ADVANCEMENTS IN LIGHT FIELD-BASED LAPAROSCOPES

by

Elliott Kwan

_____

A Dissertation Submitted to the Faculty of the

JAMES C. WYANT COLLEGE OF OPTICAL SCIENCES

In Partial Fulfilment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by **Elliott Yi-Zhi Kwan**, titled *Advancements in Light Field-based Laparoscopes* and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

/ Hong Hua /
_____   Date: Nov 14, 2022
Professor Hong Hua

_____   Date: Nov 14, 2022
Professor Rongguang Liang

_____   Date: Nov 14, 2022
Professor Travis W. Sawyer

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____   Date: Nov 16, 2022
Professor Hong Hua
Dissertation Committee Chair
Wyant College of Optical Sciences

ACKNOWLEDGEMENTS

I wish to express my most sincere gratitude and appreciation to my PhD advisor, Prof. Hong Hua. Learning about displays and exploring imaging optics under her expert guidance has been a special experience. Her ideas, advice, feedback, and patience were all vital to the accomplishments of my graduate school journey. I would like to thank her for all the support and training in overcoming my research challenges. From this experience, I have grown to be more independent and have developed the technical skillset to further the development of optical technologies.

I would like to acknowledge all my committee members. Prof. Judith Su and Prof. Leilei Peng made my comprehensive exam a positive experience. Prof. Rongguang Liang and Prof. Travis Sawyer were very accommodating to my needs and provided useful discussions and suggestions for this dissertation.

I would also like to acknowledge my lab colleagues. Yi Qin provided me with the lens design of my first prototype. Hekun Huang answered many of my questions when I was learning about light field concepts. Xuan Wang helped me figure out uncertainties in the lens design software. Jeremy Katz was tremendously helpful since we both worked on laparoscope projects. Our discussion on tolerancing was essential to the operation of my second prototype. I also appreciate all my other lab colleagues for any general assistance and for simply bonding in the lab.

DEDICATION

I dedicate my dissertation work to my wife Kristy Ujiiye for all her emotional support. Even from a long distance, her love, patience, and enthusiasm kept me going every day.

I also dedicate this to my family. I could not have asked for better parents. They have supported me in every possible way, reminding me to stay healthy, serving as my role models, and providing encouragement and life advice every step of the way. I also could not have asked for better sisters. They are always around when I need them most.

Last but not least, I dedicate this to my dog Neveah for making sure I get exercise during her daily walks, for sticking by my side 24/7, and keeping me safe with her barking at any potential strangers.

1   TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In the 20th century, rigid laparoscopes revolutionized surgery such that minimally invasive procedures are now the norm. However, these systems only provide surgeons with a two-dimensional (2D) view of the operative field and are subject to two major optical limitations: (1) the absence of binocular vision results in restricted depth perception and (2) the field of view (FOV) is restricted to the local operating region to ensure high image spatial resolution. Performing surgery through a monitor without depth perception is challenging and requires extensive training. Meanwhile, surgical accidents that occur outside of the limited FOV and have gone unnoticed may cause unnecessary trauma to the patient. In this dissertation, two novel optical designs were developed to address the two limitations and further advance this technology. The conceptualization, lens design, prototyping, calibration, and processed results are discussed for both designs.

The first design is a programmable aperture light field laparoscope. It was used to investigate and explore the requirements of three-dimensional depth information extraction in a monocular form factor. Compared to state-of-the-art dual objective stereoscopic laparoscopes, this form factor preserves more design volume for transmitting more of the object scene's light field. A programmable aperture is used to preserve the laparoscope's conventional high resolution 2D imaging and upon demand, capture the light field. The light field information enables this system to view the object scene from different viewing angles, digitally refocus, and generate depth maps for surgical guidance in post processing.

A second-generation design called a tri-aperture monocular laparoscope was then developed to address the depth perception and FOV limitations simultaneously. This system uses two displaced apertures and a custom prism to capture the stereoscopic views

simultaneously, which can then be processed to generate absolute depth maps. Meanwhile, a wide FOV for situational awareness is captured via a central third aperture. It provides 2D vision over an area 2x as large as that of the stereoscopic views. Such a system may pave the way towards restoring the binocular and large, foveated FOV qualities of human vision within the minimally invasive surgical setting.

1   INTRODUCTION

Endoscopies, or minimally invasive surgery (MIS), are essential to modern day medicine. For instance, during the COVID-19 pandemic, the number of endoscopy procedures reduced to 5-20%, resulting in a 58% decrease of the weekly number of cancers detected [1]. The ability to investigate, diagnose, and perform surgery minimally invasively inside the cavities of the body with an endoscope is now the norm. These procedures are more specifically named based on the body part to be observed. One common example is a laparoscopy, which is surgery performed on organs within the abdomen or pelvis and visualized using a laparoscope.

The laparoscope most commonly has the shape of a rigid long tube approximately 10-12 mm in diameter. Only a small incision and a trocar are required to insert this instrument into the body. Inside the tube, there is a ring of fiber optics to transmit light from an external source onto the organs of interest and a set of lens groups to image the operative field onto a camera sensor outside of the patient's body. The laparoscopic optics and sensor technology has continually improved throughout the 20th century and can now potentially achieve 8K ultra-high definition imaging [2]. This allows surgeons to see capillary features on the surface of the large intestine.

Although these conventional systems provide a high quality two-dimensional (2D) view of the operative field, they are subject to two major optical limitations: (1) the absence of binocular vision results in restricted depth perception and (2) the field of view (FOV) is restricted to the local operating region to maintain high image spatial resolution. Performing surgery through a monitor displaying only monocular depth cues causes challenges of eye-hand coordination and requires extensive training. It is equivalent to

navigating three-dimensional (3D) space with only one eye. Meanwhile, surgical accidents that occur outside of the limited FOV and have gone unnoticed may cause unnecessary trauma to the patient. The laparoscope needs to be inconveniently moved to view and keep track of these areas.

To overcome the first limitation, 3D capable laparoscopes have been developed. In one study, 451 3D laparoscopic surgeries were compared to their 2D counterpart, and the results indicated that surgeries were completed more quickly in 3D [3]. It also noted that tactile feedback was retained given the provided depth perception, thus improving precision, accuracy, and shortening the learning curve. Similar studies have shown that providing 3D information is most beneficial to novice surgeons and reduces strain [4–6]. 2D MIS is still the current standard, but 3D MIS may become prevalent as it continues to develop.

To overcome the second limitation, wide field of view (WFOV) laparoscopes have been developed. The goal of these systems is to provide as much visualization of blind zones as possible such that surgical accidents can be avoided, and all abnormalities are detected. Providing a panoramic view as seen in open surgery would be ideal and has been proposed [7]. However, for any conventional imaging system, the FOV and spatial resolution are inversely proportional. Sacrificing the spatial resolution to increase the FOV would be unacceptable to surgeons, especially now that they've become accustomed to the high image quality of commercial 2D systems. This tradeoff has been overcome by a multiresolution foveated laparoscope (MRFL) design [8–12]. Ongoing ex vivo and in vivo developments for this system indicate that this is a promising approach. Although this system incorporates the WFOV, it still lacks the 3D capability.

Implementing either 3D imaging or 2D WFOV imaging capability into laparoscopes has been explored extensively. Alternatively, a laparoscope that can capture both capabilities has only been minimally explored and could be the future direction of MIS. Little work has been done in this area because it requires a novel optical architecture that can acquire all this information. For instance, at least three camera systems would be required to capture two stereoscopic views and one WFOV. Trying to fit three optical systems into the limited volume constrained by the mechanical housing of the laparoscope is a key challenge. The research in this dissertation focuses on ultimately developing a laparoscope that can provide both capabilities. Such a system could restore the stereoscopic and WFOV qualities of human vision within the environment of MIS.

1.1 Dissertation Contribution

In this dissertation, two novel optical systems were developed to address the limitations of conventional laparoscopy and to further advance this technology. The conceptualization, first-order design, lens design, prototyping, calibration, and processed results are discussed for both systems.

The first system is a programmable aperture light field laparoscope. It was used to investigate and explore the requirements of 3D depth information extraction in a monocular form factor. Compared to state-of-the-art dual objective stereoscopic laparoscopes, this form factor preserves more design volume for transmitting more of the object scene's light field. A programmable aperture is used to preserve the laparoscope's conventional high resolution 2D imaging and upon demand, capture the light field. The light field information enables this system to view the object scene from different viewing angles, digitally refocus, and generate depth maps for surgical guidance in post processing.

16

Based on what was learned from the first design, a second system called a tri-aperture monocular laparoscope was then developed to address the depth perception and FOV limitations concurrently. This system uses two displaced apertures and a custom prism to capture the stereoscopic views simultaneously, which can then be processed to generate anaglyphs for 3D viewing and absolute depth maps. Overlapping crosstalk between the stereoscopic views is diminished by incorporating a strategically placed vignetting aperture. Meanwhile, a WFOV for situational awareness is captured via a central third aperture. It provides 2D vision over an area 2x as large as that of the stereoscopic views for peripheral awareness.

To enable the functionality of these two systems, a generalized ray error calibration was developed for as-built light field cameras. The calibration accounts for vignetting, transverse ray errors, as well as pupil aberration, and can be applied to light field camera modeling of arbitrary pupil sampling systems. The calibration method is based on measurements of fiducial markers on a checkerboard for modeling the imaging properties of light field cameras. The utility of this calibration is demonstrated by applying it to the two prototypes of this dissertation to recover relative 3D light field computational imaging capability and calibrated stereoscopic views for accurate 3D viewing and depth mapping.

1.2 Dissertation Contents

Following this chapter of INTRODUCTION, Chapter 2 BACKGROUND AND RELATED WORK presents the preparatory information necessary for understanding the dissertation contributions. This includes general theory and a discussion of existing technologies that have addressed the 3D and WFOV limitations of laparoscopy.

Chapter 3 DEPTH PERCEPTION FROM A PROGRAMMABLE APERTURE

LIGHT FIELD LAPAROSCOPE presents a uniquely modified laparoscope system for studying depth extraction in the monocular formfactor. It includes an analysis of achievable depth mapping resolution, a prototype for light field acquisition, and processing of the light field for depth assessment.

Chapter 4 A PRISM-BASED TRI-APERTURE LAPAROSCOPE FOR MULTI-VIEW ACQUISITION presents an original laparoscope system for acquiring stereoscopic and WFOV content in the monocular formfactor. The complete optical design process is demonstrated, from conceptualization to 3D and WFOV viewing. This is the most significant contribution of this dissertation.

Chapter **Error! Reference source not found. Error! Reference source not found.** presents the generalized theory and procedure to model and calibrate an arbitrary pupil sampling system as a light field camera. This calibration is demonstrated in both laparoscopes of Chapters 3 and 4 to enable their functionality.

Chapter 6 CONCLUSION AND FUTURE WORK summarizes the contributions and lessons learned and discusses future directions.

Appendix A includes a published peer-reviewed paper titled "High resolution, programmable aperture light field laparoscope for quantitative depth mapping."

Appendix B includes a published peer-reviewed paper titled "Prism-based tri-aperture laparoscopic objective for multi-view acquisition."

Appendix C includes a published peer-reviewed paper titled "Calibration of transverse ray and pupil aberrations for light field cameras."

2   BACKGROUND AND RELATED WORK

Implementing either 3D imaging or 2D WFOV imaging capability into laparoscopes has been explored extensively. To develop a system that incorporates both capabilities, the existing systems that have one of these capabilities should be understood first. This chapter discusses the different classes of technologies enabling these existing systems and then assesses which classes are most suitable to be combined into a single system to achieve both capabilities.

2.1   WFOV Capable Endoscopes

The fundamental goal of WFOV capable endoscopes is to properly image the area of interest onto a camera sensor. This has been accomplished in different ways, which will be discussed in order of complexity.
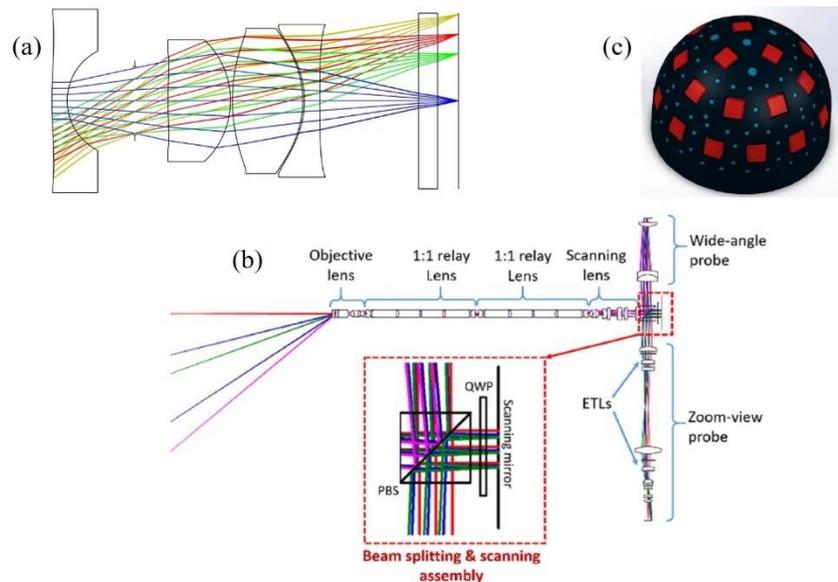
Fig. 2.1 Methods of achieving WFOV capable endoscopes: (a) traditional lens design [13]**, (b) beam splitting system** [11]**, (c) multicamera system** [14]

Fig. 2.1(a) shows an example of increasing the endoscope objective's FOV and numerical aperture using traditional lens design methods. This particular design has a 140° WFOV at a working distance of 15 mm and a high optical resolution associated with a

lower f-number of 2.2 [13]. The redesign process consisted of conventional methods, such as reoptimizing the lens radii, thickness, glass type, surface type, spacing, and number of lenses. Ultimately, the focal length was reduced sufficiently such that the image of the newly desired FOV fits onto an appropriately sized sensor.

Reducing the focal length allows the sensor to capture more of the object field, but if the sensor format is unchanged, the trade-off is a reduction in spatial resolution because the sensor's pixels now sample a larger distance in object space. This trade-off is also seen in all conventional optical zoom capable camera systems, including the commercial endoscopes made by Karl Storz and Stryker, as well as laparoscopes using liquid lenses [15] to change system focal length. For a system that is sensor limited, the pixel size could be reduced to recover the resolution. However, smaller pixels typically result in a lower signal-to-noise ratio and dynamic range, so this approach is dependent on more advanced sensor technology. Also, medical grade endoscopes are already using top of the line camera sensors for maximum image quality, so finding higher quality sensors would be challenging.

A few methods based on foveated imaging have been proposed to overcome the resolution and FOV trade-off. One of them extended the traditional lens design approach by controlling the lens distortion such that the center of the WFOV has higher magnification and resolution while the outer region of the WFOV is minified with lower resolution [16]. However, the lens design for this approach appeared to require much tighter tolerances for fabrication. Another foveated approach called the multi-resolution foveated laparoscope (MRFL) utilizes a beam splitting system to capture the conventional FOV and the WFOV onto two different sensors, as shown in Fig. 2.1(b) [8–12]. The

objective lens creates an intermediate image of the WFOV, the relays relay this intermediate image to outside of the patient's body, and the scanning lens collimates the image. Then a beam splitter splits the light so that a zoom-view probe can image the conventional FOV to fill a dedicated sensor and maintain the high resolution of state-of-the-art laparoscopes while a wide-view probe has the shorter focal length to image the WFOV onto another dedicated sensor. With this architecture, the resolution versus FOV trade-off is eliminated. The MRFL has been demonstrated in animal trials and is a promising technology for 2D WFOV MIS.

The last classification of WFOV capable endoscopes is multicamera systems, in which two or more cameras are incorporated at the tip of the endoscope to capture additional FOV. For instance, one design places two cameras adjacent to one another, one for side viewing and one for forward viewing [17]. In another example called the Third Eye Panoramic device, two side viewing cameras were added by incorporating them into a cap that clips on the tip of the endoscope [18]. However, in these systems, the extra camera or cap results in an increase of the endoscope's total volume, which is typically undesired for MIS. There is also another type of multicamera system for endoscopy inspired by the compound eye of insects [14], as shown in Fig. 2.1(c). In this design, 24 miniature pinhole cameras were integrated on a 5 mm radius hemisphere to point them in different directions for WFOV imaging. Although the size of this design is appropriate for endoscopy, the image quality is ultimately limited by the diffraction limit of the pinhole cameras and the performance of the miniature sensors.

2.2   3D Capable Endoscopes

Endoscopes that recover 3D information come in a variety of form factors. These

endoscopes typically require an additional optical component to capture the necessary data for determining the depth of objects in the FOV. The type of component used determines if the 3D information can be displayed as stereoscopic images or simply a depth map. Providing both options for the surgeon would be ideal such that they can visualize the scene in 3D and check the depth map for tasks requiring high precision. In addition, the type of optical component also determines how easily it can be integrated into the endoscope housing and results in different form factors. Based on the 3D capability and engineering constraints, some of these form factors are more prevalent. The various form factors can be classified into three different categories [19].
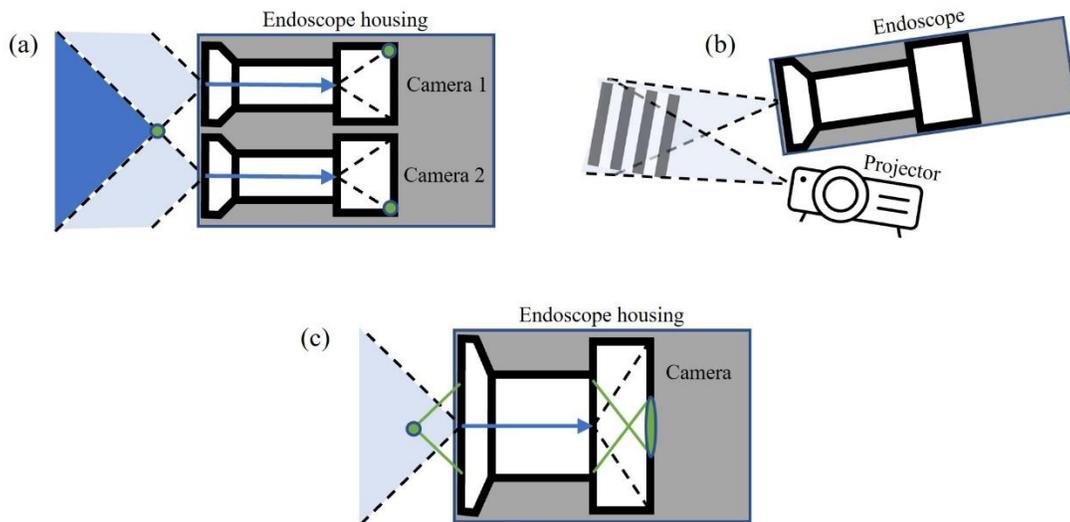
Fig. 2.2 The main types of 3D capable endoscopes: (a) dual camera stereo endoscope, (b) structured light projection endoscope, (c) monocular endoscope with uniaxial 3D depth cues.

Fig. 2.2(a) shows the conceptual form factor of a stereo endoscope, in which two cameras are built into the endoscope housing. This is the most prevalent category of 3D endoscopes because the stereo camera concept has been well established in 3D photography. These types of endoscopes have been sufficiently developed such that they are now commercially available. The depth of object points is determined by triangulation

based off the images from each camera. However, there are several challenges that can be visualized in the figure. If the two cameras are parallel to each other, only a portion of their FOVs that overlap is useful for 3D imaging. This results in wasted sensor area. The cameras could be tilted towards each other to maximize the overlapping FOV, but that would be harder to fabricate. Another major challenge is cramming two independent optical systems inside the limited diameter housing of the endoscope. This results in smaller lenses, which can be more difficult to work with, may require tighter mechanical tolerances, and reduces the numerical aperture and diffraction limited image performance of each optical system. Relative to that of the conventional 2D endoscopes, the image quality of each stereoscopic image could be further reduced if two smaller less capable sensors are used for the two camera systems.

Fig. 2.2(b) shows a regular endoscope used in combination with a projection system to reconstruct 3D surface profiles. By analysing the warped projection pattern due to the non-planar object surface, the surface profile can be reconstructed. To use this concept in MIS, the projection system would ideally need to be built into the endoscope housing. However, the projection system still requires separation from the imaging system because the reconstruction algorithm relies on a triangulation between them. Also, this category of 3D endoscopes can only directly provide a depth map instead of the stereoscopic images. For these reasons, this approach is least popular among the three main types of 3D capable endoscopes. Although, structured light projection is most useful for 3D mapping in scenarios where the surgical area has little contrast or surface texture.

Fig. 2.2(c) shows a traditional monocular endoscope design that achieves 3D capability by assessing depth from defocus, a uniaxial 3D depth cue. The larger the amount of

defocus, the closer or farther away the object is located. Compared to the dual camera stereo endoscope, this specific method can determine 3D information for the entire FOV. The major benefit of this type of system is that it maintains the conventional endoscope optics such that the redesign for 3D capability is minimized. However, it requires a physical mechanism to scan the focal planes of the camera and determine the in-focus object depths. Alternatively, an additional piece of hardware can be inserted into this camera system to record the light field, or the set of individual light rays, emanating from the object scene. The focal planes can then be scanned digitally in post processing using the light field information. Moreover, the captured light field contains the stereoscopic views that the surgeons can use to visualize 3D surgery. Other uniaxial depth cues exist such as depth from controlled optical aberrations and time of flight. However, these methods are unable to directly provide stereoscopic images, so they would be less desirable for the surgeon.

2.3 Combining WFOV and 3D Capability

Combining WFOV and 3D capability into an endoscope has only been demonstrated in one literature example [20]. This example is an extension of the dual camera stereo endoscope in Fig. 2.2(a). The 3D stereoscopic field of view (SFOV) was unchanged while the 2D WFOV was obtained by utilizing and stitching together the non-overlapping FOVs of the stereo cameras together with the SFOV. This strategy requires splitting the image sensor area between the SFOV and WFOV and results in a trade-off between them. This is not ideal since the surgeon would expect to visualize surgery with an adequate SFOV that has high pixel resolution, but a portion of the pixels are now reserved for the WFOV. Increasing the WFOV for better peripheral awareness is limited with this method since it would reduce the size or resolution of the SFOV. Regardless of this issue, the ability to

24

combine the two capabilities was shown, and the groundwork was laid for such a system.

To improve upon this combined capable system, the methods and form factors from the previous two sections were assessed to identify the type of architecture that would be most supportive of these two capabilities. For WFOV capable endoscopes, the system can either be monocular and uniaxial or composed of multiple cameras facing different directions. For 3D capable endoscopes, the system can either be composed of two smaller separate optical systems facing the same direction or a larger monocular system utilizing uniaxial 3D depth cues. Based on these two observations, it is apparent that a monocular architecture can support both capabilities.

The monocular architecture offers many benefits that justify its selection. Since this architecture is equivalent to the one used for conventional 2D endoscopes, implementing the additional capabilities in this kind of system would only require modification of traditional endoscopic lens design and fabrication techniques. By maintaining the conventional design, the image quality of state-of-the-art 2D endoscopes that surgeons are already using can be ideally preserved. This architecture also preserves the central optical design volume that would typically be wasted to house two independent stereoscopic optical systems. The endoscope housing already limits the design volume, so preserving as much of it as possible is desired for implementing additional imaging capabilities. Meanwhile, this type of endoscope would maximize the SFOV and avoid the issue with partially overlapping FOVs from two cameras. Finally, the monocular architecture also enables the option of incorporating the beam splitting system to provide custom image sensors for maximizing the performance of each capability. For these reasons, the research presented in this dissertation focuses on developing a unique monocular architecture and

calibration for 3D and WFOV endoscopy.

3 DEPTH PERCEPTION FROM A PROGRAMMABLE APERTURE LIGHT FIELD LAPAROSCOPE

To develop a unique system for dual capability, first the monocular architecture using uniaxial 3D depth cues needs to be fully understood. This chapter proposes a high resolution, programmable aperture light field laparoscope for quantitative depth mapping. The idea is to insert a programmable aperture into a modified traditional monocular laparoscope to sample and record the light field captured by the objective lens. The light field information is then used to show the stereoscopic views that can be captured from a monocular system and to assess depth from defocus in post processing rather than with a physical scanning mechanism. From the design of this system, the constraints that must be met to achieve sufficient depth resolution for MIS are revealed. The work in this chapter is published in Appendix A, and any additional details mentioned here are supplementary.

3.1 Optical Limitations of Existing Light Field Endoscopes
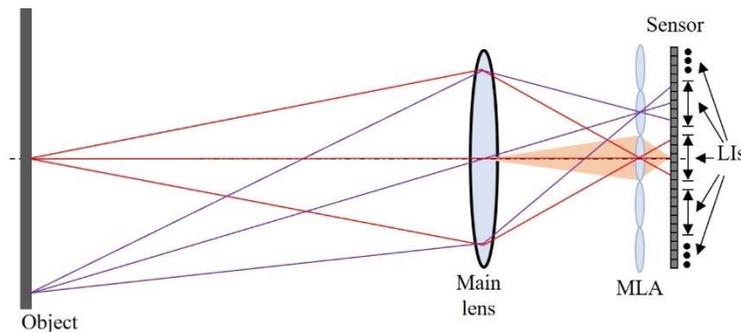
Fig. 3.1 Working principal of light field camera 1.0

There are a few monocular light field (LF) endoscopes that already exist in literature, including the LF otoscope [21], laryngoscope [22], and endoscope [23]. They were all designed by simply inserting a microlens array (MLA) at an image plane of the endoscopic imaging optics, as shown in Fig. 3.1. This design is known as LF camera 1.0 and is

27

thoroughly discussed, along with the methods used to process the captured light field, in [24].

The LF camera 1.0 records a portion of the LF by sampling the individual light rays captured by the main lens. In this design, the light rays are recorded by the sensor and the direction of each one is defined by a microlens and a sensor pixel. The microlenses are spatially sampling the rays from the image plane while the sensor pixels are angularly sampling the ray bundles diverging from the image plane. Once the light rays are recorded, they can be traced forward or backward to refocus at any image plane in post processing. For instance, they can be traced backward from the sensor to the MLA plane to reconstruct that image plane.

The spatial resolution of the reconstructed image at the MLA plane depends on the sampling of the light rays. To increase signal to noise ratio, the microlenses sample a bundle of rays and focus them together onto a sensor pixel. As a result, the lateral resolution of the reconstructed image plane is reduced to the MLA pitch. Meanwhile, the angular resolution is determined by the effective sampling of the diverging ray bundles by the sensor's pixel pitch. For a larger MLA pitch and longer focal length microlenses, the sensor can be repositioned accordingly to increase the effective sampling and achieve a higher angular resolution. The MLA design results in a tradeoff between the spatial and angular resolution of the sampled LF. The reduced lateral resolution of reconstructed image planes is a significant limitation for endoscopy since surgeons are already accustomed to the high resolution of current 2D systems.

3.2   Proposed Optical Approach
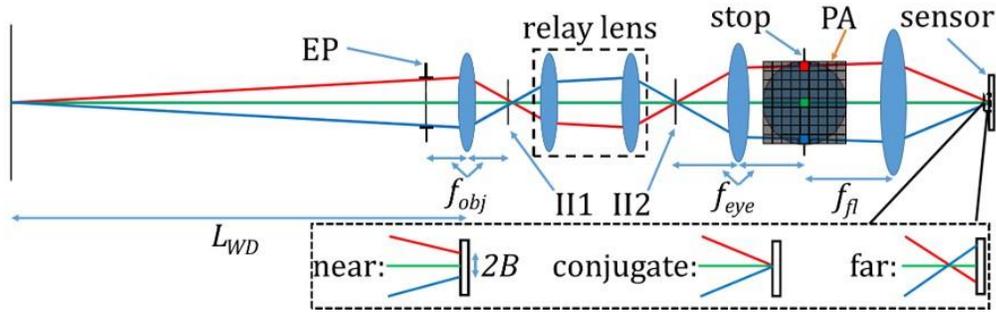
Fig. 3.2 Schematic layout of the PALFL [25]

To sample the LF without reducing the lateral resolution due to the MLA, this section proposes to apply a programmable aperture (PA) approach [26,27] to the endoscope instead. The programmable aperture light field laparoscope (PALFL) schematic layout is shown in Fig. 3.2. The system is composed of an objective lens, a 1:1 relay lens group, an eyepiece, a programmable aperture, a focusing lens, and a sensor. The objective lens with a focal length of f_{obj} images the desired FOV to form intermediate image 1 (II1). Next, for a rigid laparoscope, a 1:1 relay lens is necessary to extend the insertion length of the imaging probe within the limit of the housing tube diameter and relay the image to outside of the patient's body at intermediate image 2 (II2). To fit the objective lens and relay lens within the standard 10 mm-diameter housing of laparoscopes, the objective lens is designed to be image-space telecentric with its entrance pupil (EP) placed at its front focal point while the relay lens group is designed to be double telecentric. The eyepiece with a focal length of f_{eye} projects the image toward optics infinity for direct viewing or further imaging. In the meantime, the eyepiece forms a conjugate image of the objective EP, labeled as "stop", at which the programmable aperture is placed. Opening a given region of the PA component allows the focusing lens, with a focal length of f_{fl}, to image different bundles of rays from the object onto the sensor.

29

By selectively opening different sub-apertures (e.g. three instantaneous sub-apertures are highlighted by the Red, Green and Blue pixels in Fig. 3.2) sequentially, the sensor captures different light ray angles incident upon the EP from the same object point. As illustrated by the zoomed view at the sensor, depending on the depth of the object of interest, the rays through the different sub-apertures may be imaged at the same pixel when the object depth is optically conjugate to the senor or at different pixels when it is either nearer or further than the conjugate depth. Such disparity information recorded by the sub-aperture images is to be used for reconstructing the depth map of the object field or refocusing the image at different depth. In other words, the disparity information is used to determine depth from defocus. Also note that the Red and Green pixels of the PA are effectively sampling light rays that would be captured by a stereo camera, or a pair of individual cameras, located at the EP and with a baseline equal to the diameter of the EP.

The significant advantage of this design over existing LF endoscopes using MLA is that spatial and angular resolutions of the captured LF images are only subject to the limits of the sensor resolution and the pitch of the PA, respectively. Another worth-noting feature of a PALFL system is its hybrid capability. The system's instantaneous aperture can be switched between sub-aperture LF capture state and a normal capture state where a centered, regular-sized aperture is operated to capture a conventional 2D full-resolution, full FOV image that is the same as a conventional laparoscope. This capability provides a surgeon with the option, on demand, to receive guidance through the visualization of depth information.

Another interesting aspect of the PA approach is that the size and pattern of the sub-aperture can be customized based on what is needed. To match the throughput of existing

30

LF endoscopes, sub-apertures can span multiple adjacent pixels in the PA while sensor pixels can be binned. In the case of insufficient illumination, the span can be further extended at the cost of depth-of-field or depth mapping range, and high angular resolution can still be maintained by allowing sequential sub-aperture regions to overlap. The drawback of a sequential capture is the cost of speed, but the ever-increasing frame rates of imaging sensors can well overcome this limitation. Also, multiplexed light field acquisition [26,27], which uses patterns spanning multiple regions of the PA per frame, can be implemented to increase signal-to-noise ratio and allow for faster frame rates.

3.3   Depth Mapping Resolution

A key aspect to the design of a PALFL system is to achieve adequate depth mapping resolution. This mainly depends on the maximal angular separation of the rays through the centers of the sub-apertures, which establishes the maximal baseline equivalent to a stereo system, and the minimally detectable ray separation of the imaging system. For the convenience of quantifying the depth resolution of different systems, the numerical aperture at the nominal working distance, NA_{WD}, in the object space is used to characterize the maximal angular separation of the sub-apertures, and the equivalent sensor spatial resolution in the object space, B_{obj}, of the system is used to quantify the minimally detectable ray separation. It is assumed that distinguishing the three separated rays in Fig. 3.2 and confidently detecting a depth offset from the sensor conjugate depth, L_{WD}, minimally requires a 2-pixel separation ($2B$ at the sensor or $2B_{obj}$ at L_{WD}) between the Red and Blue rays on the sensor. A higher depth resolution can be achieved by digitally interpolating pixel data and refining the location of rays that land in between two pixels, but this possibility is not demonstrated here. Using similar triangles with bases located at

31

L_{WD} and the EP and a Taylor series expansion for simplification, the depth resolution of a PALFL design is derived:

$$d_\pm \approx \frac{B_{obj}}{NA_{WD}}\left[1\pm\frac{2B_{obj}}{D_{EP}}\right], \tag{3.1}$$

where d_+ and d_- represent the absolute distances from the sensor conjugate depth, L_{WD}, to the closest resolvable depths away from and towards the EP, respectively, and D_{EP} is the EP diameter. Given the pixel resolution, B, of the sensor and first-order optics specifications, without considering the effects of diffraction and aberration, B_{obj} and NA_{WD} are defined as:

$$B_{obj} = \frac{\left(L_{WD} - f_{obj}\right)f_{eye}}{f_{obj}f_{fl}}B, \quad NA_{WD} \approx \frac{D_{EP}/2}{L_{WD} - f_{obj}}. \tag{3.2}$$
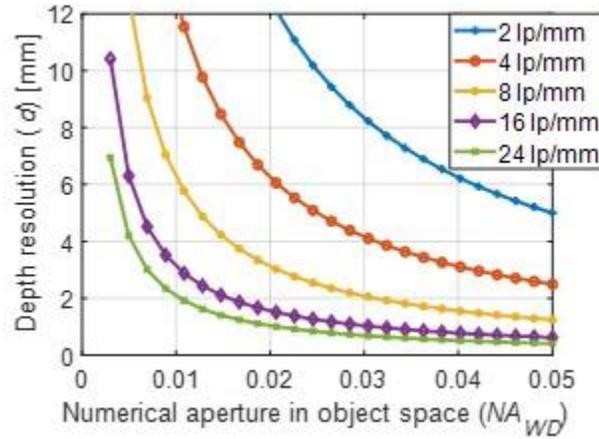
Fig. 3.3 Plot of achievable depth resolution in the laparoscopic environment for different NA_{WD} and $1/(2B_{obj})$ [25]

Fig. 3.3 plots the average depth resolution, d, of d_+ and d_- in relation to NA_{WD} for systems of different spatial resolutions in the object space. At a nominal working distance of 50 mm, the NA_{WD} of a standard monocular laparoscope is ~0.003 while the 5 mm baseline of a state-of-the-art stereo laparoscope (with a 12 mm diameter rod) by Intuitive Surgical produces an equivalent NA_{WD} of ~0.05. The object-space spatial resolution here is

32

quantified by the minimally discernable pair of line features per unit distance (lps/mm), equivalent to $1/(2B_{obj})$. The object-space resolution of a commercial laparoscope is 2-6 lps/mm and the diffraction limited resolution of the multi-resolution foveated laparoscope reported in [9] is ~12 lps/mm.

Fig. 3.3 suggests that implementing a LF laparoscope using standard laparoscope optics, with a spatial resolution of 4 lps/mm and NA_{WD} less than 0.01, can yield a depth resolution of worse than 12 mm. The combination of NA_{WD} of 0.015 and resolution of 6 lps/mm provides a depth resolution of ~5.5 mm, which can be useful for surgeons to determine the proximity of their surgical tools, but inadequate for accurately visualizing anatomical structures. Achieving sub-mm depth resolution with light field method requires substantial improvements in both the object-space resolution and numerical aperture of standard 2D laparoscopes. On the other hand, achieving this resolution in a LF laparoscope with dimensions like a stereo laparoscope seems possible.

