to the wavelength. The unambiguous depth range can be increased through multi-wavelength interferometry [10]. In addition, multi-wavelength interferometry has also been proposed to compensate for stray fluctuations in the refractive index of the medium. These interferometric techniques are mostly used for surface quality control, such as in specifying the surface roughness and scratch-dig of optical surfaces which are relatively smooth. On rougher surfaces, speckle dominates over the interference pattern therefore these methods cannot be used directly. A way of solving this problem is to perform two measurements under identical conditions but using two different wavelengths. The difference between the two acquired images then reveals the interference pattern [10].

Another popular configuration is the Mirau interferometer shown in Fig. 1.3 which is often used in surface profilometry. In this case, the two arms of the interferometer share the same optical path and are nearly the same length. This results in a compact arrangement, making it possible to incorporate in a microscope objective. Furthermore, this feature relaxes the coherence length requirement of a suitable light source. As a result, this kind of interferometer is suitable for white-light interferometry. In white-light interferometry, a source with a comparatively short coherence length (compared to lasers) is used thus reducing the cost of the system. Due to the short coherence length of the light, interference occurs only when the path lengths along each arm are nearly the same, with the intensity peaking when they are exactly the same length. To find the depth of the target therefore, a series of images are acquired while stepping the target

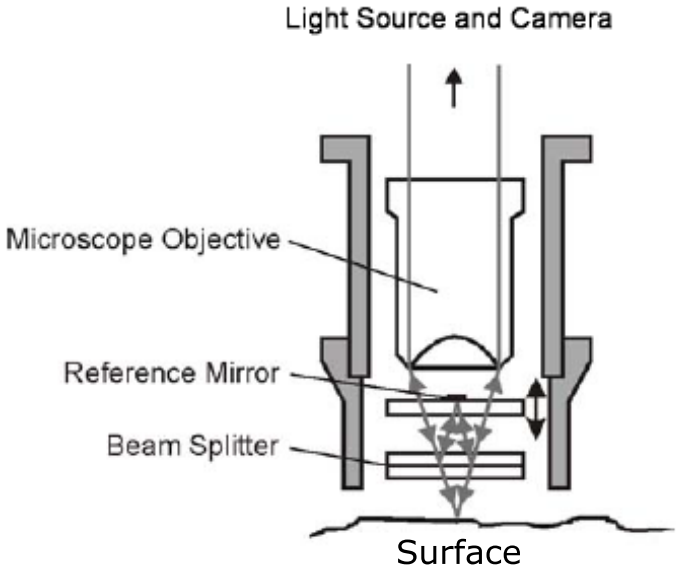


Fig. 1.3 Mirau interferometer. Image taken from [10].

or the interferometer itself along the z direction. The depth at each pixel location is then equal to the depth which maximises the intensity of the pixel. This technique can achieve a resolution of $0.1nm$ over a $100\mu m$ depth range and it is often employed in micro- and nanotechnology [10]. White-light interferometry is also the principle behind optical coherence tomography (OCT), which is widely used in the medical field due to its unrivalled tissue penetration depth [11].

The principal advantages of these techniques are their accuracy and resolution. Their main disadvantages are their limited depth range, cost and their complexity which is accentuated further by their sensitivity to environmental factors such as temperature, air currents and vibrations.

1.3.5 Moiré topography

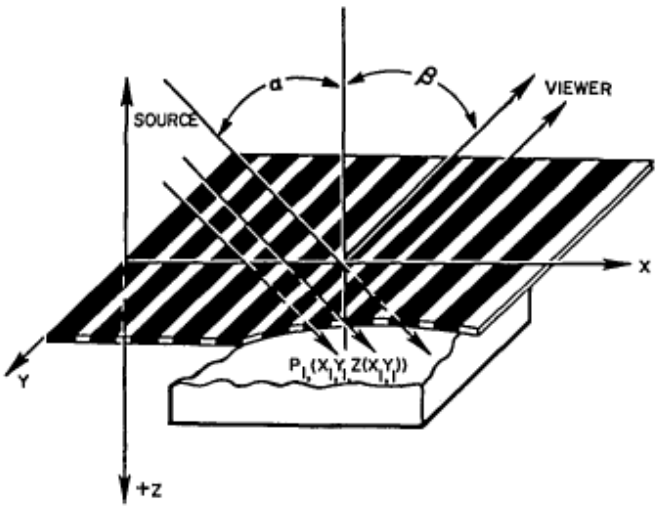


Fig. 1.4 Moiré topography. Image taken from [12].

Moiré topography was originally proposed by Takasaki [13] in the 1970s and operates on the principal of interference. A typical arrangement is shown in Fig. 1.4 where a laser is used to illuminate a sample at an angle α through a grating. The reflected light is then viewed through the same grating at an angle β as shown. The resultant fringe pattern which exists on the surface of the object as seen by the detector is given by [12]:

$$I_M(x_1, y_1) = A\left[1 + \frac{1}{2} \cdot \cos\left(\frac{2\pi\left[\tan(\alpha) + \tan(\beta)\right] z(x_1, y_1)}{p}\right)\right] \tag{1.1}$$

where p denotes the period of the grating, (x_1, y_1) denote the lateral coordinates of a point on the surface such as P_1 and $z(x_1, y_1)$ denotes the height or depth of the surface as a function of the lateral coordinates. Eq. 1.1 represents a series of contours from which the depth can be estimated by unwrapping the phase term:

$$\theta = \frac{2\pi\left[\tan(\alpha) + \tan(\beta)\right] z(x_1, y_1)}{p} \tag{1.2}$$

The main advantage of this technique lies in its simplicity despite the fact that an illumination source is still required. On the other hand, the basic technique displays relatively low accuracies. Several techniques, such as phase-locked fringe detection [12] can be used to improve this, however these techniques significantly increase the complexity of the technique. Furthermore, the basic technique also suffers from phase ambiguity caused by phase wrapping and is also incapable of discerning the sign of a concavity. Phase ambiguity problems [14] limit the size of discontinuities in the depth profile at any point (x, y) to:

$$z(x, y) < \frac{p}{\left[\tan(\alpha) + \tan(\beta)\right]} \tag{1.3}$$

consequently, Moiré topography is typically used to measure largely flat surfaces with small and smooth variations in depth. Finally, it should be noted that defocus further lowers the accuracy of this method, therefore the depth of field can pose a limit on the maximum measurable range. Despite its disadvantages, Moiré topography is extensively used in medicine to perform anatomical measurements, for instance, in the diagnosis of scoliosis.

1.3.6 Shape from focus or defocus

Defocus is a well known phenomenon in conventional optical system. Given that this is proportional to the distance of the object with respect to the focal plane, measuring the defocus (or blur) gives a measure of the depth of the object. There are several implementations which rely on this phenomenon, including passive and active systems. In passive systems, blur is evaluated on the texture and features of the scene. The performance of passive shape from focus systems is therefore dependent on the scene itself. In active systems, patterned illumination is projected onto the target surface,

creating texture artificially. This technique reduces the reliance on the scene, hence making the technique more versatile. Traditional shape from focus/defocus techniques require the acquisition of two or more images of the scene in order to have a reference and hence make it possible to assess the blur at different locations in the scene. Some techniques where a single image is used to infer depth (or shape) from defocus (and other monocular depth cues) can be found in the literature [15], however, these techniques in general are less robust and less accurate.

Another technique which relies on defocus to infer depth or shape is the well-known focus-stacking or Z-stack. In this method, a stack of images of the scene is acquired at definite depth intervals. Instead of quantifying the amount of blur in the scene, the sharpness of the image is assessed locally in each of the acquired frames and the most in-focus regions are then stitched together to produce the final reconstruction. The local depth of the scene is equal to the depth at which the frame which maximises the sharpness there has been acquired. Sharpness is assessed using several methods, including simple high-pass and Sobel filters; and wavelet-based techniques [16]. Z-stacking is nowadays implemented in most commercial high-end microscopes.

The principal advantage of these techniques is their simplicity because no additional optics are necessary beside the imaging optics. The most commonly used implementation is, however, Z-stacking and for this, a precise mechanised Z stage is necessary. Unlike stereo vision and other triangulation-based techniques, this technique has only one view point and therefore it does not suffer from occlusion. Finally, defocus based technique typically shows low to moderate precisions in the range of one part per 400 [3].

1.3.7 Confocal microscopy

The confocal principle dates back to the 1950s and is a very well known and extensively used principle. In conventional epi-illumination microscopy, the whole sample is illuminated. Light from both the in-focus plane and the out-of-focus planes will therefore make it to the detector. This reduces the contrast and hence the quality of the acquired images. If an aperture (typically a pinhole) is placed at the image point as shown in Fig. 1.5, the out-of-focus light is blocked and only light coming from the in-focus plane is imaged. This improves the contrast and the axial resolution at the expense of a reduction in optical throughput. A modest improvement in the lateral resolution is also obtained. Mathematically, this would be equivalent to deconvolution, however, given that this happens before detection, the noise amplification associated with computational deconvolution is avoided.

This technique is exclusively employed in microscopy, where 3D images of an object are acquired by optical sectioning. This implies that confocal microscopy can also reconstruct internal features of an object besides its surface features. This makes it highly versatile for biological applications, hence its extensive usage. In metrology,

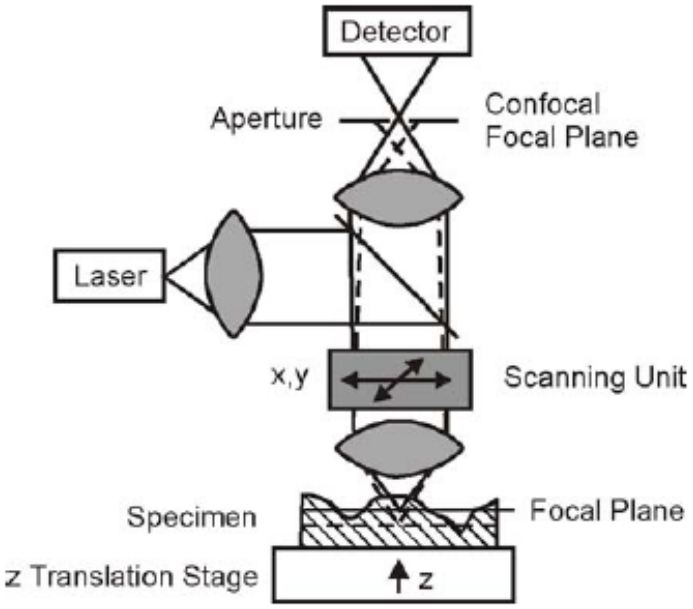


Fig. 1.5 Confocal microscope. Image taken from [10].

confocal microscopy is used to profile surfaces and multi-layered structures such as silicon dies [10]. Surveying an extended sample, however, necessarily entails scanning in all three dimensions of space. The lateral scanning is most commonly accomplished by moving the sample laterally or by means of scanning mirrors. Recently, digital micro-mirror devices have also been used for this purpose. As a result, this technique is suitable mostly to image static samples. Furthermore, the additional scanning hardware required contributes substantially to the cost and the complexity of this technique.

1.3.8 Selective plane illumination microscopy

Selective plane illumination microscopy (SPIM) [17] operates on the same principle as confocal microscopy; that is by limiting of out-of-focus light from reaching the detector. To achieve this goal however, in SPIM, instead of using a pin hole, the illumination is expanded into a thin light sheet using a cylindrical lens and this is then used to illuminate only the in-focus plane. This way, no light is reflected or emitted from the out-of-focus planes. Unlike confocal microscopy, SPIM does not result in a lower optical throughput. Furthermore, it does not require lateral scanning although scanning in the axial dimension is still required. This is SPIM's principal advantage. On the other hand, the lateral resolution is not improved and the axial resolution is defined by the thickness of the light sheet. Also, in SPIM, the illumination and the imaging path have to be separate. In a SPIM microscope therefore, two objectives are required, whereas in a confocal microscope only one. The main application of this technique is in 3D imaging of fluorescent samples such as the imaging of zebrafish hearts [18].

1.3.9 Light field imaging

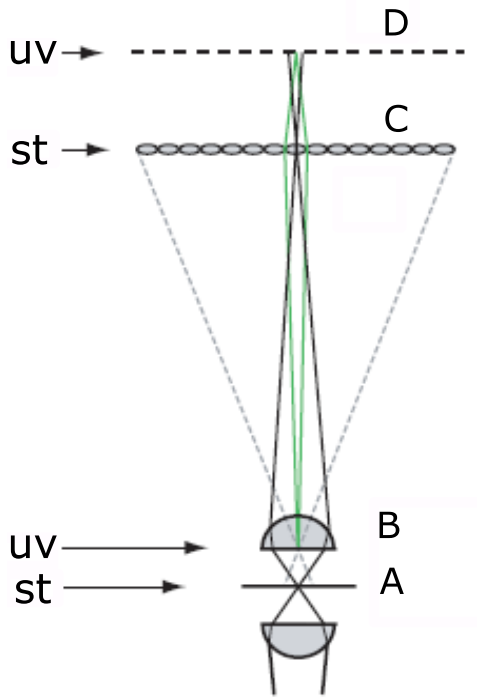


Fig. 1.6 Light field microscope. (A) sensor plane, (B) objective plane, (c) microlens array plane, and (D) detector plane. *uv* and *st* are conjugate planes. Image taken from [19].

Another technique of interest is light field microscopy [19] or photography [20] which is sometimes also classified as a geometric technique (such as triangulation for instance). To implement a light field microscope, all that is required is to place a microlens array in the primary image plane of the microscope as shown in Fig. 1.6. This creates two planes (st and uv) conjugate to the sample plane and the objective as shown. As a result, each microlens corresponds to a different point on the objective and hence captures a different view of the sample. The acquired image (called a light field) therefore consists of a number of small images (called subimages) of the sample from different perspectives. The number of subimages is equal to the number of microlenses in the microlens array. By extracting one pixel from each subimage, different perspectives of the sample can be created. By adding each individual subimage, an orthographic view with a shallow DOF is obtained and finally, by shearing the light field (extracting different pixel positions from each subimage); refocusing can also be performed. Furthermore, all this additional functionality is obtained from a single snapshot, therefore this technique is suitable to image moving samples.

This technique is growing in popularity both in macroscopic and microscopy and has been exploited for the imaging of neural activity in zebrafish brains over an extended DOF [21]. Its principal disadvantage is the drastic reduction in optical resolution incurred by the use of the microlens array. Whereas optical resolution is not critical in macroscopic imaging, in microscopy, this is often paramount. Consequently, the use of

light field imaging in microscopy is restricted to those applications where resolution is not critical.

1.3.10 Computational 3D imaging techniques

Traditionally, the imaging process is performed solely by the optics. In computational imaging (also known as hybrid imaging), the optics are modified (typically by engineering the aperture stop of the system) in such a manner as to produce an intermediate image with some desired properties. The final image is then obtained after post-processing the acquired intermediate image. Part of the imaging process can therefore be considered to have been devolved to computation. The technique proposed here (CKM) falls into this category, therefore these techniques will be discussed in more detail than the others.

One such technique was proposed by Quirin *et. al.* [22] where two different pupil functions (or equivalently PSFs) were used: the double helix PSF (DH-PSF) [23, 24] and the cubic PSF [1].

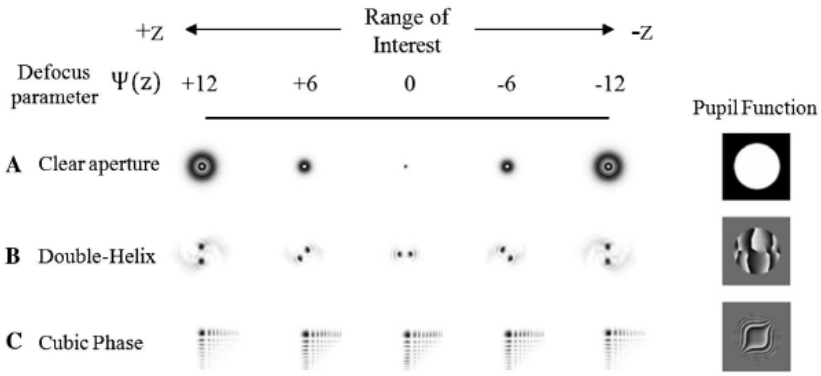


Fig. 1.7 Comparison between double-helix PSF and cubic PSF. Image taken from [22].

These two PSFs are shown in Fig. 1.7 together with the airy disk PSF created by a clear aperture for comparison. As one can see, the DH-PSF which is generated using a phase-only mask obtained from the superposition of Gauss-Laguerre modes presents two bright lobes which rotate with defocus. This makes the DH-PSF suitable for ranging over an extended DOF. On the other hand, the cubic PSF which is generated using a cubic phase mask (CPM) is invariant with defocus. This makes the cubic PSF suitable for the acquisition of the intensity of the scene over an extended DOF. In [22] therefore, two imaging channels are used to acquire two images of the scene. One coded with a CPM and the other with the DH-PSF's mask. The intensity image given by the CPM channel is then used to locally deconvolve the image acquired through the DH-PSF channel thus yielding the DH-PSF with which that region of the scene is convolved. From the obtained DH-PSF, the depth in that region can therefore be estimated by measuring the angle subtended by the two bright lobes. It should be noted, however, that deconvolving by the scene is an ill-posed problem due to absent frequencies in the object's spectrum; therefore, this operation poses some difficulties.

This method was demonstrated in macro imaging where both the intensity image of the scene and a depth map with a range error of $0.4mm$ over an extended DOF were achieved. Its main disadvantage lies in the fact that two different pupil functions are necessary. To implement this, in [22] the use of a spatial light modulator (SLM) is proposed and acquisition was performed in a sequential fashion. If single snapshot acquisition is desired, two separate imaging paths are unavoidable in order to accommodate the two phase masks. This incurs additional complexity of the optics and a significant increase in cost - a phase mask alone costs £2,500 to manufacture at the time of writing. A variation of this technique is proposed in [23] where one image is acquired using the DH-PSF and the other with a clear aperture instead of a CPM. This does reduce the cost because only one phase mask is necessary, however it does not permit exploitation of the extended DOF provided by the DH-PSF. Besides, two separate imaging channels or otherwise an SLM are still required. Yet another variation was presented in [25] where only the DH-PSF channel is used. This solves the problem of complexity, however both the quality of the recovered depth and the intensity image is lower.

The use of amplitude-only masks has also been proposed in the literature for 3D imaging. For example, in [26] an amplitude-only mask was designed in order to engineer the response of the system to defocus. Using this together with a statistical model, the depth map and the intensity image of the scene can be acquired from a single image over an extended DOF. The process however cannot be automated and the recovered depth maps are of low accuracy, therefore it is proposed mainly for image processing applications such as image segmentation. The principal disadvantage of this technique which is common to all systems which make use of an amplitude mask, is the reduction in optical throughput.

Finally, in [27], a deformed binary grating placed in the aperture plane has been used to alter the focal point of the resultant diffraction orders. In this way, each diffraction order corresponds to a different depth in the sample space. Furthermore, it is suggested that the binary grating can be replaced by a phase grating, thus preserving the optical throughout and the resolution. This technique, however, is inherently monochromatic. Furthermore, the large drop in the intensity of higher diffraction orders with respect to the zeroth-order can pose a problem in imaging orders higher than the ± 1. This implies that this technique can image only a few depth levels simultaneously. It should be noted that this technique does not necessitate any post processing, therefore it is not strictly a computational imaging technique.

1.4 Background material

In this section, an overview of WC and the theory behind it is presented. This work is based on WC; consequently an understanding of it is required in order to follow the

material in the subsequent chapters. In addition, important practical considerations and the artefact problem in WC in particular, will also be discussed in this section since the mitigation of these artefacts is a principal objective of this study.

1.4.1 Basic wavefront coding system

WC is a hybrid imaging technique. As explained earlier, this means that part of the imaging process is devolved to computation whereas the other part is devolved to the optics. This particular hybrid imaging technique was introduced by Dowski *et. al.* in 1995 [1] where by means of the stationary phase approximation, the pupil function $f_p(\cdot)$ which renders the optical transfer function (OTF) independent of defocus has been derived. Following the derivation given in [1], this pupil function is given by:

$$f_p(x,y) = P(x,y) \cdot e^{\left[\alpha \cdot \left(x^3+y^3\right)\right]} \tag{1.4}$$

where x, y are the pupil coordinates normalised by the radius of the pupil, $P(x,y)$ is the amplitude of the pupil and α is referred to as the coding strength. Notice that the only difference between this pupil function and the pupil function of a conventional optical system is the addition of the cubic phase term: $\alpha \cdot (x^3 + y^3)$. This implies that all that is required to implement WC is the addition of a phase mask in the aperture plane. This mask is known as the cubic phase mask. An important feature of this mask is the fact that it is phase-only. As a result, the optical throughput is not reduced.

The stationary phase expression of the OTF given in [1] in 1D is:

$$OTF(u,\psi) \approx \left(\frac{\pi}{12 \mid \alpha u \mid}\right)^{1/2} \cdot e^{\left[i\frac{\alpha u^3}{4}\right]} \cdot e^{\left[-i\frac{\psi^2 u}{3\alpha}\right]} \tag{1.5}$$

where ψ is the defocus parameter and u is normalised spatial frequency. As one can see from Eq. 1.5, with the exception of a linear phase term, the OTF is independent of W_{20} as desired. Furthermore, in [1], the author argues that by choosing a large value for the coding strength, this term can be reduced to negligible levels and hence the OTF can be viewed as essentially independent of W_{20}. The 2D OTF can be calculated easily from the 1D expression because the CPM is separable. The in-focus modulation transfer function (MTF) and the corresponding PSF, which will be referred to as the cubic PSF here, are shown in Fig. 1.8. These were numerically computed for an α of 3. From these plots, another important feature of the OTF generated by the CPM besides its insensitivity to defocus can be observed. This is the fact that although the MTF is significantly suppressed, it presents no zeros in the pass band. This implies that aberrating the pupil by a CPM does not result in any loss of detail (*i.e.* resolution is preserved). Furthermore, this property makes the MTF particularly suitable for deconvolution, which is necessary in order to restore contrast.

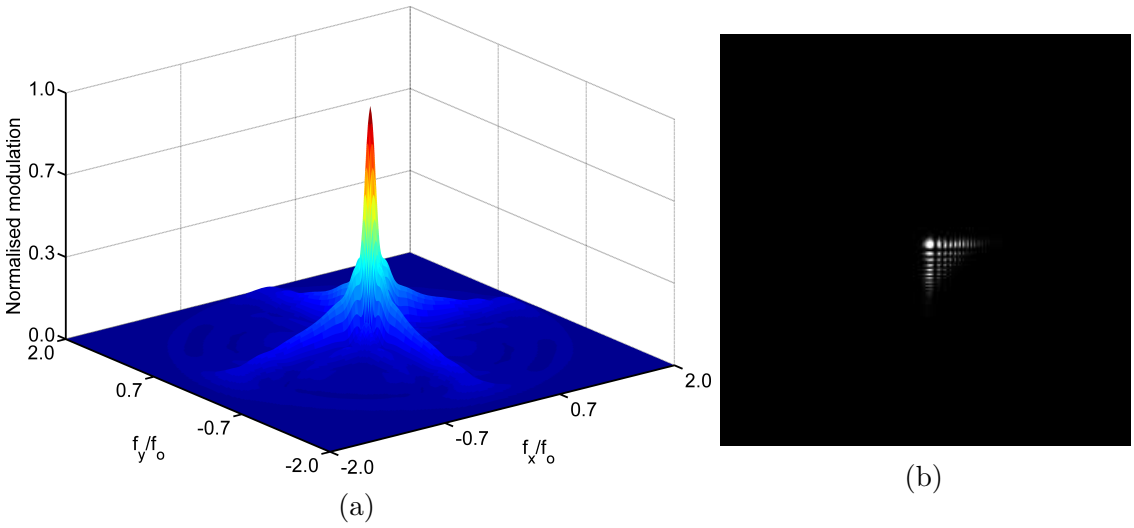


Fig. 1.8 (a) MTF of a wavefront coding system and (b) the cubic PSF. f_x and f_y are the spatial frequencies and f_o is the coherent cut-off frequency.

From a geometric point of view, the operation of the CPM can be intuitively understood by considering the caustic. This is shown in Fig. 1.9 where the CPM was placed in the aperture of a $70mm$ focal-length lens (not shown in the figure for clarity) which was then illuminated using a collimated light beam from the left. A large α value ($\sim 900waves$) was used in order to accentuate the shape of the caustic. Without the CPM, all the light rays which make it through the aperture of the lens will converge to a single point in space at the focal point. On the other hand, with the introduction of the CPM, one notices that the focal point of a given bundle of rays depends on the point on the CPM from which it emanates. For instance, in Fig. 1.9, bundle B which lies at the centre of the CPM focuses at the original focal point of the lens (Z_B) whereas bundles A and C focus at point Z_A to the right and Z_C to the left of the original focal point respectively. This implies that the introduction of the CPM results in a focal point (or focal points) distributed over a range of axial positions rather than at single one. As a result, the DOF is extended.

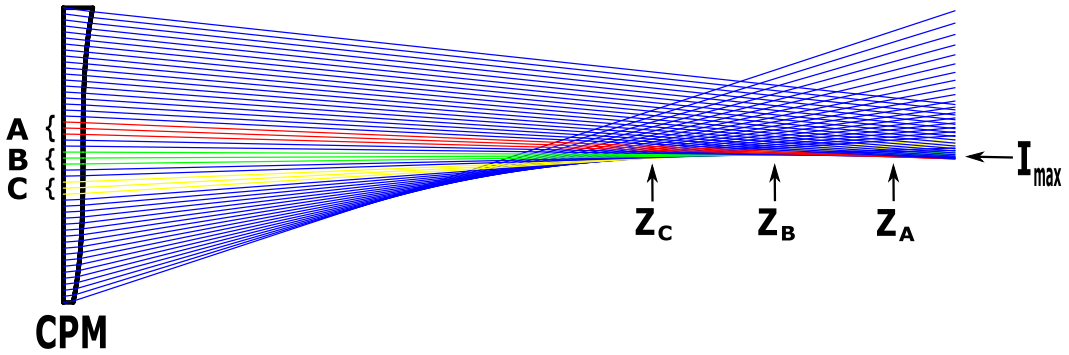


Fig. 1.9 Caustic generated by a CPM. $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ are bundles of rays along the phase mask which are focued at $\boldsymbol{Z_A}, \boldsymbol{Z_B}$ and $\boldsymbol{Z_C}$ respectively; $\boldsymbol{CPM}$ is a cubic phase mask; and, $\boldsymbol{I_{max}}$ is the point where the maximum intensity (peak intensity of the PSF) occurs.

Lastly, the WC imaging process will be described. This is illustrated in Fig. 1.10. As shown, a CPM is fitted in the aperture stop of an optical system - in this case a

finite conjugate single-lens arrangement. Capturing an image of an object will therefore result in a blurred intermediate image because the scene is convolved with the extended cubic PSF. This is the optical part of WC. Once the intermediate image is acquired,

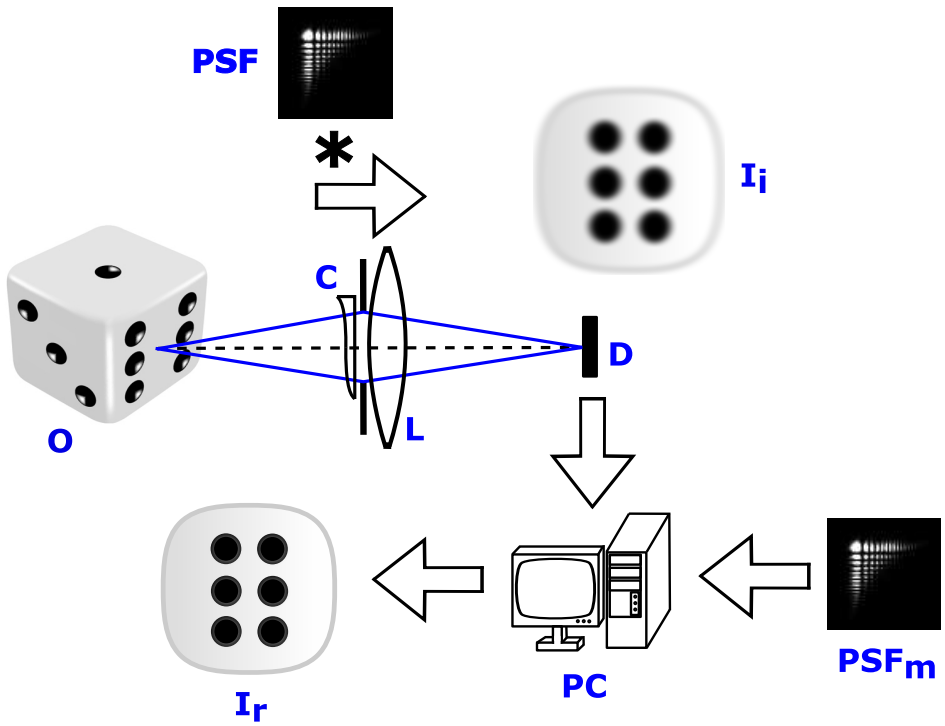


Fig. 1.10 Wavefront coding imaging process. **O**: object, **C**: cubic phase mask, **L**: lens, **D**: detector, $\boldsymbol{I_i}$: intermediate image, $\boldsymbol{PSF}$: point spread function, **PC**: computer, $\boldsymbol{PSF_m}$: pre-acquired (measured) point spread function and $\boldsymbol{I_r}$: recovered image.

deconvolution is necessary in order to remove the blur. This is typically computed using a simple Wiener filter:

$$H_w = \frac{OTF^*}{|\,OTF\,|^2 + k} \tag{1.6}$$

where k is a regularisation parameter which represents the noise-to-signal ratio (NSR). Besides the system itself, this parameter depends also on the scene and therefore it is difficult to known *a priori*. For this reason, it is often chosen empirically and is assumed to be constant, which is tantamount to assuming the noise to be white.

To compute Eq. 1.6, knowledge of the OTF of the system is required. This is typically obtained by acquiring a PSF at some axial position. Given that the cubic PSF is approximately invariant with depth, the exact position where the PSF is measured is not important. In conventional WC therefore, a PSF close to the in-focus position is often used. Once the intermediate image has been deconvolved, a sharp image of the scene is obtained. This will be referred to as the recovered image in this text, whereas the deconvolution process is sometimes referred to as the WC recovery process. Note that since the PSF does not change with defocus, then the intermediate image will also be approximately unchanged with defocus. This implies that a sharp image of the scene can be obtained irrespective of the defocus in the scene. In simpler terms, this implies that the DOF of the optical system has been increased (extended). In fact, WC

is considered to be a technique for extending the DOF of the system without inuring a throughput and resolution penalty. This is in line with the requirements of this work hence the 3D imaging technique proposed in this work was based on WC technology.

1.4.2 Practical considerations in wavefront coding - noise and post-recovery artefacts

Before adopting WC, there are some important practicalities which one should consider carefully. One of the most important considerations is the signal-to-noise ratio (SNR). Clearly, since WC entails aberrating the pupil by the introduction of a CPM, the resultant MTF will be suppressed, as one can easily observe from Fig. 1.8a. Consequently, the SNR of the recovered image given by a WC system will be less than that given by a conventional clear-aperture system close to the focus. On the other hand, if the defocus is significant, the SNR and spectral content of the WC image will be better. As a result, if one considers the SNR over the entire extended DOF provided by a WC system it is found that a WC system performs better than an equivalent conventional clear-aperture system [28].

Another important consideration is the choice of the coding strength α of the CPM. The higher the value of α is, the more suppressed the resultant MTF will be and hence the lower the SNR. A small value is therefore desirable. On the other hand, if a small α is used, the linear term in the OTF (Eq. 1.5) is significant. This implies that if the scene has a variable depth profile within the FOV, the capture image will appear distorted, since parts of the image which lie at different depths will appear shifted laterally by an amount approximately proportional to their depth squared. In WC, therefore, eliminating this distortion necessarily entails sacrificing SNR. This trade-off is particularly important in microscopy because the shallow DOF often results in large variations in defocus across the FOV.

(a)

(b)

(c)

(d)

Fig. 1.11 Artefacts in wavefront coding. (a) scene, (b) coded (intermediate) image, (c) sub-optimal recovery (post-recovery artefacts present) and (d) optimal recovery (no artefacts present).

Finally, the as yet unsolved artefact problem in WC will be considered. This is particularly important in WC and is probably the reason which is preventing full exploitation of this technology. For instance in [29], the authors deemed WC unfit for their application due to this problem. In order to explain the artefact problem, numerically simulations using an α of 3 are shown in Fig. 1.11. Fig. 1.11a shows the original scene and Fig. 1.11b shows the intermediate or coded image which would be acquired using a WC system. The scene was assumed to be flat (single depth) and the defocus was set to 1 wave. If the scene is recovered using the in-focus PSF, as is customary in WC, the recovered image will display artefacts as shown in Fig. 1.11c. By comparing this to the original scene shown in Fig. 1.11a, it is evident that this is far from a faithful reproduction of the scene.

These artefacts occur as a result of the fact that although the cubic PSF exhibits a remarkable resistance (or insensitivity) to defocus, it is not completely unresponsive to it. In fact, in order to eliminate these artefacts, it is necessary to recover the scene with a PSF measured at the same depth (or defocus) as that in the scene itself [2]. This is shown in Fig. 1.11d where the coded image has been deconvolved with a PSF which includes 1 wave of defocus like the scene itself. This implies that if the scene has a varying depth profile (different depths in different regions) as is often the case in practice, then optimal recovery will require a range of PSFs and not a single one as is customary in WC. Notice further that the approximation of the OTF given by Eq. 1.5 does not describe these artefacts. This is due to the fact that these artefacts are caused by modulation of the phase transfer function [30] as their appearance suggests. By taking the stationary phase approximation, therefore, one inherently assumes that this effect is negligible.

To date, most research in WC aimed at solving the discussed issues has focused on alternatives or alteration to the CPM such as the logarithmic mask [31, 32], radially symmetric masks [33] and the generalised cubic mask [34, 35]. These phase masks alleviate the problem, however they typically incur some other penalty. For instance, the generalised cubic mask eliminates the issues caused by the linear term in the OTF, however it has a reduced DOF compared to the CPM and is still an asymmetric phase mask, hence it still gives rise to other forms of distortion. The artefact problem, on the other hand, is largely unexplored. In [36], the use of an average phase transfer function (PTF) measured over a range of axial positions is proposed as a solution. This mitigates the problem, however artefacts can still be observed by eye. Another approach can be found in [37] where a metric proportional to the amount of artefacts present in the recovered image is proposed. By deconvolving the scene by a number of PSFs acquired over a range of depths and minimising this metric, optimal recovery can be performed. This, however, is suitable only for flat scenes due to the artefacts not being localised. One therefore concludes that a satisfactory solution to this problem has yet to be identified.

1.5 Introduction to Complementary Kernel Matching

In this section, an introduction to the solution identified in this study will be presented, in order to set the scene for the investigation conducted in the following chapters of this thesis. Following the discussion above, one concludes that the key to a solution lies in estimating the depth of the sample at every region in the FOV. This is because: (1) ranging is an indispensable feature of any 3D imaging system, therefore the depth requires measuring and (2) if the depth of the sample is known across the FOV, then

each region of the intermediate image can be recovered with the matching PSF; thus, post-recovery artefacts and the distortion caused by the linear phase term in the OTF would be eliminated. This in turn would permit the use of WC to extend the DOF of a system without the penalty of a reduced image quality.

The problem therefore boils down to exclusively measuring the depth of the sample. To do this, in this work it was decided to make use of the linear phase term in the OTF of a WC system. To date, this phase term has been either ignored or otherwise suppressed, mainly by selecting large α values at the expense of SNR as previously discussed. From Eq. 1.5, we notice that regions of the scene at different depth will be imaged laterally shifted by an amount proportional to: $\psi^2/3\alpha$ where ψ is the defocus aberration caused by the depth of the sample. It therefore follows that if this lateral shift can be measured, then, a measure of the depth of the sample in that region would also have been acquired.

Unfortunately, encoding the z dimension (depth) as a lateral shift implies that there is ambiguity between the lateral coordinates of a point in the scene and its axial coordinate. Consequently, additional information is required to solve for the 3D position of any given point in the scene. The easiest way of acquiring this information is to acquire a second image of the scene with a different response to depth or defocus - *i.e.* the relationship between depth in the scene and shift in each of the two acquired images must be different. In this way, corresponding point in each image can be matched and the relative shift between them calculated in a manner analogous to the way disparity is measured in stereo vision. Several ways of accomplishing this goal are proposed and investigated in this work. For the reasons discussed in the following chapter, the best way of doing this turns out to be to acquire two images with an offset in defocus between them. This results in a very simple optical setup and it is also possible to acquire the two images simultaneously on the same sensor hence allowing single-snapshot operation. We refer to this technique as: Complementary kernel matching or CKM for short.

Compared to the other techniques considered in the literature review above, CKM does not require complex optics to implement and is very easy to incorporate in an off-the-shelf system. In fact, it is simpler to implement than stereo because it does not require two detectors placed some distance apart. Like stereo vision, it does not require scanning and therefore is suitable for real-time 3D imaging. However, unlike stereo vision and all other techniques which rely on triangulation, CKM is inherent monocular and therefore it does not suffer from occlusion. This is a unique advantage of this technique. Furthermore, with the exception of light field imaging, none of the techniques considered exhibit an extended DOF and hence none is capable of acquiring in a single-snapshot both an intensity image of the scene and its depth. Light field can do this, however, unlike CKM, this is achieved at the expense of optical resolution which is of major concern in microscopy. As the conducted experiments show, CKM displays moderate accuracy in depth measurements comparable to that achieved by other passive

techniques such as stereo vision and focus-stacking (depth from focus/defocus). Finally, it is also important to note that CKM provides a definitive solution to the long-standing artefacts problem in WC for the first time, thus paving the way for full exploitation of this technology.

1.6 Contributions of this study

Following is a list of the contributions made by this study:

- Proposed a new 3D imaging technique (Complementary Kernel Matching) capable of acquiring both the irradiance of a scene and its depth from a single-snapshot over an extended DOF without incurring a reduction in optical resolution and efficiency. As stated in the Acknowledgements, the conjugate-masks CKM including the recovery algorithm for extended scenes was proposed by Dr. G. Carles. All mathematical analyses including those of the conjugate-masks CKM and the other CKM configurations considered were conceived by the author. All experiments were conducted by the author.
- By means of CKM, the long standing image quality problem (artefacts and distortion) in WC was definitively solved.
- Proposed the use of the Airy transform for the analyses of WC systems. This was used to derive the exact paraxial expression for the cubic PSF in 1D. In principle, these analyses could be extended to any system with a dominant third-order aberration.
- Proposed the use of the cubic PSF for 3D point localisation applications together with the biplane CKM setup. This was conceived by the author independently and before CKM was proposed.
- Using the analyses in (3), it was shown that translating the aperture augments the 3D position information content of the cubic PSF.

1.7 Conclusion

In conclusion, the goal of this work has been defined and is the development of a simple imaging technique which is capable of acquiring both an intensity image of the scene and its depth in a single snapshot without incurring a reduction in optical throughput and/or lateral resolution. To this end, a novel technique (CKM) based on WC has been proposed and developed. From the conducted literature review, one concludes that CKM has several advantages over the competition. These include: single-snapshot operation; simplicity and relatively low cost; the ability to acquire both the depth and

the intensity of the scene over an extended DOF and its inherent immunity to occlusion which effects triangulation-based 3D imaging techniques. Additionally, CKM also solves the long-standing image quality problem which has plagued WC systems to date. In the following sections, the theory related to this technique is developed, several ways of implementing it are investigated and ultimately, the technique was successfully tested experimentally, thus fulfilling the objectives set for this study.

Chapter 2

Theory

In this chapter, the cubic PSF is modelled in space under the paraxial approximation only, in order to investigate its properties; in particular, whether or not there are secondary effects affecting its translation in response to defocus. This was necessary because to date, only approximate expressions for the cubic PSF are available in the literature. Following this investigation, the mathematical formalism of the CKM technique is presented and various ways of generating disparity are proposed. These are then analysed using various tools and compared in order to establish the most promising. Next, various techniques of measuring disparity are discussed and the ones used later during the experimental phase formalised. Finally, the CKM recovery algorithms for extended scenes and spare point-scenes are presented in detail.

2.1 Space model of the cubic PSF

To date, most of the analysis of WC systems has been performed using numerical simulations and/or approximate analysis. For instance, in [30], the OTF was modelled using a geometric approach from which inferences about the artefacts and the cut-off frequency of the system were made. The stationary phase method originally proposed by Dowski in [1] has been used to approximate the OTF in [38] and the PSF in [39]. A similar method was used to approximate the MTF in [40]. To the best of my knowledge, the only exact analysis in the literature was conducted by Somayaji *et. al.* [41] where the exact expression of the OTF has been derived from first principle. This expression gives a clear description of the artefacts; however, it does not reveal any additional information regarding the shift compared to the seminal 1995 paper by Dowski.

Given that this study is concerned with a secondary effect (WC artefacts) which is not revealed by the stationary phase approximation as discussed in Section 1.4.1, and that the proposed solution (CKM) relies on a largely unexplored phenomenon (the translation of the PSF with defocus), it was deemed appropriate to conduct exact analysis and modelling of the cubic PSF in space. The analysis were conducted in the

space domain because CKM relies on the spatial translation of the PSF to infer depth and hence to mitigate the post-recovery artefacts. Furthermore, the use of the Airy transform [42, 43] - a recently defined integral transform with various properties similar to those of the Fourier transform - is introduced as a mathematical tool for the analysis of WC system for the first time. As discussed below, this results in a systematic method of analysing WC systems, thus facilitating the process. Finally, in this section it is shown for the first time that the shift generated by the CPM in response to defocus is decoupled from the morphology of the PSF; hence, in principle, it is possible to increase the responsivity (shift per wave of defocus) without major loss of frequency content, or equivalently, with a smaller reduction in the cut-off frequency. Note that the cut-off frequency of a WC system is solely defined by the size of the aperture under ideal conditions (SNR $= \infty$). In practice, however, the SNR is finite due to unavoidable noise sources such as shot noise and quantization noise. Furthermore, frequencies which lie below the noise floor are hard to recover especially because deconvolution amplifies the noise in the recovered image. Algorithms which are capable of recovering some of these frequencies exist however the fidelity of the recovered spectrum still depends on the SNR. In this work, a common definition of the cut-off frequency was adopted which states that the cut-off frequency is the highest frequency in the normalised MTF which is greater than or equal to the noise floor.

The 1D pupil function of a WC system using a CPM as shown in Fig. 2.1 can be expressed as below where the curvature of the wavefront due to defocus has been included:

$$f_p\left(x_p\right) = \begin{cases} 1.0 \times e^{\left[i2\pi\left(Wx_p^2+\alpha x_p^3\right)\right]}, & \text{if } |x_p| < p \\ 0.5 \times e^{\left[i2\pi\left(Wx_p^2+\alpha x_p^3\right)\right]}, & \text{if } |x_p| = p \\ 0, & \text{if } |x_p| > p \end{cases} \tag{2.1}$$

x_p denotes the pupil coordinate, p denotes the half-width of the pupil, $\alpha = {}^{\alpha_o}/_{p^3}$ $[\mu m^{-3}]$ is the cubic aberration introduced by the CPM, where α_o $[{}^{\mu m}/_{\mu m}$ or $Waves]$ is the peak cubic aberration at the edge of the pupil expressed as OPD in multiples of the wavelength; $W = {}^{W_{20}}/_{p^2}$ $[\mu m^{-2}]$ is the defocus aberration, where W_{20} $[Waves]$ is the Seidel defocus coefficient.

The amplitude PSF (APSF) is therefore given by:

$$APSF\left(x_i\right) = \mathscr{F}\left\{f_p\left(x_p\right)\right\} = \int_{-\infty}^{\infty} \left\{ \text{rect}\left(\frac{x_p}{2p}\right) \cdot e^{\left[i2\pi\left(Wx_p^2+\alpha x_p^3\right)\right]} \cdot e^{\left[-i2\pi x_i x_p\right]} \right\} dx_p \tag{2.2}$$

where $x_i = {}^{x}/_{\lambda z_i}$ $[\mu m^{-1}]$ is the image-side coordinate; x is the space coordinate on the image side and z_i is the image distance. $\mathscr{F}\{\cdot\}$ denotes the Fourier transform operator

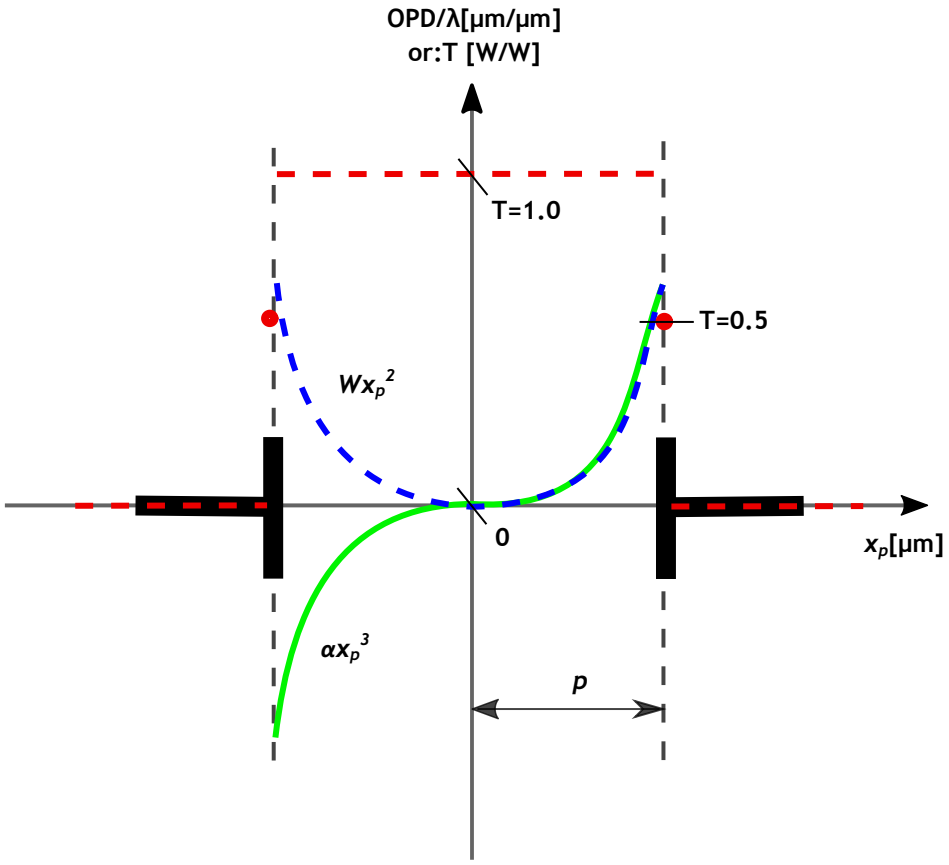


Fig. 2.1 Pupil function of a wavefront coding system in 1D showing the optical path difference (OPD) expressed as multiples of the wavelength (λ) introduced by the CPM, αx_p^3, (green curve), the OPD introduced by the curvature of the wavefront due to defocus, Wx_p^2, (blue curve) and the transmittance, T, (red curve). p is half the pupil width.

and rect $(\cdot)$ is the Rectangular or Pulse function. Defining:

$$\zeta = (6\alpha\pi)^{1/3} \tag{2.3}$$

$$x'_p = \zeta x_p \tag{2.4}$$

then, Eq. 2.2 can be re-rewritten as:

$$APSF\,(x_i) = \int_{-\infty}^{\infty} \left\{ \text{rect}\left(\frac{x'_p}{2\zeta p}\right) \cdot e^{\left[i\left(\frac{x'^3_p}{3}+\left(\frac{W\zeta}{3\alpha}\right)(x'_p)^2+\left(\frac{-x_i\zeta^2}{3\alpha}\right)x'_p\right)\right]} \right\} \frac{dx'_p}{\zeta} \tag{2.5}$$

Letting:

$$a = \frac{W\zeta}{3\alpha} \tag{2.6}$$

$$x'_i = \frac{-x_i\zeta^2}{3\alpha} \tag{2.7}$$

$$x''_p = x'_p + a \tag{2.8}$$

Eq. 2.5 can be re-written in a more revealing form as shown below:

$$APSF\left(x_i\right) = \frac{e^{\left[i\left(\frac{2a^3}{3} - ax_i'\right)\right]}}{\zeta} \int_{-\infty}^{\infty} \left\{ \left[\text{rect}\left(\frac{x_p'' - a}{2\zeta p} \right) \cdot e^{\left[\frac{i}{3}x_p''^3\right]} \right] \cdot e^{\left[ix_p''\left(x_i' - a^2\right)\right]} \right\} dx_p'' \quad (2.9)$$

where Eq. 2.7 was not applied to the left hand side of Eq. 2.5 as this will be reversed later. Here we note that Eq. 2.9 is an inverse Fourier transform integral; therefore, letting:

$$H\left(x_p''\right) = \text{rect}\left(\frac{x_p'' - a}{2\zeta p} \right) \cdot e^{\left[\frac{i}{3}x_p''^3\right]} \quad (2.10)$$

$$h\left(r\right) = \mathscr{F}^{-1}\left\{H\left(x_p''\right)\right\}\left(r\right) \quad (2.11)$$

where r is a dummy variable, then, by Eq. 2.9, the amplitude PSF can be expressed as:

$$APSF\left(x_i\right) = \frac{e^{\left[i\left(\frac{2a^3}{3} - ax_i'\right)\right]}}{\zeta} \cdot h\left(x_i' - a^2\right) \quad (2.12)$$

The intensity PSF, which is what we are interested in, is thus given by Eq. 2.13 below where Eq. 2.6 and Eq. 2.7 have been substituted back in.

$$PSF\left(x_i\right) = \frac{1}{\zeta^2} \cdot \left| h\left(\left[\frac{-x_i\zeta^2}{3\alpha}\right] - \left[\frac{W\zeta}{3\alpha}\right]^2 \right) \right|^2 \quad (2.13)$$

From Eq. 2.13, one can notice that irrespective of what function $h\left(\cdot\right)$ is, the PSF will experience a shift proportional to the square of the defocus as shown by the independent variable of function $h\left(\cdot\right)$. Equating this to zero and substituting for peak aberrations at the edge of the pupil yields:

$$\left[\frac{-x_i\zeta^2}{3\alpha}\right] - \left[\frac{W\zeta}{3\alpha}\right]^2 = 0 \quad (2.14)$$

$$\implies -x_i = \frac{-x}{\lambda z_i} = \frac{W^2}{3\alpha} = \frac{{W_{20}}^2}{3\alpha_o p} \quad (2.15)$$

The spatial shift, δx, experienced by the PSF is therefore given by the equation:

$$\delta x = \frac{-\lambda z_i {W_{20}}^2}{3\alpha_o p} = \frac{-2\lambda f_{\#} {W_{20}}^2}{3\alpha_o} = \frac{-2{W_{20}}^2}{3\alpha_o f_o} \quad (2.16)$$

where $f_{\#}$ denotes the f-number of the system and f_o the incoherent cut-off frequency. The negative sign in Eq. 2.16 indicates that the shift will occur in the negative direction as shown in Fig. 2.2. Assuming a separable (square) aperture, the total shift, δS, as shown in Fig. 2.2 is given by:

$$\delta S = \frac{\sqrt{2}\lambda z_i {W_{20}}^2}{3\alpha_o p} = \frac{2\sqrt{2}\lambda f_{\#} {W_{20}}^2}{3\alpha_o} = \frac{2\sqrt{2}{W_{20}}^2}{3\alpha_o f_o} \quad (2.17)$$

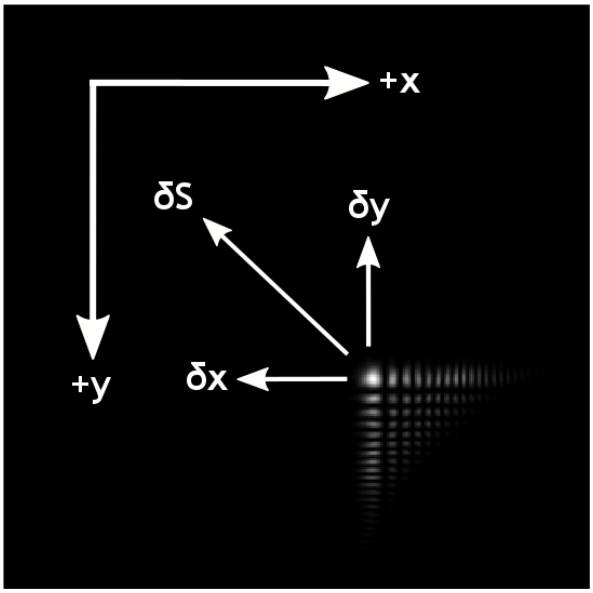


Fig. 2.2 Translation of the point spread function with defocus. $\delta S = \sqrt{\delta x^2 + \delta y^2}$

which is in agreement with the original derivation given by Dowski in [1] using the stationary phase approximation.

