UCLA

UCLA Electronic Theses and Dissertations

Title

Multi-dimensional Optical Imaging

Permalink

https://escholarship.org/uc/item/4zt4v3q0

Author

Cui, Qi

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Multi-dimensional Optical Imaging

A dissertation submitted in partial satisfaction of the

requirements for the Degree of Philosophy

in Bioengineering

by

Qi Cui

2024

ABSTRACT OF THE DISSERTATION

Multi-dimensional Optical Imaging

by

Qi Cui

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2024

Professor Liang Gao, Chair

Imaging systems capture light rays contain rich information, which can be described by the plenoptic function, $P(x, y, z, u, v, \lambda, t)$—where x, y, z represent spatial coordinates; u, v denote emittance angles; λ signifies wavelength; and t denotes time. Given a finite photon budget, it is crucial for an imaging system to maximize the information yield from each captured image. Yet, traditional cameras only capture two-dimensional spatial data (x, y), neglecting the wealth of information.

Capturing multi-dimensional information presents significant challenges, primarily due to the complexity of mapping high-dimensional datacubes onto a two-dimensional detector array. This mapping process introduces a fundamental trade-off among various axes of information, such as spatial, angular, and spectral dimensions, which can adversely affect imaging speed, resolution,

and other critical parameters. Balancing of these factors often leads to compromises in one aspect to enhance another, pose a significant challenge in designing and implementing a multi-dimensional imaging system.

In response to these challenges, this dissertation presents three innovative multi-dimensional optical imaging systems. The first system, snapshot hyperspectral light field imaging utilizing image mapping spectrometer (LF-IMS), represents a five-dimensional (x, y, u, v, λ) imaging system. It is uniquely designed to maintain full light throughput, enabling the capture of detailed three-dimensional spatial and spectral information without sacrificing efficiency. The second system, snapshot hyperspectral light field tomography (Hyper-LIFT), leverages compressed sensing to facilitate five-dimensional (x, y, u, v, λ) imaging. This approach significantly alleviates the tradeoff between different dimensions of information, allowing for capturing an input scene with a more compact sensor, thereby greatly reducing the volume of data generated during image acquisition. The third system, the tunable image projection spectrometer (TIPS), is a Fourier-domain line-scan hyperspectral imager with a tunable compression ratio. Compared to state-of-the-art spatial-domain pushbroom hyperspectral cameras, TIPS requires much fewer measurements and provides a higher light throughput.

Furthermore, this dissertation will explore the impact of optical aberrations on the image quality in light field imaging, providing insights into how these imperfections influence the overall imaging performance. A lens design pipeline is proposed to mitigate key aberrations, and its effectiveness is demonstrated through the design of a light field endoscope. Additionally, a rapid calibration method has been proposed for a compact hyperspectral camera, termed the image mapping spectrometer (IMS), reducing the calibration time from weeks to hours.

The dissertation of Qi Cui is approved.

Jun Chen

Song Li

Tzung Hsiai

Liang Gao, Committee Chair

University of California, Los Angeles

2024

Table of Contents

List of Figures

viii

Acknowledgements

I would like to extend my deepest gratitude to Professor Liang Gao, my PhD advisor, for his invaluable guidance, patience, and support throughout my doctoral journey. His mentorship have been essential to my development and success.

Additionally, my heartfelt thanks go to my dissertation committee members: Prof. Song Li, Prof. Jun Chen, and Prof. Tzung Hsiai. Their ongoing support has been crucial to my PhD journey, and their insightful feedback has greatly enhanced the quality of my work.

I am also immensely grateful for the opportunity to work with my lab mates and collaborators. The knowledge and experiences we shared have been incredibly beneficial to my growth as a researcher.

Lastly, I express my deepest appreciation to my family and friends for their endless love, support, and encouragement.

This journey has indeed been an extraordinary adventure, made all the more meaningful by your presence in my life.

Chapter 1 Introduction

In the realm of optical imaging, the quest to fully harness the rich information carried by light rays demands progress beyond traditional two-dimensional imaging techniques. The plenoptic function, $P(x, y, z, u, v, \lambda, t)$, presents a comprehensive model encompassing spatial, angular, spectral, and temporal dimensions, offering a complete description of a scene. With a limited photon budget, the image system must assure that each detected light ray provides as much information as possible. However, a conventional camera detects only two-dimensional (2D) spatial information (x, y), throwing away much of the information actually carried by a light ray and therefore leading to an inefficient optical measurement.

This dissertation emphasizes two critical methodologies to address these challenges: light field imaging and hyperspectral imaging. Light field imaging, leveraging spatial multiplexing, simultaneously records spatial (x, y) and angular (u, v) dimensions of light rays. This process yields four-dimensional (4D) light field data, from which depth maps (x, y, z) and multi-view images can be generated through further processing, offering a comprehensive view of the scene's three-dimensional structure.

On the other hand, hyperspectral imaging extends beyond traditional methods by capturing both spatial (x, y) and spectral (λ) information. This approach provides a detailed spectral signature for each pixel, greatly enhancing the ability to distinguish and characterize materials based on their unique spectral responses, which enables more precise analysis and identification in a variety of applications ranging from environmental monitoring to medical diagnostics.

Light field and hyperspectral imaging represent the forefront of efforts to fully exploit the

informational potential of light. By addressing the limitations of traditional imaging and exploring these advanced methodologies, this dissertation aims to improve the capabilities of optical imaging systems, creating opportunities for research and application across various disciplines.

1.1 Light field imaging

Light field imaging marks a revolutionary advancement in computational photography, capturing both the spatial and angular dimensions of light rays in a scene. This approach significantly surpasses the capabilities of traditional photography which is limited to recording the intensity of light on a single plane, by comprehensively recording the light field. This field embodies the flow of light in all directions across every point in space, offering distinct advantages including the ability to refocus images digitally and achieve an extended depth of field (DOF).

The concept of light field imaging was first introduced by Gabriel Lippmann in 1908 with his proposal of integral photography [1], but it was not until the early 21st centuries that computational photography research led to practical implementations of light field cameras. These advancements were propelled by the development of algorithms capable of processing the complex data captured in light fields, alongside the creation of hardware capable of recording this information, notably through the integration of microlens arrays above the sensor. These arrays facilitate the capture of not only the intensity but also the vector direction of incoming light rays.

Pioneering research by Ng et al. at Stanford University demonstrated the feasibility of light field photography using a handheld plenoptic camera, marking a significant milestone in the field [2]. This work provided a framework for capturing and processing light fields with conventional

imaging sensors, opening pathways for novel photographic techniques and applications. Light field imaging was commercially made available by a company named Lytro, which released the first consumer light field camera in 2011. The Lytro camera allowed users to refocus images after they were taken, showcasing the potential of light field technology for everyday photography.

Beyond its implications for photography, light field imaging holds great promise for a variety of applications across object recognition, machine vision, and biomedical imaging [3-9]. In the field of microscopy, light field microscopy enables rapid, three-dimensional capture of biological processes in a single shot, significantly reducing sample damage, and eliminating the need for extensive mechanical scanning.

In this dissertation, I will discuss how optics aberration affect the image quality of light field imaging, then proposed a lens design pipeline and demonstrated with a design of light field endoscope.

1.2 Hyperspectral imaging

Hyperspectral imaging (HSI) is an advanced imaging technology that captures images across a wide spectrum, extending from the visible to the near-infrared regions. Unlike conventional imaging, which records only three primary colors (RGB), HSI collects data in hundreds of narrow, adjacent spectral bands, which provides continuous spectrum for every pixel and enables classification for different materials through their unique spectral signatures. HSI typically separates spatial and spectral information by employing a dispersion element, such as a grating or prism.

Originally developed for remote sensing, hyperspectral imaging has found broad applications in other fields as well, such as biomedical imaging and machine vision [10-19]. In this dissertation, I will first focus on a compact spectral imager with full-light throughput, known as the image mapping projection spectrometer. A rapid calibration method for this spectrometer will be introduced, which notably shortens the calibration time from weeks to mere hours. Additionally, a compressed-sensing Fourier domain spectrometer named the Tunable Image Projection Spectrometer (TIPS) is introduced. When compared to the contemporary spatial-domain pushbroom hyperspectral cameras, TIPS requires significantly fewer measurements and achieves higher light throughput.

Although light field and hyperspectral imaging techniques continue to mature, combining them on a single platform is non-trivial. Xiong et al. [20] developed a hyperspectral light field imaging system by using two cameras to acquire the spectral and light field data, respectively. However, the two-camera setup is plagued by misalignment errors. Alternatively, Zhu et al. [21] used a 5×6 spectrally filtered camera array to measure the hyperspectral light field data. Despite being able to provide high-resolution images, their system suffers from a bulky setup and low-light

efficiency. Finally, in a most recent implementation, Zhu et al. [22] introduced light field imaging into Fourier transform imaging spectroscopy. The resultant system features a compact form factor. Nonetheless, because of the reliance on polarization, the light throughput is halved when imaging unpolarized natural scenes.

In response to this unmet need, I proposed two five-dimensional systems. The first, snapshot hyperspectral light field imaging utilizing image mapping spectrometer (LF-IMS), which is designed for full light throughput and enables efficient measurement of detailed spatial and spectral information. The second, snapshot hyperspectral light field tomography (Hyper-LIFT), employs compressed sensing to significantly lessen the compromise between the dimensions of information and the volume of data generated.

Chapter 2 Light field imaging

Light field cameras have been employed in myriad applications thanks to their 3D imaging capability. By placing a microlens array in front of a conventional camera, one can measure both the spatial and angular information of incoming light rays and reconstruct a depth map. Due to its superior 3D imaging capability, the light field camera has been employed in various applications such as biomedical imaging, object recognition, and machine vision.

2.1 Fundamentals of light field imaging

Light field cameras are categorized into two main types: unfocused light field (ULF) cameras and focused light field (FLF) cameras. Figure 1 shows the corresponding schematics. As shown in Fig. 1(a), in a ULF camera, three point objects $S1$, $S2$, and $S3$ are first imaged by the main lens, forming intermediate image points $S'1$, $S'2$ and $S'3$. These intermediate image points are then reimaged by the microlens array (MLA) onto a detector array. Because the distance from the MLA to the detector array is equal to the focal length of the MLA, the ULF camera essentially images the pupil associated with each microlens. We use (u, v) and (x, y) to denote the Cartesian coordinates at the pupil plane and the MLA, respectively. The captured raw images ($M1$, $M2$, and $M3$ in Fig. 1(a)) can be re-arranged as a 4D datacube (x, y, u, v), which is also referred to as a light field (LF) [23]. A 2D x-u slice of the LF is termed an epipolar plane image (EPI). As an example, Fig. 1(b) shows three EPIs associated with point $S1$, $S2$, and $S3$, respectively. The corresponding depths can then be deduced by estimating the slope of lines in the EPIs. A refocused image at a given depth can be reconstructed from an integral projection of the 4D LF along a trajectory in the EPIs. Reconstructing images at all depths creates a focal stack

of images, and an extended depth of field (DOF) image can be rendered by fusing all the reconstructed images [24].
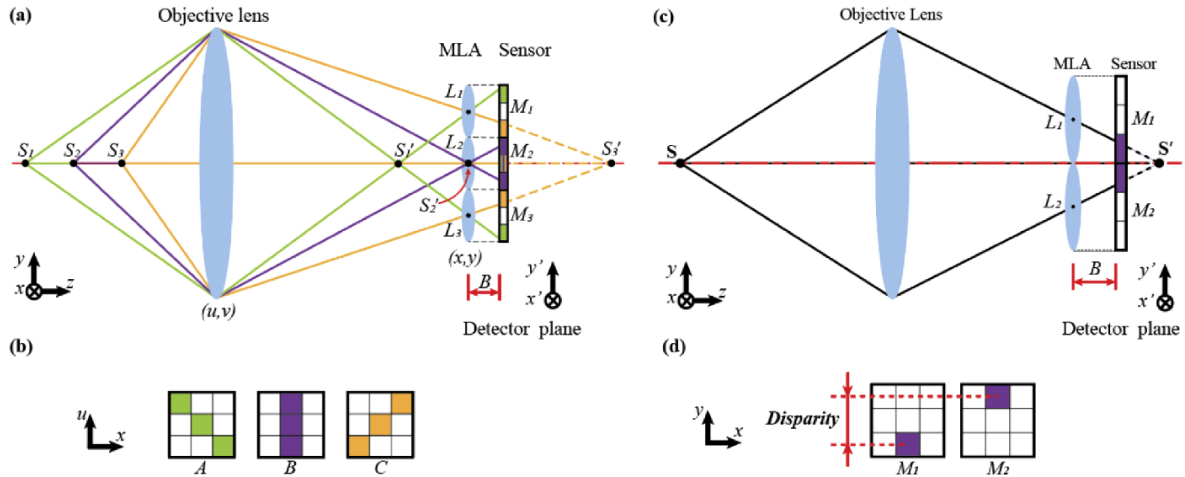
Figure 1.: Ray models of light field cameras. (a) ULF camera. (b) EPIs associated with point $S1$, $S2$ and $S3$ in (a). (c) FLF camera. (d) Perspective images imaged by microlens $L1$ and $L2$ in (c).