3.4   Prototype and Experimental Setup
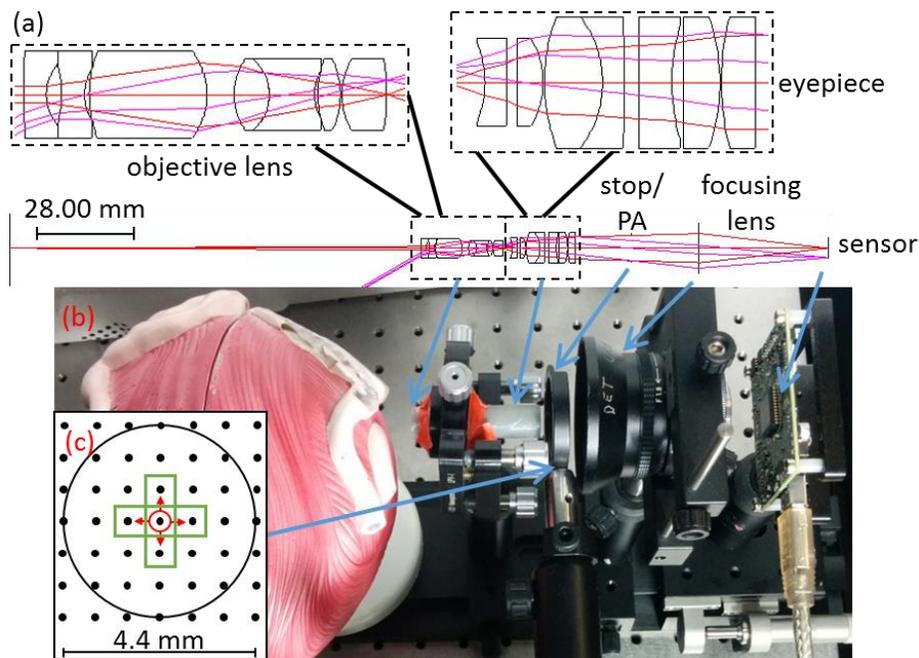
Fig. 3.4 (a) Optical layout of prototype, (b) construction of benchtop system, and (c) manual PA sampling scheme [25]**.**

Fig. 3.4(a) shows the optical layout of a prototyped bench-top PALFL system for proof of concept. An f/2.5 objective lens with a focal length of 1.8 mm from an existing laparoscope developed in [9] was repurposed for this prototype. The diameter of this objective lens group is 5.7 mm, which is small enough to allow space for fiber illumination and lens housing to build a standard 10-mm diameter rigid laparoscope as demonstrated in [9]. The optical system inside the rigid laparoscopic tube in [9] also consists of several groups of relay optics to relay the image of the objective to the distal end of the tube for further imaging. As the objective and the relay were optimized and custom-made independently, they can be used separately without the other and different number of the relay optics can be added or removed without affecting the optical performance. When building the PALFL bench prototype, we removed the relay optics for simplicity and only used the objective along with a newly added eyepiece, a PA, and a focusing lens as the relay optics does not add or change the imaging function of the system. The objective lens was originally optimized for an L_{WD} of 120 mm, a D_{EP} of 0.8 mm, and lens diameters < 6 mm, resulting in an effective NA_{WD} of ~0.003. However, for this PALFL prototype the objective lens was used at an L_{WD} of 20 mm. Although this distance is short for laparoscopy, it yields an NA_{WD} as large as ~0.022 if the full EP is sampled and produces an NA_{WD} that is more comparable to that of stereo laparoscopes. The relay lenses were omitted to simplify the lens design and optical alignment of this prototype. A 10 mm focal length eyepiece built with stock lenses was optimized to meet sufficient performance over a 60° full FOV and expanded the 0.8 mm EP of the objective lens to a 4.4 mm stop where a PA could be inserted. Note that the eyepiece diameter can be much larger than that of the objective and relay because

34

it is outside of the patient's body. These modules were aligned to a commercial focusing lens with a focal length of 25 mm and 1/3" color CCD sensor (1.3 MP Dragonfly2 from Point Grey). The pixel resolution of the sensor is 1280x960, and the color pixel size is 3.75x3.75 μm^2. Using Eq. (3.2), we can estimate that the theoretical spatial resolvability of the system in the object space is ~33 lps/mm. Using Eq. (3.1), the depth resolution of the prototype can potentially reach ~0.69 mm if the sub-aperture images are sampled at the full aperture and the optics perform at its full resolution.

Fig. 3.4(b) shows the prototype. The objective lens and eyepiece were assembled in a 3D printed opto-mechanical housing, as shown in the grey cylinder. Instead of using a digital PA, a physical iris mounted on a two-axis linear stage was employed. Fig. 3.4(c) illustrates the angular sampling scheme bounded by the stop. The grid of black dots represents the locations that would be sampled sequentially by the pitch of the sub-apertures and determines the angular resolution. The iris, indicated by the red circle with arrows, moves to each sampling location. An illuminated bladder model object is placed near an L_{WD} of 20 mm, as shown in Fig. 3(b). On the image side, the sensor was adjusted to the new conjugate image position.

Since the preexisting objective lens was not optimized for this short L_{WD}, aberrations and vignetting were introduced. To minimize degradation of data due to this issue, an effective LF calibration based on the aberration correction theory presented in [24] was developed and applied post-data capture. Similarly, this LF calibration can minimize the impact of aberrations from relay lenses. Since the focus here is the PALFL concept, this calibration is briefly summarized hereby and will be thoroughly discussed in Chapter 5. The calibration process began with a step of calibrating the amount of vignetting across the

field of view by capturing the LF data of a flat Lambertian source extending across the full

FOV. By comparing the peripheral sub-aperture images to the center one, the vignetting

was quantified and minimized via multiplication for future LF data sets. Following the step

of removing the vignetting effects, residual aberrations were minimized next using an

analogous process. The LF data of a checkerboard extending across the full FOV was

taken. By comparing the peripheral sub-aperture images to the center one, the aberrations

were quantified and minimized via lateral shifting of pixels for future LF data sets.

3.5   Data Capture
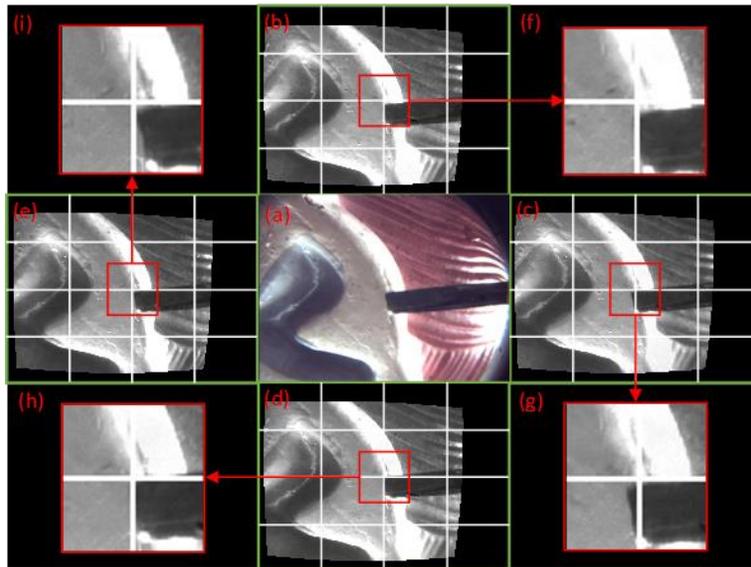
Fig. 3.5 LF data: (a) uncalibrated center and (b-e) calibrated peripheral sub-aperture images, and (f-i) magnified views of ray separations [25]**.**

Fig. 3.5(a) through 3.5(e) show the captured LF data organized into sub-aperture images

bordered in green according to the sample scheme shown in Fig. 3.4(c). Any two selected

sub-aperture images form a stereo image pair, where the parallax occurs along the line

connecting the corresponding subapertures. The captured scene consists of a part of the

bladder model and a screwdriver placed in front within the FOV to simulate a laparoscopic

surgical tool. For scaling reference, the width of the screwdriver is 3 mm while the

background bladder model is minified since it is farther away. The center sub-aperture image, Fig. 3.5(a), is uncalibrated and colored and was used as a reference for LF calibration. For the peripheral sub-aperture images, Figs. 3.5(b) through 3.5(e), the calibrated greyscale results extracted from the green color channel were shown along with white grid lines representing matching locations on the sensor for reference. Each of the original sub-aperture images has a high pixel resolution of 1280x960 pixels, which is the same as that of the native sensor. Due to the LF calibration, the FOV of the peripheral sub-apertures was cropped as seen in these images. Fig. 3.5(f) through 3.5(i) show magnified images of a small region, marked by a Red box on each of the corresponding sub-aperture images, 3.5(b) through 3.5(e), respectively. The small but slightly different displacements of the screwdriver relative to the white reference grids in the different sub-aperture images help to visualize the ray separations described in Fig. 3.2 and validate that the screwdriver is in front of the nominal working distance, L_{WD}.

The optical performance of the built prototype was limited by the quality of the stock lenses in the eyepiece and the use of a generic focusing lens. Therefore, we only utilized the greyscale images converted from the green color channel for further data processing to eliminate the effects of chromatic aberration, and we only used the center five angular samples to avoid severe vignetting and abberation-blurring, which increases significantly for sub-apertures farther from the optical axis. These five samples of sub-aperture images, however, are adequate to demonstrate the minimum data needed to achieve maximum data processing speeds and depth sensitivity from x or y-oriented image features in a PALFL system.

The angular sampling dimensions for the data in Fig. 3.5 were determined

37

experimentally. A 1 mm diameter iris was found to produce sufficient sub-aperture image quality and depth-of-field for the object distances of interest. A 0.91 mm pitch between the sub-apertures at the stop provided a balance between enough light ray separation at different object depths, absence of sub-aperture image aberration, and aliasing during digital refocusing.
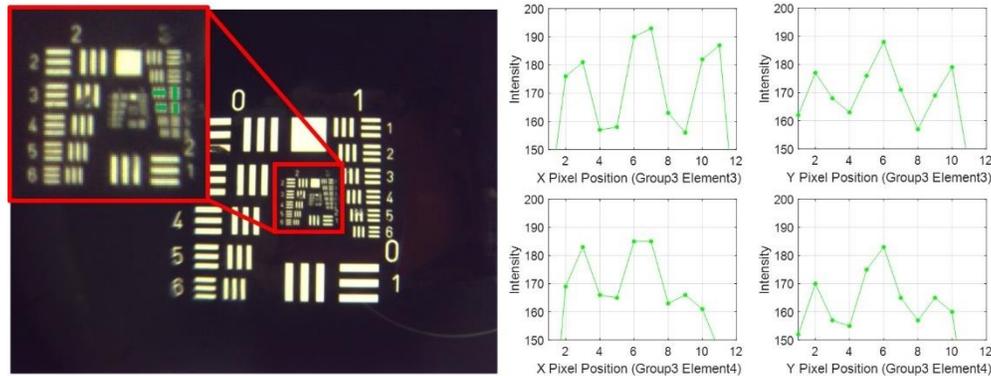
Fig. 3.6 The center sub-aperture image and intensity profiles of a 1951 USAF resolution target placed at an L_{WD} of 20 mm [25]**.**

The diffraction limited spatial resolution of the sub-apertures was measured using a 1951 USAF resolution target (groups 0-3) placed at an L_{WD} of 20 mm. Fig. 3.6 shows the center sub-aperture image, a zoomed in view of groups 2 and 3, and green channel intensity profiles along group 3, element 3 and 4. The bars in element 3 are clear while in element 4, they begin to diminish. This indicates that the diffraction limited spatial resolution is in between these two elements, which is ~10.7 lp/mm. Although the sub-aperture spatial resolution is limited by diffraction, the higher pixel sensor resolution is not wasted because it enables more precise measurement of disparity between sub-aperture images and will also be used for the conventional laparoscope, where the PA is fully opened and the optical resolution is higher.

3.6 Data Processing
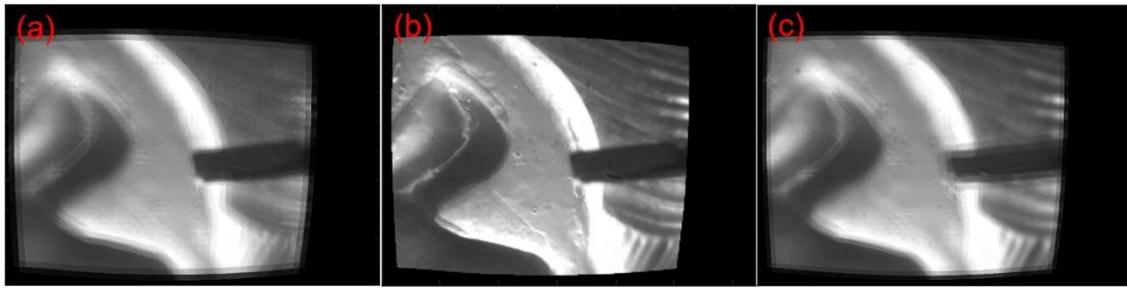
Fig. 3.7 Digital refocusing to depths: (a) near, (b) medium, and (c) far [25]

A modified open source code [28] was used to process the calibrated LF data for digital refocusing and to generate depth maps. Fig. 3.7 demonstrates digital refocusing for three image planes corresponding to near, medium, and far object distances. At near focus, the screwdriver is identifiable while the background is blurry. At medium focus, the white protrusion on the bladder model becomes clear. At far focus, the screwdriver and white protrusion are defocused while the pink line features on the right side are beginning to defocus. Because of the minimum angular sampling for this experiment, when refocusing to one extreme depth, the opposite one shows some aliasing, as seen by the edges of the defocused screwdriver when the focus is far.
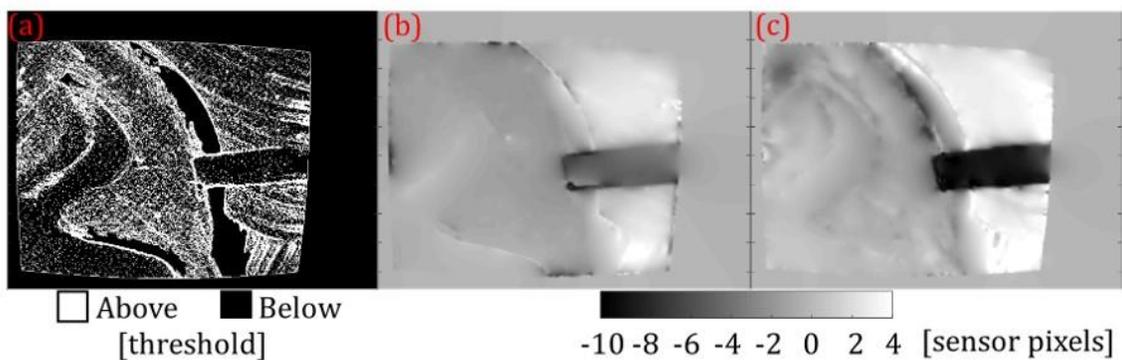
Fig. 3.8 (a) Intensity gradient thresholding of Fig. 3.5(a) for depth mapping noise reduction. Relative depth reconstruction maps based on (b) depth from focus contrast and (c) depth from multi-view correspondence feature matching [25]**.**

39

Fig. 3.8(a) was constructed by applying an intensity gradient threshold to Fig. 3.5(a) to highlight pixels containing strong image features for confident depth estimation. The depth was then estimated at those pixels while the other pixels were nullified. These null regions were interpolated based on the nearest confident depth estimation to construct a full depth map. This strategy reduced noisy depth estimations. Fig. 3.8(b) and 3.8(c) show full depth maps generated from algorithms based on focus contrast and on correspondence feature matching, respectively. For each object point, the depth estimation is obtained by measuring at the sensor the separation between light rays captured by adjacent sub-apertures (in units of sensor pixels). A negative pixel value indicates the separation occurred in the opposite direction, as shown in the zoomed view of Fig. 3.2 when comparing the ray separation from near and far images. Greyscale color illustrates that darker is closer and brighter is farther, allowing determination of relative depth.

Both depth maps identify the correct objects at three different depths, according to Fig. 3.7. However, depending on the image feature characteristics [28] and error from defocus aliasing, the algorithms perform differently. In the focus contrast map shown in Fig. 3.8(b), aliasing resulted in inconsistent depth estimation between the screwdriver's edges and body. Also, aliasing likely caused slight inconsistency between the two algorithms in the depth estimation of the farthest layer of depth. Therefore, the feature matching algorithm performs better for larger depth ranges. On the contrary, for the grey valley and surrounding white region on the left side of the FOV where aliasing is absent, the focus contrast map provides a smoother depth reconstruction.

3.7 Quantitative Depth Mapping
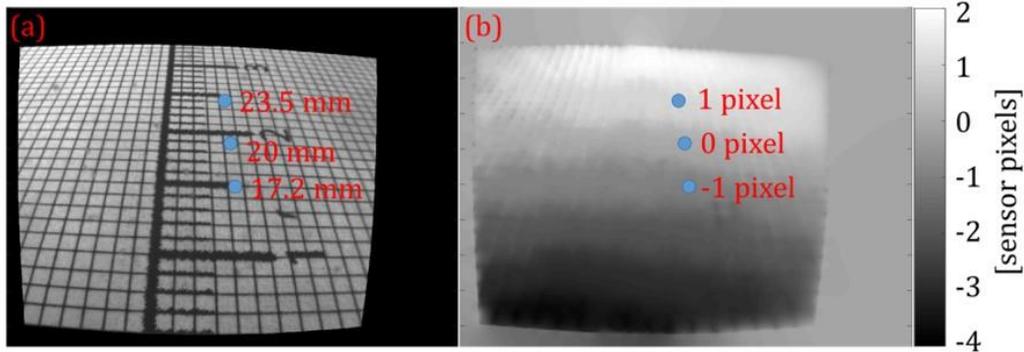
Fig. 3.9 A (a) tilted ruler object and its (b) measured depth map create a lookup table for converting ray separations to absolute depth values [25]**.**

A lookup table method was created to enable conversion of depth maps from the pixels measuring ray separations to absolute, quantitative depth values and to validate depth resolution based on the system design. Fig. 3.9(a) shows the center sub-aperture view of a 45° tilted ruler providing 0.7 mm depth intervals across the vertical FOV. After applying the same LF calibrations as those in the bladder model experiment, a smooth focus contrast depth map was generated in Fig. 3.9(b). Based on the measured ray separation, Fig. 3.9(b) highlights the pixels corresponding to $d_\pm$ and the L_{WD} of 20 mm. The corresponding pixels were found in Fig. 3.9(a), and knowing the ruler dimensions, the units were converted to physical depth.

These results were compared to our derived depth resolution study in Section 3.3. Due to the optical performance limitations discussed earlier, we experimentally determined the following prototype specifications. Knowing the real image to object magnification and manufacturer pixel size (B), the sensor spatial resolution in the object space, B_{obj}, of the current prototype was calculated to be 21.3 lps/mm for the center angular samples. Calculating B_{obj} using dimensions known in Fig. 3.6 yields a similar result. We measured

the equivalent D_{EP} from the angular samples shown in Fig. 3.5 to be 0.345 mm and the equivalent NA_{WD} of the sampled data to be 0.0074 for an L_{WD} of 20 mm. From Eq. (3.1), d_+ and d_- are 3.6 and 2.7 mm, respectively. Measured from the labeled data points in Fig. 3.9(a) and 3.9(b), the ±1 sensor pixel depths corresponding to d_+ and d_- are separated from the 0 sensor pixel depth on the ruler by +5 and -4 intervals, respectively. Knowing the depth between each interval on the tilted ruler is 0.7 mm, they correspond to measured depth resolutions of 3.5 mm and 2.8 mm, respectively, resulting in a maximum percent error of 3.7% in comparison to the theoretical values. Because depth estimation may be nonuniform depending on the algorithm used and the variation of an object's texture density, the percent error can fluctuate for different objects throughout the FOV. Nevertheless, the results presented here demonstrate the potential of the PALFL while depth estimation algorithms are continually being improved.

3.8   Summary

In this chapter, a PALFL was conceptualized to obtain high spatial resolution LF data up to that of the camera sensor for refocusing and quantitative depth mapping, without trading off angular resolution. By taking advantage of the PA's flexibility, this hybrid system integrates the high performance of existing 2D endoscopes with 3D LF imaging. Theory was then developed to analyze, compare, and design laparoscopes regarding adequate depth resolution. A bench-top prototype using an existing laparoscope objective demonstrated proof of concept by capturing stereo images in the LF and performing quantitative depth mapping according to the depth resolution theory.

From this work, the monocular architecture using uniaxial 3D depth cues is fully understood. Although strategies were mentioned for increasing the data capture rate, the

most significant limitation of the PA approach is the need for time sequential imaging of the stereo or subaperture images. Any time delay for 3D viewing would be undesirable in real time surgery. This issue needs to be addressed when expanding upon and modifying the design to account for WFOV and 3D capabilities.

4   A PRISM-BASED TRI-APERTURE LAPAROSCOPE FOR MULTI-VIEW ACQUISITION

The PALFL will now be modified to support WFOV and 3D capability. Since 3D viewing and reconstruction only minimally requires one stereo image pair, all the subapertures except for three can be disregarded. This chapter presents the design and prototype of a novel tri-aperture monocular laparoscopic objective (TAMLO) that can acquire both stereoscopic views for depth information and a WFOV for situational awareness. The stereoscopic views are simultaneously captured via a shared objective with two displaced apertures and a custom prism. Overlapping crosstalk between the stereoscopic views is diminished by incorporating a strategically placed vignetting aperture. Meanwhile, the wide FOV is captured via a central third aperture of the same objective and provides a 2D view of the surgical field 2x as large as the area imaged by the stereoscopic views. We also demonstrate how the WFOV provides a reference data set for stereo calibration, which enables absolute depth mapping in our experimental prototype. The work in this chapter is published in Appendix B, and any additional details mentioned here are supplementary.

4.1 Schematic Design of a TAMLO
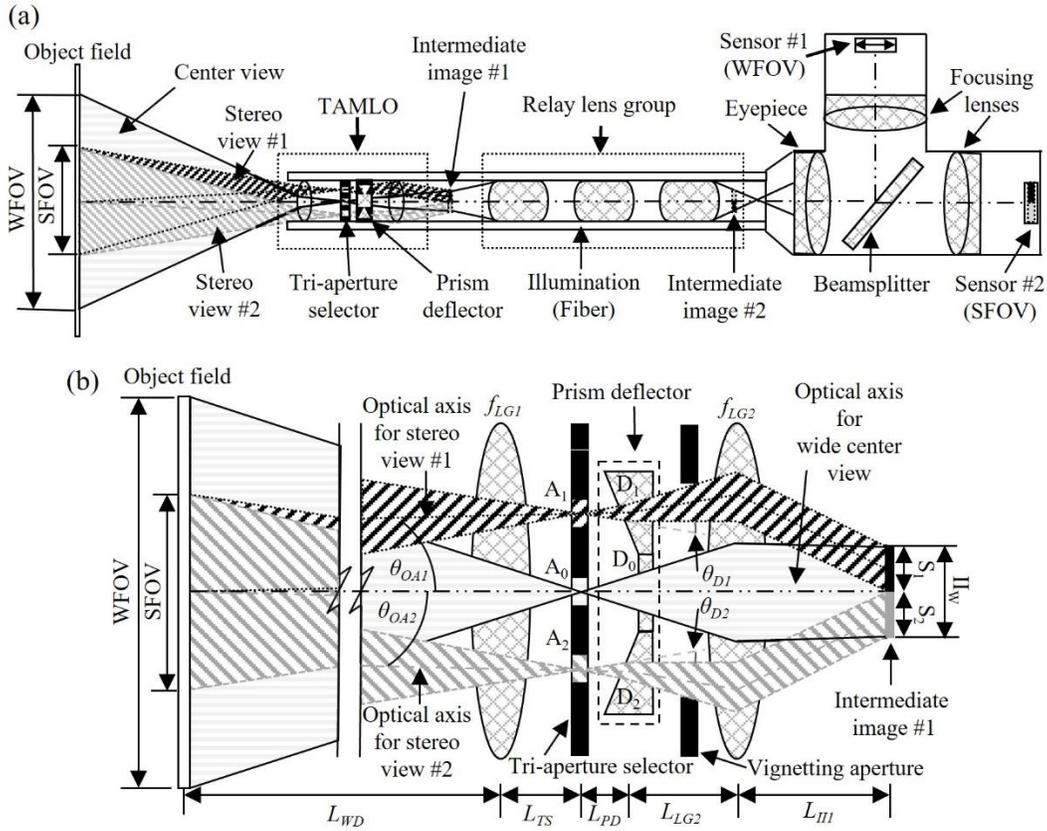
Fig. 4.1 (a) Proposed schematic layout of the prism-based tri-aperture monocular laparoscope design and (b) a magnified view of the TAMLO design with optical layout and key parametric specifications [29]**.**

Fig. 4.1(a) shows the schematic layout of the proposed prism-based tri-aperture monocular laparoscope design, and Fig. 4.1(b) fully illustrates the optical layout and key parametric specifications of the TAMLO design, which is the most significant contribution of this dissertation. Adopting the convention of light traveling from left to right, the TAMLO images the object field through three laterally displaced apertures and forms three different views of the object on the intermediate image #1, corresponding to a WFOV image of a large object field and two stereo FOV (SFOV) images of a smaller overlapping object field. Following the TAMLO, conventional laparoscope relay optics is utilized to relay the

45

intermediate image #1 and forms an intermediate image #2 outside the patient's body. After relaying the intermediate image #1 to the intermediate image #2, the spatial arrangement of the three views captured by the TAMLO is still preserved. Depending on the specific requirements and priorities of a particular laparoscope design, the three views may be captured by a single or multiple imaging sensors, which leads to different possible designs of the imaging probes. For instance, as illustrated in Fig. 4.1(a), we may adopt a scheme similar to the dual-channel imaging probes in the MRFL system. An eyepiece collimates the light from the intermediate image #2 and a beamsplitter then splits the collimated light into two imaging paths, one for capturing the WFOV image and one for the SFOV images. The type of beamsplitter used will depend on the type of tri-aperture selector used to separate the overlapping WFOV and SFOV images formed on the intermediate image #1. The two SFOV images through the two side-view apertures are recorded simultaneously on each half of the sensor #2, while the WFOV image through the center aperture is captured by the sensor #1. As long as the multiple views captured by the TAMLO are constrained within the maximally allowed diameter, II_W, of the intermediate image #1, the relay lens group only needs to be designed to support a FOV matching II_W. The relay lens must also support the maximal ray angle incident on the intermediate image #1 from all of the views. To avoid severe light loss from vignetting, it is preferred that the TAMLO and relay lens group are designed to be nearly telecentric at both intermediate image #1 and #2. It is worth noting that the TAMLO may be utilized alone for the option of a chip-on-tip form factor.

As schematically illustrated in Fig. 4.1(b), the TAMLO mainly consists of a front lens group with a focal length of f_{LG1}, a tri-aperture selector, a prism deflector, and a back lens

group with a focal length of f_{LG2}. The two lens groups are placed in front and behind the selector-deflector assembly to provide sufficient degrees of freedom during lens design optimization. The distances L between adjacent components are constrained by the method of optomechanical mounting and then precisely determined by the optical optimization. The chief ray bundles for the three views, which are highlighted by the different shaded regions in Fig. 4.1(b), are ideally maintained in separate regions of the two lens groups so that the local regions of the lenses can be optimized to the respective viewing angle. The aperture stop plane of the TAMLO is located at the tri-aperture selector, which consists of an on-axis central aperture A_0 and two decentered apertures of A_1 and A_2. A different viewing angle of the object field is seen by each of the three different aperture stops as indicated by the labeled optical axes for the wide center view and the stereo views #1 and #2, respectively. The prism deflector located adjacent to the tri-aperture selector is made up of individual prisms D_i corresponding to each aperture stop A_i. The central prism D_0 is effectively a thin plane parallel plate and does not change the outgoing ray angles from A_0, so a WFOV image, II_W, of the object field is formed at the intermediate image #1 and is centered about the central optical axis. D_0 could be removed leaving an air space, but instead is present to provide structural support and to manufacture the prism deflector as one piece. The side prisms, D_1 and D_2, bend rays transmitting through them by deflection angles of θ_{D1} and θ_{D2}, respectively, in opposite directions so that the stereo view images, S_1 and S_2, also located at the intermediate image #1, are laterally translated apart to the opposite sides of the central optical axis. Without the side prisms, S_1 and S_2 would be spatially overlapping on the exact same region about the central optical axis.

47

4.1.1 WFOV and SFOV Design

The two lens groups with focal lengths of f_{LG1} and f_{LG2} are optimized to support the WFOV captured by A_0 and the SFOV captured by apertures A_1 and A_2. To ensure that the WFOV and stereo views can be imaged by the same relay lens within a confined volume required by a rigid laparoscope, the dimensions of the stereo view images, S_1 and S_2, need to be constrained to the same circular region as the WFOV image, II_W. Furthermore, to avoid crosstalk between the two stereo views, S_1 and S_2 should not overlap at the intermediate image #1. Therefore, the maximum SFOV covered by both the S_1 and S_2 along the direction of the stereo aperture displacement, $SFOV_{\overline{A_1A_2}}$, shall satisfy:

$$SFOV_{\overline{A_1A_2}} \leq \frac{WFOV_{\overline{A_1A_2}}}{2} \tag{4.1}$$

where $WFOV$ is the maximum wide FOV of the objective. $WFOV$ is expressed as $2\tan^{-1}\dfrac{II_W}{2f_{TAMLO}}$, where f_{TAMLO} is the effective focal length of the objective, given as $\dfrac{1}{f_{LG1}} + \dfrac{1}{f_{LG2}} - \dfrac{(L_{TS}+L_{PD}+L_{LG2})}{f_{LG1}f_{LG2}}$, and II_W is the maximally allowed diameter of the intermediate image #1. The maximum SFOV along the axis orthogonal to the stereo aperture displacement can be as large as that of the maximum WFOV. To maximize the SFOV without causing overlapping between S_1 and S_2, the optimal value of the prism deflection angles, θ_{D1} and θ_{D2}, are determined by:

$$\theta_{D1} = -\theta_{D2} \approx \tan^{-1}\frac{II_W}{4f_{LG2}}. \tag{4.2}$$

4.1.2 Tri-aperture Selector

To separate the overlapping WFOV and SFOV images, the tri-aperture selector, which

48

determines the choice of the beamsplitter to be used in Fig. 4.1(a), may be implemented from different types of technologies that either block or encode the transmitting light. Blocking technologies include a mechanical shutter or a liquid-crystal device (LCD) that allows localized control of light transmission through a sub-region by switching the corresponding region on or off. Encoding technologies include a custom polarization device or color filter that allows localized control of light transmission through a sub-region by encoding different polarization states or spectral filters across the tri-aperture selector. If a blocking technology is used, either A_0 or $A_{1,2}$ is blocked in a time-sequential fashion so that the WFOV and SFOV images can be alternately captured by a single sensor, which eliminates the need for a beamsplitter and a second imaging probe in Fig. 4.1(a) and leads to a simpler system design with lower hardware cost. If an encoding technology is used, A_0 can be encoded oppositely to $A_{1,2}$. For example, orthogonal polarizers may be utilized for the apertures A_0 and $A_{1,2}$. Then, a corresponding polarizing or dichroic beamsplitter matching the encoded tri-aperture selector is used. This results in simultaneous capture of the WFOV and SFOV images by the two sensors illustrated in Fig. 4.1(a). The tradeoff of using this technology is half of each view's irradiance due to the encoding filters.

4.1.3   Prism Deflector

There are multiple variations of the prism deflector design that vary in manufacturability and light manipulation. For the example in Fig. 4.1(b), the back faces of the prisms D_0 through D_2 are co-aligned vertically while the front faces of D_1 and D_2 are tilted oppositely to achieve the proper amount of light ray bending. Since only one side of the prisms requires angled faces, the three prisms can be manufactured as one piece through diamond turning. After deflection, the light rays must pass through different portions of the

49

TAMLO's back lens group. Ideally, the ray bending by the prism deflectors is desired to be independent of ray incident angle. Realistically, the net ray bending by the prism deflectors is derived using Snell's law [30]:

$$\theta_{D1} = -\theta_{D2} = \alpha - \sin^{-1}\left[\sqrt{n^2 - \sin^2\theta_i}\sin\alpha - \cos\alpha\sin\theta_i\right] - \theta_i, \tag{4.3}$$

where α is the angle of the prism, n is the index of the prism material, and θ_i is the incident angle of the incoming light ray. Since $\theta_{D1,2}$ is dependent on a field or incident angle, $S_{1,2}$ are distorted accordingly, but they will be calibrated in post-processing. Combining Eqs. (4.2) and (4.3), the prism design can be approximated.

4.1.4 Image and Depth Mapping Resolution

Acquiring good image performance and sufficient depth resolution is critical to the TAMLO's functionality. The diameter of each aperture stop A_i determines the corresponding F/# or numerical aperture of each view, and thus the cut-off spatial frequency or limiting resolution of the objective. Therefore, within the optics volume constraints for a rigid laparoscope, they should be maximized for optimal optical resolution.

In the meanwhile, the lateral separation between the centers of the aperture stops A_1 and A_2, denoted as BL_{TS}, ultimately determines the effective baseline, *EBL,* of the stereo views and thus the depth resolution of the system. The *EBL* can be found by determining the lateral separation between the centers of the entrance pupils, which are optically conjugate to the aperture stops A_1 and A_2 through the first lens group and is expressed as:

$$EBL = \frac{BL_{TS} * f_{LG1}}{f_{LG1} - L_{TS}}, \tag{4.4}$$

where $BL_{TS} = \overline{A_1 A_2} = 2\overline{A_{1,2}A_0}$ and L_{TS} is the axial displacement of the aperture stop from

50

the first lens group. The location of the entrance pupils, denoted by L_{EP}, is found by imaging the tri-aperture selector through the first lens group and is expressed as $L_{EP} = L_{TS} f_{LG1}/(f_{LG1} - L_{TS})$. The *EBL* can also be described in object space as:

$$EBL = \left(L_{WD} + L_{EP} \right)\left(\left| \tan\left(\theta_{OA1}\right) \right| + \left| \tan\left(\theta_{OA2}\right) \right) \right|, \tag{4.5}$$

where θ_{OA1} and θ_{OA2} define the optical axis directions of the stereo views with respect to the central optical axis, and L_{WD} is the working distance optically conjugate to the intermediate image #1 through the TAMLO objective.

The depth resolution of the SFOV system is determined analogously to the PALFL. The average depth resolution, d, is given by:

$$d \approx \frac{2L_{WD}}{EBL} \left| \frac{-f_{LG1} f_{LG2} + t f_{LG1} + L_{WD}\left(f_{LG1} + f_{LG2} - t\right)}{-f_{LG1} f_{LG2}} \right| P, \tag{4.6}$$

where P is the limiting resolution or equivalent pixel size at the intermediate image #1 and $t = L_{TS} + L_{PD} + L_{LG2}$. The first fraction corresponds to the triangular geometry between the object field and the effective baseline while the second fraction corresponds to the magnification of the pixel from intermediate image #1 to the object field. The variables in this equation must be chosen properly to obtain adequate depth resolution for laparoscopic surgery. An *EBL* of 4 mm is standard for commercial stereo endoscopes and a pixel magnification of ~18 from intermediate image #1 to the object field is reasonable for recording the object field with an appropriately sized sensor. Using these constants, Fig. 4.2 plots the depth resolution rendered by a TAMLO as a function of the equivalent pixel size at the intermediate image #1 for three different working distances of 30, 60, and 120 mm. With a typical working distance of about 50 mm, a standard 2D laparoscope with an HD resolution sensor covers a circular object field of about 60 mm in diameter in a spatial

51

resolution up to 16 lps/mm in the object space. Consider the SFOV images are expected to provide the same circular field coverage as a standard 2D laparoscope with a single HD resolution sensor, the equivalent pixel size on the intermediate image #1 falls in the range from 1.5 to 3 µm, depending on the maximally allowed diameter of the intermediate image #1 due to package constraints. A pixel resolution between 0.5 and 4 µm on the intermediate image #1 can provide a depth resolution from 0.14 to 4.37 mm depending on the working distance. This indicates a properly designed TAMLO can provide sufficient depth resolution for surgical guidance.

Fig. 4.2 The depth resolution range of the TAMLO as a function of pixel size at intermediate image #1 for three different working distances [29]**.**

4.1.5   Stereo Image Crosstalk Solutions

The TAMLO must be able to acquire the three views without interference. The stereo images S_1 and S_2 are translated apart by the prism deflector according to the designed SFOV. However, as illustrated by Fig. 4.1, the object field is larger than the SFOV, which results in image points present outside of each designed stereo image $S_{1,2}$ due to lack of a field stop. The undesired image points from one stereo image will overlap the opposite stereo image across the central optical axis, resulting in crosstalk. To address this issue, a vignetting aperture can be placed after the prism deflector, as shown in Fig. 4.1(b). At this

location, the stereo ray bundles have diverged enough so that individual fields can be vignetted. The shape of the vignetting aperture depends on the prism deflector design. For the example in Fig. 4.1(b), an annulus vignetting aperture is required to allow the WFOV ray bundle to pass while blocking any stereo rays outside of the illustrated stereo ray bundle regions on the sides toward the central optical axis. Alternatively, if the two stereo views were oppositely polarized, a sensor with matching polarization on each half of the sensing area can eliminate crosstalk.