The final step in evaluating the PSF is to evaluate the function $h(\cdot)$. For generality, let a generic aperture function be defined as: $f_{apt}\left(\frac{x_p - t}{s}\right)$ where $s > 0$ and t are arbitrary scaling and translation constants respectively. For instance, in the case of a square-shaped aperture as considered above, the aperture function is given by:

$$f_{apt}\left(\frac{x_p - t}{s}\right)\Big|_{\substack{s=2p \\ t=0}} = \text{rect}\left(\frac{x_p}{2p}\right) \tag{2.18}$$

then, from Eq. 2.10 and Eq. 2.11, we have:

$$h(r) = \mathscr{F}^{-1}\left\{ f_{apt}\left(\frac{\left(\frac{x_p'' - a}{\zeta}\right) - t}{s}\right) \cdot e^{\left[\frac{i}{3}x_p''^3\right]} \right\}(r) \tag{2.19}$$

$$\implies h(r) = \mathscr{F}^{-1}\left\{ f_{apt}\left(\frac{x_p'' - a - \zeta t}{s\zeta}\right)\right\}(r) * \mathscr{F}^{-1}\left\{ e^{\left[\frac{i}{3}x_p''^3\right]}\right\}(r) \tag{2.20}$$

where $*$ denotes the convolution product. Letting also:

$$F_{apt}(r) = \mathscr{F}^{-1}\left\{ f_{apt}\left(\frac{x_p - t}{s}\right)\Big|_{\substack{s=1 \\ t=0}}\right\}(r) = \mathscr{F}^{-1}\left\{f_{apt}(x_p)\right\}(r) \tag{2.21}$$

then, the function $h(\cdot)$ is given by:

$$h(r) = |s\zeta|\left\{\left[F_{apt}(s\zeta r) \cdot e^{[i(a+\zeta t)r]}\right] * Ai(r)\right\} \tag{2.22}$$

where $Ai(\cdot)$ denotes the homogeneous Airy function of the first kind. Evaluating Eq. 2.22 from first principles is quite mathematically involved, however, we note that Eq. 2.22 fits the definition of a relatively recently defined integral transform known as

the Airy transform which, in a manner analogous to the Fourier transform, is denoted as $\mathscr{A}_a\{\cdot\}$ for the forward and $\mathscr{A}_a^{-1}\{\cdot\}$ for the inverse transform [42, 44]. The subscript $a \in \mathbb{R}^+$ is a constant used to scale the basis of this integral transform which are given by:

$$\omega_a(r) = \frac{1}{|a|} Ai\left(\frac{r}{a}\right) \tag{2.23}$$

The Airy transform of a function $f(\cdot)$ whose Fourier transform is $F(\cdot)$ is given in [44] as:

$$\gamma(r) = \mathscr{A}_a\{f(r)\} = f(r) * \omega_a(r) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{e^{i\left[\left(\frac{a^3\phi^3}{3}+\phi r\right)\right]} \cdot F(\phi)\right\} d\phi \tag{2.24}$$

We can therefore write Eq. 2.22 as:

$$h(r) = |s\zeta|\,\mathscr{A}_1\left\{F_{apt}(s\zeta q)\cdot e^{[i(a+\zeta t)q]}\right\}(r) \tag{2.25}$$

where q and r are dummy variables. By applying the scaling property of the Airy transform, we can also write Eq. 2.25 as:

$$h(r) = |s\zeta|\,\mathscr{A}_{s\zeta}\left\{F_{apt}(q)\cdot e^{\left[i\left(\frac{a+\zeta t}{s\zeta}\right)q\right]}\right\}(s\zeta r) \tag{2.26}$$

At this point, we make an important observation. From Eq. 2.13, it is clear that the morphology of the intensity PSF is exclusively dependent on the function $h(\cdot)$. Furthermore, from Eq. 2.25 (or Eq. 2.26), the only effect the defocus has on this function is given by the phase term:

$$e^{[i(a+\zeta t)q]} = e^{\left[i\left(\frac{W\zeta+3\alpha\zeta t}{3\alpha}\right)q\right]} \tag{2.27}$$

From the equation above, we notice further that if the aperture function is shifted by an amount: $t = \frac{-pW_{20}}{3\alpha_o}$ for a 1D aperture (or under the assumption of a separable $f_{apt}(\cdot)$, by $t = \frac{-\sqrt{2}pW_{20}}{3\alpha_o}$ along the 45^0 degree direction with respect to the CPM axes for a 2D aperture), then this phase term reduces to one. This means that the only effect the defocus has on the shape of the PSF is describable by a shift of the aperture function. This is consistent with what is reported in [39] where the highly idealised unbounded aperture case (*i.e.* ignoring the aperture function) was analysed through a combination of approximations and ray tracing. On the other hand, the fact that the lateral shift experienced by the PSF in response to defocus, which is given by Eq. 2.17, is not effected by shifting the aperture function, was not considered. This implies that the shape of the PSF - and therefore the cut-off frequency of the system - is decoupled from the responsivity of the PSF which we define as:

$$\frac{\partial(\delta S)}{\partial W_{20}} = \frac{-2\sqrt{2}\lambda z_i W_{20}}{3\alpha_o p} \tag{2.28}$$

This is of particular relevance for CKM because in CKM maximising the responsivity is of crucial importance, since this is directly linked to the precision with which the depth can be inferred and thus also the extent or effectiveness with which WC artefacts can be mitigated. Increasing the responsivity can only be achieved by operating the system at a large W_{20} as shown by Eq. 2.28. This, however, has the highly undesirable effect of reducing the cut-off frequency (potentially by a substantial amount especially for off-axes frequencies) and thus the quality of the recovered images. The fact that shifting the aperture can be used to compensate for the reduction in the cut-off frequency without effecting the translation of the PSF eliminates this compromise; thus, it allows the responsivity to be increased without lowering the image quality. Of course, shifting the aperture is not always feasible in practice without stopping down the optics. For instance, if we consider a microscope objective, then, shifting the aperture might be difficult. On the other hand, if we consider a zoom lens, then this scenario is more practical. Another practical difficulty in utilising the shift of the aperture to enhance the performance of CKM lies in the fact that having an off-axes aperture would increase field dependence for wide FOV systems, as well as introduce distortion. The use of barrel distortion was identified in [45] as a feasible solution to this problem. This would require the use of a modified phase mask, which is not foreseen to be problematic from a practical point of view. The alternative would be to simply restrict the FOV, which in applications such as inspection might not constitute a problem whatsoever.

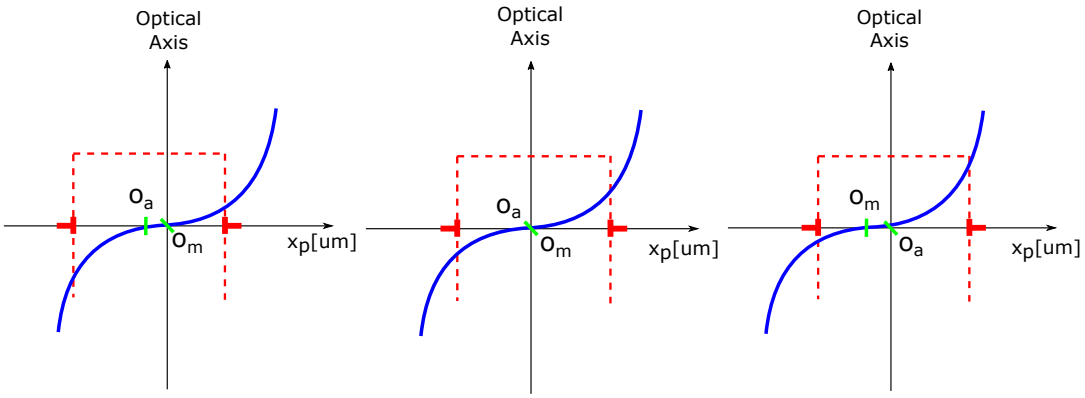


Fig. 2.3 Translation of the pupil with respect to the optical axis: (a) centred CPM and shifted aperture; (b) both CPM and the aperture centred (reference); and, (c) centred aperture and shifted pupil. O_m and O_a denote the origin of the mask and the aperture respectively.

As a final note, it is important to distinguish between shifting the aperture of a system and shifting the CPM itself [46], as depicted in Fig. 2.3. Fig. 2.3a shows the aperture shifted with respect to the optical axis of the optical system, whereas the CPM is centred. This is the case discussed above. On the other hand, the reverse situation is depicted in Fig. 2.3c. This case is very easily analysed by considering the fact that a shifted CPM will result in a pupil function of:

$$f_{pup}\left(x_p\right) = f_{apt}\left(x_p\right) \cdot e^{\left[i2\pi\alpha\left(x_p - t\right)^3\right]} = f_{apt}\left(x_p\right) \cdot e^{\left[i2\pi\alpha\left(-t^3 + 3t^2 x_p - 3t{x_p}^2 + {x_p}^3\right)\right]} \tag{2.29}$$

Clearly, from Eq. 2.29, the only effect shifting the CPM has, besides introducing piston $\left(e^{\left[-i2\pi\alpha t^3\right]}\right)$ and tilt $\left(e^{\left[i6\pi\alpha t^2 x_p\right]}\right)$, is that of introducing an offset in defocus given by the term: $e^{\left[i6\pi\alpha t {x_p}^2\right]}$. This defocus term acts merely like a defocus offset in the specimen plane and therefore it will affect both the morphology of the PSF (*i.e.* cut-off) as well as the translation of the PSF in response to defocus. In conclusion, the decoupling between the translation (and hence the translation rate) of the PSF with defocus and its morphology cannot be achieved by decentring the CPM instead of the aperture with respect to the optical axis of the system.

To conclude this section, we can express the PSF as follows by substituting Eq. 2.25 into Eq. 2.13.

$$PSF\left(x\right) = \frac{\left|s\zeta\right|^2}{\zeta^2} \cdot \left| \mathscr{A}_1 \left\{ F_{apt}\left(s\zeta q\right) \cdot e^{\left[i(a+\zeta t)q\right]} \right\} \left(\left[\frac{-x\zeta^2}{3\lambda z_i \alpha}\right] - \left[\frac{W\zeta}{3\alpha}\right]^2 \right) \right|^2 \tag{2.30}$$

For a rectangular aperture function, which is the main concern of this study, we have:

$$F_{apt}\left(r\right) = 2\frac{sin\left(\frac{r}{2}\right)}{r} = \text{sinc}\left(\frac{r}{2}\right) \tag{2.31}$$

Substituting in Eq. 2.30 and setting $s = 2p$ gives:

$$PSF\left(x\right) = 4p^2 \cdot \left| \mathscr{A}_1 \left\{ \text{sinc}\left(p\zeta q\right) \cdot e^{\left[i(a+\zeta t)q\right]} \right\} \left(\left[\frac{-x\zeta^2}{3\lambda z_i \alpha}\right] - \left[\frac{W\zeta}{3\alpha}\right]^2 \right) \right|^2 \tag{2.32}$$

which by Euler's formula, can also be written as:

$$PSF\left(x\right) = \left(\frac{1}{\zeta^2}\right) \cdot \left| \mathscr{A}_1 \left\{ \left(\frac{e^{\left[i(p\zeta+a+\zeta t)q\right]} - e^{\left[-i(p\zeta-a-\zeta t)q\right]}}{q} \right) \right\} \left(\left[\frac{-x\zeta^2}{3\lambda z_i \alpha}\right] - \left[\frac{W\zeta}{3\alpha}\right]^2 \right) \right|^2 \tag{2.33}$$

The last step is therefore to evaluate the Airy Transform in Eq. 2.33. To do this, we use the following property of the Airy transform [44].

$$\mathscr{A}_\alpha \left\{ qf\left(q\right) \right\}\left(r\right) = r\varphi\left(r\right) - \alpha^3 \varphi''\left(r\right) \tag{2.34}$$

$$\text{where } \varphi\left(r\right) = \mathscr{A}_\alpha \left\{ f\left(q\right) \right\}\left(r\right) \tag{2.35}$$

Defining:

$$C_p = p\zeta + a + \zeta t \tag{2.36}$$

$$C_m = p\zeta - a - \zeta t \tag{2.37}$$

$$\varphi\left(r\right) = \mathscr{A}_1 \left\{ \left(\frac{e^{\left[iC_p q\right]} - e^{\left[-iC_m q\right]}}{q} \right) \right\}\left(r\right) \tag{2.38}$$

and given that:

$$\mathscr{A}_1\left\{\left(e^{[iC_pq]}-e^{[-iC_mq]}\right)\right\}(r)=e^{\left[i\left(C_pr+\frac{C_p{}^3}{3}\right)\right]}-e^{\left[-i\left(C_mr+\frac{C_m{}^3}{3}\right)\right]} \tag{2.39}$$

then, the necessary Airy transform, $\varphi(\cdot)$, is given by the solution to the following non-homogeneous, second order differential equation with variable coefficients:

$$r\varphi(r)-\varphi''(r)=e^{\left[i\left(C_pr+\frac{C_p{}^3}{3}\right)\right]}-e^{\left[-i\left(C_mr+\frac{C_m{}^3}{3}\right)\right]} \tag{2.40}$$

To solve Eq. 2.40, we note that the homogeneous equation, $r\varphi(r)-\varphi''(r)=0$, is the well known Airy equation; therefore, its solutions are: $Ai(r)$ and $Bi(r)$. The Wronksian of these functions is also well known and is: $W\{Ai(r),Bi(r)\}=\frac{1}{\pi}$. The particular solution can therefore be found by the method of variation of parameters and is given by:

$$\begin{aligned}\varphi(r)=\;& C_1Ai(r)+C_2Bi(r)+..\\ &..-\pi Ai(r)\int_1^r\left\{Bi(t)\left(e^{\left[-i\left(C_mt+\frac{C_m{}^3}{3}\right)\right]}-e^{\left[i\left(C_pt+\frac{C_p{}^3}{3}\right)\right]}\right)\right\}dt+..\\ &..+\pi Bi(r)\int_1^r\left\{Ai(t)\left(e^{\left[-i\left(C_mt+\frac{C_m{}^3}{3}\right)\right]}-e^{\left[i\left(C_pt+\frac{C_p{}^3}{3}\right)\right]}\right)\right\}dt\end{aligned} \tag{2.41}$$

where C_1 and C_2 are constants which depend on the initial conditions. For instance, since: $\lim_{x\to\infty}\{PSF(x)\}=0$, in a damped decaying fashion, then, we can consider the initial conditions: $\varphi(-\infty)=0$ and $\varphi\prime(-\infty)=0$. For these initial conditions, C_1 and C_2 are given by the following equations:

$$\begin{aligned}C_1&=\pi\int_{-\infty}^1\left\{Bi(t)\left(e^{\left[-i\left(C_mt+\frac{C_m{}^3}{3}\right)\right]}-e^{\left[i\left(C_pt+\frac{C_p{}^3}{3}\right)\right]}\right)\right\}dt\\ C_2&=-\pi\int_{-\infty}^1\left\{Ai(t)\left(e^{\left[-i\left(C_mt+\frac{C_m{}^3}{3}\right)\right]}-e^{\left[i\left(C_pt+\frac{C_p{}^3}{3}\right)\right]}\right)\right\}dt\end{aligned}$$

Finally, substituting Eq. 2.41 into Eq. 2.33 gives the exact analytical expression for the intensity PSF of a WC system using a CPM and a rectangular aperture in 1D. This is given by:

$$PSF(x)=\left(\frac{1}{\zeta^2}\right)\cdot\left|\varphi\left(\left[\frac{-x\zeta^2}{3\lambda z_i\alpha}\right]-\left[\frac{W\zeta}{3\alpha}\right]^2\right)\right|^2 \tag{2.42}$$

Finally, it is worth mentioning that the PSF can also be expressed as a series of Airy polynomials [44] which are a special case of Gould-Hopper polynomials of the first kind [47–49]. This expression is still not very tractable in itself; however, as discussed above, the use of the Airy transform for the analyses of WC systems reveals important information.

2.2 Verification of the PSF model

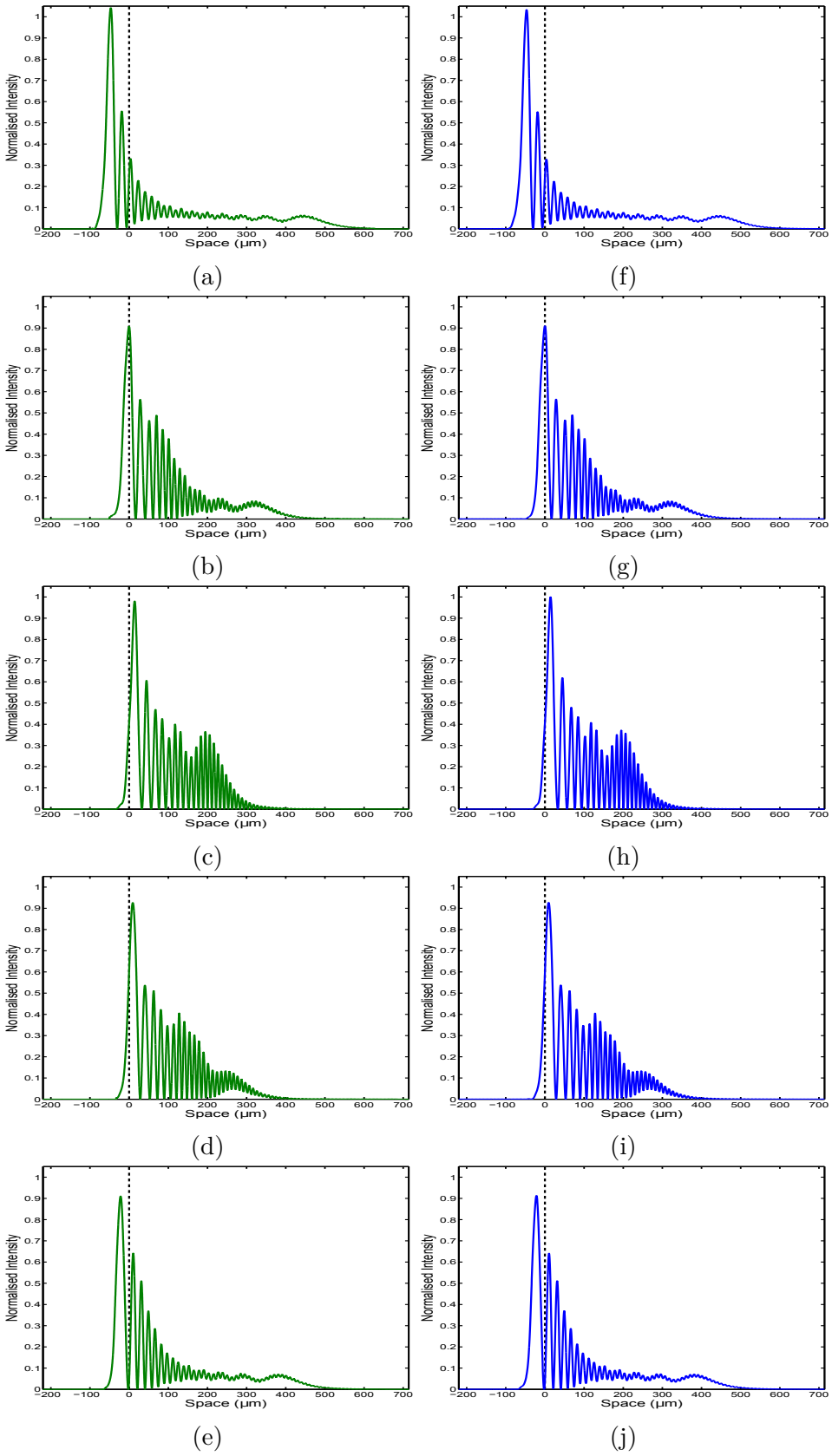


Fig. 2.4 Plots of the space model of the PSF compared to the Fourier transform PSF. (a) - (e) PSF given by Eq. 2.42 and (f) - (j) PSF given by the Fourier transform for different defocus values. (a) and (f) for a defocus of −7.00 waves; (b) and (g) for a defocus of −3.50 waves; (c) and (h) for a defocus of 0 waves; (d) and (i) for a defocus of 1.75 waves; and, (e) and (j) for a defocus of 5.25 waves.

The scope of this section is to verify the space model of the PSF and the deductions made from it in Section 2.1. The simplest way of doing this is perhaps to compare the PSF given by Eq. 2.42 to numerically computed PSFs using a fast Fourier transform (FFT) algorithm. This is shown in Fig. 2.4 for an $\alpha_o = 5$ waves and a $1.4\text{NA}, 40\times$ system. Each row in the figure corresponds to a defocus of: $W_{20} = -7.00, -3.50, 0, 1.75$ and 5.25 waves respectively from top to bottom. A tube length of 200mm was assumed, as is common for various commercial microscopes. The initial conditions used to evaluate Eq. 2.42 were: $\varphi(-147) = 0$ and $\varphi'(-147) = 0$. The value -147 was selected arbitrarily in order to stabilise the solution where, given that the solution is highly oscillatory, numerical instability can be problematic. As previously mentioned, however, if numerical considerations are ignored, any arbitrary large value should yield the correct solution since the intensity of the PSF tends to zero at $\pm\infty$. The maximum RMS error between the numerically computed PSFs and the PSFs given by Eq. 2.42 was 1% and this happens at a $W_{20} = 0$ where the PSF exhibits the steepest slopes. This fact suggests that this error can be attributed to sampling and other numerical phenomena. Furthermore, the simulations predict a shift of: $-1.269{W_{20}}^2\ \mu m$ whereas Eq. 2.42 predicts a shift of: $-1.270{W_{20}}^2\ \mu m$ demonstrating further agreement between the two.

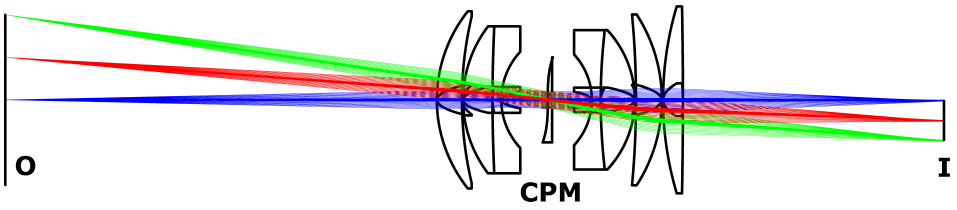


Fig. 2.5 Double Gauss lens setup simulated to investigate the effect of displacing the aperture stop on the cubic PSF. **O**: Object plane, **CPM**: cubic phase mask, and, **I**: image plane. The stop is in the same plane as the front surface of the CPM.

To verify the inferences regarding the pupil shift made in Section 2.1, it was deemed appropriate to consider a more practical scenario rather than mere numerical simulations based on the FFT. This is due to the fact that, as previously mentioned, displacing the aperture can potentially introduce significant aberration, which will inevitably affect the cut-off of the cubic PSF. To achieve this, a double Gauss lens with a circular aperture $4mm$ in diameter and a focal ratio of 8.56 was simulated using an optical design package (*Zemax-EE*) which takes into consideration the aberrations introduced by displacing the aperture stop. The aperture stop was placed in the same plane as the front surface of the CPM. A CPM with an α of ~ 8.3 waves was simulated by encoding the appropriate sag (height of an optical surface with respect to a flat datum) on a BK7 (refractive

index: ~ 1.5168 at $587nm$) substrate. The simulated setup is shown in Fig. 2.5 where different colours denote different fields. The wavelength of light was taken to be $587nm$.

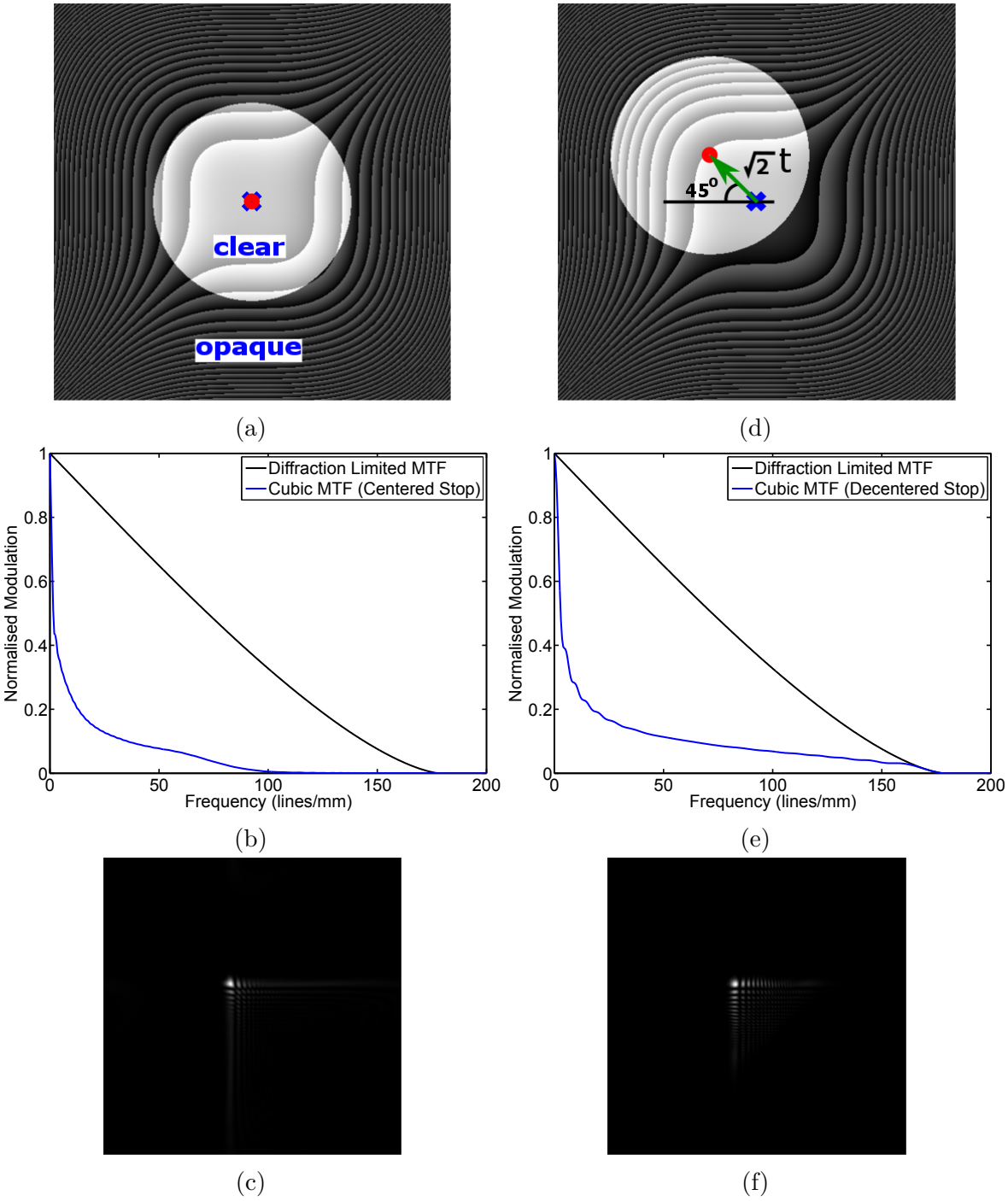


Fig. 2.6 MTF and PSF for centred and decentred aperture stop under ~ -12.5 waves of defocus; $\alpha = \sim 8.3$ waves. (a) centred aperture stop; (b) on-axis defocused MTF with a centred aperture stop; (c) PSF corresponding to the MTF in (b); (d) decentred aperture stop; (e) on-axis defocused MTF with a decentred aperture stop showing that the effect of defocus is nullified; (f) PSF corresponding to the MTF in (e).

To verify that displacing the aperture compensates for defocus by restoring the cut-off frequency to its in-focus value, the object was displaced by $-120mm$ (negative was defined as away from the front surface of the lens) away from the in-focus position. This corresponds to ~ -12.5 waves of defocus. The on-axis MTF and the PSF obtained

at this high level of defocus are shown in Fig. 2.6b and Fig. 2.6c respectively, where, as one can observe, the cut-off frequency is reduced to $\sim 50\%$ of the in-focus cut-off frequency. From the analysis presented in Section 2.1, to compensate for this defocus, the pupil should be displaced by: $t = \frac{-pW_{20}}{3\alpha_o} = +1.00mm$ along the vertical and the horizontal axes of the CPM as shown in Fig. 2.6d with respect to the centred aperture (Fig. 2.6a). This was simulated in *Zemax* and the resultant MTF and corresponding PSF are shown in Fig. 2.6e and Fig. 2.6f respectively. As one can observe, the displaced pupil restores the cut-off frequency to its in-focus value; thus, despite the fact that there is a large defocus offset of 12.5 waves, no loss of detail results. This confirms the hypothesis discussed in Section 2.1.

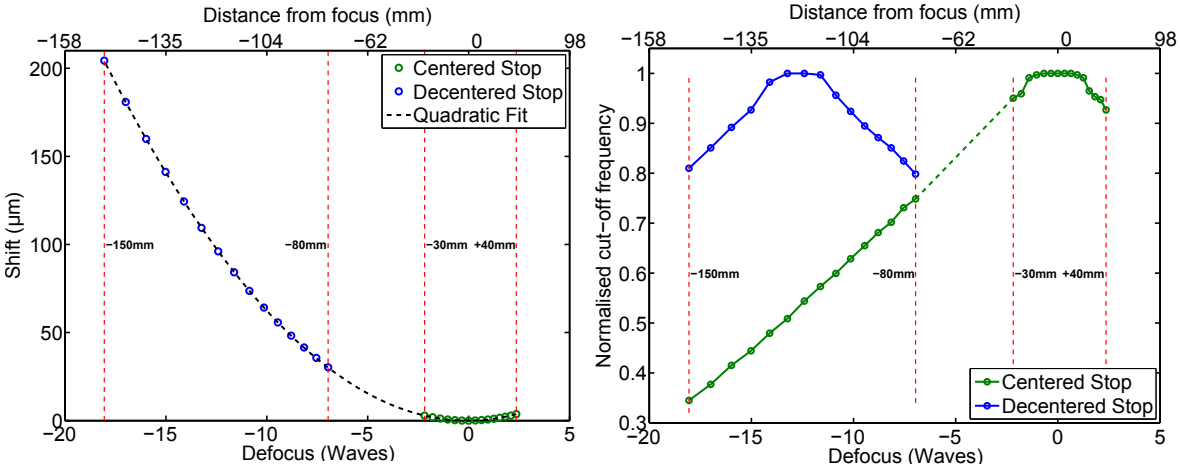


Fig. 2.7 (a) shift and (b) cut-off frequency of a CPM system with a centred and a decentred aperture stop plotted against defocus W_{20} in waves (bottom x-axis). The top x-axis shows the depth in mm corresponding to the W_{20} on the bottom x-axis.

Finally, to verify that displacing the aperture affects the cut-off frequency and does not affect the lateral translation of the PSF, the object was displaced in steps of $5mm$ over a $\pm 35mm$ depth range around the focal point of the decentred aperture system and for comparison, the centred aperture system as well. The shift and the cut-off frequency were then measured and are shown plotted on the same axes in Fig. 2.7a and Fig. 2.7b respectively. As shown in Fig. 2.7a, the shift of the decentred and centred aperture cases lies on the same quadratic. This implies that indeed, decentring the aperture does not affect the translation of the PSF. The cut-off frequency, which was defined as the highest frequency above the noise floor for an SNR of $100dB$, on the other hand, is amplified, as can be seen by comparing the blue to the green curve in Fig. 2.7b. Notice that over the $\pm 35mm$ depth range, the cut-off for the decentred aperture case drops more than the cut-off for the centred aperture case. This is not caused by the decentring of the aperture; instead, it is due to the fact that for the decentred aperture, the system is operated at $120mm$ away from the original focus position. As a result, a depth range of $\pm 35mm$ translates into a larger W_{20} range for the decentred aperture case than it does for the centred aperture case. This should be clearly visible in Fig. 2.7.

Finally, we conclude this section by stating that for the considered optical system, an aperture shift of $1.00mm$ will give a cubic PSF with a responsivity of:

$$\frac{\partial(\delta S)}{\partial W_{20}} = \frac{-2\sqrt{2}\lambda z_i (W_{20} - 12.5)}{3\alpha_o p} = (14.28 - 1.14W_{20}) \quad [\mu m/\text{Wave}] \tag{2.43}$$

and with no reduction in cut-off frequency where W_{20} is the defocus with respect to the in-focus position (*i.e.* maximum cut-off) of the decentred aperture case. To achieve the same responsivity with a centred aperture, the cut-off frequency would have to be reduced by $\sim 45\%$, which is unacceptable for most applications. On the other hand, to achieve the same cut-off frequency with a centred aperture, then the responsivity would be:

$$\frac{\partial(\delta S)}{\partial W_{20}} = \frac{-2\sqrt{2}\lambda z_i W_{20}}{3\alpha_o p} = -1.14W_{20} \quad [\mu m/\text{Wave}] \tag{2.44}$$

where W_{20} is with respect to the original focus position of the system. In other words, if we consider a specific case, for instance, $W_{20} = -1$ wave, decentring the aperture will result in a responsivity 13.5 times larger than that given by a centred aperture but with the same cut-off frequency. The same responsivity with a centred aperture will result in a reduction in the cut-off frequency of $\sim 53\%$.

2.3 Formalisation of the Complementary Kernel Matching (CKM) technique

Following the brief introduction presented in Section 1.5, a formal (mathematical) treatment of the CKM concept is presented in this section. To do this, the exact OTF of a WC system using a CPM needs to be considered. Fortunately, unlike the PSF, for which a model had to be derived from first principles as presented in the previous two sections, the exact expression for the OTF can be already found in the literature [41]. This is given in 1D by the following equation:

$$H_{1D}(u,\psi) = \begin{cases} \left(\frac{\pi}{48\alpha u}\right)^{1/2} e^{\left[i2\alpha u^3\right]} e^{\left[\frac{i2\psi^2 u}{3\alpha}\right]} \times A(u,\psi)\, e^{[\Phi(u,\psi)]}, & 0 < |u| \leq 1 \\ 1, & |u| = 0 \end{cases} \tag{2.45}$$

$u = {}^{f}/_{f_c}$ is the normalised spatial frequency where f denotes spatial frequency and f_c denotes the cut-off frequency; $\psi = 2\pi W_{20}$ is the defocus in *rad*; and, $\alpha = 2\pi\alpha_o$ is the peak cubic aberration of the CPM also in *rad*. Eq. 2.45 assumes a pupil radius of 1 - i.e. pupil coordinates should be normalised by the radius of the pupil. The function:

$A(u,\psi)\,e^{[\Phi(u,\psi)]}$ is given by:

$$A(u,\psi) = \left|C\left[b(u,\psi)\right] - C\left[a(u,\psi)\right] + iS\left[b(u,\psi)\right] - iS\left[a(u,\psi)\right]\right| \tag{2.46}$$

$$\Phi(u,\psi) = \angle\left\{C\left[b(u,\psi)\right] - C\left[a(u,\psi)\right] + iS\left[b(u,\psi)\right] - iS\left[a(u,\psi)\right]\right\} \tag{2.47}$$

$$\text{where: } a(u,\psi) = \left(\frac{12\alpha u}{\pi}\right)^{1/2}\left(\frac{\psi}{3\alpha} - (1-|u|)\right) \tag{2.48}$$

$$b(u,\psi) = \left(\frac{12\alpha u}{\pi}\right)^{1/2}\left(\frac{\psi}{3\alpha} + (1-|u|)\right) \tag{2.49}$$

and $C[\cdot]$ and $S[\cdot]$ are the Fresnel cosine and sine integrals respectively. Assuming a separable pupil function, the 2D OTF is then given by: $H(u,v,\psi) = H_{1D}(u,\psi) \times H_{1D}(v,\psi)$ where u and v are the two spatial frequencies.

Consider a scene with intensity of $I(x,y)$ for each point (x,y) in space. This can be expressed in continuous space as:

$$i(x,y) = \iint_S \left\{I(x,y)\,\delta(\gamma - x, \zeta - y)\right\}\,d\gamma\,d\zeta \tag{2.50}$$

where S is some region of space. Assuming further that the scene has a varying defocus profile given by: $W(x,y)\ \ [rad]$, then, the image captured by a WC system, $i_{cap}(\cdot)$, using a CPM as its coding element can be expressed with the aid of Eq. 2.45 in frequency as follows:

$$\mathscr{F}\left\{i_{cap}(x,y)\right\} = I_{cap}(u,v) = \iint_S \left\{I(x,y)\,e^{[2\pi i(xu+yv)]}H(u,v,W(x,y))\right\}\,dx\,dy \tag{2.51}$$

Under the assumption of an infinite SNR, the Wiener filter reduces to an inverse filter. The recovered image, $i_{rec}(x,y)$, in frequency can therefore be written as:

$$I_{rec}(u,v) = \mathscr{F}\left\{i_{rec}(x,y)\right\} = \frac{I_{cap}(u,v)}{H(u,v,\psi_o)} \tag{2.52}$$

where ψ_o represents the defocus at which the PSF used for the recovery has been acquired. Note that in this case, we are considering conventional WC recovery and hence, a single PSF acquired at some axial position (typically at or close to the in-focus position) is used to recover the capture image irrespective of the defocus in the scene. Expanding Eq. 2.52 yields:

$$I_{rec}(u,v) = \iint_S \left\{I(x,y)\,e^{\left[i\left(\left(2\pi x f_c + \frac{2W(x,y)^2}{3\alpha} - \frac{2{\psi_o}^2}{3\alpha}\right)u + \left(2\pi y f_c + \frac{2W(x,y)^2}{3\alpha} - \frac{2{\psi_o}^2}{3\alpha}\right)v\right)\right]} \times \ldots \right.$$
$$\left. \ldots \times \left(\frac{A(u,v,W(x,y))}{A(u,v,\psi_o)}\right) e^{[\Phi(u,v,W(x,y)) - \Phi(u,v,\psi_o)]}\right\}\,dx\,dy \tag{2.53}$$

where f_c is the cut-off frequency. Defining:

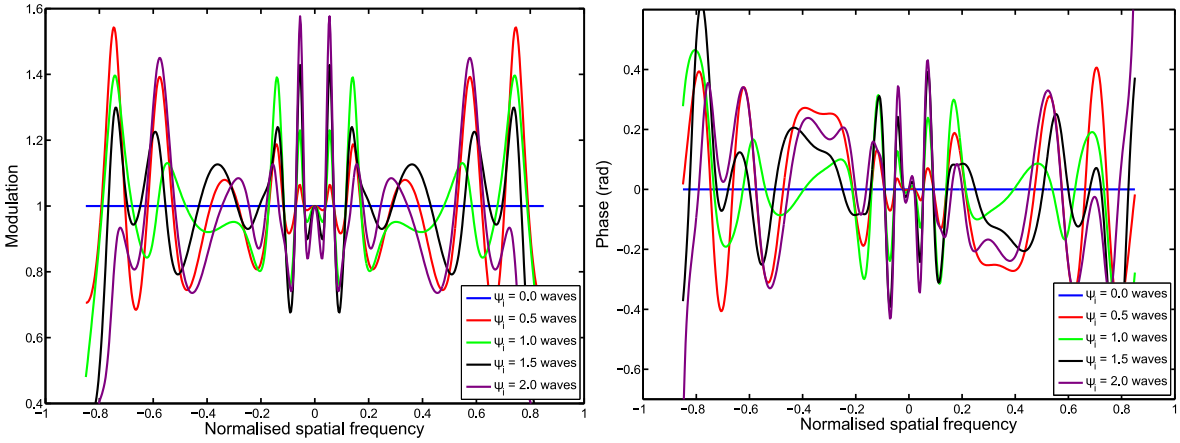
$$H_r\left(u,v,\psi_i,\psi_o\right)=\left(\frac{A\left(u,v,\psi_i\right)}{A\left(u,v,\psi_o\right)}\right)e^{\left[\Phi\left(u,v,\psi_i\right)-\Phi\left(u,v,\psi_o\right)\right]} \tag{2.54}$$

where $H_r\left(\cdot\right)$ will be referred to as the "residual transfer function" and its inverse Fourier transform, $\mathscr{F}^{-1}\left\{H_r\right\}=h_r\left(x,y,\psi_i,\psi_o\right)$, as the "residual point-spread function", then, the recovered image can be expressed in space as:

$$\begin{aligned} &i_{rec}\left(x,y\right)= \\ &\iint_S\left\{I\left(x,y\right)\delta\left(\gamma-x-\frac{W\left(x,y\right)^2}{3\pi\alpha f_c}+\frac{{\psi_o}^2}{3\pi\alpha f_c},\zeta-y-\frac{W\left(x,y\right)^2}{3\pi\alpha f_c}+\frac{{\psi_o}^2}{3\pi\alpha f_c}\right)*\ldots\right. \\ &\left.\ldots*h_r\left(\gamma,\zeta,W\left(x,y\right),\psi_o\right)\right\}d\gamma\,d\zeta \end{aligned} \tag{2.55}$$

In (conventional) WC, the CPM is assumed to render the PSF completely insensitive to defocus, which is tantamount to ignoring all the terms which depend on the defocus in Eq. 2.55. Under this assumption, the recovered image (Eq. 2.55) is identical to the scene (Eq. 2.50) which is the desired outcome. However, this assumption only holds for very large α values. A large α value, however, results in a lower SNR, which is of critical importance to the quality of the recovered image, especially when one considers the fact that inverse filtering amplifies the noise [50]. Consequently, in practice, a smaller α is desirable and is often used.

Investigating this further by comparing Eq. 2.50 to Eq. 2.55, one can immediately notice two important differences between the scene and the recovered image. The first concerns the residual point-spread function. From Eq. 2.54, we note that if $\psi_o=\psi_i$, then, $H_r\left(u,v,\psi_i,\psi_o\right)=1\ \forall\ \left(u,v\right)$ and therefore, the residual point-spread function (or equivalently the residual transfer function) has no effect on the recovered image. In any other eventuality, however, $H_r\left(\cdot\right)$ will distort the spectrum of the recovered image. In fact, the effect of the residual point-spread function on the recovered image manifests as post-recovery image artefacts which substantially degrade the quality of the recovered image, as discussed in the previous chapter. Shown in Fig. 2.8 are plots of the amplitude and phase of the residual transfer function for: $\alpha=3$ waves, $\psi_o=0$ and a range of ψ_i values as indicated. From Fig. 2.8a, which shows the magnitude of the residual transfer function, one can see that for all $\psi_i\neq\psi_o$, the residual transfer function will amplify or attenuate different frequencies of the scene's spectrum by different amounts; thus, distorting it. More importantly, the highly oscillatory phase spectrum of the residual transfer function shown in Fig. 2.8b implies that different harmonics of the scene's spectrum will be shifted in space by different amounts, giving rise to the replication artefacts which we observe in WC.



(a) (b)

Fig. 2.8 Plots of the modulation of the OTF for different ψ_i values and $\psi_o = 0$. (a) modulation (multiplication) of the amplitude of the scene's spectrum which corresponds to the magnitude of the residual transfer function and (b) modulation (addition) of the scene's phase spectrum which corresponds to the phase spectrum of the residual transfer function.

The second difference consists of a shift in space along the axes of the CPM given by:

$$-\frac{W(x,y)^2}{3\pi\alpha f_c} + \frac{{\psi_o}^2}{3\pi\alpha f_c} = -\frac{2W_{20}(x,y)^2}{3\alpha_o f_c} + \frac{2{W_o}^2}{3\alpha_o f_c} \tag{2.56}$$

where $W_o = {}^{\psi_o}/_{2\pi}$ $[waves]$. Comparing the first term to Eq. 2.16, one can notice that this is the shift introduced by the cubic PSF. The second term is the shift of the PSF used to perform the deconvolution (recovery). This implies that unless $\psi_o = \psi_i$, in which case the deconvolution is optimal, then, different (x, y) points in the scene will experience an additional shift in the lateral coordinates in proportion to the depth (or defocus) of the scene at that particular point. Consequently, by deconvolving the scene with a range of PSFs measured over a range of ψ_o values and finding the ψ_o value which gives a shift of zero (or from a more practical point of view, which minimises the shift), then, it is possible to: (1) eliminate (or reduce) the artefacts from the recovered image and (2) estimate the depth at each location in the scene -*i.e.* obtain a depth map of the scene, $\bar{W}(x,y)$, such that: $\bar{W}(x,y) \sim W(x,y)$. These two points are essentially the objective of the CKM technique.

Achieving these objectives, however, necessitates measuring the shift given by Eq. 2.56. This cannot be done from a single image because the fact that depth information (Z coordinate) is encoded as a shift in the lateral dimensions (X and Y-coordinates) implies that we cannot distinguish between them - or in other words, the Z dimension is coupled to the X and the Y dimensions. The problem is solved in CKM by capturing two images of the scene in which the depth at a given (x, y) point in the scene gives rise to a different shift (magnitude and/or direction). The disparity (*i.e.* difference in shift) between the image of a point or region of the scene as it appears in the two captured images will then give the shift induced by the depth only. The

generation of this disparity, which has far-reaching consequences on various important aspects of the technique, is the subject of the following sections.

2.4 Generating disparity

As previously mentioned, in order to solve for the depth, and hence eliminate the artefacts, CKM requires two images in which the depth of the scene generates a different shift (magnitude and/or direction). To solve for the depth, the difference in shift, which is referred to as disparity, between the image of each point or region of the scene as it appears in each of the two recovered images must then be estimated. The method used to generate disparity has important repercussions on the performance of the technique itself, as well as the complexity of the optics necessary for its implementation and the calibration procedure. In fact, CKM was originally conceived with the idea of generating disparity based on different shift directions (conjugate-masks approach); however, a proof-of-concept experiment soon revealed that an alternative way of generating disparity was necessary in order to make the technique practical and worth the extra complexity. Two other ways of generating disparity based on modulating the magnitude of the shift (the biplane method) and both the magnitude and direction of the shift (the conjugate-astigmatic-masks method) were thus investigated and are presented in the following subsections.

2.4.1 Conjugate-masks method

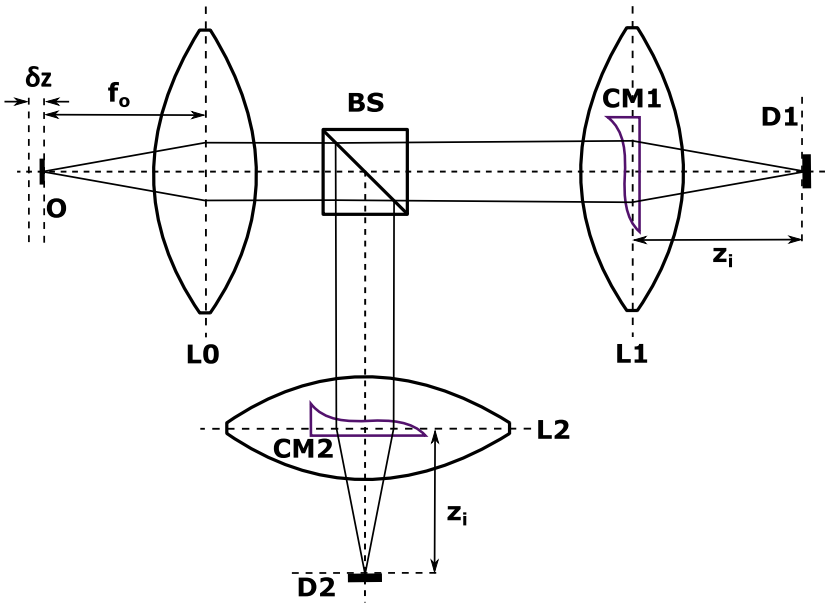
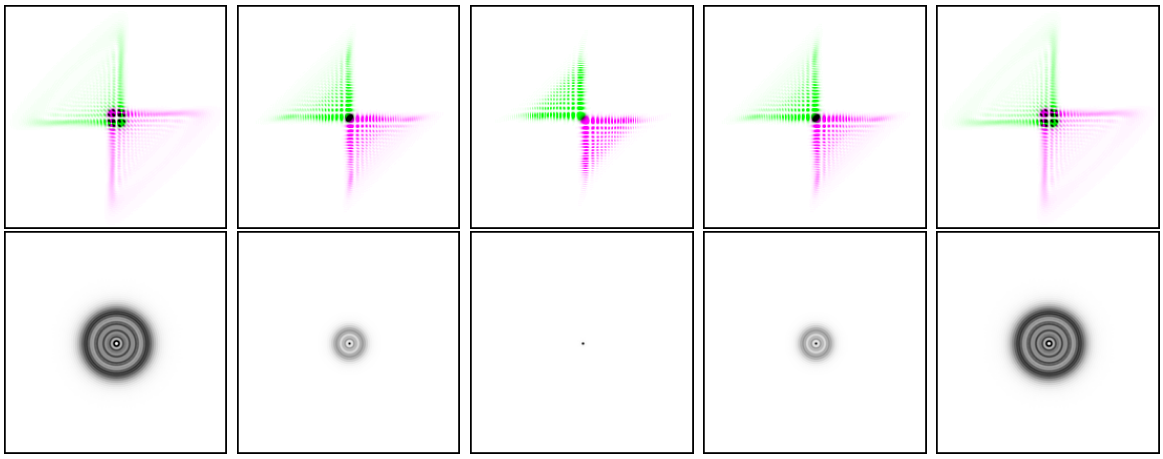


Fig. 2.9 Optical setup for the generation of disparity by the conjugate-masks method. **O**: object; **L0**: collimating lens; **BS**: beam splitter; **CM2** and **CM1**: cubic phase masks rotated by 180° with respect to each other; **L1** and **L2**: imaging lenses; z_i: image distance; f_o: object distance; δz: longitudinal defocus (positive in the shown direction); and, **D1** and **D2**: detectors.

In the conjugate-masks method, disparity is generated by modulating the direction of the shift. In other words, the shift in each of the two captured images is of the same magnitude but in diametrically opposite directions. This can be achieved by acquiring one of the images with a CPM and the other with its complex conjugate, where the complex conjugate of a CPM can be generated by simply rotating the mask through an angle of 180°. A possible way of implementing this method is therefore shown in Fig. 2.9. Lens $L0$ collimates the light coming from the object and a beam splitter is then used to generate two separate imaging paths. Two other lenses fitted with CPMs ($CM1$ and $CM2$) rotated by an angle of 180° with respect to each other are then used to image the object onto two detectors ($D1$ and $D2$) which could either be two separate ones or different portions of the same detector. An alternative implementation is to use a distorted diffraction grating as proposed in [27]. This technique would eliminate the necessity of replicating the pupil [51]; however, the optical efficiency would be substantially reduced. Furthermore, this technique does not permit full usage of the detector area.



(a) (b) (c) (d) (e)

Fig. 2.10 Comparison of the PSFs generated by the conjugate-masks setup. The top row depicts the two PSFs of the two imaging paths superimposed (with different colours). The bottom row shows the airy disk of a conventional optical system for comparison. The defocus from (a) to (e) is: −5.0, −2.5, 0, 2.5 and 5.0 waves respectively. An α of 5.0 waves was used.

Irrespective of the optical setup used to implement the conjugate-masks method, the PSFs of the two images should look as shown in Fig. 2.10 for an α of 5.0 waves and a range of defocus values. As shown, the PSFs move by equal magnitudes in opposite directions. The disparity between the two PSFs is therefore related to the defocus or depth of the point source object. Assuming a separable aperture, the shift of the cubic PSF is given by Eq. 2.17. From Eq. 2.55 and Eq. 2.56, the disparity is therefore given by:

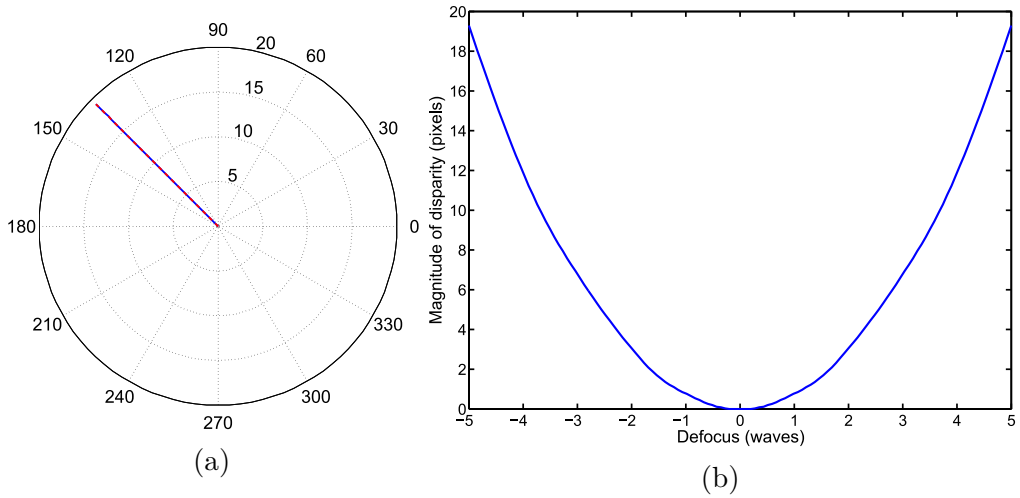
$$\text{Disparity} = \delta D = \frac{2\sqrt{2}\lambda z_i {W_{20}}^2}{3\alpha p} \approx \left(\frac{\sqrt{2} z_i p^3}{6\alpha\lambda}\right) \cdot \left(\frac{\delta z}{f_o\,(f_o + \delta z)}\right)^2 \tag{2.57}$$

where α is the peak cubic aberration in waves, p is the radius of the aperture stop and λ is the imaging wavelength. Furthermore, typically, $f_o \gg \delta z$, therefore Eq. 2.57 can

be approximated as:

$$\delta D \approx \left(\frac{\sqrt{2} M p^3}{6 \alpha \lambda {f_o}^3} \right) \cdot \delta z^2 \tag{2.58}$$

where M is the lateral magnification. From this equation, one can notice that the disparity is quadratically related to the peak defocus aberration, whereas it exhibits a "quasi-quadratic" relationship with longitudinal defocus (δz). This is confirmed by numerical simulations as shown in Fig. 2.11.