Unlike the ULF camera, the FLF camera directly images the object, rather than pupils, onto the detector array. There are two types of FLF cameras: the Keplerian and Galilean [25]. Figure 1(c) shows the schematic of a Galilean FLF camera. The spacing (B) between the MLA and the detector is smaller than the focal length of the MLA. In contrast, B is larger than the focal length of the MLA in the Keplerian configuration. The depth information can be derived from the disparities between adjacent perspective images (Fig. 1(d)), and an all-in-focus image can be reconstructed by projecting all the pixels in the raw image back to the intermediate image plane.

2.2 Aberrations in light field objective lens

Although the depth calibration method and ray tracing model of light field camera have been extensively studied [26–33], the optical design of its main lens has yet to be exploited. Because of the unique optical architecture of light field cameras, the handling of lens aberrations and vignetting is significantly different from conventional lens design methods [34,35]. To address this unmet need, we systematically analyzed the effect of aberrations and vignetting on the fidelity of reconstructed images and developed a design pipeline for the main lens of light field cameras. While the proposed lens design pipeline is generally applicable to all light field cameras, we focus on a niche application in endoscopy.

When designing an imaging lens, although aberrations and vignetting are usually unwanted, they are not equally weighted in the tolerancing budget. Here we limit our discussion to third-order Seidel aberrations and ignore defocus and wavefront tilt. The conventional optical design prioritizes the correction of aberrations, which increase the spot size at the image plane (i.e., spherical aberration, coma, astigmatism, field curvature). Particularly, when field curvature W_{222} exists, a flat object plane is imaged to a curved surface. Because the detector plane is flat, field-dependent defocus is then introduced to the final image. In the periphery field, the blur so induced is so severe that it often overshadows other aberrations. More problematically, field curvature is more difficult to correct for than other Seidel aberrations—common approaches such as lens bending/splitting and stop shifting cannot be applied because field curvature depends on only the power and refractive index of lenses if the system is free of astigmatism. Therefore, in conventional optical design, field curvature is considered one of the toughest aberrations, and correcting for it normally leads to a bulky setup. By contrast, vignetting reduces the irradiance of the image but not the resolution, and it can be numerically corrected for in postprocessing. For this reason, vignetting is a less-concerned factor compared with Seidel

aberrations.

Unlike conventional cameras that capture only the 2D (x, y) information of a scene, light field cameras measure a 4D (x, y, u, v) datacube and derive the depth from light ray angles. Therefore, designing the main lens needs a new standard. Particularly, the field curvature and vignetting must be assessed in 3D (x, y, z) rather than 2D (x, y). Figure 2 shows a light field camera with field curvature. The object is imaged by the main lens to a curved surface, as indicated by the black dashed line. The depth of field of the microlens array (MLA), denoted by DRM, determines the depth range of the main lens, while the DRM itself depends on the detector pixel size and the numerical aperture (NA) of the MLA [36]. Provided that the entire curved intermediate image is located within the DRM, the shape of the surface can be recovered through calibration. As a result, the field curvature can be numerically corrected by digital refocusing, and it can be loosely tolerated in light field cameras.
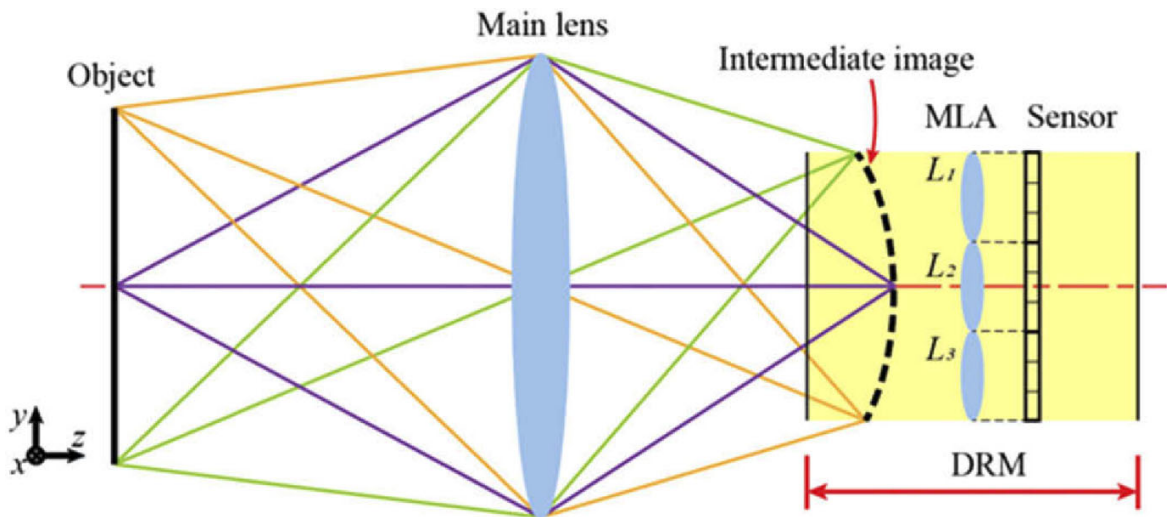
Figure 2. Field curvature in a light field camera. DRM, depth range of the microlens array; MLA, microlens array.

By contrast, in light field cameras, vignetting must be minimized. Because light field cameras estimate depths using the light ray angles, the loss of the angular information due to vignetting will reduce the number of views in the EPIs. To elaborate on this effect, we performed a

simulation using Zemax (Zemax, LLC). Figure 3 shows the shaded model of an ULF camera. The object is a point source. We use a 4F system as the main lens, which consists of two paraxial lenses ($f = 15$ mm) and a physical stop. The stop is placed at the Fourier plane of the first lens (i.e., back focal plane). To match the NA of the main lens and the MLA, we set the stop diameter to 1.38 mm. To introduce vignetting, we place another aperture of the same diameter at a location 10 mm after the stop. A MLA ($f = 0.65$ mm, lens pitch $= 60$ μm) locates at the back focal plane of the second lens, and a detector array is placed at the back focal plane of the MLA. The pixel size of the detector array is 4 μm.
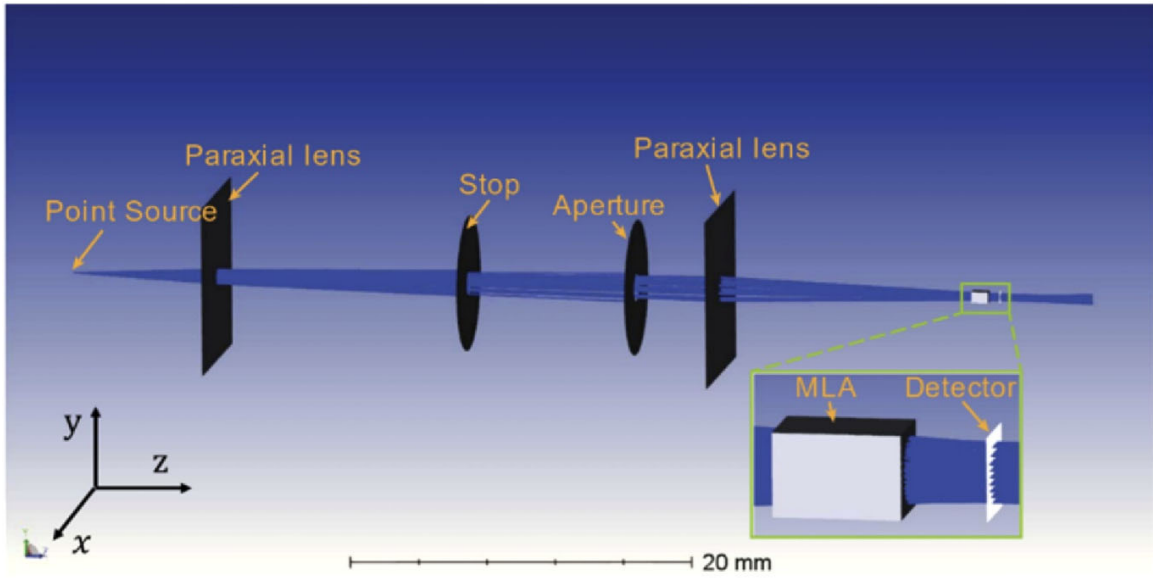
Figure 3. Shaded model of an unfocused light field (ULF) camera. MLA, microlens array.

We define the vignetting factor η as:

$$\eta = 1 - E/Eu, \tag{1}$$

where E and Eu denote the total irradiance received by the detector array with and without vignetting, respectively, and η is zero if the image is unvignetted. In the simulation, the point source was placed at the front focal plane of the first lens, and we scanned it along the x-axis at

10

13 different locations from 0 mm to 1.2 mm with a step size of 0.1 mm. At each step, we traced 100,000 light rays to form a raw image and rendered an EPI at v = 0 and y = 0. Figure 4(a) shows three representative raw images at x = 0 mm, 0.6 mm, 1.2 mm, and their corresponding EPIs. The results indicate that although the slope of the line feature in the EPIs does not change, the number of pixels that forms the line (i.e., views) reduces as vignetting increases. The relation between the vignetting factor and the number of views is shown in Fig. 4(b). We calculated the number of views by enumerating the non-zero pixels in the EPI after image binarization. The light field camera reconstructs depth by estimating the slope of line features in EPIs through linear regression. The standard error of fitting can be computed by:

$$SE = \sqrt{\frac{\sum(b_i - \widehat{b}_\iota)^2}{\sum(a_i - \bar{a})^2}} \cdot \sqrt{\frac{1}{n-2}}, \tag{2}$$

where *SE* is the standard error, *n* is the number of observations, *ai* is an independent variable for the ith observation, $\bar{a}$ is the mean, *bi* is a dependent variable for the ith observation, and $\widehat{b}_\iota$ is the estimated value of b_i. Equation 2 implies that the standard error decreases as the number of observations increases. In light field cameras, vignetting reduces the number of views in EPIs, resulting in a larger regression error and, therefore, a reduced depth accuracy. Particularly, when the number of detector pixels associated with a microlens is small, vignetting dramatically increases the regression error.
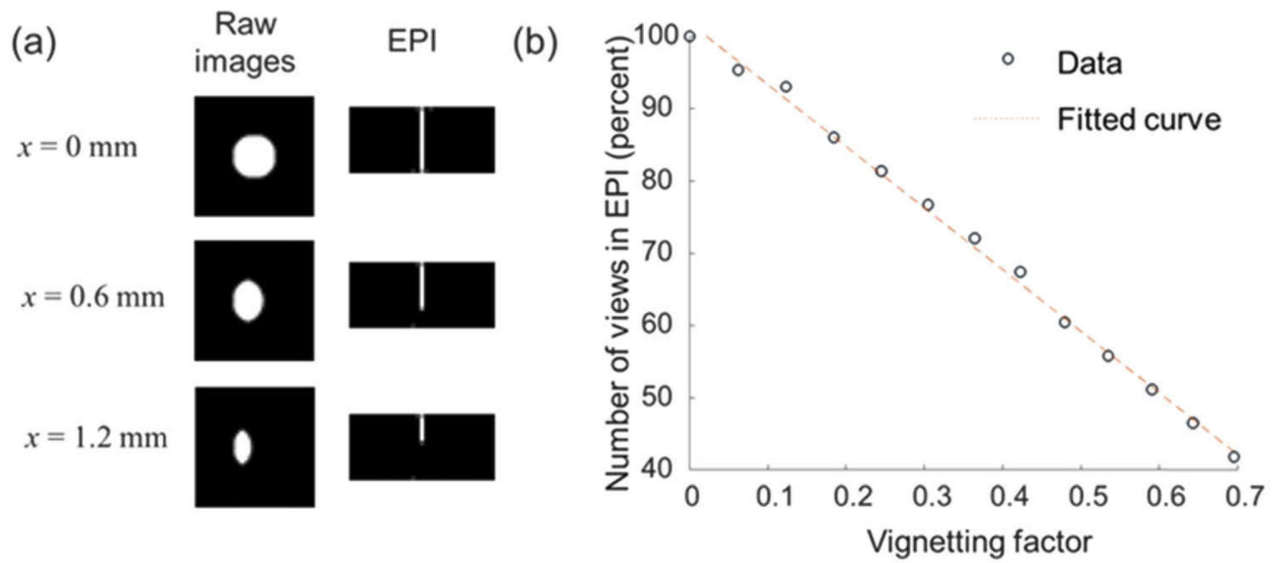
Figure 4. Vignetting and number of views in epipolar plane images (EPIs). (a) Three representative raw images and corresponding EPIs at x = 0 mm, 0.6 mm, 1.2 mm. (b) Number of views in an EPI vs. vignetting factor.