4.2   Tri-aperture Objective Lens Design

4.2.1   First Order Specifications

Table 4-1 First order lens design specifications for TAMLO

Specification	Value	Specification	Value
Working distance	120 mm	Effective stereo baseline	4 mm
Stereoscopic full FOV	26 deg.	Wavelengths	625, 506, 456 nm
Wide full FOV	39 deg.	Object resolution for stereoscopic view	6.25 lps/mm
Effective focal length	7 mm	Object resolution for wide view	2.1 lps/mm
Entrance pupil diameter per aperture	1.2 mm	Mechanical housing diameter	12 mm
Telecentricity	Image space telecentric	Maximum diameter of lenses	8 mm

Based on the analytical relationships and various constraints described in Section 4.1, we derived the first order specifications of the TAMLO design, which are listed in Table 4-1. To ensure adequate field coverage by both the WFOV and SFOV images, we chose a working distance of 120 mm. With a full FOV of 39° and 26° diagonally for the wide and stereo views, the system captures a circular region with a diameter of about 85mm and

53

54mm, respectively. To ensure adequate depth resolution, an *EBL* of 4 mm was chosen. We aim to fit the TAMLO prototype for a standard laparoscopic trocar and thus limited the mechanical housing diameter for the objective to be 12mm and the maximum optics diameter to be 8mm to account for housing and fiber illumination of 1 mm thickness each. We further limited the maximally allowed diameter of the intermediate image #1, *II$_W$*, to be 4.4 mm to allow direct capture of the images with a standard 1/3" imaging sensor and future development of the relay and imaging optics. These FOV and image size constraints led to an effective focal length of 7 mm for the objective. Between the two lens groups, we set an *f$_{LG1}$* of 47.5 mm for aberration compensation and balancing out the ray bending throughout the system and an *f$_{LG2}$* of 8 mm for sufficient power in a Petzval objective design. Since prisms are commonly made from N-BK7 glass material, a refractive index of 1.517 was selected for the prism deflector. The design of the side prisms was approximated using Eqs. (4.2) and (4.3) and an isosceles shaped prism for simplicity such that $\theta_i = \alpha$ for an incident ray parallel to the central optical axis. Based on the defined specifications, the bending $\theta_{D1,2}$ of the stereo chief rays is ±7.8° to adequately separate the stereo images, and the associated prism angle α was found to be ~15°. Using an HD sensor with 2.2µm pixels, Fig. 4.2 indicates that this design will provide ~2.5 mm depth resolution at a working distance of 120 mm, and a higher depth resolution can be achieved at a shorter working distance. Compared to the MRFL prototypes [8–12], the main tradeoff of implementing the stereo apertures was the reduction of the WFOV to about half. In other words, the challenge of the TAMLO design is to balance the optical performance between the SFOV and WFOV. The entrance pupil diameter for all three apertures was set to 1.2 mm, leading to an F/# of 5.8. The entire objective, however, is effectively F/1.35 because

it supports, in a monocular form factor, larger ray angles that come from the stereo aperture stops. The target spatial resolution in the object space was set to be 2.1 lps/mm and 6.25 lps/mm for the WFOV and SFOV, respectively. The object resolution specification is weighted lower for the WFOV than the SFOV because it is mainly used for peripheral awareness. Meanwhile, the constraints applicable to conventional rigid laparoscopes were also met. The optical design is constrained for image space telecentricity so that relay lenses can be easily inserted after the objective lens.

4.2.2 Lens Design Process

The starting point of the TAMLO lens design was based on the existing MRFL and commercial 3D endoscope objectives. In all lenses and the prism deflector, rays were constrained within a 7.2 mm clear aperture diameter, or 90% of the maximum lens diameter. This diameter is slightly larger than that of the MRFL to provide additional design volume for the SFOV ray paths. For the SFOV, the object field was sampled across the +x and +/-y region with an aspect ratio of 4:3 corresponding to the sensor because each stereo view system is bilaterally, rather than rotationally, symmetric. Some of these field points along the edge are noted in Fig. 4.3(d). In the middle stage of the design process, custom lenses and prism deflector designs were allowed to determine the maximum achievable image performance and avoid local minimum solutions. Throughout this phase, the size of the SFOV was maintained at the same size as the conventional 2D laparoscope while the other first order specifications were adjusted accordingly based on what was practical, given the required laparoscope constraints and the incorporation of stereo apertures. Since the size of the intermediate image II_W can vary and be magnified accordingly after being relayed, it was kept to < 5.6 mm diameter to avoid vignetting and

55

ensure possible integration with relays. In the late stage, further constraints were implemented to convert the custom components into manufacturable ones. Tolerance sensitivity reduction was also applied to produce realistic lens shape factors. In addition, the lens housing would be 3D printed, so optomechanical tolerances were loosened accordingly to ensure assembly variation was accounted for. The custom lenses were then converted to stock lenses to lower costs and achieve rapid prototyping. Image performance was slightly reduced as a result but can be restored in future versions.

4.2.3   Manufacturable Lens Design
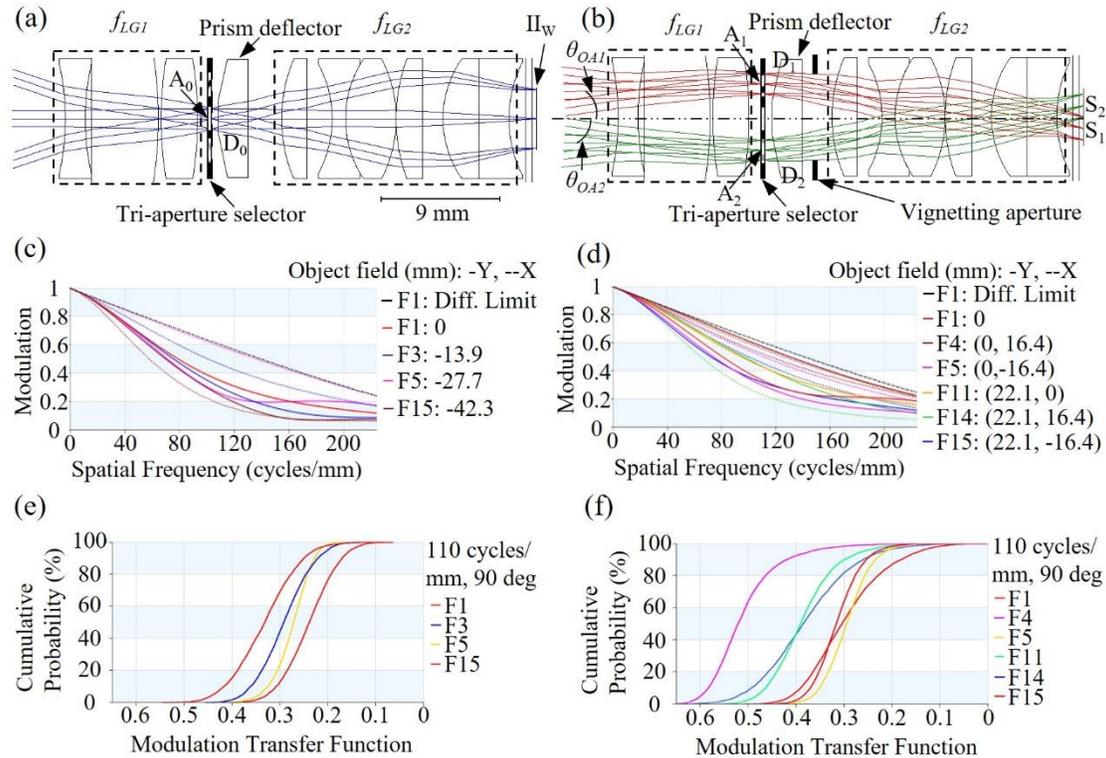
Fig. 4.3 Manufacturable TAMLO lens design for (a) WFOV and (b) SFOV acquisition. Corresponding (c, d) polychromatic MTFs and (e, f) tolerance analyses indicate sufficient performance for prototyping [29]**.**

Table 4-2 Lens prescription for design in Fig. 4.3

Surface #		Curvature radius (mm)	Thickness (mm)	Index	Abbe #	Y Decenter (mm)	Alpha tilt angle (°)
0	Object	Plano	120				
1	Lens 1	-18.86	1.5	1.517	64.2		
2		18.86	0.474				
3	Lens 2	Plano	5	1.785	25.7		
4		-31.39	0.215				
5	Lens 3	27.43	2.74	1.517	64.2		
6		-27.43	1.016				
7	Multiview selector	Plano	A_0: 0.75			0	
			A_1: 1.011			2.5103	
			A_2: 1.011			-2.5103	
8	Multiview deflector	Plano	D_0: 2.261	1.517	64.2	0	0
			D_1: 2.001			2.5103	-11.7967
			D_2: 2.001			-2.5103	11.7967
9		Plano	2.276			D_0: 0	
						D_1: 2.5103	
						D_2: -2.5103	
10	Lens 4	11.37	1.93	1.517	64.2		
11		Plano	1.869				
12	Lens 5	-8.6	1	1.847	23.8		
13		8.13381	3.8	1.806	40.9		
14		-9.73765	0				
15	Lens 6	12.92	2	1.517	64.2		
16		Plano	1.519				
17	Lens 7	7.85	2.7	1.785	25.7		
18		Plano	2.25	1.785	25.7		
19		9.42	1.238				
20	Sensor cover glass	Plano	0.47	1.517	64.2		
21		Plano	0.35				
22	Intermediate image 1	Plano					

The manufacturable TAMLO lens design to be prototyped is shown in Fig. 4.3(a) and 4.3(b) for WFOV and SFOV acquisition, respectively, and the lens prescription is shown

in Table 4-2. Both figures show the same set of monocular lenses, and all of them except one are stock components. It was found that the achromatic doublet near the middle and the field lens should be kept in the meniscus shape to maintain good image performance. Since meniscus lenses are uncommon in stock lenses, the achromat was custom made and the field lens was formed using two singlets of the same glass. For the prism deflector, a reversed deflection design was chosen so that it could be more easily manufactured as one piece. It is essentially an obtuse angle prism with the top flattened out. In this real lens design, $\theta_{D1,2}$ is ±6.22° for a horizontal incident ray, and the associated prism angle α is 11.9°. Compared to these real values, the corresponding theoretical values from the beginning of this section are slightly different because they did not account for real thick optics and relied on approximations for simplification. Yet, those values were a good starting point for the prism design. Furthermore, the dispersion from the prism deflector was accounted for. The prism deflector design can be thought of as a segmented lens. For example, the one in Fig. 4.1(b) approximates a concave-plano lens, and the one in Fig. 4.3(b) approximates a convex-plano lens. Because the prism deflector looks like a conventional lens, its dispersion is similar to the lens it approximates. This dispersion was suppressed by using conventional dispersion compensation from the other monocular lenses with different glass types during the lens optimization process.

Fig. 4.3(a) also shows the light rays from the WFOV transmitting through the central aperture stop A_0 and plane parallel plate D_0. Because there is no deflection from D_0, the WFOV system is modeled as rotationally symmetric, and the center of II_W is on the lens optical axis. As the chief rays travel to II_W, they are collimated. This indicates the system is image space telecentric. Telecentricity is one of the image quality limiting constraints,

which can be removed if using the TAMLO as a chip-on-tip system, where the sensor is placed at intermediate image #1. Telecentricity is maintained in this prototype design to demonstrate design feasibility. Similarly, Fig. 4.3(b) shows the light rays from the SFOV transmitting through the top and bottom aperture stops $A_{1,2}$. On the prism deflector's left side, the ray bundles from each view were constrained so that they only interact with their corresponding prism surface. On the prism deflector's right side, all ray bundles share the same flat surface. $D_{1,2}$ bend the corresponding ray bundles according to Eqs. (4.2) and (4.3) so that the SFOV images are translated to the upper or lower side of the optical axis without surpassing the WFOV boundary, thus allowing for simultaneous stereo image pair acquisition on a single sensor. Compared to Fig. 4.1(b), the prism design here deflects the corresponding stereo images $S_{1,2}$ to the opposite sides of the optical axis rather than to the same side.

By having lens groups in front and behind the prism deflector, the optical system has sufficient degrees of freedom to achieve a balanced image performance between the two imaging modalities. Comparing Fig. 4.3(a) and 4.3(b) further illustrates that for the lenses closest to the tri-aperture selector, the SFOV ray bundles only occupy the outer local portions of the lenses while the WFOV ray bundles mainly occupy the central local portion. This indicates that these lenses have more flexibility to impact the imaging modalities separately, and aspheric surfaces can add additional degrees of freedom. The polychromatic MTFs in Fig. 4.3(c) and 4.3(d) corresponding to the WFOV and SFOV, respectively, show that adequate image performance can be achieved with this lens design. Although astigmatism impacts the WFOV especially at the higher frequencies, only peripheral awareness is essential rather than high resolution. Thus, slightly lower contrast

is acceptable in the WFOV system. Quantitatively, the MTFs indicate that the lowest modulation for the WFOV at 37 lps/mm (2.1 lps/mm in object space) is 0.7 and for the stereo view at 109 lps/mm (6.25 lps/mm in object space) is 0.24. The cutoff frequency at 227 lps/mm corresponds to the sensor (Allied Vision Alvium 1800 U-500c) used for capturing intermediate image #1. Using stock lens and 3D printing tolerances, the tolerance analyses for the WFOV and SFOV systems in Fig. 4.3(e) and 4.3(f), respectively, confirm that this design will maintain adequate performance after assembly. The modulation will be greater than 0.1 at 110 lps/mm (6.12 lps/mm in object space) for both systems. According to Table 4-1, this approximately meets the object resolution criteria for the SFOV and exceeds it for the WFOV.

4.2.4 Stereo Image Crosstalk Reduction
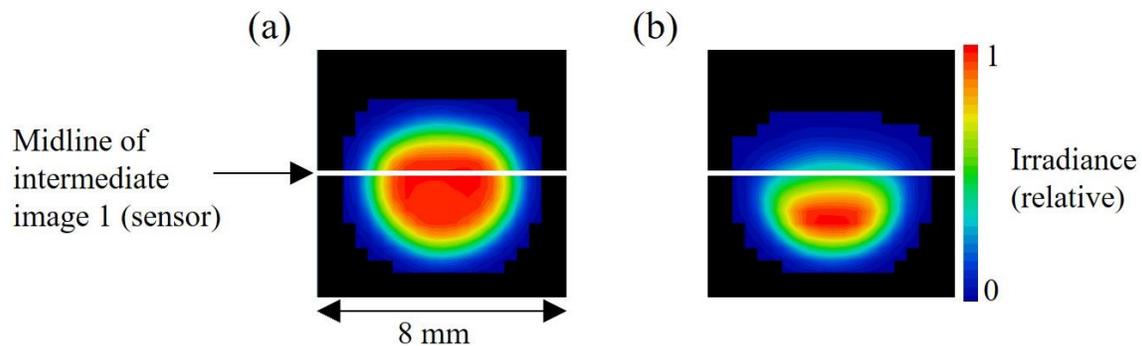
Fig. 4.4 Simulation of S_1 image size and overlap when SFOV is extended with (a) no vignetting and (b) inserted vignetting aperture [29]**.**

As described in Section 4.1.5, although $S_{1,2}$ are translated apart, they are larger than as designed in Fig. 4.3(b) because the object field is larger than the chosen SFOV. This results in overlapping crosstalk between $S_{1,2}$. The amount of overlap is simulated by seeing how much S_1 crosses onto the upper half of the sensor, as shown in Fig. 4.4(a), when the SFOV is extended by a large amount. The boundaries of the larger S_1 indicate when the edge

60

apertures of the TAMLO's lenses begin to vignette S_1. By symmetry, S_2 would overlap just as much on the lower half of the sensor. The amount of overlap past the midline is significant and would corrupt a major portion of the designed $S_{1,2}$ area. Since it is not possible to limit the size of $S_{1,2}$ with a field stop, a vignetting strategy is implemented here. Because the prism deflector deflects the images to opposite sides of the optical axis, a circular vignetting aperture can be inserted right after it to significantly reduce the overlapping crosstalk, as shown in Fig. 4.3(b). Comparing to Fig. 4.3(a), the placement of this vignetting aperture will not interfere with the rays from the WFOV system. This technique can preserve most of the designed $S_{1,2}$, as simulated in Fig. 4.4(b), where most of the overlapping crosstalk is diminished after insertion of the vignetting aperture. Note that in this demonstration, the relative irradiance is 0.5 at the midline of intermediate image #1 because the circular vignetting aperture was designed to half vignette there. If the residual crosstalk needs to be further reduced, the vignetting can be increased and a calibration in post-processing could recover the irradiance that was lost in the designed $S_{1,2}$ area. For prism deflector designs that deflect to the same half of the sensor, such as the one in Fig. 4.1(b), the circular vignetting aperture would not work because it would vignette at the edges of the sensor instead of the center where the overlapping crosstalk occurs. Instead, a similar vignetting solution could be achieved with an annulus vignetting aperture.

4.3 Prototype Assembly
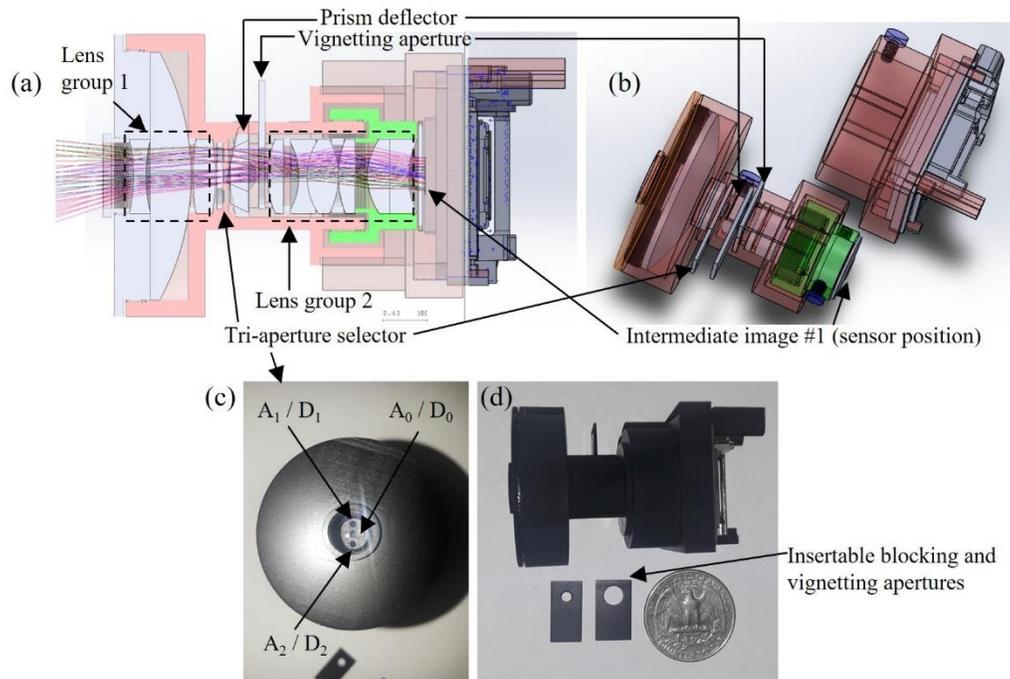
Fig. 4.5 TAMLO (a, b) optomechanical housing design and (c, d) prototype assembly [29]

A basic lens housing was designed and 3D printed for assembling the TAMLO prototype, as shown in Fig. 4.5. The second stock lens prescription did not come in the same diameter as the other lenses, resulting in the large housing in the front. The housing contains railings to align the tri-aperture selector, prism deflector, and sensor along the same axis. To separate the overlapping WFOV and SFOV images, the tri-aperture selector was a manual shutter that blocked either A_0 or $A_{1,2}$. Rectangular aperture blockers were simply inserted into a slot of the housing, resulting in time-sequential acquisition between the two imaging modalities. There is an additional slot after the prism to insert a vignetting aperture to reduce the overlapping stereo image crosstalk. For prototype evaluation, a real sensor was mounted at intermediate image #1. The entrance pupils corresponding to A_{0-2} can be seen clearly in the frontal view of Fig. 4.5(c).

4.4 Raw Data Capture
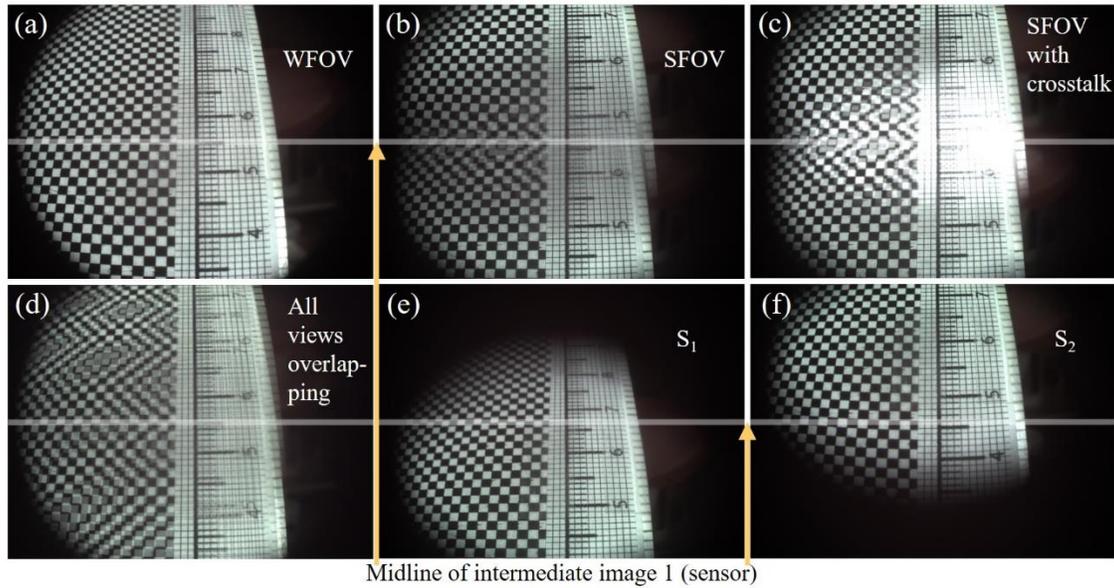
Fig. 4.6 TAMLO prototype raw data: (a) WFOV, SFOV simultaneous stereo image capture (b) with and (c) without vignetting aperture, (d) overlapping WFOV and SFOV images, (e) S1 and (f) S2 captured independently without vignetting aperture [29].

Fig. 4.6 illustrates the raw data acquired from the working TAMLO prototype. The object field is a ruler lying on a planar checkerboard that is tilted so that the object depth linearly increases as a function of image height. For all the images, distortion can be observed by looking at the curvatures of lines that should be straight. Fig. 4.6(a) shows the WFOV image while the SFOV apertures are blocked. Along the vertical axis, ~5.5 cm of the ruler can be seen. Fig. 4.6(b) shows the SFOV images captured by both stereo apertures simultaneously with the vignetting aperture in place. The stereo images were translated by the prism deflector to the top and bottom half of the sensor without exceeding the WFOV image. Each of the stereo images sees ~2.5 cm of the ruler. Thus, in quantitative comparison, the WFOV shows twice the SFOV in the vertical or baseline direction when the stereo images are captured simultaneously. Fig. 4.6(c) shows the same stereo images taken with the same exposure settings but without the vignetting aperture. Along the

midline of the sensor, the strong presence of the overlapping crosstalk reduces the contrast and the sum of the irradiance results in saturated pixels. There is still some residual crosstalk in Fig. 4.6(b), but it has been significantly reduced, and the vignetting aperture size can be further optimized in future prototypes. Fig. 4.6(d) shows the overlap between all three views without a method of blocking or encoding either A_0 or $A_{1,2}$, thus resulting in unusable data. Fig. 4.6 (e) and 4.6(f) show S_1 and S_2, respectively, captured independently without the vignetting aperture. They demonstrate the extent of overlap that causes the crosstalk. Overall, the image quality of these figures appears sufficient, as predicted during the lens design phase.

4.5   Calibration Overview

To calculate correct disparity and absolute depth maps, the stereo systems require camera parameter and distortion calibration. For a conventional stereo system with two independent cameras, methods for calibrating camera parameters and distortion have been thoroughly developed [31]. The conventional calibration assumes each of the cameras has rotational symmetry, so the lens distortion can be modeled with a radial polynomial. The TAMLO effectively creates two virtual stereo cameras with their optical axes tilted from each other, but their distortion model is no longer rotationally symmetric. Instead, because the TAMLO captures each stereo image with an off-axis aperture, the distortion model is bilaterally symmetric and can have additional distortion from the finite thickness of the prism deflector. Analytically calibrating the unique distortions in the TAMLO would require rigorous theoretical analysis. Alternatively, a numerical solution can be developed by taking advantage of the additional WFOV data, which was captured with rotational symmetry.

Fig. 4.7 Images of (a) WFOV and (b) SFOV before and (c) after thin lens modeling and distortion calibration. Features outlined in red and the image region labeled S1U correspond to the same region of checkerboard squares in the object field [29]**.**

The goal of our calibration was to obtain the intrinsic parameters of the TAMLO optics and model the imaging process as a projection by an ideal thin lens along with distortion correction. The process is summarized here and will be fully discussed in Chapter **Error! R eference source not found.**. First the WFOV system was calibrated using the well-established method in [31] so that it could be modeled as a pinhole camera with radial distortion correction. Because the apertures $A_{1,2}$ are in the same plane as A_0, it can be assumed that their representative pinhole models also lie in the same plane as the one for A_0. A planar checkerboard was then placed perpendicular to the optical axis of the TAMLO lens and at the working distance conjugate to the image sensor. This object field was captured by II_W and $S_{1,2}$, as shown in Fig. 4.7(a) and 4.7(b), respectively. S_2 is like S_1 so it isn't shown. Using the WFOV calibration data, II_W was undistorted (II_{WU}), as shown in Fig. 4.7(c). Within the designed SFOV, corresponding image features outlined in red between II_W and $S_{1,2}$ were determined so that the light rays in the SFOV system could be digitally bent by translating $S_{1,2}$, pixel by pixel, to the corresponding pixel coordinates that contain the matching image in II_{WU}. In other words, the light rays from the SFOV system were digitally bent so that they focused with the calibrated chief rays in II_{WU}. This is illustrated in Fig. 4.7(c), where the calibrated stereo image S_{1U} is directly overlapping II_{WU}

after digital bending and summed together for visualization. S_{1U} and the area of II_{WU} underneath S_{1U} look the same, so the brightness is doubled after summation. This technique effectively removes both the distortion from the stereo images and the translation from the prism deflector and converts the TAMLO into a thin lens model. The amount of digital bending is stored for each pixel of $S_{1,2}$ in a lookup table for calibrating any subsequent stereo images. Although the lookup table was generated from a 2D object field, it applies to 3D object fields because each pixel of $S_{1,2}$ corresponds to unique object angles defined by the 3D object point location and $A_{1,2}$. The final step of this calibration was to determine the parameters of the ideal thin lens model. The image distance was already determined from the focal length of the WFOV system's pinhole model. The object distance to any point on the planar checkerboard placed at the conjugate working distance could be determined using the extrinsic parameters from the pinhole model. Knowing object and image distance, the effective focal length of the thin lens model was found from the thin lens equation. To find the baseline between the pinhole models of $A_{1,2}$, two object points at different depths were captured by $S_{1,2}$, which were then calibrated using the lookup table. First order ray tracing was performed from the two known object points to the stereo pinhole models of unknown baseline, refracted by the effective focal length, and then further traced to the corresponding image points in $S_{1U,2U}$. The baseline could then be algebraically solved. Conceptually, this calibration recovers an ideal thin lens model that obtains depth from defocus.

4.6   3D Viewing and Depth Mapping
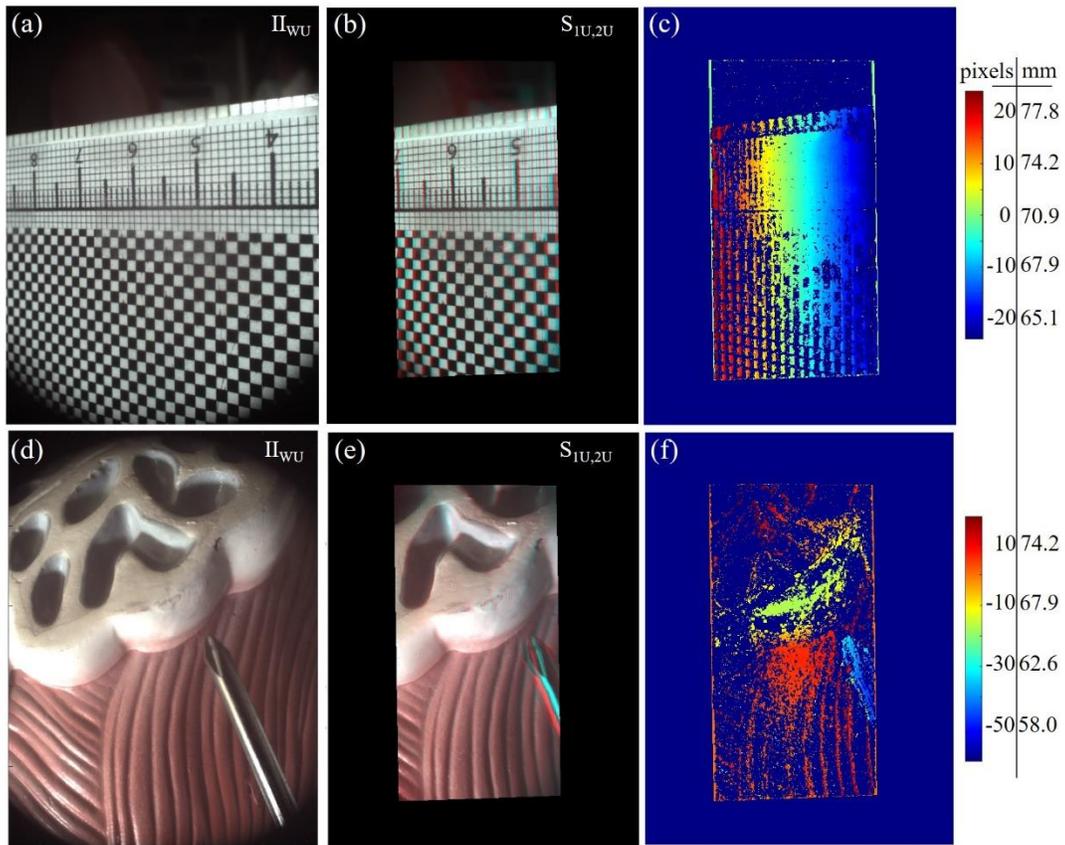
Fig. 4.8 Fully calibrated TAMLO results of a tilted ruler (top row) and a 3D bladder model (bottom row): (a, d) WFOV, (b, e) SFOV images overlaid as an anaglyph, (c, f) depth maps in units of pixel disparity and absolute depth [29]**.**

After calibration, fully processed TAMLO results of the tilted ruler and a 3D bladder model were generated to complete the proof of concept. The following results were rotated counterclockwise by 90° from the original image orientation so that the stereo views can be displayed with parallax along the horizontal direction and can be viewed with 3D glasses. Fig. 4.8(a) and 4.8(d) show the undistorted WFOV images, as indicated by the straightened lines of the ruler and checkerboard. Fig. 4.8(b) and 4.8(e) show the calibrated SFOV images overlaid as a red and cyan anaglyph, which demonstrates parallax based on the difference in disparity between corresponding object points. Close observation of Fig.

4.8(b) illustrates a reversal in the arrangement of the cyan and red colors from the 4.5 to 7 cm tick marks. This indicates the center of the image has zero disparity and is the conjugate working distance to the image sensor while the right and left of the image are closer and farther away, respectively. Similarly, Fig. 4.8(e) illustrates large disparity at the screwdriver, indicating that it is much closer than the bladder model. Fig. 4.8(d) and 4.8(e) demonstrate good image quality for both WFOV and SFOV imaging in a surgical setting. Finally, the calibrated stereo images were processed to produce accurate depth maps in Fig. 4.8(c) and 4.8(f). The color bars have units of pixel disparity, which were then converted to absolute depth values in millimeters as shown on the right of the color bars using the thin lens model parameters found during calibration. Although the original lens design had a 120 mm working distance, tolerances in the 3D printed resulted in a backward shift of the sensor, so the conjugate working distance or 0 pixel disparity in these figures is located at ~71 mm. According to Fig. 4.2, the depth resolution increases to ~1.5 mm at this closer working distance. Fig. 4.8(c) shows the linear change in depth corresponding to the tilted ruler without any depth resolution artifacts, thus confirming depth mapping ability. Similarly, Fig. 4.8(d) shows the closer distance of the screwdriver and the correct surface profile of the bladder model.

4.7   Summary

In this chapter, a novel prism-based tri-aperture monocular laparoscopic objective was conceptualized, designed, prototyped, and calibrated. This system achieved WFOV and SFOV imaging with sufficient image quality. Compared to the SFOV, the WFOV sees 2x the object field along the baseline axis. Overlapping crosstalk between the stereo images was also addressed. The calibration of the stereo views using the rotationally symmetric

WFOV image as a reference was then introduced. Completion of the calibration enabled removal of distortion from the WFOV and SFOV images, which were then processed to generate accurate, absolute depth maps. The TAMLO certainly demonstrates the potential for optically combining WFOV and SFOV imaging in a compact system. Such a system may pave the way towards restoring the binocular and large, foveated FOV qualities of human vision within the minimally invasive surgical setting.

5 CALIBRATION OF TRANSVERSE RAY AND PUPIL

ABERRATIONS FOR LIGHT FIELD CAMERAS

To produce the processed results shown in Chapters 3 and 4, the respective prototypes were calibrated using a custom LF calibration technique that was developed for generic LF cameras. This chapter presents this generalized calibration of transverse ray and pupil aberrations for LF cameras. The calibration step is necessary to meaningfully interpret the captured raw LF data. More specifically, the accuracy of reconstructing depth maps or performing digital refocusing in LF cameras largely depends on how well the spatial and angular samples of light rays can be obtained.

For instance, ray sample errors induced by optical aberrations in a LF camera reduce the contrast of any reconstructed image. To recover the contrast and depth plane, the LF can be digitally corrected by using the ray tracing data from the nominal lens design, if available [24]. However, most commonly nominal lens prescription is not accessible to end users. Additionally, even if available, due to tolerances in optomechanical design, the ray tracing data can be inaccurate. Similarly, vignetting has been shown to impact light field camera calibration [32] and depth mapping error [33]. Because aberrations and vignetting alter the light field, they confuse light field depth reconstruction algorithms [28]. Therefore, all factors impacting the light field must be considered and calibrated for proper recovery of light field camera functionality.

In this chapter, we introduce a calibration method based on measurements of fiducial markers on a checkerboard for modeling the imaging properties of light field cameras. The calibration accounts for vignetting, transverse ray errors, as well as pupil aberration, and can be applied to light field camera modeling of arbitrary pupil sampling systems. We then

demonstrate the capability of this calibration by calibrating the TAMLO prototype from Chapter 4, which technically captures simultaneous stereo views via artificially induced transverse ray errors. The work in this chapter is published in Appendix C, and any additional details mentioned here are supplementary.

5.1 Mathematical Model for Light Field Camera Calibration

A light field camera angularly samples the geometrical light rays of a 3D scene according to the directions of the rays apparently emitted by the scene, which leads to the well-known 4D light field function, $L(u,v,s,t)$, for representing the ray radiance as a function of ray position (s,t) and direction (u,v) [34]. Among the various existing light field cameras, there exist two fundamental architectures—Light Field (LF) 1.0 (a.k.a. plenoptic 1.0) and Light Field 2.0 (a.k.a. plenoptic 2.0). In cameras based on the LF 1.0 architecture, an aperture array or a micro-lens array (MLA) is typically placed at the back focal plane of an objective lens and a detector is placed at the position that is optically conjugate to the stop of the objective lens. In this architecture, the MLA pitch determines the spatial resolution and forms the ST-plane and the sensory pixel pitch determines the view resolution and forms the UV-plane. In cameras based on the LF 2.0 architecture, the aperture array or the MLA is placed at a position conjugate to the main stop of the objective and the detector is placed at a location optically conjugate to the intermediate image plane through the MLA. In this architecture, the MLA pitch determines the view resolution and forms the UV-plane, and the sensor pixel pitch determines the spatial resolution and forms the ST-plane. The mathematical model and calibration process discussed below assumes a LF 2.0 architecture, but it is generally applicable to LF 1.0 architecture with adaptation.

71

5.1.1 Subaperture Images in Light Field Camera
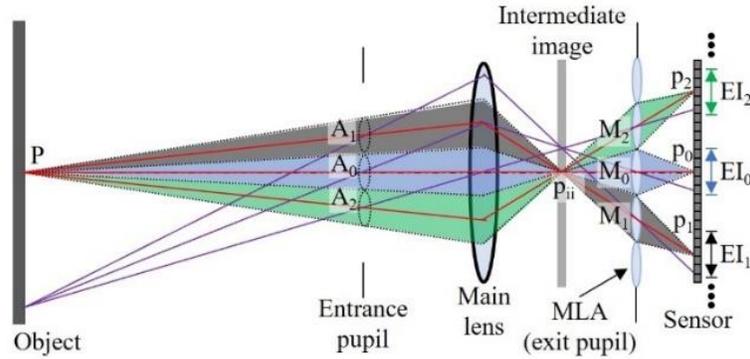
Fig. 5.1 Working principle of light field camera 2.0 [35]

Fig. 5.1 illustrates an MLA-based LF 2.0 camera design, consisting of a main lens, an MLA, and a sensor [36]. The light field of a 3D object is captured by the main lens, forming an intermediate image plane. The MLA is placed behind the intermediate image of the main lens, and the sensor is placed at a location optically conjugate to the intermediate image plane. Each lenslet of the MLA sees a different viewing angle of the intermediate image and forms a corresponding elemental image (EI) on a different portion of the sensor. Because the lenslet apertures limit the ray bundles of each viewing angle, the stop and exit pupil are defined at the MLA plane, which is optically conjugate to the entrance pupil of the main lens.

In the configuration in Fig. 5.1, the object's light field is sampled spatially by the EIs on the sensor and angularly by the MLA, corresponding to the ST-plane and UV-plane for ray position and direction sampling in the 4D LF function, respectively. Each of the lenslet apertures can be mapped to a subaperture location on the entrance pupil, and each of the EIs can be mapped to a portion of the intermediate image. For instance, the light field of point P in Fig. 5.1 is captured through three subapertures, A_0, A_1, and A_2, on the entrance pupil or equivalently through three lenslets, M_0, M_1, and M_2. It is imaged onto the same

point p_{ii} on the intermediate image plane, and onto three spatially separated pixels, p_0, p_1, and p_2, on three elemental images, EI_0, EI_1, and EI_2. Once a light field is captured on the sensor, the raw image can be reorganized into individual EIs according to these conjugate planes for post-processing and 3D reconstruction.