(a) (b)

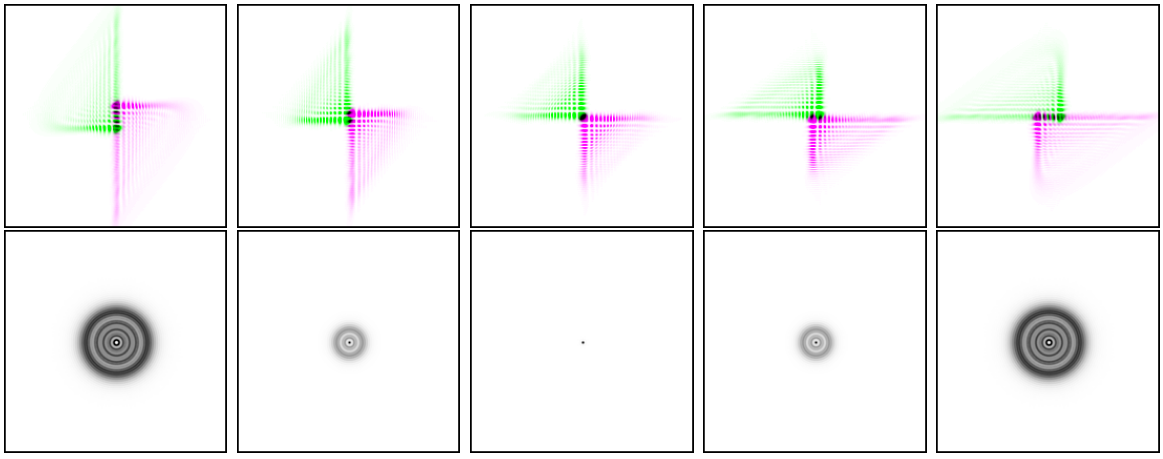
Fig. 2.11 Plots of the disparity between the PSFs of the conjugate-masks method. (a) polar plot of the disparity illustrating both direction and magnitude. The blue line is for negative defocus, while the red line is for positive defcus (lines are superimposed). (b) magnitude against defocus plot of the disparity.

From these analyses, we notice two important facts. Firstly, as shown in Fig. 2.11a, the conjugate-masks method generates identical disparity for positive and negative defocus. This implies that we cannot distinguish between them and as a result, only half of the operational DOF is viable. Secondly, the conjugate-masks method has a responsivity (given by the derivative of Eq. 2.57) of zero at the in-focus position which gradually increases as the defocus increases. This means that the conjugate-masks method does not allow the depth to be inferred with precision at small defocus values. Consequently, use of this method of generating disparity would limit the operation of the technique to less than half ($\sim 35\%$ from experimental observations) of the total DOF provided by the CPM in a conventional WC arrangement. Given that this is a rather significant inefficiency, a way of addressing the problem was needed and this led to the development of the following methods of generating disparity.

2.4.2 Conjugate-astigmatic-masks method

This method relies on the same phenomenon as the conjugate-masks approach; however, the addition of some astigmatism to the phase mask permits the sign of the defocus to be discerned. As a result, using this method, the operational DOF is, in principle,

doubled. The optical setup used for the conjugate-masks setup can be used for the conjugate-astigmatic-masks unchanged except for the design of the sag of the phase masks, which must include some amount of astigmatism.



(a) (b) (c) (d) (e)

Fig. 2.12 Comparison of the PSFs generated by the conjugate-astigmatic-masks setup. The top row depicts the two PSFs of the two imaging paths superimposed (with different colours). The bottom row shows the airy disk of a conventional optical system for comparison. The defocus from (a) to (e) is: -5.0, -2.5, 0, 2.5 and 5.0 waves respectively. An α of 5.0 waves and 2.5 waves of astigmatism was used.

The PSFs of the two captured images generated by this method are shown superimposed in Fig. 2.12 for various defocus values. Comparing this to the PSFs generated by the conjugate-masks method (Fig. 2.10), one can notice that the only difference lies in the fact that in this case, the PSFs do not translate along the 135° line. Instead, the PSFs move along a parabola while rotating with respect to each other; thus, following a distinct path for negative and positive defocus. This is shown in Fig. 2.12 where one can observe that the PSFs go through an angle of $\sim 270°$ for a defocus ranging from -5 to 5 waves. This rotation is caused by a different rate of translation along the cardinal axes of the CPM as a result of the introduced astigmatism. Therefore, in this case, to analyse the disparity, the translation along each axis needs to be considered separately. From Eq. 2.16, we can express the shift along each axis in the presence of astigmatism as:

$$\delta x = \frac{-\lambda z_i (W_{20} + W_a)^2}{3\alpha p} \tag{2.59}$$

$$\delta y = \frac{-\lambda z_i (W_{20} - W_a)^2}{3\alpha p} \tag{2.60}$$

where δx and δy denote the shift along the x and y axis of the CPM respectively, α is the peak cubic aberration of the CPM, W_a denotes the amount of astigmatism encoded on the phase mask in waves and z_i is the image distance. Keeping present that rotating

the mask is equivalent to a negative α value, then, the disparity, δD, can be expressed as:

$$\mid \delta D \mid = 2\sqrt{\delta x^2 + \delta y^2} = \left(\frac{2\lambda z_i}{3\alpha p}\right) \cdot \sqrt{\left(2{W_{20}}^4 + 12{W_{20}}^2{W_a}^2 + 2{W_a}^4\right)} \tag{2.61}$$

$$\angle \delta D = tan^{-1}\left(\frac{\delta x}{\delta y}\right) = tan^{-1}\left(\frac{(W_{20}+W_a)^2}{(W_{20}-W_a)^2}\right) \tag{2.62}$$

which can be expressed in terms of longitudinal defocus, δz, by substituting: $W_{20} = {\delta z p^2}/{(2\lambda f_o\,(f_o+\delta z))}$. W_a (in waves) would be produced by manufacturing the appropriate sag on the phase mask so it would be known accurately.

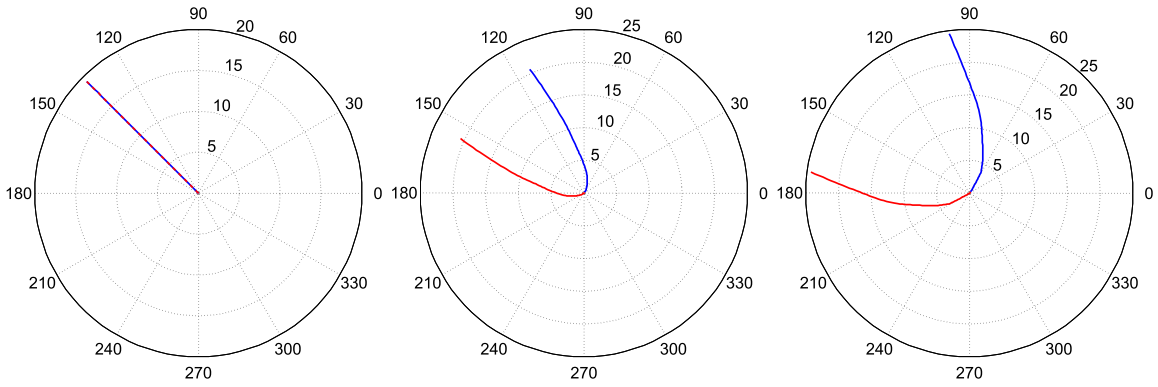


Fig. 2.13 Polar plots of the disparity between the two PSFs of the conjugate-astigmatic-masks method. (a) for 0 waves of astigmatism; (b) for 1.0 waves of astigmatism; and, (c) for 2.0 waves of astigmatism. The blue lines are for negative defocus, while the red lines are for positive defcus.

Numerical simulations of the disparity generated by the conjugate-astigmatic-masks method are shown in Fig. 2.13. Fig. 2.13a shows the disparity for $W_a = 0$. This is equivalent to the conjugate-masks method discussed in the preceding section and as one can observe, the disparity generated by negative (blue line) and positive (red line) defocus are identical and hence indistinguishable. Fig. 2.13b shows the disparity for $W_a = 1$ and as one can see, in this case, the disparity is distinct for each defocus sign. Finally, Fig. 2.13c shows the disparity for $W_a = 2$ and as one can see, the larger the amount of astigmatism introduced in the phase mask, the more distinct the disparity for each defocus sign is; hence, the easier it is to tell them apart. Of course, the larger the amount of astigmatism, the lower the cut-off frequency of the system as a whole will be and hence the lower the quality of the recovered image. This poses a maximum limit on the amount of astigmatism that can be used. As far as the responsivity is concerned, the conjugate-astigmatic-masks method does not show a zero value at the in-focus position. The responsivity, however, still decreases with decreasing defocus. One therefore concludes that, whereas this method permits the sign of the defocus to be discerned, and hence the DOF provided by the CPM can be exploited more

efficiently compared to the conjugate-masks method, it still requires replication of the pupil (*i.e.* two CPMs) and its responsivity is still dependent on the depth of the scene itself. It therefore solves only one of the shortcomings of the conjugate-masks method.

2.4.3 Biplane method

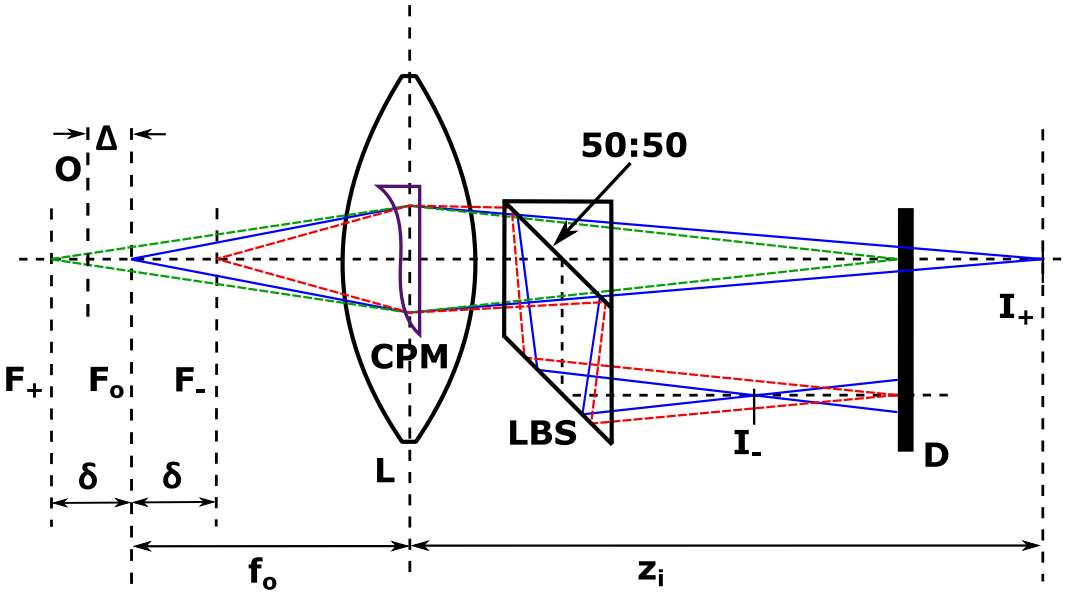
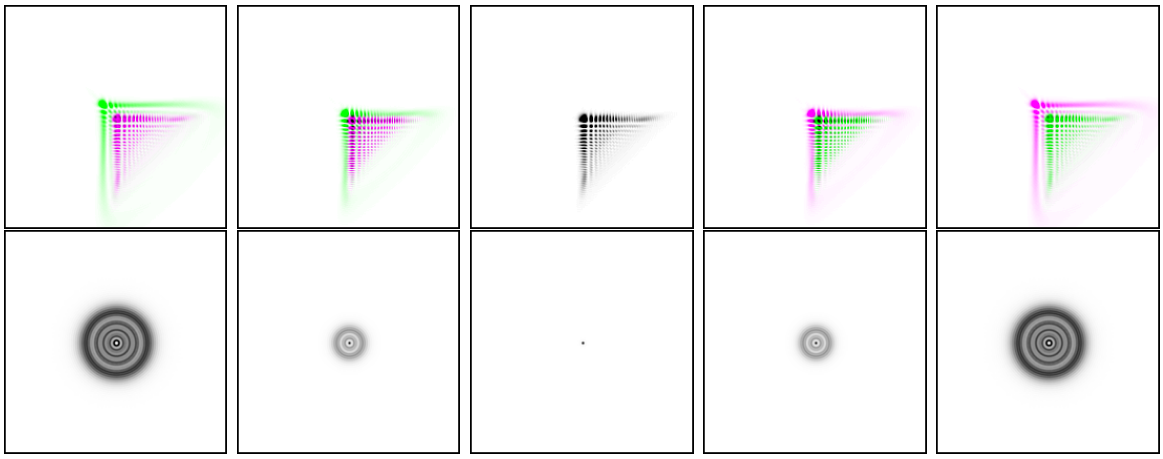


Fig. 2.14 Optical setup for the generation of disparity by the biplane method. **O**: a generic object plane or position along the z-axis; $\boldsymbol{F_o}$: original focal plane; **Δ**: longitudinal defocus (in μm) with respect to the original focal plane (positive in the shown direction); $\boldsymbol{F_-}$ and $\boldsymbol{F_+}$: the two new focal planes; $\boldsymbol{\delta}$: difference in focal distance between the two new focal planes and the original focal plane; $\boldsymbol{f_o}$: original focal length; **L**: imaging lens; **CPM**: cubic phase mask; **LBS**: later beam splitter; $\boldsymbol{I_-}$ and $\boldsymbol{I_+}$: images formed along the two image paths for an object situated in the original focal plane; $\boldsymbol{z_i}$: original image distance; and, **D**: detector. The green and red dashed lines show conjugate planes. The solid blue line denotes the imaging path.

The last method of generate disparity which will be considered is the biplane method. In this case, we do not rely on the direction of the shift as in the conjugate-masks method or the direction and magnitude as in the conjugate-astigmatic-masks method. Instead, this method relies on the dependence of the rate of translation of the cubic PSF with defocus. From Eq. 2.17, we know that the shift of the cubic PSF is a quadratic function of defocus (W_{20}). This means that if a defocus offset is introduced between the two acquired images, then, a disparity will result. This is the concept exploited in the biplane method.

Fig. 2.14 shows the optical setup (assuming a finite conjugate system) used to implement the biplane method. In this case a single phase mask is fitted in the aperture stop of the imaging lens and a lateral beam splitter is used to create both an offset in defocus between the two images, as well as to split the imaging path into two. The two images required for CKM can thus be easily acquired on a single detector as shown. It should be noted that some additional optics, such as the introduction of a lens or

glass slab or slabs in one of the imaging paths, might be necessary, depending on the magnitude of the desired defocus offset. The difference in defocus is set such that two focal planes (F_+ and F_-) are created around the original focal plane (F_o) at equal distances (δ) away from it.



(a) (b) (c) (d) (e)

Fig. 2.15 Comparison of the PSFs generated by the biplane setup. The top row depicts the two PSFs of the two imaging paths superimposed (with different colours). The bottom row shows the airy disk of a conventional optical system for comparison. The defocus from (a) to (e) is: -5.0, -2.5, 0, 2.5 and 5.0 waves respectively. An α of 5.0 waves was used and the difference in defocus between the two imaging paths was set to 2.5 waves.

The generated PSFs are as shown superimposed in Fig. 2.15 for a range of defocus. Comparing these to the PSFs generated by the two methods considered previously, one can notice that the PSFs in this case have the same orientation; thus, they translate in the same direction. Starting again from Eq. 2.17, the shift experienced by the PSF with positive defocus offset (imaging path shown by a dashed green line in Fig. 2.14), PSF_+, can be expressed as:

$$\delta S_+ = \frac{\sqrt{2}\lambda z_i^+ W_{20}^{+2}}{3\alpha p} \tag{2.63}$$

where z_i^+ is the image distance of the positive defocus offset path and is given by:

$$z_i^+ = z_i - \delta M^2 = f_o M - \delta M^2 \tag{2.64}$$

where $M = {}^{z_i}/_{f_o}$ is the magnification of the original geometry. The defocus, W_{20}^+, can be expressed as:

$$W_{20}^+ \approx \frac{(\Delta - \delta)\, p^2}{2\lambda\,(f_o + \delta)\,(f_o + \Delta)} \tag{2.65}$$

Similarly, for the PSF with a negative defocus offset (imaging path shown by a red dashed line in Fig. 2.14), PSF_-, we can write:

$$\delta S_- = \frac{\sqrt{2}\lambda z_i^- {W_{20}^-}^2}{3\alpha p} \tag{2.66}$$

$$z_i^- = z_i + \delta M^2 = f_o M + \delta M^2 \tag{2.67}$$

$$W_{20}^- \approx \frac{(\Delta+\delta)\,p^2}{2\lambda\,(f_o-\delta)\,(f_o+\Delta)} \tag{2.68}$$

The disparity is therefore given by:

$$\delta D = \delta S_+ - \delta S_- \approx \left(\frac{\sqrt{2}\;p^3\;M}{12\alpha\lambda(f_o+\Delta)^2}\right)\cdot\left(\frac{(\delta-\Delta)^2\,(f_o-\delta M)}{(\delta+f_o)^2} - \frac{(\delta+\Delta)^2\,(f_o+\delta M)}{(\delta-f_o)^2}\right) \tag{2.69}$$

Assuming $f_o \gg \delta M$, (for a typical implementation on a $40\times$, 0.5NA system, $f_o \sim 40\times\delta M$), then we can approximate Eq. 2.69 as:

$$\delta D \approx \left(\frac{-\sqrt{2}\;p^3\;M}{3\alpha\lambda f_o^3}\right)\cdot(\delta\Delta) \tag{2.70}$$

where one can notice that not only is the disparity proportional to the magnification, M; and inversely proportional to the imaging wavelength, λ; the peak cubic aberration, α; and, the f-number cubed, ${f_\#}^3$; but it also is proportional to the defocus offset, δ, which we introduce. This is important because whereas the other parameters are stipulated by other requirements and thus cannot always be changed without compromising the performance of the system, δ is not. As a result, the defocus offset provides a completely decoupled degree of freedom which can be adjusted without compromise in such a way as to maximise the disparity. This advantage is not provided by any other disparity generation method considered.

Another important consideration revealed by Eq. 2.70 is the fact that the disparity is a linear function of the longitudinal defocus Δ rather than a quadratic, as was the case with the conjugate-masks method, or quasi-quadratic as in the conjugate-astigmatic-masks method. This implies that the sign of the disparity mirrors that of the defocus and thus, it is easy to differentiate negative and positive defocus. Moreover, the fact that the disparity is linearly related to the defocus implies that the responsivity (derivative of the disparity with respect to the defocus) is constant. Therefore, whereas the previous two methods display low responsivity for small defocus values, making them unsuitable close to the in-focus position, the biplane method does not.

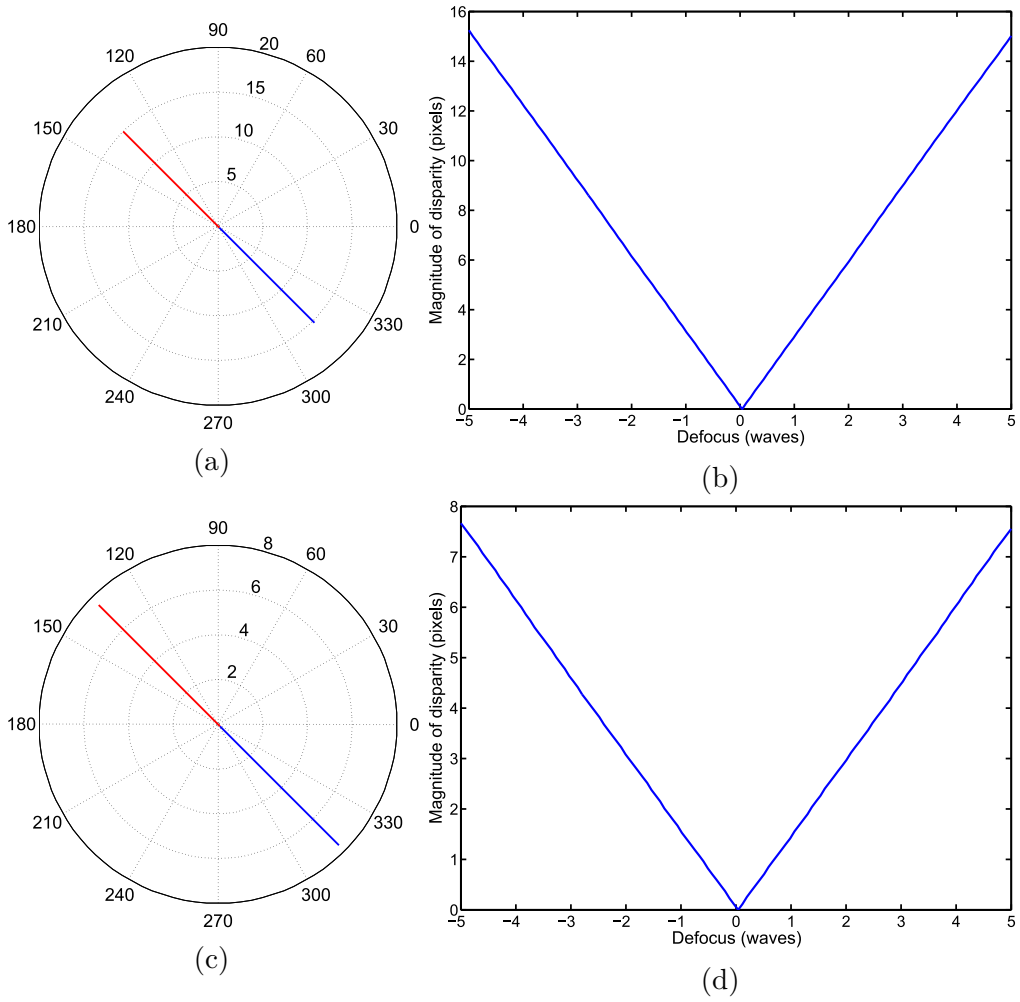


Fig. 2.16 Plots of the disparity between the two PSFs of the biplane method. (a) polar plot of the disparity showing both magnitude and direction with $\delta = 2.0$ waves and $\alpha = 5$ waves; (b) corresponding magnitude against defocus plot; (c) polar plot of the disparity with $\delta = 2.0$ waves and $\alpha = 5$ waves; (d) corresponding magnitude against defocus plot. The blue and red lines in the polar plots are for negative and positive defcus respectively.

Plots of the disparity obtained from numerical simulations are shown in Fig. 2.16 for 2.0 and 1.0 waves of defocus difference and an α of 5.0 waves. The polar plots reveal that indeed, the sign of the disparity reflects the sign of the defocus as predicted by the analyses. On the other hand, by comparing the magnitude of the disparity to that given by the other two methods, one can notice that the disparity generated by biplane method is smaller for larger values of defocus, which constitutes an obvious disadvantage. Equating Eq. 2.57 to Eq. 2.70, we deduce that the biplane method will generate a disparity greater than or equal to the conjugate-masks method for: $\mid \Delta \mid \leq 2\delta$ for a given geometry and peak cubic aberration.

2.4.4 Comparison of the various disparity generation techniques

As discussed above, the method used to generate disparity not only impacts the performance of the technique itself, but also its implementation. Careful analysis is therefore in order so as to identify the best compromise between practicality and performance. In this section, a qualitative comparison of the three disparity generation methods mentioned above is given and the best compromise selected.

Consider first the optical setup and complexity of implementation, the conjugate-masks and the conjugate-astigmatic-masks techniques require two CPMs and the replication of the aperture plane as shown in Fig. 2.9 to implement. An appropriately deformed diffraction grating [27] can also be used; however, this would compromise the optical efficiency and the efficient usage of the sensor area. Alternatively, a single CPM and a single imaging path could be used and the two necessary images would then have to be acquired sequentially by rotating either the sample or the CPM. This has the obvious disadvantage of precluding single-snapshot operation, which is one of the most alluring features of this technique. Furthermore, precision (highly repetitive) positioning (rotation) equipment would also be essential; thus, making this option rather unattractive from a practical perspective. In contrast, the biplane method requires a single CPM and no replication of the aperture plane without sacrificing optical efficiency, sensor area or single-snapshot operation and without requiring any positioning equipment. The only additional optical component necessary is a lateral beam splitter, which can be replaced by two cameras if desired with no additional complexity.

Next we consider the disparity and the responsivity of the various disparity generation methods. These are of particular importance because these parameters are directly related to the performance of CKM. For the conjugate-masks method, the disparity is a quadratic function of defocus and occurs along a single direction (along the unity-slope line with respect to the CPM axes). Since a quadratic has two solutions for a given disparity, then, this technique is sensitive only to the magnitude of the disparity. In other words, the sign of the defocus cannot be deduced and as a consequence, the operational depth range (depth range over which the depth of the scene/sample can be inferred reliably) is half of that provided by the CPM in a conventional setting. Even worse, the responsivity is proportional to the defocus. This implies that whereas a large disparity is generated at large defocus values, close to the in-focus position, the responsivity is low; therefore, unless the defocus in the scene is substantial, it cannot be measured reliably using this method. This reduces the operational depth range of this method further, resulting in an operational depth range $\sim 40\%$ of that provided by the CPM.

On the other hand, the conjugate-astigmatic-masks method solves both of the issues mentioned above by modulating the direction of the shift. By measuring both the magnitude and direction of the disparity, therefore, both the magnitude and sign of the defocus can be inferred. The responsivity of this method is still dependent on the defocus; however, it is non-zero at the in-focus position. Small amounts of defocus in the scene can therefore potentially still be reliably measured. The biplane method generates an approximately linear disparity with defocus, therefore the responsivity is constant (*i.e.* independent of defocus). Like the conjugate-astigmatic-masks methods, the sign of the defocus in the biplane method is given by the direction of the disparity. However, unlike the conjugate-astigmatic-masks method, in the biplane method, the sign of the defocus is given by two diametrically opposite directions. This increases the robustness and also facilitates the inference of the sign of the defocus, especially for small defocus values. This permits both the conjugate-astigmatic-masks and the biplane methods to exploit the entire operational depth range provided by the CPM. On the downside, the biplane method generates the smallest disparity at large defocus values compared to both the conjugate-masks and conjugate-astigmatic-masks. The conjugate-astigmatic-masks method can generate disparities of a larger magnitude than that generated by the conjugate-masks method depending on the amount of astigmatism introduced however, one should exercise caution because astigmatism lowers the SNR.

Finally, we consider redundancy in the acquired information for each method. Starting with the conjugate-masks and the conjugate-astigmatic-masks methods, we notice that although the two acquisitions have different shifts with respect to each other, they contain exactly the same information regarding the radiance of the sample or the scene. This is because the two images are acquired with identical pupil functions except for the sign of the α of the CPM. Conversely, the biplane method has two focal planes. This implies that depending on the depth of the sample, parts of the sample that might be out-of-focus in one of the images will be in-focus in the other. As a consequence, the biplane method displays a higher overall frequency content as well as a better SNR in the reconstructed image. On the other hand, one would expect the two images to be less similar in the biplane method compared to the other two methods. This makes it harder to precisely match features between them and thus to measure the disparity.

Based on this discussion, we conclude that the biplane method is the best of the three options. This is mainly because it has the largest operational depth range and it also is the easiest to implement and align. Experimental work will therefore be focused on this implementation with the exception of the initial proof-of-concept experiment discussed in Chapter 3.

2.5 Statistical investigation of the various disparity generation techniques - Cramér-Rao lower bound

An inherent capability of any 3D imaging technique such as CKM is that of localisation in 3D space. In the case of CKM, besides the 3D position of a particular feature, information about the irradiation of the sample is also obtained. The objective of this section, however, is to assess the 3D localisation performance (precision) of the various implementations (disparity generation methods) of the CKM technique. This allows the comparison of CKM to other established techniques and thus to establish its place in the broad spectrum of 3D imaging techniques - which has been taken to include 3D super-resolution techniques.

The parameters which affect the localisation precision of a 3D imaging technique are specific to the technique itself, therefore precision analysis tend to be technique-specific as well. For instance, the analyses proposed in [52–54] are applicable only to depth from defocus (DFD). Given that each theoretical framework has its own set of assumptions, comparing between techniques can be difficult and perhaps even narrow in scope. One generic framework which solves this issue was proposed by Oder *et al.* in [55] and is based on the Cramér-Rao lower bound (CRLB). The CRLB is a generic statistical tool through which the minimum error of an unbiased estimator $\hat{\boldsymbol{\theta}} = \left[\hat{\theta_1}, \hat{\theta_1}, .., \hat{\theta_n}\right]^T$ of some parameter $\boldsymbol{\theta} = [\theta_1, \theta_1, .., \theta_n]^T$ $\left(i.e.\ \langle\hat{\boldsymbol{\theta}}\rangle = \boldsymbol{\theta}\right)$, can be calculated for a given process [56]. The process itself is represented by the probability density function $p\,(y;\boldsymbol{\theta})$ where y denotes the observations. The $(k,l)^{th}$ element of the $n \times n$ Fisher information matrix, $I_{k,l}$, is therefore defined as:

$$I_{k,l}\,(\boldsymbol{\theta}) = E\left[\frac{\partial\ ln\ p\,(y;\boldsymbol{\theta})}{\partial\theta_k} \cdot \frac{\partial\ ln\ p\,(y;\boldsymbol{\theta})}{\partial\theta_l}\right] \tag{2.71}$$

By the Cramér-Rao inequality, we can write:

$$Var\left(\hat{\theta}_m\right) \geq I^{-1}_{m,m}\,(\boldsymbol{\theta}) \tag{2.72}$$

It therefore follows naturally that the smallest possible error in the estimate of parameters $\boldsymbol{\theta}$ - *i.e.* the CRLB - is given by:

$$CRLB_{\hat{\boldsymbol{\theta}}} = \delta_{\hat{\boldsymbol{\theta}}} = \sqrt{diag\,(\boldsymbol{I}\,(\boldsymbol{\theta}))^{-1}} \tag{2.73}$$

where $diag\,(\boldsymbol{I}\,(\boldsymbol{\theta}))^{-1}$ denotes the vector composed of the reciprocal of the diagonal elements of the Fisher matrix.

In computing the CRLB for CKM, an isolated point source scene will be assumed despite the fact that CKM can also be applied to extended scenes - and perhaps that

could even be considered to be its primary application. Other scene models have been considered in the literature for different techniques, such as the Gaussian prior on the scene's slopes taken in [54] for single-image DFD. These are, however, not as mainstream as the assumption of an isolated point-source scene. Furthermore, the CRLB is affected by the spectral content of the scene where the higher the spectral content, the lower the CRLB [22]. In this sense, the computed CRLB can be taken as the smallest error achievable by CKM for a given geometry, SNR and over all possible scenes or samples. Nevertheless, the primary objective of this section is to compare CKM and its various implementations to other established techniques. Consequently, as long as the same assumptions regarding the scene are taken for each technique, then a meaningful comparison can still be made. Lastly, in computing the CRLB for CKM, the post-processing (Wiener deconvolution) was not taken into consideration. This is because Wiener deconvolution does not improve the SNR nor increase the information content hence the lower bound on the uncertainty still holds. One could argue that the SNR is improved by band-limiting the image, however, that could be applied to any technique. Therefore, as far as comparing different techniques is concerned, it was deemed inconsequential.

The image of an isolated sub-resolution point is obviously the PSF therefore the observations in the case consist of the photon count per pixel over an N-pixel long array of pixels. If we assume the photon count of the k^{th} pixel to be Poisson distributed, then the probability density function is given by:

$$p(c_1, c_2, .., c_N; \boldsymbol{\theta}) = \prod_{k=1}^{N} \left(\frac{\nu_{\boldsymbol{\theta},k}^{c_k} e^{-\nu_{\boldsymbol{\theta},k}}}{c_k!} \right) \tag{2.74}$$

where $\boldsymbol{\theta} = [x, y, z]$ denotes 3D space, c_n denote the photon count in the n^{th} pixel and $\nu_{\boldsymbol{\theta},n} = \mu_{\boldsymbol{\theta},n} + \beta_{\boldsymbol{\theta},n}$ denotes the expected - or the "true" for an unbiased estimator - photon count of the n^{th} pixel; $\mu_{\boldsymbol{\theta},n}$ denotes the mean number of signal photons and $\beta_{\boldsymbol{\theta},n}$ the number of background photons. Substituting Eq. 2.74 into Eq. 2.71 and simplifying yields the Fisher information matrix as follows:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \sum_{k=1}^{n} \left[\left(\frac{1}{\nu_{\boldsymbol{\theta},k}} \right) \cdot \left(\frac{\partial \nu_{\boldsymbol{\theta},k}}{\partial \boldsymbol{\theta}} \right)^T \cdot \left(\frac{\partial \nu_{\boldsymbol{\theta},k}}{\partial \boldsymbol{\theta}} \right) \right] \tag{2.75}$$

thus the CRLB can be evaluated. Notice that whereas the data model adopted here (known as the "Poisson data model" [57]) takes into consideration the pixelation of the detector, shot noise and the background count (which contributes to the shot noise), it omits the read noise. Although the read noise can have a non-negligible effect depending on the number of signal photons and on the effective pixel size, the Poisson data model is nonetheless often used in the literature for comparison purposes [58] mainly due to its relative simplicity. Thorough analysis of the Poisson data model and other data models which take into consideration the read noise of the detector can be found in [57].

As one can notice from the above discussion, this framework makes no assumptions about the morphology of the PSF, therefore it is applicable to any PSFs [55], making it very versatile. In fact, it has been extensively exploited in the literature to analyse and compare various 3D PSFs and various microscopy techniques including: the astigmatic PSF [59], multifocal plane microscopy (MUM) [60], 4Pi microscopy [61] and even asymmetric engineered PSF such as the DH-PSF [62, 63, 58]. Here we extended this body of work to include the cubic PSF. For the conventional (non-engineered or clear aperture) PSF, tractable expressions exit. This is rarely the case for engineered PSFs, therefore the PSF has to be computed numerically [63]. The simulations conducted here assume a $40\times, 0.50$NA system and an image size of 265×265 pixels with $162.5nm$ square pixels. The photon count was set to a total of 4,000 signal photons and 5 background photons at a wavelength of $530nm$. For techniques which necessitate the acquisition of more than one image of the emitter (CKM requires two), the total number of signal photons was split equally between the number of acquired images. Also, the total Fisher information is given by the sum of the Fisher information in each acquired image [60, 64]. The PSF was calculated using the high NA model proposed in [59, 65]. The PSF as a function of axial position is thus given by:

$$PSF\left(x,y,z\right) = \left|\mathscr{F}\left\{f\left(k_x,k_y\right)\cdot e^{\left[2\pi i\cdot K_z\left(k_x,k_y\right)\cdot z\right]}\right\}\right|^2 \tag{2.76}$$

where z, y, z are space coordinates; k_x and k_y are the reciprocal coordinates; $f\left(k_x,k_y\right)$ is the complex aperture function (NA limited) and $K_z\left(k_x,k_y\right) = \sqrt{\left(n/\lambda\right)^2 - \left({k_x}^2 + {k_y}^2\right)}$; n being the refractive index of the medium and λ is the wavelength of light. The PSF given by Eq. 2.76 has to be normalised by the energy and then multiplied by the number of photons.

We start by computing the CRLB for the conventional Gaussian PSF and for a single cubic PSF to highlight the differences between them. Two techniques which rely on the Gaussian PSF, the astigmatic PSF and the MUM technique, are considered next as two examples of standard or established 3D localisation techniques. The CRLB for the various implementations of CKM, a technique which relies on the cubic PSF, was computed next. In addition, the biplane CKM implementation was also implemented with a shifted aperture. As discussed in Section 2.1 and Section 2.2, decentring the aperture function results in a higher responsivity without compromising the cut-off frequency. This is expected to increase the amount of depth information in the PSF and hence to lower the CRLB. Plots of the CRLB for each of the three dimensions of space (σ_x, σ_y and σ_z) as a function of axial position (depth) for each of these PSFs and techniques are shown in Fig. 2.17. Furthermore, the average CRLBs ($\bar{\sigma_x}, \bar{\sigma_y}$ and $\bar{\sigma_z}$) are shown tabulated in Tbl. 2.1 for various depth ranges as a figure of merit. This is mainly intended to facilitate comparison between the different techniques and PSFs considered.

As can be seen from Fig. 2.17a, the Gaussian PSF displays an exceptionally low σ_x and σ_y close to the in-focus position. This is due to the small size of the PSF -

hence the higher SNR - at this depth. On the other hand, at the focus, σ_z displays a singularity due to the symmetry of the PSF around the focal point. This implies that close to the focus, the Gaussian PSF has very little depth information. As the depth starts increasing, the PSF increases in size rapidly causing σ_z to drop rapidly as well. This is, however, accompanied by a drop in SNR which eventually causes σ_z to start increasing rapidly again. As a consequence, the depth range over which an adequately low uncertainty in all three dimensions of space is obtained is severely limited, making the Gaussian PSF unsuitable for 3D localisation. In contrast, the cubic PSF's transverse intensity profile changes only slightly with depth. This implies that the drop in SNR is small for a relatively large distance away from the focal position. As a result, the in-focus lateral localisation precision is conserved over a large depth range, as can be observed from Fig. 2.17b where the CRLBs for a cubic PSF generated using a CPM with an α of 3 waves is shown. On the other hand, this same resilience to defocus of the cubic PSF means that the cubic PSF has little depth information. This is clearly visible in Fig. 2.17b from the wide region of high uncertainty close to the focus; at which point, the Fisher information is zero giving rise to a singularity as with the Gaussian PSF. Lastly, it must be stated that this behaviour is entirely expected given that the cubic PSF was engineered to extended the DOF of an optical system, or in other words, to be insensitive to depth, which is tantamount to minimising depth information.

As discussed above, the singularity at the focal point produced by the Gaussian PSF severely limits its depth information content and hence its utility for 3D localisation applications. As a result, extensive research has been dedicated to the solution of this problem. One of the earliest solutions entails the intentional introduction of astigmatism in the system, often by means of a cylindrical lens. This breaks the symmetry of the PSF around the focal point, hence the singularity is eliminated. This also permits the discrimination of the sign of the depth with respect to the focal point, therefore the operational depth range is extended further. The CRLB for such a scenario has been computed for a PSF with a difference of $\sim 5.65\mu m$ between the sagittal and tangential foci as shown in Fig. 2.17c.

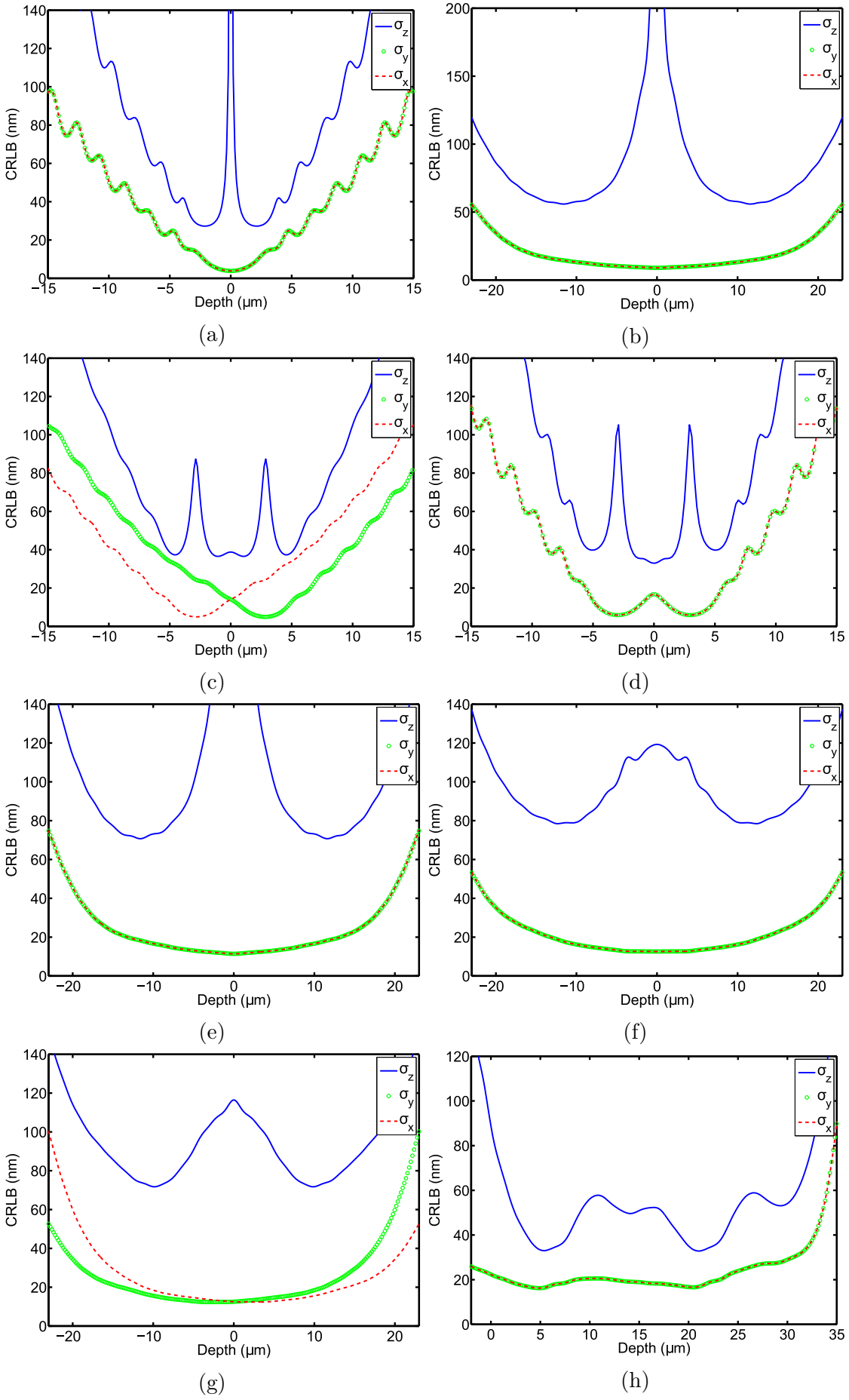


Fig. 2.17 Crámer-Rao lower bound for various techniques. (a) Gaussian PSF, (b) cubic PSF, (c) astigmatic PSF, (d) multifocal plane microscopy, (e) conjugate-masks technique, (f) biplane technique, (g) conjugate-astigmatic-masks and (h) biplane with shifted aperture. σ_x, σ_y and σ_z denote the Crámer-Rao lower bound (CRLB) along the x, y and z dimensions of space.

Another solution to the problem is the so-called MUM technique. In this case, two cameras are used to capture an image of the emitter simultaneously. The image distance for each camera is adjusted so that the corresponding two focal planes are some distance apart. This has also been simulated, which together with the astigmatic PSF, represents two commonly used or established techniques for comparison. The distance between the two focal planes was set to $\sim 5.95\mu m$ and the CRLB for MUM is shown in Fig. 2.17d. Note that the CRLB for these two techniques depends on the amount of astigmatism and on the distance between the focal planes respectively. The values used here are typical experimental values take from [66] and scaled accordingly by the NA.

As can be seen from Fig. 2.17c and Fig. 2.17d, the astigmatic PSF and the MUM technique do not display a singularity at the focal point as expected. Furthermore, their σ_z is flatter over a wider depth range. This makes them suitable for 3D localisation over roughly a depth range of $\pm 5.0\mu m$ in this case. The mean CRLBs increase if a larger depth range is considered, as tabulated in Tbl. 2.1. The two techniques show a similar uncertainty in depth, whereas the MUM displays $\sim 40\%$ lower uncertainty in the lateral dimensions over a $\pm 5\mu m$ depth range compared to the astigmatic PSF.

We next consider the CKM implementations where an α of 3 waves has been taken for all simulations. Considering first the conjugate-masks method, one immediately notices that the singularity in σ_z which the single CPM displays is preserved, as can be seen in Fig. 2.17e. This is expected because the responsivity of this method is zero at the focal point. Moreover, in common with all the other disparity generation methods, this method necessitates the capture of two images, hence the number of photons has to be split between two images resulting in a lower overall SNR. These two facts coupled together result in higher CRLBs for the conjugate-masks method compared to the single CPM. Furthermore, the wide central region of high uncertainty, coupled with the fact that the disparity is symmetric around the focal point, limits the operational depth range of this method. This confirms the deductions made in Section. 2.4 from the analysis of the disparity.

In contrast, the conjugate-astigmatic-masks method and the biplane methods do not display a singularity at the focal point in σ_z. This is because both methods break the symmetry of the disparity around the focal point. Alternatively, this can be deduced from the fact that neither of these methods displays zero responsivity at the focal point. Both methods display a similar σ_z as can be seen from Fig. 2.17g and Fig. 2.17f, where the difference in distance between the two focal planes of the biplane method was set to $7.70\mu m$ to coincide with the experiments conducted. For the conjugate-astigmatic-masks, the amount of astigmatism introduced produces $6.36\mu m$ difference between the sagittal and the tangential foci. Such a large amount of astigmatism and distance between planes for the biplane method is possible due to the CPM's insensitivity to defocus.

Technique	$\bar{\sigma_x}$ (nm)	$\bar{\sigma_y}$ (nm)	$\bar{\sigma_z}$ (nm)	Depth range (μm)
Conventional PSF	60.45	60.45	129.29	-20 - +20
Conventional PSF	40.99	40.99	90.33	-15 - +15
Conventional PSF	11.93	11.93	44.68	-5 - +5
Conjugate-masks	19.63	19.63	126.58	-20 - +20
Conjugate-masks	22.11	22.11	82.11	+5 - +20
Biplane	22.12	22.12	82.12	-20 - +20
Conjugate-astigmatic-masks	20.80	20.80	89.77	-20 - +20
Biplane with shifted aperture	20.21	20.21	48.14	+2 - +32
Astigmatic PSF	61.51	61.51	125.31	-20 - +20
Astigmatic PSF	42.24	42.24	87.14	-15 - +15
Astigmatic PSF	15.88	15.88	47.36	-5 - +5
MUM	71.01	71.01	153.06	-20 - +20
MUM	43.55	43.55	100.72	-15 - +15
MUM	9.37	9.37	50.53	-5 - +5

Table 2.1 Mean Crámer-Rao lower bound along the x ($\bar{\sigma_x}$), y ($\bar{\sigma_y}$) and z ($\bar{\sigma_z}$) dimensions of space against depth range for various techniques.

On the other hand, the conjugate-astigmatic-masks method displays higher lateral localisation precisions, σ_x and σ_y, compared to the biplane method. The only depth range where the conjugate-masks method shows a slightly lower uncertainty ($\sim 7\%$ lower) than either of these two methods is roughly between $5\mu m$ and $15\mu m$ (or $-5\mu m$ and $-15\mu m$). This however is offset by a $\sim 70\%$ smaller operational depth range. Consequently, one

concludes that the conjugate-masks method offers the poorest compromise between uncertainty and depth range. On the other hand, the biplane method shows the best compromise of all three disparity generations methods considered.

Finally, a comparison between the biplane CKM method and the established techniques introduced earlier -the astigmatic PSF and the MUM technique- will be made. Comparing Fig. 2.17f to Fig. 2.17c and Fig. 2.17d, it is immediately apparent that both the astigmatic PSF and the MUM display a σ_z approximately 50% smaller than the biplane CKM over a $\pm 5.0 \mu m$. Furthermore, the σ_x and a σ_y of the MUM technique are approximately 50% less than those of the biplane CKM, whereas the astigmatic PSF shows a slightly smaller lateral uncertainty compared to the biplane CKM. However, in contrast to the biplane CKM, both techniques display a rapid increase in uncertainty beyond this depth range. In fact, the uncertainty of both techniques in all three dimensions surpasses that of the biplane CKM within a depth range of $\sim \pm 7.0 \mu m$ and continues to increase rapidly with depth. One therefore concludes that the biplane CKM has scope if one is interested in performing 3D localisation over an extended depth range (by a factor of ~ 4 for the parameters considered here). Finally, in Section. 2.1, shifting the aperture function has been proposed as a means of increasing the responsivity of the cubic PSF without compromising the cut-off frequency. This is expected to result in a smaller σ_z. In order to verify this, the biplane method was simulated again with the aperture function shifted by $+1.30mm$ along the cardinal direction of the CPM. The resultant CRLBs are shown in Fig. 2.17h and as one can observe, the uncertainty in depth has been reduced by $\sim 50\%$ compared to the biplane with a centred aperture function. In fact, the biplane CKM with a decentred aperture has CRLBs comparable to those of the astigmatic PSF over a depth range three times as large. On the downside, besides the practical difficulties involved in displacing the aperture function of, for instance, an objective, the amount by which the aperture function can be displaced is also limited by the NA of the system; therefore, the applicability of this technique is limited.

In conclusion, these quantitative analyses confirm that amongst the three disparity generation methods considered, the biplane method offers the best compromise between performance and simplicity. Furthermore, for the considered parameters, we find that this method has a localisation depth range roughly four times as large as that of the astigmatic PSF and the MUM technique with lower CRLBs in all three dimensions of space outside a depth range of roughly $\pm 7.0 \mu m$. On the other hand, the biplane CKM with an aperture function decentred by $+1.30mm$ shows a σ_z comparable to that of the astigmatic PSF and the MUM technique and a slightly higher σ_x and σ_y over a localisation depth of approximately three times as large. We therefore conclude that there is scope for this technique mainly when the primary objective is 3D localisation over an extended depth range rather than an improvement in the 3D localisation precision.

2.6 Measuring disparity

In the previous sections, several ways of generating disparity have been proposed, analysed and the best performing one (the biplane method) identified. Focusing therefore exclusively on the biplane method, in this section, what is known as the "correspondence problem" in stereo vision will be tackled. The correspondence problem is defined in the field of stereo vision as the determination of the locations in each of the two camera images that are the projection of the same physical point in space [67]. Similarly, in CKM, we have two otherwise identical images with different local translations in proportion to the depth of the scene at that point. We then seek to measure the disparity between corresponding points in each image. This is essentially the same correspondence problem tackled in stereo vision, consequently, all the methods used in stereo vision can in principle also be adapted for CKM. Here we first consider extended scenes where the disparity is to be estimated at each point in the FOV, thus resulting in a dense depth map. We then turn our attention to measuring the disparity under the assumption of a sparse scene composed of points. In the latter, the disparity is to be estimated at the location of the points only, hence it is a simpler problem. Furthermore, this would be more akin to 3D localisation in microscopy rather than solving the correspondence problem in stereo vision. Conceptually, however, they are essentially the same problem, therefore they have been grouped together in this section.

2.6.1 Extended scenes

In a nutshell, to solve the correspondence problem, the two images must be compared by means of some similarity metric and corresponding points in each identified. Once this is done, the disparity between them can be measured in pixels. Myriad ways of doing this can be found in the extensive body of stereo vision literature. These can be broadly divided into two categories: local methods and global methods. In local methods, the image is segmented according to some scheme and then, each segment (or region) of the image is processed separately. Examples of local methods are block matching [68, 69], gradient-based methods (also known as optical flow) [70, 71] and feature matching [72, 73]. Conversely, global methods exploit constraints and criteria which are applicable to the entire image at once - hence the nomenclature. Some well known global methods include dynamic programming [74] and graph cuts [75, 76]. As we shall shortly discuss, the correspondence problem is ill-posed and cannot be solved in regions of the scene where there are no features or texture. Since global methods operate over the entire image at once, they are less sensitive to regional absence of features and texture than local methods. This is in fact their principal advantage and also the main purpose for which they were developed. However, this advantage is gained at a significantly greater computational complexity compared to local methods [77]. Given that this is a first attempt at the implementation of the CKM technique, it was

decided to focus on the simpler local disparity measuring techniques; however, global methods by themselves or in conjunction with local methods definitely merit further consideration.

As far as local methods are concerned, gradient-based methods (optical flow) are computationally simple, however they have a theoretical limit on their accuracy of half a pixel since the local derivatives are only valid over that range [67]. This excludes them from usage for our application because the disparities we intend to measure in CKM are of the order of a pixel per micron of depth on a $40\times, 0.5$NA system. Ways of circumventing this theoretical limit exist, but these methods involve interpolation or are otherwise iterative algorithms [78], and thus, the complexity of the algorithm is increased significantly. Feature-based methods result in depth maps of lower density compared to other local and global methods, however they are robust against discontinuities in the depth profile of the scene and against regions lacking in texture. Furthermore, these methods tend to be relatively efficient computationally therefore they are appealing for real-time implementations. They however require well identifiable features in order for segmentations algorithms to operate correctly. Given that the main application considered here is microscopy, where features in the sample can be faint, this requirement was deemed excessively limiting; hence, for a first attempt, the use of feature-based methods was not favoured.

This leaves block matching methods, which are the simplest of all methods considered and thus the most efficient. These methods estimate disparity at a given point in one of the images by seeking to match a small region (neighbourhood or support) around that point (referred to as the "template") to similar small regions in the other image known as the "search image". In order to perform the matching, a cost has to be computed using some metric. Commonly used metrics include: Normalised cross correlation (NCC), sum of absolute differences (SAD) and the sum of squared differences (SSD) [67]. The process of computing the cost over a neighbourhood of each point rather than exclusively at the point is referred to as cost aggregation and it increases the uniqueness of each point as well as the robustness of the algorithm to noise. Again, several different support functions can be found in the literature, ranging from the traditional convolution with a 2D Gaussian function to 2D shiftable windows of various kinds with fixed or variable size [77]. Here we focus on the traditional approach (*i.e.* convolution with a 2D Gaussian function), where the support is defined by the standard deviation or the full-width half-maximum (FWHM) of the Gaussian. This is mainly due to the simplicity and computational efficiency of this approach. The cost function that was used is the SSD; however, although the results presented here were obtained using this, several others were tested and no significant improvement in the depth maps was observed.