To further illustrate the effect of vignetting on depth accuracy, we defocused the point source by 6 mm towards the first lens, and we scanned it under the same conditions. Because the depth of the point source has changed, the line in the EPI is tilted with respect to the vertical axis, and it is not aligned with the detector pixels. As a result, ambiguities are introduced by sampling. Three representative raw images and corresponding EPIs at x = 0 mm, 0.6 mm, 1.2 mm are shown in Fig. 5(a). At each step, we computed the slope of the line in the EPI. The relation between the slope regression error and the vignetting factor is shown in Fig. 5(b).
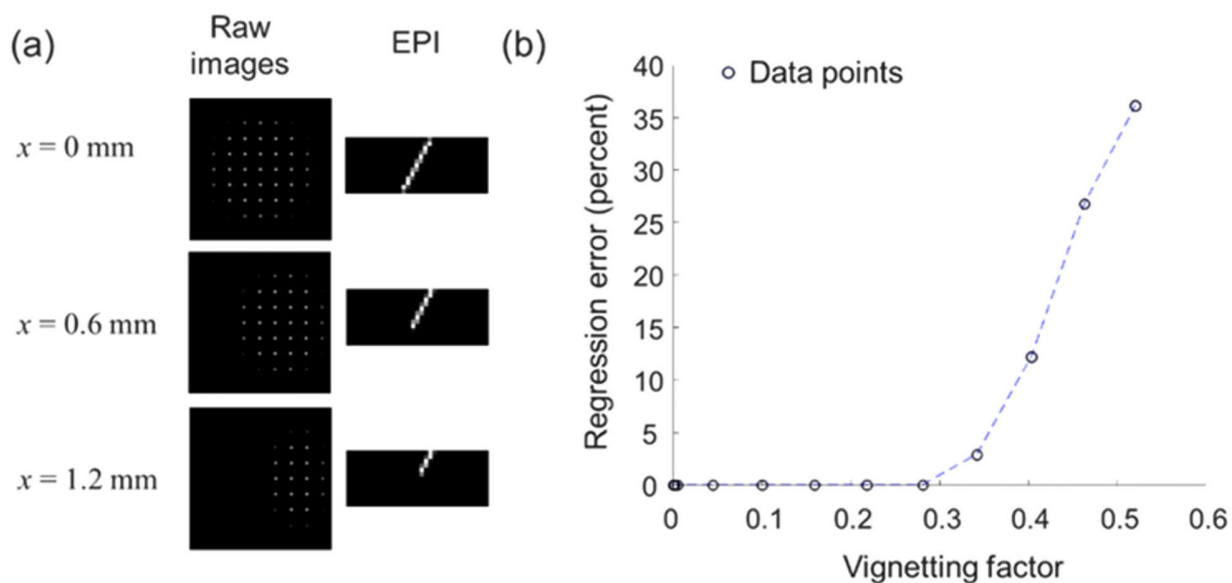
Figure 5. Vignetting and slope regression error of the line feature in epipolar plane images (EPIs). (a) Three representative raw images and corresponding EPIs at x = 0 mm, 0.6 mm, 1.2 mm. (b) Slope regression error vs. vignetting factor.

It is worth mentioning that the slope regression error is also dependent on aberrations and noises. When aberrations exist, the image of a point source is no longer a sharp point, and the shape of the line in the EPI may be distorted. On the other hand, noises affect the intensity of the views and the background pixels. In both cases, a sufficient number of views is critical for faithful depth reconstruction. Therefore, vignetting must be minimized in light field cameras.

Finally, we validated the effect of vignetting through a real experiment. The optical setup of an unfocused light field camera is shown in Fig. 6(a). We used a 4F system as the main lens, which consists of two 50 mm focal length achromatic doublets (Thorlabs, AC254-050-A-ML). A 4.8 mm diameter stop was placed at the Fourier plane to match the NA of the main lens and the MLA. An MLA with a 50 μm pitch was placed at the back focal plane of the second lens, and the spacing between the MLA and the camera (Lumenera, Lt965R) is equal to the MLA focal length. A flat printed grid pattern was used as the object, and it is located near the front focal plane of the main lens. An adjustable aperture was positioned 12 mm before the camera, and its diameter

was set to be 2.8 mm, 4 mm, and 5 mm to create different levels of vignetting. We captured a raw image for each aperture diameter and a baseline image when the aperture was removed (i.e. no vignetting). A representative raw image when the aperture diameter = 4 mm and the baseline image are shown in Fig. 6(b), each including two magnified subfields. Compared to the baseline, Area 2 from the raw image when the aperture diameter = 4 mm shows vignetted pupils. Next, we calculated the vignetting factor and generated a disparity map for each image, followed by computing the root-mean-squared error (RMSE) for each disparity map. Note that a depth map can be further rendered based on disparity-to-depth calibration. The resultant disparity maps are shown in Fig. 7. The experimental results indicate that the disparity RMSE increases as the vignetting factor increases. Therefore, depth accuracy would be jeopardized if vignetting exists.
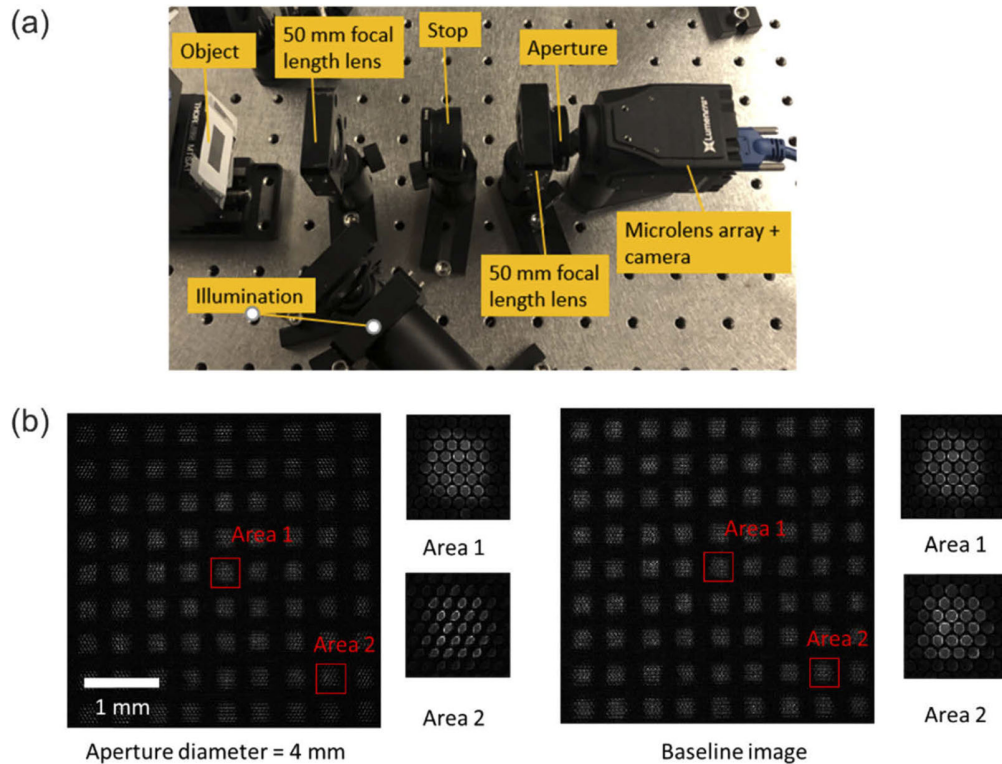
Figure 6. Experimental setup and raw images of a flat printed grid pattern object. (a) Optical setup. (b) A raw image when the aperture diameter = 4 mm and the baseline image with two magnified subfields.

14
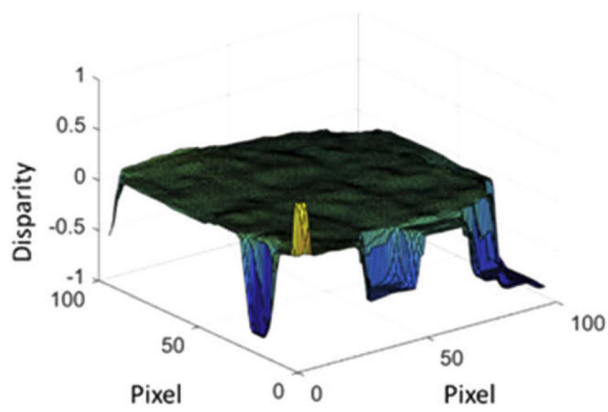
Vignetting factor = 0.46, RMSE = 0.1843
Aperture diameter = 2.8 mm
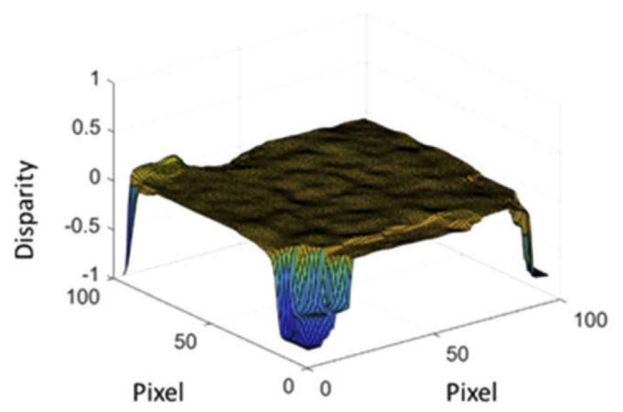
Vignetting factor = 0.236, RMSE = 0.118
Aperture diameter = 4 mm
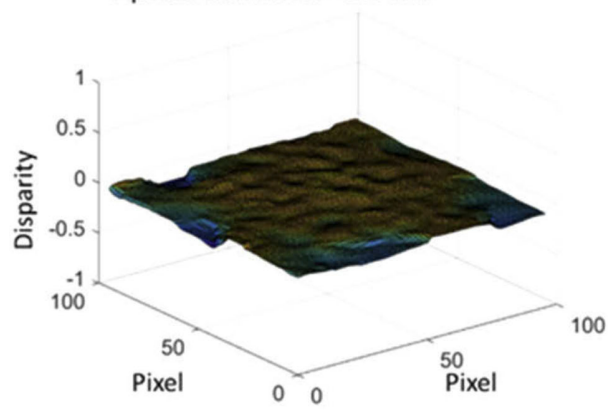
Vignetting factor = 0.047, RMSE = 0.033
Aperture diameter = 5 mm
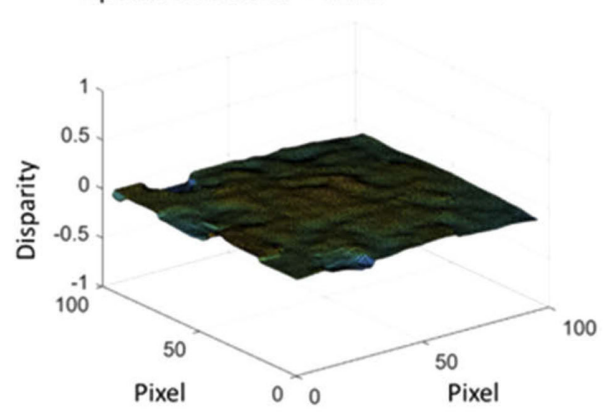
Vignetting factor = 0, RMSE = 0.03
Ground truth

Figure 7. Disparity maps for each aperture diameter. RMSE, root-mean-squared error.

2.3 Lens design for light field cameras

Compared to conventional cameras, light field cameras can tolerate field curvature but are sensitive to vignetting. The field curvature coefficient W220 can be separated into two terms.

$$W_{220} = \frac{1}{2}W'_{222} + W_{220p},\qquad(3)$$

where W'_{222} is proportional to astigmatism and W_{220p} is Petzval curvature. Without astigmatism, the field curvature reduces to Petzval curvature. Because Petzval curvature depends on only the power and refractive index of lenses, it is insensitive to most aberration correction methods (e.g., lens bending/splitting, stop shifting). The primary method to flat Petzval surface is to add negative power lenses and create air spaces in between. However, it makes the system bulky and expensive. Therefore, releasing the tolerance on the field curvature can greatly reduce the system complexity and design constraints. For example, if we use a single ball lens as the main lens in a light field camera, all off-axis aberrations would be eliminated [37]. Digitally correcting for the remaining field curvature provides an ideal solution to achieve a large field of view with a high resolution. To minimize vignetting in a light field camera, we put a constraint on the lens aperture:

$$a \geq |y| + |\bar{y}|,\qquad(4)$$

where a is the radius of the aperture, and $|y|$ and y are the chief ray height and marginal ray height at the aperture position, respectively. In addition, we force the telecentricity of the main lens in the image space.