For the purpose of calibration and further depth reconstruction, the image formation process shown in Fig. 5.1 can be simplified into a model consisting of a virtual thin lens located on the entrance pupil and an array of subapertures located on the virtual thin lens. The subaperture locations can be found by tracing the subaperture chief rays through the lenslet centers of the MLA or the equivalent array elements in a LF camera system. We can then map each of the raw EIs on the sensor plane onto an equivalent EI on the image plane of the virtual thin lens. Fig. 5.2(a) and 5.2(b) show the mapping of the EI_0 and the EI_1 captured by the lenslets M_0 and M_1 through the subapertures A_0 and A_1 onto the intermediate image plane of the virtual main lens, respectively. In the mapping process, the EIs on the sensor plane are repositioned onto the virtual intermediate image plane such that their matching image pixels (such as p_0 and p_1) from the sensor plane are now virtually overlapping. The virtual overlapping represents the expected image formation from the virtual thin lens. During the mapping, a unit magnification is assumed between the raw EIs on the sensor plane and the equivalent EIs on the intermediate image plane, except for the sign from flipping the EIs to match the orientation of the intermediate image. The equivalent focal length of the virtual thin lens accounts for the difference in magnification between the raw EIs and their corresponding intermediate image induced by the lenslet magnification.

Fig. 5.2 Light field reorganized into (a) center subaperture image and (b) peripheral subaperture images from a virtual thin lens model [35]**.**

For generalization, these equivalent EIs obtained through remapping will be referred to as subaperture images for the remainder of the chapter. We calibrate the light field system based on this subaperture image setup. Note that the raw light field data from a LF 1.0 camera design [24] can be reorganized into these subaperture images in an analogous fashion. Also note that the raw light field data from a programmable aperture-based light field camera design [26] is already organized according to the subaperture image setup. Therefore, our proposed calibration method is applicable to any pupil sampling or light field camera design after the raw data is organized into subaperture images.

5.1.2   Computational Model for Depth Reconstruction
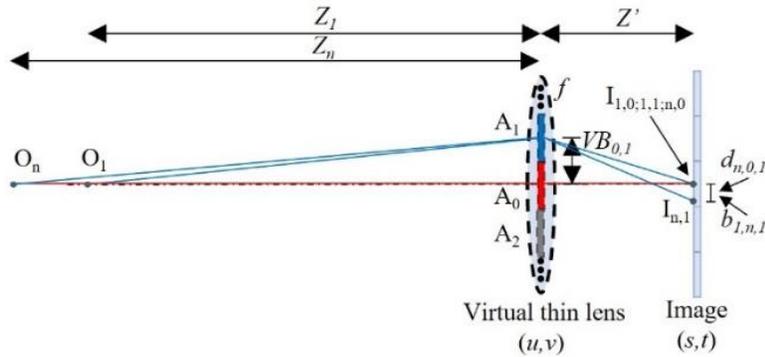
Fig. 5.3 Ray tracing for two known object points through two subapertures [35]**.**

To reconstruct the depth map of a 3D scene from recorded subaperture images, the

parameters of the virtual thin lens as well as the subaperture locations need to be determined, as shown in Fig. 5.3. f is the equivalent focal length of the virtual thin lens, Z' is the distance between the virtual thin lens and the subaperture images, and A_i is the center location of ith subaperture where $i=0$ corresponds to the aperture centered with the optical axis of the virtual lens. $VB_{i,j}$ represents the baseline distance between two subapertures, A_i and A_j. Here A_i corresponds to (u,v) coordinates while the pixels of the subaperture images correspond to (s,t) coordinates in the 4D light field function, $L(u,v,s,t)$.

O_1 is defined as an object point optically conjugate to the image plane, while O_n is an arbitrary point at a different depth than O_1. The depths of these object points are denoted as Z_1 and Z_n, respectively. Their images by the ith subaperture onto its corresponding subaperture image are denoted as $I_{1,i}$ and $I_{n,i}$, respectively. For instance, rays from O_1 and O_n going through the center subaperture A_0 highlighted in red are imaged onto the overlapping points $I_{1,0}$ and $I_{n,0}$, respectively. The ray from O_1 going through the subaperture A_1 highlighted in blue is imaged onto $I_{1,1}$, which is also overlapping the previous two points according to image formation of the conjugate object. The ray from O_n going through A_1 is imaged onto $I_{n,1}$. The pixel displacement between $I_{n,0}$ and $I_{n,1}$ for the point O_n is defined as the disparity of the subaperture images captured by subaperture A_0 and A_1. To generalize, the image disparity of the point O_n between the ith and j^{th} subaperture images, denoted as $d_{n,i,j}$, is defined as the pixel displacements between the subaperture image points $I_{n,i}$ and $I_{n,j}$.

For simplicity, let us consider a pair of subapertures A_i and A_j along the t-axis, where A_i is the lower subaperture and A_j is the higher subaperture. The resulted image disparity for point O_n is also along the t-axis, and is calculated as $d_{n,i,j} = t_{n,i} - t_{n,j}$. The disparity $d_{n,i,j}$

is positive when $Z_n > Z_l$, is zero when $Z_n = Z_l$, and is negative when $Z_n < Z_l$. By ray tracing minimally through two subapertures, the depth of O_n can be reconstructed by:

$$Z_n = \left(-\frac{d_{n,i,j}/VB_{i,j} + 1}{Z'} + \frac{1}{f} \right)^{-1}.$$

(5.1)

The depth reconstruction in Eq. (5.1) only uses disparity information along the t-axis. For depth reconstruction from subapertures and disparity along the orthogonal axis, the process is analogous.

5.1.3 Calibration from Center and Peripheral Subaperture Systems

To obtain the calibration parameters, the first step is to utilize the chief rays of the fields going through the center subaperture in Fig. 5.2(a). This data is treated as light passing through a conventional pinhole camera model at A_0 in Fig. 5.3. The image projection process is described by:

$$sI_{n,0} = K[R \ T]O_n,$$

(5.2)

where s is a scale factor, $I_{n,0}$ is the 2D image point, K is the camera intrinsic projection matrix, $[R \ T]$ are the rotation and translation extrinsic parameters to relate the world coordinate system to the camera coordinate system, and O_n is the 3D object point defined in the world coordinates. Solving for K, R, and T is achieved using conventional monocular camera calibration [31]. The x and y focal lengths, f_x and f_y, of the pinhole camera are extracted from K to determine the distance from the pinhole to the sensor along the normal connecting them, or Z' in Fig. 5.3. $Z' = f_x$ for disparity analysis along the s-axis and $Z' = f_y$ for disparity analysis along the t-axis.

The calibrated center subaperture system (CCSS) is then utilized to calculate the distance from the representative pinhole model to the conjugate object plane, or Z_l in Fig.

5.3. A flat checkerboard object is placed both parallel to the sensor and at the best object focus position based on the defocus cues from the finite sized A_0. To assess parallelism, the checkerboard is imaged by the CCSS, and given the known checkerboard dimensions, its extrinsic parameters in relation to the calibrated pinhole model are calculated in closed form [37]. R is assessed to determine if the checkerboard requires any alignment adjustments to achieve parallelism. After alignment, the extrinsic parameters are calculated once more. To determine Z_l, the image point corresponding to the optical axis of A_0 is found from K and converted to camera coordinates using Eq. (5.2).

Next, the determined parameters Z_l and Z' are used to calculate focal length f. Because the peripheral subaperture samples were physically acquired on the same plane as the center subaperture, it is assumed that their respective pinhole models are also located at the same (u,v) plane as the center subaperture pinhole model in Fig. 5.3. The peripheral pinhole models at A_i must also incorporate ray bending such that the location of the peripheral subaperture images along the image plane is properly predicted. The combined set of pinhole models validates and forms the virtual main lens with focal length f, which is solved for by using the thin lens equation:

$$\frac{1}{f} = \frac{1}{Z_1} + \frac{1}{Z'}.$$
(5.3)

Finally, the distance between two subapertures is determined by $VB_{i,j} = \left| VB_{0,i} - VB_{0,j} \right|$, where the total distance is split into two segments each defined by A_0 and A_i as shown in Fig. 5.3. To define each segment, we consider two calibration checkerboards located at the depths of Z_l and Z_n and two respective feature points O_1 and O_n, one on each of the checkerboards. The camera coordinates of these two feature points are determined by

applying the CCSS to the corresponding image points from A_0. Then we illustrate in Fig. 5.3 that for the same two feature points, the pixel separation, $b_{1,n,i}$, between the corresponding image points from the ith subaperture is a function of $VB_{0,i}$ and the distance between O_1 and O_n. This relationship is determined by combining two equations obtained through ray tracing from O_1 and O_n through A_i to the image plane and algebraically solving for $VB_{0,i}$:

$$VB_{0,i} = -\frac{\dfrac{b_{1,n,i}}{Z'} + \dfrac{Y_n}{Z_n} - \dfrac{Y_1}{Z_1}}{-\dfrac{1}{Z_n} + \dfrac{1}{Z_1}}, \tag{5.4}$$

where $b_{1,n,i} = \pm(t_{n,i} - t_{1,i})$ is the amount of separation along the t-axis between the two image points. For sign convention, if Z_n is larger than Z_1, the positive sign is used and vice versa. Y_1 and Y_n are the y-camera coordinate of the feature points. For subaperture sampling distance along the orthogonal axis, the process is analogous.

This calibration method achieves the highest depth mapping accuracy for object depths near Z_1 and Z_n because these are the reference depths used for determining the calibration parameters. Z_1 is set by the optical conjugates of the camera while Z_n can be chosen based on the application to maximize depth mapping confidence at a second region. Since Eqs. (5.1) and (5.4) were defined based on the assumption that the entrance pupil is at the virtual lens, the amount of depth mapping error for objects away from Z_1 and Z_n is then dependent on the amount of separation between the entrance pupil and virtual thin lens that would more precisely model the real system.

The theory in the above sections assume the real system performs according to ideal, first-order imaging. The following subsections address how real factors impacting the light

field are accounted for during calibration.

5.1.4 Vignetting

To maximize the usability of the captured light field information, vignetting is mitigated. A flat Lambertian surface is used to fill the full field of view of the camera. The image should ideally have uniform irradiance, so any vignetting can be quantified and calibrated. Without vignetting, here the light field would be $L(u,v,s,t) = c$, where L is the radiance and c is a constant due to the Lambertian property. With vignetting, $L_v(u,v,s,t) \leq c$. In a real system, we can assume the center subaperture A_0 image has the minimum amount of vignetting, where $L(0,0,s,t) = c$ for field points with equal to or less than half vignetting because the chief rays are undisturbed. Using the A_0 image as the reference, vignetting $V(u,v,s,t)$ for each peripheral subaperture image is quantified by:

$$V(u,v,s,t) = L_v(u,v,s,t) / L(0,0,s,t) . \tag{5.5}$$

For subsequently recorded light fields, $L_v(u,v,s,t)$ is divided by $V(u,v,s,t)$ to recover the non-vignetted $L(u,v,s,t)$. For field points outside of the half vignetted field of view, vignetting is only corrected up to the amount of vignetting in the A_0 image.

5.1.5 Transverse Ray Aberrations

Distortion from the main lens is removed first in preparation for correction of transverse ray aberrations. Because distortion is characterized by the intersections of the chief rays at the image plane, the amount of main lens distortion can be observed in a raw center subaperture A_0 image. Since the A_0 system is rotationally symmetric, its distortion can be represented by a radial polynomial model:

$$\begin{aligned} s_d &= s + s[k_1(s^2 + t^2) + k_2(s^2 + t^2)^2] \\ t_d &= t + t[k_1(s^2 + t^2) + k_2(s^2 + t^2)^2] \end{aligned} . \tag{5.6}$$

The distortion coefficients k_1 and k_2 are determined simultaneously during the calculation [37] of the intrinsic and extrinsic parameters of the CCSS from Section 2.C. Eq. (6) is then used to convert $L(0,0,s_d,t_d)$ to the undistorted A_0 image $L(0,0,s,t)$. During transverse ray aberration calibration, the peripheral subaperture images will be referenced to the undistorted chief rays in $L(0,0,s,t)$, resulting in the removal of distortion from the entire captured light field.
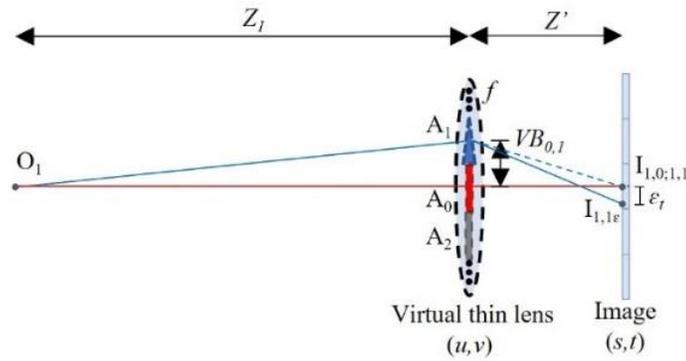
Fig. 5.4 Transverse ray error illustrated by replacing a peripheral subaperture with a thin prism to represent additional ray bending [35]**.**

In the presence of transverse ray aberration from the main lens, the light rays from the peripheral subapertures will intersect the image plane at different locations according to the transverse ray errors $\varepsilon_s(u,v,X,Y)$ and $\varepsilon_t(u,v,X,Y)$. ε_s and ε_t vary with subaperture location (u,v) and with object point location (X,Y) in the object plane conjugate to the image plane. Fig. 5.4 illustrates the error by replacing A_1 with a thin prism to represent additional ray bending from transverse ray aberration. The dashed blue ray is from ideal ray bending while the solid blue ray is from real ray bending. For a conjugate object point O_1, ε_s and ε_t are defined as the distance from the chief ray intersection or center subaperture image point $I_{1,0}$ to the corresponding peripheral subaperture image point $I_{1,i\varepsilon}$. This is illustrated by $I_{1,1\varepsilon}$ in Fig. 5.4. The ideal subaperture image point $I_{1,i}$ can be recovered by digitally moving $I_{1,i\varepsilon}$

by -ε_s and -ε_t. Since ε_s and ε_t are now defined at the image plane, they will be reparameterized as $\varepsilon_s(u,v,s_\varepsilon,t_\varepsilon)$ and $\varepsilon_t(u,v,s_\varepsilon,t_\varepsilon)$, where $(s_\varepsilon,t_\varepsilon)$ is transverse ray aberrated image space coordinates.

Using a checkerboard at the object plane of O$_1$, the errors can be measured for each corner point to build $\varepsilon_s(u,v,s_\varepsilon,t_\varepsilon)$ and $\varepsilon_t(u,v,s_\varepsilon,t_\varepsilon)$ across the full field of view. The measurement is accomplished by comparing each peripheral subaperture image $L(u,v,s_\varepsilon,t_\varepsilon)$ to the center subaperture image $L(0,0,s,t)$ and quantifying the ε_s and ε_t of corresponding subaperture image points. For peripheral subaperture image points without a corner feature, ε_s and ε_t are linearly interpolated from the nearest four corners. The aberrated light rays can then be digitally bent in post-processing by warping the peripheral subaperture images according to ε_s and ε_t to recover $L(u,v,s,t)$. This digital ray bending process is analogous to aberration correction in traditional optical lens design, where lens design software optimizes and minimizes transverse ray error by converging all light rays to their respective chief ray at the image plane.

For the calculation of Z_1 in Chapter 5.1.3, the checkerboard is supposed to be placed exactly at the real conjugate object plane. In practice, it may be slightly displaced, so transverse ray error calibration is also used to account for this. If the image plane location remains the same, the displacement will introduce additional transverse ray error between corresponding subaperture image points due to defocus. This error is added to the errors from the other aberrations and will be corrected simultaneously. In other words, a point O$_1$ slightly displaced from distance Z_1 will digitally focus at Z' after transverse ray error calibration. As a result of the additional digital ray bending, the virtual lens focal length f is slightly adjusted according to Eq. (5.3), where Z_1 is updated based on the displacement

of O_1. Therefore, the same placement of the checkerboard is used for transverse ray error calibration and calculation of the effective Z_1.

The monochrome aberrations have been addressed, and transverse ray error from chromatic aberrations can be dealt with in similar fashion. In this case, all subaperture images from the red, green, and blue channels separated by the Bayer filter would be compared to the green center subaperture image and warped accordingly. This type of color correction has been analogously demonstrated in [12].

5.1.6 Pupil Aberrations
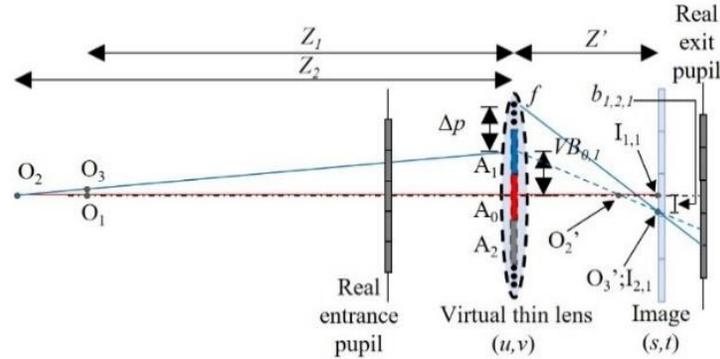
Fig. 5.5 Ray trace in the presence of pupil aberration and zero transverse ray aberration [35]

Although the transverse ray aberrations for the virtual main lens are corrected, pupil aberrations may still exist and impact the captured light field, as illustrated in Fig. 5.5. Pupil aberration is analogous to transverse ray aberration, except that deformation from the entrance to exit pupil is observed instead of deformation from the object to image plane. Points O_1 and O_3 are on the object plane conjugate to the image plane. A real ray from object point O_2 passes through the point where O_3 would be and then through a section of the real entrance pupil. It is then bent by the real main lens, which results in a translation Δp of the transmitted ray due to pupil aberration as shown at A_1 of the virtual main lens.

Because there is no transverse ray aberration here and the ray appears to have originated from O_3, the pupil aberrated ray intersects the conjugate image point location O_3' or the same $I_{2,1}$ from Fig. 5.3. The aberration-free ray indicated by the dashed blue line also intersects O_3'. These two rays are extended to the real exit pupil to further illustrate their differences due to pupil aberration. The aberration-free ray passes through the image point O_2' conjugate to O_2 while the pupil aberrated ray does not. This indicates that digital refocusing or ray tracing to the correct image point or depth will be inaccurate with pupil aberrated light field data. Meanwhile, a ray from O_1 going through A_1 will still land at $I_{1,1}$.

After transverse ray aberration correction from Chapter 5.1.5, pupil aberration is calibrated directly during the calculation of $VB_{0,i}$ from Eq. (5.4). Figure 5 shows that the separation $b_{1,2,1}$ is unaffected in the presence of pupil aberration without transverse ray aberration. Based on this property, the first-order, aberration-free ray can be restored by calculating $VB_{0,i}$ as normal. Since pupil aberration is radially dependent, a light field uniformly sampled at the exit pupil becomes nonuniformly sampled at the entrance pupil and virtual thin main lens. Thus, $VB_{0,i}$ must be calculated for each subaperture A_i to restore the aberration-free virtual main lens.

5.2   Calibration Procedure for Aberrated Light Field Cameras
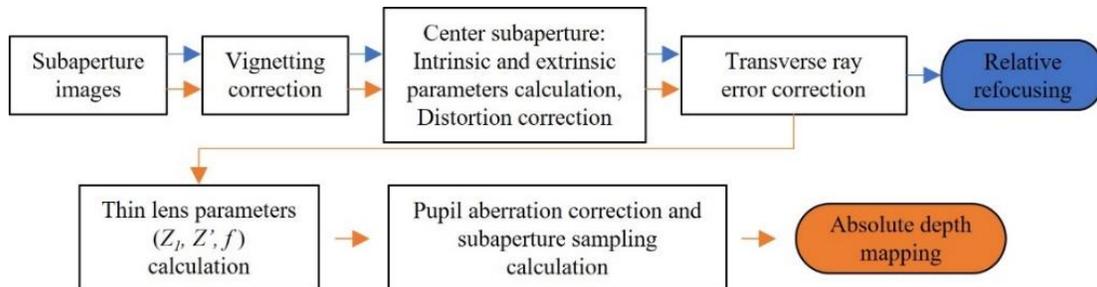
Fig. 5.6 Flowchart of aberrated light field camera calibration for recovering relative digital refocusing capability and absolute depth mapping capability [35]

In practice, aberrated light field camera calibration follows the flowchart in Fig. 5.6. If vignetting is observed in the camera's captured light field data, vignetting correction is first applied. Subaperture images of a flat Lambertian surface filling the full field of view are acquired to calculate $V(u,v,s,t)$ according to Eq. (5.5), which is then used to unvignette subsequently recorded light fields by division.

Then, center subaperture A_0 is modeled as a pinhole camera via monocular camera calibration. The CCSS is generated by capturing and processing a sufficient set of subaperture images from A_0, where each image is of an arbitrarily oriented checkerboard. For subsequently recorded subaperture images from A_0, distortion correction is performed using the CCSS's estimated distortion coefficients and Eq. (5.6).

Next, transverse ray error correction is performed. A checkerboard is placed both parallel to the sensor and at the best object focus position Z_1 based on the defocus cues from the finite sized A_0. Using the CCSS, parallelism is achieved by alignment of the checkerboard until the calculated extrinsic rotation matrix R is equal to the identity matrix. Aberrated peripheral subaperture images $L(u,v,s_\varepsilon,t_\varepsilon)$ and undistorted center subaperture image $L(0,0,s,t)$ of the checkerboard are then acquired. They are compared to quantify $\varepsilon_s(u,v,s_\varepsilon,t_\varepsilon)$ and $\varepsilon_t(u,v,s_\varepsilon,t_\varepsilon)$, which are used to recover transverse ray error corrected light fields $L(u,v,s,t)$ by digitally warping subsequently recorded $L(u,v,s_\varepsilon,t_\varepsilon)$.

In applications where only relative refocusing is needed, the calibration process can end here. If pupil aberration is absent or minimal, refocusing capability is recovered after transverse ray error correction. If pupil aberration is significant, refocused images at depths away from the image plane at Z' may have lower image quality, and the relative depth information may be slightly altered. To avoid this and recover absolute depth mapping

capability, the full calibration process is required.

Continuing with the calibration, the thin lens model parameters are calculated. First, Z' is determined by the pinhole focal lengths from the intrinsic matrix K of the CCSS. Second, Z_1 is determined using the same undistorted center subaperture image $L(0,0,s,t)$ of the checkerboard from transverse ray error correction. The image point corresponding to the optical axis of A_0 is found from K and converted to camera coordinates $(0,0,Z_1)$ using Eq. (5.2). Third, virtual main lens focal length f is then calculated using Eq. (5.3).

Finally, pupil aberration correction and subaperture sampling calculation are performed. The checkerboard is translated away from Z_1 to Z_2. The checkerboard parallelism to sensor is checked again by observing R. The translation must be large enough such that the separation $b_{1,2,i}$ is detectable by the sampling of the sensor. Since distortion and transverse ray error have already been quantified in a lookup table, all subaperture images for this new checkerboard position are then acquired and warped accordingly. Object points O_1 and O_2 are found by selecting two corner image points, $I_{1,0}$ and $I_{2,0}$, where each one is closest to the optical axis of A_0 in the two A_0 images of the checkerboard at Z_1 and Z_2. By selecting the two object points in the center of the field of view and close in (X,Y), depth mapping errors can be minimized in that region. Arbitrary points farther apart could be used, but this can introduce errors due to the entrance pupil at the thin lens assumption. Using the CCSS, the camera coordinates of O_1 and O_2 are found. Then the corresponding peripheral subaperture image points $I_{1,i}$ and $I_{2,i}$ are located. Separation $b_{1,2,i}$ is calculated from the pixel coordinates of these image points. $VB_{0,i}$ is then calculated for each peripheral subaperture A_i using Eq. (5.4).

With all the calibration parameters and lookup tables determined, any pair of raw

subaperture images from this aberrated LF camera can be corrected for vignetting, distortion, and transverse ray error. The corrected pair of subaperture images are then processed to generate a disparity $d_{n,i,j}$ map across the full field of view, which is converted to an absolute depth map using the known calibration parameters and Eq. (5.1).

5.3   Application and Experimental Results

Using the prism-based tri-aperture camera developed in Chapter 4, this section shows the full calibration process to recover relative refocusing and absolute depth mapping. It also simultaneously shows how the calibration could enable light field imaging in novel camera designs that capture light rays with artificially induced transverse ray aberration.

5.3.1   Prism-based Tri-aperture Camera

Fig. 5.7 Conceptual model of the prism-based tri-aperture laparoscopic objective for wide and stereo field of view image acquisition from Chapter 4. Note that the prism deflector introduces artificial transverse ray error, analogous to the thin prism in Fig. 5.4 [35]**.**

Fig. 5.7 reviews the conceptual model of the prism-based tri-aperture laparoscopic objective from Chapter 4. This system presents an alternative method of capturing portions of the light field. Its purpose is to provide the surgeon with stereo vision (SFOV), absolute depth mapping, and an additional wide field of view (WFOV) to enhance surgical awareness. Because digital refocusing is not necessary here, full light field capture isn't required. Instead, this system only acquires the light rays from the center A_0 and two

peripheral A_1 and A_2 subapertures at the stop of the objective. In the optical design of this system, a prism is introduced adjacent to the stop to simultaneously capture the peripheral stereo subaperture images S_1 and S_2. The overlapping center subaperture wide view image II_W is captured separately with a time or spatially multiplexed tri-aperture selector. Notice that the ray bending from the prism here is analogous to Fig. 5.4, where a thin prism was used to represent additional ray bending from transverse ray error. In other words, the real prism in this system artificially introduces transverse ray error to translate S_1 and S_2 onto opposite halves of the sensor. Compared to conventional stereo cameras with two individual lens systems, calibrating the distortion in these stereo views without additional reference information is not straightforward because there is no rotational symmetry. There's also additional distortion due to the finite thickness of the prism deflector. Instead, the rotationally symmetric II_W can then be utilized to remove the artificial transverse ray error, remove the distortion from the stereo views, and obtain the calibration parameters for absolute depth mapping according to our full light field calibration process in Fig. 5.6.

5.3.2 Vignetting
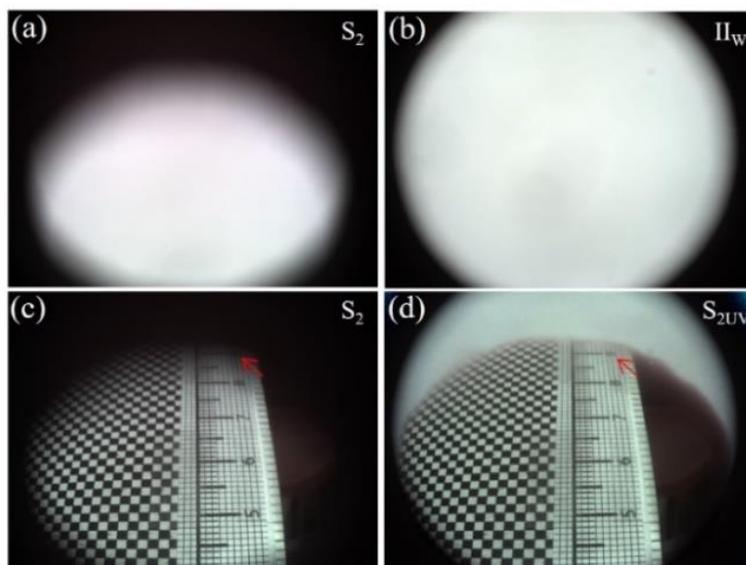
Fig. 5.8 Raw subaperture images (a) S_2 and (b) II_W of a flat Lambertian surface. S_2 of an object scene (c) before and (d) after vignetting correction [35].

First, vignetting correction was skipped in the full calibration because the prototype was designed with zero vignetting across the desired SFOV. However, vignetting correction can still be demonstrated for object points outside of the desired SFOV. Fig. 5.8(a) and 5.8(b) show raw subaperture images S_2 and II_W of a flat Lambertian surface, respectively. V was calculated by dividing S_2 by II_W, pixel by pixel, according to Eq. (5.5). Fig. 5.8(c) shows S_2 of an object scene, and the number 9 along the ruler is vignetted. V is ~.23 in this region of the image. To correct for this, S_2 is divided by V, pixel by pixel, to recover the unvignetted S_{2UV} in Fig. 5.8(d). The number 9 is now clearly visible after the removal of vignetting. For the rest of Chapter 5.3, raw subaperture images are used directly since vignetting correction is unnecessary for this prototype.

5.3.3 Center Subaperture Image Calibration
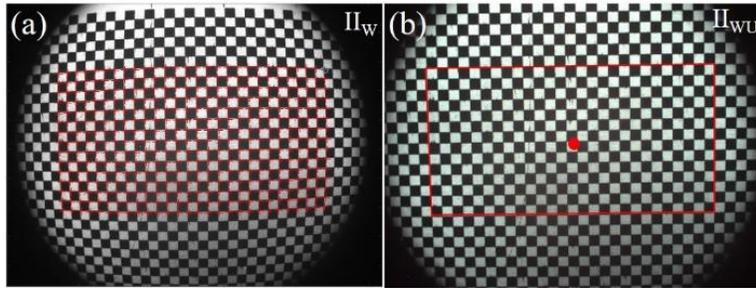
Fig. 5.9 Checkerboard placed both parallel and conjugate to the sensor. (a) Raw center subaperture image II_W and (b) undistorted center subaperture image II_{WU} [35].

Next, the CCSS was generated by capturing and processing 18 II_W images from A_0. Each image was from a random orientation of a checkerboard. The estimated intrinsic and extrinsic properties and the distortion coefficients of the pinhole model resulted in an overall mean reprojection error of 0.46 pixels. Next, the checkerboard was placed both parallel and conjugate to the sensor, as shown in II_W of Fig. 5.9(a). R is approximately

equal to the identity matrix, indicating parallelism:

$$R = \begin{bmatrix} 1.0000 & -0.0051 & 0.0048 \\ 0.0049 & 0.9985 & 0.0542 \\ -0.0050 & -0.0542 & 0.9985 \end{bmatrix}. \tag{5.7}$$

Next, II_W was undistorted as shown in II_{WU} of Fig. 5.9(b) using the CCSS. For instance, before correction the distortion is -7.3% at the top right checkerboard corner feature highlighted in red. After correction, the distortion is negligible as indicated by the overlapping red straight lines. II_{WU} represents the undistorted chief rays.

5.3.4 Transverse Ray Error Correction
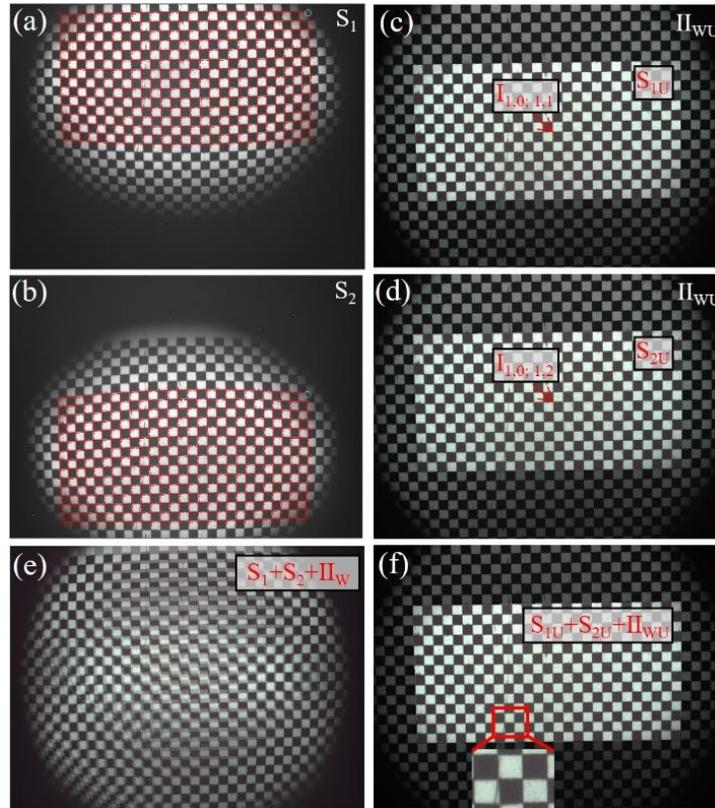
Fig. 5.10 Checkerboard placement unchanged from Fig. 5.9. Raw peripheral subaperture images (a) S_1 and (b) S_2. The image points highlighted in red in each subaperture image of Fig. 5.9 and Fig. 5.10 correspond to the same group of object points. The translation of the red grid between II_W, S_1, and S_2 illustrates the amount of artificially induced transverse ray error. Distortion and transverse ray error corrected images (c) S_{1U} and (d) S_{2U} using II_{WU} as the

reference chief rays. Full aperture images before (e) and after (f) transverse ray error correction [35]**.**

The checkerboard placement remained unchanged, and S_1 and S_2 were captured from A_1 and A_2, as shown in Fig. 5.10(a) and 5.10(b), respectively. The grid of image points highlighted in red in each subaperture image of Fig. 5.9 and Fig. 5.10 correspond to the same group of object points. The vertical translation of the red grid between II_W, S_1, and S_2 illustrates the amount of artificially induced transverse ray error. The grid in S_1 and S_2 is translated ~400 pixels up and down, respectively. The difference in distortion is also clearly observed. For II_W, it is rotationally symmetric, while for S_1 and S_2, it is bilaterally symmetric. S_1 and S_2 were compared to II_{WU} to quantify ε_s and ε_t. Then, S_1 and S_2 were digitally warped and translated accordingly to simultaneously remove their distortion and induced transverse ray error. Since the checkerboard for these images is at the conjugate object depth Z_1, corresponding subaperture image points should be overlapping. This is illustrated by the transverse ray error corrected images S_{1U} and S_{2U} of Fig. 5.10(c) and 5.10(d), respectively. II_{WU} is in the background while S_{1U} and S_{2U} overlap the corresponding region, indicating that the peripheral subaperture rays now focus together with their corresponding chief rays after transverse ray error correction. The amount of residual transverse ray error across the full field of view after correction is within ~2 pixels, depending on the accuracy of the feature recognition algorithm used to pinpoint the checkerboard corners and the interpolation of transverse ray error for object points in between checkerboard corners. For example, the zoomed view in Fig. 5.10(f) shows one of the squares was slightly miswarped.

For evaluation, image contrast of full aperture images can be quantified before and after transverse ray error correction. As illustrated in Fig. 5.7 and digitally generated in Fig.

5.10(e), the pre-corrected full aperture image is when S_1, S_2, and IIw are averaged together at the image plane. Here, the contrast at the checkerboard frequency is effectively zero because the artifacts due to the aberration result in aliasing. The post-corrected full aperture image in Fig. 5.10(f) looks equivalent to its subaperture image components, as expected. Full aperture imaging has been restored with an image contrast around 49%. Individually, subaperture images S_{1U}, S_{2U}, and II_{WU} have image contrast values of 46%, 46%, and 59%, respectively. This indicates that the full aperture image is limited by the image quality of its individual subaperture images.

5.3.5   Calibration of Thin Lens Parameters

The thin lens model parameters were calculated next. Since the disparity in this prototype only occurs along the t-axis, $Z' = f_y$. From the intrinsic matrix K of the CCSS, $Z' = 3413.4$ pixels or 7.506 mm based on the sensor's 2.2 μm pixels. The effective focal length of the real lens design is 7 mm, so Z' is appropriate for the object distance in Fig. 5.9(a). To determine Z_1, the image point corresponding to the A_0 optical axis was found from K, as shown by the red point in Fig. 5.9(b). Using the CCSS, the camera coordinate of this image point was calculated as $(0,0,Z_1) = (0,0,71.253)$ mm. Out of the entire field of view, the maximum and minimum object depths were also calculated to be 72.214 mm and 69.935 mm, respectively, indicating a slight tilt of the checkerboard. Z_1 is in the middle of this range and approximately represents all object points across the field of view, so some of this tilt error may be carried over to the depth mapping results. Then, from Eq. (5.3), $f = 6.79$ mm.

5.3.6 Subaperture Sampling
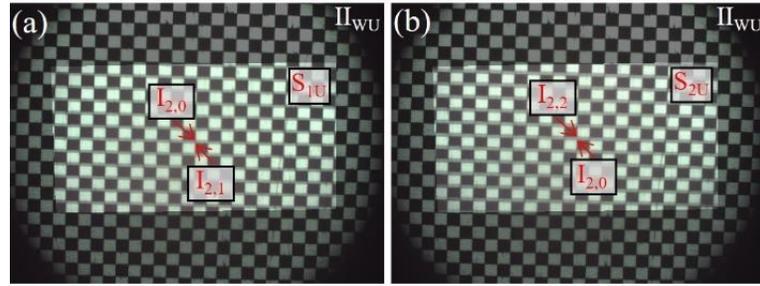
Fig. 5.11 Checkerboard moved 1 cm toward camera and new subaperture images captured. Corrected images (a) S_{1U} and (b) S_{2U} generated using transverse ray error lookup tables. These images were fused with II_{WU}. In comparison, disparity is observed in opposite directions relative to II_{WU} [35].