In formalising this, let $G_\sigma(x,y)$ be a 2D Gaussian function with a standard deviation of σ and a mean of zero:

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \tag{2.77}$$

where (x,y) are the space dimensions in a reference plane - here taken to be the original image plane at a focal distance of f_o (see Fig. 2.14). The sum of square differences in the reference plane is given by:

$$\begin{aligned} \beta(x,y,\rho) &= \left[i_{r+}\left(\frac{x_+M_+}{M}, \frac{y_+M_+}{M}, \rho\right) - i_{r-}\left(\frac{x_-M_-}{M}, \frac{y_-M_-}{M}, \rho\right)\right]^2 \\ \implies \beta(x,y,\rho) &= \left[i_{r+}(x,y,\rho) - i_{r-}(x,y,\rho)\right]^2 \end{aligned} \tag{2.78}$$

where assuming the focal lengths of the two focal planes of the biplane setup to be $(f_o \pm \delta)$ and the original image distance to be z_i, then, $M = {}^{z_i}/_{f_o}$, $M_+ = {}^{(z_i - \delta M^2)}/_{(f_o+\delta)}$ and $M_- = {}^{(z_i + \delta M^2)}/_{(f_o-\delta)}$ are the original magnification and the magnification of each channel respectively. (x_+, y_+) are the image coordinates of the focal plane with a positive defocus offset, (x_-, y_-) are the image coordinates of the focal plane with a negative defocus offset and (x,y) are the image coordinates of the original focal plane which is here taken as the reference plane. $i_{r+}(\cdot)$ is the image recovered from the image captured with a positive defocus offset, $i_{r-}(\cdot)$ is the image recovered from the image captured with a negative defocus offset and ρ is the axial position with respect to the reference focal plane at which the PSFs used to deconvolve the images were acquired. The cost function, $f(x,y,\rho)$, for any point, (x,y), in the reference image plane and a given ρ can therefore be expressed as:

$$f(x,y,\rho) = \iint_S G_\sigma(\gamma - x, \nu - y) \cdot \beta(\gamma,\nu,\rho)\, d\gamma\, d\nu \tag{2.79}$$

where S is a region of space. Notice that because $G_\sigma(\cdot)$ is a symmetric function, then, $f(x,y,\rho) = (G_\sigma * \beta)(x,y)\,|_\rho$. This allows the cost function to be computed very efficiently. Letting, therefore, the depth profile of the sample be $r(x,y)$ with respect to the reference focal plane, the defocus profile of the sample in either of the captured images, $W_+(x,y)$ and $W_-(x,y)$, can be approximated by:

$$W_\pm(x,y) \approx \frac{(r(x,y) \mp \delta)\, p^2}{2\lambda(f_o \pm \delta)(f_o + r(x,y))} \tag{2.80}$$

where λ is the wavelength of light and p is the radius of the entrance pupil. Similarly, the defocus at which the PSF was acquired on each of the two imaging channels has been captured, ψ_{o+} and ψ_{o-}, can be approximated by:

$$\psi_{o\pm}(\rho) \approx \frac{(\rho \mp \delta)\, p^2}{2\lambda(f_o \pm \delta)(f_o + \rho)} \tag{2.81}$$

From Eq. 2.55, therefore, we can express the recovered images as:

$$i_{r\pm}(x,y,\rho) = \iint_S \left\{ I(x,y)\,\delta\left(\gamma - x - \frac{W_\pm(x,y)^2}{3\pi\alpha f_{c\pm}} + \frac{\psi_{o\pm}(\rho)^2}{3\pi\alpha f_{c\pm}}, \zeta - y - \frac{W_\pm(x,y)^2}{3\pi\alpha f_{c\pm}} + \frac{\psi_{o\pm}(\rho)^2}{3\pi\alpha f_{c\pm}}\right) * \ldots \right.$$
$$\left. \ldots * h_r(\gamma,\zeta,W_\pm(x,y),\psi_{o\pm}(\rho)) \right\} d\gamma\, d\zeta \tag{2.82}$$

where $f_{c\pm} \approx {}^{2p_e}/_{[\lambda(z_i \mp \delta M^2)]}$ with p_e denoting the radius of the exit pupil. From Eq. 2.80 and Eq. 2.81, we note that $\forall(x,y)$ such that $r(x,y) = \rho$, $W_\pm(x,y) = \psi_{o\pm}(\rho)$. Furthermore, from Eq. 2.54, the residual point spread function satisfies: $h_r(x,y,W_\pm(x,y),\psi_{o\pm}(\rho)) = \delta(x,y)$. From Eq. 2.82, therefore, it follows that $\forall(x,y)$ such that $r(x,y) = \rho$, $i_{r+}(x,y,\rho) = i_{r-}(x,y,\rho)$ hence $\beta(x,y,\rho) = 0$. On the other hand, $\forall(x,y)$ such that $r(x,y) \neq \rho$, disparity between $i_{r+}(\cdot)$ and $i_{r-}(\cdot)$ exists thus $\beta(x,y,\rho) > 0$ therefore the cost function yields a larger score. Assuming therefore that $r(x,y)$ is a single-valued, smooth (at least over the σ of $G_\sigma(\cdot)$) function, then, we conclude that $\forall(x,y)$ such that $r(x,y) = \rho$, the cost function $f(x,y,\rho)$ will yield a minimum. By minimising the cost function over a range of ρ values, $[\rho_{min}, \rho_{max}]$, then, the depth profile of the sample can be estimated. This can be expressed as:

$$r(x,y) = \underset{\rho\in[\rho_{min},\ \rho_{max}]}{\arg\min} \{f(x,y,\rho)\} \tag{2.83}$$

Here we are assuming that the cost function yields a global minimum $\forall(x,y)$ such that $r(x,y) = \rho$ however this omits the complex effects of $h_r(\cdot)$ in Eq. 2.82. In principle, the residual point spread function, $h_r(\cdot)$, which accounts for the post-recovery artefacts, could cause $f(\cdot)$ to yield a minimum when this condition is not satisfied. This is particularly relevant in areas of the image with no features and/or texture or with features and/or texture of comparable contrast to the artefacts themselves. For most practical situations, however, the impact of the artefacts on the score of the cost function is negligible compared to that of legitimate image features and texture, therefore this assumption is found to hold well. Furthermore, areas of the image with little texture and features are translation invariant. This implies that despite the fact that disparity exists between the images in these regions, the images would still be very similar (*i.e.* $i_{r+}(x,y,\rho) \approx i_{r-}(x,y,\rho)$) irrespective of ρ. As a result, the cost function is flat over ρ hence the depth in featureless and textureless regions cannot be reliably estimated. This problem is common to all techniques which rely on intensity matching to infer depth. Lastly, the assumption of a smooth $r(x,y)$ highlights the susceptibility of this method to rapid changes in depth (or equivalently discontinuities in $r(x,y)$) which typically, but not necessarily, accompany a change in intensity. This is because for points close to a discontinuity in depth, the neighbourhood over which the cost

function is evaluated would encompass a wide range of depths; thus, rendering the score of the cost function non-representative of the intended condition.

Once the depth profile of the sample has been estimated from Eq. 2.83, the intensity image, $i_{rec}(x,y)$, can then be reconstructed by stitching together parts of the two recovered images according to some scheme. In principle, $\forall(x_o,y_o) \in (x,y)$ such that $r(x_o,y_o) = \rho_o$, $i_{r+}(x_o,y_o,\rho_o) = i_{r-}(x_o,y_o,\rho_o)$. However, in practice, despite the fact that the CPM increases the DOF significantly, the more out-of-focus the sample is, the more suppressed the higher frequencies are. Since for $\rho_o > 0$, $i_{r+}(x_o,y_o,\rho_o)$ is more in-focus than $i_{r-}(x_o,y_o,\rho_o)$ and vice versa, then, we choose to compose the final intensity image from the most in-focus parts of both acquired images, and thus, exploit the fact that we have two focal planes in the biplane configuration to further improve the quality of the reconstruction. Note that for $\rho_o = 0$, then both images are equally out-of-focus, therefore either one of them can be used, or else their mean. Assuming that the mean image is used, this can be expressed as:

$$i_{rec}(x,y) = \begin{cases} i_{r+}(x,y,r(x,y)) & \forall(x,y) \;\; s.t. \;\; r(x,y) > 0 \\ [i_{r+}(x,y,r(x,y)) + i_{r-}(x,y,r(x,y))]/2 & \forall(x,y) \;\; s.t. \;\; r(x,y) = 0 \\ i_{r-}(x,y,r(x,y)) & \forall(x,y) \;\; s.t. \;\; r(x,y) < 0 \end{cases} \tag{2.84}$$

In conclusion, Eq. 2.83 and Eq. 2.84 constitute a 3D topographic image of a sample over an extended DOF.

2.6.2 Sparse point-scenes

For the case of a sparse scene or sample consisting of points of a size comparable to the resolution limit, the problem of measuring disparity is significantly simplified. This is because the features in the captured images -essentially PSFs- are predefined. Moreover, owing to the sparsity condition, localising them with adequate accuracy is also relatively easy. By far, the most commonly used techniques to achieve this goal are: least-square fitting (LSF) and Maximum likelihood estimation (MLE). These techniques entail fitting a model of the PSF to the acquired image of a fluorophore in order to estimate its position. Non-fitting algorithms such as virtual window centre of mass used in software tools such as QuickPALM [79, 80] which relies on centroiding and Fourier-based techniques which rely on phase estimation rather than position [81], can be found in the literature. These techniques have the advantage of not requiring a model and are also simpler and faster; however, they tend to achieve lower accuracies than their fitting counterparts. For this reason, they are mostly employed for applications such as deep-tissue imaging where the PSF is distorted due to aberrations and/or refractive index variations within the sample, and thus, it is difficult to model. Given that the main objective of this work is to compare this technique to other established

techniques and that fitting algorithms are by far the most frequently employed in these techniques, in this work, we focus exclusively on fitting algorithms.

Amongst MLE and LSF, the MLE is known to be the estimator which theoretically achieves the CRLB (maximum theoretical precision). In fact, in [82], where an exhaustive comparison of MLE and LSF can be found, MLE was found to yield precisions 25% closer to the CRLB than LSF for a wide range of SNRs. MLE, however, requires detailed knowledge of the noise sources of the detector and rigorous characterisation of their statistics. This can be particularly difficult if, as is the case here, an sCMOS detector is used because the pixels of an sCMOS detector have different noise statistics. On the other hand, LSF can be implemented with less detailed knowledge of the noise and is still widely used in the literature; hence, LSF was preferred.

LSF relies on the minimisation of the weighted squared-difference between a model of the data and the acquired noisy data. We can therefore define the cost function for the k^{th} particle as follows:

$$f_k\left(x_o, y_o, \sigma_x, \sigma_y, \theta, A, b\right) = \sum_{(x,y)\in w_k} \left\{ \frac{\left[I_o\left(x - x_o, y - y_o, \sigma_x, \sigma_y, \theta, A, b\right) - I_m\left(x, y\right)\right]^2}{I_o\left(x - x_o, y - y_o, \sigma_x, \sigma_y, \theta, A, b\right)} \right\} \tag{2.85}$$

where (x, y) are pixel indices; w_k is a window of a predefined size which encompasses the k^{th} particle ideally close to its centre; $I_o\left(\cdot\right)$ is the model of the data; and $I_m\left(\cdot\right)$ is the measured data including noise -*i.e.* a cubic PSF. In principle, one could localise a cubic PSF directly, however, this would require specialised algorithms. Here, the acquired cubic PSFs were deconvolved using a Wiener filter, thus reducing them to Gaussian-shaped points in order to be able to use conventional Gaussian-fitting algorithms. The denominator on the right hand side of the equation is the weight and it equates to the expected variance where a Poisson noise model has been assumed. The model of the data used here is given by:

$$\begin{aligned} I_o\left(x, y, \sigma_x, \sigma_y, \theta, A, b\right) &= Ae^{\left[-\left(ax^2 - 2bxy + cy^2\right)\right]} + b \\ a &= \frac{\left(\cos\theta\right)^2}{2{\sigma_x}^2} + \frac{\left(\sin\theta\right)^2}{2{\sigma_y}^2} \\ b &= \frac{\sin 2\theta}{4{\sigma_x}^2} + \frac{\sin 2\theta}{4{\sigma_y}^2} \\ c &= \frac{\left(\sin\theta\right)^2}{2{\sigma_x}^2} + \frac{\left(\cos\theta\right)^2}{2{\sigma_y}^2} \end{aligned} \tag{2.86}$$

which describes a generic elliptical Gaussian function centred at $(0, 0)$, where: A is the amplitude, θ essentially describes its orientation, σ_x and σ_y are the standard deviations and b is an offset, which, in this context would model the background. Notice that the smaller $I_o\left(\cdot\right)$ is for a given pixel, the higher the weight assigned to it. In this context, this implies that such a cost function weighs the tails of the PSF heavily. This would not be a problem if the data model, $I_o\left(\cdot\right)$, is a rigorous model of the data, as is often

the case for non-engineered (conventional) and non-aberrated PSFs. If, however, the model of the PSF is approximate, then, weighted LSF can be problematic. In this case, $I_m(\cdot)$, is a deconvolved cubic PSF for which an elliptical Gaussian function is clearly only an approximation. As a result, no weights were used [83]. The actual cost function used can therefore be expressed as:

$$f_k(x_o, y_o, \sigma_x, \sigma_y, \theta, A, b) = \sum_{(x,y)\in w_k} \left\{ \left[I_o(x - x_o, y - y_o, \sigma_x, \sigma_y, \theta, A, b) - I_m(x, y) \right]^2 \right\} \tag{2.87}$$

Letting, therefore, $\Theta = [x_o, y_o, \sigma_x, \sigma_y, \theta, A, b] \in \Theta_s$ where Θ_s is the parameter space, the position (centroid) of the k^{th} fluorophore or particle, (x_{ok}, y_{ok}), is given by:

$$(x_{ok}, y_{ok}) \in \Theta_k = \underset{\Theta\in\Theta_s}{\arg\min} \left\{ f_k(x_o, y_o, \sigma_x, \sigma_y, \theta, A, b) \right\} \tag{2.88}$$

where Θ_k are the fit parameters for the k^{th} particle. Eq. 2.88 was minimised using the well known Levenberg-Marquadt algorithm [84]. Initial (coarse) estimates of the particle location were obtained by thresholding the intensity and then performing a weighted-centroid on the detected points. Note that during the deconvolution operation, the image is band-limited; therefore, high frequency noise is reduced making intensity thresholding viable. The initial estimate of the background, b, was taken to be the mean of the entire image and the initial A was taken to be the mean of the four central pixels around the point given by the weighted-centroid. The orientation, θ, of the deconvolved point, $I_m(\cdot)$, depends exclusively on the orientation of the CPM which is fixed. As a result, an empirically determined fixed value of θ was used. This reduces the number of parameters in the optimisation, therefore it aids the algorithm to converge both faster and more reliably. Lastly, although the optimisation yields various other useful parameters such as $(\sigma_{xk}, \sigma_{yk}) \in \Theta_k$, here, only (x_{ok}, y_{ok}) were considered, while the rest were discarded. For instance, $(\sigma_{xk}, \sigma_{yk}) \in \Theta_k$ describe the ellipticity of the localised particle; therefore, from these parameters, it is possible to tell whether the localised particle is a single particle or not, and thus, to discard or retain the result accordingly. These considerations, however, are beyond the scope of the investigation conducted here, therefore no such checks were performed. Instead, appropriate measures such as low bead density were taken in the experiment in order to avoid such problems.

In conclusion, Eq. 2.88 gives a means of localising particles in an image. To determine the disparity as required by CKM, particles have to be localised in the two acquired images, and then, the difference between the positions of corresponding particles taken. Once the disparity is known, the depth and hence the lateral position of the particle can be calculated. The algorithm used to implement this will be discussed in the following sections.

2.7 Algorithmic representation of the CKM recovery process

After discussing in some detail the concepts behind this technique, in this section, the algorithms used to implement it will be discussed. As mentioned in Section 2.6, the procedure used to measure disparity for extended scenes is different from that used to measure disparity for sparse scenes consisting of sub-resolution points, although they are conceptually equivalent. For this reason, the two cases will again be treated separately.

2.7.1 Extended scenes

For extended scenes, the flow diagram representing the CKM recovery algorithm is shown in Fig. 2.18 where the quantities shown are matrices or in the continuous case, 2D functions of space. The notation: $Q\left(\langle condition \rangle\right) = p$, has been used in the figure to denote that those elements of matrix Q, for which the condition: $\langle condition \rangle$, is satisfied are assigned the value p, where p is a scalar. Similarly, the notation: $Q = P\left(\langle condition \rangle\right)$, implies that those elements of matrix Q for which the condition: $\langle condition \rangle$, is satisfied are assigned the same (or the corresponding) elements in matrix P where matrices Q and P have the same dimensions. The condition: $\langle condition \rangle$, is itself a matrix of the same dimensions as matrices Q and P generated by comparing element-wise two matrices of the appropriate dimensions.

The CKM recovery algorithm requires two sets of data which need to be pre-acquired. These are:

1. A set of PSFs acquired at definite intervals over the designed DOF. In Fig. 2.18, the size of the z-step is denoted by δz and the total number of acquired PSFs is $2k_m$. The PSFs are therefore referred to by an index, k, such that: $-k_m \leq k \leq k_m$ where $k \in R^+$. This index therefore denotes the depth at which the PSF has been acquired where the depth is given by: $k \cdot \delta z$

2. A set of calibration data which maps the two halves of the detector onto one another at every depth. This is necessary because the images acquired on either half of the detector will be misaligned. Consequently, unless corrected, the displacement between them will not reflect the translation experience by the cubic PSF in response to defocus, which is what we are ultimately interested in measuring. This calibration takes the form of a geometric transform of some kind. In Fig. 2.18, this is denoted as a piece-wise linear (PWL) transform, since this kind of geometric transform was found to yield best results.

From this data, the input data structures required by the CKM recovery algorithm are obtained. These are:

1. Two sets (one for each of the two acquired images of the scene) of $2k_m$ images of the PSF acquired at steps of δz along the designed DOF. These can be represented as 3D arrays: $PSF_+(N_r, N_c, k)$ and $PSF_-(N_r, N_c, k)$ where $-k_m \leq k \leq k_m$, $k \in R^+$; and, the size of the PSF image at each k (or depth step) in pixels is $N_r \times N_c$. From these, two corresponding sets of $2k_m$ OTFs can be obtained by simply taking the Fourier transform of each PSF set. These can also be represented as 3D arrays: $H_+(N_r, N_c, k)$ and $H_-(N_r, N_c, k)$ where $H_+(\cdot)$ and $H_-(\cdot)$ denote the OTFs corresponding to $PSF_+(\cdot)$ and $PSF_-(\cdot)$ respectively. For conciseness, these will be denoted as: $H_+(k\delta z)$ and $H_-(k\delta z)$ hereupon - that is, the OTF as a function of depth $(k\delta z)$ only.

2. A set of $2k_m$ PWL transforms which map each of the two acquired images of the scene onto each other at every depth step. This eliminates any shifts and distortion caused by misalignment and other geometrical effects leaving only the shift induced by the CPM which is what need to be measured. These can be represented as an array of PWL transform: $PWL(k\delta z)$ where $-k_m \leq k \leq k_m$, $k \in R^+$. In other words, for a given depth step k, or equivalently, an axial position (or depth) of $k\delta z$, the corresponding PWL transform which eliminates all stray shifts between the two acquired images at that depth is: $PWL(k\delta z)$.

The CKM recovery algorithm therefore proceeds as follows. First, two images of the scene (I_{c+} and I_{c-}) are acquired on either half of the detector as shown in Fig 2.14. These are then deconvolved by their respective PSFs at the depth of the current iteration $(k\delta z)$. The deconvolution is denoted in Fig. 2.18 by the inverse filters: $\frac{1}{H_+(k\delta z)}$ and $\frac{1}{H_-(k\delta z)}$. The recovered images (I_{r+} and I_{r-}) are then transformed using the PWL transform corresponding to the depth of the current iteration ($PWL(k\delta z)$) to eliminate stray translations caused by misalignment, as previously explained. The transformed images (I'_{r+} and I'_{r-}) are then used to compute the the sum of square differences as given by Eq. 2.78. Finally, the metric M for the current iteration is computed by aggregating (Eq. 2.79).

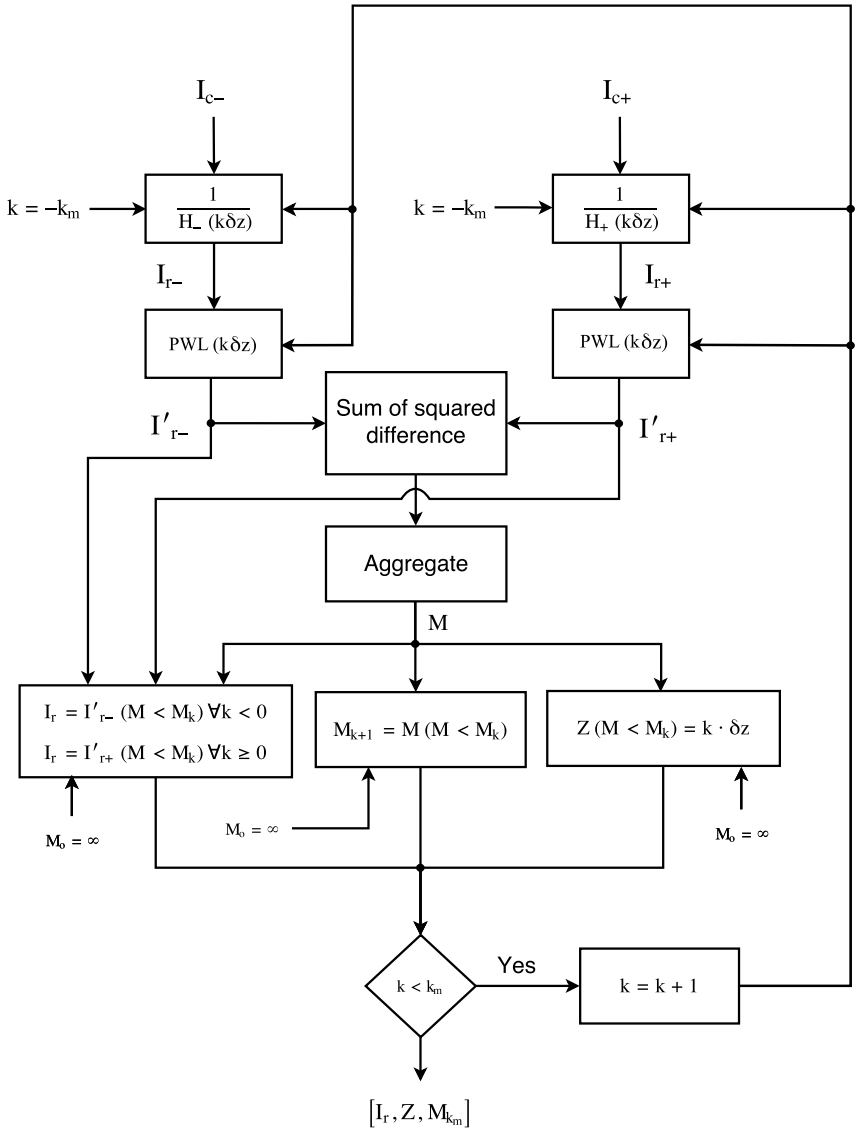


Fig. 2.18 Flow diagram of the CKM recovery algorithm for extended scenes

The next step is to compute the recovered image I_r according to the metric of the current iteration. This is done by keeping those pixels in the two recovered images where the metric M is minimum. After all iterations have been processed, I_r will be composed of pixels from I'_{r+} and I'_{r-} for which the metric yields a minimum over the entire depth range considered; hence, fulfilling Eq. 2.84. Note that the equations in Fig. 2.18 show a slight variation compared to Eq. 2.84 for the case $k = 0$; however, both versions are applicable and both yield very similar results. Next, the depth map (denoted by $Z(\cdot)$ in the figure) is updated by setting those pixels of the depth map for which the metric M is minimum to the depth corresponding to the current iteration - *i.e.* $\delta z \cdot k$ for the k^{th} iteration. After the last iteration, the depth map will consist of the depths at which the metric is minimum over the entire depth range considered; thus, as previously explained, the depth of the scene at each (x, y) point. This fulfils Eq. 2.83. Note that for obvious reasons, the depth map saturates at: $\pm k_m \cdot \delta z$; therefore, one must ensure that the depth of the scene lies within these bounds.

The final step in each iteration of the algorithm is to update the metric in order to ensure that the global minimum is obtained at the end of all iterations. This is done by comparing M to the metric obtained in the previous iteration and keeping only those pixels which have a smaller value in the current iteration. This is expressed in Fig. 2.18 as: $M_{k+1} = M\,(M < M_k)$. Note that the value of the metric itself is not utilised in the recovery process except to perform the minimisation. It can, however, be utilised in a post-processing step in order to identify regions in the scene where the depth cannot be reliably, inferred such as textureless or featureless regions. This is because for such regions, the metric will display a steady value over the entire depth range. This was not implemented in this work and is suggested as further work.

2.7.2 Sparse point-scenes

The flow diagram for the case of a sparse scene consisting of point sources is shown in Fig. 2.19. As in the case of extended scenes, some pre-measurements need to be taken in order to implement this algorithm. These are listed below:

1. The in-focus PSF of each of the two channels.
2. The alignment PWL geometric transform in order to align the two channels.
3. The relationship between generated disparity and actual depth which will then be used as a dictionary (or look-up table) in order to convert the measured disparity to depth. This is denoted by D_z in the figure.
4. The relationship between the generated disparity and the translation experienced by each cubic PSF in each channel with respect to its in-focus position. This is then utilised as a dictionary in order to compute the actual (x, y) coordinates of the emitter from the measured centroid of the point in each of the two acquired images. The dictionaries for the positive and the negative channels are denoted by D_{xy+} and D_{xy-} respectively in Fig. 2.19.

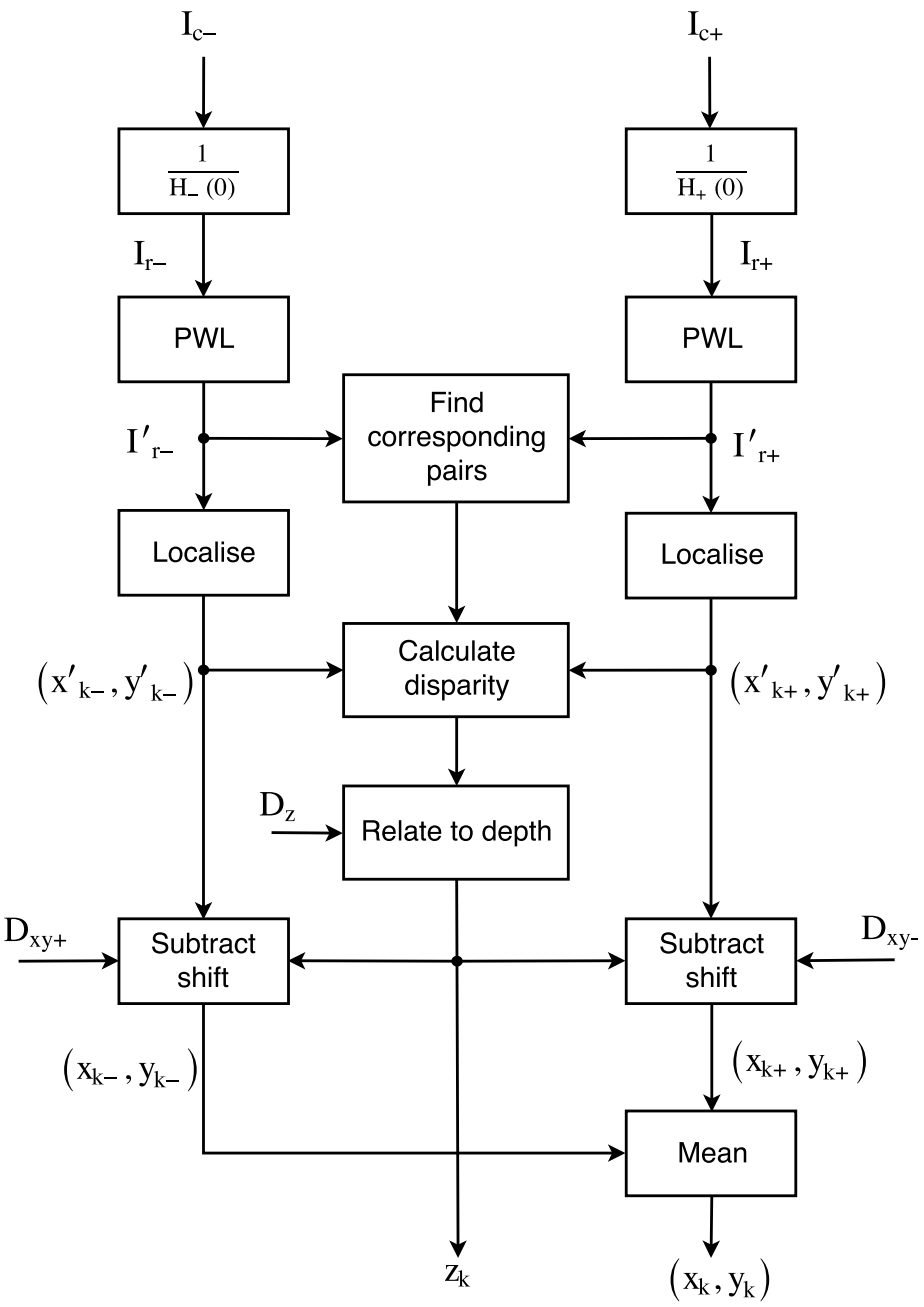


Fig. 2.19 Flow diagram of the 3D point localisation algorithm.

The process starts by the acquisition of two images of the scene on either half of the detector, as was the case for extended scenes. In this case, however, the recovered images are deconvolved by their respective PSFs. Moreover, since the features of the image (points) are known *a priori*, the deconvolution can be done by the in-focus PSFs only as denoted in the figure by the inverse filters: $\frac{1}{H_+(0)}$ and $\frac{1}{H_-(0)}$ where $H_+(\cdot)$ and $H_-(\cdot)$ denote the in-focus OTFs. The cubic PSFs in the captured images are thus reduced to Gaussian-shaped points in the recovered image which are easier to localise accurately. These are then aligned by means of a pre-measured PWL transform which, as already mentioned, eliminates translations and misregistrations caused by misalignment. Note that in this case, the PWL was measured only at the in-focus position. This was found to be adequate due to the additional calibration data D_z, D_{xy+} and D_{xy-}.

The next step is to determine the correspondence between the images of the same emitter in each of the two recovered images. To achieve this goal, a coarse affine transform obtained by matching intensities in each image was applied. This transform maps the images of the emitter in each image to one another albeit coarsely. Corresponding images will thus appear close to each other (typically within ~ 1.0 pixels); thus, allowing them to be paired based on a minimum distance criterion. The particles were then centroided by means of a weighted-centroid algorithm and the pairs of corresponding centroids were then transferred back to their original location by reversing the coarse affine transform. This eliminates any inaccuracies such a transform might introduce. Note that this method of establishing correspondence is sufficiently adequate in this case due to the assumption of a smooth depth profile. In any other eventuality, a minimum distance criterion alone cannot guarantee proper pairing. To solve this problem, other constraints can be applied such as a predefined direction (a curve or region) within which a corresponding pair must lie.

After correspondence has been established and the centroids of the image of the emitter in either image coarsely estimated, localisation is performed using the LSF method given by Eq. 2.88. The centroids obtained by the weighted centroiding performed in the previous step are used as the initial estimates of the centroid of each emitter in either image. This step yields an accurate estimate of the centroids $\left(\left(x'_{k-}, y'_{k-}\right)\right.$ and $\left.\left(x'_{k+}, y'_{k+}\right)\right)$ from which the disparity is then calculated. This is done by calculating the magnitude of the vector connecting the centroid of the emitter in one of the images to the other -*i.e.* $\left(\sqrt{\left(x'_{k+} - x'_{k-}\right)^2 + \left(y'_{k+} - y'_{k-}\right)^2}\right)$. Using the dictionary D_z and the disparity, the actual axial position, Z_k, of the emitter can thus be inferred. The depth, or the disparity itself depending on how the calibration has been performed, can then be used to find the amount by which the estimated centroids are displaced from the actual lateral position of the emitter. This is performed by means of the dictionaries D_{xy+} and D_{xy-} which relate disparity or depth to the shift of the individual cubic PSFs. After applying this correction, the resultant centroids $((x_{k-}, y_{k-})$ and $(x_{k+}, y_{k+}))$ of the image of the emitter in either image will be the same within the limits of precision; thus, the position of the emitter can be estimated by taking the average of the two. Alternatively, given that an estimate of the depth is available at this point, the lateral position of the emitter could be estimated from the more in-focus image only. The latter was found to yield a slight improvement in precision at large defocus values.

In conclusion, at the end of this process, an estimate of the 3D position (x_k, y_k, z_k) of the k^{th} emitter is obtained. Except for the deconvolution and alignment step, the process would then have to be repeated for all other emitters in the FOV. This, however, can be largely parallelised; therefore, it can be performed in a reasonable time frame, especially if coded in a low-level language such as C.

2.8 Conclusions

In this chapter, the theoretical aspects of this work have been investigated. The cubic PSF has been thoroughly modelled in the space domain using a novel approach based on the Airy transform. This led to a closed-form representation of the cubic PSF under the paraxial approximation only, which should be fit for publication. Furthermore, through this model, it was confirmed that the cubic PSF does not experience any additional translation besides those predicted by conventional approximate models. This theoretical framework also revealed that the morphology of the cubic PSF and the amount of translation it experiences are decoupled and that they can be controlled separately by decentring the aperture function. Based on simulations, it has been shown that this can be used to amplify the shift of the cubic PSF without sacrificing SNR; a compromise which is crucial to the CKM technique.

The mathematical formulation of the CKM technique was discussed next and several ways of generating the necessary disparity have been proposed and investigated. This investigation revealed that the biplane method is the easiest to implement, as well as the method which performs best. A statistical framework (CRLB) commonly used in localisation microscopy and other 3D imaging techniques was then utilised to compare the various disparity generation methods considered amongst themselves and also to compare them to other established 3D PSFs. Throughout this investigation, it was confirmed that the biplane offers the best compromise between DOF and precision. Furthermore, it was also confirmed that decentring the aperture increases the information content (i.e. lower the CRLB) as expected. This investigation also revealed that CKM is likely to display lower depth precision than the considered standard 3D PSFs for the same number of detected photons close to the focus. This is mainly due to the fact that cubic PSF is larger, therefore it exhibits a lower SNR. The lateral precision is also moderately lower close to the focus; however, it remains constant over a significantly larger DOF than the competition. Its main application would therefore be 3D localisation over a large depth range.

Finally, methods to measure the disparity for the case of extended scenes and sparse point-scenes, have been proposed and investigated. In this case, it was decided to focus on the most basic algorithms since this is a first attempt at the implementation of the technique in question. Furthermore, although satisfactory results have been obtained using these methods, several better alternatives or additions have been suggested. Lastly, the algorithms as implemented during the course of this work were presented and discussed.

Chapter 3

Proof-of-concept: numerical simulations and experiment

Given that no literature of direct relevance to this technique is available, an extensive investigation was deemed necessary in order to assess its feasibility before undertaking the actual implementation. With this goal in mind, numerical simulations were conducted, as presented in this chapter. These are intended to give an idea of what performance to expect from CKM under various scenarios such as lack of features and low SNR. Additionally, a proof-of-concept experiment was conducted in order to gain further insight into the practical difficulties involved in its implementation. This experiment was conducted before the emergence of the biplane CKM method, therefore the conjugate-masks method was used instead. Nevertheless, by means of the theoretical framework presented earlier, extrapolation of the results and conclusions to the biplane CKM technique is possible. The investigation presented in this chapter is for extended scenes only, which are the primary concern of this study. Sparse point-source scenes are considered in Chapter 4.

3.1 Numerical simulations

As stated, the principal objective of these numerical simulations is to investigate the viability of this technique under a low SNR and for scenes which lack features. Moreover, a brief comparison between the biplane CKM method and the conjugate-masks CKM method was made in order to confirm numerically the theoretical predictions made in the previous chapter. The conjugate-astigmatic-masks CKM was not considered because it requires specialised phase masks to implement and, as with the conjugate-masks method, replicating the aperture plane is still necessary. Consequently, from a practical stand point, the complexity was deemed to outweigh the advantage. The intensity of the simulated scene is shown in Fig. 3.1a which is an image of fungus spores. A ramp depth profile ranging from -2.5 to $+2.5$ waves from left to right as shown in

Fig. 3.1b was assumed. It should be noted that whereas the use of a natural intensity distribution (intensity image) avoids unnatural slopes and hence allows a more realistic assessment of the performance of the technique to be made, the fact that the depth profile does not correlate with the intensity image could be considered to be unrealistic. This is therefore to be considered as a limitation of the numerical simulations conducted here. The coded images were generated by point-wise convolving the scene with a PSF generated at a depth equal to the depth of the scene at that point. The PSFs were calculated by taking the Fourier transform of the appropriate pupil function; therefore, the paraxial approximation is implicit. A coding strength α of 3 waves was taken for all simulations and a difference in defocus of 2 waves was taken for the biplane CKM.

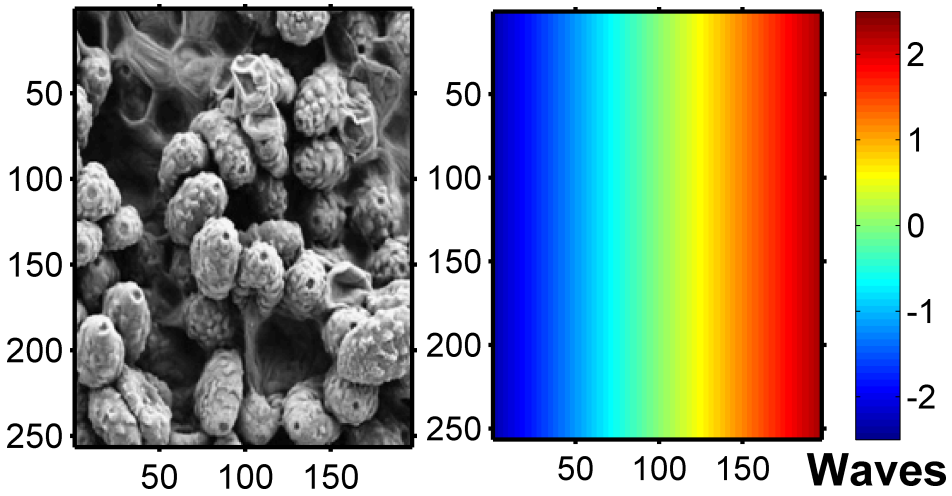


Fig. 3.1 Scene used for numerical simulations: (a) intensity image (fungus spores [85]) and (b) depth profile (ramp from left to right).

Consider first the comparison between the biplane CKM and the conjugate-masks CKM. For this purpose, the SNR is of no relevance as long as it is identical for both techniques, therefore no noise was added to the images. The image recovered using the conventional wavefront recovery procedure (*i.e.* single PSF deconvolution) is shown in Fig. 3.2a for reference. From this image, one can clearly observe the sinusoidal post-recovery artefacts discussed earlier. Besides, it is also evident that the image presents a significant amount of blur due to the mismatch between the recovery PSF and the actual PSF in different regions of the image. As a result, the quality of the recovered image is poor. Fig. 3.2c shows the CKM recovery given by the conjugate-masks method where a conspicuous improvement in image quality can be observed. The raw depth map given by the conjugate-masks method is shown in Fig. 3.2b. By comparing this to the actual depth profile of the scene, one can see that a substantial amount of noise is present in the depth map. This is expected and as discussed in Chapter 2, the cause of this lies in the fact that this technique is unable to differentiate the sign of the defocus. As a result, the measured magnitude of the defocus will be correct; however, the sign will oscillate randomly between negative and positive.

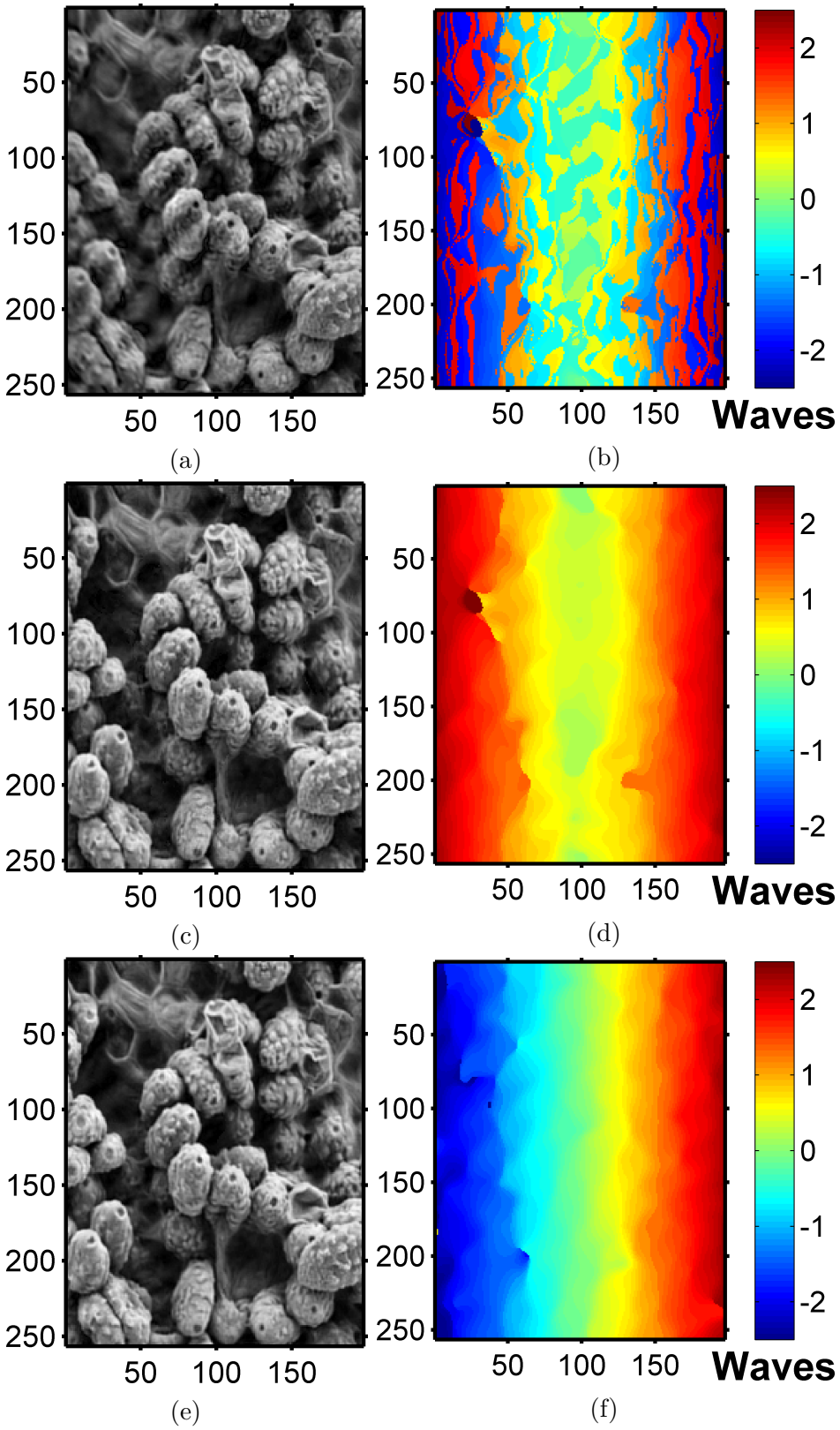


Fig. 3.2 Biplane CKM compared to conjugate-masks CKM: (a) conventional recovery, (b) raw depth map, (c) recovered intensity image and (d) absolute of the depth map given by conjugate-masks CKM. (e) recovered intensity image, and (f) raw depth map given by biplane CKM.

If one takes the absolute of the depth map, this noise is eliminated, as shown in Fig. 3.2d. This, however, also results in the loss of information about the depth of the scene. On the other hand, the biplane CKM is not prone to this ambiguity in sign, therefore

such difficulties should be eliminated. The recovered image given by the biplane CKM method and the raw depth map are shown in Fig. 3.2e and Fig. 3.2f respectively. As one can see, whereas the recovered image is comparable to that given by the conjugate-masks CKM, the depth map is superior because it does not display the mentioned noise. In addition, both negative and positive defocus in the scene are correctly identified, thus no information about the scene's depth is lost. These results are in accordance with the theoretical predictions presented earlier.

The last point which shall be considered is the effect of the different responsivity of the two techniques. As stated earlier, the conjugate-masks CKM has a responsivity of zero near the focus which results in its inability to infer depth at small defocus values. On the other hand, the biplane method has a constant responsivity therefore this problem is once more eliminated. To investigate this, the absolute value of the depth maps given by either method were averaged along the columns and plotted as shown in Fig. 3.3 together with the actual depth profile of the scene, which serves as a ground truth. As one can see from this figure, the conjugate-masks CKM does display a large error in depth over a defocus range of $\sim \pm 0.8$ waves as expected. Conversely, the biplane CKM follows the ground truth over the entire defocus range. The RMS error in depth for the biplane and the conjugate-masks methods were 0.1 and 0.14 waves respectively, therefore the conjugate-masks method gives an RMS error 40% larger than the biplane CKM. This is also in agreement with the theoretical predictions.

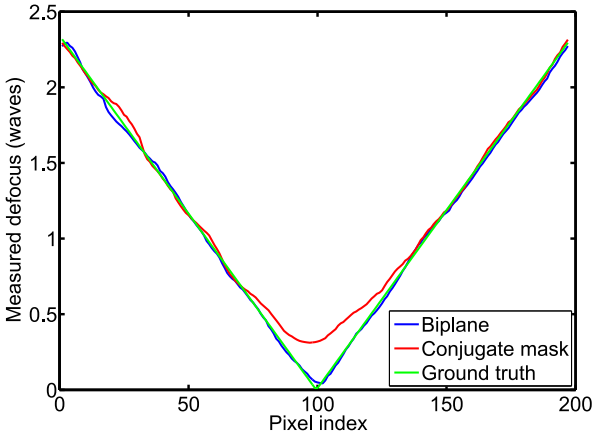


Fig. 3.3 Comparison of the depth profile given by the biplane CKM and the conjugate-masks CKM.

Next, the feasibility of this technique and its performance under low SNR conditions will be explored. To do this, Poisson noise was added to the image as per the following equation:

$$SNR = 20 \cdot \log\left(\sqrt{\bar{N}}\right) \text{ dB} \tag{3.1}$$

where $\bar{N} = k \cdot \bar{I}$ is the mean number of detected photons; k denotes the conversion factor of the detector in e^- per analogue-to-digital unit (ADU) and $\bar{I}$ denotes the mean intensity. Shown in Fig. 3.4a and Fig. 3.4b are the recovered intensity image and depth map using the biplane CKM technique (the conjugate-masks CKM will

not be considered any further) for an SNR of 30dB. As one can see, the recovered image displays a substantial amount of noise which is accentuated by the deconvolution process. The depth map, however, still reflects the true depth of the sample. The RMS error for an SNR of 20dB and 30dB was found to be 0.27 and 0.16 waves respectively. These correspond to an increase in RMS error of $\sim 170\%$ and $\sim 60\%$ respectively with respect to the ideal case ($SNR = \infty$). We therefore conclude that the CKM technique is robust against noise, although a significant increase in RMS error is to be expected.

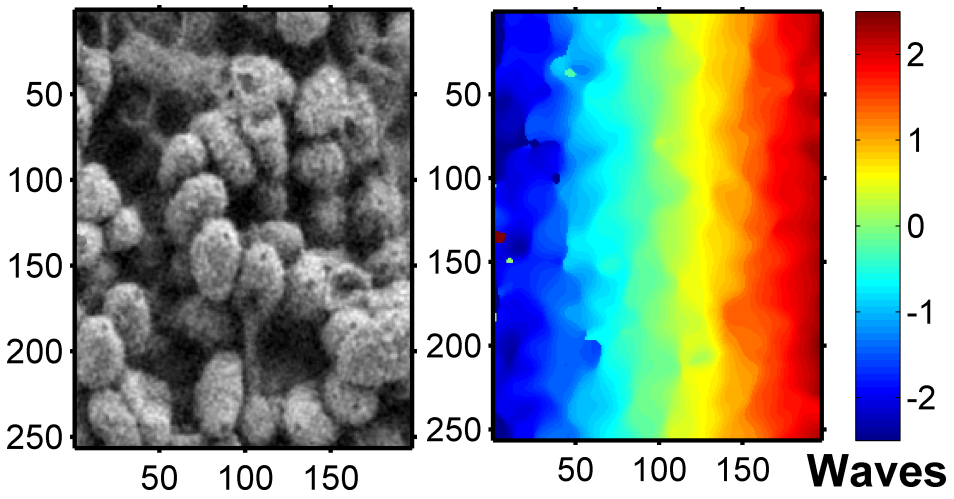


Fig. 3.4 Numerical simulation of CKM at low SNR (30dB): (a) recovered intensity image, and (b) depth map.

Finally, the susceptibility of this technique to the texture and features of the scene was investigated. As stated earlier, just like stereo vision and focus/defocus-based techniques such as Z-stacking and depth or shape from defocus, CKM's ability to infer depth depends on the amount and the quality of the texture and features in the scene. If none are present, then the correspondence between the acquired pair of images cannot be established and hence the depth cannot be measured. On the other hand, the more texture the scene has and the higher the contrast, the easier it is to establish correspondence, hence the smaller the RMS error of the depth map will be. Furthermore, unlike the other techniques mentioned, in CKM, errors in the calculated depth have the additional repercussion of lowering the quality of the recovered image due to suboptimal deconvolution. In order to investigate this, the intensity in the central region of the scene was set to its mean intensity. A pair of coded images was then simulated with an offset in defocus as required for the biplane CKM and ultimately recovered. No noise was added in order to isolate the effect which lack of features has on the technique. The recovered intensity image is shown in Fig. 3.5a and the corresponding depth map is shown in Fig. 3.5b. As one can see, in regions of the scene where there are features, the quality of the recovered image still shows a significant improvement over the conventional WC recovery. However, in the central region where the scene has no features, artefacts are clearly visible. This implies that as expected,

CKM is appreciably less effective in removing artefacts in areas of the scene where features and/or texture are lacking. This can also be observed from the depth map where an error in the inferred depth can easily be observed close to the centre of the scene. The RMS error in this case is 0.21 waves, which corresponds to an increase of 112% with respect to the ideal case without the features of the central region of the scene blacked out.

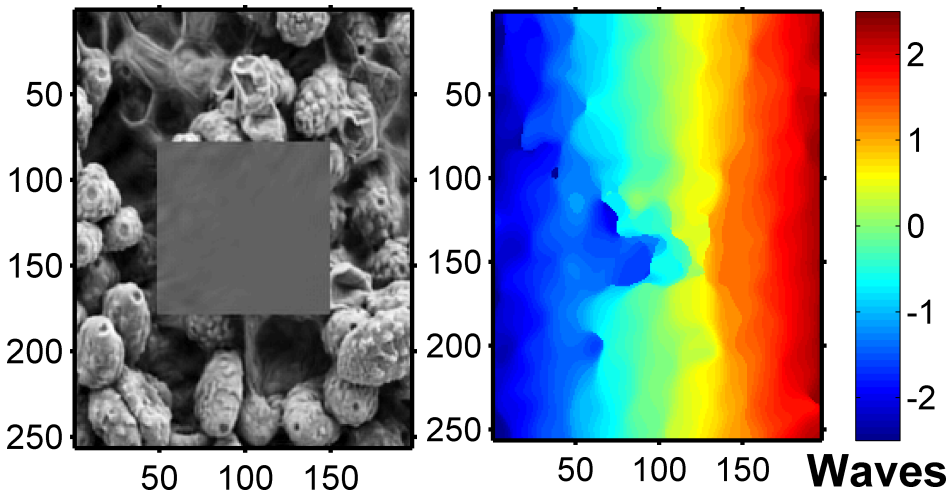


Fig. 3.5 Numerical simulation of CKM for texture-lacking scenes: (a) recovered intensity image, and (b) depth map.

3.2 Proof-of-concept experiment

As has already been mentioned, the objective of this experiment is to gain insight into the difficulties of implementing this technique in practice. It also serves the purpose of experimentally confirming the theoretical predictions. Furthermore, these investigations were conducted using the conjugate-masks method because the biplane method was developed later. As a result, the problem of rotating the CPM in order to generate the complex conjugate of the pupil function had to be tackled. This can be done by replicating the pupil or by mechanically rotating the CPM in a highly repetitive manner; however, both such techniques are quite complicated or necessitate precision positioning equipment. A simpler way of achieving this goal is to use an SLM. An SLM provides a means of electronically rotating the CPM without physically moving any part of the setup itself, thus eliminating the need for precision positioning equipment and ensuring repeatability. Furthermore, this also provides the flexibility of changing the focus of the system electronically, as well as to experiment with different coding strengths in order to establish the most appropriate value. The SLM used in this experiment was a Holoeye LC2002, which is of the thin twisted nematic type operable in transmission. The detector used was a Hamamatsu Orca CCD (C4742-95).