Figure 8 illustrates the proposed optical design pipeline, which differs from the conventional standard in two aspects: first, the field curvature is not a primary design constraint and can be loosely tolerated, while vignetting must be strictly minimized. Second, optimization must be

performed in 3D (*x, y, z*) rather than 2D (*x, y*)—we must account for all object points within both the depth range (*z*) and FOV (*x, y*). In practice, given radial symmetry, it is justified to sample object points only in the *y-z* plane. During optimization, we assign each (*y, z*) object point to a system configuration. We then perform ray tracing in each configuration and calculate the corresponding vignetting factor. Lastly, we construct a *y-z* vignetting factor map and compute the mean. We use this value as the metric to evaluate vignetting of the system.
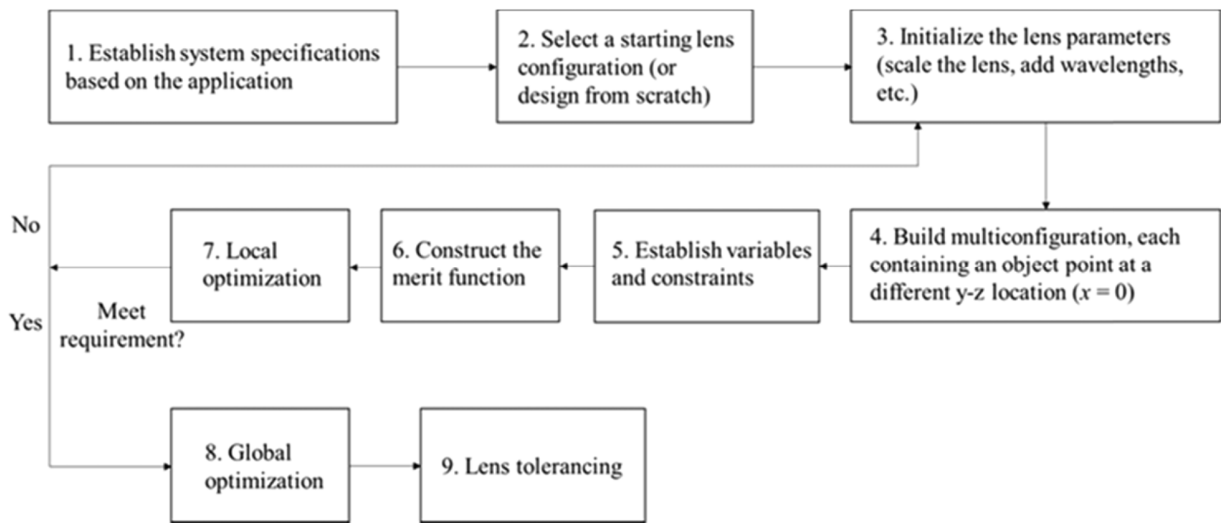
Figure 8. Optical design pipeline for light field cameras. Due to correction of aberrations/vignetting in a 3D space, our design pipeline yields optimized optical performance for computational refocusing and parallax-based depth estimation.

17

2.4 Design example

To demonstrate the implementation of the proposed pipeline, we designed the main lens for a light field endoscope using Zemax. The requirements are listed as follows: object space numerical aperture (NA) is 0.024, working distance is 65 mm. Field of view (FOV) is 10 mm and magnification is 0.2. The diameter is 5 mm, depth range is 6 mm, and total length > 200 mm. We selected a double Gauss lens as the initial configuration to reduce odd aberrations, followed by scaling down the lens to the required diameter. Next, nine object points within the depth range (z) and the FOV (x, y) were chosen to build the multi-configuration, as summarized in Fig. 9. The working distance (WD) is defined as the distance between an object point and the first surface of the main lens. We inserted a dummy surface after the nominal image plane (where the marginal ray height = 0 mm) in each configuration, which serves as the real image plane. Due to field curvature, defocus is introduced for off-axis object points. During optimization, the location of the dummy surface was set as a variable, and each configuration was optimized independently to compensate for the field-dependent defocus. In this way, the effect of field curvature is excluded in the merit function for image quality optimization.

1.	Config 1		Config 2		Config 3		Config 4		Config 5		Config 6		Config 7		Config 8		Config 9	
2.	0.000		7.000		10.000		0.000		7.000		10.000		0.000		7.000		10.000	
3.	65.000		65.000		65.000		62.000		62.000		62.000		68.000		68.000		68.000	
4.	-0.024	V	-0.903	V	-1.818	V	-0.024	V	-0.975	V	-1.991	V	-0.023	V	-0.839	V	-1.671	V

1: Configuration numbers   4: Dummy surface's location with respect
2: Object height (mm)   to the nominal image plane (mm)
3: Working distance (mm)   V: variables in optimization

Figure 9. Multi-configuration in lens optimization.

Next, we built the merit function based on design specifications. The activated operands are summarized in Figure 10. The variables consist of the radius of surface curvature and the central thickness between adjacent surfaces. Only spherical surfaces are used for each lens element. The optimization process is divided into two steps: local optimization and global optimization. During the local optimization, the paraxial magnification is defined using operand PMAG, RECI, ABLT, and ABGT. The desired magnification of the main lens is 0.2. We used operand AXCL to minimize the axial color, while other aberrations (spherical aberration, coma, astigmatism, distortion, and lateral color) are optimized together to minimize the root-mean-squared (RMS) spot size using default operand TRAC. Particularly, we limited vignetting by image space telecentricity. The operand RAID was used to confine the chief ray angle (CRA) to the last surface of the lens. In addition, the semi-diameter of the lens group was limited by operand MXSD, and the air and glass thicknesses were constrained by operand MNCA, MXCA, MNEA, MNCG, MXCG, and MNEG. During the global optimization, we made two changes: first, we replaced operand TRAC with operand OPDX to minimize the RMS wavefront error. Second, the glass type of each element was set as "substitute" for better performances the real image.

Local optimization	PMAG, RECI, ABLT, ABGT, AXCL, TRAC, RAID, MXSD, MNCA, MXCA, MNEA, MNCG, MXCG, MNEG
Global optimization	PMAG, EFFL, ABLT, ABGT, AXCL, OPDX, RAID, MXSD, MNCA, MXCA, MNEA, MNCG, MXCG, MNEG

Figure 10. Activated operands in the merit function.

To meet the length requirement, we further used Hopkins rod lenses as the relay lens. The desired magnification of the relay lens is 1. We started with two thick doublets, which are symmetric about the stop. As a result, the lens does not introduce coma, distortion, and lateral color. We used the same merit function as that in the main lens, except the object space telecentric was forced to match the pupil. The variables consist of the radius of curvature of each surface and the spacing between adjacent surfaces. After optimization, we duplicated the lenses to extend the relay optics to the required length.
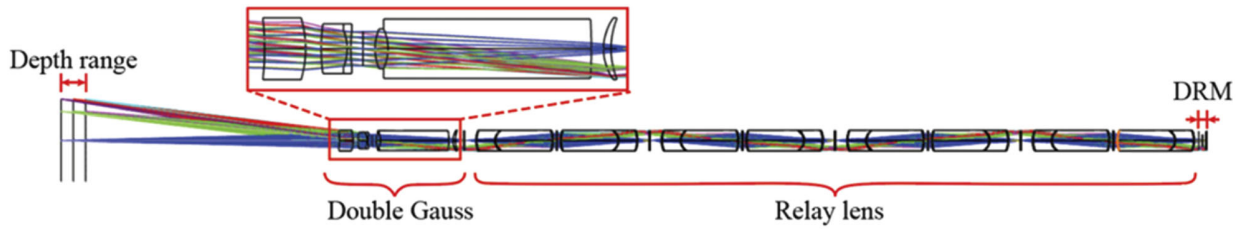
Figure 11. Optical setup of the endoscope.

The schematic of the final endoscope is shown in Fig. 11. The effective focal length (EFFL) of the system is 14.6 mm, and the total length (TOTR) is 212 mm. The back focal length is 3 mm, and the paraxial magnification is ~ 0.206. Figure 12 shows spot diagrams of three configurations when working distance = 65 mm and object height = 0 mm, 7 mm, 10 mm, respectively, and the corresponding modulation transfer functions (MTFs) are shown in Fig. 13. Finally, we performed ray tracing to calculate vignetting factors for all object points within the depth range and the FOV, and the result is shown in Fig. 14, where the pixel value represents the normalized percentage of unvignetted rays. The mean of this map is 0.99, implying that only one percent of total rays are vignetted. The resultant design, therefore, maximizes the depth reconstruction fidelity.
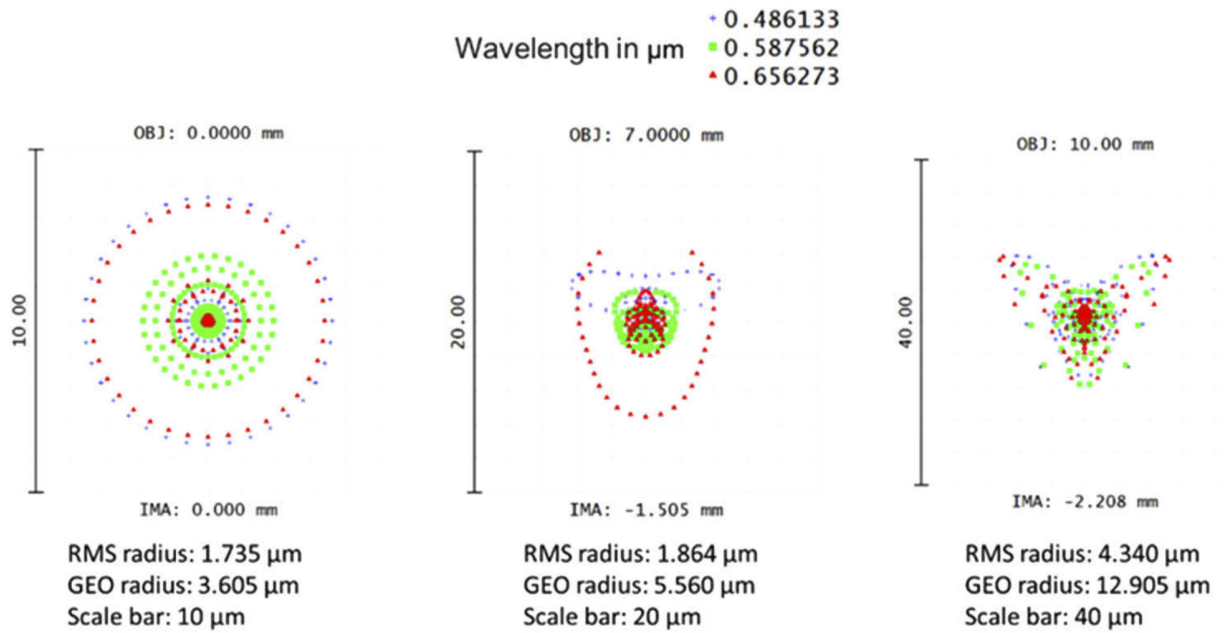
Figure 12. Spot diagrams corresponding to three configurations in which working distance = 65 mm, object height = 0 mm, 7 mm, 10 mm, respectively.
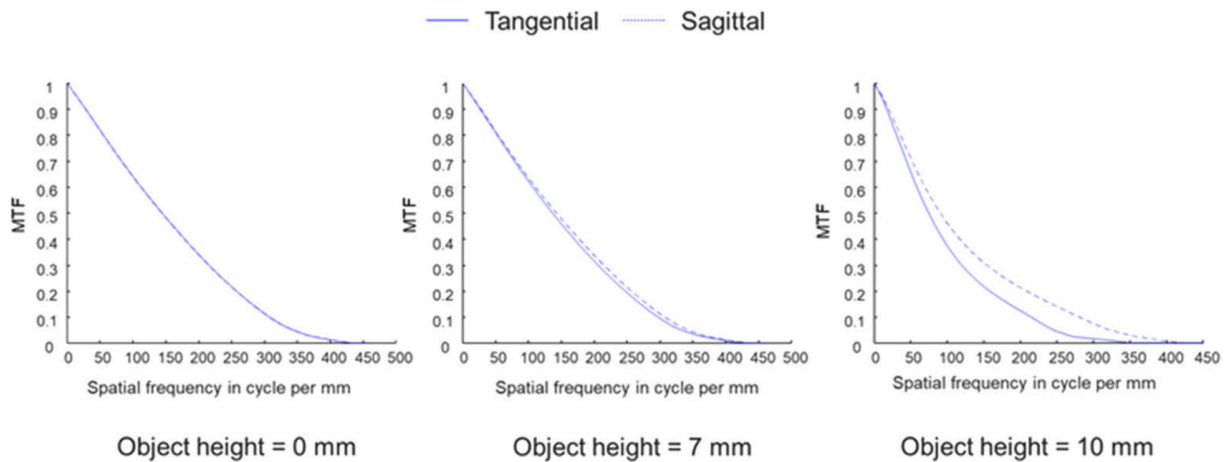
Figure 13. Modulation transfer functions (MTFs) corresponding to three configurations in which working distance = 65 mm, object height = 0 mm, 7 mm, 10 mm, respectively.