To perform the pupil aberration correction and subaperture sampling calculation, the checkerboard was placed ~1 cm in front of the conjugate object depth to minimize depth mapping error around this range. The three raw subaperture images were acquired for this new object depth and warped accordingly for distortion and transverse ray error correction. Fig. 5.11(a) and 5.11(b) show the corrected images S_{1U} and S_{2U}, respectively, fused together with II_{WU}. Comparing the two figures, disparity due to defocus is observed in opposite directions relative to II_{WU}, as expected since A_1 and A_2 are on opposite sides of A_0. In each figure, the magnitude of disparity is constant throughout the SFOV, indicating that the checkerboard was simply translated along the central optical axis.

Corner image points $I_{1,0}$ and $I_{2,0}$ were selected in Fig. 5.10 and Fig. 5.11. Using the CCSS, the camera coordinates of O_1 and O_2 are (0.3906, -0.4499, 71.2217) mm and (-0.2556, -0.1793, 58.7407) mm, respectively. Corresponding points $I_{1,i}$ and $I_{2,i}$ were located in Fig. 5.10 and Fig. 5.11. From the image coordinates and pixel size, separations $b_{1,2,i}$ were calculated. $b_{1,2,1}$ is -0.07524 mm, and $VB_{0,1}$ was then calculated as 2.2643 mm. $b_{1,2,2}$ is 0.0297 mm, and $VB_{0,2}$ was then calculated as -2.42 mm. Thus, the total $VB_{1,2}$ between A_1 and A_2 is 4.6843 mm. This value is within the expected range because the prototype was

designed for a real baseline separation between A_1 and A_2 of 5.02 mm. $VB_{0,2}$ is slightly

larger than $VB_{0,1}$ because the two apertures were slid into a slot of the optomechanical

housing, where mechanical tolerance resulted in a decenter of the stereo apertures.

Assessment and quantification of the pupil aberration before and after correction is difficult

here because the pupils are not imaged in this calibration. Instead, we must observe the

quality of depth estimation results shown in the next section. A summary of the calibrated

system parameters is shown in Table 5-1.

Table 5-1 Summary of calibrated system parameters for depth mapping

Parameter	Value
f	6.79 mm
Z'	7.51 mm
$VB_{1,2}$	4.68 mm
Z_1	71.2 mm
Z_2	58.7 mm

5.3.7 Calibration Results
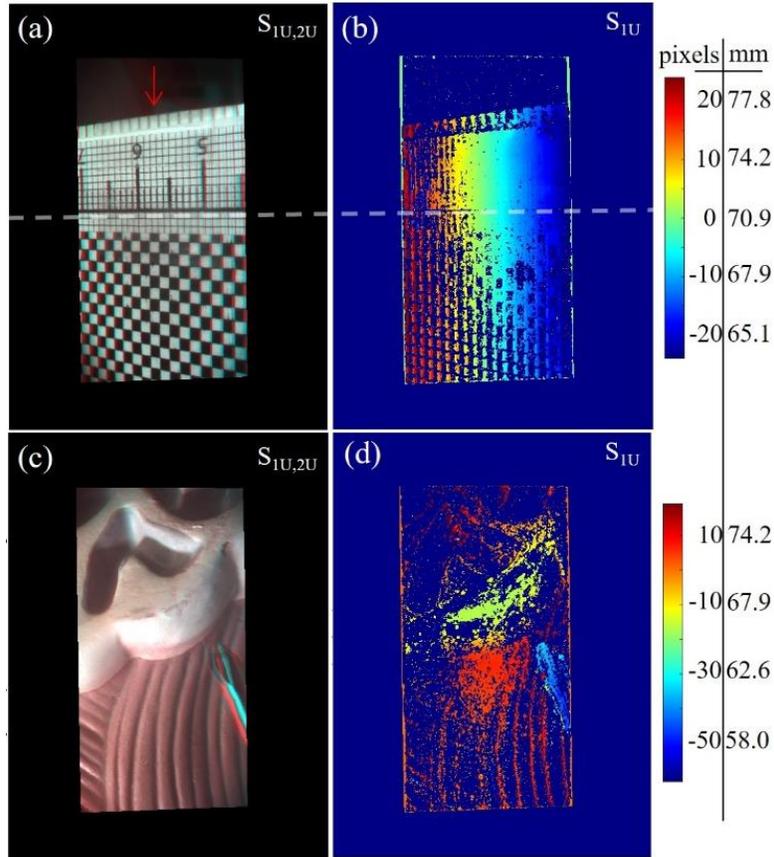
Fig. 5.12 Checkerboard was tilted to produce a linear change in depth. (a) S_{1U} and S_{2U} overlaid together as a red-cyan anaglyph and rotated counterclockwise 90° for viewing with 3D glasses. (b) S_{1U} converted into a disparity $d_{n,2,1}$ map in pixel units and into an absolute depth Z_n map in mm units. Process was repeated for a 3D bladder model (c) and (d) [35].

Once all the calibration parameters were determined, any subsequently captured pair of S_1 and S_2 images could be processed for absolute depth mapping. The checkerboard with a ruler placed on top was tilted away from the camera to produce a linear change in depth along the vertical axis, as shown in Fig. 5.8(c). S_1 and S_2 were captured for this new object scene, warped into S_{1U} and S_{2U} based on the quantified ε_s and ε_t lookup tables, overlaid together as a red-cyan anaglyph, and rotated counterclockwise 90° such that the disparity is along the horizontal axis and can be viewed with 3D glasses, as shown in Fig. 5.12(a). Then, S_{1U} was chosen as the reference image, and for each image point, the disparity $d_{n,2,1}$

in pixel units to the corresponding image point in S_{2U} was calculated as shown in Fig. 5.12(b). Using Eq. (5.1) and the previously determined calibration parameters, $d_{n,2,1}$ was converted to absolute depth Z_n in mm units. This process was repeated for a second object scene, a 3D bladder model, in Fig. 5.12(c) and 5.12(d).

As discussed in Fig. 5.6, relative refocusing is recovered after transverse ray error correction. In Fig. 5.12(a), the focus position is at 0 pixel disparity or around the 5.75 tick mark of the ruler, as indicated by the red arrow. By translating S_{1U} and S_{2U} properly and fusing them together with II_{WU}, the focus position can be digitally changed. This is demonstrated in Fig. 5.13(a) and 5.13(b), where the focus position has been changed to the 4.5 and 7 tick mark, respectively. The out of focus blurring increases further away from the focus position as expected. The defocus only occurs along the vertical axis according to the captured subapertures, and aliasing is noticeable due to the three minimum subapertures collected. Nonetheless, relative refocusing recovery is demonstrated and is further illustrated in [25].
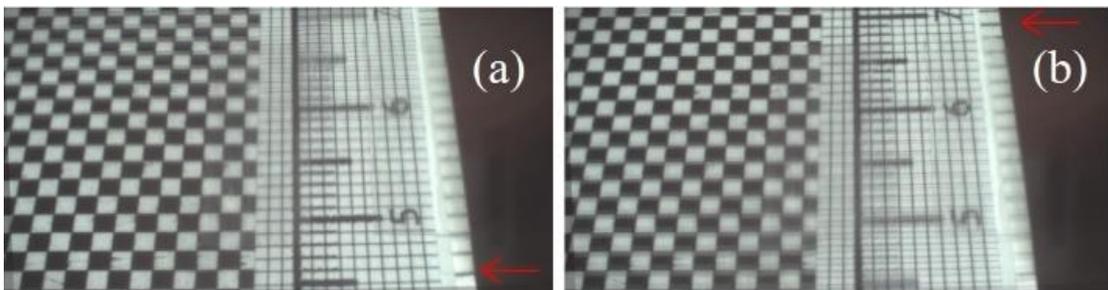
Fig. 5.13 Relative refocusing to (a) near and (b) far depths after transverse ray error correction [35]**.**

Depth estimation from the CCSS and the calibrated stereo apertures were compared to assess the proposed calibration. Since the entire object field was lying on a checkerboard plane, any object point's depth can be calculated with the CCSS. The depth of object points along the transparent white dashed line in Fig. 5.12(b) were calculated with the CCSS and

compared to the results from the triangulation method of A_1 and A_2, as shown in Fig. 5.14. The triangulation method estimates the overall shape of the object field correctly, indicating that absolute depth mapping capability is recovered after aberration calibration. Linear regression, excluding the one outlier in the data due to noise, calculated a slope less than 1. This means the triangulation method is overestimating and underestimating at the opposite ends of the depth range and has an overall tilt error. This is likely due to either a slight tilt in the calibration checkerboard, as determined by how close Eq. (5.7) is to the identity matrix, or the entrance pupil at the thin lens assumption. Nonetheless, this tilt error could be calibrated out as well. Without doing so, the maximum difference in depth mapping estimation is ~0.75 mm in this example.
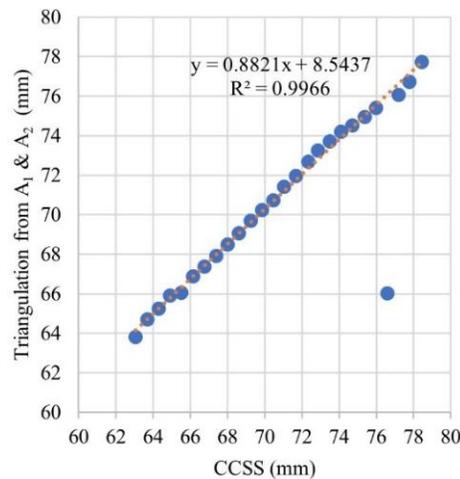
Fig. 5.14 Comparison of depth estimation between the CCSS and the triangulation from A_1 and A_2 indicates that absolute depth mapping capability is recovered after aberration calibration [35]**.**

5.4   Summary

In this chapter, we proposed a virtual thin lens model and calibration to recover the first order performance of an as-built light field or pupil sampling camera. The calibration utilizes center subaperture images to define the imaging and distortion parameters of the model. It then compares center and peripheral subaperture images to quantify vignetting

and transverse ray errors. Finally, it corrects pupil aberration by calculating the ideal peripheral subaperture coordinates based on object to image point relationships. After calibration, relative refocusing or absolute depth mapping capability is restored. This was demonstrated in the prism-based tri-aperture camera, indicating that calibration can enable the functionality of novel light field camera designs with artificially introduced transverse ray error, such as the TAMLO.

6   CONCLUSION AND FUTURE WORK

6.1   Conclusion

Rigid endoscopic technology has already come a long way to produce high quality 2D imaging for MIS. However, the absence of both binocular vision and wide FOV imaging in a rigid 2D endoscope prevents surgeons from utilizing the depth perception and the wide foveated FOV capabilities of human vision. In this dissertation, a novel optical design solution was explored to capture the necessary imaging data for restoring human vision in MIS. By using advancements in LF-based laparoscopes, a monocular design solution was found for efficient image capture. The programmable aperture light field laparoscope revealed the necessary endoscopic design constraints for quantitative 3D depth mapping. Then, the prism-based tri-aperture laparoscope was invented to capture both high quality stereo images and a WFOV with 2x the FOV of the binocular FOV. Finally, a generalized LF camera calibration was created to enable the functionality of these novel laparoscopes. The optical engineering research developed in this dissertation provides supporting evidence that binocular vision and WFOV imaging can be simultaneously provided to the surgeon, and that further investigation into these types of systems is worthwhile.

6.2   Future Work

Future work on the light field-based laparoscopes may include the following:

(1) The stereo apertures in the TAMLO collect larger ray angles from the object field than typical endoscopes. An investigation is required to determine the design difficulty of the rigid rod lens relay groups with higher numerical aperture corresponding to the TAMLO.

(2) The TAMLO design can be applied to chip-on-tip systems, as demonstrated in the TAMLO prototype. For chip-on-tip systems, the telecentricity constraint is no longer

necessary, which makes the lens design easier. This particular application seems promising.

(3) The TAMLO's proper tri-aperture selector hardware needs to be implemented and configured in software for automated stereoscopic and WFOV image capture.

(4) The LF calibration code needs to be revised for automation and speed. Also, additional calibration can be added to improve depth estimation accuracy.

(5) A bio-inspired multiband sensor could be implemented into the TAMLO for simultaneous 3D stereoscopic, WFOV, and NIR fluorescence imaging.

REFERENCES

1. M. D. Rutter, M. Brookes, T. J. Lee, P. Rogers, and L. Sharp, "Impact of the COVID-19 pandemic on UK endoscopic activity and cancer detection: A National Endoscopy Database Analysis," Gut **70**, 537–543 (2021).

2. H. Yamashita, H. Aoki, K. Tanioka, T. Mori, and T. Chiba, "Ultra-high definition (8K UHD) endoscope: our first clinical success," Springerplus **5**, (2016).

3. R. Sinha, M. Sundaram, S. Raje, G. Rao, M. Sinha, and R. Sinha, "3D laparoscopy: Technique and initial experience in 451 cases," Gynecol Surg **10**, 123–128 (2013).

4. B. Alaraimi, W. El Bakbak, S. Sarker, S. Makkiyah, A. Al-Marzouq, R. Goriparthi, A. Bouhelal, V. Quan, and B. Patel, "A randomized prospective study comparing acquisition of laparoscopic skills in three-dimensional (3D) vs. two-dimensional (2D) laparoscopy," World J Surg **38**, 2746–2752 (2014).

5. G. Currò, G. La Malfa, S. Lazzara, A. Caizzone, A. Fortugno, and G. Navarra, "Three-Dimensional Versus Two-Dimensional Laparoscopic Cholecystectomy: Is Surgeon Experience Relevant?," Journal of Laparoendoscopic & Advanced Surgical Techniques **25**, 566–570 (2015).

6. P. V. Tomazic, F. Sommer, A. Treccosti, H. R. Briner, and A. Leunig, "3D endoscopy shows enhanced anatomical details and depth perception vs 2D: a multicentre study," European Archives of Oto-Rhino-Laryngology **278**, 2321–2326 (2021).

7. P. Roulet, P. Konen, M. Villegas, S. Thibault, and P. Y. Garneau, "360° Endoscopy Using Panomorph Lens Technology," Endoscopic Microscopy V **7558**, 75580T (2010).

8. Y. Qin, H. Hua, and M. Nguyen, "Multiresolution foveated laparoscope with high resolvability," Opt Lett **38**, 2191–2193 (2013).

9. Y. Qin, H. Hua, and M. Nguyen, "Characterization and in-vivo evaluation of a multi-resolution foveated laparoscope for minimally invasive surgery," Biomed Opt Express **5**, 2548–2562 (2014).

10. Y. Qin and H. Hua, "Optical design and system engineering of a multiresolution foveated laparoscope," Appl Opt **55**, 3058–3068 (2016).

11. Y. Qin and H. Hua, "Continuously zoom imaging probe for the multi-resolution foveated laparoscope," Biomed Opt Express **7**, 1175–1182 (2016).

12. J. I. Katz, S. Lee, and H. Hua, "Improved multi-resolution foveated laparoscope with real-time digital transverse chromatic correction," Appl Opt **59**, G79–G91 (2020).

13. D. Kim, S. Chang, and H. Kwon, "Wide Field-of-View , High-Resolution Plastic Lens Design with Low F-Number for Disposable Endoscopy," Photonics **8**, 89 (2021).

14. O. Cogal, S. Member, and Y. Leblebici, "An Insect Eye Inspired Miniaturized Multi-Camera System for Endoscopic Imaging," IEEE Trans Biomed Circuits Syst **11**, 212–224 (2017).

15. S. Lee, M. Choi, E. Lee, K. Jung, J. Chang, and W. Kim, "Zoom lens design using liquid lens for laparoscope," Opt Express **21**, 1751–1761 (2013).

16. X. Dallaire and S. Thibault, "Design of a foveated wide-angle endoscopic lens," Optical Engineering **55**, 047106 (2016).

17. C.-H. Lin, L.-J. Hsiao, J.-T. Hsaio, and H. Y. Lin, "Front view and panoramic side view videoscope lens system design," Appl Opt **53**, H146 (2014).

18. M. Rubin, L. Lurie, K. Bose, and S. H. Kim, "Expanding the view of a standard colonoscope with the Third Eye® Panoramic™ cap," World J Gastroenterol **21**, 10683–10687 (2015).

19. J. Geng and J. Xie, "Review of 3-D endoscopic surface imaging techniques," IEEE Sens J **14**, 945–960 (2014).

20. D. T. Kim, C. H. Cheng, D. G. Liu, K. C. J. Liu, and W. S. W. Huang, "Designing a New Endoscope for Panoramic-View with Focus-Area 3D-Vision in Minimally Invasive Surgery," J Med Biol Eng **40**, 204–219 (2020).

21. N. Bedard, T. Shope, A. Hoberman, M. A. Haralam, N. Shaikh, J. Kovacevic, N. Balram, and I. Tosic, "Light field otoscope design for 3D in vivo imaging of the middle ear," Biomed Opt Express **8**, 260–272 (2017).

22. S. Zhu, P. Jin, R. Liang, and L. Gao, "Optical design and development of a snapshot light-field laryngoscope," Optical Engineering **57**, 023110 (2018).

23. J. Liu, D. Claus, T. Xu, T. Keßner, A. Herkommer, and W. Osten, "Light field endoscopy and its parametric description," Opt Lett **42**, 1804–1807 (2017).

24. R. Ng, "Digital light field photography," PhD. Thesis, Stanford University (2006).

25. E. Kwan, Y. Qin, and H. Hua, "High resolution, programmable aperture light field laparoscope for quantitative depth mapping," OSA Contin **3**, 194–203 (2020).

26. C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography," ACM Trans Graph **27**, 55 (2008).

27. C. Zuo, J. Sun, S. Feng, M. Zhang, and Q. Chen, "Programmable aperture microscopy: A computational method for multi-modal phase contrast and light field imaging," Opt Lasers Eng **80**, 24–31 (2016).

28. M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," Proceedings of the IEEE International Conference on Computer Vision 673–680 (2013).

29. E. Kwan and H. Hua, "Prism-based tri-aperture laparoscopic objective for multi-view acquisition," Opt Express **30**, 2836 (2022).

30. J. E. Greivenkamp, *Field Guide to Geometrical Optics* (SPIE, 2004).

31. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans Pattern Anal Mach Intell **22**, 1330–1334 (2000).

32. L. Mignard-Debise and I. Ihrke, "A Vignetting Model for Light Field Cameras With an Application to Light Field Microscopy," IEEE Trans Comput Imaging **5**, 585–595 (2019).

33. Q. Cui, S. Zhu, and L. Gao, "Developing an optical design pipeline for correcting lens aberrations and vignetting in light field cameras," Opt Express **28**, 33632 (2020).

34. M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH* (1996), pp. 1–8.

35. E. Kwan and H. Hua, "Calibration of transverse ray and pupil aberrations for light fieldcameras," Appl. Opt. **61**, 6974–6984 (2022).

36. T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," J Electron Imaging **19**, 021106 (2010).

37. J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," (2004).

APPENDIX A: High Resolution, Programmable Aperture Light Field Laparoscope for Quantitative Depth Mapping

Elliott Kwan, Yi Qin, and Hong Hua

High resolution, programmable aperture light field laparoscope for quantitative depth mapping

ELLIOTT KWAN, YI QIN, AND HONG HUA* ⓘ

3D visualization and imaging system laboratory, College of Optical Sciences, University of Arizona, 1630 E University Blvd., Tucson, AZ 85721, USA
**hhua@optics.arizona.edu*

Abstract: Recent applications have shown that light field imaging can be useful for developing uniaxial three-dimensional (3D) endoscopes. The immediate challenges in implementation are a tradeoff in lateral resolution and acquiring enough depth information in the physically limited environment of minimally invasive surgery. Here we propose using programmable aperture light field imaging in laparoscopy to capture 3D information without sacrificing the camera sensor's native, high spatial resolution. This hybrid design utilizes a programmable aperture to preserve the conventional laparoscope's functionality and, upon demand, to compute a depth map for surgical guidance. A working prototype is demonstrated.

1. Introduction

Conventional laparoscopic systems provide surgeons with a two-dimensional (2D) view of the operative field, which limits depth cues due to the lack of binocular vision and causes loss of accurate depth perception and challenges of eye-hand coordination [1]. On the other hand, 3D laparoscopes offering more accurate depth cues such as binocular vision have gained significant popularity, especially when integrated with robotic surgery. Studies have reported less fatigue and more accurate and faster surgical performances with 3D laparoscopy [2,3].

There exist several different methods for implementing 3D laparoscopy, including dual-sensor stereo, single-sensor stereo, single-sensor 3D imaging via structured light, and uniaxial 3D imaging [4]. Engineering these methods for 3D laparoscopes faces unique design limitations. Dual-sensor stereo systems that integrate a pair of imaging optics and sensors to acquire binocular images of the surgical field into the constrained, standardized endoscope housing are one of the popular means that have been successfully adopted in commercial systems. Compared to the 4k ultra-high definition image quality of state-of-the-art 2D laparoscopes, improving optical performance in these systems is still challenging due to the naturally constrained dimensions and effective numerical aperture (NA). Single-sensor stereo systems capture stereo images with a single sensor by means of split-channel optics and result in compromise of resolution. Single-sensor 3D imaging systems via structured light record distorted structured light to determine 3D surface profile, but at the cost of an extra projection path. Uniaxial 3D acquisition techniques that extract 3D depth information using monocular endoscopes with a single optics channel have one advantage of maintaining the similar form factor to the monocular endoscopes and several strategies such as time of flight measurement, shape from defocus or shading, and various active illumination methods have been investigated actively [4].

Recently, another uniaxial 3D acquisition method called light field (LF) imaging was applied to minimally invasive surgery, such as LF otoscope [5], laryngoscope [6], and endoscope [7]. Capturing the LF of a surgical field requires recording both the spatial and angular information of the light rays from a 3D object and thus enables an imaging system to digitally refocus post-image capture, extend the depth-of-field, and acquire depth information [8]. In addition, LF

capture can be implemented with a simple addition of a microlens array (MLA) to its original monocular imaging optics. The existing LF endoscopes [5–7], however, are subject to several major limitations. The most important limitation is their substantially reduced spatial and limited angular resolution due to tradeoffs between ray position and angular sampling. Furthermore, they are often dedicated 3D systems incapable of acquiring high-resolution 2D images, require splitting the imaging path, or are limited to acquiring depth information for a specific environment.

Here we propose a design of a high resolution, programmable aperture light field laparoscope (PALFL) and demonstrate its utility for quantitative depth mapping. By adopting a programmable aperture (PA) instead of an MLA for capturing light fields in a time-multiplexing fashion [9,10], the proposed laparoscope design is able to address the above-mentioned limitations of existing LF endoscopes. In this case, the spatial resolution of the acquired light fields is only subject to the limit of the image sensor and the undesirable tradeoff between spatial and angular resolution is removed.

2. Optical approach

Figure 1 illustrates the schematic layout of our proposed PALFL design, which consists of an objective lens, a 1:1 relay lens group, an eyepiece, a programmable aperture, a focusing lens, and a sensor. The objective lens with a focal length of f_{obj} images the entire field of view (FOV) of an object and forms intermediate image 1 (II1). The 1:1 relay lens is necessary for rigid laparoscopes to extend the insertion length of the imaging probe within the limit of the housing tube diameter and relay the image to outside of the patient's body at intermediate image 2 (II2). To fit the objective lens and relay lens within the standard 10 mm-diameter housing of laparoscopes, the objective lens is designed to be image-space telecentric with its entrance pupil (EP) placed at its front focal point while the relay lens group is designed to be double telecentric. The eyepiece with a focal length of f_{eye} projects the image toward optics infinity for direct viewing or further imaging. In the meantime, the eyepiece forms a conjugate image of the objective EP, labeled as "stop", at which the programmable aperture is placed. Opening a given region of the PA component allows the focusing lens, with a focal length of f_{fl}, to image different bundles of rays from the object onto the sensor.

Fig. 1. Conceptual model of the PALFL.

By selectively opening different sub-apertures (e.g. three instantaneous sub-apertures are highlighted by the Red, Green and Blue pixels in Fig. 1) sequentially, the sensor captures different light ray angles incident upon the EP from the same object point. As illustrated by the zoomed view at the sensor, depending on the depth of the object of interest, the rays through the different sub-apertures may be imaged at the same pixel when the object depth is optically conjugate to the senor or at different pixels when it is either nearer or further than the conjugate depth. Such disparity information recorded by the sub-aperture images is to be used for reconstructing the depth map of the object field or refocusing the image at different depth.

One significant advantage of a PALFL design over existing LF endoscopes using MLA is that spatial and angular resolutions of the captured LF images are only subject to the limits of the sensor resolution and the pitch of the PA, respectively, while existing LF endoscopes are subject to the tradeoffs between the spatial resolution of the images and the angular resolution of ray direction samples. Another worth-noting feature of a PALFL system is its hybrid capability. The system's instantaneous aperture can be switched between sub-aperture LF capture state and a normal capture state where a centered, regular-sized aperture is operated to capture a conventional 2D full-resolution, full FOV image that is the same as a conventional laparoscope. This capability provides a surgeon with the option, on demand, to receive guidance through the visualization of depth information.

Another interesting aspect of the PA approach is that the size and pattern of the sub-aperture can be customized based on what is needed. To match the throughput of existing LF endoscopes, sub-apertures can span multiple adjacent pixels in the PA while sensor pixels can be binned. In the case of insufficient illumination, the span can be further extended at the cost of depth-of-field or depth mapping range, and high angular resolution can still be maintained by allowing sequential sub-aperture regions to overlap. The drawback of a sequential capture is the cost of speed, but the ever-increasing frame rates of imaging sensors can well overcome this limitation. Multiplexed light field acquisition [9,10], which uses patterns spanning multiple regions of the PA per frame, can be implemented to increase signal-to-noise ratio and allow for faster frame rates.

3. Depth mapping resolution

A key aspect to the design of a PALFL system is to achieve adequate depth mapping resolution. This mainly depends on the maximal angular separation of the rays through the centers of the sub-apertures, which establishes the maximal baseline equivalent to a stereo system, and the minimally detectable ray separation of the imaging system. For the convenience of quantifying the depth resolution of different systems, we use the numerical aperture at the nominal working distance, NA_{WD}, in the object space to characterize the maximal angular separation of the sub-apertures, and we use the equivalent sensor spatial resolution in the object space, B_{obj}, of the system to quantify the minimally detectable ray separation. We assume that distinguishing the three separated rays in Fig. 1 and confidently detecting a depth offset from the sensor conjugate depth, L_{WD}, minimally requires a 2-pixel separation ($2B$ at the sensor or $2B_{obj}$ at L_{WD}) between the Red and Blue rays on the sensor. A higher depth resolution can be achieved by digitally interpolating pixel data and refining the location of rays that land in between two pixels, but this possibility is not demonstrated here. Using similar triangles with bases located at L_{WD} and the EP and a Taylor series expansion for simplification, the depth resolution of a PALFL design is derived:

$$d_{\pm} \approx \frac{B_{obj}}{NA_{WD}} \left[1 \pm \frac{2B_{obj}}{D_{EP}} \right], \tag{1}$$

where d_+ and d_- represent the absolute distances from the sensor conjugate depth, L_{WD}, to the closest resolvable depths away from and towards the EP, respectively, and D_{EP} is the EP diameter. Given the pixel resolution, B, of the sensor and first-order optics specifications, without considering the effects of diffraction and aberration, B_{obj} and NA_{WD} are defined as:

$$B_{obj} = \frac{(L_{WD} - f_{obj})f_{eye}}{f_{obj}f_{fl}}B, \quad NA_{WD} \approx \frac{D_{EP}/2}{L_{WD} - f_{obj}}. \tag{2}$$

Figure 2 plots the average depth resolution, d, of d_+ and d_- in relation to NA_{WD} for systems of different spatial resolutions in the object space. At a nominal working distance of 50 mm, the NA_{WD} of a standard monocular laparoscope is ~0.003 while the 5 mm baseline of a state-of-the-art stereo laparoscope (with a 12 mm diameter rod) by Intuitive Surgical produces

an equivalent NA_{WD} of ~0.05. The object-space spatial resolution here is quantified by the minimally discernable pair of line features per unit distance (lps/mm), equivalent to $1/(2B_{obj})$. The object-space resolution of a commercial laparoscope is 2-6 lps/mm and the diffraction limited resolution of the multi-resolution foveated laparoscope reported in [11] is ~12 lps/mm.
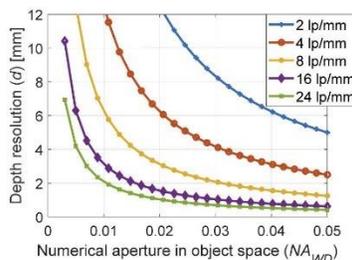

Fig. 2. Plot of achievable depth resolution in the laparoscopic environment for different NA_{WD} and $1/(2B_{obj})$

Figure 2 suggests that implementing a LF laparoscope using standard laparoscope optics, with a spatial resolution of 4 lps/mm and NA_{WD} less than 0.01, can yield a depth resolution of worse than 12 mm. The combination of NA_{WD} of 0.015 and resolution of 6 lps/mm provides a depth resolution of ~5.5 mm, which can be useful for surgeons to determine the proximity of their surgical tools, but inadequate for accurately visualizing anatomical structures. Achieving sub-mm depth resolution with light field method requires substantial improvements in both the object-space resolution and numerical aperture of standard 2D laparoscopes. On the other hand, achieving this resolution in a LF laparoscope with dimensions like a stereo laparoscope seems possible.

4. Prototype and experimental setup

Figure 3(a) shows the optical layout of a prototyped bench-top PALFL system for proof of concept. An f/2.5 objective lens with a focal length of 1.8 mm from an existing laparoscope developed in [11] was repurposed for this prototype. The diameter of this objective lens group is 5.7 mm, which is small enough to allow space for fiber illumination and lens housing to build a standard 10-mm diameter rigid laparoscope as demonstrated in [11]. The optical system inside the rigid laparoscopic tube in [11] also consists of several groups of relay optics to relay the image of the objective to the distal end of the tube for further imaging. As the objective and the relay were optimized and custom-made independently, they can be used separately without the other and different number of the relay optics can be added or removed without affecting the optical performance. When building the PALFL bench prototype, we removed the relay optics for simplicity and only used the objective along with a newly added eyepiece, a PA, and a focusing lens as the relay optics does not add or change the imaging function of the system. The objective lens was originally optimized for an L_{WD} of 120 mm, a D_{EP} of 0.8 mm, and lens diameters < 6 mm, resulting in an effective NA_{WD} of ~0.003. However, for this PALFL prototype the objective lens was used at an L_{WD} of 20 mm. Although this distance is short for laparoscopy, it yields an NA_{WD} as large as ~0.022 if the full EP is sampled and produces an NA_{WD} that is more comparable to that of stereo laparoscopes. The relay lenses were omitted to simplify the lens design and optical alignment of this prototype. A 10 mm focal length eyepiece built with stock lenses was optimized to meet sufficient performance over a 60° full FOV and expanded the 0.8 mm EP of the

objective lens to a 4.4 mm stop where a PA could be inserted. Note that the eyepiece diameter can be much larger than that of the objective and relay because it is outside of the patient's body. These modules were aligned to a commercial focusing lens with a focal length of 25 mm and 1/3" color CCD sensor (1.3 MP Dragonfly2 from Point Grey). The pixel resolution of the sensor is 1280×960, and the color pixel size is 3.75×3.75 μm^2. Using Eq. (2), we can estimate that the theoretical spatial resolvability of the system in the object space is ~33 lps/mm. Using Eq. (1), the depth resolution of the prototype can potentially reach ~0.69 mm if the sub-aperture images are sampled at the full aperture and the optics perform at its full resolution.
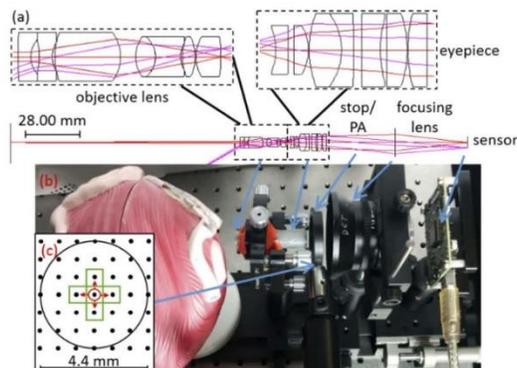

Fig. 3. (a) Optical layout of prototype, (b) construction of benchtop system, and (c) manual PA sampling scheme.

Figure 3(b) shows the prototype. The objective lens and eyepiece were assembled in a 3D printed opto-mechanical housing, as shown in the grey cylinder. Instead of using a digital PA, a physical iris mounted on a two-axis linear stage was employed. Figure 3(c) illustrates the angular sampling scheme bounded by the stop. The grid of black dots represents the locations that would be sampled sequentially by the pitch of the sub-apertures and determines the angular resolution. The iris, indicated by the red circle with arrows, moves to each sampling location. An illuminated bladder model object is placed near an L_{WD} of 20 mm, as shown in Fig. 3(b). On the image side, the sensor was adjusted to the new conjugate image position.

Since the preexisting objective lens was not optimized for this short L_{WD}, aberrations and vignetting were introduced. To minimize degradation of data due to this issue, an effective LF calibration based on the aberration correction theory presented in [8] was developed and applied post-data capture. Similarly, this LF calibration can minimize the impact of aberrations from relay lenses. Since the focus here is the PALFL concept, this calibration is briefly summarized hereby and will be thoroughly discussed in a future paper. The calibration process began with a step of calibrating the amount of vignetting across the field of view by capturing the LF data of a flat Lambertian source extending across the full FOV. By comparing the peripheral sub-aperture images to the center one, the vignetting was quantified and minimized via multiplication for future LF data sets. Following the step of removing the vignetting effects, residual aberrations were minimized next using an analogous process. The LF data of a checkerboard extending across the full FOV was taken. By comparing the peripheral sub-aperture images to the center one, the aberrations were quantified and minimized via lateral shifting of pixels for future LF data sets.

5. Data capture

Figures 4(a) through 4(e) show the captured LF data organized into sub-aperture images bordered in green according to the sample scheme shown in Fig. 3(c). The captured scene consists of a part of the bladder model and a screwdriver placed in front within the FOV to simulate a laparoscopic surgical tool. For scaling reference, the width of the screwdriver is 3 mm while the background bladder model is minified since it is farther away. The center sub-aperture image, Fig. 4(a), is uncalibrated and colored and was used as a reference for LF calibration. For the peripheral sub-aperture images, Figs. 4(b) through 4(e), the calibrated greyscale results extracted from the green color channel were shown along with white grid lines representing matching locations on the sensor for reference. Each of the original sub-aperture images has a high pixel resolution of 1280×960 pixels, which is the same as that of the native sensor. Due to the LF calibration, the FOV of the peripheral sub-apertures was cropped as seen in these images. Figures 4(f) through 4(i) show magnified images of a small region, marked by a Red box on each of the corresponding sub-aperture images, 4(b) through 4(e), respectively. The small but slightly different displacements of the screwdriver relative to the white reference grids in the different sub-aperture images help to visualize the ray separations described in Fig. 1 and validate that the screwdriver is in front of the nominal working distance, L_{WD}.
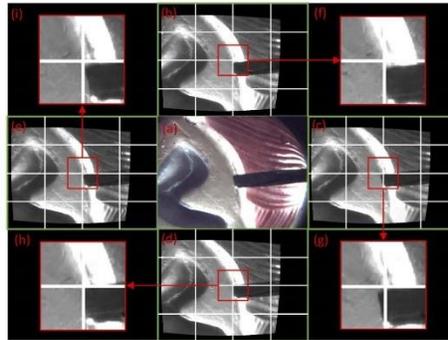
Fig. 4. LF data: (a) uncalibrated center and (b-e) calibrated peripheral sub-aperture images, and (f-i) magnified views of ray separations.

The optical performance of the built prototype was limited by the quality of the stock lenses in the eyepiece and the use of a generic focusing lens. Therefore, we only utilized the greyscale images converted from the green color channel for further data processing to eliminate the effects of chromatic aberration, and we only used the center five angular samples to avoid severe vignetting and aberration-blurring, which increases significantly for sub-apertures farther from the optical axis. These five samples of sub-aperture images, however, are adequate to demonstrate the minimum data needed to achieve maximum data processing speeds and depth sensitivity from x or y-oriented image features in a PALFL system.

The angular sampling dimensions for the data in Fig. 4 were determined experimentally. A 1 mm diameter iris was found to produce sufficient sub-aperture image quality and depth-of-field for the object distances of interest. A 0.91 mm pitch between the sub-apertures at the stop provided a balance between enough light ray separation at different object depths, absence of sub-aperture image aberration, and aliasing during digital refocusing.