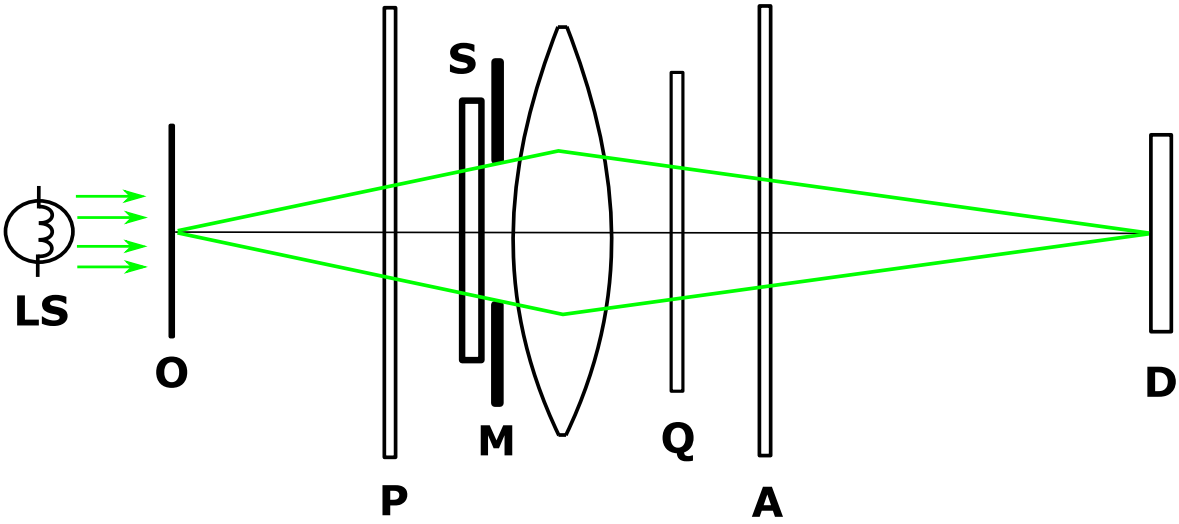
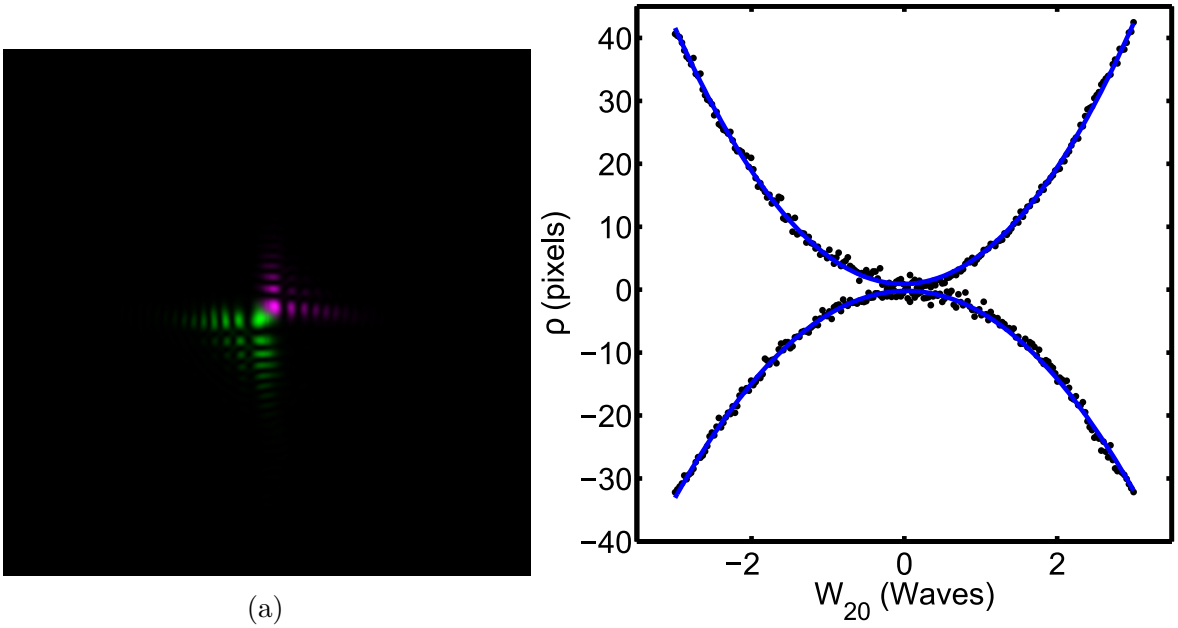


Fig. 3.6 Experimental setup for the proof-of-concept experiment. **LS**: fibre-coupled tungsten light source filtered at $540nm$, **O**: sample or object plane, **P**: polariser, **S**: spatial light modulator, **M**: aperture mask (iris), **Q**: quarter-wave plate, **A**: analyser and **D**: detector (Hamamatsu C4742-95).

Fig. 3.6 shows a schematic of the experimental setup which is essentially a single-lens finite conjugate system with a nominal magnification of approximately 5. The SLM requires some ancillary optics to operate; namely: a polariser, an analyser and a quarter-wave plate. The function of these is discussed in further detail in the following section where the procedure used to calibrate the SLM is covered. Given that the SLM has a rectangular aperture, an iris was used to produce a circular aperture. This and the SLM were placed in close proximity to the lens (f/15 singlet lens of focal length 300mm) in order to act as the aperture stop of the system. The light source used was a fibre-coupled white tungsten light source which was then filtered at $540nm$. The filtering is primarily required because the calibration of the SLM is strictly only valid at the calibration wavelength. To a lesser extent, this is also necessary because WC is in principle a monochromatic technique, although it can still be employed satisfactorily for narrowband applications.

Following calibration of the SLM, CPMs with an intended coding strength, α, of 3 waves were generated and a set of 200 PSFs was acquired for each rotation at different axial position. The focus was varied in steps of 0.03 waves over a range of ± 3 waves by introducing a quadratic phase profile of the appropriate magnitude in the aperture stop of the system by means of the SLM. The PSFs were acquired by imaging a $1\mu m$ pinhole illuminated by a focused laser beam of the appropriate wavelength. A laser is necessary in this case in order to acquire the PSFs at a sufficiently high SNR. This is important because these PSFs will be used to perform the recovery. After acquiring the PSFs, the offset was removed and the relative shift between them was measured for each set by correlating all PSFs with the in-focus one. The acquired PSFs close to the focus position are shown in Fig. 3.7a and the measured shift for each set is shown in Fig. 3.7b.



(a) (b)

Fig. 3.7 Proof-of-concept experiment: (a) PSFs (green acquired with the pupil function (CPM) and purple with its complex conjugate (rotated CPM)) and (b) shift of the PSFs (shift of the purple PSF was taken as positive).

From 3.7a it is immediately apparent that the two PSFs acquired with the pupil function and its complex conjugate have other differences beside the expected 180° relative rotation. Furthermore, from Fig. 3.7b, one can also notice that the shift of the PSF acquired with the complex conjugate of the pupil function displays a larger shift. Given that the magnitude of the shift is inversely proportional to α and that the extent of the PSF in space is proportional to α, one concludes that both these observations indicate that the α of the complex conjugate pupil function is smaller than that of the pupil function. To confirm this theory, the coding strengths were measured using a wavefront sensor (Shack-Hartmann sensor) and the obtained results were found to be compatible with the observations. Despite this, this discrepancy was not corrected because in actual implementation, the pupil has to be replicated; therefore, such variations are likely to occur due to imperfect matching between the two channels. As a result, it was deemed beneficial to conduct the experiment in the presence of the discrepancy in question. This matter is treated in more detail in the following section.

From Fig. 3.7b, it is also possible to notice that the measured shifts (black dots) display a substantial amount of noise (or jitter), which is more detrimental to this technique than variations in parameters such as α. The cause of this is the refresh rate of the SLM. This implies that this jitter is dependent on the exposure time setting of the camera, which makes it difficult to calibrate for. Note that the actual implementation is envisaged to make use of a glass phase mask rather than an SLM, so this would not constitute a problem. In order to mitigate this, a quadratic polynomial was fitted to the measured shifts (blue curves in Fig. 3.7b) and the PSFs were then repositioned to lie on them. This eliminates the error associated with the refresh rate of the SLM from

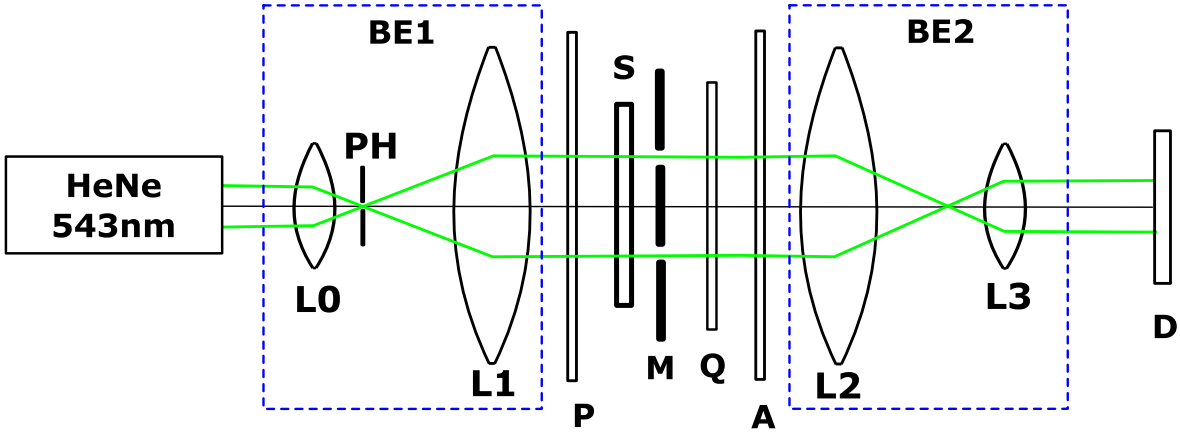
the measured PSFs but not from the subsequently acquired scenes, therefore it is not a definitive solution.

Once the PSFs were acquired and calibrated, two images of the scene were acquired in succession with the pupil function and its complex conjugate and deconvolved by their respective set of PSFs. The metric discussed in Subsection 2.6.2 was evaluated in each case and minimised over the depth range considered. This is essentially the same algorithm discussed in Subsection 2.7.2 with the exclusion of the alignment step (PWL). In this case, since the images were acquired in succession through the same imaging channel (*i.e.* exactly the same train of optics), it is safe to assume that no misalignment between the two acquired images exists. This is in fact an additional benefit of utilising an SLM and sequential acquisition to perform CKM. On the down side, the obvious disadvantage lies in the fact that this precludes single-snapshot operation which is one of the main advantages of this technique. This, however, is of no concern here because the primary objective of this experiment is to verify the feasibility of the technique.

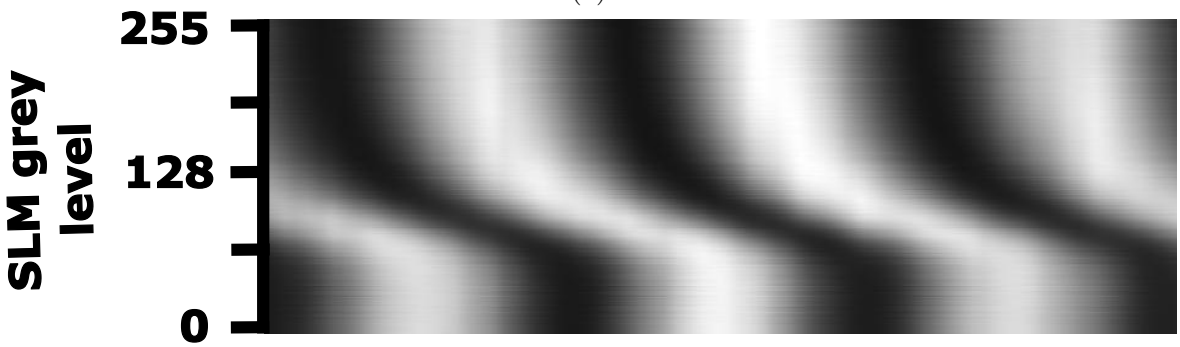
3.2.1 Calibration of the spatial light modulator

The SLM was calibrated using the method proposed in [86]. Here SLMs are of interest exclusively from an application point of view, therefore the mathematical framework (Jones matrices) describing their operation and other similar details were deemed beyond the scope of this work. Consequently, only those physical phenomena which have a direct impact on the operation of the SLM will be discussed.

An SLM is composed of a number of liquid crystal cells -pixels. Liquid crystals are birefringent and their birefringence, Δn, defined as the maximum difference between the refractive indices of the material is dependent on the potential difference applied across them. This is because applying a voltage across a layer of liquid crystals causes the molecules to align with the field, hence altering the refractive index. The retardation introduced by a slab (or a pixel) of such material is given by: $\Delta n \times t$ where t denotes the thickness of the slab or pixel. For twisted nematic SLMs, besides a change in birefringence, the polarisation of the light transmitted or reflected by the SLM is also dependent on the potential difference applied across the pixel. This is because, as the name implies, in twisted nematic SLMs, the molecules are arraigned in a helical fashion when the potential difference across the pixel is zero. As the potential difference increases, this helical arraignment starts unwinding, causing the polarisation of the output light to turn with the alignment of the molecules.



(a)



(b)

Fig. 3.8 Calibration of the SLM. (a) optical setup. **BE1**: beam expander 1 (**L0**: 0.4NA objective, **PH**: 1.0μm pinhole, and, **L1**: 30mm lens), **P**: polariser, **S**: spatial light modulator (twisted nematic), **M**: double slit mask, **Q**: quarter-wave plate, **A**: analyser, **BE2**: beam expander 2 (**L2**: 300mm lens, and, **L3**: 75mm lens), and, **D**: detector (Hamamatsu C4742-95). The spatial light modulator and the double slit mask should be placed as close to each other as possible. (b) detected fringes as a function of grey level. (Diagrams are not to scale)

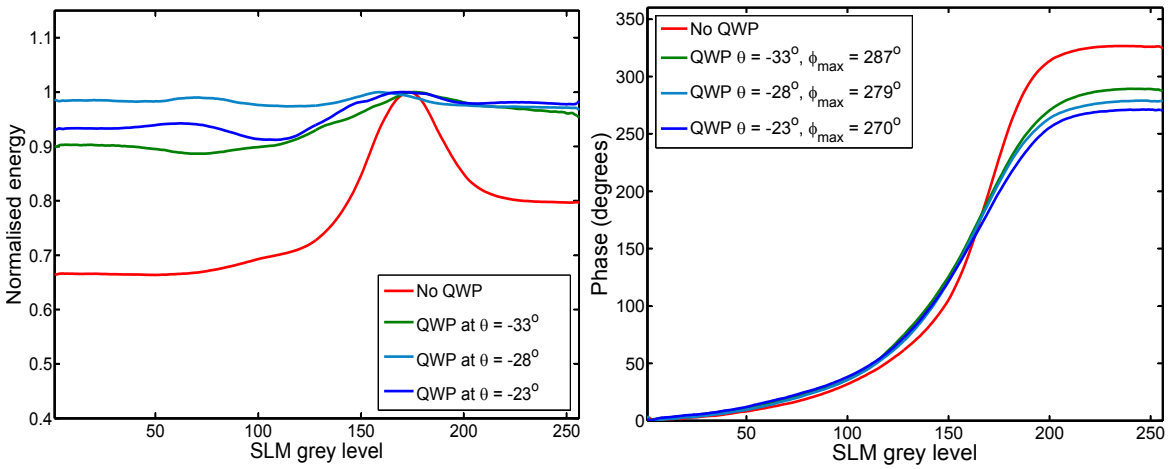
In this case, the SLM used is a transmission-type, thin, twisted nematic SLM. Twisted nematic type SLMs were originally designed as display devices for equipment such as projectors, therefore their primary use is for amplitude modulation rather than phase. In the past, these devices were manufactured with thick pixels, hence an adequate phase modulation depth could be achieved despite the fact that the change in birefringence is small. With the advent of technology, however, the same amount of amplitude modulation became possible with thin pixels and thus most modern twisted nematic SLMs are manufactured with thin pixels like the one used here. To achieve an adequate phase modulation depth with these devices, the SLM has to be sandwiched between a polariser and an analyser as shown in Fig. 3.8a [87, 88]. Doing this, however, results in coupling between amplitude and phase modulation [89] which is detrimental to application where phase-only modulation is required. This coupling is a consequence of the fact that the polarisation of light transmitted by each pixel of such an SLM depends on the potential difference across it, as previously discussed. Several methods have been proposed in the literature to mitigate this problem, such as the double quarter-wave plate arrangement [90, 91] and the more recent and simpler single quarter-wave plate configuration employed in [90]. Using this method, it is possible to maximise the phase

modulation depth while keeping amplitude modulation to minimal levels, thus achieving the necessary phase-only modulation. This is done by including a quarter-wave plate in between the SLM and the analyser as shown in Fig. 3.6 and Fig. 3.8a, and setting the angles of the polariser, the quarter-wave plate and the analyser appropriately.

Fig. 3.8a shows the setup used to calibrate the SLM - *i.e.* measure phase modulation as a function of grey level. Note that the grey level of a pixel of the SLM is proportional to the voltage applied across it. The angles of the polariser and the analyser were set to the experimental values determined in [86] for the same model of SLM as the one used in this experiment. In principle, the input-director angle of the SLM should be measured experimentally, as discussed in [92, 90], for optimal results; however, satisfactory results were obtained nonetheless. The SLM display was then split into two regions: a left region and a right region; and a double-slit mask was then placed close to the SLM as shown where the left slit coincides with the left region of the SLM's display and the right with the right. A $543nm$ HeNe laser was expanded to an appropriate size using a beam expander and spatially filtered (BE1) and this was then used to illuminate the SLM. This constitutes a double-slit interferometer, where by changing the grey level of one of the regions of the SLM with respect to the other, the path length difference between the light emerging from either slit is changed. Finally, a second beam expander (BE2) was used to magnify the resultant interference pattern before detection. The resultant interference pattern for different relative grey levels (between the left and the right halves of the SLM) is shown in Fig. 3.8b.

The angle of the quarter-wave plate was determined empirically by minimising the amplitude modulation over the entire range of grey levels (0 - 255 since each pixel is addressed by an 8-bit number). This was done by sweeping the grey level of the right half of the SLM display in steps of ten, while keeping the grey level of the left half at zero and acquiring an image of the resultant interference pattern for each grey level setting. By summing each acquired frame, the total detected energy can be estimated. This was repeated for a range of quarter-wave plate angle settings and the quarter-wave plate was then set at that angle which displays the least variation in detected energy over the entire grey level range. This is shown in Fig. 3.9a where the shown angles of the quarter-wave plate are with respect to the angle of the input-director, which has been assumed to be 45° to the horizontal (in [86], this angle is given as 46.5°) where positive angles are clockwise. From Fig. 3.9a, it can easily be seen that the amplitude modulation is minimised when the quarter-wave plate is set at an angle of $-28°$. In this case, the maximum variation in energy was found to be $\sim 5\%$. In Fig. 3.9a, the cases where the angle of the quarter-wave plate was set to $-28° \pm 5°$ are also shown and as one can observe, even a small error in the angle of this component significantly increases the amplitude modulation. Finally, the case where the quarter-wave plate has been omitted was also included in Fig. 3.9a, in order to demonstrate the coupling

between phase and amplitude modulation displayed by twisted nematic SLMs when sandwiched between a polariser and an analyser.



(a) (b)

Fig. 3.9 SLM calibration: (a) amplitude modulation and (b) phase modulation against grey level setting on the SLM. θ denotes the angle of the slow axis of the quarter wave plate and ϕ_{max} denotes the maximum phase modulation attainable. The angles are measured with respect to the angle of the SLM's input-director ($\sim 45°$). Negative angles are anti-clockwise.

Once amplitude modulation has been minimised, the phase modulation as a function of SLM grey level is to be measured. As previously mentioned, changing the grey level of one of the regions of the SLM while keeping the grey level of the other fixed at zero introduces a relative phase difference between the wavefront of the light emerging from the slits. As a result, the interference fringes will translate laterally, as shown in Fig. 3.8b, for all grey levels. From basic interferometer theory, we can therefore calculate the phase introduced by the SLM for each grey level from:

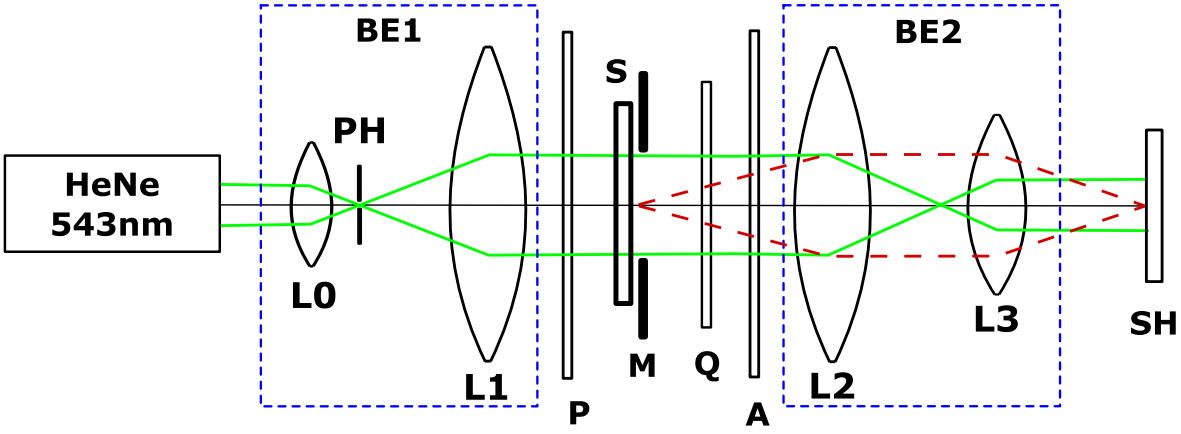
$$\Phi_g = \frac{2\pi y_g}{w} \tag{3.2}$$

where y_g is the lateral translation of the fringes with respect to the position of the fringes when the grey level of both halves of the SLM is the same -*i.e.* zero phase difference- and w is the period of the fringes which should not change with grey level. By sweeping the grey level from 0 to 255 and measuring the period and position of each maxima in the resultant interference pattern from which y_g can then be calculated, the phase delay introduced by the SLM for each grey level was deduced. This is shown in Fig. 3.9b, where, as one can see, the maximum phase delay which can be introduced for the optimal quarter-wave plate angle ($-28°$) is 279° ($\sim {}^{3\pi}/_{2}$ rad), which is consistent with what is reported in [86]. Consequently, any phase greater than this value has to

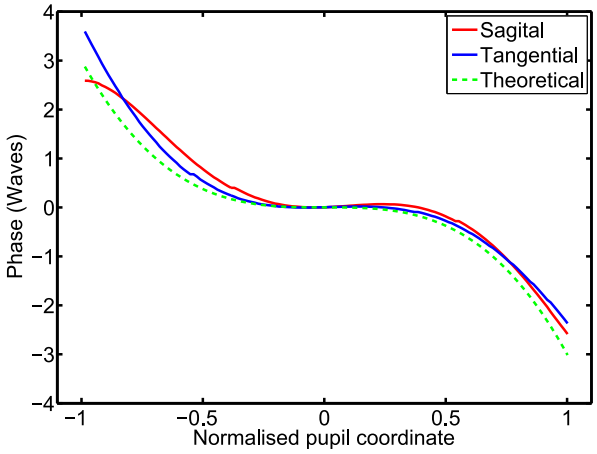
be rounded. This was done according to the equation below:

$$\theta_d = \begin{cases} \theta_d \text{ if } 0° \leq \theta_d \leq 279° \\ 279° \text{ if } 279° < \theta_d \leq 319° \\ 0° \text{ if } 319° < \theta_d \leq 360° \end{cases} \tag{3.3}$$

Finally, the cases where the quarter-wave plate was set to an angle of $-28° \pm 5°$ and the case where no quarter-wave plate was used are also shown to illustrate the effect of the quarter-wave plate on the phase modulation depth. A maximum phase modulation depth of $\sim 320°$ is obtained when no wave plate is used. From this, it is clear that by introducing a quarter-wave plate, phase modulation depth is sacrificed in favour of a reduced amplitude modulation.



(a)



(b)

Fig. 3.10 Testing the SLM. (a) SLM test setup. **BE1**: beam expander 1 (**L0**: 0.4NA objective, **PH**: 1.0μm pinhole, and, **L1**: 30mm lens), **P**: polariser, **S**: spatial light modulator (twisted nematic), **M**: aperture mask, **Q**: quarter-wave plate, **A**: analyser, **BE2**: beam expander 2 (**L2**: 300mm lens, and, **L3**: 75mm lens), and, **SH**: Shack-Hartmann wavefront sensor (WFS150). The spatial light modulator and the aperture mask should be placed as close to each other as possible. (Diagram not to scale). (b) measured sagital and tangential phase profiles for a CPM with $\alpha = 3$.

This was deemed advantageous because, given that the cubic phase profile is smooth, rounding is not expected to have a severe impact. On the other hand, since CKM relies on intensity matching between two images acquired with different coding, the effect of the amplitude modulation in the aperture stop can be significant.

The final step is to test the calibration of the SLM. Given that the SLM in this case will be used to generate the necessary cubic phase profile necessary for a CPM, it was deemed appropriate to test the SLM on this phase pattern directly. The optical setup used to achieve this goal is shown in Fig. 3.10a.

Comparing this to the calibration setup shown in Fig. 3.8a, one can see that the only difference is the use of a wavefront sensor (also known as a Shack-Hartmann (SH) sensor) instead of a detector. The SH is to be illuminated with a collimated light beam. This was generated by expanding a $543nm$ laser beam using a beam expander and spatial filter (BE1). A second telescope or beam expander (BE2) was then used to resize the beam in order to fill most of the area of the SH sensor. The focal lengths of the lenses of this beam expander were chosen so that the SH is in a plane conjugate to the aperture stop (*i.e.* the SLM) of the system. To do this, the aperture stop and the SLM were laterally illuminated by a diffused light source and the SH sensor was then positioned in the plane where a sharp image of the SLM's dark regions (inter-pixel regions) and the aperture mask were visible. This ensures that the wavefront at the aperture stop is the same as that sensed by the SH.

A CPM with an α of 3 was therefore generated and the corresponding grey levels necessary to implement it on the SLM were obtained from the calibration curve (Fig. 3.9b). Fig. 3.10b shows the measured phase profiles along the vertical and the horizontal directions. Also shown in Fig. 3.10b is the theoretical (or expected) phase profile, which should be identical in both directions. Note that the wavefront sensor used decomposes the incident wavefront into Zernike polynomials and outputs the Zernike coefficients. For this reasons, the discontinuities in the phase profile produced by rounding (Eq. 3.3) are not visible in the plots. Furthermore, due to the alignment of the setup, the measured wavefront will include tilts which are not actually generated by the SLM, thus the peak aberration cannot be measured directly. This, however, can be easily estimated by fitting a cubic polynomial to the measured phase profile and then estimating α from the cubic term. This was done and the measured coding strength in the vertical and in the horizontal directions of the CPM were found to be 3.04 waves and 3.17 waves respectively. The corresponding values for the rotated CPM were found to be 3.98 waves and 2.49 waves.

From these measurements, one concludes that there is significant variation in the coding strength of the intended mask and that which is actually generated, particularly for the non-rotated one. It is interesting to note that the error in the measured α becomes appreciably smaller for CPMs with larger α values. This suggests that other (fixed) aberrations in the system are not negligible compared to a CPM with an α of 3

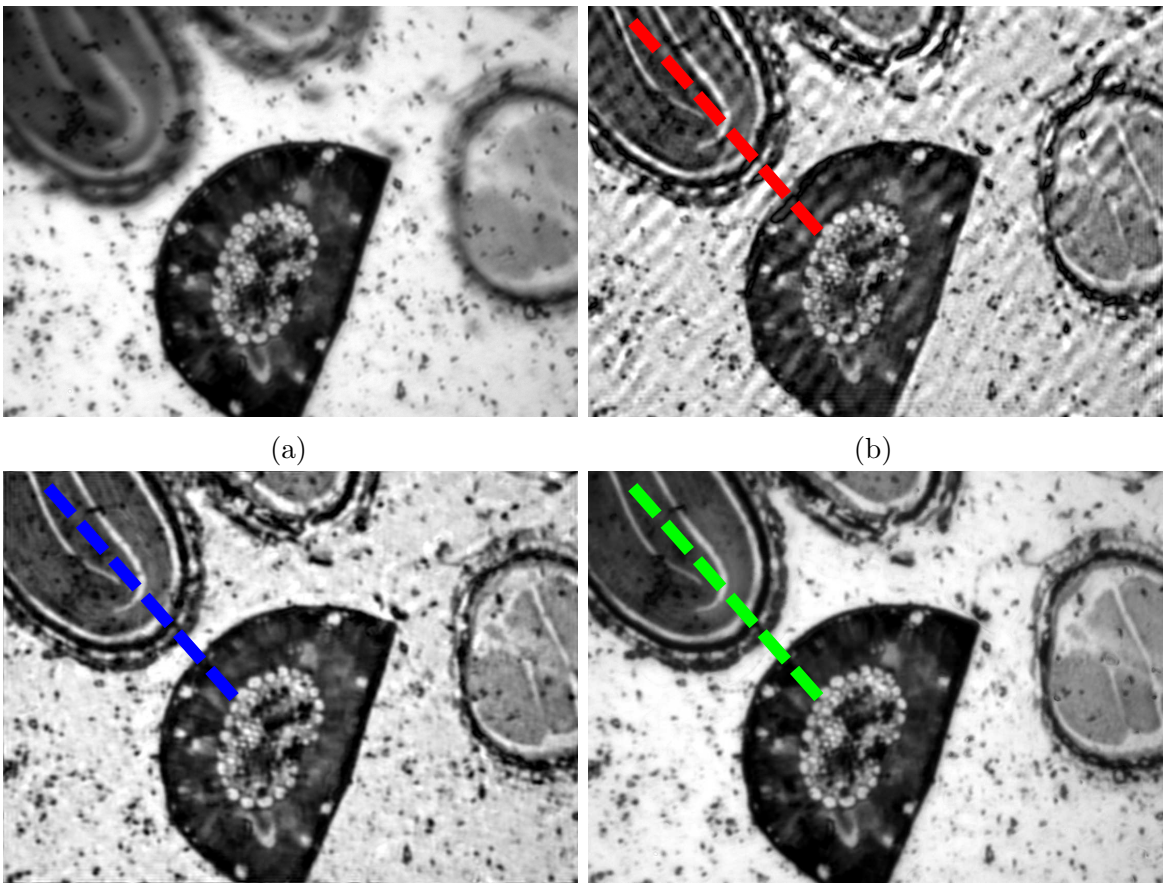
waves. As α is increased, the CPM aberration dominates; therefore, the error in the coding strength becomes smaller. Furthermore, it should be stated that the calibration method used here operates on the fundamental assumption that all pixels of the SLM are identical and will respond in identical fashion. This follows from the fact that only those pixels which lie within the width of the two slits of the double-slit mask are actually calibrated. While this greatly simplifies the calibration procedure because it renders it insensitive to mechanical vibrations, air currents and other practicalities, it does not take into account regional variations in the response of the SLM. This is believed to be the major cause of the observed error in coding strength. Nevertheless, as previously stated, it is interesting to investigate the performance of CKM under such a mismatch between mask parameters because such variations are to be expected to some extent. As a result, no further calibration was undertaken. Finally, from Eq. 2.16 we know that the shift of the cubic PSF is inversely proportional to α. Therefore, from the measured α values, one can calculate that the difference in shift between the two PSFs at a defocus of ± 3 waves will be ~ 10.6 pixels. This is in good agreement with the shifts measured directly from the PSFs which are shown in Fig. 3.7b.

3.2.2 Results

The main objectives of this experiment are to demonstrate: (1) the feasibility of the CKM recovery algorithm on real images acquired on a real detector in the presence of noise and other practicalities; (2) the removal of the post-recovery artefacts associated with WC and hence the ability to recover an image of a quality superior to that achievable with conventional WC; and (3) the ability to recover the depth information of a sample for a pair of complementary coded images.

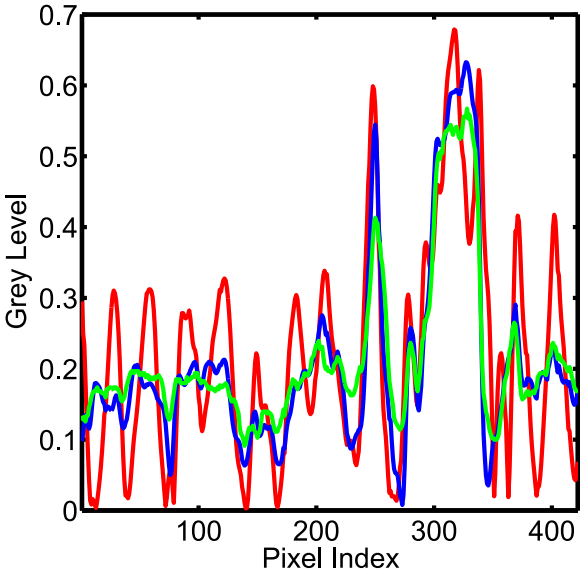
Shown in Fig. 3.11a is the NA limited image of a sample of a section of plant stem superimposed on another sample of a section of mustard seeds. These two samples were superimposed on each other to create a step-change in defocus, as one can observe from Fig. 3.11a, where the mustard seeds in the background are out-of-focus whereas the stem section is in-focus. Fig. 3.11b shows the image given by conventional WC. This was obtained by acquiring a coded image with the CPM and deconvolving it with the in-focus PSF only. As one can see, in this case we obtain strong post-recovery artefacts which, as expected, show up as low frequency sinusoidal variations in intensity. The image recovered by applying the CKM algorithm is shown in Fig. 3.11c. As one can easily notice, the artefacts have been completely eliminated. For comparison, a Z-stack was captured and the reconstruction is shown in Fig. 3.11d. This was obtained from 201 images captured at equally spaced intervals over a defocus range of ± 3 waves. The reconstruction was performed using a commercial Z-stacking software (Helicon focus V5.3). Furthermore, line profiles taken along the dotted lines shown on the conventional WC recovery, the CKM reconstruction and the Z-stack reconstruction are shown in Fig. 3.11e. From this, we conclude that the CKM reconstruction obtained

from two snapshots and which can be obtained from a single one as proven later in this thesis, is comparable to that given by a Z-stack from 201 snapshots. The line profiles further prove this point, although the Z-stack does show fewer variations than the CKM recovery. It is, however, hard to speculate on what the reason for this might be, because the Z-stack reconstruction algorithm is proprietary.



(a) (b)

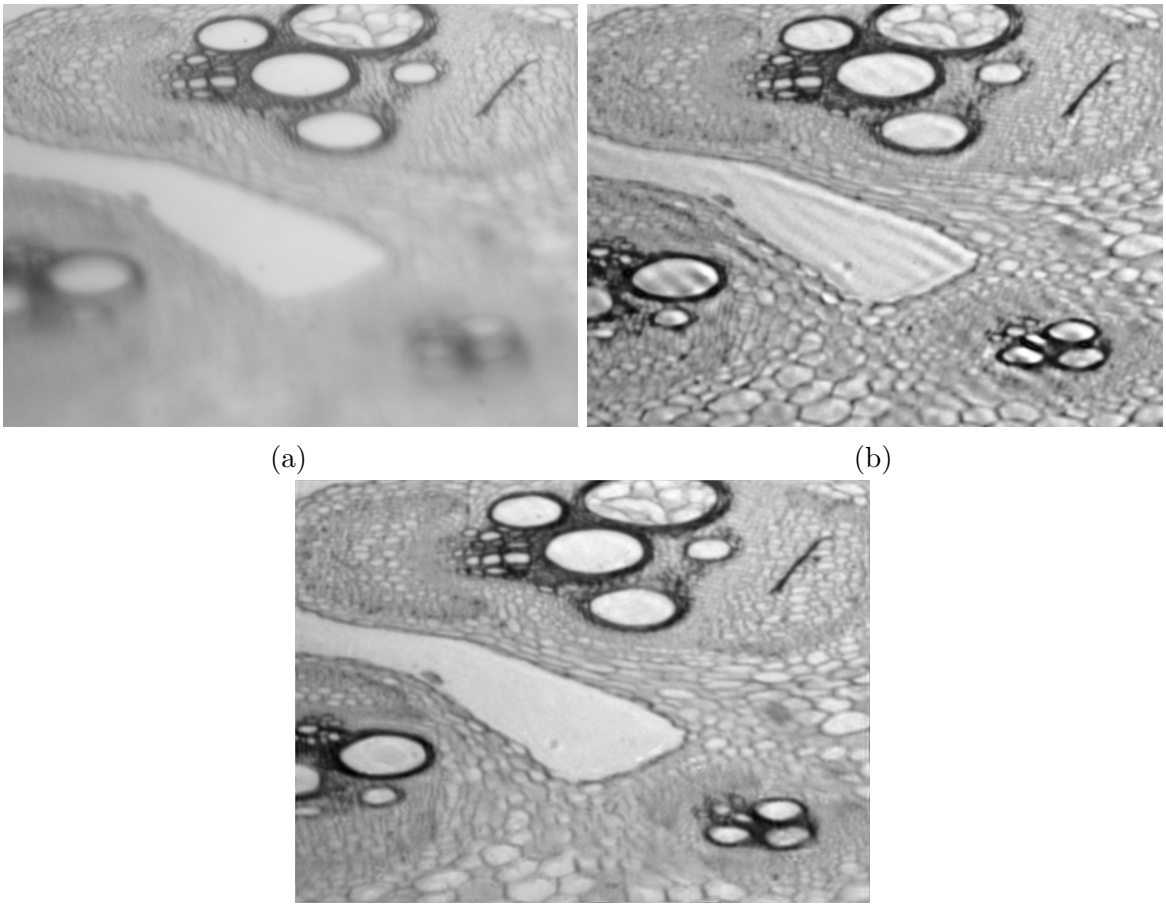
(c) (d)



(e)

Fig. 3.11 Artefact removal: stem section and seeds. (a) NA limited, (b) conventional wavefront coding recovery, (c) CKM recovery, (d) Z-Stack reconstruction and (e) line profiles taken along the lines of the corresponding colour shown in figures (b) to (d).

Fig. 3.12 is another example showing artefact removal by CKM. In this case, the sample is a tilted section of petiole. Unlike the previous sample where a step-change (or two level) depth profile was induced, in this case, a linear depth gradient was emulated by tilting the slide, as one can see from the NA limited image shown in Fig. 3.12a. Fig. 3.12b and Fig. 3.12c show the conventional WC recovery and the CKM recovery respectively. Again, one can see that the conventional single PSF recovery yields appreciable post-recovery artefacts which are not present in the CKM recovery. Note that the presence of wavefront coding artefacts in the image tend to give the impression that the image has higher contrast. This is likely because they are most prominent near the edges. In Fig. 3.12b and Fig. 3.12c, the contrast was adjusted such that the two images have identical contrast. We therefore conclude that the CKM algorithm is practical on real images and that it does yield the expected improvement in image quality by eliminating post-recovery artefacts.



(a) (b)

(c)

Fig. 3.12 Artefact removal: tilted petiole. (a) NA limited, (b) conventional wavefront coding recovery and (c) CKM recovery.

The final objective of this investigation is to explore the possibility of estimating or measuring the depth of a sample through this technique. This was done by imaging a systematic scene with well identifiable features. For this reason, a tilted dot-distortion target was used, as shown in the NA limited image illustrated in Fig. 3.13a. In this case, it was only possible to measure depth on either the negative or the positive defocus

range. This is expected because, as previously discussed, the conjugate-masks CKM does not permit discrimination of the sign of the defocus.

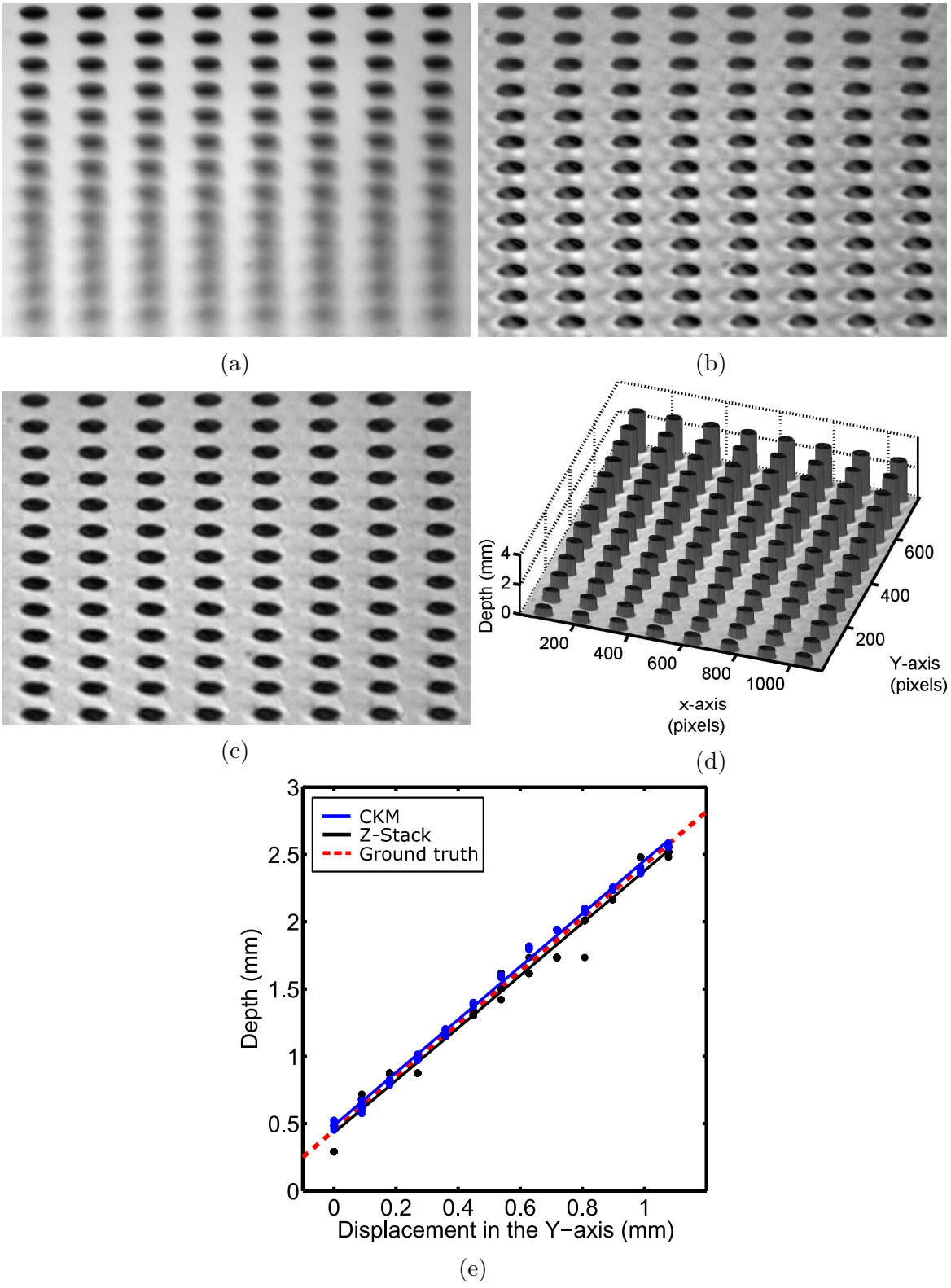


Fig. 3.13 Depth measurement: tilted distortion target. (a) NA limited, (b) conventional wavefront coding recovery, (c) CKM recovery, (d) processed 3D reconstruction (white regions were set to zero and depth was averaged over the area of each dot) and (e) slope estimates.

An attempt was made to apply the conjugate-masks CKM on a sample with a depth spanning both negative and positive defocus and the expected abrupt changes in the

depth map due to the fluctuations in sign were observed. Another interesting point is the fact that the error in depth close to the focus was found to be excessively large. These two points confirm the theoretical deductions presented in Chapter 2. The conventional single PSF WC recovery and the CKM recovery are shown in Fig. 3.13b and Fig. 3.13c respectively where artefact removal and the associated improvement in image quality are again evident.

As previously discussed, CKM relies on image features to infer depth. This implies that the depth map in regions where the image has no features, such as the areas of the distortion target where there are no dots and within the dots themselves is unreliable. To reduce the effect of this, only the depth of the dots was considered. This was done by averaging the depth map over the area of each dot and taking that value as its depth. The regions where there are no dots were then set to zero. The obtained 3D reconstruction using this method is shown in Fig. 3.13d for illustration purposes. Fig. 3.13e shows the depths of the dots as given by CKM and by a Z-stack for comparison. In this case, the Z-stack was performed using only a sharpness estimator (Sobel filter). The generated depth map was then post-processed in the same way as that generated by CKM. This allows us to compare the accuracy with which depth can be estimated from the sharpness of an image to that which can be obtained by measuring the translation of a cubic PSF without weighing in the effect of additional post-processing. A ground truth slope was then established by taking the ratio of the mean horizontal and the mean vertical separation between the dots. This was found to be 1.976 ± 0.002. A reliable measure of the slope can be obtained this way due to the relatively low NA of the system, which permits substantial tilt to be induced. The slope of the distortion target from the Z-stack and the CKM measurements was obtained by least-square fitting a line to the depths of the dots as shown in Fig. 3.13e. These were found to be 1.95 ± 0.02 and 2.00 ± 0.01 respectively, which are both consistent with the ground truth. Moreover, given that the error in the estimate of the slope given by CKM is $+1.2\%$, whereas that given by the Z-stack is -1.3%, we can further claim that in this case, CKM displays an accuracy comparable to that of a Z-stack. Of course, additional post-processing might improve the depth accuracy of both techniques. Finally, it should be noted that such a comparison is scene-dependent, since both techniques rely on the features and texture of the scene.

3.3 Conclusion

In conclusion, numerical simulations were conducted and it was confirmed that the biplane method is superior to the conjugate-masks method, which is in agreement with the theory. The biplane method was therefore adopted for the actual implementation, as discussed in Chapter 4. Furthermore, through numerical simulation, it was established that CKM is robust against noise and lack of features and texture in the scene despite

the fact that in both cases, a significant increase in the RMS error of the depth is to be expected. A lowering of the quality of the recovered intensity image is also to be expected in both cases. For low SNR conditions, the quality of the recovered image is degraded by noise amplification following deconvolution. For scenes lacking features and texture, besides the noise amplification, the quality of the recovered image is degraded further by suboptimal recovery, which results in the incomplete removal of the post-recovery artefacts.

A proof-of-concept experiment was also conducted which confirmed that CKM is feasible on experimental images in the presence of optical aberrations, mismatch between the coding strengths of the CPM in the two imaging channels and other practicalities. Artefact removal has been demonstrated experimentally and the error in the slope of a tilted dot distortion target was found to be 1.2%, which is comparable to that given by a Z-stack. This experiment also demonstrates that CKM can be satisfactorily implemented using experimental PSF, which is an important consideration. Although the experiment was conducted using the conjugate-masks CKM, it is safe to assume that the conclusions also apply for the biplane method especially since both numerical simulations and theoretical considerations suggest that the latter should perform better. Finally, given that this experiment was performed by acquiring the two required images of the scene sequentially, single-snapshot operation was not investigated. This, therefore, is the only major component of this technique which was not tested during this phase of this work. This was eventually implemented successfully, as discussed in Chapter 4.

Chapter 4

Experimental verification of the biplane CKM technique

In this chapter, the experimental verification of the biplane CKM method is presented in the context of microscopy. Results for both extended imaging and point-localisation performed on synthetic samples are presented. The optical setup is discussed first in Section 4.1 because this is common to both extended imaging and point localisation. The calibration procedure used for extended imaging is discussed next, to which an individual section (Section 4.2) has been dedicated, mainly due to its complexity. The conducted imaging on extended samples and the obtained results are discussed next in Section 4.3. The calibration procedure for point localisation, experimental point localisation and the obtained results are presented in Section 4.4. The performance of this technique is also compared to that of established point localisation techniques of comparable complexity in the aforementioned section. Finally, the conclusions extracted from this experiment and future work are discussed in Section 4.5.

4.1 Optical setup

A representation of the optical setup used in this experiment is shown in Fig. 4.1 for widefield fluorescence imaging. For the case of brightfield imaging in reflection, the dichroic was replaced by a 50:50 beam splitter and a $532nm$ ($25nm$ FWHM) filter, was then placed in the path of the epifluorescence illumination in order to achromatise the illumination. Normal Köhler illumination filtered at $532nm$ using the aforementioned filter was used to perform brightfield imaging in transmission. The setup was built around a commercial inverted microscope (Nikon Eclipse Ti-E) using a modified $40\times, 0.75NA$ objective (Nikon Plan Apochromat) with a DOF of $1.1\mu m$. The modifications made to the objective were necessary in order to fit the CPM inside it and this was accomplished by means of several 3D printed parts. The dismantled objective together with the CPM and its mount are shown in Fig. 4.2 where the objective was

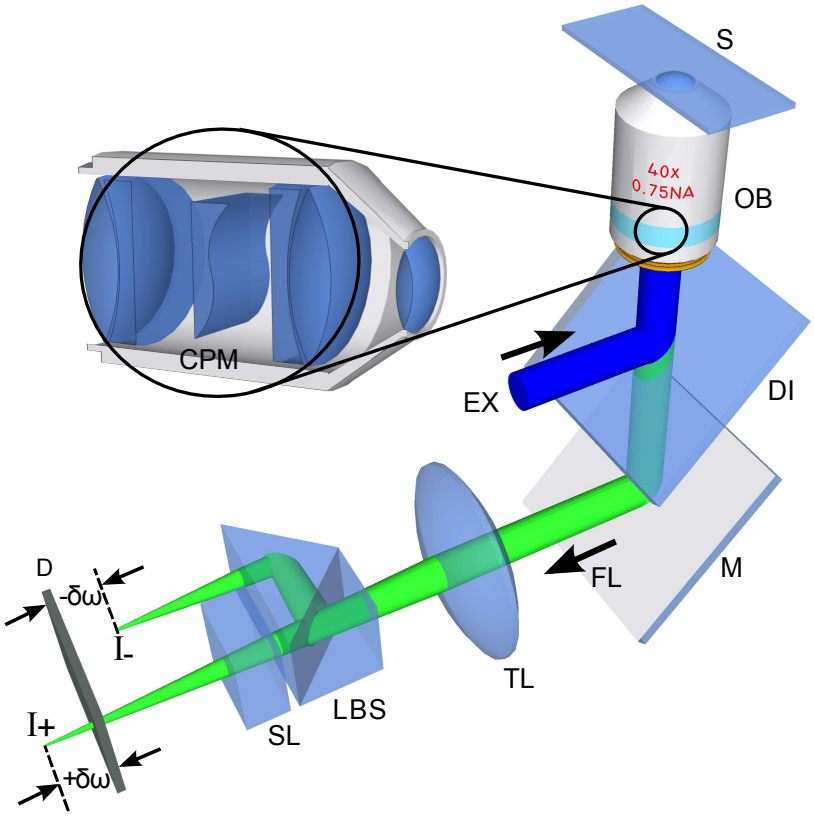


Fig. 4.1 Epifluorescence optical setup. **CPM**: cubic phase mask, **S**: specimen, **OB**: objective (40×, 0.75NA), **DI**: dichoric, **M**: mirror, **EX**: excitation ($480nm$), **FL**: fluorescence ($532nm$), **TL**: tube lens, **LBS**: lateral beam splitter, **SL**: glass slab, **D**: detector, $\boldsymbol{\delta\omega}$: defocus offset, $\mathbf{I}_+$: image with positive defocus offset, $\mathbf{I}_-$: image with negative defocus offset

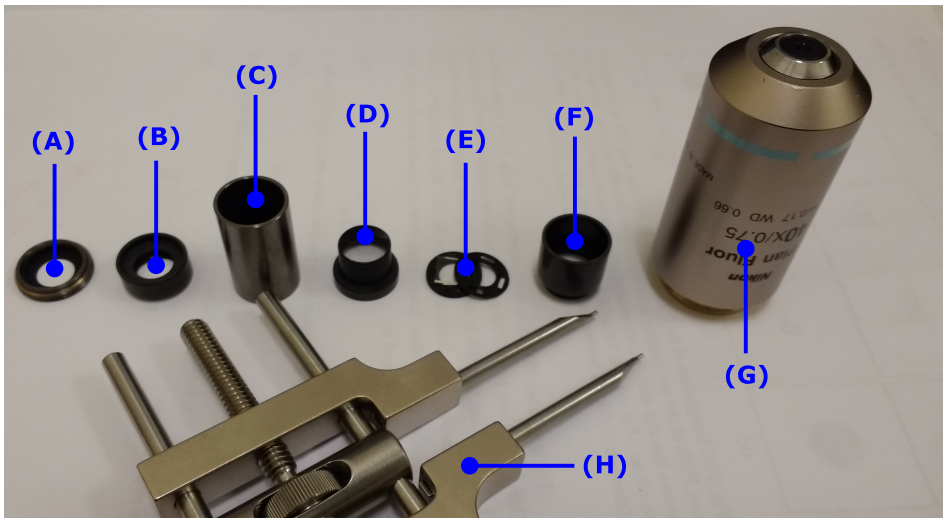


Fig. 4.2 Mounting the CPM inside an objective. **(A)**: back aperture; **(B)**: first optical element; **(C)**: spacer; **(D)**: second optical element; **(E)**: CPM and 3D printed mounts; **(F)**: third optical element; **(G)**: barrel and front lens of the objective; and **(H)**: tool used to open the objective.