21
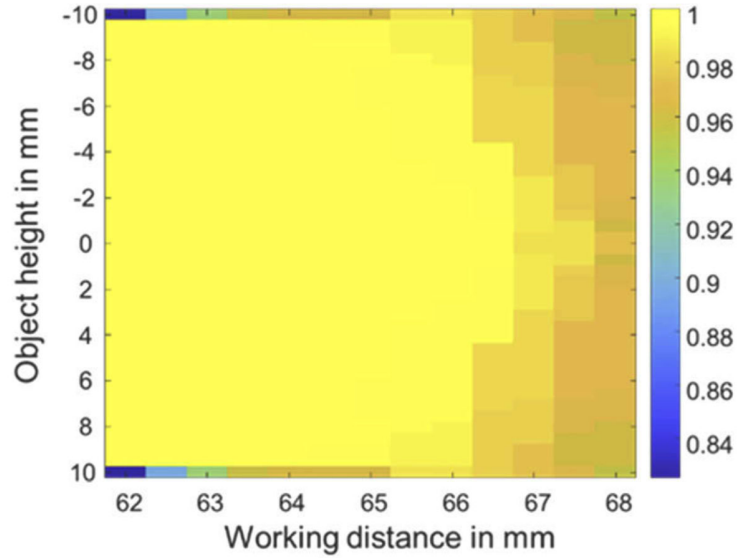
Figure 14. Vignetting factor map within the depth range and the FOV.

In summary, we systemically studied the effect of field curvature and vignetting on light field depth reconstruction accuracy. We show that the field curvature in light field cameras can be loosely tolerated, while vignetting must be minimized to assure high reconstruction fidelity. To incorporate this finding into the lens design process, we developed a pipeline that optimizes the optical performance of light field cameras in a 3D space, facilitating the computational refocus and parallax-based depth estimation. We expect this work will lay the foundation for future light field camera lens design development, particularly in biomedical applications where diagnosis and treatment heavily rely on the accuracy of the 3D measurement [38–40].

Noteworthily, our current optical design pipeline is applicable to only ray optics models. This premise holds valid for light field cameras with a relatively small aperture, such as a light field endoscope. For large NA imaging, to account for the diffraction effect that occur when recording the light field, we must adapt the design process for a wave optics model [41] instead. This study is out of the scope of current work, and we will leave it for future investigation.

Chapter 3 Hyperspectral imaging

Hyperspectral imaging (HSI) has been extensively employed in myriad applications such as biomedical imaging, remote sensing, and machine vision. HSI is a functional combination of a two-dimensional (2D) camera and a spectrometer, acquiring both the spatial (x, y) and spectral (λ) information of a scene. While capable of acquiring a hyperspectral datacube(x, y, λ), most existing HSI techniques require scanning either in the spatial [42–44] or spectral domain [45–48]. Because only a small portion of the hyperspectral datacube can be seen at a time by these imagers, the scanning mechanism significantly jeopardizes the light throughput and hinders the imaging of dynamics.

In contrast, snapshot HSI techniques overcome the above limitations by acquiring all hyperspectral datacube voxels in parallel. Representative techniques encompass coded aperture snapshot spectral imaging (CASSI) [49–51], computed tomography imaging spectrometry (CTIS) [52,53], and image mapping spectrometry (IMS) [54–61]. While all these techniques operate in a snapshot format, only IMS features 100% light throughput while maintaining a compact form factor and high computational efficiency, making it suitable for real-time field imaging applications.

3.1 Image mapping spectrometer (IMS)

The IMS operates by mapping a three-dimensional (3D) hyperspectral datacube to a 2D camera through an angled mirror array, referred to as an image mapper. The underlying principle is detailed in Ref. [62]. In brief, the image mapper splits an image into strips and redirects them to different locations on the 2D camera. Because the mirror facets on the array have varied tilts, a blank region is created between adjacent sliced images. A dispersion element, such as a prism,

then disperses the sliced image and fills this blank region with spectrum. In this way, each pixel on the 2D camera is encoded with the unique spatial and spectral information of the original hyperspectral datacube, and remapping generates the image. The image remapping during reconstruction requires a lookup table, which connects each voxel in the hyperspectral datacube to a pixel in the raw image. The calibration of this lookup table is nontrivial. Previous calibration methods, such as edge alignment [54] and point scan [62], suffer from either low image quality or slow calibration speed. For example, the edge alignment method images a target with a sharp edge and constructs the lookup table by aligning the image slices with respect to this feature. Because image slices experience different levels of distortion, the lookup table built upon local feature alignment cannot be faithfully applied to the global image, leading to a reduced image quality and spectrum accuracy [63]. In contrast, the point scan method builds the table by illuminating one datacube voxel at a time, followed by pinpointing the centroid of the impulse response image on the 2D camera. Because the calibration is performed on each datacube voxel, the reconstructed image quality and spectral accuracy are superior to those achieved by edge alignment. However, due to the reliance on scanning, the calibration process is time consuming, typically taking hundreds of hours to complete.

To overcome the above limitations, we herein present a fast and accurate calibration method, which we refer to as slit scan. Slit scan can correct for the same image slice distortion as with point scan; however, it does not have the need for prolonged 2D scanning, thereby significantly reducing the calibration time to tens of minutes. Moreover, we quantitatively analyzed the primary artifact in the IMS, the striped image, and we provided several solutions to correct it. The radiometric calibration was detailed in Ref [62].
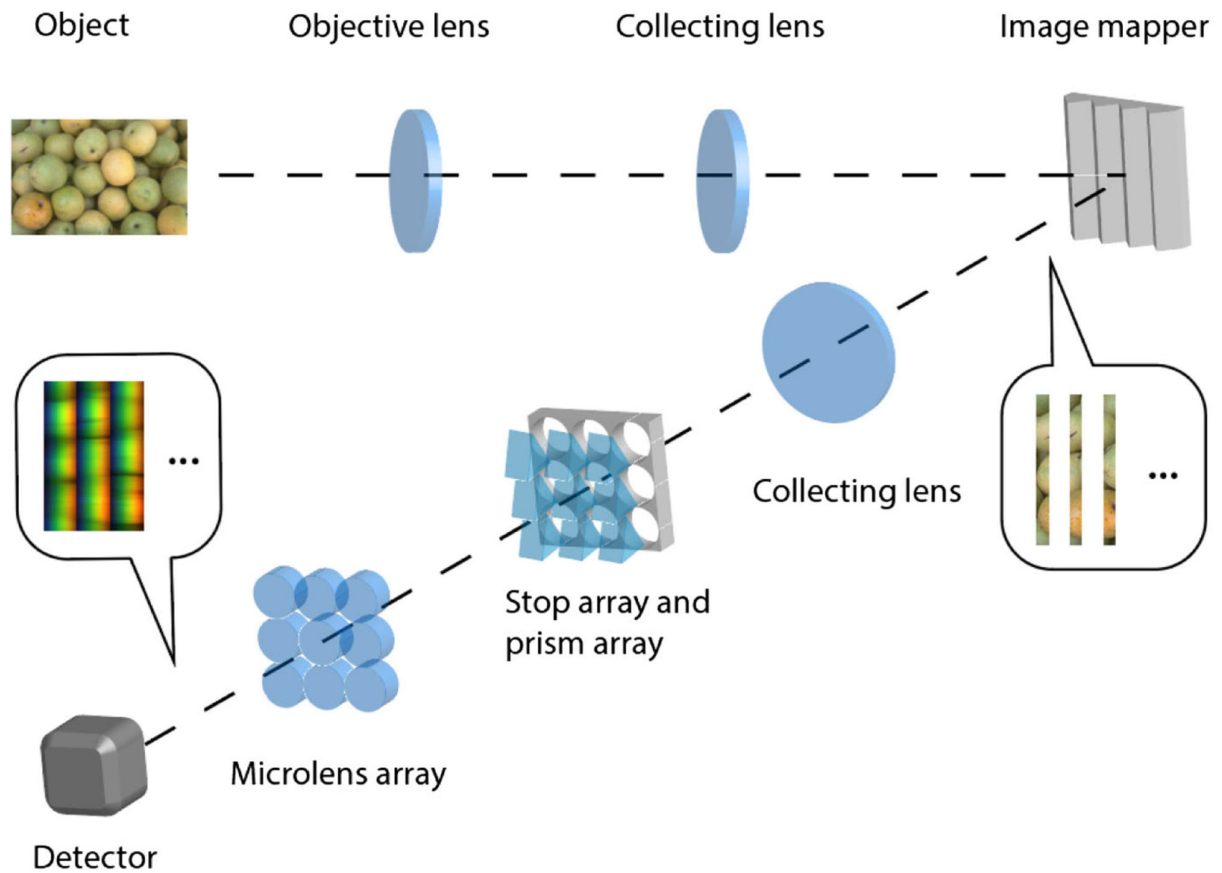
Figure 15. Schematic of the IMS.

A schematic of the IMS is shown in Fig. 15. The mapping mirror is placed at a plane that conjugate with both the object and the 2D camera. The object is imaged by an objective lens and a collecting lens, and an intermediate image is formed at the image mapper. As a result, each mirror facet on the image mapper reflects a slice of the intermediate image. Because the angles of the mirror facets are grouped into periodic blocks, the reflected image slices are directed to the correspondent subfields. In our experiment, a total of 480 image slices are grouped into 5×8 subfields, each containing 12 image slices. Within a subfield, there is a blank region between adjacent image slices. After passing through a dispersion prism, the spectrum of each image slice occupies this blank region. The dispersion prism array in our system works in a visible light range (400–700 nm). A bandpass filter is used to limit the input spectral range, thereby avoiding

25

spectral overlap between adjacent sliced images. Microlens array with 5×8 microlenses is used to focus the final image on the camera. Under monochromatic illumination, the width of each image slice is one pixel on the camera, and the width of a blank region between two adjacent image slices is 40 pixels. Under white-light illumination, these 40 pixels represent 40 different color channels in the visible spectral range. The parameters of the IMS are detailed in Refs. [58,59].

3.2 IMS calibration methods

The IMS maps a 3D hyperspectral datacube (x, y, λ) to a 2D image $(x\ , y\ )$ on the camera. If the point spread function (PSF) of the IMS is a delta function, the mapping between a voxel in the datacube and a pixel in the image would be one to one. The imaging process can be described by the following equation:

$$I(x, y) \ = \ T \cdot O(x, y, \lambda), \tag{5}$$

where I is the captured 2D raw image, O is the 3D hyperspectral datacube, and T is the mapping matrix of the IMS. The goal of calibration is to determine the inverse of T. With that being known, the original datacube can be easily reconstructed based on

$$O(x, y, \lambda) \ = \ T^{-1} I(x, y). \tag{6}$$

In practice, rather than calculating the mapping matrix, we experimentally built a lookup table, which contains the index of each voxel in the datacube and locations of the corresponding pixel in the image.