110

The diffraction limited spatial resolution of the sub-apertures was measured using a 1951 USAF resolution target (groups 0-3) placed at an L_{WD} of 20 mm. Figure 5 shows the center sub-aperture image, a zoomed in view of groups 2 and 3, and green channel intensity profiles along group 3, element 3 and 4. The bars in element 3 are clear while in element 4, they begin to diminish. This indicates that the diffraction limited spatial resolution is in between these two elements, which is ~10.7 lp/mm. Although the sub-aperture spatial resolution is limited by diffraction, the higher pixel sensor resolution is not wasted because it enables more precise measurement of disparity between sub-aperture images and will also be used for the conventional laparoscope, where the PA is fully opened and the optical resolution is higher.
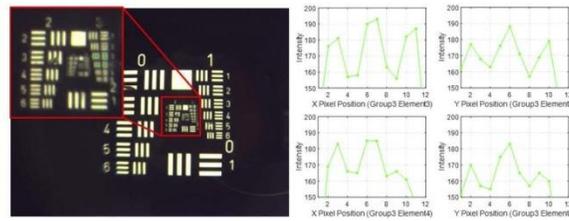
Fig. 5. The center sub-aperture image and intensity profiles of a 1951 USAF resolution target placed at an L_{WD} of 20 mm.

6. Data processing

A modified open source code [12] was used to process the calibrated LF data for digital refocusing and to generate depth maps. Figure 6 demonstrates digital refocusing for three image planes corresponding to near, medium, and far object distances. At near focus, the screwdriver is identifiable while the background is blurry. At medium focus, the white protrusion on the bladder model becomes clear. At far focus, the screwdriver and white protrusion are defocused while the pink line features on the right side are beginning to defocus. Because of the minimum angular sampling for this experiment, when refocusing to one extreme depth, the opposite one shows some aliasing, as seen by the edges of the defocused screwdriver when the focus is far.
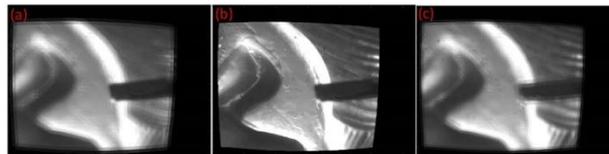
Fig. 6. Digital refocusing to depths: (a) near, (b) medium, and (c) far (see also Visualization 1)

Figure 7(a) was constructed by applying an intensity gradient threshold to Fig. 4(a) to highlight pixels containing strong image features for confident depth estimation. The depth was then estimated at those pixels while the other pixels were nullified. These null regions were interpolated based on the nearest confident depth estimation to construct a full depth map. This strategy reduced noisy depth estimations. Figures 7(b) and 7(c) show full depth maps generated from algorithms based on focus contrast and on correspondence feature matching, respectively. For

each object point, the depth estimation is obtained by measuring at the sensor the separation between light rays captured by adjacent sub-apertures (in units of sensor pixels). A negative pixel value indicates the separation occurred in the opposite direction, as shown in the zoomed view of Fig. 1 when comparing the ray separation from near and far images. Greyscale color illustrates that darker is closer and brighter is farther, allowing determination of relative depth.
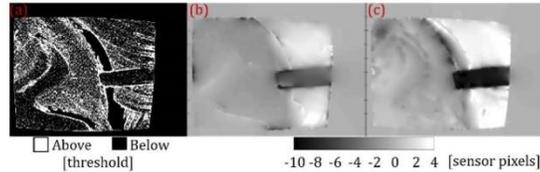
Fig. 7. (a) Intensity gradient thresholding of Fig. 4(a) for depth mapping noise reduction. Relative depth reconstruction maps based on (b) depth from focus contrast and (c) depth from multi-view correspondence feature matching.

Both depth maps identify the correct objects at three different depths, according to Fig. 6. However, depending on the image feature characteristics [12] and error from defocus aliasing, the algorithms perform differently. In the focus contrast map shown in Fig. 7(b), aliasing resulted in inconsistent depth estimation between the screwdriver's edges and body. Also, aliasing likely caused slight inconsistency between the two algorithms in the depth estimation of the farthest layer of depth. Therefore, the feature matching algorithm performs better for larger depth ranges. On the contrary, for the grey valley and surrounding white region on the left side of the FOV where aliasing is absent, the focus contrast map provides a smoother depth reconstruction.

7. Quantitative depth mapping

A lookup table method was created to enable conversion of depth maps from the pixels measuring ray separations to absolute, quantitative depth values and to validate depth resolution based on the system design. Figure 8(a) shows the center sub-aperture view of a $45°$ tilted ruler providing 0.7 mm depth intervals across the vertical FOV. After applying the same LF calibrations as those in the bladder model experiment, a smooth focus contrast depth map was generated in Fig. 8(b). Based on the measured ray separation, Fig. 8(b) highlights the pixels corresponding to $d_\pm$ and the L_{WD} of 20 mm. The corresponding pixels were found in Fig. 8(a), and knowing the ruler dimensions, the units were converted to physical depth.

Fig. 8. A (a) tilted ruler object and its (b) measured depth map create a lookup table for converting ray separations to absolute depth values.

These results were compared to our derived depth resolution study in Sec. 3. Due to the optical performance limitations discussed earlier, we experimentally determined the following prototype

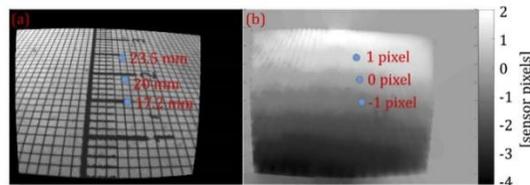specifications. Knowing the real image to object magnification and manufacturer pixel size (B), the sensor spatial resolution in the object space, B_{obj}, of the current prototype was calculated to be 21.3 lps/mm for the center angular samples. Calculating B_{obj} using dimensions known in Fig. 5 yields a similar result. We measured the equivalent D_{EP} from the angular samples shown in Fig. 4 to be 0.345 mm and the equivalent NA_{WD} of the sampled data to be 0.0074 for an L_{WD} of 20 mm. From Eq. (1), d_+ and d_- are 3.6 and 2.7 mm, respectively. Measured from the labeled data points in Figs. 8(a) and 8(b), the ± 1 sensor pixel depths corresponding to d_+ and d_- are separated from the 0 sensor pixel depth on the ruler by $+5$ and -4 intervals, respectively. Knowing the depth between each interval on the tilted ruler is 0.7 mm, they correspond to measured depth resolutions of 3.5 mm and 2.8 mm, respectively, resulting in a maximum percent error of 3.7% in comparison to the theoretical values. Because depth estimation may be nonuniform depending on the algorithm used and the variation of an object's texture density, the percent error can fluctuate for different objects throughout the FOV. Nevertheless, the results presented here demonstrate the potential of the PALFL while depth estimation algorithms are continually being improved.

8. Conclusion

In conclusion, a PALFL was conceptualized to obtain high spatial resolution LF data up to that of the camera sensor for refocusing and quantitative depth mapping, without trading off angular resolution. By taking advantage of the PA's flexibility, this hybrid system integrates the high performance of existing 2D endoscopes with 3D LF imaging. Theory was then developed to analyze, compare, and design laparoscopes regarding adequate depth resolution. A bench-top prototype using an existing laparoscope objective demonstrated proof of concept by performing quantitative depth mapping according to the depth resolution theory. Using our understanding of this prototype, the next generation PALFL will incorporate many improvements. We will optimize the optical system to achieve high performance at its full aperture, incorporate a liquid crystal array in either a transmissive or reflective mode with multiplexed LF acquisition capability to acquire data up to the sensor frame rate, include relay lenses to extend the optical system, and redesign the system to have a working distance and maximum baseline similar to current commercial stereo endoscopes.

Funding

Disclosures

Dr. Hong Hua has a disclosed financial interest in Magic Leap Inc. that was not involved in the work reported here. The terms of this arrangement have been properly disclosed to The University of Arizona and reviewed by the Institutional Review Committee in accordance with its conflict of interest policies.

References

1. R. Sinha, M. Sundaram, S. Raje, G. Rao, M. Sinha, and R. Sinha, "3D laparoscopy: Technique and initial experience in 451 cases," Gynecol. Surg. **10**(2), 123–128 (2013).
2. B. Alaraimi, W. El Bakbak, S. Sarker, S. Makkiyah, A. Al-Marzouq, R. Goriparthi, A. Bouhelal, V. Quan, and B. Patel, "A randomized prospective study comparing acquisition of laparoscopic skills in three-dimensional (3D) vs. two-dimensional (2D) laparoscopy," World J. Surg. **38**(11), 2746–2752 (2014).
3. G. Currò, G. La Malfa, S. Lazzara, A. Caizzone, A. Fortugno, and G. Navarra, "Three-Dimensional Versus Two-Dimensional Laparoscopic Cholecystectomy: Is Surgeon Experience Relevant?" J. Laparoendosc. Adv. Surg. Tech. **25**(7), 566–570 (2015).
4. J. Geng and J. Xie, "Review of 3-D endoscopic surface imaging techniques," IEEE Sens. J. **14**(4), 945–960 (2014).
5. N. Bedard, T. Shope, A. Hoberman, M. A. Haralam, N. Shaikh, J. Kovacevic, N. Balram, and I. Tosic, "Light field otoscope design for 3D in vivo imaging of the middle ear," Biomed. Opt. Express **8**(1), 260–272 (2017).

113

6. S. Zhu, P. Jin, R. Liang, and L. Gao, "Optical design and development of a snapshot light-field laryngoscope," Opt. Eng. **57**(02), 1 (2018).
7. J. Liu, D. Claus, T. Xu, T. Keßner, A. Herkommer, and W. Osten, "Light field endoscopy and its parametric description," Opt. Lett. **42**(9), 1804 (2017).
8. R. Ng, "Digital light field photography," PhD. Thesis, Stanford University (2006).
9. C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography," ACM Trans. Graph. **27**(3), 1 (2008).
10. C. Zuo, J. Sun, S. Feng, M. Zhang, and Q. Chen, "Programmable aperture microscopy: A computational method for multi-modal phase contrast and light field imaging," Opt. Lasers Eng. **80**, 24–31 (2016).
11. Y. Qin, H. Hua, and M. Nguyen, "Characterization and in-vivo evaluation of a multi-resolution foveated laparoscope for minimally invasive surgery," Biomed. Opt. Express **5**(8), 2548 (2014).
12. M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," *Proc. IEEE Int. Conf. Comput. Vis.*, 673–680 (2013).

APPENDIX B: Prism-based Tri-aperture Laparoscopic Objective for Multi-view Acquisition

Elliott Kwan and Hong Hua

Prism-based tri-aperture laparoscopic objective for multi-view acquisition

ELLIOTT KWAN AND HONG HUA*

3D Visualization and Imaging Systems Laboratory, James C. Wyant College of Optical Sciences, University of Arizona, 1630 E University Blvd., Tucson, AZ 85721, USA
hhua@optics.arizona.edu

Abstract: This paper presents the design and prototype of a novel tri-aperture monocular laparoscopic objective that can acquire both stereoscopic views for depth information and a wide field of view (FOV) for situational awareness. The stereoscopic views are simultaneously captured via a shared objective with two displaced apertures and a custom prism. Overlapping crosstalk between the stereoscopic views is diminished by incorporating a strategically placed vignetting aperture. Meanwhile, the wide FOV is captured via a central third aperture of the same objective and provides a 2D view of the surgical field 2x as large as the area imaged by the stereoscopic views. We also demonstrate how the wide FOV provides a reference data set for stereo calibration, which enables absolute depth mapping in our experimental prototype.

1. Introduction

The conventional optical design of a rigid laparoscope comprises of an optical objective lens and rod lens relays, and this form has been used for minimally invasive surgery since the beginning while novel designs are continually being developed. The conventional design has provided surgeons with excellent two-dimensional (2D) image quality over the operative field. However, two major optical limitations arise with conventional 2D laparoscopes: (1) the absence of binocular vision results in restricted depth perception and (2) the field of view (FOV) and spatial resolution are inversely proportional. The lack of depth information requires extensive training for physicians to become efficient with a 2D operative view. Meanwhile, the second limitation constrains the FOV to cover just the surgical area to maintain sufficient image resolution. Hence, complications that occur outside the surgical area would not be seen unless the laparoscope is physically moved. To improve upon these surgeries, these two limitations and their corresponding optical design solutions have been explored separately in literature.

To recover depth perception, various methods have been investigated, including dual-sensor stereo, single-sensor stereo, single-sensor 3D imaging via structured light, and uniaxial 3D imaging [1]. Commercially, stereoscopic endoscopes with dual-channel object-relay optics and dual imaging sensors, like the DaVinci and Endoeye Flex 3D [2,3], have been popularized through successful demonstration of 3D vision and depth perception in the live surgical setting. Academically, other types of stereo endoscopes have shown potential for the future. The ones that can acquire stereo vision in a uniaxial, single camera system are attractive because they preserve the limited design volume constrained by the laparoscope housing. The 3D-MARVEL system highlights this by using a monocular system with a dual aperture comprising of complementary multi-bandpass filters [4]. A generalized multi-aperture approach can acquire multiple views from different directions with varying baselines, leading to demonstrations of light field endoscopes [5,6]. For instance, Kwan et al. demonstrated light field laparoscope design, which captures each view by sampling the entrance pupil with a programmable aperture placed at the stop of the laparoscope [5]. Instead of using multiple apertures, another way to acquire multiple views is to place a multifaceted prism [7,8] or a microprism array (MPA) [9] in front of the endoscope. These prisms bend the light rays from the stereo views towards the camera lens so that each stereo

image can be captured on one half of the image sensor. Although all these types of laparoscopes can acquire depth information, they still are limited by the FOV and spatial resolution tradeoff.

To overcome the second major optical limitation, the FOV versus spatial resolution tradeoff has been eliminated using a multi-resolution foveated laparoscope (MRFL) [10–14]. A MRFL system simultaneously captures high-resolution zoomed-in and wide-angle zoomed-out views through a shared objective-relay tube and two separate imaging probes via beam splitting and 2D scanning. Surgeons can then use the high-resolution zoomed-in view for surgical operation while the wide-angle view provides peripheral awareness for preventing patient injury from accidental collisions of surgical instruments outside the surgical area. The MRFL has demonstrated successful 2D wide FOV (WFOV) minimally invasive surgery in animal trials, but additional implementation of depth perception recovery has not been attempted.

A laparoscope that can capture both 2D WFOV and 3D depth information has only been minimally explored and could be a promising development for minimally invasive surgery. Kit et al. recently attempted to utilize two independent stereo cameras to achieve 3D vision while creating a larger 2D FOV image by stitching the non-overlapping FOVs of the stereo cameras [15]. It required trading off a portion of the stereo FOV (SFOV) to gain a 2D WFOV. This situation is suboptimal because the surgeon would want to perform surgery with a maximal SFOV while the WFOV is only used periodically for peripheral awareness. Nonetheless, the concept was demonstrated and provides groundwork for future developments.

In this paper, we present the optical design and prototype of a novel tri-aperture monocular laparoscopic objective (TAMLO) for acquiring SFOV views for 3D vision and a 2D WFOV view for peripheral awareness. The monocular form factor with a sufficiently large clear aperture is chosen over a conventional stereo system with dual-channel objective-relay optics to preserve the design volume that is limited by the diameter of the laparoscope housing. More specifically, having two independent optical systems consumes more design volume due to the required edge apertures from each system and the opaque space between them. Instead, a monocular system can capture two stereo views from two laterally displaced aperture stops while preserving the central design volume for capturing the WFOV through a central aperture stop. The tri-aperture layout is a simplification of the multi-aperture approach of light field endoscopes [5] and retains the benefits of 3D depth acquisition without the cost of spatial or temporal tradeoffs. To enable simultaneous acquisition of the stereo views on a single sensor, we also incorporate a custom prism that can support the tri-aperture layout inside the TAMLO. As a result, the stereoscopic views are simultaneously captured via a shared objective with two displaced apertures and a custom prism while the 2D WFOV view is captured via a central third aperture of the same objective [16]. The rest of the paper is organized as follows. Section 2 presents the schematic design and key first-order analytical considerations for the proposed TAMLO optics. Section 3 discusses the optical design for a TAMLO prototype and the solution to resolve the crosstalk issue between the stereo views. Section 4 presents the opto-mechanical design and prototype assembly along with raw images of the SFOV and WFOV views captured by our experimental prototype. Section 5 demonstrates a method of stereo distortion calibration and camera modeling assisted by the WFOV view, the results of calibrated stereo views, and the absolute depth mapping results.

2. Schematic design of a TAMLO

Figure 1(a) shows the schematic layout of the proposed prism-based tri-aperture monocular laparoscope design, and Fig. 1(b) fully illustrates the optical layout and key parametric specifications of the TAMLO design, which is the main contribution of this paper. Adopting the convention of light traveling from left to right, the TAMLO images the object field through three laterally displaced apertures and forms three different views of the object on the intermediate image #1, corresponding to a wide FOV (WFOV) image of a large object field and two stereo FOV (SFOV) images of a smaller overlapping object field. Following the TAMLO, conventional

117

laparoscope relay optics is utilized to relay the intermediate image #1 and forms an intermediate image #2 outside the patient's body. After relaying the intermediate image #1 to the intermediate image #2, the spatial arrangement of the three views captured by the TAMLO is still preserved. Depending on the specific requirements and priorities of a particular laparoscope design, the three views may be captured by a single or multiple imaging sensors, which leads to different possible designs of the imaging probes. For instance, as illustrated in Fig. 1(a), we may adopt a scheme similar to the dual-channel imaging probes in our previous MRFL system. An eyepiece collimates the light from the intermediate image #2 and a beamsplitter then splits the collimated light into two imaging paths, one for capturing the WFOV image and one for the SFOV images. The type of beamsplitter used will depend on the type of tri-aperture selector used to separate the overlapping WFOV and SFOV images formed on the intermediate image #1. The two SFOV images through the two side-view apertures are recorded simultaneously on each half of the sensor #2, while the WFOV image through the center aperture is captured by the sensor #1. As long as the multiple views captured by the TAMLO are constrained within the maximally allowed diameter, Π_W, of the intermediate image #1, the relay lens group only needs to be designed to support a FOV matching Π_W. The relay lens must also support the maximal ray angle incident on the intermediate image #1 from all of the views. To avoid severe light loss from vignetting, it is preferred that the TAMLO and relay lens group are designed to be nearly telecentric at both intermediate image #1 and #2. It is worth noting that the TAMLO may be utilized alone for the option of a chip-on-tip form factor.
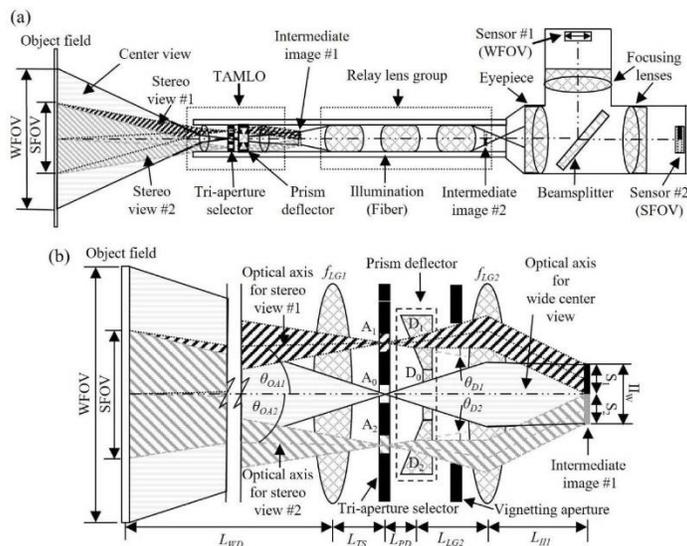

Fig. 1. (a) Proposed schematic layout of the prism-based tri-aperture monocular laparoscope design and (b) a magnified view of the TAMLO design with optical layout and key parametric specifications.

As schematically illustrated in Fig. 1(b), the TAMLO mainly consists of a front lens group with a focal length of f_{LG1}, a tri-aperture selector, a prism deflector, and a back lens group with a focal length of f_{LG2}. The two lens groups are placed in front and behind the selector-deflector assembly to provide sufficient degrees of freedom during lens design optimization. The distances L between adjacent components are constrained by the method of optomechanical mounting and then precisely determined by the optical optimization. The chief ray bundles for the three views, which are highlighted by the different shaded regions in Fig. 1(b), are ideally maintained in separate regions of the two lens groups so that the local regions of the lenses can be optimized to the respective viewing angle. The aperture stop plane of the TAMLO is located at the tri-aperture selector, which consists of an on-axis central aperture A_0 and two decentered apertures of A_1 and A_2. A different viewing angle of the object field is seen by each of the three different aperture stops as indicated by the labeled optical axes for the wide center view and the stereo views #1 and #2, respectively. The prism deflector located adjacent to the tri-aperture selector is made up of individual prisms D_i corresponding to each aperture stop A_i. The central prism D_0 is effectively a thin plane parallel plate and does not change the outgoing ray angles from A_0, so a WFOV image, Π_W, of the object field is formed at the intermediate image #1 and is centered about the central optical axis. D_0 could be removed leaving an air space, but instead is present to provide structural support and to manufacture the prism deflector as one piece. The side prisms, D_1 and D_2, bend rays transmitting through them by deflection angles of θ_{D1} and θ_{D2}, respectively, in opposite directions so that the stereo view images, S_1 and S_2, also located at the intermediate image #1, are laterally translated apart to the opposite sides of the central optical axis. Without the side prisms, S_1 and S_2 would be spatially overlapping on the exact same region about the central optical axis.

The two lens groups with focal lengths of f_{LG1} and f_{LG2} are optimized to support the WFOV captured by A_0 and the SFOV captured by apertures A_1 and A_2. To ensure that the WFOV and stereo views can be imaged by the same relay lens within a confined volume required by a rigid laparoscope, the dimensions of the stereo view images, S_1 and S_2, need to be constrained to the same circular region as the WFOV image, Π_W. Furthermore, to avoid crosstalk between the two stereo views, S_1 and S_2 should not overlap at the intermediate image #1. Therefore, the maximum SFOV covered by both the S_1 and S_2 along the direction of the stereo aperture displacement, $SFOV_{\overrightarrow{A_1A_2}}$, shall satisfy:

$$SFOV_{\overrightarrow{A_1A_2}} \leq \frac{WFOV_{\overrightarrow{A_1A_2}}}{2}, \tag{1}$$

where *WFOV* is the maximum wide FOV of the objective. *WFOV* is expressed as $2\tan^{-1}\frac{\Pi_W}{2f_{TAMLO}}$, where f_{TAMLO} is the effective focal length of the objective, given as $\frac{1}{f_{LG1}} + \frac{1}{f_{LG2}} - \frac{(L_{TS}+L_{PD}+L_{LG2})}{f_{LG1}f_{LG2}}$, and Π_W is the maximally allowed diameter of the intermediate image #1. The maximum SFOV along the axis orthogonal to the stereo aperture displacement can be as large as that of the maximum WFOV. To maximize the SFOV without causing overlapping between S_1 and S_2, the optimal value of the prism deflection angles, θ_{D1} and θ_{D2}, are determined by:

$$\theta_{D1} = -\theta_{D2} \approx \tan^{-1}\frac{\Pi_W}{4f_{LG2}}. \tag{2}$$

To separate the overlapping WFOV and SFOV images, the tri-aperture selector, which determines the choice of the beamsplitter to be used in Fig. 1(a), may be implemented from different types of technologies that either block or encode the transmitting light. Blocking technologies include a mechanical shutter or a liquid-crystal device (LCD) that allows localized control of light transmission through a sub-region by switching the corresponding region on or off. Encoding technologies include a custom polarization device or color filter that allows localized control of light transmission through a sub-region by encoding different polarization states or

119

spectral filters across the tri-aperture selector. If a blocking technology is used, either A_0 or $A_{1,2}$ is blocked in a time-sequential fashion so that the WFOV and SFOV images can be alternately captured by a single sensor, which eliminates the need for a beamsplitter and a second imaging probe in Fig. 1(a) and leads to a simpler system design with lower hardware cost. If an encoding technology is used, A_0 can be encoded oppositely to $A_{1,2}$. For example, orthogonal polarizers may be utilized for the apertures A_0 and $A_{1,2}$. Then, a corresponding polarizing or dichroic beamsplitter matching the encoded tri-aperture selector is used. This results in simultaneous capture of the WFOV and SFOV images by the two sensors illustrated in Fig. 1(a). The tradeoff of using this technology is half of each view's irradiance due to the encoding filters.

There are multiple variations of the prism deflector design that vary in manufacturability and light manipulation. For the example in Fig. 1(b), the back faces of the prisms D_0 through D_2 are co-aligned vertically while the front faces of D_1 and D_2 are tilted oppositely to achieve the proper amount of light ray bending. Since only one side of the prisms requires angled faces, the three prisms can be manufactured as one piece through diamond turning. After deflection, the light rays must pass through different portions of the TAMLO's back lens group. Ideally, the ray bending by the prism deflectors is desired to be independent of ray incident angle. Realistically, the net ray bending by the prism deflectors is derived using Snell's law [17]:

$$\theta_{D1} = -\theta_{D2} = \alpha - \sin^{-1}\left[\sqrt{n^2 - \sin^2\theta_i}\sin\alpha - \cos\alpha\sin\theta_i\right] - \theta_i, \tag{3}$$

where α is the angle of the prism, n is the index of the prism material, and θ_i is the incident angle of the incoming light ray. Since $\theta_{D1,2}$ is dependent on a field or incident angle, $S_{1,2}$ are distorted accordingly, but they will be calibrated in post-processing. Combining Eqs. (2) and (3), the prism design can be approximated.

Acquiring good image performance and sufficient depth resolution is critical to the TAMLO's functionality. The diameter of each aperture stop A_i determines the corresponding F/# or numerical aperture of each view, and thus the cut-off spatial frequency or limiting resolution of the objective. Therefore, within the optics volume constraints for a rigid laparoscope, they should be maximized for optimal optical resolution. In the meanwhile, the lateral separation between the centers of the aperture stops A_1 and A_2, denoted as BL_{TS}, ultimately determines the effective baseline, EBL, of the stereo views and thus the depth resolution of the system. The EBL can be found by determining the lateral separation between the centers of the entrance pupils, which are optically conjugate to the aperture stops A_1 and A_2 through the first lens group and is expressed as:

$$EBL = \frac{BL_{TS} * f_{LG1}}{f_{LG1} - L_{TS}}, \tag{4}$$

where $BL_{TS} = \overline{A_1A_2} = 2\overline{A_{1,2}A_0}$ and L_{TS} is the axial displacement of the aperture stop from the first lens group. The location of the entrance pupils, denoted by L_{EP}, is found by imaging the tri-aperture selector through the first lens group and is expressed as $L_{EP} = L_{TS}f_{LG1}/(f_{LG1} - L_{TS})$. The EBL can also be described in object space as:

$$EBL = (L_{WD} + L_{EP})(|\tan(\theta_{OA1})| + |\tan(\theta_{OA2})|), \tag{5}$$

where θ_{OA1} and θ_{OA2} define the optical axis directions of the stereo views with respect to the central optical axis, and L_{WD} is the working distance optically conjugate to the intermediate image #1 through the TAMLO objective. Meanwhile, the depth resolution of the SFOV system is determined analogously to the light field laparoscope in [5]. The average depth resolution, d, is

120

given by:

$$d \approx \frac{2L_{WD}}{EBL} \left| \frac{-f_{LG1}f_{LG2} + tf_{LG1} + L_{WD}(f_{LG1} + f_{LG2} - t)}{-f_{LG1}f_{LG2}} \right| P, \tag{6}$$

where P is the limiting resolution or equivalent pixel size at the intermediate image #1 and $t = L_{TS} + L_{PD} + L_{LG2}$. The first fraction corresponds to the triangular geometry between the object field and the effective baseline while the second fraction corresponds to the magnification of the pixel from intermediate image #1 to the object field. The variables in this equation must be chosen properly to obtain adequate depth resolution for laparoscopic surgery. An *EBL* of 4 mm is standard for commercial stereo endoscopes and a pixel magnification of ∼18 from intermediate image #1 to the object field is reasonable for recording the object field with an appropriately sized sensor. Using these constants, Fig. 2 plots the depth resolution rendered by a TAMLO as a function of the equivalent pixel size at the intermediate image #1 for three different working distances of 30, 60, and 120 mm. With a typical working distance of about 50 mm, a standard 2D laparoscope with an HD resolution sensor covers a circular object field of about 60 mm in diameter in a spatial resolution up to 16 lps/mm in the object space. Consider the SFOV images are expected to provide the same circular field coverage as a standard 2D laparoscope with a single HD resolution sensor, the equivalent pixel size on the intermediate image #1 falls in the range from 1.5 to 3 μm, depending on the maximally allowed diameter of the intermediate image #1 due to package constraints. A pixel resolution between 0.5 and 4 μm on the intermediate image #1 can provide a depth resolution from 0.14 to 4.37 mm depending on the working distance. This indicates a properly designed TAMLO can provide sufficient depth resolution for surgical guidance.
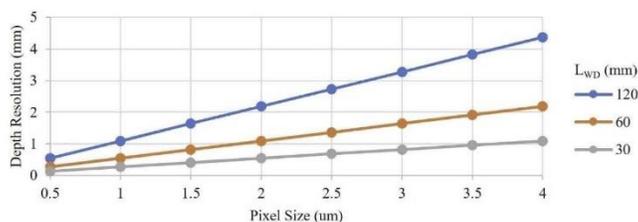

Fig. 2. The depth resolution range of the TAMLO as a function of pixel size at intermediate image #1 for three different working distances.

Finally, the TAMLO must be able to acquire the three views without interference. The stereo images S_1 and S_2 are translated apart by the prism deflector according to the designed SFOV. However, as illustrated by Fig. 1, the object field is larger than the SFOV, which results in image points present outside of each designed stereo image $S_{1,2}$ due to lack of a field stop. The undesired image points from one stereo image will overlap the opposite stereo image across the central optical axis, resulting in crosstalk. To address this issue, a vignetting aperture can be placed after the prism deflector, as shown in Fig. 1(b). At this location, the stereo ray bundles have diverged enough so that individual fields can be vignetted. The shape of the vignetting aperture depends on the prism deflector design. For the example in Fig. 1(b), an annulus vignetting aperture is required to allow the WFOV ray bundle to pass while blocking any stereo rays outside of the illustrated stereo ray bundle regions on the sides toward the central optical axis. Alternatively, if the two stereo views were oppositely polarized, a sensor with matching polarization on each half of the sensing area can eliminate crosstalk.

121

3.   Tri-aperture objective lens design

Based on the analytical relationships and various constraints described in Section 2, we derived the first order specifications of the TAMLO design, which are listed in Table 1. To ensure adequate field coverage by both the WFOV and SFOV images, we chose a working distance of 120 mm. With a full FOV of 39° and 26° diagonally for the wide and stereo views, the system captures a circular region with a diameter of about 85 mm and 54 mm, respectively. To ensure adequate depth resolution, an *EBL* of 4 mm was chosen. We aim to fit the TAMLO prototype for a standard laparoscopic trocar and thus limited the mechanical housing diameter for the objective to be 12 mm and the maximum optics diameter to be 8 mm to account for housing and fiber illumination of 1 mm thickness each. We further limited the maximally allowed diameter of the intermediate image #1, II_W, to be 4.4 mm to allow direct capture of the images with a standard 1/3" imaging sensor and future development of the relay and imaging optics. These FOV and image size constraints led to an effective focal length of 7 mm for the objective. Between the two lens groups, we set an f_{LG1} of 47.5 mm for aberration compensation and balancing out the ray bending throughout the system and an f_{LG2} of 8 mm for sufficient power in a Petzval objective design. Since prisms are commonly made from N-BK7 glass material, a refractive index of 1.517 was selected for the prism deflector. The design of the side prisms was approximated using Eqs. (2) and (3) and an isosceles shaped prism for simplicity such that $\theta_i = \alpha$ for an incident ray parallel to the central optical axis. Based on the defined specifications, the bending $\theta_{D1,2}$ of the stereo chief rays is ±7.8° to adequately separate the stereo images, and the associated prism angle α was found to be ~15°. Using an HD sensor with 2.2 µm pixels, Fig. 2 indicates that this design will provide ~2.5 mm depth resolution at a working distance of 120 mm, and a higher depth resolution can be achieved at a shorter working distance. Compared to the MRFL prototypes [10–14], the main tradeoff of implementing the stereo apertures was the reduction of the WFOV to about half. In other words, the challenge of the TAMLO design is to balance the optical performance between the SFOV and WFOV. The entrance pupil diameter for all three apertures was set to 1.2 mm, leading to an F/# of 5.8. The entire objective, however, is effectively F/1.35 because it supports, in a monocular form factor, larger ray angles that come from the stereo aperture stops. The target spatial resolution in the object space was set to be 2.1 lps/mm and 6.25 lps/mm for the WFOV and SFOV, respectively. The object resolution specification is weighted lower for the WFOV than the SFOV because it is mainly used for peripheral awareness. Meanwhile, the constraints applicable to conventional rigid laparoscopes were also met. The optical design is constrained for image space telecentricity so that relay lenses can be easily inserted after the objective lens.

Table 1. First order lens design specifications for TAMLO

Working distance	120 mm	Effective stereo baseline	4 mm
Stereoscopic full FOV	26 deg.	Wavelengths	625, 506, 456 nm
Wide full FOV	39 deg.	Object resolution for stereoscopic view	6.25 lps/mm
Effective focal length	7 mm	Object resolution for wide view	2.1 lps/mm
Entrance pupil diameter per aperture	1.2 mm	Mechanical housing diameter	12 mm
Telecentricity	Image space telecentric	Maximum diameter of lenses	8 mm

The starting point of the TAMLO lens design was based on the existing MRFL and commercial 3D endoscope objectives. In all lenses and the prism deflector, rays were constrained within a 7.2 mm clear aperture diameter, or 90% of the maximum lens diameter. This diameter is slightly larger than that of the MRFL to provide additional design volume for the SFOV ray paths. For the SFOV, the object field was sampled across the + x and +/-y region with an aspect ratio

of 4:3 corresponding to the sensor because each stereo view system is bilaterally, rather than rotationally, symmetric. Some of these field points along the edge are noted in Fig. 3(d). In the middle stage of the design process, custom lenses and prism deflector designs were allowed to determine the maximum achievable image performance and avoid local minimum solutions. Throughout this phase, the size of the SFOV was maintained at the same size as the conventional 2D laparoscope while the other first order specifications were adjusted accordingly based on what was practical, given the required laparoscope constraints and the incorporation of stereo apertures. Since the size of the intermediate image II_W can vary and be magnified accordingly after being relayed, it was kept to < 5.6 mm diameter to avoid vignetting and ensure possible integration with relays. In the late stage, further constraints were implemented to convert the custom components into manufacturable ones. Tolerance sensitivity reduction was also applied to produce realistic lens shape factors. In addition, the lens housing would be 3D printed, so optomechanical tolerances were loosened accordingly to ensure assembly variation was accounted for. The custom lenses were then converted to stock lenses to lower costs and achieve rapid prototyping. Image performance was slightly reduced as a result but can be restored in future versions.
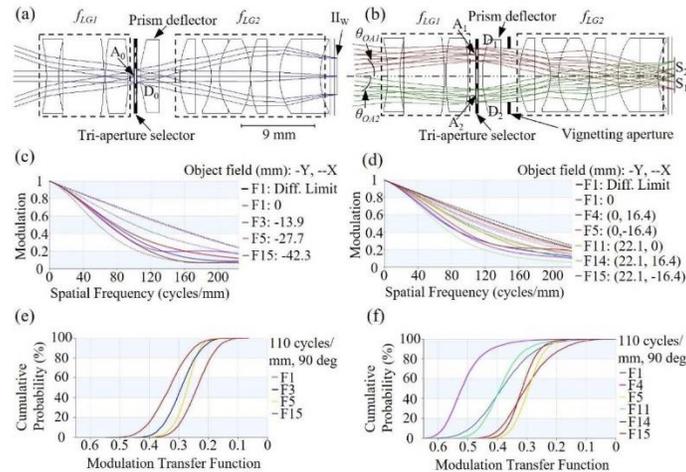

Fig. 3. Manufacturable TAMLO lens design for (a) WFOV and (b) SFOV acquisition. Corresponding (c, d) polychromatic MTFs and (e, f) tolerance analyses indicate sufficient performance for prototyping.

The manufacturable TAMLO lens design to be prototyped is shown in Figs. 3(a) and 3(b) for WFOV and SFOV acquisition, respectively, and the lens prescription is shown in Table 2. Both figures show the same set of monocular lenses, and all of them except one are stock components. It was found that the achromatic doublet near the middle and the field lens should be kept in the meniscus shape to maintain good image performance. Since meniscus lenses are uncommon in stock lenses, the achromat was custom made and the field lens was formed using two singlets of the same glass. For the prism deflector, a reversed deflection design was chosen so that it could be more easily manufactured as one piece. It is essentially an obtuse angle prism with the

top flattened out. In this real lens design, $\theta_{D1,2}$ is $\pm 6.22°$ for a horizontal incident ray, and the associated prism angle α is $11.9°$. Compared to these real values, the corresponding theoretical values from the beginning of this section are slightly different because they did not account for real thick optics and relied on approximations for simplification. Yet, those values were a good starting point for the prism design. Furthermore, the dispersion from the prism deflector was accounted for. The prism deflector design can be thought of as a segmented lens. For example, the one in Fig. 1(b) approximates a concave-plano lens, and the one in Fig. 3(b) approximates a convex-plano lens. Because the prism deflector looks like a conventional lens, its dispersion is similar to the lens it approximates. This dispersion was suppressed by using conventional dispersion compensation from the other monocular lenses with different glass types during the lens optimization process.