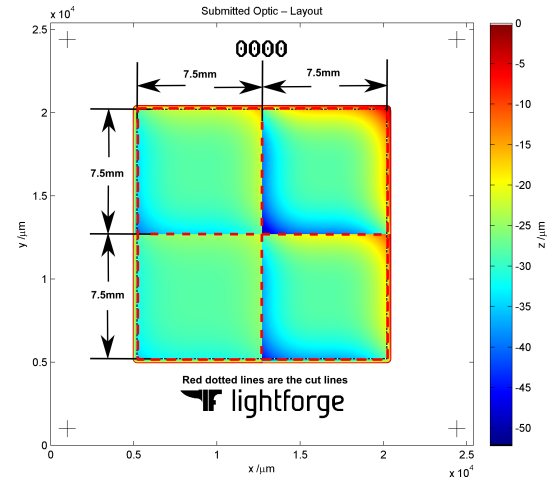
dismantled by unscrewing the back aperture of the objective and pulling out the optical elements and other components from the back in succession. With reference to Fig. 4.2, to assemble the objective, the CPM is enclosed in the 3D printed mounts (E) and then this is placed inside the hollow space in (F). The whole assembly is then placed inside

the barrel of the objective (G). Finally, elements (A) to (D) are stacked in reverse order inside the barrel as well and the back aperture is screwed tight in place using an appropriate tool (H).

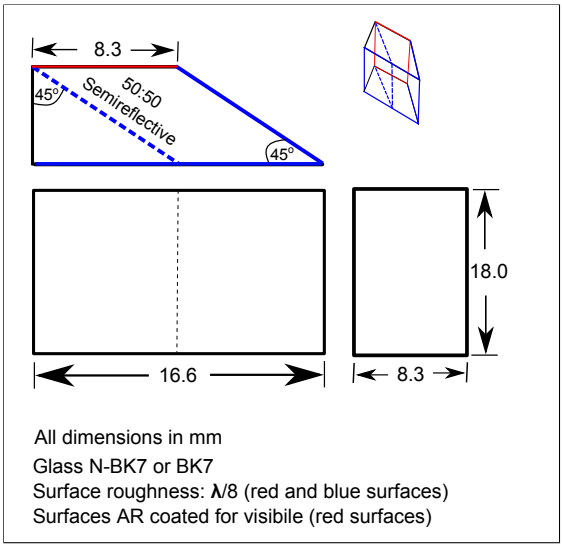
As a side-effect of these modifications, the objective was stopped down to roughly 70% of its original effective aperture diameter. Although an accurate measurement of this is quite difficult to make (especially with the CPM in place) without considerable knowledge of the optics of the objective itself, an estimate was obtained by measuring the shift of the cubic PSF with depth. This reduction in effective aperture diameter resulted in a numerical aperture of roughly 0.5 and hence a DOF of $\sim 2.4\mu m$ as per the definition of DOF given by Eq. 4.1 where λ_o is the wavelength of the illumination, NA is the numerical aperture of the objective under the assumption that the numerical aperture of the condenser is equal to or greater than that of the objective, n is the refractive index (1.0 in this case), M is the lateral magnification and finally, e is the pixel pitch.

$$DOF = \frac{\lambda_o \times n}{\mathrm{NA}^2} + \frac{n \times e}{M \times \mathrm{NA}} \tag{4.1}$$

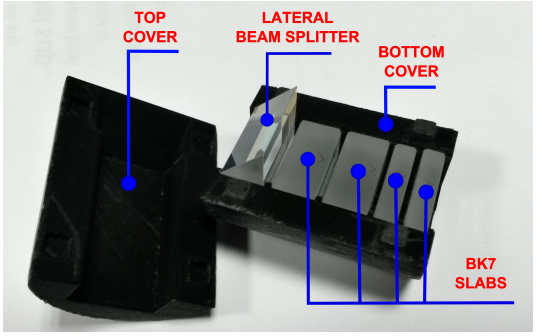
On the other hand, the position of the CPM within the objective is of critical importance. In the ideal case, it should be placed exactly in the aperture stop in order to ensure uniform coding of all fields. Any deviation from this condition will result in field dependence - *i.e.* the PSF will vary across the FOV. Placing the mask exactly in the aperture stop is, however, difficult to achieve in practice. This is because, owing to their commercial sensitivity, the design details of a commercial objective are not available to the general public. As a result, the exact position of the aperture stop is unknown and can only be roughly estimated. Furthermore, not every plane within the objective is accessible due to the physical construction of the objective itself. In this experiment, given that the number of accessible plane between the internal optical elements of the objective was small, an exhaustive approach was taken. In other words, the mask was placed at each accessible planes within the objective and the amount of field variation assessed by sampling the PSF across the FOV. The mask was then placed at the plane where field dependence was minimum. Despite this, some remnant field dependence was observed and this was mitigated through calibration, as discussed in the following section.



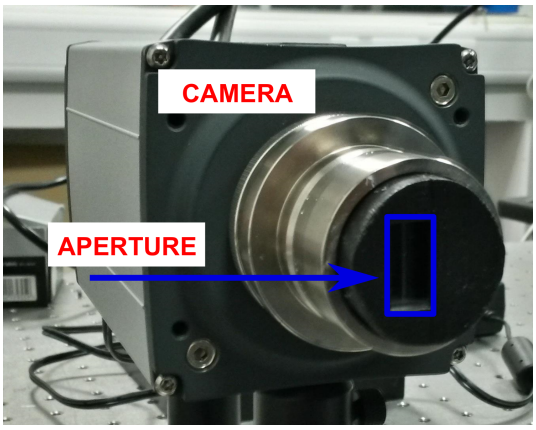
(a)



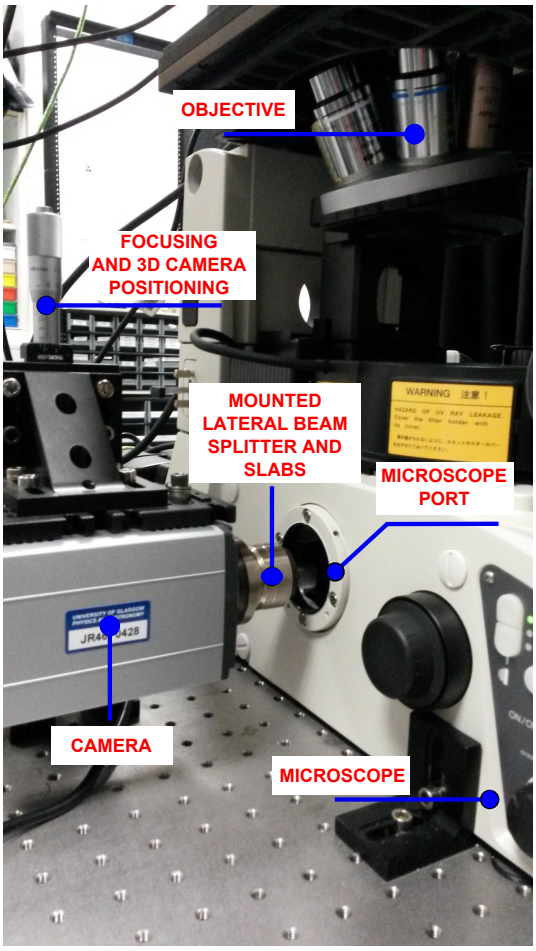
(b)



(c)



(d)

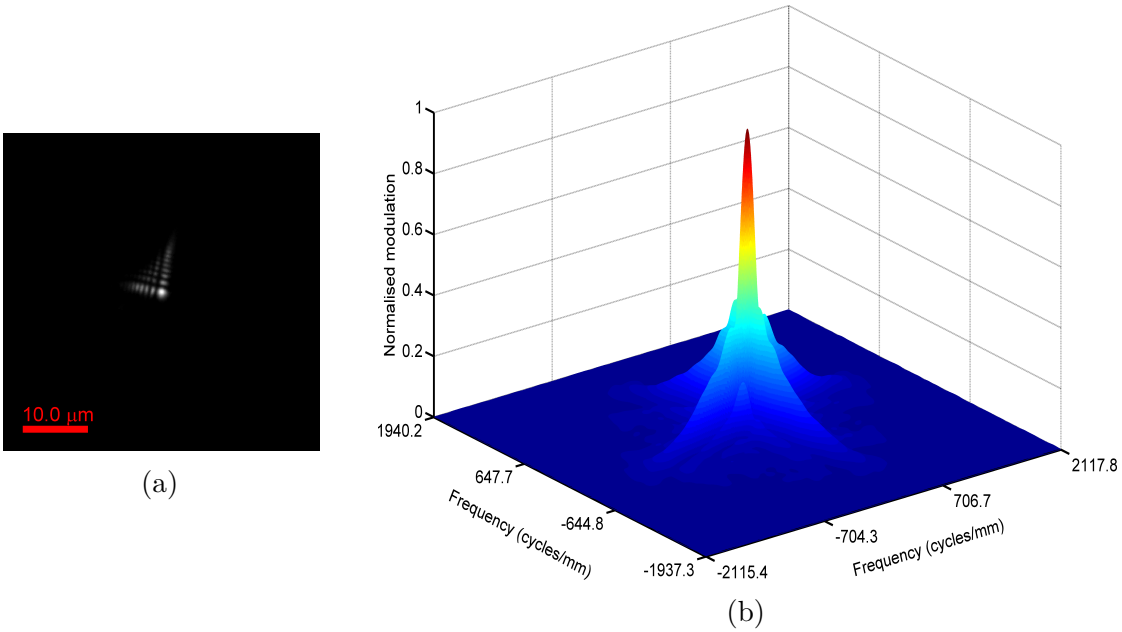


(e)

Fig. 4.3 Components of the optical setup: (a) layout of the mask substrate, (b) design of the lateral beam splitter, (c) 3D printed housing and assembly of the slabs and the lateral beam splitter, (d) slabs and lateral beam splitter mounted on the camera showing the aperture (half of the detector), and (e) slabs and lateral beam splitter mounted on the camera on the microscope port.

The CPM used in this experiment was custom-designed for a wavelength of $532nm$ and manufactured by PowerPhotonic Ltd. on a $1mm$ thick fused silica substrate (refractive index of 1.46071 at $532nm$) using a process with a sag resolution of $1nm$ and a pixel size of $10\mu m \times 10\mu m$. Fig. 4.3a shows a diagram of the design for illustration

purposes. As shown, four CPMs with α of 3.0, 5.3, 7.7 and 10.0 waves respectively were fitted onto the available substrate area. The substrate was then diced into four quadrants, thus giving four separate CPMs with different α values. Following some experimentation with different masks, it was decided to use the one with the smallest α value of 3 waves in order to maximise the magnitude of the shift, as discussed in Section 2.1.



(a) (b)

Fig. 4.4 Experimental (a) PSF (scale bar: $10\mu m$), and (b) MTF. In each case, the dimensions are referred to the object side.

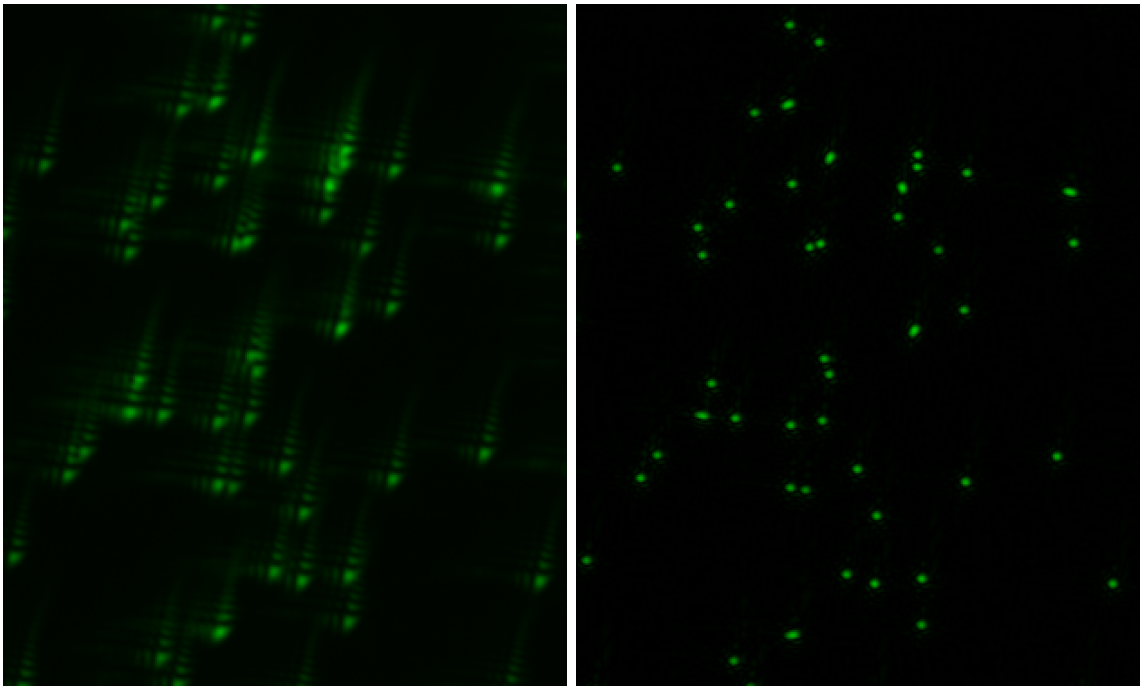
As discussed earlier, CKM necessitates two acquisitions (I_+ and I_-) of the scene which for a "true" single-snapshot implementation have to be acquired simultaneously. There are several ways of performing this, for instance, by using two cameras and a beam splitter. In this experiment, however, a single camera (Andor Zyla 5 MP, $6.5\mu m$ pixel pitch) was used in order to make the setup more compact. The sensor area was then split in two using a lateral beam splitter custom designed to match the sensor size, as shown in Fig. 4.3b. A difference in defocus, $2\delta\omega$, of $11.2mm$ was then generated by means of additional custom-cut BK7 glass slabs as shown in Fig. 4.3c and Fig. 4.3e. As one would expect, the addition of glass slabs introduce spherical and other aberrations; however, owing to the fact that the beam on the image side of a microscope is slow (*i.e.* high $f_\#$ since the tube lens has a long focal length ($200mm$ for Nikon microscopes)), this was not found to be a major problem. Lastly, the PSF and MTF are shown in Fig. 4.4a and Fig. 4.4b respectively. These were acquired by imaging $0.39\mu m$ fluorescent beads affixed to a coverslip close to the in-focus position.

4.2 Calibration for extended imaging

As discussed in Section 2.3, in order to infer depth, CKM relies on the measurement of depth-induced disparity between two shifted, but otherwise identical, images of the sample (I_+ and I_-). Moreover, this disparity is measured by matching the intensities in either image over a neighbourhood of pixels. This implies that any difference in the distortion between the two images, and any relative shift between them which is not depth-induced, has a direct impact on the measured depth. Given that the achievable responsivity (translation in the lateral dimensions on the image side [μm] per change in axial position around the focal point on the object side [μm]) is often small, the error in depth introduced by these factors can be significant. It therefore is of the utmost importance to have excellent correspondence between the two acquisitions. This constitutes one of the two objectives of the calibration procedure described here. The other objective is the characterisation of the system over the extended DOF of interest; or, in other words, the acquisition of the set of PSFs necessary to implement the algorithm described in Section 2.6.

In practice, achieving a good enough correspondence between the two images, or equivalently, the two halves of the detector (a process which will henceforth be referred to as *field calibration*), revealed to be challenging mainly due to the complex distortion which resulted. The problem is further compounded by the fact that this distortion was found to be dependent on the axial position (*i.e.* depth). The probable source of this distortion is the modifications applied to the objective coupled with the fact that the CPM was not placed exactly in the aperture stop as was previously mentioned. Furthermore, in extended DOF imaging, the objective is operated at a significant distance away from the intended focal point. This reduces the efficacy of its aberration correction, thus giving rise to additional aberrations which would normally be suppressed [93]. Finally, as is often the case for an inverted microscope, the Nikon Eclipse Ti-E microscope used in this experiment focuses by displacing the objective rather than the sample. Given that a CPM was fitted inside the objective, the beam in the infinity space of the microscope is no longer collimated. This can potentially enhance the dependence of the distortion on the axial position. More importantly, this might also have repercussions on the performance of the calibration itself, since this constitutes a key difference between the calibration procedure and the actual measurement procedure. Further enquiry into the matter is, however, necessary in order to make definitive claims.

The calibration procedure adopted in this experiment is based on the 2D calibration method presented in [94] using fluorescent beads and the similar procedure followed in [95] where a sub-resolution pinhole array was used instead. In this case, a flat (single depth) synthetic scene consisting of ~ 600 sub-resolution fluorescent beads ($0.39\mu m$ diameter) scattered across the FOV was used, where the beads act as fiducial markers.



(a) (b)

Fig. 4.5 (a) captured image of sub-resolution ($0.39\mu m$ diameter) beads, and (b) recovered (deconvolved) image of the same beads.

A through-focus series over a depth range of $\pm 10\mu m$ was then acquired at steps of $0.1\mu m$. The image of one of the sub-resolution beads in each of the two acquisitions (I_+ and I_-) at each depth step, which constitutes the PSF with which that image is convolved, was then isolated, centred and thresholded to remove the offset, and finally used to deconvolve its respective image. In the absence of any distortion and field dependence, the two recovered images should be identical -*i.e.* no relative shift between them should exist. This is because each of the two acquisitions has been deconvolved with a PSF measured at the same depth as that of the beads in it. Any remnant disparity (or misregistration) between the two recovered images can therefore be attributed to distortion and other non-depth-induced displacements, thus it must be compensated for.

To this end, corresponding beads in the two images acquired at each depth step were identified and localised with subpixel accuracy. A piecewise linear (affine) transform [96, 97] was then extracted from the location of the fiducial markers in such a manner as to map the position of the fiducial markers in one of the recovered images to those of the corresponding ones in the other. Localisation of the beads was performed on the recovered images because deconvolution converts the triangular cubic PSFs to Gaussian-shaped points (see Fig. 4.5) and the latter are more apt for accurate localisation. The position of the beads was thus found by least-square fitting a 2D Gaussian function to each recovered PSF. Since this amounts to over $240,000$ (~ 600 beads $\times 2$ images $\times 200$ axial positions) fits for the calibration alone, it was found to be rather computationally intensive and could not be performed in a realistic time frame using a high-level language

such as Matlab. Consequently, the necessary code was developed in C and subsequently interfaced to other code written in Matlab.

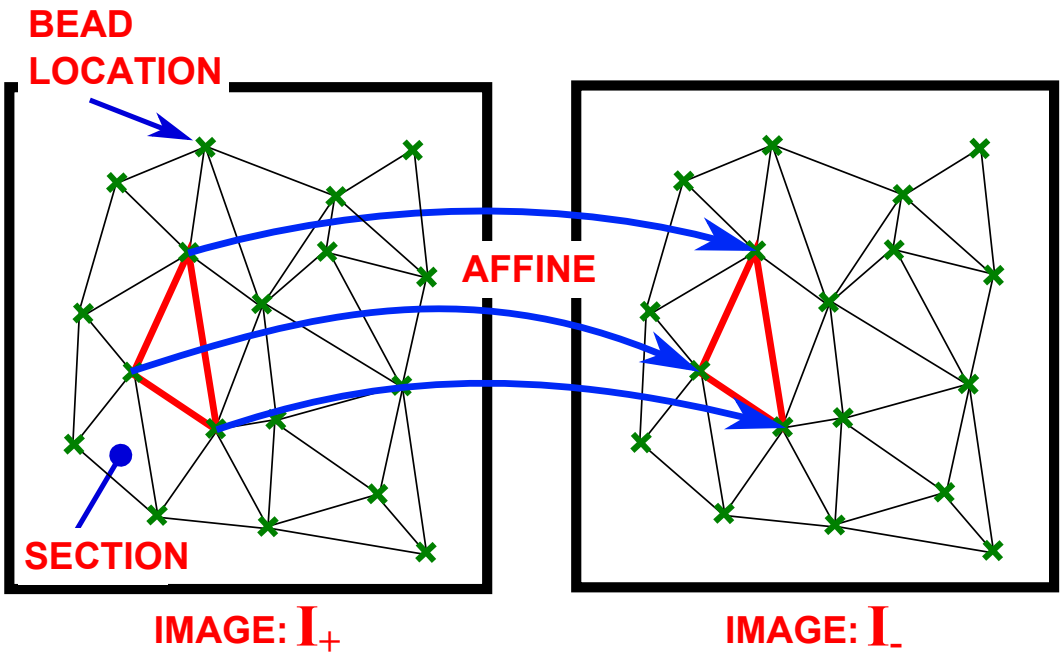


Fig. 4.6 Mapping field of view between the two captured images I_- and I_+ (not all possible sections are shown).

In theory, the only distortion which should result is simple projective distortion and a change in magnification (scaling); however, as mentioned previously, complex field-dependent distortion was observed. As a result, simple global (*i.e.* applicable to the entire FOV/image) geometric transforms, such as affine and projective transforms, were found to be inadequate. On the other hand, a piecewise linear transform segments the FOV into small triangular sections with vertices at three neighbouring fiducial markers, as shown in Fig. 4.6. An affine transform is then calculated in such a manner as to map each triangular section in one of the images to its corresponding one in the other image. The division of the FOV here is key because over a small section of the FOV, the field dependence is small enough for a simple geometric transform such as affine to hold. This permits adequate mapping between the two images to be achieved without rigorous modelling of the distortion. The same issue was observed in [94, 95] and was solved by means of a local weighted mean transform [98] which, similar to the piecewise linear transform used in this experiment, acts on portions of the FOV rather than globally. A piecewise linear transform was preferred because it can be calculated and applied faster than the local weighted mean and because it was found to yield almost identical (in some cases even better) results, while requiring less trimming at the edges of the image. The mean residual error calculated by applying the calibration to the images from which it was calculated and then re-localising the beads was $3.2nm$ which corresponds to an error in depth of $\sim 0.34nm$.

Sample preparation is of critical importance to the success of the calibration procedure described above. Beads suspended in a fluid will experience Brownian motion

and drift, which can result in appreciable error. Furthermore, even if a liquid viscous enough to reduce these effects to negligible levels had to be used, the sample is still likely to contain beads at different depths, rendering the calibration impossible. This problem was solved by producing a single layer of beads immobilised on a coverslip. In order to do this, $5\mu l$ of beads solution (1% solids, 0.1% Tween 20 and deionised water) was diluted in $70\mu l$ of isopropyl. The mixture was then shaken for a few minutes to ensure homogeneity and a drop was then placed on a $24 \times 60mm$, #1 1/2 coverslip. This was then re-flown until the liquid evaporated, leaving only the beads dispersed on the coverslip. Isopropyl was used because it is volatile and to prevent the Tween 20 from crystallising; thus, ensuring a homogeneous refractive index throughout the sample. Finally, a region of the sample with little aggregation was identified and imaged. Obviously, the flatness of the sample, which is critical to the calibration procedure just described, is directly affected by any tilt in the placement of the sample and/or any bending of the coverslip. A way to circumvent this problem is to image only a small region of the sample over which it can be assumed to be flat, despite any tilts which might be present. This region would then have to be scanned in order to calibrate the entire FOV. In this experiment, this could not be done, however, for higher numerical apertures, following this procedure is advisable [99].

As mentioned earlier, the second objective of the calibration procedure is the acquisition of the set of PSFs (PSF_+ and PSF_-) necessary to run the CKM algorithm (*i.e.* system characterisation). Since the PSF depends on the imaging modality whereas field calibration does not, several attempts were made to acquire these PSFs separately from the field calibration data. These were, however, unsuccessful because it is difficult to match with precision the PSFs to their corresponding field calibration data unless they were acquired simultaneously on the same sample. The choice of bead diameter was made with this in mind. As stated previously, the diameter of the beads was chosen to be smaller than the resolution limit ($636nm$). This enables the acquisition of PSFs concurrently with field calibration data. The problem of matching PSFs to field calibration data is therefore eliminated. A relatively isolated bead was thus identified and its two images were taken as the PSFs at that particular depth. Before the PSFs were used to deconvolve images, they were centred in exactly the same manner as in the field calibration, the offset was removed by subtracting a dark frame and the total intensity was then normalised to unity. These are essentially the same PSFs used to deconvolve the field calibration data prior to the localisation of the beads. The downside of this approach is that it is only applicable to fluorescence imaging because beads in brightfield do not yield PSFs of a sufficiently high quality. This is not an issue for field calibration, however, due to spectral differences between brightfield and fluorescence, it is for the system characterisation, since the PSFs acquired during this phase of the calibration will be used for deconvolution. In this work, PSFs acquired in fluorescence were used to deconvolve both fluorescence and brightfield images and were found to yield recoveries of satisfactory quality; however, this leaves scope for improvement.

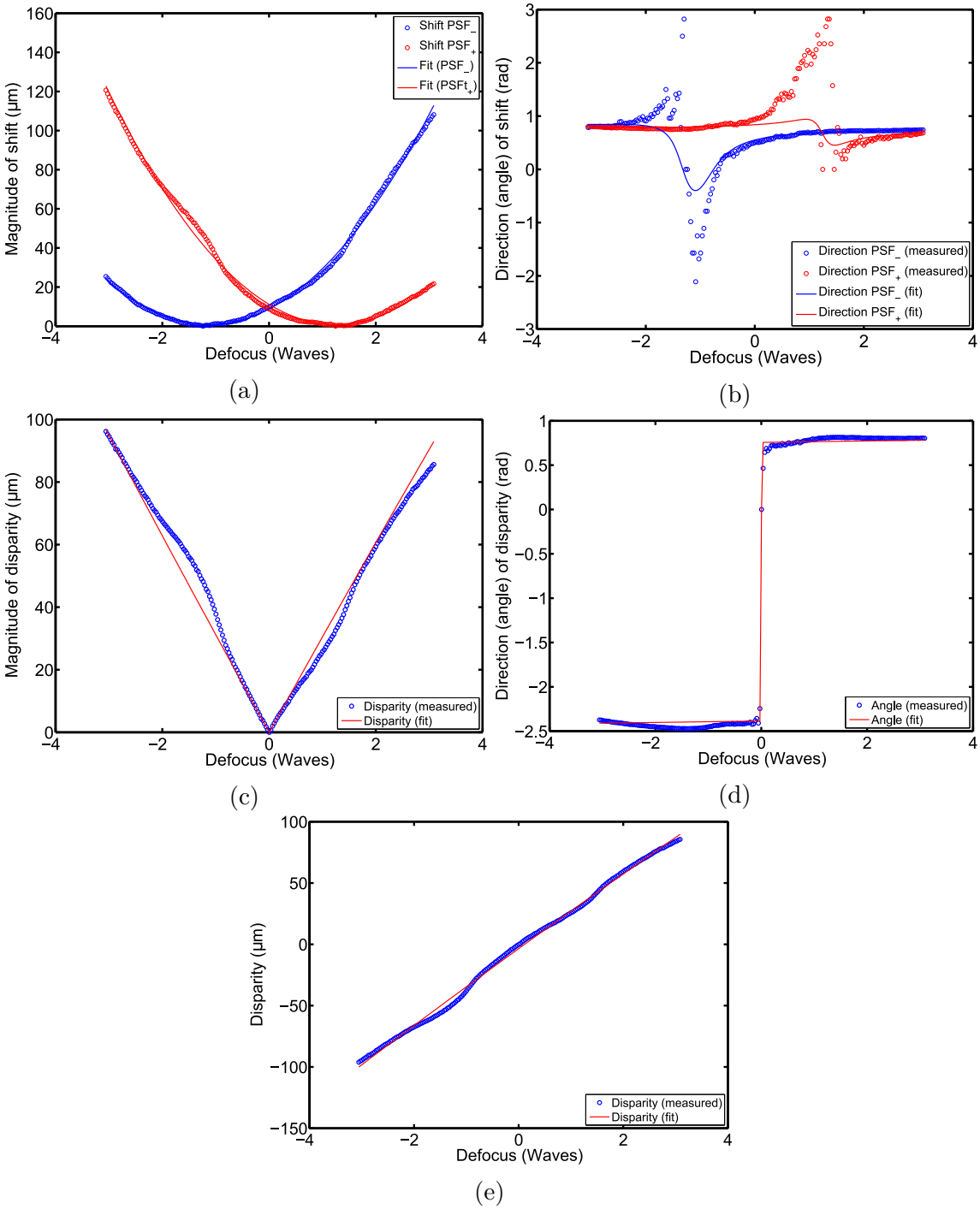


Fig. 4.7 (a) magnitude of the individual shift of the PSF with positive defocus offset (PSF_+) and the PSF with negative defocus offset (PSF_-), (b) direction of the shifts represented by the angle subtended between the shift vectors and the horizontal axis of the image, (c) magnitude of the disparity (*i.e.* difference between the individual shifts), (d) direction of the disparity represented by the angle subtended between the disparity vector and the horizontal axis of the image, and (e) the disparity (magnitude and approximate direction).

The individual shifts of the PSF with positive defocus offset (PSF_+) and the PSF with negative defocus offset (PSF_-) were measured by correlating the PSFs with a reference PSF in each case. These are shown in Fig. 4.7a and, as expected, the shift is approximately a quadratic function of defocus expressed as peak aberration at the edge of the pupil (*i.e.* Seidel's defocus coefficient W_{20}). To convert between longitudinal

defocus, δz, and W_{20}, an NA of 0.5 was taken. Moreover, to do this, knowledge of the in-focus position is also required; however, this cannot be inferred from the morphology of the PSF, since the cubic PSF does not respond to defocus - unlike a Gaussian PSF. To find the in-focus position, therefore, the longitudinal defocus was assumed to be proportional to W_{20}; hence, under this assumption, the relationship between longitudinal defocus and the shift is also quadratic. It is important to note that this assumption only holds close to the focus (*i.e.* $|\delta z| \ll f$ with f being the focal length) as the following equation shows:

$$W_{20} \approx \frac{\delta z r^2}{2f(f+\delta z)\lambda} \approx \frac{\delta z r^2}{2f^2\lambda}\, Waves \tag{4.2}$$

where r is the radius of the aperture and λ is the wavelength. In this experiment, the total extended DOF is several orders of magnitude smaller than f, therefore, this approximation holds. A reference PSF was thus chosen arbitrarily close to the apparent focus. All the PSFs were then correlated with it and the shifts with respect to it thus measured. The positions at the turning points were then estimated by fitting a quadratic polynomial to the measured shifts against longitudinal defocus plots. The positions at the turning points represent the in-focus position of the two imaging channels. We can therefore write:

$$\begin{aligned} TP_+ &\approx Z_o + \delta z_o \\ TP_- &\approx Z_o - \delta z_o \end{aligned} \tag{4.3}$$

where TP_+ and TP_- are the axial positions at the turning points, Z_o is the nominal focus position and δz_o is the longitudinal defocus offset introduced by the lateral beam splitter and slabs combination referred to the object side. From Eqs. 4.3, the nominal focus position, Z_o, can then be estimated by taking the mean of TP_+ and TP_-, thus longitudinal defocus can be converted to W_{20}. Note also that the defocus offset introduced by the beam splitter and slabs is given by:

$$\delta\omega \approx \frac{\delta z_o r^2}{2f(f+\delta z_o)\lambda} \tag{4.4}$$

which was found to be 1.26 $Waves$. Finally, given that this translation (or shift) is a vector quantity, the direction is also of relevance and this is shown in Fig. 4.7b as the angle subtended between the sift vectors and the horizontal axis. As one can observe from this figure, the translation occurs in approximately the same direction and therefore the angle does not change much, except close to the focus. This anomaly is caused by the presence of aberrations in the system, in particular astigmatism, as discussed in Subsection 2.4.2.

The disparity between them (*i.e.* resultant shift, or equivalently, the difference between the shift of PSF_- and PSF_+) is shown in Fig. 4.7c. As expected, the magnitude

of the disparity displays an approximately linear relationship with the defocus, albeit some deviation from this is evident. This can be attributed to aberrations present in the system to which it is particularly susceptible given the modest coding strength (α) of 3 $Waves$ used in this experiment. The slope of the disparity which corresponds to the responsivity was found to be 2.90 pixels/$Wave$ or equivalently 18.85$\mu m/Wave$. Close to the in-focus position, 1 $Wave \sim 2.03\mu m$, therefore, the responsivity can also be expressed as: $9.30\mu m/\mu m$, or equivalently, 1.43 $Pixels/\mu m$. The direction of the disparity vector is shown in Fig. 4.7d, from which one can see that the disparity switches direction with defocus. This is what permits inference of the sign of the defocus as discussed in Subsection 2.4.3, and hence, the utilisation of the full extended DOF provided by the CPM. All fitted curves in Fig. 4.7b to Fig. 4.7e were obtained by fitting quadratic polynomials to the individual shifts of the PSFs as shown in Fig. 4.7a and are intended to give an idea of the ideal (*i.e.* aberration-free) response. Fig. 4.7e depicts the magnitude of the disparity and its direction under the assumption that for negative defocus, the disparity is negative and vice versa. This is intended to illustrate the predicted approximately linear relationship between the disparity and the defocus, as derived in Subsection 2.4.3.

To recapitulate, following the calibration procedure just described, for each $0.1\mu m$ step within the depth range of interest ($\pm 10\mu m$): (1) field calibration data which is used to compensate for any kind of distortion which might be present at that particular depth, and (2) two PSFs (PSF_- and PSF_+) were acquired. For a generic sample, therefore, two images, I_- and I_+, are acquired simultaneously once. These two images are then deconvolved by their respective PSF (I_- by PSF_- and I_+ by PSF_+) acquired at some depth within the depth range of interest. Following deconvolution, the field calibration at that depth is applied in order to compensate for any stray shifts caused by any distortion present. A metric proportional to the disparity is then evaluated for that particular depth. This is then repeated for each depth step and the metric minimised over the entire depth range of interest. The depth which minimises the metric at a particular (x, y) location in the sample is equal to the depth of the sample at that location. This is essentially the algorithm discussed in Subsection 2.7.1.

4.3 Extended imaging

Using the optical setup and the calibration procedure discussed earlier, several extended samples intended to demonstrate the various facets of this technique and to gauge its performance in matter of image quality/artefact removal and depth measurement were imaged. On a generic sample with complex geometry, it can be difficult and quite laborious to obtain a suitable ground truth. For this reason, a tilted chrome-on-glass distortion target consisting of dots with a $5\mu m$ diameter and $10\mu m$ centre-to-centre separation, was imaged (in transmission) in order to assess the performance of CKM in

inferring depth. The purpose of the tilt is to induce a range of depths in the sample. The diffraction limited image (*i.e.* image taken with a clear aperture) of the sample is shown in Fig. 4.8a. This was acquired using a non-modified $0.75NA, 40\times$ objective. The single PSF (conventional) WC recovery is shown in Fig. 4.8b and Fig. 4.8c for each of the two acquired images. As one can see from these images, single PSF recovery results in noticeable artefacts besides the noise amplification inherent to WC.

For comparison, a Z-Stack of the sample was also taken. The Z-Stack reconstruction is shown in Fig. 4.9a, a colourmap of the depth as given by the Z-Stack is shown in Fig. 4.9b, and finally, the 3D topographic reconstruction is shown in Fig. 4.9c. The equivalent images obtained by CKM are shown in Fig. 4.9d, Fig. 4.9e and Fig. 4.9f respectively. In each case, the raw depth map was average over the area of each dot and the mean depth was taken to be the depth of the dot. This was necessary because both the Z-Stack and CKM display artefacts in their depth maps. The SNR in the images shown in Fig. 4.8 and Fig. 4.9 was $60dB$ by the definition of SNR given by Eq. 4.5 below:

$$SNR = 20 \times log_{10}\left(\frac{\text{Image mean} - \text{Offset}}{\text{Noise standard deviation}}\right) \tag{4.5}$$

Comparing Fig. 4.9a to Fig. 4.9d, one can see that the CKM reconstruction displays a lower SNR than its Z-Stack counterpart. Apart from that, the two images are of comparable quality. On the other hand, comparing the CKM recovery to that given by conventional wavefront (Fig. 4.8b and/or Fig. 4.8c), the improvement in image quality due to the suppression of artefacts is apparent. The correlation between the depth profile given by CKM and that given by the Z-Stack is conspicuous; however, this merits further consideration.

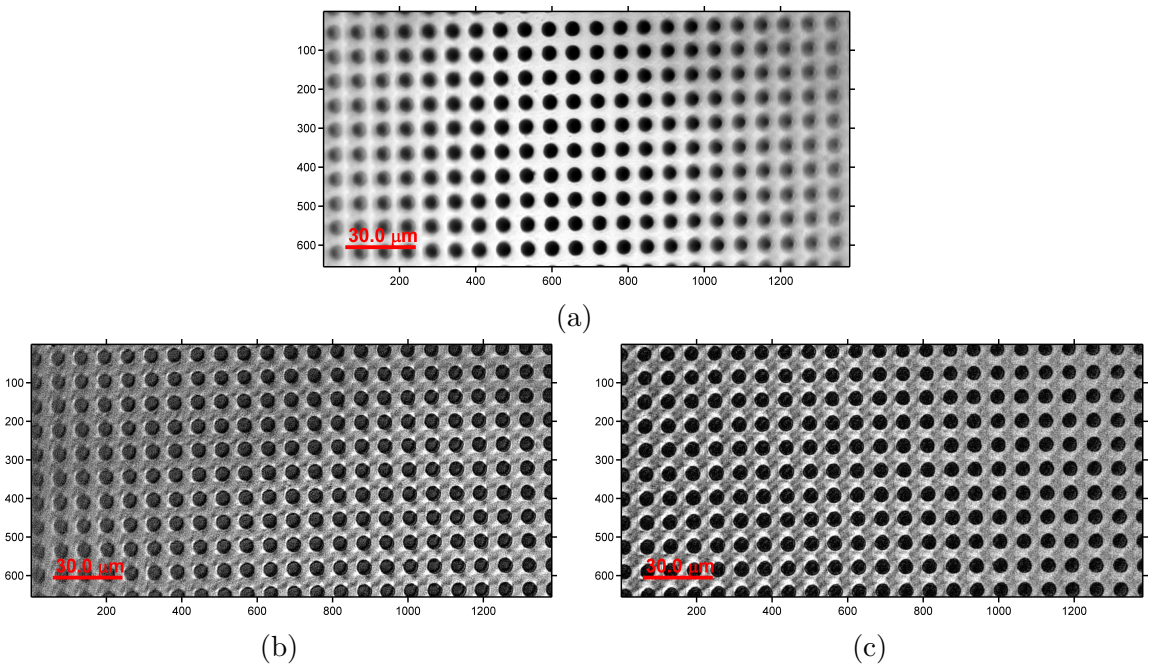


Fig. 4.8 Tilted distortion target: (a) Diffraction limited ($0.75NA, 40\times$), (b) Single PSF recovery from I_-, and (c) Single PSF recovery from I_+.

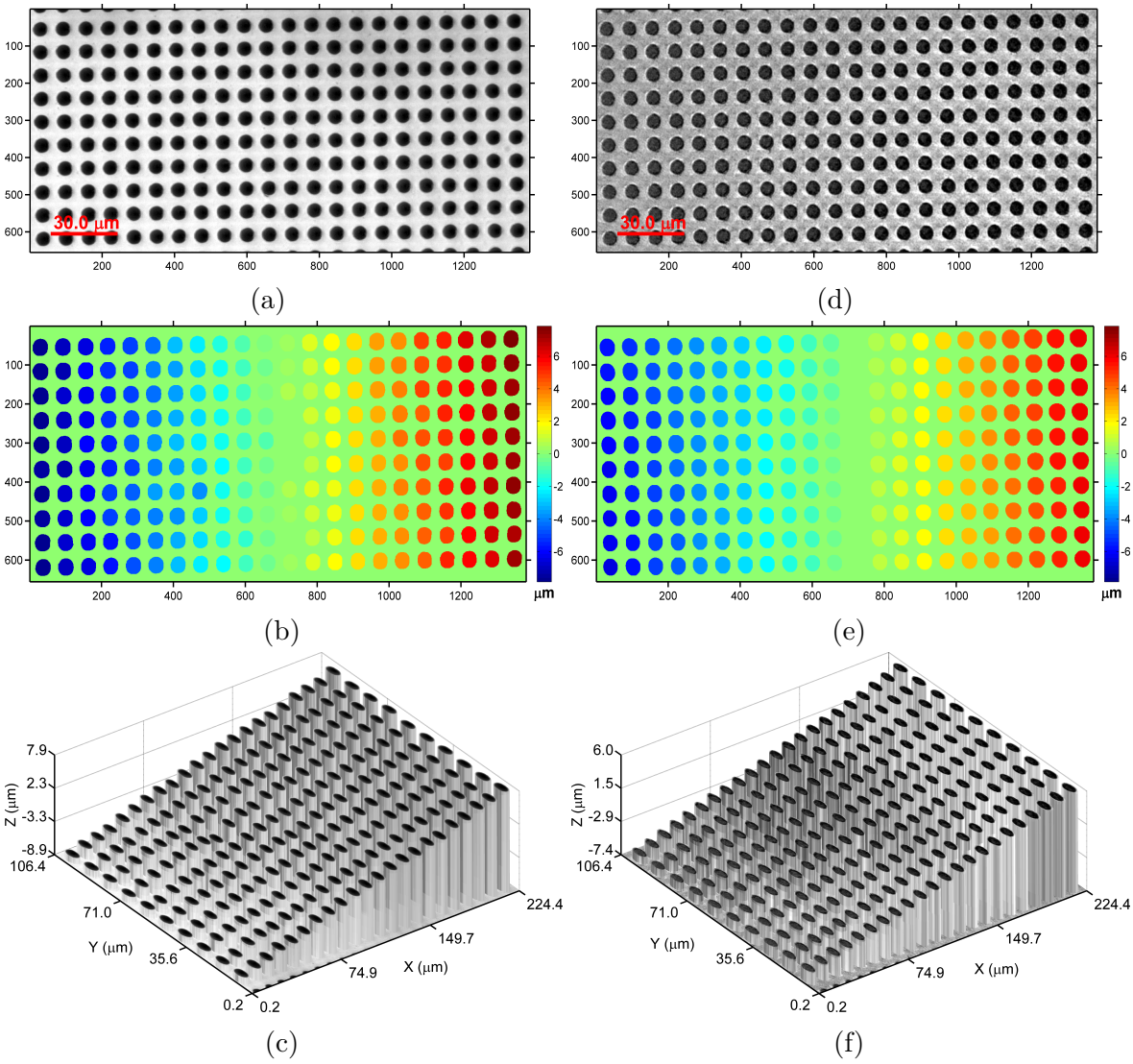


Fig. 4.9 Tilted distortion target: (a) Z-Stack reconstruction, (b) Z-Stack depth profile, (c) Z-Stack 3D reconstruction, (d) CKM reconstruction, (e) CKM depth profile, and (f) CKM 3D reconstruction. Z-Stack step size: $0.1\mu m$.

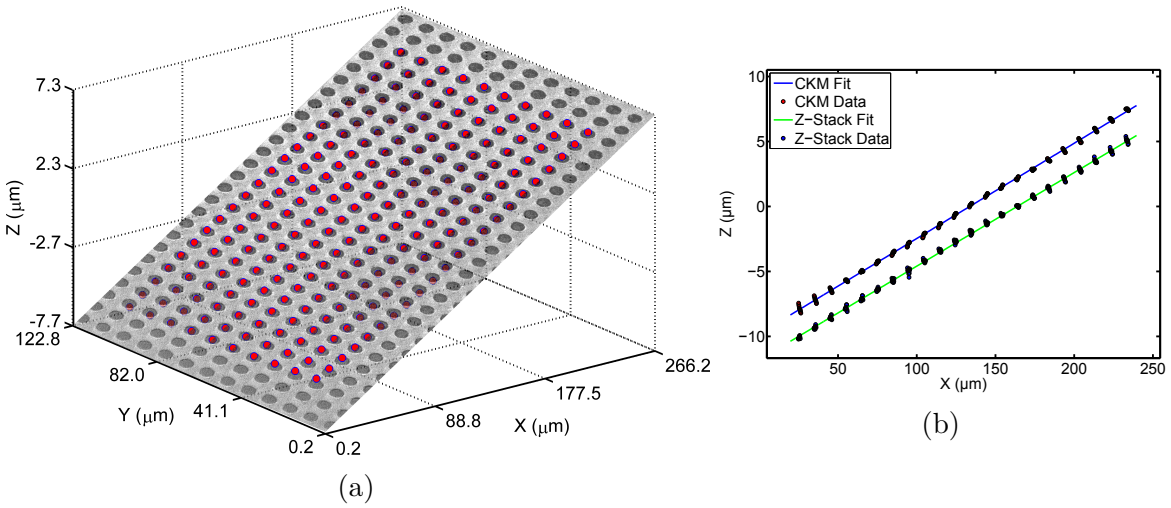


Fig. 4.10 Error analysis on extended scenes: (a) measured dot depths (red markers) and fitted plane (image of distortion target), and (b) comparison between CKM and Z-Stack.

To calculate the error in the depth as inferred by CKM, a ground truth needs to be established. The obvious way of doing this is to take the depth given by the Z-Stack as the ground truth. This, however, is not ideal because the Z-Stack, like CKM, relies on the features in the image to infer depth and hence is susceptible to depth artefacts [100–102] which for individual dots might still be significant. In this case, the depth profile of the distortion target is known to be a plane. Consequently, a plane was fitted using a least-square fit to the depths of the dots given by CKM, as shown in Fig. 4.10a. Any deviation from this plane was then assumed to be random (*i.e.* "uncalibratable") error caused by noise and other factors which were not accounted for. A plane was fitted in the same way to the depths of the dots as given by the Z-Stack and a comparison between the slopes caused by the tilt was made. The experiment was repeated for different SNRs and the results are tabulated in Tbl. 4.1.

SNR (dB)	Error (μm)	Difference in slope w.r.t. Z-Stack (%)
60	±0.15	-4.61
50	±0.19	-4.89
40	±0.20	-3.55
30	±0.25	-4.95

Table 4.1 Error in depth of the biplane CKM for extended samples with respect to SNR

From Tbl. 4.1, one can notice that the error in depth is inversely proportional to the SNR, which is as expected. Moreover, one can see that there is a ∼5.0% difference in slope between the CKM estimate and that obtained from the Z-Stack. In this case, it cannot be established whether this difference is an underestimate by CKM, an overestimate by the Z-Stack or a combination of both; however, given that it varies only slightly with SNR, one concludes that it is probably a systemic error. In principle, therefore, this error can be eliminated by calibration. A potential cause of this error lies in the fact that the calibration was performed using an epifluorescence setup, whereas

the imaging in this case was performed in transmission. This implies that there are spectral differences between the calibration and the actual imaging, which is known to effect the responsivity, and hence, the measured depth. Further investigation is, however, necessary in order to definitively identify the cause.

An important consideration is the effect which the NSR ("K factor") in the Wiener filter used to recover the captured (coded) images has on the depth map given by CKM. In WC, the NSR is commonly assumed to be independent of frequency [103] although sometimes, a power law distribution [104] is used instead. In this experiment, no substantial difference in the quality of the recovered images given by either method was observed; therefore, a constant NSR was used for simplicity. Furthermore, the NSR dependence on the scene itself amongst other things. Consequently, without making assumptions on the scene, it is impossible to know the NSR *a priori*. This is usually dealt with by manually tweaking this parameter in order to optimise the recovered image.

On the other hand, whereas the frequency dependence of the NSR can be adequately addressed as explained above, the dependence of the NSR on defocus (or depth) is rarely considered in WC. This is because in conventional WC, the coding strength, α, of the CPM can be freely chosen to match the required depth range. As a result, over the depth range of interest, variations in the MTF with depth can be reduced to negligible levels. A single NSR value can therefore be used to recover an image irrespective of the defocus in the scene. On the contrary, in CKM, the choice of α (or $\delta\omega$ in case of the biplane CKM) is restricted by shift considerations. Because of this, the depth range is often at the limit of what the CPM can handle without significant loss of frequencies. This implies that over the depth range of interest, variations in the MTF with depth can be significant (especially at higher frequencies), and thus, a single value of NSR might not be suitable. In fact, using a single NSR value over the entire depth range was found to give a difference of ∼20% between the slope of the distortion target as given by CKM and that given by the Z-Stack for the $60dB$ case. To circumvent this problem, an empirically-determined profile of NSR values was used as given by Eq. 4.6.

$$NSR = C \cdot \left\{ \left(1.25 \times 10^{-4}\right) Z^4 - \left(2.51 \times 10^{-5}\right) Z^3 + ... \right. \\ \left. ... + \left(1.88 \times 10^{-6}\right) Z^2 - \left(6.28 \times 10^{-8}\right) Z + 1 \right\} \quad (4.6)$$

where Z is depth in μm and C is a constant chosen to optimise the recovered image, as previously discussed. The choice of a quartic polynomial was made based on the reasoning that the NSR close to the focus position should change slowly, whereas it should increase rapidly at the extremities of the depth range. This is because close to the focus position, the MTF is still resistant to defocus, and hence, is suppressed only slightly. Conversely, the MTF at larger defocus is suppressed significantly and rapidly.

A quartic was found to adequately emulate this behaviour. The values of C used are tabulated in Tbl. 4.2.

SNR (dB)	C
60	4.5×10^{-4}
50	6.0×10^{-4}
40	9.0×10^{-4}
30	22.0×10^{-4}

Table 4.2 NSR parameter values used in the image recovery for different SNR values

To further illustrate this technique, several other natural -*i.e.* non-synthetic- samples were imaged in different imaging modalities (fluorescence and reflection). In each case, a Z-Stack of the sample was taken for comparison. In addition, the diffraction limited (non-coded aperture) and the conventional WC images are illustrated as tabulated in Tbl. 4.3. The diffraction limited images were acquired using an unmodified $40\times, 0.75$NA objective. The raw depth maps given by CKM were smoothed using a simple Gaussian filter or a guided filter [105] in order to reduce noise-induced artefacts in the depth map. These filters give very similar results, with the guided filter showing a slightly better performance when sharp transitions in the depth map coincide with transitions in the intensity image. Care, however, needs to be exercised in choosing the guided filter parameters because excessive filtering will give rise to erroneous edges in the depth map.

The Z-Stack reconstructions were performed using NIS-Elements version 4.10 by Nikon. In this case, owing to the complex geometry of the samples, the procedure used to find the error in depth described above for the distortion target could not be applied. As a result, the standard deviation of the difference between the depth maps given by CKM and the Z-Stack was taken as a figure of merit (referred to as "error" hereafter) for the correlation between the two techniques. As stated before, however, both CKM and the Z-Stack produce artefacts in their depth maps; therefore, one cannot attribute the error solely to CKM. Another source of error lies in the fact that the Z-Stack and CKM were performed using two different objectives. Although care has been taken to use similar objectives for both, the effects of the modifications applied to the objective used for CKM are not easily matched. Consequently, some differences in translation, viewing angle and distortion in general will inevitably result. To mitigate these effects, an intensity-based affine transform was applied to the Z-Stack images and depth maps in such a manner as to register them to their CKM counterparts. This registration, however, is not perfect, mainly due to differences between the CKM and the Z-Stack reconstructions. As a result, the error is augmented artificially.

Finally, an important consideration which has a profound impact on the performance of CKM is the amount of texture and/or features present in the sample itself. As discussed earlier, much like a Z-Stack, CKM can only infer depth in areas of the sample where features and/or texture are present because these permit the local registration

Sample	Imaging modality	Results	Error (μm)
Groove on a 2 Euro coin	Reflection	Fig. 4.11 \4.12	±0.59
Silicon die	Reflection	Fig. 4.13 \4.14	±0.84
Lily pollen grains	Fluorescence	Fig. 4.15 \4.16	±1.22
Schwann (neurilemma) cells on a tilted slide	Fluorescence	Fig. 4.17 \4.18	±3.17
Atomic force microscope cantilevers	Reflection	Fig. 4.19 \4.20	n/a

Table 4.3 Summary of the imaged extended samples.

of the image to be performed, and therefore, the shifts induced by the CPM to be measured. In featureless and textureless regions, no depth information can be obtained. This is demonstrated by the cantilever sample which has been included for this purpose. Owing to the fact that this sample is entirely featureless, depth could only be inferred at the edges, and as expected, the CKM reconstruction (Fig. 4.20c) is inferior to that obtained for other samples. At the other extreme, the coin sample has a substantial amount of texture throughout, therefore, a good depth map was obtained over the whole sample (or FOV). In fact, most of the error occurs close to the edge due to the a large transition in depth in a small region of space. The pollen grains sample has enough features to infer depth in the region of the image occupied by the grains. The rest of the FOV is background so it was masked. In this case, most of the error occurs at the interface between the grain and the background. The Z-Stack also predicts changes in depth as large as $5\mu m$ on the surface of the grain, which is probably artefactual. The Schwann cells sample is similar to the pollen sample; therefore, most of the error was found to occur at the interface between the cells and the background. In this case, however, such interfaces are much more numerous, therefore, their contribution to the error figure is substantially larger than in the pollen grains case. For the silicon die, CKM was found to struggle in the green region of the sample, mainly due to the lack of features there. In contrast, owing to the granularity of the copper deposits, a good depth estimate was obtained on the copper tracks. The error quoted in Tbl. 4.3 for this sample was evaluated on the copper tracks only. On the entire sample, the error increases to $\pm 1.54\mu m$. The Z-Stack in this case was found to perform well on the entire sample. In each case, the removal of post-recovery artefacts can be appreciated by comparing the CKM reconstruction with the conventional WC images.