As mentioned, two methods have been developed for constructing the lookup table: edge alignment [54] and point scan [62]. Edge alignment simultaneously extracts all image slices in the raw image and rearranges them to form a spectral channel. In this case, the lookup table

26

contains the locations of image slices both in the datacube and in the raw image. The sequence of image slices in the raw image is determined by the design of the image mapper, and such knowledge is known as priors. Due to distortion, the image slices in the raw image are tilted and exhibit different magnification ratios. A common approach to extracting these image slices is to capture an empty field under uniform illumination, followed by image binarization. After this procedure, each image slice is isolated, and a simple curve fitting reveals the shape contour. Next, a target with a sharp edge that is perpendicular to the mirror facet direction is imaged, and all the resultant images slices are aligned with respect to this edge. The image processing pipeline is illustrated in Fig. 16 (a). Finally, the starting coordinates of the aligned image slices in the raw image are recorded to fill in the lookup table.
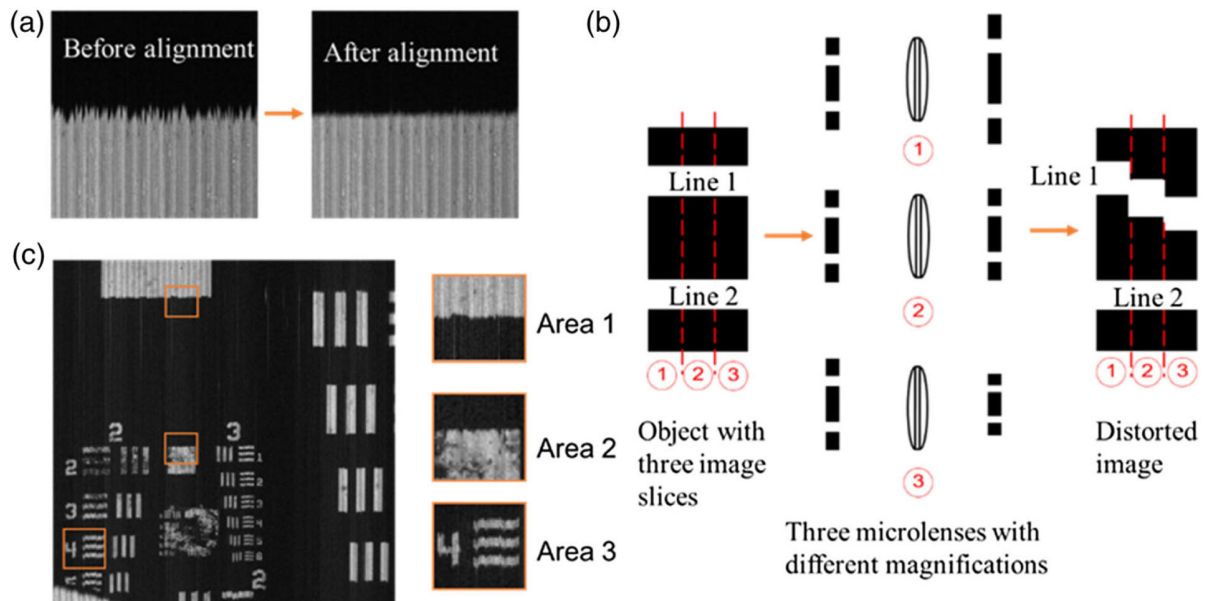
Figure 16. Edge alignment method. (a) Image slices alignment. (b) Distorted edge due to the variation in the magnification ratio of adjacent image slices. (c) Reconstructed image at 532 nm. The three insects show the zoomed view of the boxed areas. Area 3 shows the bars of group 2 element 4, indicating 5.66 lp/mm.

Despite being fast in implementation, the major drawback of edge alignment is that the

reconstructed image suffers from distorted edges. This is because the image slices experience varied magnification ratios when they pass through different lenslets in the array. As illustrated in Fig. 16(b), an object consisting of two straight lines is divided into three image slices. These three adjacent image slices are imaged by different lenslets with varied magnification ratios. If we extract these imaged slices from the raw image and align them with respect to line 2, the edge of line 1 would appear distorted. We further show a real-world example when imaging a US air force (USAF) resolution target. The reconstructed image through edge alignment is shown in Fig. 16(c). Area 3 shows the bars of group 2 element 4, indicating 5.66 lp/mm. All image slices were aligned with respect to a horizontal line that crosses the central field of view(FOV). As shown in the zoomed inset, Area 2 is close to the reference edge and thus has a high reconstruction quality. In contrast, Area 1 and Area 3 located at the upper and bottom FOV exhibit distorted edges.

In comparison, the point scan method constructs the lookup table by illuminating hypercube voxels one at a time and recording the location of the impulse response on the 2D camera. To calibrate the entire hypercube, one must raster scan a monochromatic point light source—such as a pinhole illuminated by narrow band light—across the 2D FOV. Because this method experimentally identifies the mapping relation for each voxel in the hypercube, the errors induced by the image distortion and ununiform magnification are corrected. However, the implementation of the point scan method is time consuming. For example, to calibrate an image of 500×500 pixels at a given wavelength, a total of 139 h is needed, using a dwell time of 2 s at each scanning step.

3.3 Slit scan calibration

To overcome previous limitations and enable fast and accurate IMS calibration, we herein present slit scan, a method that maintains the accuracy of point scan but at a speed orders of magnitude faster. Simply put, rather than scanning a point source across the 2D FOV, we use a uniformly illuminated slit and scan only the axis parallel to the mirror facet to construct the lookup table.

Our method has been made possible due to the unique optical architecture of the IMS, where the image mapper slices the incident field into strips and projects them to different locations of the detector array. Upon illumination by a thin slit source that is perpendicular to the slice direction, the image of the slit is divided into a series of points, which create impulse responses separable in space in the raw image. Pinpointing the centroids of these impulse responses then generates the lookup table for all the voxels illuminated in parallel. In essence, this method uses the image mapper to perform point scans on all the points of the slit simultaneously.

The experimental setup is shown in Fig. 17(a). A10 μm optical slit (Lenox Laser, G-SLIT-1 DISC-10) was placed at the object plane of the IMS, where the intermediate image of the slit at the image mapper is perpendicular to the mirror facet. The slit is mounted on a motorized translation stage (Thorlabs, MTS50-Z8) for scanning. We chose the width of the slit to match with the size of a resolution cell at the object plane. The image of the slit is then sliced by the mirror facets. Because the length of the slit is longer than the width of the FOV, each mirror facet creates a unique response on the detector array. A representative raw image is shown in Fig. 17(b). By scanning the slit along the direction of mirror facets and recording the locations of all responses, we obtained a complete lookup table.
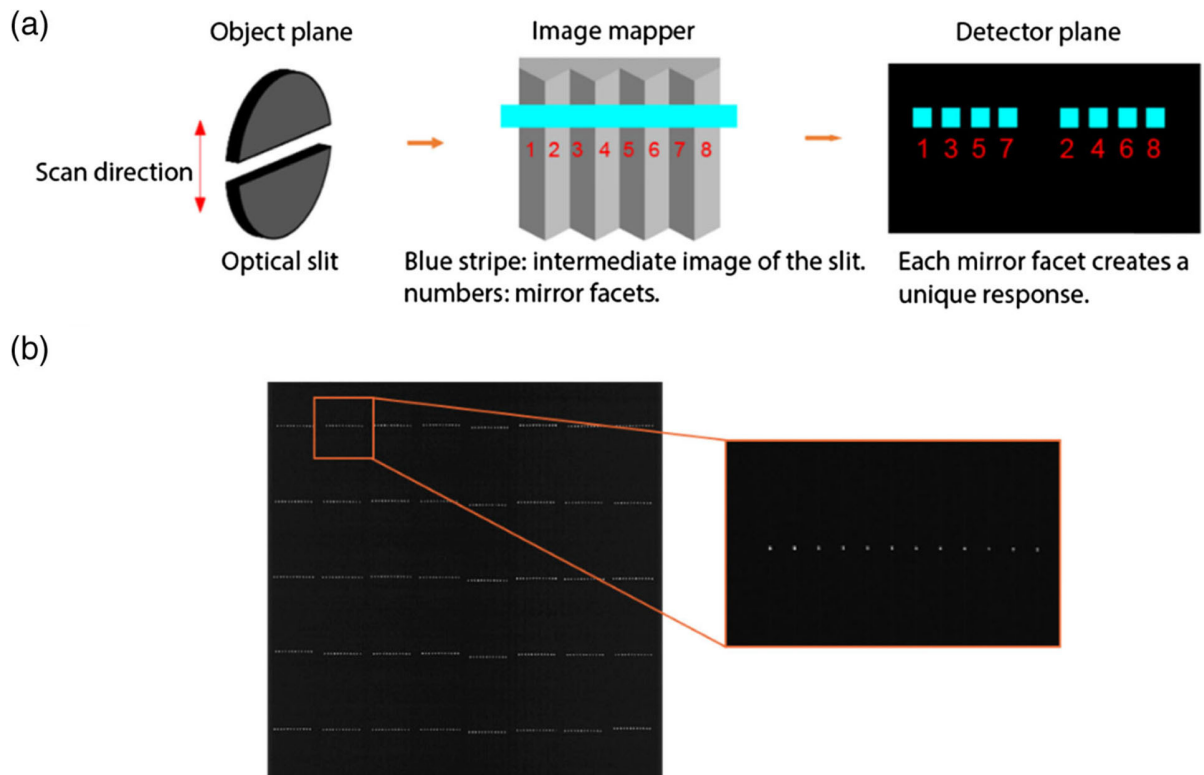
Figure 17. Slit scan method. (a) Image formation. (b)Mapped image of the sliced slit.

Compared with the edge alignment and point scan methods, slit scan offers two prominent advantages. First, slit scan calibrates the mapping for each voxel in the hypercube, and therefore it automatically corrects for the image distortion and varied magnification in the IMS. A reconstructed image of a USAF resolution target at 532 nm is shown in Fig. 18. Compared with the result obtained through edge alignment in Fig. 17(d), the reconstructed image shows no distorted edges. To quantitatively compare the reconstruction quality, we calculated the standard deviation of the edge in Area 1 from both results. Here we define the standard deviation of a horizontal edge as the variation in its pixel index along the vertical axis. The standard deviation (STD) of the edge reconstructed through edge alignment is 1.47 pixels, compared with 0.55 pixels obtained through slit scan, the reconstruction accuracy has been improved by 60%. The STD of the edge in Area 2 is 0.31 pixel, and the STD of the edge of one bar in Area 3 is 0.56

pixel in Fig. 18. Second, slit scan can be implemented and performed rapidly. In our experiment, a wavelength layer of the hypercube (i.e., a spectral channel image) has 445×480 voxels, where 480 is the total number of mirror facets on the image mapper. Because the scanning pitch is one voxel, only 445 steps are required. Again, with a 2s dwell time for each step, calibration takes 30 min to build the entire lookup table. In contrast, the point scan method needs to scan a total of 445×480 steps. Therefore, our method increases the calibration speed by a factor of 480, which is equal to the total number of mirror facets on the image mapper. This dramatic speed improvement, which we refer to as the parallelization advantage, becomes more critical for high-resolution IMSs with a large number of mirror facets.
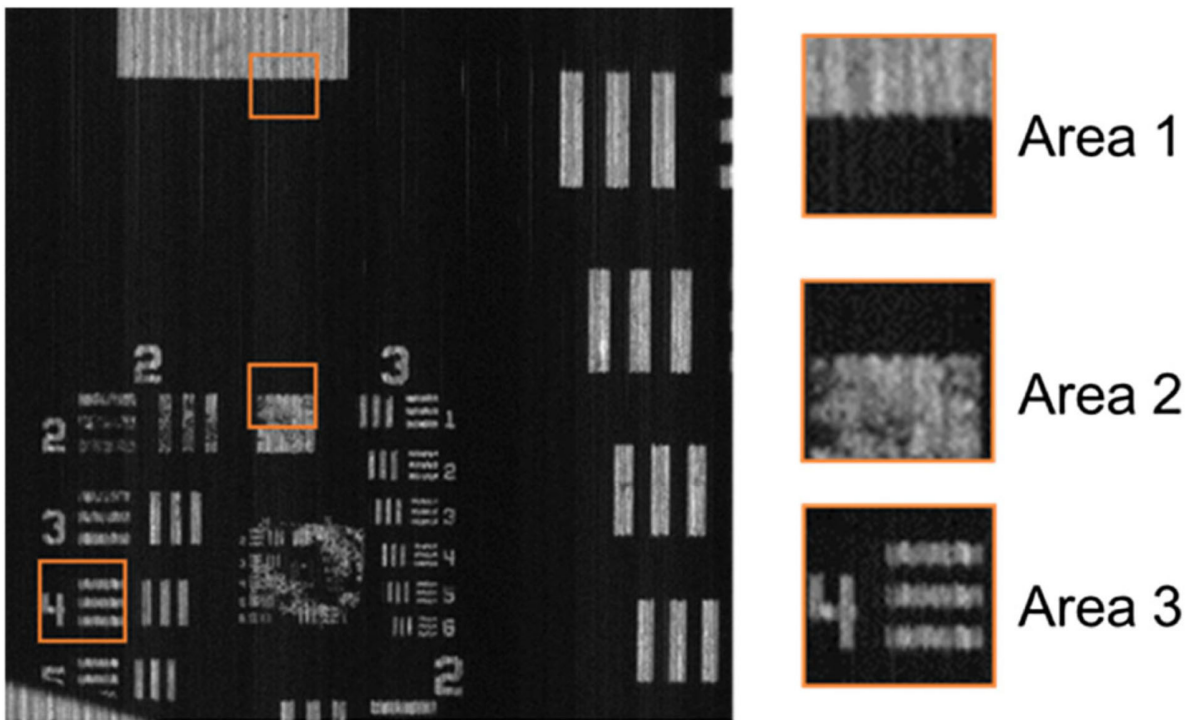
Figure 18. Reconstructed image of a USAF resolution target through slit-scan calibration. Compared to the result in Fig. 2(c), the standard deviation of the edge in Area 1 decreases from 1.47 pixels to 0.55 pixels.