Figure 3(a) also shows the light rays from the WFOV transmitting through the central aperture stop A_0 and plane parallel plate D_0. Because there is no deflection from D_0, the WFOV system is modeled as rotationally symmetric, and the center of II_W is on the lens optical axis. As the chief rays travel to II_W, they are collimated. This indicates the system is image space telecentric. Telecentricity is one of the image quality limiting constraints, which can be removed if using the TAMLO as a chip-on-tip system, where the sensor is placed at intermediate image #1. Telecentricity is maintained in this prototype design to demonstrate design feasibility. Similarly, Fig. 3(b) shows the light rays from the SFOV transmitting through the top and bottom aperture stops $A_{1,2}$. On the prism deflector's left side, the ray bundles from each view were constrained so that they only interact with their corresponding prism surface. On the prism deflector's right side, all ray bundles share the same flat surface. $D_{1,2}$ bend the corresponding ray bundles according to Eqs. (2) and (3) so that the SFOV images are translated to the upper or lower side of the optical axis without surpassing the WFOV boundary, thus allowing for simultaneous stereo image pair acquisition on a single sensor. Compared to Fig. 1(b), the prism design here deflects the corresponding stereo images $S_{1,2}$ to the opposite sides of the optical axis rather than to the same side.

By having lens groups in front and behind the prism deflector, the optical system has sufficient degrees of freedom to achieve a balanced image performance between the two imaging modalities. Comparing Figs. 3(a) and 3(b) further illustrates that for the lenses closest to the tri-aperture selector, the SFOV ray bundles only occupy the outer local portions of the lenses while the WFOV ray bundles mainly occupy the central local portion. This indicates that these lenses have more flexibility to impact the imaging modalities separately, and aspheric surfaces can add additional degrees of freedom. The polychromatic MTFs in Figs. 3(c) and 3(d) corresponding to the WFOV and SFOV, respectively, show that adequate image performance can be achieved with this lens design. Although astigmatism impacts the WFOV especially at the higher frequencies, only peripheral awareness is essential rather than high resolution. Thus, slightly lower contrast is acceptable in the WFOV system. Quantitatively, the MTFs indicate that the lowest modulation for the WFOV at 37 lps/mm (2.1 lps/mm in object space) is 0.7 and for the stereo view at 109 lps/mm (6.25 lps/mm in object space) is 0.24. The cutoff frequency at 227 lps/mm corresponds to the sensor (Allied Vision Alvium 1800 U-500c) used for capturing intermediate image #1. Using stock lens and 3D printing tolerances, the tolerance analyses for the WFOV and SFOV systems in Figs. 3(e) and 3(f), respectively, confirm that this design will maintain adequate performance after assembly. The modulation will be greater than 0.1 at 110 lps/mm (6.12 lps/mm in object space) for both systems. According to Table 1, this approximately meets the object resolution criteria for the SFOV and exceeds it for the WFOV.

As described in Section 2, although $S_{1,2}$ are translated apart, they are larger than as designed in Fig. 3(b) because the object field is larger than the chosen SFOV. This results in overlapping crosstalk between $S_{1,2}$. The amount of overlap is simulated by seeing how much S_1 crosses onto the upper half of the sensor, as shown in Fig. 4(a), when the SFOV is extended by a large

Table 2. Lens prescription for design in Fig. 3

Surface #		Curvature radius (mm)	Thickness (mm)	Index	Abbe #	Y Decenter (mm)	Alpha tilt angle (°)
0	Object	Plano	120				
1	Lens 1	-18.86	1.5	1.517	64.2		
2		18.86	0.474				
3	Lens 2	Plano	5	1.785	25.7		
4		-31.39	0.215				
5	Lens 3	27.43	2.74	1.517	64.2		
6		-27.43	1.016				
7	Multiview selector	Plano	A_0: 0.75 A_1: 1.011 A_2: 1.011			0 2.5103 -2.5103	
8	Multiview deflector	Plano	D_0: 2.261 D_1: 2.001 D_2: 2.001	1.517	64.2	0 2.5103 -2.5103	0 -11.7967 11.7967
9		Plano	2.276			D_0: 0 D_1: 2.5103 D_2: -2.5103	
10	Lens 4	11.37	1.93	1.517	64.2		
11		Plano	1.869				
12	Lens 5	-8.6	1	1.847	23.8		
13		8.13381	3.8	1.806	40.9		
14		-9.73765	0				
15	Lens 6	12.92	2	1.517	64.2		
16		Plano	1.519				
17	Lens 7	7.85	2.7	1.785	25.7		
18		Plano	2.25	1.785	25.7		
19		9.42	1.238				
20	Sensor cover glass	Plano	0.47	1.517	64.2		
21		Plano	0.35				
22	Intermediate image 1	Plano					

amount. The boundaries of the larger S_1 indicate when the edge apertures of the TAMLO's lenses begin to vignette S_1. By symmetry, S_2 would overlap just as much on the lower half of the sensor. The amount of overlap past the midline is significant and would corrupt a major portion of the designed $S_{1,2}$ area. Since it is not possible to limit the size of $S_{1,2}$ with a field stop, a vignetting strategy is implemented here. Because the prism deflector deflects the images to opposite sides of the optical axis, a circular vignetting aperture can be inserted right after it to significantly reduce the overlapping crosstalk, as shown in Fig. 3(b). Comparing to Fig. 3(a), the

125

placement of this vignetting aperture will not interfere with the rays from the WFOV system. This technique can preserve most of the designed $S_{1,2}$, as simulated in Fig. 4(b), where most of the overlapping crosstalk is diminished after insertion of the vignetting aperture. Note that in this demonstration, the relative irradiance is 0.5 at the midline of intermediate image #1 because the circular vignetting aperture was designed to half vignette there. If the residual crosstalk needs to be further reduced, the vignetting can be increased and a calibration in post-processing could recover the irradiance that was lost in the designed $S_{1,2}$ area. For prism deflector designs that deflect to the same half of the sensor, such as the one in Fig. 1(b), the circular vignetting aperture would not work because it would vignette at the edges of the sensor instead of the center where the overlapping crosstalk occurs. Instead, a similar vignetting solution could be achieved with an annulus vignetting aperture.
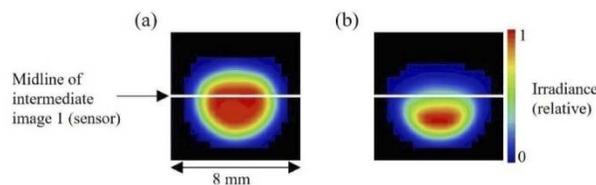

Fig. 4. Simulation of S_1 image size and overlap when SFOV is extended with (a) no vignetting and (b) inserted vignetting aperture.

4. Prototype assembly and raw data

A basic lens housing was designed and 3D printed for assembling the TAMLO prototype, as shown in Fig. 5. The second stock lens prescription did not come in the same diameter as the other lenses, resulting in the large housing in the front. The housing contains railings to align the tri-aperture selector, prism deflector, and sensor along the same axis. To separate the overlapping WFOV and SFOV images, the tri-aperture selector was a manual shutter that blocked either A_0 or $A_{1,2}$. Rectangular aperture blockers were simply inserted into a slot of the housing, resulting in time-sequential acquisition between the two imaging modalities. There is an additional slot after the prism to insert a vignetting aperture to reduce the overlapping stereo image crosstalk. For prototype evaluation, a real sensor was mounted at intermediate image #1. The entrance pupils corresponding to A_{0-2} can be seen clearly in the frontal view of Fig. 5(c).

Figure 6 illustrates the raw data acquired from the working TAMLO prototype. The object field is a ruler lying on a planar checkerboard that is tilted so that the object depth linearly increases as a function of image height. For all the images, distortion can be observed by looking at the curvatures of lines that should be straight. Figure 6(a) shows the WFOV image while the SFOV apertures are blocked. Along the vertical axis, ~5.5 cm of the ruler can be seen. Figure 6(b) shows the SFOV images captured by both stereo apertures simultaneously with the vignetting aperture in place. The stereo images were translated by the prism deflector to the top and bottom half of the sensor without exceeding the WFOV image. Each of the stereo images sees ~2.5 cm of the ruler. Thus, in quantitative comparison, the WFOV shows twice the SFOV in the vertical or baseline direction when the stereo images are captured simultaneously. Figure 6(c) shows the same stereo images taken with the same exposure settings but without the vignetting aperture. Along the midline of the sensor, the strong presence of the overlapping crosstalk reduces the contrast and the sum of the irradiance results in saturated pixels. There is still some residual crosstalk in Fig. 6(b), but it has been significantly reduced, and the vignetting aperture size can
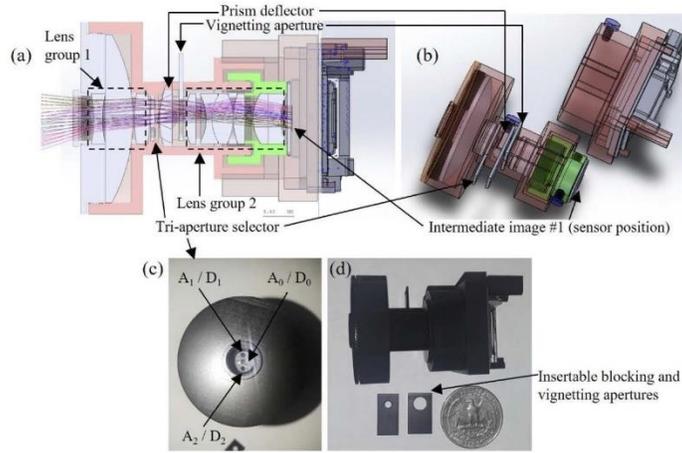
126

Fig. 5. TAMLO (a, b) optomechanical housing design and (c, d) prototype assembly
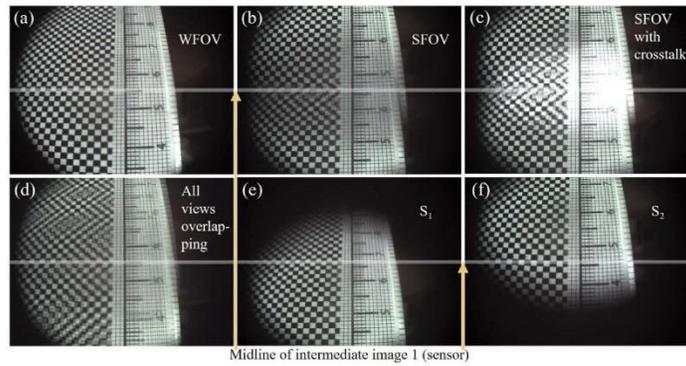

Fig. 6. TAMLO prototype raw data: (a) WFOV, SFOV simultaneous stereo image capture (b) with and (c) without vignetting aperture, (d) overlapping WFOV and SFOV images, (e) S_1 and (f) S_2 captured independently without vignetting aperture.

be further optimized in future prototypes. Figure 6(d) shows the overlap between all three views without a method of blocking or encoding either A_0 or $A_{1,2}$, thus resulting in unusable data. Figures 6(e) and 6(f) show S_1 and S_2, respectively, captured independently without the vignetting aperture. They demonstrate the extent of overlap that causes the crosstalk. Overall, the image quality of these figures appears sufficient, as predicted during the lens design phase.

127

5.   Calibration and absolute depth mapping

To calculate correct disparity and absolute depth maps, the stereo systems require camera parameter and distortion calibration. For a conventional stereo system with two independent cameras, methods for calibrating camera parameters and distortion have been thoroughly developed [18]. The conventional calibration assumes each of the cameras has rotational symmetry, so the lens distortion can be modeled with a radial polynomial. The TAMLO effectively creates two virtual stereo cameras with their optical axes tilted from each other, but their distortion model is no longer rotationally symmetric. Instead, because the TAMLO captures each stereo image with an off-axis aperture, the distortion model is bilaterally symmetric and can have additional distortion from the finite thickness of the prism deflector. Analytically calibrating the unique distortions in the TAMLO would require rigorous theoretical analysis. Alternatively, a numerical solution can be developed by taking advantage of the additional WFOV data, which was captured with rotational symmetry.

The goal of our calibration was to obtain the intrinsic parameters of the TAMLO optics and model the imaging process as a projection by an ideal thin lens along with distortion correction. The process is summarized here and will be further discussed in future work solely focused on calibration. First the WFOV system was calibrated using the well-established method in [18] so that it could be modeled as a pinhole camera with radial distortion correction. Because the apertures $A_{1,2}$ are in the same plane as A_0, it can be assumed that their representative pinhole models also lie in the same plane as the one for A_0. A planar checkerboard was then placed perpendicular to the optical axis of the TAMLO lens and at the working distance conjugate to the image sensor. This object field was captured by Π_W and $S_{1,2}$, as shown in Figs. 7(a) and 7(b), respectively. S_2 is like S_1 so it isn't shown. Using the WFOV calibration data, Π_W was undistorted (Π_{WU}), as shown in Fig. 7(c). Within the designed SFOV, corresponding image features outlined in red between Π_W and $S_{1,2}$ were determined so that the light rays in the SFOV system could be digitally bent by translating $S_{1,2}$, pixel by pixel, to the corresponding pixel coordinates that contain the matching image in Π_{WU}. In other words, the light rays from the SFOV system were digitally bent so that they focused with the calibrated chief rays in Π_{WU}. This is illustrated in Fig. 7(c), where the calibrated stereo image S_{1U} is directly overlapping Π_{WU} after digital bending and summed together for visualization. S_{1U} and the area of Π_{WU} underneath S_{1U} look the same, so the brightness is doubled after summation. This technique effectively removes both the distortion from the stereo images and the translation from the prism deflector and converts the TAMLO into a thin lens model. The amount of digital bending is stored for each pixel of $S_{1,2}$ in a lookup table for calibrating any subsequent stereo images. Although the lookup table was generated from a 2D object field, it applies to 3D object fields because each pixel of $S_{1,2}$ corresponds to unique object angles defined by the 3D object point location and $A_{1,2}$. The final step of this calibration was to determine the parameters of the ideal thin lens model. The image distance was already determined from the focal length of the WFOV system's pinhole model. The object distance to any point on the planar checkerboard placed at the conjugate working distance could be determined using the extrinsic parameters from the pinhole model. Knowing object and image distance, the effective focal length of the thin lens model was found from the thin lens equation. To find the baseline between the pinhole models of $A_{1,2}$, two object points at different depths were captured by $S_{1,2}$, which were then calibrated using the lookup table. First order ray tracing was performed from the two known object points to the stereo pinhole models of unknown baseline, refracted by the effective focal length, and then further traced to the corresponding image points in $S_{1U,2U}$. The baseline could then be algebraically solved. Conceptually, this calibration recovers an ideal thin lens model that obtains depth from defocus.

After calibration, fully processed TAMLO results of the tilted ruler and a 3D bladder model were generated to complete the proof of concept. The following results were rotated counterclockwise by 90° from the original image orientation so that the stereo views can be displayed with parallax
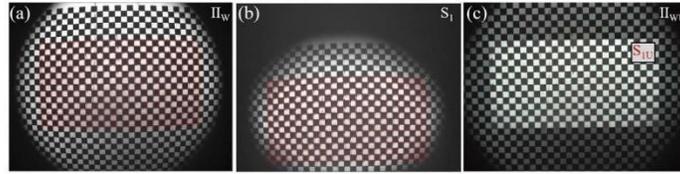
Fig. 7. Images of (a) WFOV and (b) SFOV before and (c) after thin lens modeling and distortion calibration. Features outlined in red and the image region labeled S_{1U} correspond to the same region of checkerboard squares in the object field.
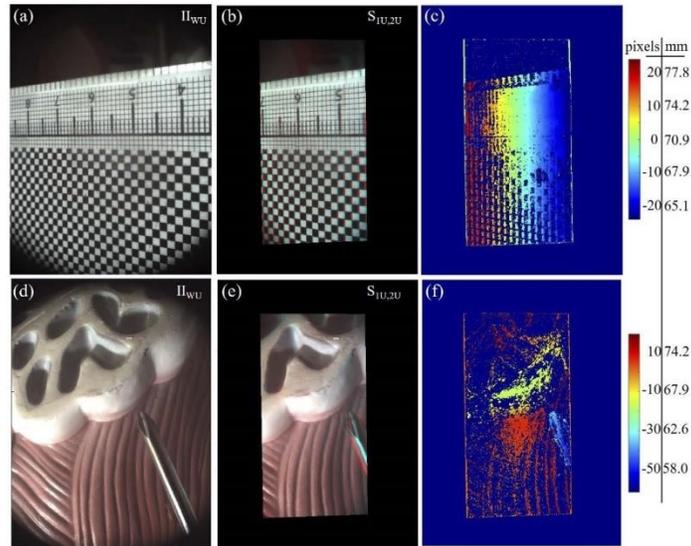

Fig. 8. Fully calibrated TAMLO results of a tilted ruler (top row) and a 3D bladder model (bottom row): (a, d) WFOV, (b, e) SFOV images overlaid as an anaglyph, (c, f) depth maps in units of pixel disparity and absolute depth.

along the horizontal direction and can be viewed with 3D glasses. Figures 8(a) and 8(d) show the undistorted WFOV images, as indicated by the straightened lines of the ruler and checkerboard. Figures 8(b) and 8(e) show the calibrated SFOV images overlaid as a red and cyan anaglyph, which demonstrates parallax based on the difference in disparity between corresponding object points. Close observation of Fig. 8(b) illustrates a reversal in the arrangement of the cyan and red colors from the 4.5 to 7 cm tick marks. This indicates the center of the image has zero disparity and is the conjugate working distance to the image sensor while the right and left of the image are closer and farther away, respectively. Similarly, Fig. 8(e) illustrates large disparity at the

129

screwdriver, indicating that it is much closer than the bladder model. Figures 8(d) and 8(e) demonstrate good image quality for both WFOV and SFOV imaging in a surgical setting. Finally, the calibrated stereo images were processed to produce accurate depth maps in Figs. 8(c) and 8(f). The color bars have units of pixel disparity, which were then converted to absolute depth values in millimeters as shown on the right of the color bars using the thin lens model parameters found during calibration. Although the original lens design had a 120 mm working distance, tolerances in the 3D printed resulted in a backward shift of the sensor, so the conjugate working distance or 0 pixel disparity in these figures is located at ~71 mm. According to Fig. 2, the depth resolution increases to ~1.5 mm at this closer working distance. Figure 8(c) shows the linear change in depth corresponding to the tilted ruler without any depth resolution artifacts, thus confirming depth mapping ability. Similarly, Fig. 8(d) shows the closer distance of the screwdriver and the correct surface profile of the bladder model.

6. Conclusion

In this paper, a novel prism-based tri-aperture monocular laparoscopic objective was conceptualized, designed, prototyped, and calibrated. This system achieved WFOV and SFOV imaging with sufficient image quality. Compared to the SFOV, the WFOV sees 2x the object field along the baseline axis. Overlapping crosstalk between the stereo images was also addressed. The calibration of the stereo views using the rotationally symmetric WFOV image as a reference was then introduced. Completion of the calibration enabled removal of distortion from the WFOV and SFOV images, which were then processed to generate accurate, absolute depth maps. The TAMLO certainly demonstrates the potential for optically combining WFOV and SFOV imaging in a compact system. Such a system may pave the way towards restoring the binocular and large, foveated FOV qualities of human vision within the minimally invasive surgical setting. In future work, the calibration details will be fully discussed, and the design of the relay lens group will be considered. Otherwise, the TAMLO can be used as a chip-on-tip system as demonstrated by our prototype. In either case, the system also needs to be configured with proper tri-aperture selector hardware and automated in software.

References

1. J. Geng and J. Xie, "Review of 3-D endoscopic surface imaging techniques," IEEE Sens. J. **14**(4), 945–960 (2014).
2. Intuitive, "Da Vinci Vision," https://www.intuitive.com/en-us/products-and-services/da-vinci/vision#%23.
3. Olympus, "ENDOEYE FLEX 3D," https://medical.olympusamerica.com/products/laparoscopes/endoeye-flex-3d.
4. S. Y. Bae, R. J. Korniski, M. Shearn, H. M. Manohara, and H. Shahinian, "4-mm-diameter three-dimensional imaging endoscope with steerable camera for minimally invasive surgery (3-D-MARVEL)," Neurophotonics **4**(1), 011008 (2017).
5. E. Kwan, Y. Qin, and H. Hua, "High resolution, programmable aperture light field laparoscope for quantitative depth mapping," OSA Continuum **3**(2), 194–203 (2020).
6. J. Liu, D. Claus, T. Xu, T. Keßner, A. Herkommer, and W. Osten, "Light field endoscopy and its parametric description," Opt. Lett. **42**(9), 1804–1807 (2017).
7. V. I. Batshev, A. Machikhin, and Y. Kachurin, "Stereoscopic tip for a video endoscope: problems in design," *Proc. SPIE 10466, 23rd Int. Symp. Atmos. Ocean Opt. Atmos. Phys.* **104664D**, (2017).
8. X. Cui, K. Bin Lim, Y. Zhao, and W. L. Kee, "Single-lens stereovision system using a prism: position estimation of a multi-ocular prism," J. Opt. Soc. Am. A **31**(5), 1074–1082 (2014).
9. S.-P. Yang, J.-J. Kim, K.-W. Jang, W.-K. Song, and K.-H. Jeong, "Compact stereo endoscopic camera using microprism arrays," Opt. Lett. **41**(6), 1285–1288 (2016).
10. Y. Qin, H. Hua, and M. Nguyen, "Multiresolution foveated laparoscope with high resolvability," Opt. Lett. **38**(13), 2191–2193 (2013).

11. Y. Qin, H. Hua, and M. Nguyen, "Characterization and in-vivo evaluation of a multi-resolution foveated laparoscope for minimally invasive surgery," Biomed. Opt. Express **5**(8), 2548–2562 (2014).

12. Y. Qin and H. Hua, "Optical design and system engineering of a multiresolution foveated laparoscope," Appl. Opt. **55**(11), 3058–3068 (2016).

13. Y. Qin and H. Hua, "Continuously zoom imaging probe for the multi-resolution foveated laparoscope," Biomed. Opt. Express **7**(4), 1175–1182 (2016).

14. J. I. Katz, S. Lee, and H. Hua, "Improved multi-resolution foveated laparoscope with real-time digital transverse chromatic correction," Appl. Opt. **59**(22), G79–G91 (2020).

15. D. T. Kim, C. H. Cheng, D. G. Liu, K. C. J. Liu, and W. S. W. Huang, "Designing a New Endoscope for Panoramic-View with Focus-Area 3D-Vision in Minimally Invasive Surgery," J. Med. Biol. Eng. **40**(2), 204–219 (2020).

16. E. Kwan and H. Hua, "Tri-Aperture Monocular Laparoscopic Objective for Stereoscopic and Wide Field of View Acquisition," in *OSA Imaging and Applied Optics Congress* (2021), p. 3Th2D.6.

17. J. E. Greivenkamp, *Field Guide to Geometrical Optics* (SPIE, 2004).

18. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000).

APPENDIX C: Calibration of Transverse Ray and Pupil Aberrations for Light Field Cameras

Elliott Kwan and Hong Hua

applied optics

Calibration of transverse ray and pupil aberrations for light field cameras

ELLIOTT KWAN AND HONG HUA*

3D Visualization and Imaging Systems Laboratory, James C. Wyant College of Optical Sciences, University of Arizona, 1630 E. University Blvd., Tucson, Arizona 85721, USA
*Corresponding author: hhua@optics.arizona.edu

The accuracy of reconstructing depth maps or performing digital refocusing in light field cameras depends largely on how well the spatial and angular samples of light rays can be obtained. Ray sample errors induced by optical aberrations in a light field camera may be digitally corrected using the ray tracing data when its nominal lens design is available. However, the most commonly nominal lens prescription is not accessible to end users. Additionally, even if available, due to tolerances in optomechanical design, the ray tracing data can be inaccurate. We propose a calibration method based on measurements of fiducial markers on a checkerboard for modeling the imaging properties of light field cameras. The calibration accounts for vignetting, transverse ray errors, as well as pupil aberration, and can be applied to light field camera modeling of arbitrary pupil sampling systems. We further demonstrate the utility of the method for calibrating a tri-aperture camera that captures simultaneous stereo views via artificially induced transverse ray errors.   © 2022 Optica Publishing Group

1. INTRODUCTION

Light field cameras capture the angular and spatial parameters of individual geometrical light rays traveling from the three-dimensional (3D) object scene to the entrance pupil of the objective lens. In comparison, conventional digital cameras integrate the same light rays or light field together at the sensor to form an image, resulting in a loss of angular information. Using a handheld digital light field camera [1], it was thoroughly demonstrated how the extra angular information from the light field could be used to implement useful and unique computational imaging techniques in post-processing, including digital refocusing and extended depth of field with higher signal-to-noise ratios.

The accuracy of reconstructing depth maps or performing digital refocusing in light field cameras depends largely on how well the spatial and angular samples of light rays can be obtained. The impact of ray aberrations in light field cameras has been investigated. Wei *et al.* demonstrated how precise addition of lens aberration can improve light field camera sampling and reduce aliasing across a range of focus distances [2]. However, when the aberrations are much more severe and complex, they certainly reduce the contrast of any image reconstructed from the light field camera. Although it is not technically an aberration, vignetting was also shown to impact light field camera calibration [3] and depth mapping error [4]. Because aberrations and vignetting alter the light field, they confuse light field depth reconstruction algorithms [5]. Therefore, all factors impacting the light field must be considered and calibrated for proper recovery of light field camera functionality.

There are a variety of existing aberration compensation methods originating from different optical technologies that sample their system's pupils. The main methods pertaining to light field aberration calibration come from work on the handheld digital light field camera [1,6–8]. Because the angular information of light rays is preserved, light field cameras have the inherent ability to digitally calibrate known aberrations in their lens design. This was demonstrated by digitally resorting the captured aberrated light field based on the deviated ray paths calculated from knowledge of the nominal lens design. The major drawback of this approach appears in more complex as-built optical systems, where the nominal lens design may not be available and the optomechanical tolerances may result in unknown combinations and magnitudes of aberrations. This aberration calibration also would not work for light field cameras built with an unknown optical system, such as a prototype using a commercial lens. Therefore, there is a need for a technique that can measure the aberrations of the as-built system to properly calibrate them.

One example is in the field of coded aperture imaging, where the optical system's pupil function was estimated by taking multiple image measurements of a test target with a set of strategically defined pupil plane masks [9]. The estimated pupil function was then converted into the system point spread function, which enabled removal of aberrations via deconvolution. However, this approach requires sampling the pupil plane

with overlapping masks, which is unconventional in light field cameras. Instead, aberrations resulting from using different immersion media in a light field microscope were measured and corrected by analyzing the captured light field of a grid of guide stars or gray code sequences [10]. Similarly, to produce accurate depth map estimation in light field cameras, a planar checkerboard object has been used as the reference to measure and calibrate field curvature and astigmatism [11]. However, these calibrations consider only transverse aberration correction, while our calibration expands upon this to include pupil error for complete restoration of an as-built aberrated light field camera.

After correcting for aberrations, light field cameras require further calibration for accurate 3D measurements. This includes decoding the raw light field data, developing models for the light field camera and lens distortion, and then calibrating the real camera by estimating the parameters of the representative model such that the errors between the model and real data are minimized [12–14]. These calibrations, among others, have been thoroughly demonstrated, compared, and improved upon. Some of the ideas and techniques from these calibrations are utilized in ours. This includes conventional monocular camera calibration [15] and vignetting correction [16]. However, our calibration does not currently compete with the prior ones. Instead, our contribution is our calibration's application towards the development of novel as-built light-field-based cameras, to the best of our knowledge.

The incorporation of robust aberration calibration in light field cameras can help reduce the image performance requirements of their optics, and therefore also the complexity and cost of these systems. For instance, this would be useful in rapid, inexpensive prototyping of light field cameras with stock lens design and 3D printing of optomechanical housing, where the loss in image quality due to limited lens choices and larger mechanical tolerances could be recovered with aberration calibration. Alternatively, this type of calibration could be used to enable unique camera designs that capture the light field in novel ways. In this paper, we present a measurement-based calibration algorithm that minimizes the aberrations of an as-built light field camera system such that it can be modeled as a virtual thin lens light field camera for restoring depth information. Section 2 begins describing this theoretical model for a non-aberrated, arbitrary pupil sampling optical system. Then we discuss how vignetting, transverse ray aberration, and pupil aberration in the real system are accounted for through measurement and generation of lookup tables. Section 3 describes the calibration procedure in practice for recovering relative refocusing and absolute depth mapping capabilities. Section 4 demonstrates experimental results of the calibration procedure by applying it to a tri-aperture camera that captures samples of the light field with artificially induced transverse ray error.

2. MATHEMATICAL MODEL FOR LIGHT FIELD CAMERA CALIBRATION

A light field camera angularly samples the geometrical light rays of a 3D scene according to the directions of the rays apparently emitted by the scene, which leads to the well-known 4D light field function, $L(u, v, s, t)$, for representing the ray radiance

as a function of ray position (s, t) and direction (u, v) [17]. Among the various existing light field cameras, there exist two fundamental architectures—Light Field (LF) 1.0 (a.k.a. plenoptic 1.0) and Light Field 2.0 (a.k.a. plenoptic 2.0). In cameras based on the LF 1.0 architecture, an aperture array or a micro-lens array (MLA) is typically placed at the back focal plane of an objective lens and a detector is placed at the position that is optically conjugate to the stop of the objective lens. In this architecture, the MLA pitch determines the spatial resolution and forms the ST-plane and the sensory pixel pitch determines the view resolution and forms the UV-plane. In cameras based on the LF 2.0 architecture, the aperture array or the MLA is placed at a position conjugate to the main stop of the objective, and the detector is placed at a location optically conjugate to the intermediate image plane through the MLA. In this architecture, the MLA pitch determines the view resolution and forms the UV-plane, and the sensor pixel pitch determines the spatial resolution and forms the ST-plane. The mathematical model and calibration process discussed below assumes an LF 2.0 architecture, but it is generally applicable to LF 1.0 architecture with adaptation.

A. Subaperture Images in Light Field Camera

Figure 1 illustrates an MLA-based LF 2.0 camera design, consisting of a main lens, an MLA, and a sensor [18]. The light field of a 3D object is captured by the main lens, forming an intermediate image plane. The MLA is placed behind the intermediate image of the main lens, and the sensor is placed at a location optically conjugate to the intermediate image plane. Each lenslet of the MLA sees a different viewing angle of the intermediate image and forms a corresponding elemental image (EI) on a different portion of the sensor. Because the lenslet apertures limit the ray bundles of each viewing angle, the stop and exit pupil are defined at the MLA plane, which is optically conjugate to the entrance pupil of the main lens.

In the configuration in Fig. 1, the object's light field is sampled spatially by the EIs on the sensor and angularly by the MLA, corresponding to the ST-plane and UV-plane for ray position and direction sampling in the 4D light field function, respectively. Each of the lenslet apertures can be mapped to a subaperture location on the entrance pupil, and each of the EIs can be mapped to a portion of the intermediate image. For instance, the light field of point P in Fig. 1 is captured through three subapertures, A_0, A_1, and A_2, on the entrance pupil or equivalently through three lenslets, M_0, M_1, and M_2. It is imaged onto the same point p_{ii} on the intermediate image plane, and onto three spatially separated pixels, p_0, p_1, and p_2, on three EIs, EI_0, EI_1, and EI_2. Once a light field is captured on
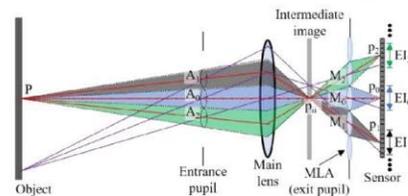

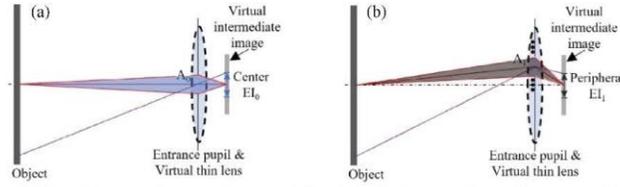**Fig. 1.** Working principle of light field camera 2.0.

Fig. 2. Light field reorganized into (a) center subaperture image and (b) peripheral subaperture images from a virtual thin lens model.

the sensor, the raw image can be reorganized into individual EIs according to these conjugate planes for post-processing and 3D reconstruction.

For the purpose of calibration and further depth reconstruction, the image formation process shown in Fig. 1 can be simplified into a model consisting of a virtual thin lens located on the entrance pupil and an array of subapertures located on the virtual thin lens. The subaperture locations can be found by tracing the subaperture chief rays through the lenslet centers of the MLA or the equivalent array elements in a light field camera system. We can then map each of the raw EIs on the sensor plane onto an equivalent EI on the image plane of the virtual thin lens. Figures 2(a) and 2(b) show the mapping of the EI_0 and EI_1 captured by the lenslets M_0 and M_1 through the subapertures A_0 and A_1 onto the intermediate image plane of the virtual main lens, respectively. In the mapping process, the EIs on the sensor plane are repositioned onto the virtual intermediate image plane such that their matching image pixels (such as p_0 and p_1) from the sensor plane are now virtually overlapping. The virtual overlapping represents the expected image formation from the virtual thin lens. During the mapping, a unit magnification is assumed between the raw EIs on the sensor plane and the equivalent EIs on the intermediate image plane, except for the sign from flipping the EIs to match the orientation of the intermediate image. The equivalent focal length of the virtual thin lens accounts for the difference in magnification between the raw EIs and their corresponding intermediate image induced by the lenslet magnification.

For generalization, these equivalent EIs obtained through remapping will be referred to as subaperture images for the remainder of the paper. We calibrate the light field system based on this subaperture image setup. Note that the raw light field data from an LF 1.0 camera design [1] can be reorganized into these subaperture images in an analogous fashion. Also note that the raw light field data from a programmable aperture-based light field camera design [19] is already organized according to the subaperture image setup. Therefore, our proposed calibration method is applicable to any pupil sampling or light field camera design after the raw data are organized into subaperture images.

B. Computational Model for Depth Reconstruction

To reconstruct the depth map of a 3D scene from recorded subaperture images, the parameters of the virtual thin lens as well as the subaperture locations need to be determined, as shown in Fig. 3; f is the equivalent focal length of the virtual thin lens, Z'
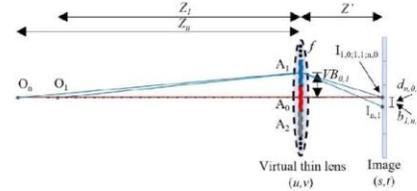

Fig. 3. Ray tracing for two known object points through two subapertures.

is the distance between the virtual thin lens and the subaperture images, and A_i is the center location of the ith subaperture, where $i = 0$ corresponds to the aperture centered with the optical axis of the virtual lens. $VB_{i,j}$ represents the baseline distance between two subapertures, A_i and A_j. Here A_i corresponds to (u, v) coordinates, while the pixels of the subaperture images correspond to (s, t) coordinates in the 4D light field function, $L(u, v, s, t)$.

O_1 is defined as an object point optically conjugate to the image plane, while O_n is an arbitrary point at a different depth than O_1. The depths of these object points are denoted as Z_1 and Z_n, respectively. Their images by the ith subaperture onto its corresponding subaperture image are denoted as $I_{1,i}$ and $I_{n,i}$, respectively. For instance, rays from O_1 and O_n going through the center subaperture A_0 highlighted in red are imaged onto the overlapping points $I_{1,0}$ and $I_{n,0}$, respectively. The ray from O_1 going through the subaperture A_1 highlighted in blue is imaged onto $I_{1,1}$, which is also overlapping the previous two points according to image formation of the conjugate object. The ray from O_n going through A_1 is imaged onto $I_{n,1}$. The pixel displacement between $I_{n,0}$ and $I_{n,1}$ for the point O_n is defined as the disparity of the subaperture images captured by subapertures A_0 and A_1. To generalize, the image disparity of the point O_n between the ith and jth subaperture images, denoted as $d_{n,i,j}$, is defined as the pixel displacements between the subaperture image points $I_{n,i}$ and $I_{n,j}$.

For simplicity, let us consider a pair of subapertures A_i and A_j along the t axis, where A_i is the lower subaperture and A_j is the higher subaperture. The resulted image disparity for point O_n is also along the t axis, and is calculated as $d_{n,i,j} = t_{n,i} - t_{n,j}$. The disparity $d_{n,i,j}$ is positive when $Z_n > Z_1$, zero when $Z_n = Z_1$, and negative when $Z_n < Z_1$. By ray tracing minimally through two subapertures, the depth of O_n can be reconstructed by

135

$$Z_n = \left(-\frac{d_{n,i,j}/VB_{i,j} + 1}{Z'} + \frac{1}{f} \right)^{-1}. \tag{1}$$

The depth reconstruction in Eq. (1) uses disparity information only along the t axis. For depth reconstruction from subapertures and disparity along the orthogonal axis, the process is analogous.