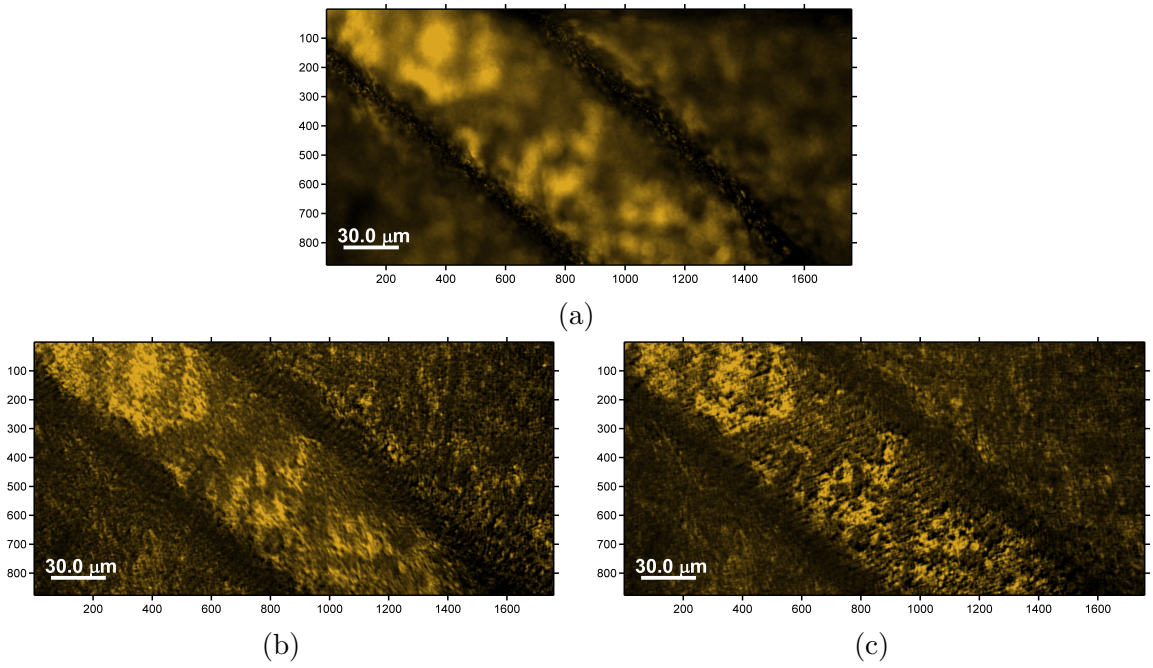


Fig. 4.11 €0.02 coin (Greek): (a) diffraction limited ($0.75NA, 40\times$), (b) single PSF recovery from I_-, and (c) single PSF recovery from I_+.

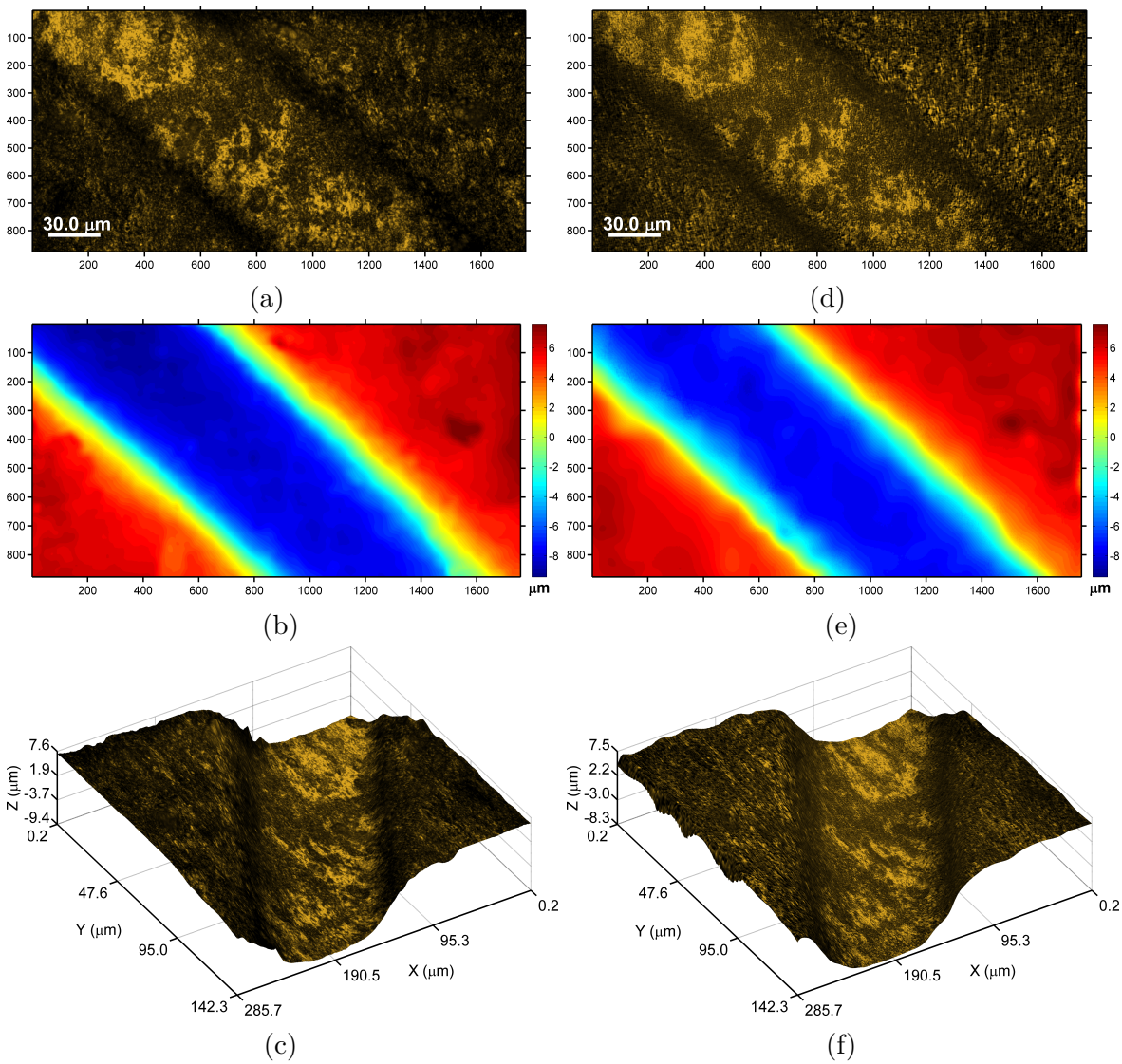


Fig. 4.12 €0.02 coin (Greek): (a) Z-Stack reconstruction, (b) Z-Stack depth profile, (c) Z-Stack 3D reconstruction, (d) CKM reconstruction, (e) CKM depth profile, and (f) CKM 3D reconstruction. Z-Stack step size: $0.1\mu m$.

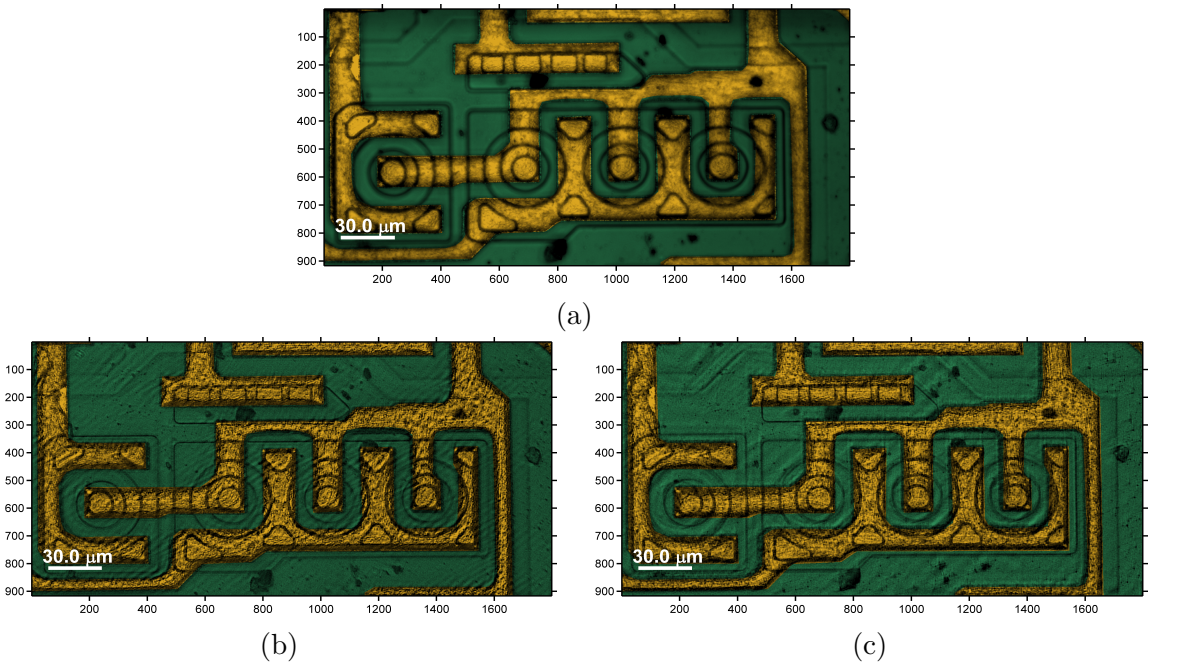


Fig. 4.13 Tracks on a silicon die: (a) diffraction limited ($0.75NA, 40\times$), (b) single PSF recovery from I_-, and (c) single PSF recovery from I_+.

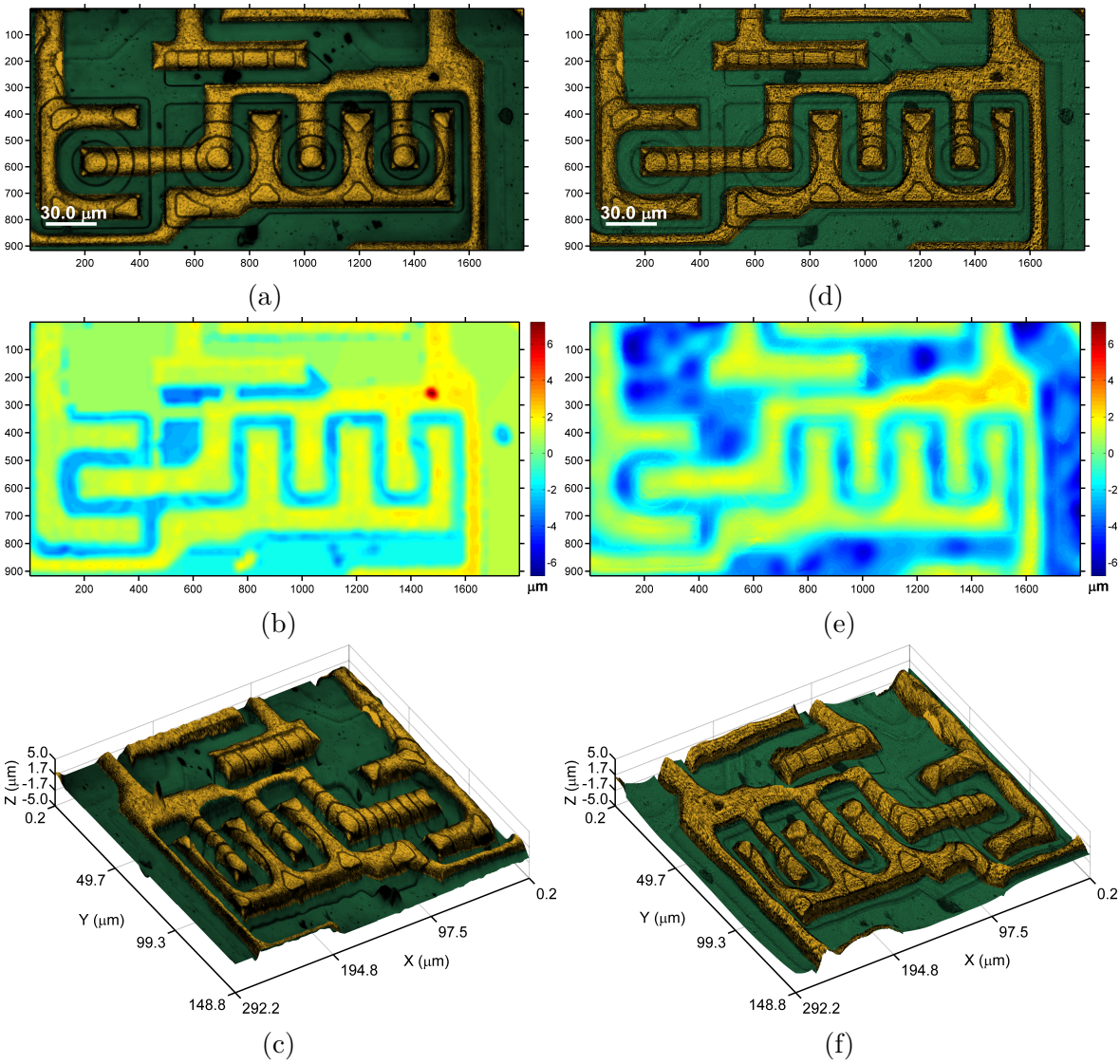


Fig. 4.14 Tracks on a silicon die: (a) Z-Stack reconstruction, (b) Z-Stack depth profile, (c) Z-Stack 3D reconstruction, (d) CKM reconstruction, (e) CKM depth profile, and (f) CKM 3D reconstruction. Z-Stack step size: $0.1\mu m$.

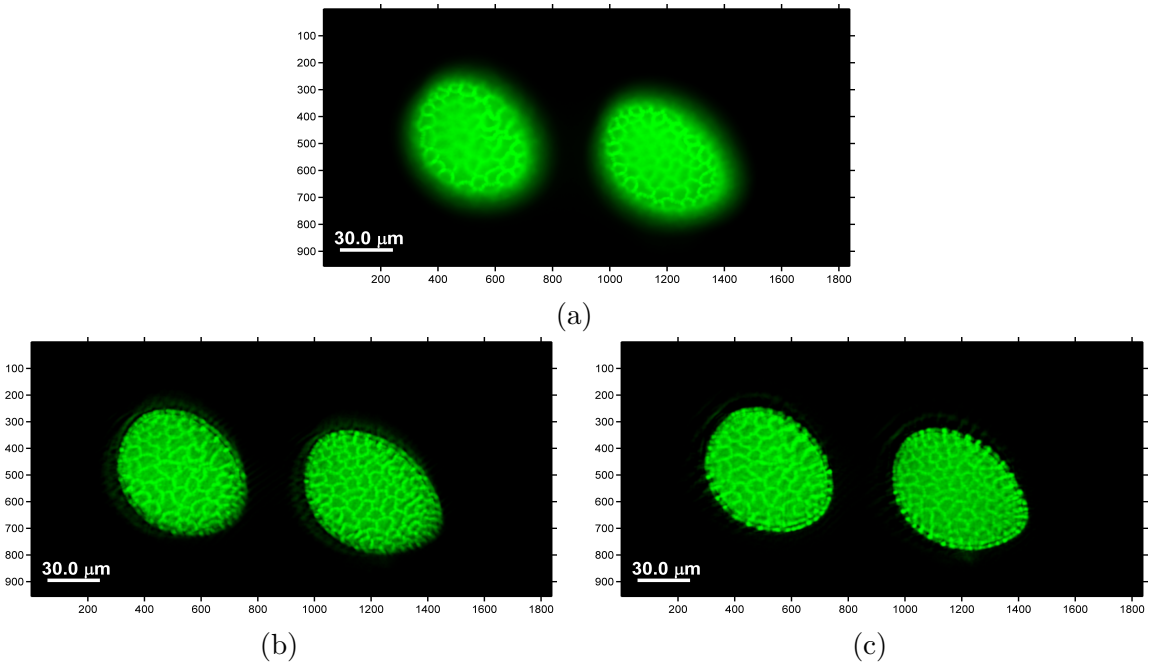


Fig. 4.15 Lily pollen grains: (a) diffraction limited $(0.75NA, 40\times)$, (b) single PSF recovery from I_-, and (c) single PSF recovery from I_+.

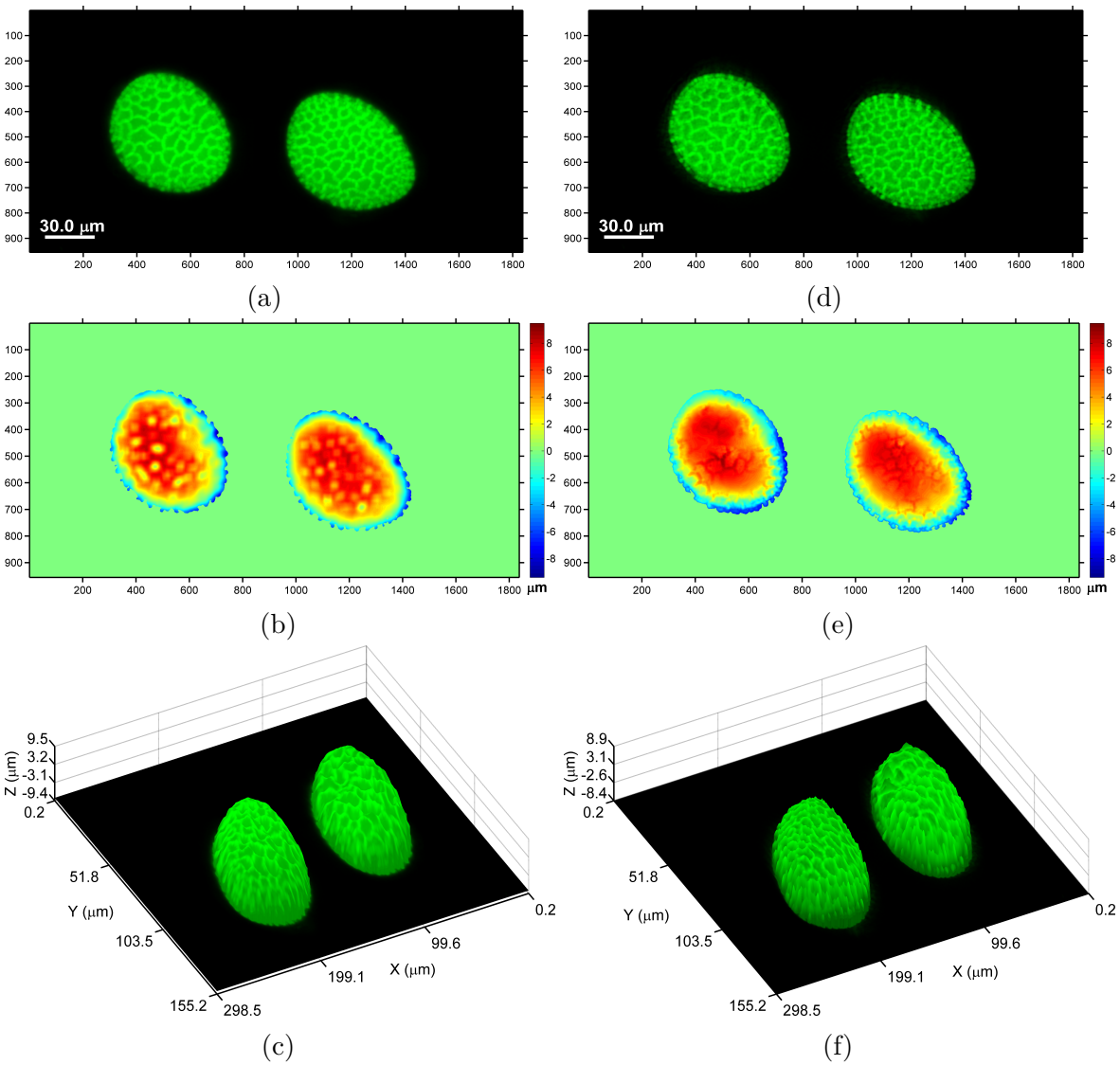


Fig. 4.16 Lilly pollen grains: (a) Z-Stack reconstruction, (b) Z-Stack depth profile, (c) Z-Stack 3D reconstruction, (d) CKM reconstruction, (e) CKM depth profile, and (f) CKM 3D reconstruction. Z-Stack step size: $0.1\mu m$.

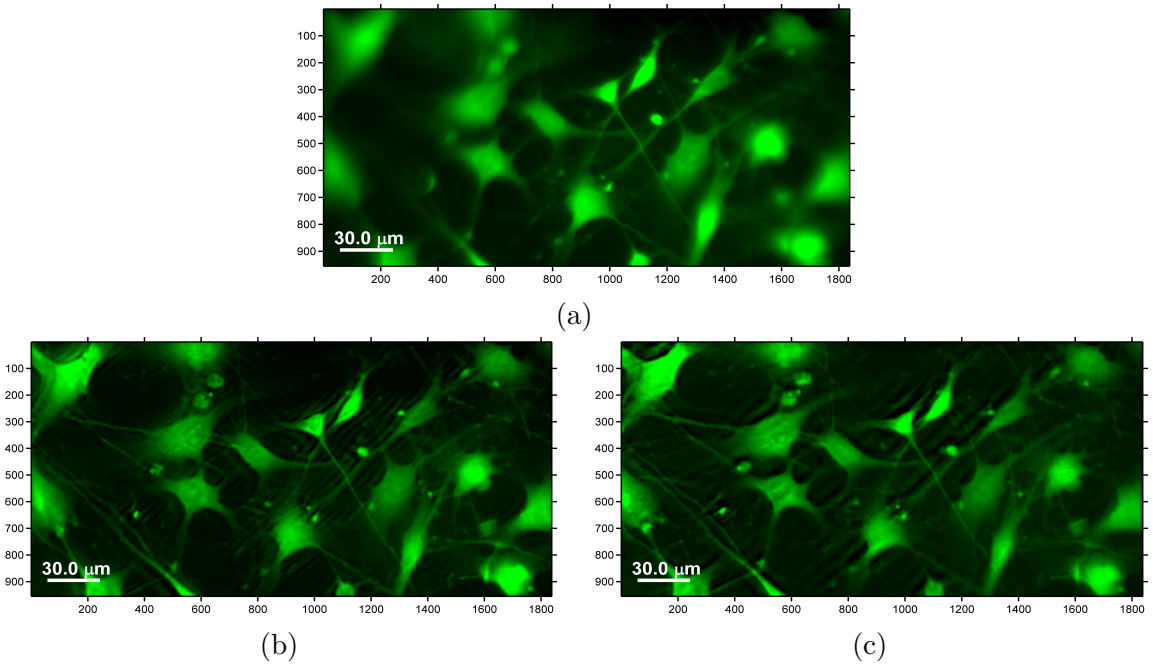


Fig. 4.17 Schwann (neurilemma) cells on a tilted slide: (a) diffraction limited $(0.75NA, 40\times)$, (b) single PSF recovery from I_-, and (c) single PSF recovery from I_+.

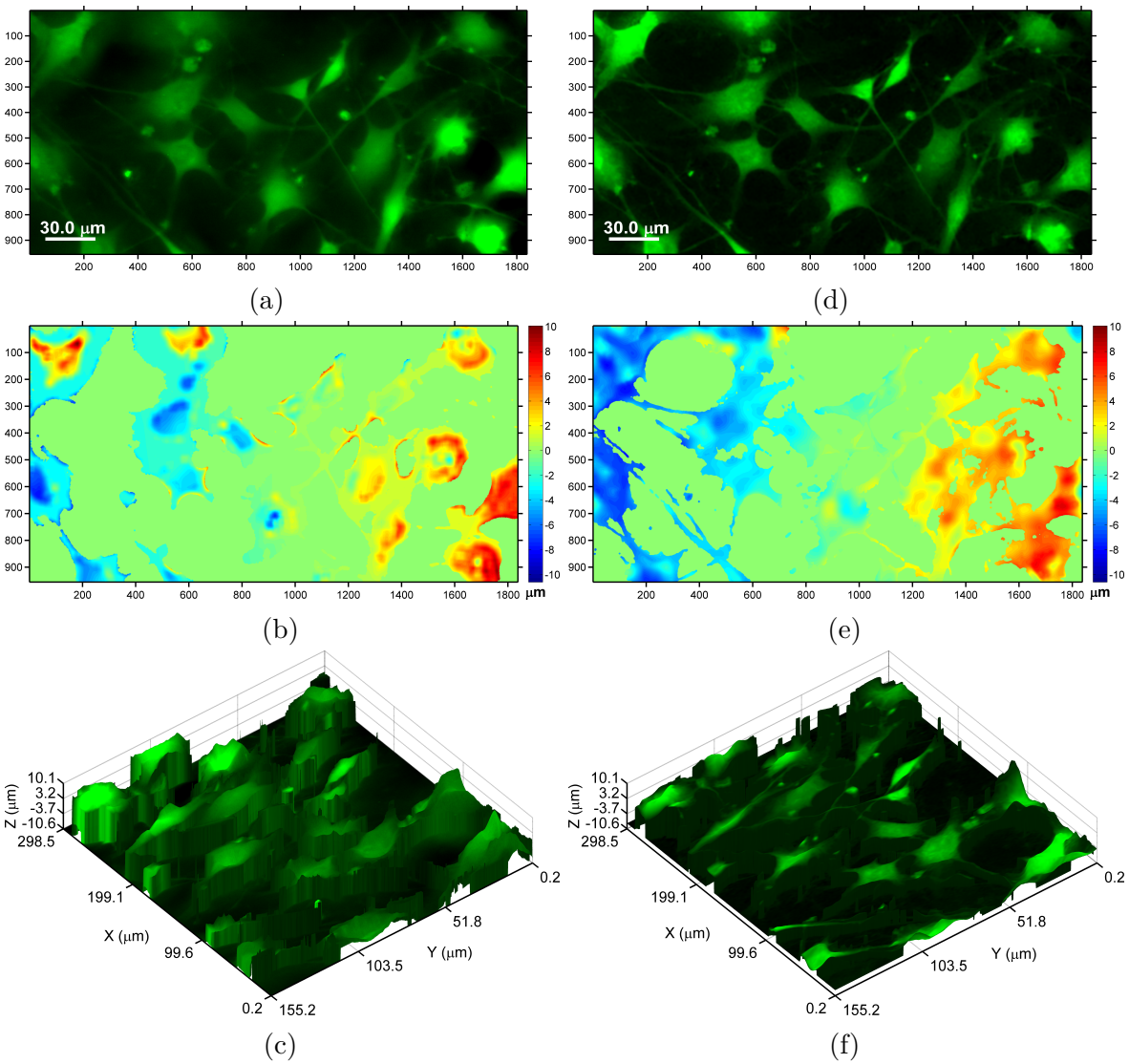


Fig. 4.18 Schwann (neurilemma) cells on a tilted slide: (a) Z-Stack reconstruction, (b) Z-Stack depth profile, (c) Z-Stack 3D reconstruction, (d) CKM reconstruction, (e) CKM depth profile, and (f) CKM 3D reconstruction. Z-Stack step size: $0.1\mu m$.

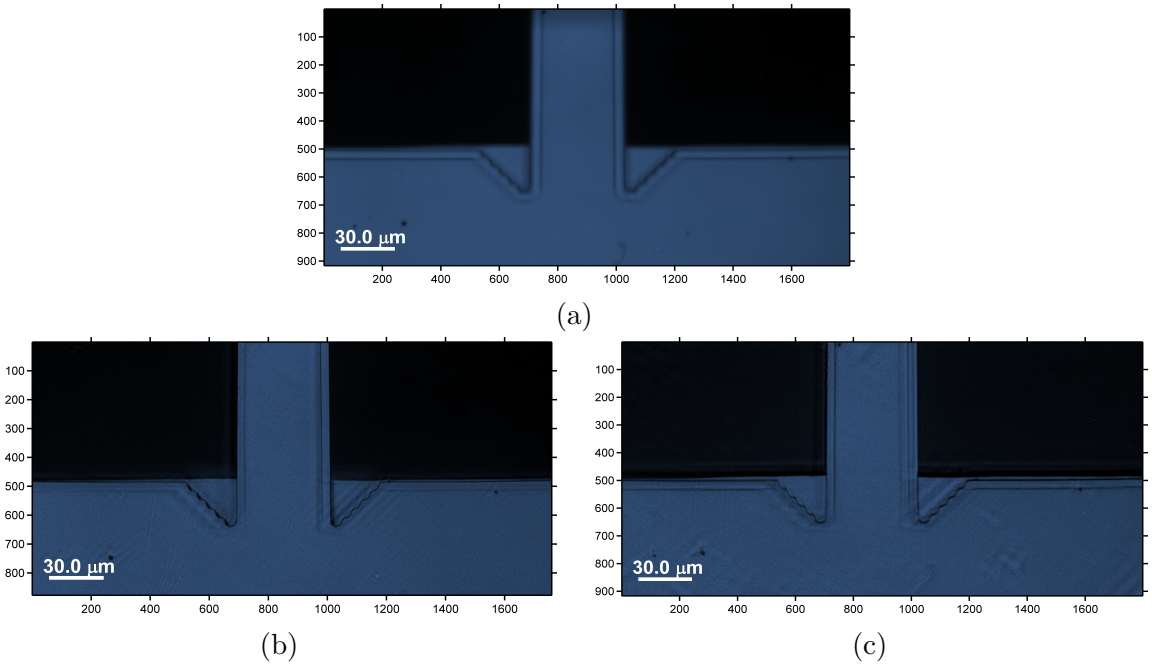


Fig. 4.19 Atomic force microscope cantilevers: (a) diffraction limited ($0.75NA, 40\times$), (b) single PSF recovery from I_-, and (c) single PSF recovery from I_+.

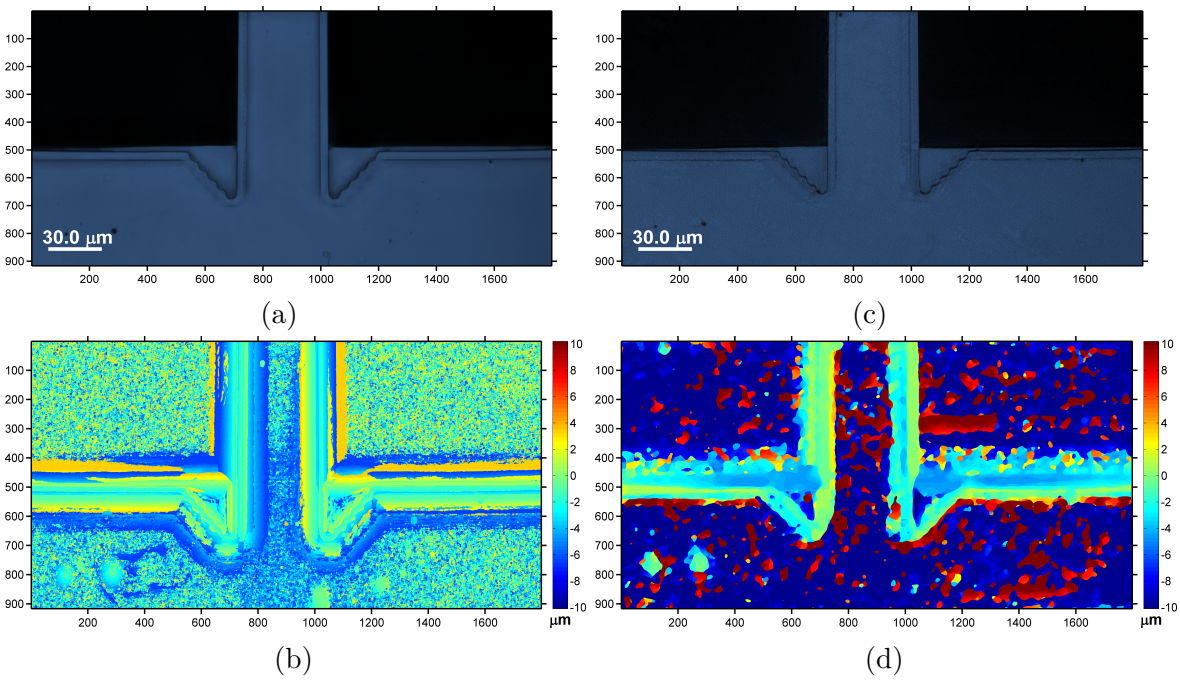


Fig. 4.20 Atomic force microscope cantilevers: (a) Z-Stack reconstruction, (b) Z-Stack depth profile, (c) CKM reconstruction, and (d) CKM depth profile. Z-Stack step size: $0.3\mu m$.

4.4 Point localisation

Owing to the large extension in DOF this technique can provide and its simplicity, there is scope to consider it for applications which rely on point or particle localisation. These include μPIV [106, 107] and localisation microscopy, which is used for super-resolution imaging and single molecule (or particle) tracking, amongst various others. The experiment reported in the section was intended as an initial evaluation of the suitability of

this technique for the mentioned applications. For this purpose, the available experimental setup originally designed for extended imaging was used despite the fact that the numerical aperture is atypically small for, for instance, super-resolution applications. Realistically scaling the results (localisation precision) to a higher numerical aperture is challenging; however, estimates can be obtained by combing measurements with numerical simulations. This approach was deemed appropriate given the scope of the experiment, and thus, it has been adopted in this evaluation.

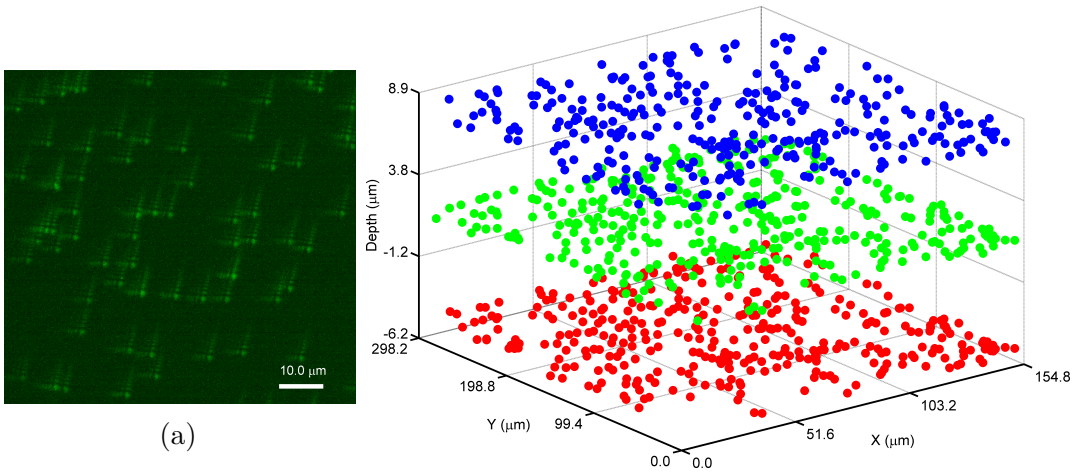
In order to accomplish this goal, a sample consisting of a single layer (*i.e.* single depth) of $0.39\mu m$ (diameter) fluorescent beads at a density of 0.009 microspheres/μm^2 was prepared. The density was kept very low in order to ensure sparseness and to avoid the additional complications associated with dense labelling [83, 108]. The samples were prepared using the procedure described in Section 4.2. These were then imaged in widefield fluorescence using the epifluorescence setup shown in Fig. 4.1, where the beads act as point emitters. By placing the sample at different axial positions and localising the beads in the FOV, the localisation precision in 3D and the accuracy of the depth measurements was determined as a function of depth. This allows us to compare the performance of this technique to that of others found in the literature; thus, fulfilling the objective of this experiment.

The average number of detected photons per emitter, N_d, was set by adjusting the integration time of the camera and was estimated from the intensity according to the following equation:

$$N_d = \frac{S}{n} \times \sum_{pixels} [(I_i - I_d)] \tag{4.7}$$

where S is the sensitivity (1.97 e^-/ADU) of the camera, I_i denotes the captured intensity image (see Fig. 4.21a), I_d is the image offset and n is the total number of emitters in the frame (~ 300). The image offset was estimated by averaging 20 dark frames taken with an empty clean coverslip in place of the sample. This was then subtracted from the image of the beads and the resultant was then thresholded to remove negative pixel values. The image was then summed, divided by the number of emitters and finally multiplied by the sensitivity, thus giving the average number of detected photons per emitter.

Localisation was performed using the algorithm described in Section 2.7. As previously noted, owing to *prior* knowledge about the shape of the PSF and to the sparseness of the sample, the full CKM technique is not required for point localisation applications. In this case, two images of the sample, I_+ and I_-, were again captured simultaneously, however, instead of deconvolving each with its respective set of 200 PSFs as per the CKM recovery algorithm, each image was deconvolved only once with its in-focus PSF, irrespective of the depth of the beads in the sample. This results in two recovered images, I_{r-} and I_{r+}. A piecewise affine transform obtained by imaging fiducial markers (beads) as described in Section 4.2 was then applied to the recovered images



(a) (b)

Fig. 4.21 Point emitter (fluorescent beads) localisation: (a) raw image of 0.39μm fluorescent beads (contrast adjusted), and (b) 3D plot of 3 layers of beads at 3 different depths.

in order to compensate for misregistrations and distortion (magnification in particular) at that particular depth. It is important to note that this does not compensate for misregistrations and depth-dependent non-linearity; however, these were taken into account in a subsequent calibration step. Each emitter in the two recovered images was then localised by least-square fitting a 2D non-isotropic (elliptical) Gaussian function where the well known Levenberg–Marquardt algorithm was used to perform the non-linear fit, as explained in Subsection 2.6.2. This yields the lateral coordinates of each emitter in each of the two recovered images. The disparity was then calculated by taking the magnitude of the vector connecting the image of the emitter in I_{r-} to its image in I_{r+}.

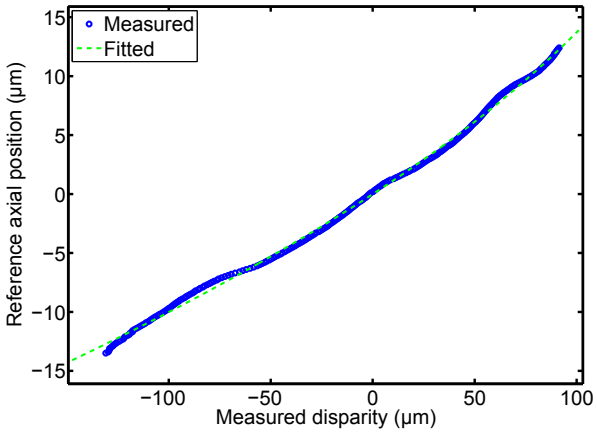


Fig. 4.22 Disparity averaged over the FOV against reference axial position.

In order to relate the measured disparity to the axial position, a second calibration step is necessary (see Subsection 2.7.2). To do this, the integration time was set to maximise the SNR and the beads were then imaged over a depth range of $\sim 30\mu m$ in steps of $0.1\mu m$. The beads were then localised as described above and a calibration curve (sometimes referred to in the literature as "look-up table" or "dictionary", which is also the term adopted in Subsection 2.7.2) relating the measured disparity to the set

axial position was constructed, as shown in Fig. 4.22. The measured disparity shown plotted in Fig. 4.22 is the average disparity of all beads across the FOV. The disparity follows a weak quadratic (*i.e.* a quadratic with the quadratic term being small compared to the linear term), therefore, a quadratic polynomial was fitted to the mean disparity and this was then used to relate disparity to axial position as desired. As a final note, it should be said that averaging the disparity over the entire FOV is not ideal if the FOV is large, because this does not take into consideration any FOV dependence. To mitigate this, rather than averaging the disparity over the entire FOV, the FOV can be divided into a number of smaller sections and a calibration curve constructed for each section individually.

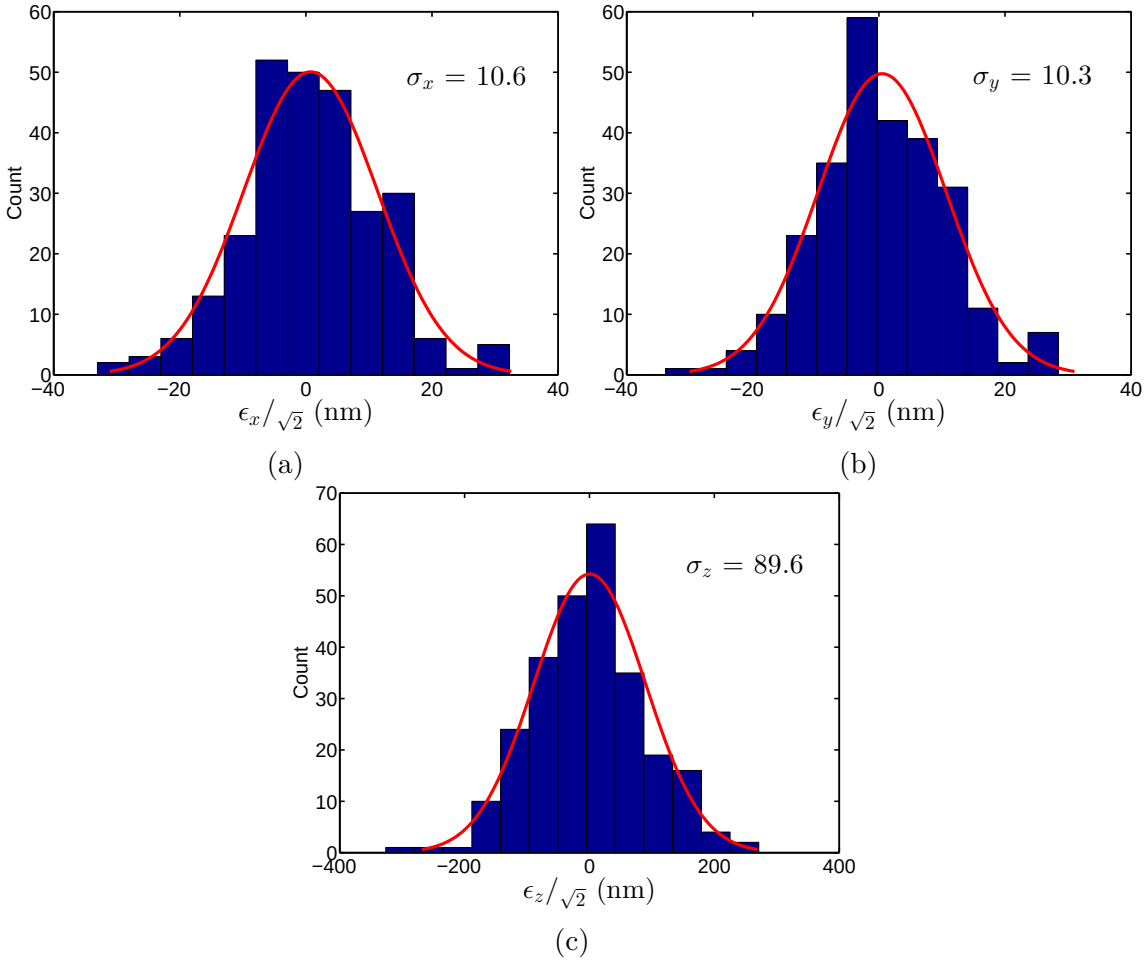


Fig. 4.23 Histograms plots of localisation errors close to the focus position: (a) histogram of errors in X position, (b) histogram of errors in Y position, and (c) histogram of errors in Z position.

In this experiment, the average number of detected photons per localisation (or per bead) was set to $\sim 4,000$. This relatively high number of detected photons was chosen because this experiment was conducted using an sCMOS camera with $\sim 3e^-$ RMS read noise, whereas EMCCDs with $< 1e^-$ RMS read noise are utilised in the vast majority of the available literature. Besides the significant difference in read-noise power between EMCCDs and sCMOS cameras, there also is the fact that each pixel in an sCMOS detector has different noise statistics due to an individual sense amplifier. This has been

shown to have an impact on the localisation precision. Despite these shortcomings, it has also been demonstrated that modern sCMOS cameras perform similarly to EMCCDs at higher photon counts (*i.e.* in the shot noise limited case) [109, 110] hence, 4,000 photons was selected and the precision was then scaled appropriately to enable comparison. Note that scaling is required anyway due to differences in wavelength and geometry between experiments.

The sample was then placed at the in-focus position and imaged over a $26\mu m$ ($-11.7\mu m$ to $14.3\mu m$) depth range at $1\mu m$ steps. In each case, a minimum of ~ 200 beads were localised in 3D as illustrated in Fig. 4.21b for 3 axial position around the focus. The precision was evaluated by capturing two sets of images of the sample at each axial position in rapid succession and then localising the beads in each of them [111]. In the absence of noise, the (x, y, z) location of each bead in each of the acquired set of images should be identical, therefore, any resultant differences must constitute error. The precision can therefore be calculated as follows:

$$\sigma_p = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(\epsilon_{p,i} - \bar{\epsilon}_p\right)^2\right]}{2\left(n-1\right)}} \tag{4.8}$$

where p denotes the dimension of space: x, y or z, $\epsilon_{p,i}$ denotes the i^{th} error along the p^{th} dimension of space (*i.e.* the difference in the positions of the i^{th} bead along the p^{th} dimension of space as inferred from each of the two acquired sets of images), $\bar{\epsilon}_p = E\left[\epsilon_{p,1}, \epsilon_{p,2}, ..., \epsilon_{p,n}\right]$ is the mean error along the p^{th} dimension, and n is the total number of localised beads. It is important to note that by Bienaymé formula, taking the difference between two such data sets will double the variance of the error under the assumption that: (1) they both have equal variances and (2) each data set is independent. The first assumption was assumed to hold because both data sets are realisations of the same process. In order to explore the second assumption further, we note that the errors, $\epsilon_{p,i}$, can be expressed as:

$$\epsilon_{p,i} = \left(P_{o,i} + \delta p_{1,i}\right) - \left(P_{o,i} + \delta p_{2,i}\right) = \delta p_{1,i} - \delta p_{2,i} \tag{4.9}$$

where $P_{o,i}$ represents the real position in the p^{th} dimension of space of the i^{th} bead; and, $\delta p_{1,i}$ and $\delta p_{2,i}$ represent the error in the position of the i^{th} bead due to noise in set 1 and set 2 respectively. From Eq. 4.9, we can therefore write:

$$Var\left[\epsilon_p\right] = Var\left[\left(\delta p_{1,1} - \delta p_{2,1}\right), \left(\delta p_{1,2} - \delta p_{2,2}\right), ..., \left(\delta p_{1,n} - \delta p_{2,n}\right)\right] \tag{4.10}$$

Since $\delta p_1 = \left[\delta p_{1,1}, \delta p_{1,2}, ..., \delta p_{1,n}\right]$ and $\delta p_2 = \left[\delta p_{2,1}, \delta p_{2,2}, ..., \delta p_{2,n}\right]$ can safely be assumed to be independent (*i.e.* $Cov\left[\delta p_1, \delta p_2\right] = 0$) due to the stochastic nature of the noise, then, it can also be assumed that:

$$Var\left[\epsilon_p\right] = Var\left[\delta p_1\right] + Var\left[\delta p_2\right] + 2Cov\left[\delta p_1, \delta p_2\right] = 2Var\left[\delta p_1\right] = 2Var\left[\delta p_2\right] \tag{4.11}$$

where as stated previously, it has been assumed that: $Var\left[\delta p_1\right] = Var\left[\delta p_2\right]$. Given that by definition the precision is: $\sigma_p = \sqrt{Var\left[\delta p_1\right]} = \sqrt{Var\left[\delta p_2\right]}$, then, from Eq. 4.11, it follows that:

$$\sigma_p = \sqrt{\frac{Var\left[\epsilon_p\right]}{2}} \tag{4.12}$$

and thus, the additional $\sqrt{2}$ factor in the denominator of Eq. 4.8 is justified. The above analyses strictly hold only if δp_1 and δp_2 are normally distributed. Since δp_1 and δp_2 are not directly observable in this experiment, this assumption is difficult to verify. However, we know that the difference of two normally distributed random variables is also normally distributed; therefore, the errors, ϵ_x, ϵ_y and ϵ_z, should be normally distributed if δp_1 and δp_2 are. The normality of the errors was therefore investigated using a one-sample Kolmogorov-Smirnov test (a.k.a. K-test) with a 5% level of significance. The results confirm that the errors are normally distributed, thus indicating that δp_1 and δp_2 also have normally distributed populations. Histogram plots of the errors taken at a depth close to the focus position are shown in Fig. 4.23a, Fig. 4.23b and Fig. 4.23c respectively together with a fitted normal distribution. In this case, the K-test for a null hypothesis that the data comes from a normal distribution yields p-values of: 0.65, 0.70 and 0.77 respectively; thus, the null hypothesis is comfortably accepted.

The measured localisation precision (standard deviation) in the lateral dimensions (σ_x and σ_y) are shown plotted in Fig. 4.24a and in the axial (σ_z) in Fig. 4.24b where the mean precisions are: $\bar{\sigma_x} = 26.44\mu m$, $\bar{\sigma_y} = 29.10\mu m$ and $\bar{\sigma_z} = 147.76\mu m$. In order to verify these results, numerical simulations were conducted using the specifications of the experimental setup. The image of the beads was modelled by the PSF, which was calculated using the high NA model proposed in [59, 65]. Only shot noise was taken into consideration in the simulations because this is expected to be the dominant source of noise. Three hundred PSFs were then generated at each depth at steps of $1.0\mu m$. These were then localised using the same procedure used in the experiment and the results are shown in Fig. 4.24c and Fig. 4.24d for the lateral and axial precisions respectively. The mean precisions are: $\bar{\sigma_x} = 28.85\mu m$, $\bar{\sigma_y} = 28.25\mu m$ and $\bar{\sigma_z} = 157.97\mu m$. Given these results, one concludes that numerical simulations and experiment results are in good agreement. This fact was used to scale the precisions obtained experimentally on the low NA setup to a higher, more typical NA, thus enabling a comparison of this technique to others found in the literature.

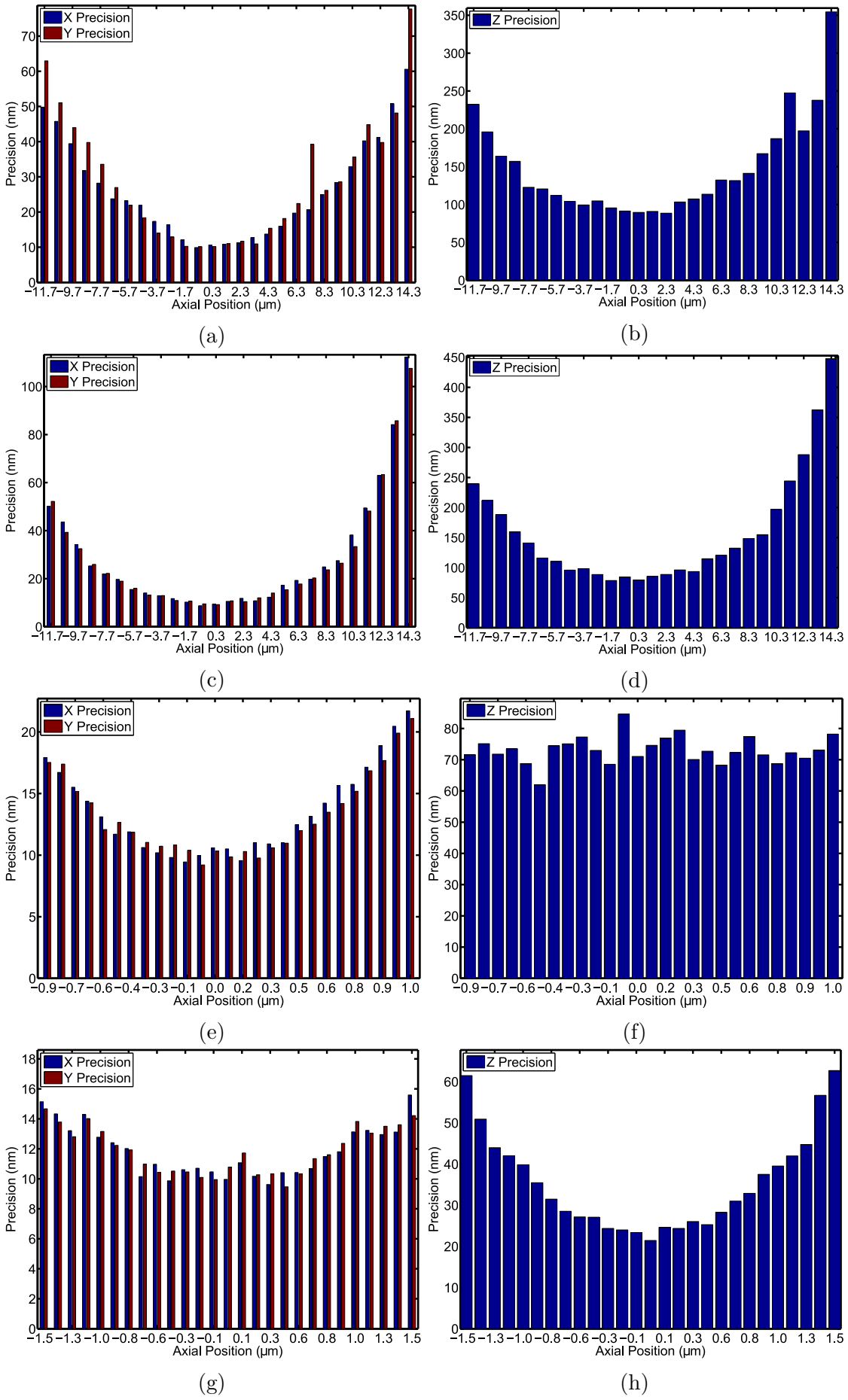


Fig. 4.24 Localisation precision at different axial positions: (a), (b) measured ($40\times$, 0.5NA, $4,000$ photons at 532nm); (c), (d) simulated ($40\times, 0.5\text{NA}, 4,000$ photons at 532nm); (e), (f) simulated ($100\times, 1.4\text{NA}, 2,000$ photons at 650nm); and, (g), (h) simulated ($100\times, 1.4\text{NA}, 2,000$ photons at 650nm). Pixel size is roughly the same in each case.