Note that it is sufficient to calibrate the lookup table at only one wavelength. As shown in Fig. 19(a), the white stripe represents an image slice on the camera, and it disperses horizontally, as denoted by the rainbow area. The lookup table at other wavelengths can be obtained by shifting the recorded locations of responses horizontally, and the shifted distance in pixels can be deduced from the chromatic dispersion curve of the prism as shown in Fig. 19(b). In order to measure the dispersion curve, an optical pinhole was placed at the object plane, and six color filters (Thorlabs, FB470-10, FB510-10, FB550-10, FB590-10, FB630-10, and FB670-10) were used sequentially to find the relationship between the wavelength and pixel locations. Considering the spectral resolution of the current system (5 nm), a spectral bandwidth of 10 nm occupies only two pixels, leading to a negligible localization error. To further minimize the localization error of the wavelength responses, we performed multiple measurements and averaging. The final curve was fitted with a second-order polynomial based on following approximation:

$$\delta(\lambda) \approx [n(\lambda) - 1]\alpha, \tag{7}$$

where λ is wavelength of light, $\delta(\lambda)$ is the wavelength dependent dispersion, $n(\lambda)$ is the refractive index of the prism glass, and α is the angle of the prism. The relation between reflective index and wavelength follows the equation:

$$n(\lambda) = A + \frac{B}{\lambda^2}, \tag{8}$$

where A and B are constant coefficients.
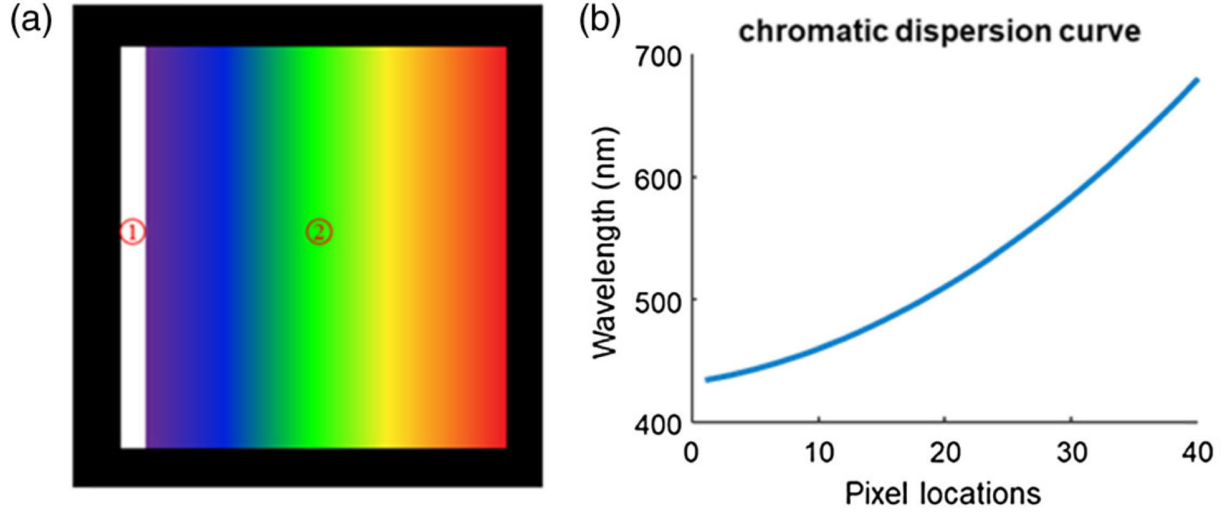
Figure 19. Wavelength calibration. (a) Illustration of a dispersed image slice on the detector. Area 1 is the image slice, and Area 2 is the spectrum. (b)Dispersion curve of the prism.

3.4 Stripe artifact

Although the slit-scan method can generate an accurate lookup table, the process itself does not correct for the image slices' intensity variations due to the reflectivity difference of the mirror facets and the defocus of the microlens array. The subsequently induced artifact, referred to as the striped image, degrades the image quality in the IMS. Here we develop a mathematical model to quantitatively analyze this artifact.

We describe the effect of mirror reflectivity and microlens defocus on the image quality using the line spread function (LSF). Although other optical aberrations may also exist in the system, we found that the defocus is the dominant factor that accounts for the stripped image artifacts in the IMS. The image of a mirror facet can be written as:

$$I_m(x,y) = L_m(x) * I'_m(x,y) \tag{9}$$

where $I_m(x,y)$ is the LSF of the mth mirror facet, * denotes convolution, and $I_m(x,y)$ depicts the geometric image of the mth mirror facet. Two representative LSFs are

33

shown in Fig. 20(a). The intensity profile along the dashed line is plotted in Fig. 20(b). The full width at half-maxima (FWHM) of these two LSFs are different. To reconstruct a spectral channel image, our calibration method extracts the peak light intensity from only one pixel and maps it back to the hypercube. The peak intensity is sensitive to both the microlens defocus and mirror reflectivity on the image mapper: the low mirror reflectivity reduces the total intensity of the LSF; on the other hand, the defocus broadens the LSF and, therefore, decreases the peak intensity. The nonuniform peak intensity introduces the stripe artifact. As an example, we show an image of a sharp edge under uniform illumination in Fig. 20(c).
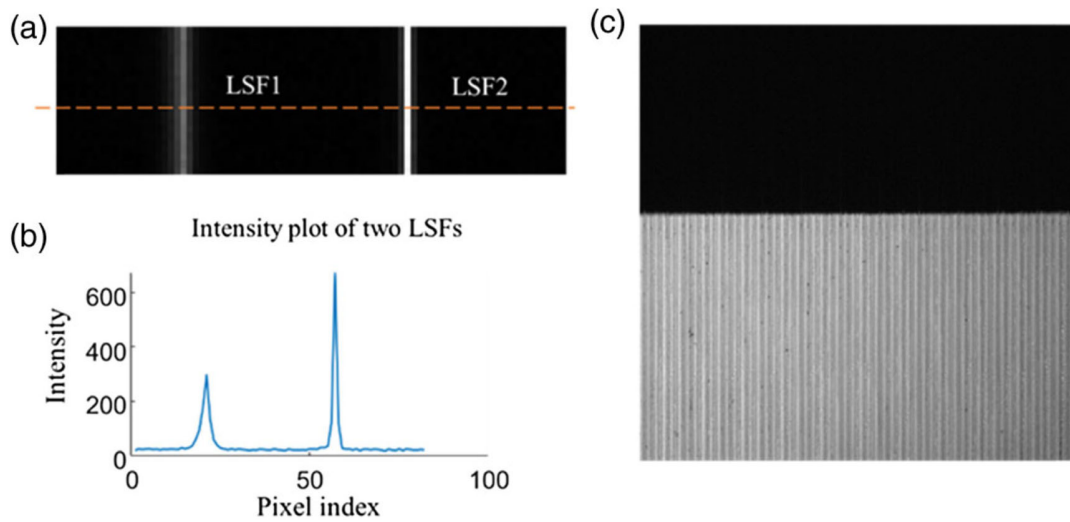
Figure 20. Line spread functions and stripe artifact. (a) Two example line spread functions (LSFs). (b) Intensity plot along the orange dashed line in (a). (c) Image of a sharp edge with stripe artifacts.

To remove the stripe artifact, we developed three methods: Fourier domain filtering, deconvolution, and datacube normalization. Fourier domain filtering is a fast method to remove the stripes of specific frequencies. The image processing pipeline is shown in Fig. 21(a). The dominant striping frequency is associated with the periodic mirror facet blocks on the image mapper as illustrated in Fig. 21(b). For simplicity, we assume the image mapper consists of three

mirror facet blocks, each containing two mirror facets. The incident field is then divided into six image slices. Slices 1, 3, and 5 are imaged by microlens 1, while slices 2, 4, and 6 are imaged by microlens 2. If the two microlenses have different levels of defocus, the intensities of these two groups of image slices would be different. Therefore, the dominant frequency of stripes in the reconstructed image is determined by the total number of mirror facets within one block. By applying notch filters in the Fourier domain, we can readily remove the correspondent stripes. However, this method cannot eliminate stripes at other frequencies, which are induced by the nonuniform reflectivity of the mirror facets and the nonuniform defocus of the microlenses. Additionally, the spatial frequency information at the same frequency range as that of the stripes is lost. As an example, we applied the Fourier domain filtering method to the image in Fig. 20(c), and the result is shown in Fig. 21(c). The stripes within the white area are effectively removed.
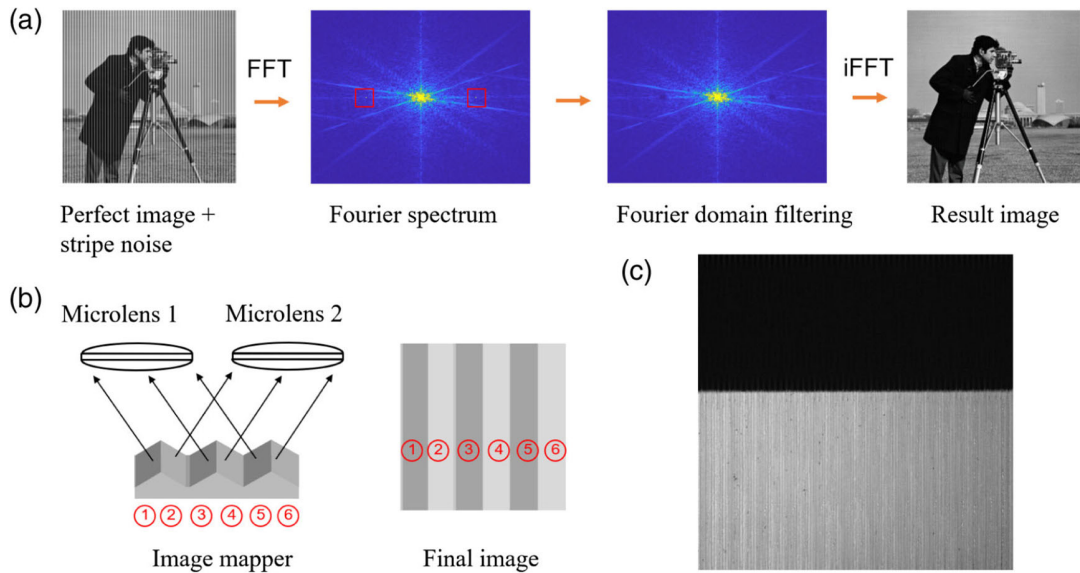
Figure 21. Image processing pipeline of Fourier domain filtering. (a) Schematic pipeline. FFT, Fourier transform; iFFT, inverse Fourier transform; red square, notch filter. The frequencies within the notch filter are blocked. (b) The dominant frequency of stripes is determined by the total number of mirror facets within one block. (c) Fourier domain filtering result of the image shown in Fig. 6(c). The stripes within the white area are effectively removed.

In contrast, deconvolution can suppress the stripe artifacts at all frequencies. By imaging an empty field illuminated by a monochromatic light source, we can simultaneously measure all combined LSFs in the raw image. This knowledge can then be used for deconvolution, compensating for the intensity variation of image slices. Additionally, deconvolution increases the spectral resolution by reducing the width of LSFs, which is also the impulse response of monochromatic light. However, because deconvolution is sensitive to noise, the performance of this method is highly dependent on the signal-to-noise ratio (SNR) of the image.

Lastly, datacube normalization captures an empty field under uniform illumination and constructs a reference hypercube. Dividing the real measurement by this reference cube yields the corrected results. The image processing pipeline is shown in Fig. 22. This method is essentially a simplified version of deconvolution, provided that the LSF of an image slice is a delta function. Therefore, the results are accurate only when the defocus is negligible.
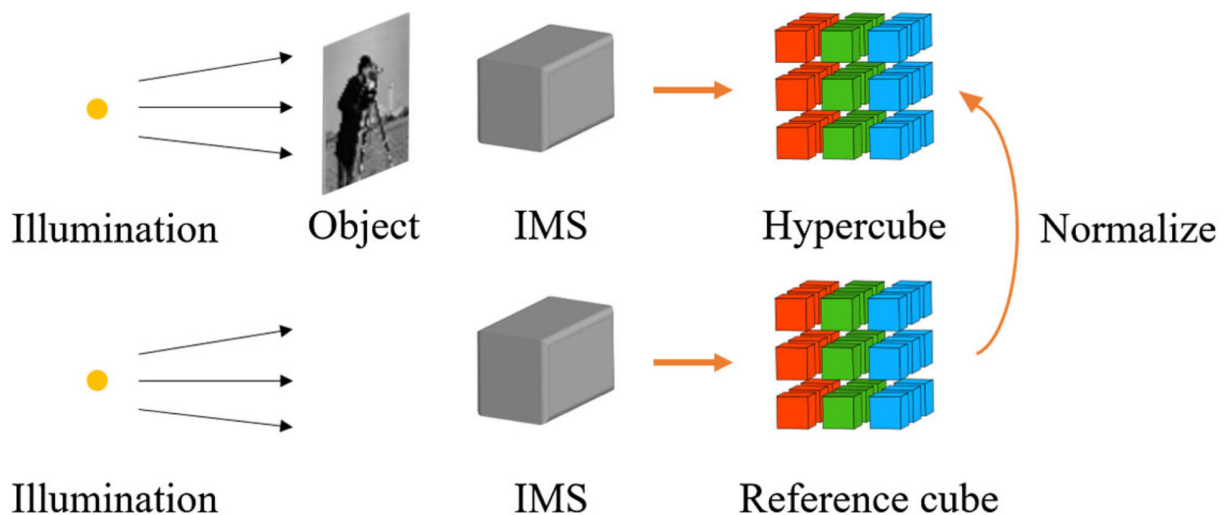
Figure 22. Image processing pipeline of datacube normalization.