C. Calibration from Center and Peripheral Subaperture Systems

To obtain the calibration parameters, the first step is to utilize the chief rays of the fields going through the center subaperture in Fig. 2(a). These data are treated as light passing through a conventional pinhole camera model at A_0 in Fig. 3. The image projection process is described by

$$s\,I_{n,0} = K\,[R\;T]\,O_n, \tag{2}$$

where s is a scale factor, $I_{n,0}$ is the 2D image point, K is the camera intrinsic projection matrix, $[R\;T]$ are the rotation and translation extrinsic parameters to relate the world coordinate system to the camera coordinate system, respectively, and O_n is the 3D object point defined in the world coordinates. Solving for K, R, and T is achieved using conventional monocular camera calibration [20]. The x and y focal lengths, f_x and f_y, of the pinhole camera are extracted from K to determine the distance from the pinhole to the sensor along the normal connecting them, or Z' in Fig. 3. $Z' = f_x$ for disparity analysis along the s axis, and $Z' = f_y$ for disparity analysis along the t axis.

The calibrated center subaperture system (CCSS) is then utilized to calculate the distance from the representative pinhole model to the conjugate object plane, or Z_1 in Fig. 3. A flat checkerboard object is placed both parallel to the sensor and at the best object focus position based on the defocus cues from the finite sized A_0. To assess parallelism, the checkerboard is imaged by the CCSS, and given the known checkerboard dimensions, its extrinsic parameters in relation to the calibrated pinhole model are calculated in closed form [15]. R is assessed to determine if the checkerboard requires any alignment adjustments to achieve parallelism. After alignment, the extrinsic parameters are calculated once more. To determine Z_1, the image point corresponding to the optical axis of A_0 is found from K and converted to camera coordinates using Eq. (2).

Next, the determined parameters Z_1 and Z' are used to calculate focal length f. Because the peripheral subaperture samples were physically acquired on the same plane as the center subaperture, it is assumed that their respective pinhole models are also located at the same (u, v) plane as the center subaperture pinhole model in Fig. 3. The peripheral pinhole models at A_i must also incorporate ray bending such that the location of the peripheral subaperture images along the image plane is properly predicted. The combined set of pinhole models validates and forms the virtual main lens with focal length f, which is solved for by using the thin lens equation

$$\frac{1}{f} = \frac{1}{Z_1} + \frac{1}{Z'}. \tag{3}$$

Finally, the distance between two subapertures is determined by $VB_{i,j} = |VB_{0,i} - VB_{0,j}|$, where the total distance is split

into two segments, each defined by A_0 and A_i as shown in Fig. 3. To define each segment, we consider two calibration checkerboards located at the depths of Z_1 and Z_n and two respective feature points O_1 and O_n, one on each of the checkerboards. The camera coordinates of these two feature points are determined by applying the CCSS to the corresponding image points from A_0. Then we illustrate in Fig. 3 that for the same two feature points, the pixel separation, $b_{1,n,i}$, between the corresponding image points from the ith subaperture is a function of $VB_{0,i}$ and the distance between O_1 and O_n. This relationship is determined by combining two equations obtained through ray tracing from O_1 and O_n through A_i to the image plane and algebraically solving for $VB_{0,i}$:

$$VB_{0,i} = -\frac{\frac{b_{1,n,i}}{Z'} + \frac{Y_n}{Z_n} - \frac{Y_1}{Z_1}}{-\frac{1}{Z_n} + \frac{1}{Z_1}}, \tag{4}$$

where $b_{1,n,i} = \pm(t_{n,i} - t_{1,i})$ is the amount of separation along the t axis between the two image points. For sign convention, if Z_n is larger than Z_1, the positive sign is used and vice versa. Y_1 and Y_n are the y-camera coordinates of the feature points. For subaperture sampling distance along the orthogonal axis, the process is analogous.

This calibration method achieves the highest depth mapping accuracy for object depths near Z_1 and Z_n because these are the reference depths used for determining the calibration parameters. Z_1 is set by the optical conjugates of the camera, while Z_n can be chosen based on the application to maximize depth mapping confidence at a second region. Since Eqs. (1) and (4) were defined based on the assumption that the entrance pupil is at the virtual lens, the amount of depth mapping error for objects away from Z_1 and Z_n is then dependent on the amount of separation between the entrance pupil and virtual thin lens that would more precisely model the real system.

The theory in the above sections assumes the real system performs according to ideal, first-order imaging. The following subsections address how real factors impacting the light field are accounted for during calibration.

D. Vignetting

To maximize the usability of the captured light field information, vignetting is mitigated. A flat Lambertian surface is used to fill the full field of view (FOV) of the camera. The image should ideally have uniform irradiance, so any vignetting can be quantified and calibrated. Without vignetting, here the light field would be $L(u, v, s, t) = c$, where L is the radiance and c is a constant due to the Lambertian property. With vignetting, $L_v(u, v, s, t) \leq c$. In a real system, we can assume the center subaperture A_0 image has the minimum amount of vignetting, where $L(0, 0, s, t) = c$ for field points with equal to or less than half vignetting because the chief rays are undisturbed. Using the A_0 image as the reference, vignetting $V(u, v, s, t)$ for each peripheral subaperture image is quantified by

$$V(u, v, s, t) = L_v(u, v, s, t)/L(0, 0, s, t). \tag{5}$$

For subsequently recorded light fields, $L_v(u, v, s, t)$ is divided by $V(u, v, s, t)$ to recover the non-vignetted $L(u, v, s, t)$. For field points outside of the half vignetted FOV,

vignetting is corrected only up to the amount of vignetting in the A_0 image.

E. Transverse Ray Aberrations

Distortion from the main lens is removed first in preparation for correction of transverse ray aberrations. Because distortion is characterized by the intersections of the chief rays at the image plane, the amount of main lens distortion can be observed in a raw center subaperture A_0 image. Since the A_0 system is rotationally symmetric, its distortion can be represented by a radial polynomial model:

$$s_d = s + s[k_1(s^2 + t^2) + k_2(s^2 + t^2)^2]$$

$$t_d = t + t[k_1(s^2 + t^2) + k_2(s^2 + t^2)^2]. \qquad (6)$$

The distortion coefficients k_1 and k_2 are determined simultaneously during the calculation [15] of the intrinsic and extrinsic parameters of the CCSS from Section 2.C. Equation (6) is then used to convert $L(0, 0, s_d, t_d)$ to the undistorted A_0 image $L(0, 0, s, t)$. During transverse ray aberration calibration, the peripheral subaperture images will be referenced to the undistorted chief rays in $L(0, 0, s, t)$, resulting in the removal of distortion from the entire captured light field.

In the presence of transverse ray aberration from the main lens, the light rays from the peripheral subapertures will intersect the image plane at different locations according to the transverse ray errors $\varepsilon_s(u, v, X, Y)$ and $\varepsilon_t(u, v, X, Y)$. ε_s and ε_t vary with subaperture location (u, v) and with object point location (X, Y) in the object plane conjugate to the image plane. Figure 4 illustrates the error by replacing A_1 with a thin prism to represent additional ray bending from transverse ray aberration. The dashed blue ray is from ideal ray bending while the solid blue ray is from real ray bending. For a conjugate object point O_1, ε_s and ε_t are defined as the distance from the chief ray intersection or center subaperture image point $I_{1,0}$ to the corresponding peripheral subaperture image point $I_{1,i\varepsilon}$. This is illustrated by $I_{1,1\varepsilon}$ in Fig. 4. The ideal subaperture image point $I_{1,i}$ can be recovered by digitally moving $I_{1,i\varepsilon}$ by $-\varepsilon_s$ and $-\varepsilon_t$. Since ε_s and ε_t are now defined at the image plane, they will be reparameterized as $\varepsilon_s(u, v, s_\varepsilon, t_\varepsilon)$ and $\varepsilon_t(u, v, s_\varepsilon, t_\varepsilon)$, where $(s_\varepsilon, t_\varepsilon)$ are transverse ray aberrated image space coordinates.

Using a checkerboard at the object plane of O_1, the errors can be measured for each corner point to build $\varepsilon_s(u, v, s_\varepsilon, t_\varepsilon)$ and $\varepsilon_t(u, v, s_\varepsilon, t_\varepsilon)$ across the full FOV. The measurement is accomplished by comparing each peripheral subaperture image $L(u, v, s_\varepsilon, t_\varepsilon)$ to the center subaperture image $L(0, 0, s, t)$ and quantifying the ε_s and ε_t of corresponding subaperture

image points. For peripheral subaperture image points without a corner feature, ε_s and ε_t are linearly interpolated from the nearest four corners. The aberrated light rays can then be digitally bent in post-processing by warping the peripheral subaperture images according to ε_s and ε_t to recover $L(u, v, s, t)$. This digital ray bending process is analogous to aberration correction in traditional optical lens design, where lens design software optimizes and minimizes transverse ray error by converging all light rays to their respective chief ray at the image plane.

For the calculation of Z_1 in Section 2.C, the checkerboard is supposed to be placed exactly at the real conjugate object plane. In practice, it may be slightly displaced, so transverse ray error calibration is also used to account for this. If the image plane location remains the same, the displacement will introduce additional transverse ray error between corresponding subaperture image points due to defocus. This error is added to the errors from the other aberrations and will be corrected simultaneously. In other words, a point O_1 slightly displaced from distance Z_1 will digitally focus at Z' after transverse ray error calibration. As a result of the additional digital ray bending, the virtual lens focal length f is slightly adjusted according to Eq. (3), where Z_1 is updated based on the displacement of O_1. Therefore, the same placement of the checkerboard is used for transverse ray error calibration and calculation of the effective Z_1.

The monochrome aberrations have been addressed, and transverse ray error from chromatic aberrations can be dealt with in a similar fashion. In this case, all subaperture images from the red, green, and blue channels separated by the Bayer filter would be compared to the green center subaperture image and warped accordingly. This type of color correction has been analogously demonstrated in [21].

F. Pupil Aberrations

Although the transverse ray aberrations for the virtual main lens are corrected, pupil aberrations may still exist and impact the captured light field, as illustrated in Fig. 5. Pupil aberration is analogous to transverse ray aberration, except that deformation from the entrance to exit pupil is observed instead of deformation from the object to image plane. Points O_1 and O_3 are on the object plane conjugate to the image plane. A real ray from object point O_2 passes through the point where O_3 would be and then through a section of the real entrance pupil. It is then bent by the real main lens, which results in a translation Δp of the transmitted ray due to pupil aberration as shown at A_1 of the virtual main lens. Because there is no transverse ray aberration here and
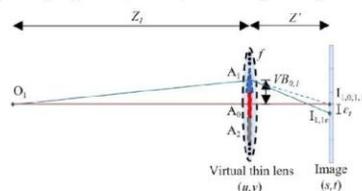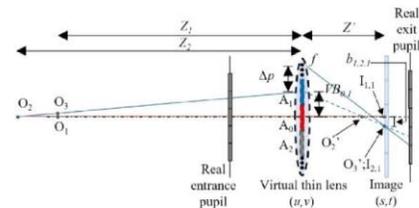

Fig. 4. Transverse ray error illustrated by replacing a peripheral subaperture with a thin prism to represent additional ray bending.

Fig. 5. Ray trace in the presence of pupil aberration and zero transverse ray aberration.

the ray appears to have originated from O_3, the pupil aberrated ray intersects the conjugate image point location O'_3 or the same $I_{2,1}$ from Fig. 3. The aberration-free ray indicated by the dashed blue line also intersects O'_3. These two rays are extended to the real exit pupil to further illustrate their differences due to pupil aberration. The aberration-free ray passes through the image point O'_2 conjugate to O_2 while the pupil aberrated ray does not. This indicates that digital refocusing or ray tracing to the correct image point or depth will be inaccurate with pupil aberrated light field data. A ray from O_1 going through A_1 will still land at $I_{1,1}$.

After transverse ray aberration correction from Section 2.E, pupil aberration is calibrated directly during the calculation of $VB_{0,i}$ from Eq. (4). Figure 5 shows that the separation $b_{1,2,1}$ is unaffected in the presence of pupil aberration without transverse ray aberration. Based on this property, the first-order, aberration-free ray can be restored by calculating $VB_{0,i}$ as normal. Since pupil aberration is radially dependent, a light field uniformly sampled at the exit pupil becomes nonuniformly sampled at the entrance pupil and virtual thin main lens. Thus, $VB_{0,i}$ must be calculated for each subaperture A_i to restore the aberration-free virtual main lens.

3. CALIBRATION PROCEDURE FOR ABERRATED LIGHT FIELD CAMERAS

In practice, aberrated light field camera calibration follows the flowchart in Fig. 6. If vignetting is observed in the camera's captured light field data, vignetting correction is first applied. Subaperture images of a flat Lambertian surface filling the full FOV are acquired to calculate $V(u, v, s, t)$ according to Eq. (5), which is then used to unvignette subsequently recorded light fields by division.

Then, center subaperture A_0 is modeled as a pinhole camera via monocular camera calibration. The CCSS is generated by capturing and processing a sufficient set of subaperture images from A_0, where each image is of an arbitrarily oriented checkerboard. For subsequently recorded subaperture images from A_0, distortion correction is performed using the CCSS's estimated distortion coefficients and Eq. (6).

Next, transverse ray error correction is performed. A checkerboard is placed both parallel to the sensor and at the best object focus position Z_1 based on the defocus cues from the finite sized A_0. Using the CCSS, parallelism is achieved by alignment of the checkerboard until the calculated extrinsic rotation matrix R is equal to the identity matrix. Aberrated peripheral subaperture images $L(u, v, s_\varepsilon, t_\varepsilon)$ and the undistorted center subaperture

image $L(0, 0, s, t)$ of the checkerboard are then acquired. They are compared to quantify $\varepsilon_s(u, v, s_\varepsilon, t_\varepsilon)$ and $\varepsilon_t(u, v, s_\varepsilon, t_\varepsilon)$, which are used to recover transverse ray error corrected light fields $L(u, v, s, t)$ by digitally warping subsequently recorded $L(u, v, s_\varepsilon, t_\varepsilon)$.

In applications where only relative refocusing is needed, the calibration process can end here. If pupil aberration is absent or minimal, refocusing capability is recovered after transverse ray error correction. If pupil aberration is significant, refocused images at depths away from the image plane at Z' may have lower image quality, and the relative depth information may be slightly altered. To avoid this and recover absolute depth mapping capability, the full calibration process is required.

Continuing with the calibration, the thin lens model parameters are calculated. First, Z' is determined by the pinhole focal lengths from the intrinsic matrix K of the CCSS. Second, Z_1 is determined using the same undistorted center subaperture image $L(0, 0, s, t)$ of the checkerboard from transverse ray error correction. The image point corresponding to the optical axis of A_0 is found from K and converted to camera coordinates $(0, 0, Z_1)$ using Eq. (2). Third, virtual main lens focal length f is then calculated using Eq. (3).

Finally, pupil aberration correction and subaperture sampling calculation are performed. The checkerboard is translated away from Z_1 to Z_2. The checkerboard parallelism to the sensor is checked again by observing R. The translation must be large enough such that the separation $b_{1,2,i}$ is detectable by the sampling of the sensor. Since distortion and transverse ray error have already been quantified in a lookup table, all subaperture images for this new checkerboard position are then acquired and warped accordingly. Object points O_1 and O_2 are found by selecting two corner image points, $I_{1,0}$ and $I_{2,0}$, where each one is closest to the optical axis of A_0 in the two A_0 images of the checkerboard at Z_1 and Z_2. By selecting the two object points in the center of the FOV and close in (X, Y), depth mapping errors can be minimized in that region. Arbitrary points farther apart could be used, but this can introduce errors due to the entrance pupil at the thin lens assumption. Using the CCSS, the camera coordinates of O_1 and O_2 are found. Then the corresponding peripheral subaperture image points $I_{1,i}$ and $I_{2,i}$ are located. Separation $b_{1,2,i}$ is calculated from the pixel coordinates of these image points. $VB_{0,i}$ is then calculated for each peripheral subaperture A_i using Eq. (4).

With all the calibration parameters and lookup tables determined, any pair of raw subaperture images from this aberrated light field camera can be corrected for vignetting, distortion, and transverse ray error. The corrected pair of subaperture images is
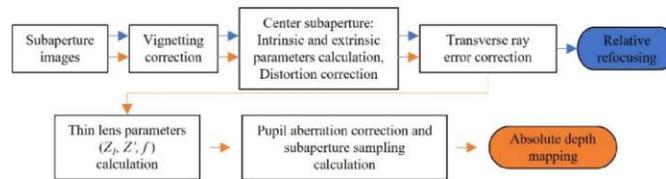

Fig. 6. Flowchart of aberrated light field camera calibration for recovering relative digital refocusing capability and absolute depth mapping capability.

then processed to generate a disparity $d_{n,i,j}$ map across the full FOV, which is converted to an absolute depth map using the known calibration parameters and Eq. (1).

4. APPLICATION AND EXPERIMENTAL RESULTS

Using a prism-based tri-aperture camera we recently developed for surgical applications [22], this section shows the full calibration process to recover relative refocusing and absolute depth mapping. It also simultaneously shows how the calibration could enable light field imaging in novel camera designs, to the best of our knowledge, that capture light rays with artificially induced transverse ray aberration.

A. Prism-Based Tri-aperture Camera

Figure 7 shows the conceptual model of a prism-based tri-aperture laparoscopic objective [22]. This system presents an alternative method of capturing portions of the light field. Its purpose is to provide the surgeon with stereo vision (SFOV), absolute depth mapping, and an additional wide FOV (WFOV) to enhance surgical awareness. Because digital refocusing is not necessary here, full light field capture is not required. Instead, this system acquires light rays from only the center A_0 and two peripheral A_1 and A_2 subapertures at the stop of the objective. In the optical design of this system, a prism is introduced adjacent to the stop to simultaneously capture the peripheral stereo subaperture images S_1 and S_2. The overlapping center subaperture wide view image II_W is captured separately with a time or spatially multiplexed tri-aperture selector. Notice that the ray bending from the prism here is analogous to Fig. 4, where a thin prism was used to represent additional ray bending from transverse ray error. In other words, the real prism in this system artificially introduces transverse ray error to translate S_1 and S_2 onto opposite halves of the sensor. Compared to conventional stereo cameras with two individual lens systems, calibrating the distortion in these stereo views without additional reference information is not straightforward because there is no rotational symmetry. There's also additional distortion due to the finite thickness of the prism deflector. Instead, the rotationally symmetric II_W can then be utilized to remove the artificial transverse ray error, remove the distortion from the stereo views, and
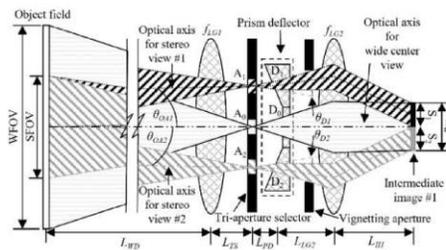

Fig. 7. Conceptual model of a prism-based tri-aperture laparoscopic objective for wide and stereo field of view image acquisition [22]. The prism deflector introduces artificial transverse ray error, analogous to the thin prism in Fig. 4.

obtain the calibration parameters for absolute depth mapping according to our full light field calibration process in Fig. 6.

B. Vignetting

First, vignetting correction was skipped in the full calibration because the prototype was designed with zero vignetting across the desired SFOV. However, vignetting correction can still be demonstrated for object points outside of the desired SFOV. Figures 8(a) and 8(b) show raw subaperture images S_2 and II_W of a flat Lambertian surface, respectively. V was calculated by dividing S_2 by II_W, pixel by pixel, according to Eq. (5). Figure 8(c) shows S_2 of an object scene, and the number 9 along the ruler is vignetted. V is ∼0.23 in this region of the image. To correct for this, S_2 is divided by V, pixel by pixel, to recover the unvignetted S_{2UV} in Fig. 8(d). The number 9 is now clearly visible after the removal of vignetting. For the rest of Section 4, raw subaperture images are used directly since vignetting correction is unnecessary for this prototype.

C. Center Subaperture Image Calibration

Next, the CCSS was generated by capturing and processing 18 II_W images from A_0. Each image was from a random orientation of a checkerboard. The estimated intrinsic and extrinsic properties and the distortion coefficients of the pinhole model resulted in an overall mean reprojection error of 0.46 pixels. Next, the checkerboard was placed both parallel and conjugate to the sensor, as shown in II_W of Fig. 9(a). R is approximately equal to the identity matrix, indicating parallelism:

$$R = \begin{bmatrix} 1.0000 & -0.0051 & 0.0048 \\ 0.0049 & 0.9985 & 0.0542 \\ -0.0050 & -0.0542 & 0.9985 \end{bmatrix}. \tag{7}$$

Next, II_W was undistorted as shown in II_{WU} of Fig. 9(b) using the CCSS. For instance, before correction, the distortion is −7.3% at the top right checkerboard corner feature highlighted in red. After correction, the distortion is negligible as indicated by the overlapping red straight lines. II_{WU} represents the undistorted chief rays.
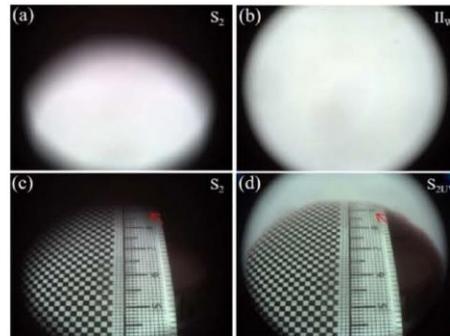

Fig. 8. Raw subaperture images (a) S_2 and (b) II_W of a flat Lambertian surface. S_2 of an object scene (c) before and (d) after vignetting correction.

D. Transverse Ray Error Correction

The checkerboard placement remained unchanged, and S_1 and S_2 were captured from A_1 and A_2, as shown in Figs. 10(a) and 10(b), respectively. The grid of image points highlighted in red in each subaperture image of Figs. 9 and 10 correspond to the same group of object points. The vertical translation of the red grid between II_W, S_1, and S_2 illustrates the amount of artificially induced transverse ray error. The grid in S_1 and S_2 is translated ∼400 pixels up and down, respectively. The difference in distortion is also clearly observed. For II_W, it is rotationally symmetric, while for S_1 and S_2, it is bilaterally symmetric. S_1 and S_2 were compared to II_{WU} to quantify ε_s and ε_t. Then, S_1 and S_2 were digitally warped and translated accordingly to simultaneously remove their distortion and induced transverse ray error. Since
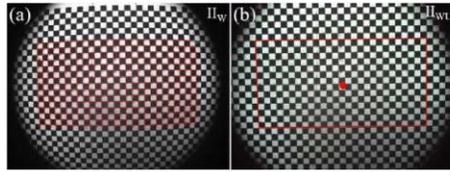

Fig. 9.   Checkerboard placed both parallel and conjugate to the sensor. (a) Raw center subaperture image II_W and (b) undistorted center subaperture image II_{WU}.
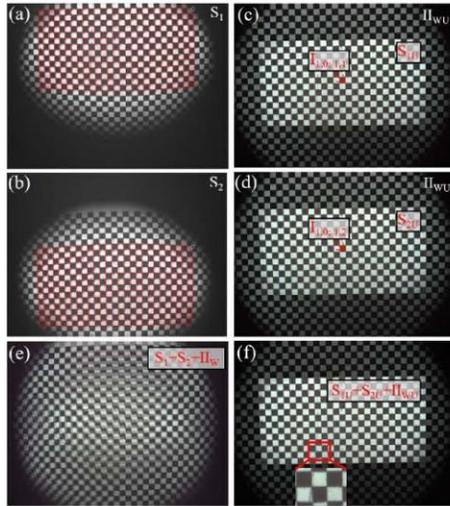

Fig. 10.   Checkerboard placement unchanged from Fig. 9. Raw peripheral subaperture images (a) S_1 and (b) S_2. The image points highlighted in red in each subaperture image of Fig. 9 and this figure correspond to the same group of object points. The translation of the red grid between II_W, S_1, and S_2 illustrates the amount of artificially induced transverse ray error. Distortion and transverse ray error corrected images (c) S_{1U} and (d) S_{2U} using II_{WU} as the reference chief rays. Full aperture images (e) before and (f) after transverse ray error correction.

the checkerboard for these images is at the conjugate object depth Z_1, corresponding subaperture image points should be overlapping. This is illustrated by the transverse ray error corrected images S_{1U} and S_{2U} of Figs. 10(c) and 10(d), respectively. II_{WU} is in the background while S_{1U} and S_{2U} overlap the corresponding region, indicating that the peripheral subaperture rays now focus together with their corresponding chief rays after transverse ray error correction. The amount of residual transverse ray error across the full FOV after correction is within ∼ 2 pixels, depending on the accuracy of the feature recognition algorithm used to pinpoint the checkerboard corners and the interpolation of transverse ray error for object points in between checkerboard corners. For example, the zoomed view in Fig. 10(f) shows one of the squares was slightly miswarped.

For evaluation, image contrast of full aperture images can be quantified before and after transverse ray error correction. As illustrated in Fig. 7 and digitally generated in Fig. 10(e), the pre-corrected full aperture image is when S_1, S_2, and II_W are averaged together at the image plane. Here, the contrast at the checkerboard frequency is effectively zero because the artifacts due to the aberration result in aliasing. The post-corrected full aperture image in Fig. 10(f) looks equivalent to its subaperture image components, as expected. Full aperture imaging has been restored with an image contrast around 49%. Individually, subaperture images S_{1U}, S_{2U}, and II_{WU} have image contrast values of 46%, 46%, and 59%, respectively. This indicates that the full aperture image is limited by the image quality of its individual subaperture images.

E. Calibration of Thin Lens Parameters

The thin lens model parameters were calculated next. Since the disparity in this prototype occurs only along the t axis, $Z' = f_y$. From the intrinsic matrix K of the CCSS, $Z' = 3413.4$ pixels or 7.506 mm based on the sensor's 2.2 µm pixels. The effective focal length of the real lens design is 7 mm, so Z' is appropriate for the object distance in Fig. 9(a). To determine Z_1, the image point corresponding to the A_0 optical axis was found from K, as shown by the red point in Fig. 9(b). Using the CCSS, the camera coordinate of this image point was calculated as $(0, 0, Z_1) = (0, 0, 71.253)$ mm. Out of the entire FOV, the maximum and minimum object depths were also calculated to be 72.214 mm and 69.935 mm, respectively, indicating a slight tilt of the checkerboard. Z_1 is in the middle of this range and approximately represents all object points across the FOV, so some of this tilt error may be carried over to the depth mapping results. Then, from Eq. (3), $f = 6.79$ mm.

F. Subaperture Sampling

To perform the pupil aberration correction and subaperture sampling calculation, the checkerboard was placed ∼1 cm in front of the conjugate object depth to minimize depth mapping error around this range. The three raw subaperture images were acquired for this new object depth and warped accordingly for distortion and transverse ray error correction. Figures 11(a) and 11(b) show the corrected images S_{1U} and S_{2U}, respectively, fused together with II_{WU}. Comparing the two figures, disparity due to defocus is observed in opposite directions relative to II_{WU},
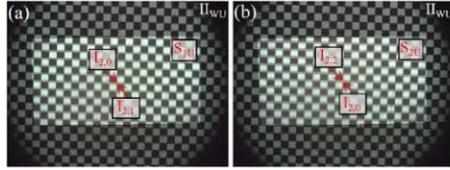
Fig. 11. Checkerboard moved 1 cm toward camera and new subaperture images captured. Corrected images (a) S_{1U} and (b) S_{2U} generated using transverse ray error lookup tables. These images were fused with II_{WU}. In comparison, disparity is observed in opposite directions relative to II_{WU}.

Table 1. Summary of Calibrated System Parameters for Depth Mapping

f	6.79 mm
Z	7.51 mm
$VB_{1,2}$	4.68 mm
Z_1	71.2 mm
Z_2	58.7 mm

as expected since A_1 and A_2 are on opposite sides of A_0. In each figure, the magnitude of disparity is constant throughout the SFOV, indicating that the checkerboard was simply translated along the central optical axis.

Corner image points $I_{1,0}$ and $I_{2,0}$ were selected in Figs. 10 and 11. Using the CCSS, the camera coordinates of O_1 and O_2 are (0.3906, −0.4499, 71.2217) mm and (−0.2556, −0.1793, 58.7407) mm, respectively. Corresponding points $I_{1,i}$ and $I_{2,i}$ are located in Figs. 10 and 11. From the image coordinates and pixel size, separations $b_{1,2,i}$ were calculated. $b_{1,2,1}$ is −0.07524 mm, and $VB_{0,1}$ was then calculated as 2.2643 mm. $b_{1,2,2}$ is 0.0297 mm, and $VB_{0,2}$ was then calculated as −2.42 mm. Thus, the total $VB_{1,2}$ between A_1 and A_2 is 4.6843 mm. This value is within the expected range because the prototype was designed for a real baseline separation between A_1 and A_2 of 5.02 mm. $VB_{0,2}$ is slightly larger than $VB_{0,1}$ because the two apertures were slid into a slot of the optomechanical housing, where mechanical tolerance resulted in a decenter of the stereo apertures. Assessment and quantification of the pupil aberration before and after correction are difficult here because the pupils are not imaged in this calibration. Instead, we must observe the quality of depth estimation results shown in the next section. A summary of the calibrated system parameters is shown in Table 1.

G. Calibration Results

Once all the calibration parameters were determined, any subsequently captured pair of S_1 and S_2 images could be processed for absolute depth mapping. The checkerboard with a ruler placed on top was tilted away from the camera to produce a linear change in depth along the vertical axis, as shown in Fig. 8(c). S_1 and S_2 were captured for this new object scene, warped into S_{1U} and S_{2U} based on the quantified ε_s and ε_t lookup tables, overlaid together as a red-cyan anaglyph, and rotated counterclockwise 90° such that the disparity is along the horizontal axis
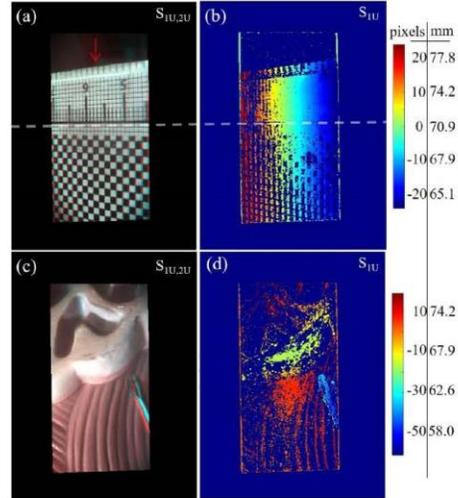

Fig. 12. Checkerboard was tilted to produce a linear change in depth. (a) S_{1U} and S_{2U} overlaid together as a red-cyan anaglyph and rotated counterclockwise 90° for viewing with 3D glasses. (b) S_{1U} converted into a disparity $d_{n,2,1}$ map in pixel units and into an absolute depth Z_n map in mm units. (c), (d) Process was repeated for a 3D bladder model.

and can be viewed with 3D glasses, as shown in Fig. 12(a). Then, S_{1U} was chosen as the reference image, and for each image point, the disparity $d_{n,2,1}$ in pixel units to the corresponding image point in S_{2U} was calculated as shown in Fig. 12(b). Using Eq. (1) and the previously determined calibration parameters, $d_{n,2,1}$ was converted to absolute depth Z_n in mm units. This process was repeated for a second object scene, a 3D bladder model, in Figs. 12(c) and 12(d).

As discussed in Fig. 6, relative refocusing is recovered after transverse ray error correction. In Fig. 12(a), the focus position is at zero pixel disparity or around the 5.75 tick mark of the ruler, as indicated by the red arrow. By translating S_{1U} and S_{2U} properly and fusing them together with II_{WU}, the focus position can be digitally changed. This is demonstrated in Figs. 13(a) and 13(b), where the focus position has been changed to the 4.5 and 7 tick marks, respectively. The out of focus blurring increases further away from the focus position as expected. The defocus occurs only along the vertical axis according to the captured subapertures, and aliasing is noticeable due to the three minimum subapertures collected. Nonetheless, relative refocusing recovery is demonstrated and is further illustrated in [23].

Depth estimations from the CCSS and the calibrated stereo apertures were compared to assess the proposed calibration. Since the entire object field was lying on a checkerboard plane, any object point's depth can be calculated with the CCSS. The depths of object points along the transparent white dashed line in Fig. 12(b) were calculated with the CCSS and compared to the results from the triangulation method of A_1 and A_2, as
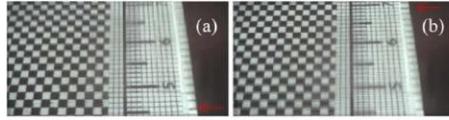
141

Fig. 13. Relative refocusing to (a) near and (b) far depths after transverse ray error correction.
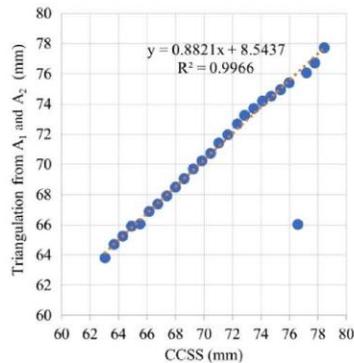

Fig. 14. Comparison of depth estimation between the CCSS and the triangulation from A_1 and A_2 indicates that absolute depth mapping capability is recovered after aberration calibration.

shown in Fig. 14. The triangulation method estimates the overall shape of the object field correctly, indicating that absolute depth mapping capability is recovered after aberration calibration. Linear regression, excluding the one outlier in the data due to noise, calculated a slope of less than one. This means the triangulation method is overestimating and underestimating at the opposite ends of the depth range and has an overall tilt error. This is likely due to either a slight tilt in the calibration checkerboard, as determined by how close Eq. (7) is to the identity matrix, or the entrance pupil at the thin lens assumption. Nonetheless, this tilt error could be calibrated out as well. Without doing so, the maximum difference in depth mapping estimation is ~0.75 mm in this example.

5. CONCLUSION

In this paper, we proposed a virtual thin lens model and calibration to recover the first-order performance of an as-built light field camera. The calibration utilizes center subaperture images to define the imaging and distortion parameters of the model. It then compares center and peripheral subaperture images to quantify vignetting and transverse ray errors. Finally, it corrects pupil aberration by calculating the ideal peripheral subaperture coordinates based on object to image point relationships. After calibration, relative refocusing or absolute depth mapping capability is restored. This was demonstrated in a prism-based tri-aperture camera, indicating that calibration can enable novel light field camera designs, to the best of our knowledge, with artificially introduced transverse ray error.

REFERENCES

1. R. Ng, "Digital light field photography," Ph.D. thesis (Stanford University, 2006).
2. L. Y. Wei, C. K. Liang, G. Myhre, C. Pitts, and K. Akeley, "Improving light field camera sample design with irregularity and aberration," ACM Trans. Graph. **34**, 152 (2015).
3. L. Mignard-Debise and I. Ihrke, "A vignetting model for light field cameras with an application to light field microscopy," IEEE Trans. Comput. Imaging **5**, 585–595 (2019).
4. Q. Cui, S. Zhu, and L. Gao, "Developing an optical design pipeline for correcting lens aberrations and vignetting in light field cameras," Opt. Express **28**, 33632–33643 (2020).
5. M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proceedings IEEE International Conference on Computer Vision* (2013), pp. 673–680.
6. X. Dallaire and S. Thibault, "Simplified projection technique to correct geometric and chromatic lens aberrations using plenoptic imaging," Appl. Opt. **56**, 2946–2951 (2017).
7. Y. Chen, X. Jin, and B. Xiong, "Optical-aberrations-corrected light field re-projection for high-quality plenoptic imaging," Opt. Express **28**, 3057–3072 (2020).
8. H. Sommer, A. Ihrig, M. Ebenau, D. Flühs, B. Spaan, and M. Eichmann, "Integral image rendering procedure for aberration correction and size measurement," Appl. Opt. **53**, 3176–3182 (2014).
9. J. Chung, G. W. Martinez, K. Lencioni, S. Sadda, and C. Yang, "Computational aberration compensation by coded aperture-based correction of aberration obtained from optical Fourier coding and blur estimation," Optica **6**, 647–661 (2019).
10. M. Levoy, Z. Zhang, and I. McDowall, "Recording and controlling the 4D light field in a microscope using microlens arrays," J. Microsc. **235**, 144–162 (2009).
11. H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, June 7–12, 2015, pp. 1547–1555.
12. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 1027–1034.
13. P. Zhou, Z. Yang, W. Cai, Y. Yu, and G. Zhou, "Light field calibration and 3D shape measurement based on epipolar-space," Opt. Express **27**, 10171–10184 (2019).
14. Y. Ji and J. Wu, "Calibration method of light-field camera for photogrammetry application," Measurement **148**, 106943 (2019).
15. J.-Y. Bouguet, "Camera calibration toolbox for MATLAB," (2004).
16. N. Asada, A. Amano, and M. Baba, "Photometric calibration of zoom lens systems," in *Proceedings, IEEE International Conference on Computer Vision* (1996), Vol. **1**, 186–190.
17. M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH* (1996), pp. 1–8.
18. T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," J. Electron. Imaging **19**, 021106 (2010).
19. C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography," ACM Trans. Graph. **27**, 55 (2008).
20. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).

142

21. J. I. Katz, S. Lee, and H. Hua, "Improved multi-resolution foveated laparoscope with real-time digital transverse chromatic correction," Appl. Opt. **59**, G79–G91 (2020).

22. E. Kwan and H. Hua, "Prism-based tri-aperture laparoscopic objective for multi-view acquisition," Opt. Express **30**, 2836–2851 (2022).

23. E. Kwan, Y. Qin, and H. Hua, "High resolution, programmable aperture light field laparoscope for quantitative depth mapping," OSA Contin. **3**, 194–203 (2020).

143