To accomplish this goal, a $100\times, 1.4$NA system was simulated. The number of photons was reduced to a total of 2,000 photons per localisation at $\lambda = 650nm$ and the pixel size of detector was set to $16\mu m$. These are the specifications of the experiment presented in [112] which is the primary competition for this technique. The difference in defocus between the two imaging channels was adjusted in order to maintain the same responsivity as in the experiment ($1.43\ pixels/\mu m$). This ensures that only the effects of the geometry, the wavelength and the number of detected photons are taken into consideration. The resultant precisions are shown in Fig. 4.24e and Fig. 4.24f, where the mean precisions are: $\bar{\sigma_x} = 13.49\mu m$, $\bar{\sigma_y} = 13.25\mu m$ and $\bar{\sigma_z} = 73.08\mu m$. By scaling the experimental precisions by the ratio of the mean simulated precision, an estimate of the precisions expected on a $100\times, 1.4$NA system from 2,000 photons/localisation at $\lambda = 650nm$ can therefore be obtained. Using this approach, the scaled precisions are: $\bar{\sigma_x} = 12.37\mu m$, $\bar{\sigma_y} = 13.65\mu m$ and $\bar{\sigma_z} = 68.35\mu m$. This approach was deemed the most reliable and comprehensive way of scaling the precisions between different system specifications. It should be noted that no rigorous study of the relationship between the localisation precision and system specifications for this technique or the cubic PSF is available in the literature or has been conducted in this work; therefore, the approximate relationships applicable for the Gaussian PSF (such as: $\sigma \propto \lambda$, and $\sigma \propto {}^{1}/\sqrt{N_d}$ where λ is the emission wavelength and N_d the number of detected photons [113, 114]) cannot be assumed to hold.

Lastly, the difference in defocus between the two imaging channels was adjusted in order to maximise the precision (*i.e.* minimise the standard deviations) without changing the other parameters. It should be noted that increasing the difference in defocus results in an increase in the responsivity and therefore a corresponding improvement in precision. On the other hand, the additional defocus also reduces the SNR, and hence, also the precision. This implies that there is an optimal value of defocus at which the precision is maximised. In simulation, this optimal value was determined by sweeping the difference in defocus and selecting the one which minimises the error in localisation. Simulations were then conducted using this value and the resultant precisions are shown plotted in Fig. 4.24g and Fig. 4.24h over a depth range of $3\mu m$. The mean precisions are: $\bar{\sigma_x} = 11.88\mu m$, $\bar{\sigma_y} = 11.91\mu m$ and $\bar{\sigma_z} = 35.42\mu m$. By scaling the measured precision as previously described, estimates of the best localisation precision attainable with this technique can be obtained. These are: $\bar{\sigma_x} = 10.88\mu m$, $\bar{\sigma_y} = 12.27\mu m$ and $\bar{\sigma_z} = 33.13\mu m$.

The mean measured precisions are shown tabulated in Tbl. 4.4, together with the typical performance of other established 3D localisation techniques commonly used in localisation microscopy of comparative complexity - *i.e.* astigmatism [115] and biplane [116]. Interferometric techniques such as the 4Pi detection scheme [117] and iPALM [118] which can achieve $\sim 10nm$ axial and $\sim 20nm$ lateral precision over a $650nm$ imaging depth have not been considered, due to their complexity, which was

Name	$\bar{\sigma_x}$ (nm)	$\bar{\sigma_y}$ (nm)	$\bar{\sigma_z}$ (nm)	Number of photons per localisation	Depth range (μm)	System specifications	Reference
Astigmatism ($\Delta^{\dagger} = 475nm$)	25	25	55	400	1.0	$1.2\text{NA}, 63\times$ $\lambda = 585nm$	[66]
Biplane ($\Delta^{\ddagger} = 500nm$)	25	25	50	480	1.5	$1.2\text{NA}, 63\times$ $\lambda = 585nm$	[66]
SB-PSF	13	13	15	2,000*	3.0	$1.4\text{NA}, 100\times$ $\lambda = 650nm$	[112]
This technique	26	29	148	4,000	26.0	$0.5\text{NA}, 40\times$ $\lambda = 532nm$	n/a

† axial distance between the sagittal and tangential foci
‡ axial distance between the foci of the two focal planes
∗ in the central lobe only

Table 4.4 Summary of point localisation results together with performance of established 3D localisation techniques. Note: quantities retrieved from the literature are approximate.

deemed to classify them in a different category from the ones considered here. On the other hand, although not as well-established as the previously mentioned techniques, the self-bending PSF (SB-PSF) [112] has been considered, due to its relevance to this work. The SB-PSF is generated from a slightly modified CPM, however, like the technique proposed here, the detection scheme used in [112] still relies on the measurement of the lateral shift. Furthermore, the disparity is generated by means of two conjugate pupil functions in a manner very similar to the conjugate-masks CKM. These similarities between the SB-PSF and the work reported here make the SB-PSF particularly relevant to this study.

As stated previously, due to the different geometries, signal levels and emission wavelength, it is difficult to compare the techniques tabulated in Tbl. 4.4. To circumvent this problem, the performance of each technique was referred to a $1.4NA, 100\times$ system with 2,000 detected photons per localisation at $\lambda = 650nm$. All figures tabulated in Tbl. 4.4 were obtained by localising beads; thus, the level of background fluorescence is expected to be similar for all. On the contrary, whereas the effective pixel size in the SB-PSF experiment and the experiment conducted here is approximately the same ($\sim 160nm$), for the astigmatic and the biplane experiments, the effective pixel size used was 1.6 times larger. This effective pixel size, however, is very close to the optimal value for Gaussian PSFs (which is equal to the standard deviation of the distribution [114]) therefore, it does not place either of the techniques at a disadvantage. As a result, this discrepancy was allowed without compensation. It should be noted that the optimal effective pixel size for either the SB-PSF or the technique proposed here have not been

Name	$\bar{\sigma_x}$ (nm)	$\bar{\sigma_y}$ (nm)	$\bar{\sigma_z}$ (nm)	Depth range (μm)	Reference
Astigmatism ($\Delta^{\dagger} = 475nm$)	12	12	27	0.7	[66]
Biplane ($\Delta^{\ddagger} = 500nm$)	14	14	27	1.1	[66]
SB-PSF	13	13	15	3.0	[112]
This technique	12 (11)*	14 (12)*	68 (33)*	1.8 (3.0)*	n/a

† axial distance between the sagittal and tangential foci
‡ axial distance between the foci of the two focal planes
∗ best performance achievable

Table 4.5 Comparison of the obtained results to other established 3D point localisation techniques. All figures are scaled for 100×, 1.4NA and 2,000 detected photons per localisation at $\lambda = 650nm$.

investigated mainly due to the difficulties posed by their complex distributions. Finally, to compensate for different wavelengths and geometry, the precisions of the astigmatic and biplane techniques were assumed to be proportional to $\lambda/\sqrt{N_d}$ [113, 114]. This approximate relationship holds (to a first-order) for Gaussian PSFs only. The scaled results are shown tabulated in Tbl. 4.5, from which a comparison amongst them is tractable.

From the table above, it is immediately apparent that the lateral precisions for all three techniques are comparable. On the other hand, the SB-PSF shows the largest depth range of $3\mu m$ followed by this technique with $1.8\mu m$. Note, however, that the depth range of this technique can be easily extended by increasing the coding strength α of the CPM and by adjusting the difference in defocus between the two imaging channels in order to compensate for the ensuing loss in SNR by an increase in responsivity and thus maintaining the precision. This capability is only offered by the technique presented here, therefore it can be considered as one of its main advantages over the other techniques. Furthermore, as previously mentioned, simulations suggest that a depth range of $3\mu m$ on a 100×, 1.4NA with a mean axial localisation error of $33.13\mu m$ is achievable even by simply increasing the difference in defocus without changing α. On the other hand, both the biplane and the astigmatic techniques have a fixed depth range beyond which the localisation error increases very rapidly; therefore, one concludes that they are mostly suitable for 3D localisation on relatively thin samples.

Regarding axial precision, it is clear that the technique proposed here performs worse than all the others considered under the specified experimental conditions. One reason for this is the fact that the cubic PSF is by far the largest (*i.e.* spreads over the largest number of pixels) PSF of all considered. The astigmatic and the biplane

techniques utilise a Gaussian PSF. Owing to the compact support of this distribution, the collected photons are concentrated on a small number of pixel, thus resulting in a high SNR. Away from the focus, the support of a Gaussian PSF increases rapidly, limiting the depth range over which a satisfactory precision can be obtained with these techniques. The SB-PSF on the other hand is essentially an apodized cubic PSF and is approximately Gaussian in shape. Furthermore, it maintains its profile over a large depth range, thus permitting localisation to be performed with high precision over a relatively large depth range. On the down side, the apodisation necessary to implement the SB-PSF reduces its maximum optical efficiency to approximately 70%, which is one of the major disadvantages of this technique. Note that in the comparison performed here, the number of detected photons was set to 2,000 photons for all techniques, effectively placing the SB-PSF at an advantage over the rest.

Besides the support of the PSF (or distribution), the responsivity (shift in the lateral dimensions per micron of depth) has a critical impact on the axial localisation precision of the SB-PSF and the technique presented here. To understand this, we can express in 1D for simplicity, the disparity in terms of the depth and the responsivity as follows:

$$D(z) = X_z^+ - X_z^- = \frac{\partial D(z)}{\partial z} \cdot z \tag{4.13}$$

where z is depth, X_z^+ and X_z^- denote the location in image coordinates of the particle in the two simultaneously acquired images at depth z, $D(z)$ is the disparity as a function of depth and $R(z) = {}^{\partial D(z)}/_{\partial z}$ is the responsivity. This readily lends itself to error analysis, therefore, we can write:

$$\begin{aligned} {\sigma_{D_o}}^2 &\approx {\sigma_{X_{z_o}^+}}^2 + {\sigma_{X_{z_o}^-}}^2 \approx \left|\frac{\partial D(z_o)}{\partial z}\right|^2 \cdot {\sigma_{z_o}}^2 \\ \implies {\sigma_{z_o}}^2 &\approx \frac{{\sigma_{X_{z_o}^+}}^2 + {\sigma_{X_{z_o}^-}}^2}{\left|\frac{\partial D(z_o)}{\partial z}\right|^2} \end{aligned} \tag{4.14}$$

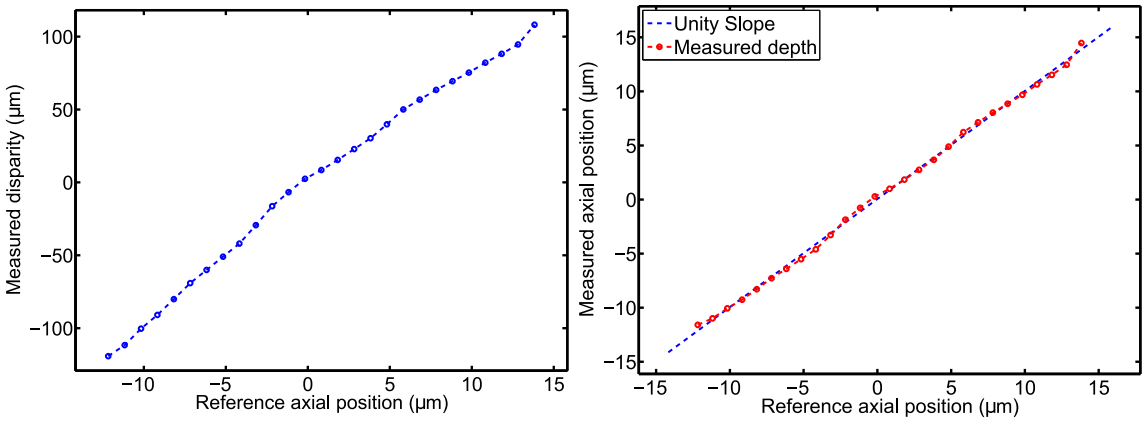
where σ_{D_o} is the error in disparity, $\sigma_{X_{z_o}^+}$ and $\sigma_{X_{z_o}^-}$ are the errors in localising the point in the two acquired images, and σ_{z_o} is the error in axial position at a particular depth $z = z_o$. Note that $\sigma_{X_{z_o}^+}$ and $\sigma_{X_{z_o}^-}$ do not correspond to the lateral localisation errors and that normality has been assumed in Eq. 4.14. Since $\sigma_{X_{z_o}^+}$ and $\sigma_{X_{z_o}^-}$ depend exclusively on the distribution and the SNR, Eq. 4.14 implies that the only control we have on the axial resolution is through the responsivity. From the data published in [112], the responsivity of the SB-PSF is: $R(z) = 56.22 \cdot z\,[\mu m/\mu m]$ whereas the scaled responsivity for this technique is: $R(z) = 25.64\,[\mu m/\mu m]$. This implies that the responsivity of the SB-PSF will exceed that of this technique for $z \geq 456nm$. Following this investigation, one therefore concludes that the SB-PSF performs better than this technique for two reasons: (1) it has a smaller support, hence a better SNR and (2) its responsivity

exceeds that of this technique, with the specified difference in defocus at a relatively small depth.

Lastly, the case of the optimal difference in defocus (Fig. 4.24g and Fig. 4.24h) will be considered, which, as previously stated, corresponds to the best performance expected from this technique. In this case, the scaled responsivity is: $R(z) = 77.18\,[\mu m/\mu m]$ therefore the responsivity of the SB-PSF will exceed that of this technique for $z \geq 1.17\mu m$, which is more than twice that of the experimental value. It is worth noting that despite the fact that this technique has a higher responsivity for $z < 1.17\mu m$, the SB-PSF still displays a precision roughly twice as good over this depth range. This can only be caused by the fact that the SB-PSF has a better SNR compared to the cubic PSF, due to a smaller support. Furthermore, this also suggests that if the cubic PSF used in this experiment is replaced by the SB-PSF while keeping the rest of the biplane CKM setup unchanged, the precision reported in [112] can be further improved.

The last criterion which will be considered is the simplicity (or complexity) of the techniques considered above. This is an important aspect of any technique, as proven by the fact that this is the central theme of various research endeavours [119, 120]. Both the astigmatic and the biplane methods have the advantage of a very simple (the simplest of all four methods considered) implementation and calibration; however, they suffer from a relatively small imaging depth which cannot be increased easily. The SB-PSF is substantially more complex to implement than the former two techniques but is still significantly less complicated than, for instance, interferometric techniques. The advantage of the SB-PSF is its large imaging depth of $3.0\mu m$, exceeded only by that of the saddle-point PSF (SP-PSF) [121], which is reported to achieve imaging depths of up to $5\mu m$ with a moderate localisation precision of $\sim 50nm$ from $\sim 3,470$ photons/localisation on a $1.4\text{NA}, 100\times$ system. The technique presented here has only one additional complication over the biplane and the astigmatic techniques: the addition of the CPM in the pupil plane. Compared to the SB-PSF, this technique is significantly simpler to implement and can provide a similar or even larger imaging depth. Furthermore, as previously mentioned, the imaging depth can easily be tailored to the application by proper selection of the design parameters δw and α -a feature not provided by the SB-PSF. Finally, in contrast to this technique, the detection scheme of the SB-PSF makes no use of deconvolution. Instead, what amounts to amplitude modulation of the pupil (apodisation), in order to suppress the side-lobes of the cubic PSF (which is what the SB-PSF is) and thus allow it to be localised directly, was employed. This simplifies the computational process at the expense of optical efficiency - a penalty which can be prohibitive for a number of applications. In fact, using the SB-PSF, the maximum attainable efficiency is $\sim 70\%$ whereas the technique proposed here does not have a theoretical maximum efficiency limit. If an SLM is used instead, the efficiency drops to $\sim 30\%$, as reported in [112], thus the use of an SLM is not an option for photon-limited applications. We therefore conclude that the technique

proposed here has a significant advantage in terms of simplicity over the SB-PSF and is of a complexity comparable to that of the astigmatic and biplane techniques.



(a) (b)

Fig. 4.25 Mean depth plots of localised beads: (a) mean disparity of beads (b) mean depth of beads.

Finally, the accuracy of the technique will be considered, although this is often left untreated in the literature. Fig. 4.25a shows the mean measured disparity of the beads at each axial position, whereas Fig. 4.25b shows the depth extracted from the disparity using the calibration curve shown in Fig. 4.22. Also shown in Fig. 4.25b is the unity-slope line which delineated the ideal response. The slope of a line fitted to the measured depths using a least-squares fit is $0.9948\mu m/\mu m$. Given that this is very close to unity, one can conclude that the linearity of the technique is good. The RMS error with respect to the unity-slope line (which can also be referred to as the accuracy) was $\sim 0.28\mu m$ (corresponding to $20.57nm$ for $1.4\text{NA}, 100\times$ system). The major cause of this inaccuracy is probably the regional dependence of the shift. This can be mitigated by performing the calibration on sections of the FOV rather than on the entire FOV, or, equivalently, by restricting the FOV over which localisation is performed. Furthermore, the localisation of the points by fitting a Gaussian function to the recovered points can potentially also cause inaccuracy because the recovered points are strictly not Gaussian in shape [82].

Following this first appreciation of the potential of this technique for 3D localisation applications, one concludes that at best, this technique yields mean localisation precisions comparable to those of the conventional astigmatic and biplane techniques but over roughly three times the depth range and with minimal additional complexity. Compared to the SB-PSF, for the same number of detected photons, the best precision achievable by this technique is roughly twice as bad in depth and comparable in the lateral dimensions; however, this is achieved using a significantly simpler optical setup. Furthermore, this technique utilises all incoming photons whereas the SB-PSF has a maximum efficiency of approximately 70%. Lastly, if the SB-PSF is used instead of the cubic PSF in the optical setup proposed here, an improvement in the localisation precision of the SB-PSF can potentially be obtained, together with simplification of the SB-PSF optical

setup. Nevertheless, with the advent of bright and very bright fluorophores such as Cy2 and Cy5.5 which yield 6,241 and 5,831 photons per switching cycle [122] respectively, attaining lateral localisation precision of $\sim 8nm$ in the lateral dimensions and $\sim 15nm$ in depth over a $3.0\mu m$ depth range using this technique without any modifications should be feasible. This, coupled with the large depth range which this technique offers, and its simplicity, brings us to the conclusion that there is scope for further investigation of this technique for 3D localisation applications.

4.5 Conclusions and future work

In conclusion, we have demonstrated 3D topographic imaging of extended scenes or samples (both natural and synthetic) with an extended DOF (> 8 times the conventional DOF) in a single-snapshot with an error of $\pm 0.15\mu m$ on a binary systematic scene consisting of regularly spaced dots from an image SNR of $60dB$. The results show that CKM eliminates WC artefacts and distortion; hence, a better image quality compared to that given by standard WC is obtained. The depth maps obtained on real samples of arbitrary shape, texture and features, are comparable to those given by a focus-stack from tens of images.

Furthermore, an initial investigation of the suitability of this technique for 3D localisation of sparse point-samples has been conducted. It has been shown experimentally that, using this technique, sub-resolution fluorescent beads can be localised with a mean precision of $26 - 29nm$ in the lateral dimensions and $148nm$ in the axial dimension over a depth range of $26\mu m$ from 4,000 photons/localisation. These results were obtained on a $40\times, 0.5$NA system using a wavelength of $532nm$. By scaling these experimental results for the optimal difference in defocus between the two acquired images and more typical system specifications, the expected best achievable precisions for the considered specifications have been determined. These are: $11-12nm$ in the lateral dimensions and $33nm$ in the axial dimension over a $3.0\mu m$ depth range from 2000 photons/localisation. These figures are for a $100\times, 1.4$NA system using a wavelength of $650nm$. Finally, a comparison between this technique and other established techniques revealed that there is scope for this technique, primarily due to its unique trade-off between localisation precision and simplicity, in addition to its large and adjustable depth range.

As far as future work is concerned, improvements to the experimental setup and/or procedure which became apparent during the course of this investigation have been suggested throughout this chapter. The most salient of these are listed below for reference:

- Focusing should be performed by displacing the sample rather than the objective, as is customary on inverted microscope. This is because once an objective is fitted with a CPM, the beam coming out of its back aperture will not be collimated

(assuming an infinity-corrected objective). It was not possible to investigate and quantify the effects of this; however, the fact that this constitutes a deviation from the ideal is conspicuous and thus it should be avoided.

- Deconvolution has been performed using a simple Wiener filter, as is customary in WC. However, the effect of the NSR parameter in the Wiener filter on the performance of the CKM technique, and a systematic way of selecting this have not been thoroughly investigated, mainly due to time constraints. Future development on this technique is therefore envisaged to include a study of this.

- Slabs of glass were used to induce the necessary difference in defocus, $\delta\omega$. Given that a significant thickness of glass ($24mm$) is needed, spherical aberrations which reduce localisation precision in point localisation and increase the differences between the two acquired images will result even though these are expected to be minimal due to the fact that the image-side beam is slow. More importantly, chromatic aberrations, which have a similar adverse effect on the performance of this technique, will also be introduced as a consequence of the introduction of the glass slabs. Assuming a $60nm$ range around a centre wavelength of $532nm$, a longitudinal chromatic interval of $\sim 22nm$ is estimated.

- Fitting the CPM inside a standard off-the-shelf objective makes the setup more compact and robust; however, this requires precise knowledge of the location of the aperture stop of the objective. Furthermore, unless this is accessible, then the CPM cannot be fitted in the required plane, resulting in field-dependent shifts and PSFs. One way of mitigating this problem is to re-image the aperture stop of the objective such that the CPM can be placed outside the objective. This permits the CPM to be positioned accurately in a plane conjugate to the aperture stop at the expense of a more complicated optical setup. More importantly, if this implementation is adopted, then, displacing the sample rather than the objective is imperative, since moving the objective necessarily entails displacing the aperture stop and thus also its conjugate plane. This effect is again expected to be small due to the large image-side DOF of a typical microscope, however, given that the shifts which are to be measured are also very small, it is best avoided.

- The flatness of the calibration target (beads attached to a coverslip) is critical to the calibration procedure used here and the higher the effective NA of the imaging system is, the more important this becomes. This, however, cannot be guaranteed when using a sample such as the one used here. One way of improving on this is to scan a small region of the sample across the FOV using a mechanised XY-translation stage, rather than imaging the entire sample at once.

- The calibration procedure used in the point localisation experiment described above involves constructing a curve relating the measured disparity to the true axial position. To do this, the average disparity measured throughout the FOV

has been used. This method ignores regional variations in disparity caused by, for instance, distortion. Consequently, a better way of doing this would be to either restrict the FOV to a smaller region if the sample which is to be imaged permits, or else to section the FOV in smaller sections and then take the average disparity over each individual section separately. This would increase the complexity of the calibration procedure; however, it is expected to result in a significant improvement in accuracy.

- The precisions quoted for typical point localisation system specifications were obtained by scaling experimental precisions obtained on a different geometry, wavelength and number of detected photons. Although these were deemed reliable due to the excellent correlation between the experiment and the numerical analysis, direct observation is still necessary. Future work is therefore necessary in order to observe these directly, as well as to investigate the performance of this technique under other practical circumstances such as background fluorescence, fluorophore overlap (multiple-fluorophore images) [83, 123], fixed-orientation dipoles [124, 125] (*i.e.* non-isotropic emitters) and the effect of photo bleaching.

- Finally, in this experiment, the beads were localised by least-square fitting a 2D Gaussian to their deconvolved image. This was done mainly due to the simplicity of the method; for instance, contrary to the MLE method, this method does not require rigorous characterisation of the detector noise [83]. However, the shape of a recovered point is not Gaussian and it is also dependent on the NSR used in the Wiener filter. Given that misspecification of PSF can result in systematic error [82], a better alternative to this might be to fit pre-acquired experimental PSFs [116]. The fitting can also be performed using the MLE method which is known to out-perform the least-square fitting if the noise characteristics of the detector are known. These modifications are envisaged to result in better accuracy and can potentially also improve the precision of the technique.

Chapter 5

Conclusion and future work

In conclusion, we have proposed a new technique dubbed Complementary Kernel Matching (CKM), which fulfils the objectives set for this project, namely:

- 3D imaging (both intensity of the scene and its (dense) depth map).
- Single-snapshot operation (no scanning required).
- No reduction in optical resolution or efficiency.
- Simple to implement or to integrate in an off-the-shelf optical system such as a standard microscope.
- Solves the long-standing artefacts and distortion problems typically associated with WC systems.

An additional benefit of this technique is its monocularity, which makes it inherently immune to occlusion - a major problem effecting triangulation-based 3D imaging techniques such as stereo vision.

The theory behind CKM has been developed and several ways of implementing it have been proposed and analysed, as presented in Chapter 2. The most promising configuration was then implemented around a standard microscope. A means of calibrating the setup was proposed and several samples were then imaged (in a snapshot) in various modalities: transmission, reflection and fluorescence. A DOF of $\sim 20\mu m$ was achieved on a $40\times, 0.5$NA system with a conventional DOF of $2.4\mu m$. The results show that CKM eliminates WC artefacts and distortion; hence, a better image quality compared to standard WC -*i.e.* single PSF recovery- is obtained. The depth maps obtained on real samples of arbitrary shape, texture and features, are comparable to those given by a focus-stack from tens of images. On a systematic scene consisting of regularly spaced dots with a linear depth gradient, an RMS error of $\pm 0.15\mu m$ in depth was obtained from an image with an SNR of $60dB$. This increases to $\pm 0.25\mu m$ for an SNR of $30dB$. Like other passive techniques, however, the performance of CKM

in measuring depth depends on the quality of the features and the texture of the scene itself.

Owing to its large and adjustable DOF and its simplicity, there is scope in considering the use of the same CKM setup for applications which involve 3D point localisation, such as 3D super-resolution techniques, for instance. An initial investigation was therefore conducted by localising sub-resolution fluorescent beads. The mean precisions measured on the $40\times, 0.5$NA setup were $148nm$ in depth and $< 30nm$ in the lateral dimensions from $4,000$ photons per localisation over a DOF of $26\mu m$. From these values, we calculate that a mean localisation precision of $< 34nm$ in depth and $< 13nm$ in the lateral dimensions from $2,000$ photons per localisation over a DOF of $3\mu m$ should be achievable on a more typical $100\times, 1.4$NA system. This DOF is appreciably larger than that provided by most 3D super-resolution techniques in common use. The recently proposed SB-PSF [112] has a similar DOF and a mean localisation precision of $\sim 15nm$ in all dimensions of space. This is, however, achieved at the expense of complexity and optical efficiency. One therefore concludes that there is scope for further investigation of this technique for 3D point localisation applications.

Despite the encouraging results obtained so far, the work conducted here is only the first step towards the finalisation of this technique. Future work has been suggested in context throughout the text. Below, the most salient future work and improvements will be summarised.

- Improvements to the experimental setup as discussed in the conclusion of Chapter 4. In particular, the use of a moving sample stage rather than a moving objective stage, which would allow the CPM to be mounted outside the objective, and a better means of producing the difference in defocus between the two channels, as required by the biplane method. The latter was implemented using glass slabs. A better way of doing this would be to use a beam splitter and two right-angle prisms in order to direct the perpendicular beam onto the appropriate half of the detector. By adjusting the position of the prisms, the separation between the two images can be adjusted; hence, any sensor size can be accommodated. Furthermore, the difference in defocus between the two imaging channels can also be adjusted. This option, therefore, has the advantage of flexibility.

- One of the least explored areas of this technique is the image processing. As stated in Chapter 2, the results presented in this thesis were obtained using a relatively basic stereo vision block matching algorithm. However, there is a wealth of literature and methods on the subject which are yet to be tested on CKM. Of particular relevance are global methods such as dynamic programming [74]. These can be used in conjunction with local methods such as block matching algorithms to alleviate the dependence of CKM on scene features and texture. Another important aspect of the processing which CKM involves is the effect

of the NSR parameter in the deconvolution. Here, this parameter was used as is customary in WC; that is, by manual optimisation. However, as discussed in Chapter 4, CKM operates at the limits of the DOF provided by the CPM for a given coding strength. As a consequence, rigorous modelling of the dependence of this parameter on depth is required.

- CKM as presented in this work is a passive technique which relies completely on the features and quality of the texture of the sample or scene itself in order to infer depth. As a result, the depth of smooth samples, such as the atomic force microscope cantilevers discussed in Chapter 4, cannot be measured. However, an active version of this technique can be realised by using structured illumination to induce texture/features in the sample. This would reduce the reliance of the sample itself at the expense of single-snapshot operation. Note that CKM would still have the advantage of monocularity and hence immunity to occlusion, over techniques such as structured light, which was introduced in Chapter 1.

- As far as 3D point localisation is concerned, the work conducted here is an initial evaluation of the capability of this technique for the mentioned application. This was conducted on sub-resolution bead samples, as is common in the literature. This implies that certain important practicalities, such as the effect of background fluorescence, dense labelling and fixed dipoles for instance, have not yet been considered. These factors are, however, application-specific, therefore they are best accounted for by directly imaging the sample of interest.

References

[1] E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Appl. Opt.*, vol. 34, no. 11, pp. 1859–1866, 1995.

[2] M. Demenikov and A. R. Harvey, "Image artifacts in hybrid imaging systems with a cubic phase mask," *Opt. Express*, vol. 18, no. 8, pp. 8207–8212, 2010.

[3] G. Sansoni, M. Trebeschi, and F. Docchio, "State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation," *Sensors*, vol. 9, no. 1, pp. 568–601, 2009.

[4] F. Chen, G. M. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," *Opt. Eng.*, vol. 39, no. 1, pp. 10–22, 2000.

[5] T. Mueller, A. Poesch, and E. Reithmeier, "Measurement uncertainty of microscopic laser triangulation on technical surfaces," *Microsc. Microanal.*, vol. 21, no. 06, pp. 1443–1454, 2015.

[6] F. Blais, "Review of 20 years of range sensor development," *J. Electron. Imaging*, vol. 13, no. 1, 2004.

[7] M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.*, vol. 23, no. 21, pp. 3837–3844, 1984.

[8] S. Lang, L. Korba, F. Blais, and M. Lecavalier, "Characterization and testing of the biris range sensor," in *Proc. IEEE I2MTC*, pp. 459–464, IEEE, 1993.

[9] G. Sansoni and E. Redaelli, "A 3d vision system based on one-shot projection and phase demodulation for fast profilometry," *Meas. Sci. Technol.*, vol. 16, no. 5, p. 1109, 2005.

[10] H. Schwenke, U. Neuschaefer-Rube, T. Pfeifer, and H. Kunzmann, "Optical methods for dimensional metrology in production engineering," *CIRP Ann. Manuf. Techn.*, vol. 51, no. 2, pp. 685–699, 2002.

[11] W. Drexler, M. Liu, A. Kumar, T. Kamali, A. Unterhuber, and R. A. Leitgeb, "Optical coherence tomography today: speed, contrast, and multimodality," *J. Biomed. Opt.*, vol. 19, no. 7, pp. 071412–071412, 2014.

[12] D. T. Moore and B. E. Truax, "Phase-locked moiré fringe analysis for automated contouring of diffuse surfaces," *Appl. Opt.*, vol. 18, no. 1, pp. 91–96, 1979.

[13] H. Takasaki, "Moiré topography," *Appl. Opt.*, vol. 9, no. 6, pp. 1467–1472, 1970.

[14] J.-T. Oh, S. Lee, and S.-W. Kim, "Scanning projection grating moire topography," vol. 3958, pp. 46–51, 2000.

[15] A. Saxena, S. H. Chung, and A. Y. Ng, "3-d depth reconstruction from a single still image," *Int. J. Comput. Vision*, vol. 76, no. 1, pp. 53–69, 2008.

[16] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser, "Extended depth-of-focus for multi-channel microscopy images: a complex wavelet approach," in *Proc. IEEE Int. Symp. Biomed. Imaging.*, pp. 660–663, IEEE, 2004.

[17] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science*, vol. 305, no. 5686, pp. 1007–1009, 2004.

[18] J. M. Taylor, J. M. Girkin, and G. D. Love, "High-resolution 3d optical microscopy inside the beating zebrafish heart using prospective optical gating," *Biomed. Opt. Express*, vol. 3, no. 12, pp. 3043–3053, 2012.

[19] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 924–934, 2006.

[20] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[21] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, *et al.*, "Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy," *Nat. Methods*, vol. 11, no. 7, pp. 727–730, 2014.

[22] S. Quirin and R. Piestun, "Depth estimation and image recovery using broadband, incoherent illumination with engineered point spread functions [invited]," *Appl. Opt.*, vol. 52, no. 1, pp. A367–A376, 2013.

[23] A. Greengard, Y. Y. Schechner, and R. Piestun, "Depth from diffracted rotation," *Opt. Lett.*, vol. 31, no. 2, pp. 181–183, 2006.

[24] R. Piestun, Y. Y. Schechner, and J. Shamir, "Propagation-invariant wave fields with finite energy," *JOSA A*, vol. 17, no. 2, pp. 294–303, 2000.

[25] R. Berlich, A. Bräuer, and S. Stallinga, "Single shot three-dimensional imaging using an engineered point spread function," *Opt. Express*, vol. 24, no. 6, pp. 5946–5960, 2016.

[26] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graphics*, vol. 26, no. 3, p. 70, 2007.

[27] P. M. Blanchard and A. H. Greenaway, "Simultaneous multiplane imaging with a distorted diffraction grating," *Appl. Opt.*, vol. 38, no. 32, pp. 6692–6699, 1999.

[28] S. Bradburn, W. T. Cathey, and E. R. Dowski, "Realizations of focus invariance in optical–digital systems with wave-front coding," *Appl. Opt.*, vol. 36, no. 35, pp. 9157–9166, 1997.

[29] D. Hong and H. Cho, "Depth-of-field extension method using variable annular pupil division," *IEEE/ASME Trans. Mechatron.*, vol. 17, no. 2, pp. 390–396, 2012.

[30] G. Muyo and A. R. Harvey, "Decomposition of the optical transfer function: wavefront coding imaging systems," *Opt. Lett.*, vol. 30, no. 20, pp. 2715–2717, 2005.

[31] S. S. Sherif, E. R. Dowski, Jr., and W. T. Cathey, Jr., "Logarithmic phase filter to extend the depth of field of incoherent hybrid imaging systems," vol. 4471, pp. 272–280, 2001.

[32] H. Zhao and Y. Li, "Optimized logarithmic phase masks used to generate defocus invariant modulation transfer function for wavefront coding system," *Opt. Lett.*, vol. 35, pp. 2630–2632, Aug 2010.

[33] S. Mezouari, G. Muyo, and A. R. Harvey, "Circularly symmetric phase filters for control of primary third-order aberrations: coma and astigmatism," *J. Opt. Soc. Am. A*, vol. 23, pp. 1058–1062, May 2006.

[34] S. Prasad, T. C. Torgersen, V. P. Pauca, R. J. Plemmons, and J. van der Gracht, "Engineering the pupil phase to improve image quality," in *Proc. SPIE*, vol. 5108, pp. 1–12, 2003.

[35] S. Prasad, V. P. Pauca, R. J. Plemmons, T. C. Torgersen, and J. van der Gracht, "Pupil-phase optimization for extended-focus, aberration-corrected imaging systems," in *Proc. SPIE*, vol. 5559, pp. 335–345, 2004.

[36] X. Mo and J. Wang, "Phase transfer function based method to alleviate image artifacts in wavefront coding imaging system," in *Proc. SPIE*, vol. 8907, pp. 89074H–89074H–10, 2013.

[37] M. Demenikov and A. R. Harvey, "Parametric blind-deconvolution algorithm to remove image artifacts in hybrid imaging systems," *Opt. Express*, vol. 18, pp. 18035–18040, Aug 2010.

[38] H. Lei, H. Feng, X. Tao, and Z. Xu, "Imaging characteristics of a wavefront coding system with off-axis aberrations," *Appl. Opt.*, vol. 45, no. 28, pp. 7255–7263, 2006.

[39] W. Zhang, Z. Ye, T. Zhao, Y. Chen, and F. Yu, "Point spread function characteristics analysis of the wavefront coding system," *Opt. Express*, vol. 15, no. 4, pp. 1543–1552, 2007.

[40] S. Bagheri, P. E. X. Silveira, and D. P. de Farias, "Analytical optimal solution of the extension of the depth of field using cubic-phase wavefront coding. part i. reduced-complexity approximate representation of the modulation transfer function," *J. Opt. Soc. Am. A*, vol. 25, no. 5, pp. 1051–1063, 2008.

[41] M. Somayaji and M. P. Christensen, "Enhancing form factor and light collection of multiplex imaging systems by using a cubic phase mask," *Appl. Opt.*, vol. 45, no. 13, pp. 2911–2923, 2006.

[42] D. V. Widder, "The airy transform," *Am. Math. Monthly*, vol. 86, no. 4, pp. 271–277, 1979.

[43] D. Babusci, G. Dattoli, and D. Sacchetti, "The airy transform and associated polynomials," *Cent. Eur. J. Phys.*, vol. 9, no. 6, pp. 1381–1386, 2011.

[44] O. Vallée and M. Soares, *Airy functions and applications to physics*. Imperial College Press, London, 2004.

[45] W. Zhang, Z. Ye, Y. Chen, T. Zhao, and F. Yu, "Ray aberrations analysis for phase plates illuminated by off-axis collimated beams," *Opt. Express*, vol. 15, no. 6, pp. 3031–3046, 2007.

[46] G. Carles, "Analysis of the cubic-phase wavefront-coding function: Physical insight and selection of optimal coding strength," *Opt. Lasers Eng.*, vol. 50, no. 10, pp. 1377–1382, 2012.

[47] G. Dattoli, C. Chiccoli, S. Lorenzutta, G. Maino, and A. Torre, "Theory of generalized hermite polynomials," *Comput. Math. Appl.*, vol. 28, no. 4, pp. 71–83, 1994.

[48] G. Yasmin, "Some properties of legendre–gould hopper polynomials and operational methods," *J. Math. Anal. Appl.*, vol. 413, no. 1, pp. 84–99, 2014.

[49] A. Torre, "Airy polynomials, three-variable hermite polynomials and the paraxial wave equation," *J. Opt.*, vol. 14, no. 4, p. 045704, 2012.

[50] K. Kubala, E. Dowski, and W. T. Cathey, "Reducing complexity in computational imaging systems," *Opt. Express*, vol. 11, no. 18, pp. 2102–2108, 2003.

[51] P. Zammit, A. R. Harvey, and G. Carles, "Extended depth-of-field imaging and ranging in a snapshot," *Optica*, vol. 1, no. 4, pp. 209–216, 2014.

[52] I. Blayvas, R. Kimmel, and E. Rivlin, "Role of optics in the accuracy of depth-from-defocus systems," *J. Opt. Soc. Am. A*, vol. 24, no. 4, pp. 967–972, 2007.

[53] R. Blendowske, "Role of optics in the accuracy of depth-from-defocus systems: comment," *J. Opt. Soc. Am. A*, vol. 24, no. 10, pp. 3242–3244, 2007.

[54] P. Trouvé-Peloux, F. Champagnat, G. Le Besnerais, and J. Idier, "Theoretical performance model for single image depth from defocus," *J. Opt. Soc. Am. A*, vol. 31, no. 12, pp. 2650–2662, 2014.

[55] S. Ram, E. S. Ward, and R. J. Ober, "A stochastic analysis of performance limits for optical microscopes," *Multidim. Syst. Sig. Process.*, vol. 17, no. 1, pp. 27–57, 2006.

[56] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.

[57] J. Chao, E. S. Ward, and R. J. Ober, "Fisher information theory for parameter estimation in single molecule microscopy: tutorial," *J. Opt. Soc. Am. A*, vol. 33, no. 7, pp. B36–B57, 2016.

[58] G. Grover, S. R. P. Pavani, and R. Piestun, "Performance limits on three-dimensional particle localization in photon-limited microscopy," *Opt. Lett.*, vol. 35, no. 19, pp. 3306–3308, 2010.

[59] J. Schnitzbauer, R. McGorty, and B. Huang, "4pi fluorescence detection and 3d particle localization with a single objective," *Opt. Express*, vol. 21, no. 17, pp. 19701–19708, 2013.

[60] A. Tahmasbi, S. Ram, J. Chao, A. V. Abraham, F. W. Tang, E. S. Ward, and R. J. Ober, "Designing the focal plane spacing for multifocal plane microscopy," *Opt. Express*, vol. 22, no. 14, pp. 16706–16721, 2014.

[61] J. Schnitzbauer, R. McGorty, and B. Huang, "4pi fluorescence detection and 3d particle localization with a single objective," *Opt. Express*, vol. 21, no. 17, pp. 19701–19708, 2013.

[62] M. Badieirostami, M. D. Lew, M. A. Thompson, and W. Moerner, "Three-dimensional localization precision of the double-helix point spread function versus astigmatism and biplane," *Appl. Phys. Lett.*, vol. 97, no. 16, p. 161103, 2010.

[63] M. D. Lew, M. A. Thompson, M. Badieirostami, and W. Moerner, "In vivo three-dimensional superresolution fluorescence tracking using a double-helix point spread function," in *Proc. SPIE*, pp. 75710Z–75710Z, 2010.

[64] S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, "High accuracy 3d quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells," *Biophys. J*, vol. 95, no. 12, pp. 6025–6043, 2008.

[65] B. Hanser, M. Gustafsson, D. Agard, and J. Sedat, "Phase-retrieved pupil functions in wide-field fluorescence microscopy," *J. Microsc.*, vol. 216, no. 1, pp. 32–48, 2004.

[66] M. J. Mlodzianoski, M. F. Juette, G. L. Beane, and J. Bewersdorf, "Experimental characterization of 3d localization techniques for particle-tracking and super-resolution microscopy," *Opt. Express*, vol. 17, no. 10, pp. 8264–8277, 2009.

[67] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, 2003.

[68] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[69] N. Lazaros, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: from software to hardware," *Int. J. Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[70] V. S. Kluth, G. W. Kunkel, and U. A. Rauhala, "Global least squares matching," in *Proc. Int'l Geoscience and Remote Sensing Symp*, vol. 2, pp. 1615–1618, IEEE, 1992.

[71] B. D. Lucas, T. Kanade, *et al.*, "An iterative image registration technique with an application to stereo vision," in *Proc. Int'l Joint Conf. Artificial Intelligence*, vol. 81, pp. 674–679, 1981.

[72] V. Venkateswar and R. Chellappa, "Hierarchical stereo and motion correspondence using feature groupings," *Int. J. Comput. Vision*, vol. 15, no. 3, pp. 245–269, 1995.

[73] C. Schmid and A. Zisserman, "The geometry and matching of curves in multiple views," in *Proc. European Conf. Computer Vision*, pp. 394–409, Springer, 1998.

[74] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *Int. J. Comput. Vision*, vol. 35, no. 3, pp. 269–293, 1999.

[75] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, 2004.

[76] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. Int'l Conf. Computer Vision*, vol. 2, pp. 508–515, IEEE, 2001.

[77] B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald, "Review of stereo vision algorithms and their suitability for resource-limited systems," *J. Real-Times Image Pr.*, vol. 11, no. 1, pp. 5–25, 2016.

[78] H. Liu, T.-H. Hong, M. Herman, T. Camus, and R. Chellappa, "Accuracy vs efficiency trade-offs in optical flow algorithms," *Comput. Vision Image Understanding*, vol. 72, no. 3, pp. 271–286, 1998.

[79] R. Henriques, M. Lelek, E. F. Fornasiero, F. Valtorta, C. Zimmer, and M. M. Mhlanga, "Quickpalm: 3d real-time photoactivation nanoscopy image processing in imagej," *Nat. Methods*, vol. 7, no. 5, pp. 339–340, 2010.

[80] A. J. Berglund, M. D. McMahon, J. J. McClelland, and J. A. Liddle, "Fast, bias-free algorithm for tracking single particles with variable size and shape," *Opt. Express*, vol. 16, no. 18, pp. 14064–14075, 2008.

[81] B. Yu, D. Chen, J. Qu, and H. Niu, "Fast fourier domain localization algorithm of a single molecule with nanometer precision," *Opt. Lett.*, vol. 36, no. 22, pp. 4317–4319, 2011.

[82] A. V. Abraham, S. Ram, J. Chao, E. Ward, and R. J. Ober, "Quantitative study of single molecule location estimation techniques," *Opt. Express*, vol. 17, no. 26, pp. 23352–23373, 2009.

[83] A. Small and S. Stahlheber, "Fluorophore localization algorithms for super-resolution microscopy," *Nat. Methods*, vol. 11, no. 3, pp. 267–279, 2014.

[84] H. Kirshner, F. Aguet, D. Sage, and M. Unser, "3-d psf fitting for fluorescence microscopy: implementation and localization application," *J. Microsc.*, vol. 249, no. 1, pp. 13–25, 2013.

[85] R. Melina, "Microscopic Worlds Gallery: Fascinating Fungi (SpongeBob Fungus)." http://www.livescience.com/17192-microscopic-worlds-gallery-fascinating-fungi.html, 2011. [Online; accessed 2016-09-26].

[86] B. Das, S. Vyas, J. Joseph, P. Senthilkumaran, and K. Singh, "Transmission type twisted nematic liquid crystal display for three gray-level phase-modulated holographic data storage systems," *Opt. Lasers Eng.*, vol. 47, no. 11, pp. 1150–1159, 2009.

[87] M. Yamauchi and T. Eiju, "Optimization of twisted nematic liquid crystal panels for spatial light phase modulation," *Opt. Commun.*, vol. 115, no. 1, pp. 19–25, 1995.

[88] G. Carles, G. Muyo, S. Bosch, and A. Harvey, "Use of a spatial light modulator as an adaptable phase mask for wavefront coding," *J. Mod. Opt.*, vol. 57, no. 10, pp. 893–900, 2010.

[89] E. Van Putten, I. Vellekoop, and A. Mosk, "Spatial amplitude and phase modulation using commercial twisted nematic lcds," *Appl. Opt.*, vol. 47, no. 12, pp. 2076–2081, 2008.

[90] V. Durán, J. Lancis, E. Tajahuerce, and Z. Jaroszewicz, "Equivalent retarder-rotator approach to on-state twisted nematic liquid crystal displays," *J. Appl. Phys.*, vol. 99, no. 11, p. 113101, 2006.

[91] M. S. Mahmud, I. Naydenova, and V. Toal, "Implementation of phase-only modulation utilizing a twisted nematic liquid crystal spatial light modulator," *J. Opt. A: Pure Appl. Opt.*, vol. 10, no. 8, p. 085007, 2008.

[92] V. Duran, J. Lancis, E. Tajahuerce, and Z. Jaroszewicz, "Cell parameter determination of a twisted-nematic liquid crystal display by single-wavelength polarimetry," *J. Appl. Phys.*, vol. 97, no. 4, p. 043101, 2005.

[93] S. F. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," *J. Opt. Soc. Am. A*, vol. 8, no. 10, pp. 1601–1613, 1991.

[94] L. S. Churchman, Z. Ökten, R. S. Rock, J. F. Dawson, and J. A. Spudich, "Single molecule high-resolution colocalization of cy3 and cy5 attached to macromolecules measures intramolecular distances through time," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 5, pp. 1419–1423, 2005.

[95] J. Min, S. J. Holden, L. Carlini, M. Unser, S. Manley, and J. C. Ye, "3d high-density localization microscopy using hybrid astigmatic/ biplane imaging and sparse image reconstruction," *Biomed. Opt. Express*, vol. 5, no. 11, pp. 3935–3948, 2014.

[96] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognit.*, vol. 19, no. 6, pp. 459 – 466, 1986.

[97] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vision Comput.*, vol. 21, no. 11, pp. 977 – 1000, 2003.

[98] A. Goshtasby, "Image registration by local approximation methods," *Image Vision Comput.*, vol. 6, no. 4, pp. 255 – 261, 1988.

[99] A. Pertsinidis, Y. Zhang, and S. Chu, "Subnanometre single-molecule localization, registration and distance measurements," *Nature*, vol. 466, pp. 647–651, Jul 2010.

[100] S. Pertuz, D. Puig, and M. A. Garcia, "Reliability measure for shape-from-focus," *Image Vision Comput.*, vol. 31, no. 10, pp. 725 – 734, 2013.

[101] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415 – 1432, 2013.

[102] P. Favaro, S. Soatto, M. Burger, and S. J. Osher, "Shape from defocus via diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 518–531, 2008.

[103] G. Carles, A. Carnicer, and S. Bosch, "Phase mask selection in wavefront coding systems: A design approach," *Opt. Lasers Eng.*, vol. 48, no. 7–8, pp. 779 – 785, 2010.

[104] T. Vettenburg, H. I. C. Dalgarno, J. Nylk, C. Coll-Llado, D. E. K. Ferrier, T. Cizmar, F. J. Gunn-Moore, and K. Dholakia, "Light-sheet microscopy using an airy beam," *Nat. Methods*, vol. 11, no. 5, pp. 541–544, 2014.

[105] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[106] S. T. Wereley and C. D. Meinhart, "Recent advances in micro-particle image velocimetry," *Annu. Rev. Fluid Mech.*, vol. 42, pp. 557–576, 2010.

[107] C. Meinhart, S. Wereley, and M. Gray, "Volume illumination for two-dimensional particle image velocimetry," *Meas. Sci. Technol.*, vol. 11, no. 6, pp. 809–814, 2000.

[108] M. Fernández-Suárez and A. Y. Ting, "Fluorescent probes for super-resolution imaging in living cells," *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 12, pp. 929–943, 2008.

[109] F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, *et al.*, "Video-rate nanoscopy using scmos camera-specific single-molecule localization algorithms," *Nat. Methods*, vol. 10, no. 7, pp. 653–658, 2013.

[110] Z.-L. Huang, H. Zhu, F. Long, H. Ma, L. Qin, Y. Liu, J. Ding, Z. Zhang, Q. Luo, and S. Zeng, "Localization-based super-resolution microscopy with an scmos camera," *Opt. Express*, vol. 19, no. 20, pp. 19156–19168, 2011.

[111] H. Deschout, K. Neyts, and K. Braeckmans, "The influence of movement on the localization precision of sub-resolution particles in fluorescence microscopy," *J. Biophotonics*, vol. 5, no. 1, pp. 97–109, 2012.

[112] S. Jia, J. C. Vaughan, and X. Zhuang, "Isotropic three-dimensional super-resolution imaging with a self-bending point spread function," *Nat. Photonics*, vol. 8, no. 4, pp. 302–306, 2014.

[113] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophys. J.*, vol. 86, no. 2, pp. 1185–1200, 2004.

[114] R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes," *Biophys. J.*, vol. 82, no. 5, pp. 2775–2783, 2002.

[115] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, 2008.

[116] M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, "Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples," *Nat. Methods*, vol. 5, no. 6, pp. 527–529, 2008.

[117] D. Aquino, A. Schönle, C. Geisler, C. V Middendorff, C. A. Wurm, Y. Okamura, T. Lang, S. W. Hell, and A. Egner, "Two-color nanoscopy of three-dimensional volumes by 4pi detection of stochastically switched fluorophores," *Nat. Methods*, vol. 8, no. 4, pp. 353–359, 2011.

[118] G. Shtengel, J. A. Galbraith, C. G. Galbraith, J. Lippincott-Schwartz, J. M. Gillette, S. Manley, R. Sougrat, C. M. Waterman, P. Kanchanawong, M. W. Davidson, *et al.*, "Interferometric fluorescent super-resolution microscopy resolves 3d cellular ultrastructure," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 9, pp. 3125–3130, 2009.

[119] S. Cho, J. Jang, C. Song, H. Lee, P. Ganesan, T.-Y. Yoon, M. W. Kim, M. C. Choi, H. Ihee, W. Do Heo, *et al.*, "Simple super-resolution live-cell imaging based on diffusion-assisted forster resonance energy transfer," *Sci. Rep.*, vol. 3, 2013.

[120] K. Xu, H. P. Babcock, and X. Zhuang, "Dual-objective storm reveals three-dimensional filament organization in the actin cytoskeleton," *Nat. Methods*, vol. 9, no. 2, pp. 185–188, 2012.

[121] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. Moerner, "Optimal point spread function design for 3d imaging," *Phys. Rev. Lett.*, vol. 113, no. 13, p. 133902, 2014.

[122] G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, and X. Zhuang, "Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging," *Nat. Methods*, vol. 8, no. 12, pp. 1027–1036, 2011.

[123] S. Daostorm, "Daostorm: an algorithm for high-density super-resolution microscopy," *Nat. Methods*, vol. 8, pp. 279–280, 2011.

[124] H. Deschout, F. C. Zanacchi, M. Mlodzianoski, A. Diaspro, J. Bewersdorf, S. T. Hess, and K. Braeckmans, "Precisely and accurately localizing single emitters in fluorescence microscopy," *Nat. Methods*, vol. 11, no. 3, pp. 253–266, 2014.

[125] J. Enderlein, E. Toprak, and P. R. Selvin, "Polarization effect on position accuracy of fluorophore localization," *Opt. Express*, vol. 14, no. 18, pp. 8111–8120, 2006.