To quantitively evaluate the stripe artifacts in the IMS, we define the striping factor as:

$$c = \frac{std}{mean} \tag{10}$$

for an image corresponding to an empty field under uniform illumination. Here mean is the mean intensity of the image and std is the standard deviation. The quotient c is zero if the image has no stripe artifact. In order to compare the performance of the proposed methods, we simulated an IMS with different levels of Gaussian noise and defocus. The Gaussian noise has a zero mean, and its standard deviation is normalized by the peak intensity in the image slice without any defocus or noise. Other simulation parameters and representative results are summarized in Fig. 23. We simulated a transparent object that alters the spectrum of the original illumination, and we generated three raw images: a reference image corresponding to an empty field uniformly illuminated by a light source with a rectangular spectrum centered at 532 nm and with a 50 nm bandwidth, an image that contains the LSFs under monochromatic illumination, and an image with the transparent object with a Gaussian transmission spectrum centered at 532 nm and with a 12 nm FWHM. Next, we reconstructed a spectral channel image at 532 nm and corrected for the stripe artifact using the three methods (Fourier domain filter, deconvolution, and datacube normalization).We calculated the metric c before and after we applied the correction. As shown in Fig. 23, Fourier domain filtering becomes more effective along with an increase in defocus, as it removes the dominant stripes introduced by the microlens array. Deconvolution outperforms other methods when the noise level is low, and datacube normalization fails when defocus is large.

37

	Gaussian Noise		Striping Factor c	Striping Factor c	Striping Factor
Wavelength			532 nm		
Number of microlenses			40		
Image slice per microlens			12		
NA of microlens			0.125		
Pixel pitch of the camera			7.4 μm		
Reflectivity of mirror facets			Uniform distribution from 95% to 100%		
Defocus of microlenses			Gaussian distribution, mean $= 0$, standard deviation $= \sigma$		
Method Name	**Standard Deviation**	**Defocus σ (in mm)**	**(before)**	**(after)**	**Change (in Percent)**
Fourier domain filtering	0	0.05	0.0158	0.0148	−6.33%
	0	0.1	0.022	0.0188	−14.55%
	0	0.2	0.0615	0.0433	−29.59%
	0.01	0.1	0.0244	0.0215	−11.89%
	0.1	0.1	0.1062	0.1056	−0.57%
Deconvolution	0	0.05	0.0158	0	−100%
	0	0.1	0.0221	0	−100%
	0	0.2	0.0615	0.0012	−98.05%
	0.01	0.1	0.0244	0.0105	−56.97%
	0.1	0.1	0.1062	0.1081	+1.79%
Datacube normalization	0	0.05	0.0158	0.004	−74.68%
	0	0.1	0.0221	0.0155	−29.86%
	0	0.2	0.0615	0.0562	−8.62%
	0.01	0.1	0.0244	0.0213	−12.7%
	0.1	0.1	0.1062	0.1503	+41.53%

Figure 23. Parameters and Stripe Artifact Correction Results of a Simulated IMS.

Finally, we performed a tolerance analysis for the microlens array defocus based on the criterion defined. We use c <0.05 as a criterion for acceptable image quality. Here we consider only the defocus and neglect other lens aberrations. To establish the relationship between the defocus of the microlens and the corresponding peak intensity in the image slice, we simulated our system using a wave propagation method, and the result is shown in Fig. 24. The wavelength used for simulation is 532 nm. All intensities were normalized by the peak intensity in the image slice without defocus. Note that the resultant curve is dependent on both the detector pixel size and the numerical aperture (NA) of the microlenses. If defocus exists, the Strehl ratio of the system decreases; however, because the signals measured by a pixel is ensquared energy, the peak intensity of the LSFs may not change as much as the Strehl ratio. On the other hand, for a fixed defocus, a smaller NA microlens has a larger depth of focus. Therefore, the image slice has a higher peak intensity compared to using a large-NA microlens. For our system, the detector pixel

38

size is 7.4 μm, and the NA of the microlens array is 0.125.
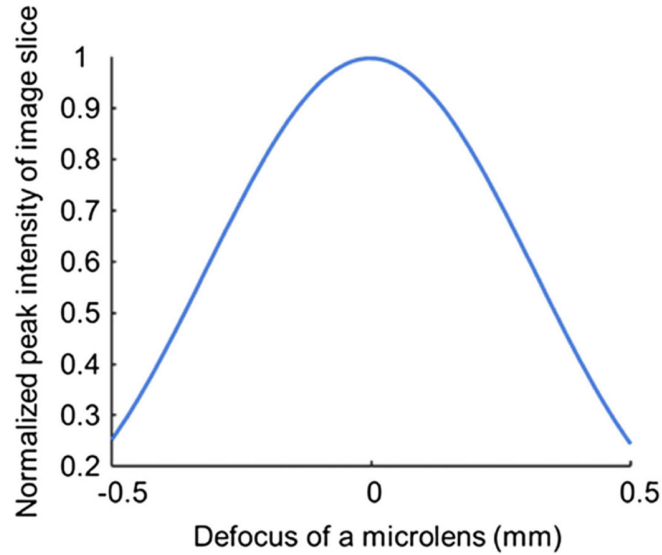
Figure 24. Normalized peak intensity in image slice versus defocus of
the microlens.

Next, we ran a Monte Carlo simulation. The parameters used for simulation are listed in Table 2.
A group of microlenses with different levels of defocus was generated, and the probability
density function (PDF) of the defocus of each microlens was assumed as a Gaussian distribution
with a zero mean and a σ standard deviation. The number of simulated microlenses in a group is
the same as the number of microlenses used in our system. The striping factor c was then
calculated for this group of microlenses based on the peak intensities, which can be deduced
from the result in Fig. 24. For each σ, we generated 1000 groups of microlenses and calculated
corresponding striping factor c. The goal of the simulation is to find a threshold σ, so that 90% of
microlens groups meet the acceptable image quality criterion (i.e., c <0.05). The results are
shown in Fig. 10. In our case, the threshold σ is 0.073 mm, indicating that 95% of microlenses
should have a defocus of less than 2 σ or 0.146 mm if operating the IMS at 532 nm.

In summary, we developed a slit-scan method for constructing a lookup table for the IMS. By

39

using an optical slit that is perpendicular to the mirror facets and scanning the FOV along one axis, we increased the calibration speed by 2 orders of magnitude. Also, we improved image reconstruction accuracy by 60% when imaging a resolution target. Moreover, we proposed three methods to reduce the stripe artifacts, and we compared their pros and cons. We expect this work to lay the foundation for future IMS development.

3.5 Tunable image projection spectrometer (TIPS)

A hyperspectral camera captures both spatial (x, y) and spectral information (λ) of input scenes, providing both intrinsic and discriminative spectral characteristics of objects for target recognition and classification. Based on the data acquiring mechanism, hyperspectral cameras are generally stratified into three categories: spatial scanning [42–44], spectral scanning [45–48], and snapshot [49–61].

While spatial-scanning hyperspectral cameras capture the spectrum of a spatial point or line at a time, their spectral-scanning counterparts acquire a two-dimensional (2D) image at each wavelength. Despite a relatively simple optical architecture, spatial-scanning hyperspectral imagers bear a lengthy acquisition—to measure a (x, y, λ) datacube, the system must perform a complete scan of all spatial locations or spectral wavelengths. In contrast, snapshot hyperspectral imagers can acquire all datacube voxels in parallel, thereby maximizing the light throughput. However, current snapshot hyperspectral systems are plagued by various problems, such as extensive computation [49-51], low image quality [52], and complex configuration [57-60]. The primary challenge for measuring a hyperspectral datacube is dimensionality reduction because most electronic detectors are in two-dimensional (2D), one-dimensional (1D), or zero dimensional (0D) format. Direct mapping a three-dimensional (3D) hyperspectral datacube (x, y, λ) to a 2D detector array often leads to a trade-off between spatial, spectral, and temporal

resolutions.

Compressed-sensing-based techniques solves this problem by utilizing the sparsity of natural scenes and measuring a hyperspectral datacube with much fewer measurements than that required by the Nyquist-Shannon sampling theorem [64-67]. Nonetheless, their applications are restricted by a static optical architecture and a fixed compression ratio. For example, an imager with a high compression ratio cannot be applied to a complex object, while an imager with a low compression ratio is ineffective in measuring a simple scene.

To address this unmet need, we present tunable image projection spectrometry (TIPS), a Fourier-domain line-scan spectral imager with a tunable compression ratio. Compared to state-of-the-art spatial-domain pushbroom hyperspectral cameras, TIPS requires much fewer measurements and provides a higher light throughput. Using a rotating Dove prism and a cylindrical field lens, TIPS scans an input scene in the Fourier domain and captures a subset of multi-angled one-dimensional (1D) en face projections of the input scene, allowing a tailored data compression ratio for a given scene. We demonstrate the spectral imaging capability of TIPS with a hematoxylin and eosin (H&E) stained pathology slide. Moreover, we showed the spectral information obtained can be further converted to depths when combining TIPS with a low-coherence full-field spectral-domain interferometer.

The schematic of TIPS is shown in Fig. 25. The object plane is located at the front focal plane of an objective lens L1 (f = 50 mm, f /# = 2). L1 and L2 (f = 50 mm, f /# = 2) together form a 4f system, and a Dove prism (Thorlabs, PS992M) serves as the system stop and locates at the Fourier plane. The Dove prism is mounted on a motorized rotation stage (Thorlabs, PRM1Z8). When the Dove prism is rotated by an angle of θ, the image of the object is rotated by 2θ. We positioned a cylindrical lens (f = 15 mm, invariant axis is along y axis) 35 mm after L2. Here the cylindrical lens plays two roles: generating 1D en face projection along y axis like those in

standard computed tomography (CT) and serving as a field lens to reimage the system stop to a slit plane (Thorlabs, S15K) along x axis. The tracing of chief and marginal rays in the two orthogonal planes is illustrated in Fig. 1(b). It is worth noting that along the y axis, the object plane is conjugated to the slit plane, while alone x axis the stop plane is conjugated to the slit plane. In y-z plane, the cylindrical lens is effectively a plane-parallel plate (PPP) and transparent to the object. The focus shift and spherical aberration introduced by the cylindrical lens are compensated by shifting the slit plane along the optical axis. In x-z plane, the cylindrical lens integrates the intermediate image along x axis to form en face projection. Moreover, it demagnifies the pupil image to increase the light throughput. After passing the slit, the line image is spectrally dispersed by a diffraction grating (600 groves/mm) and reimaged by another 4f system (L3 and L4, f = 50 mm, f /# = 1.4) to an area camera (Thorlabs, CS235MU).
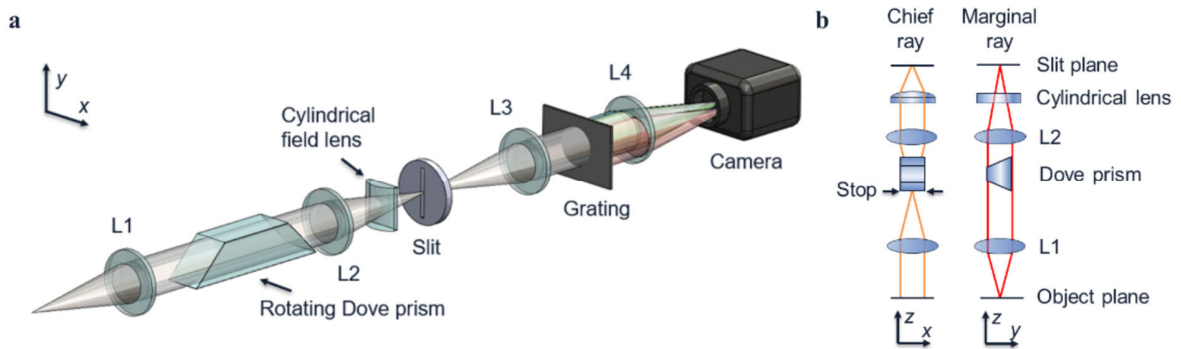
Figure 25. Optical system of a tunable image projection spectrometer (TIPS). (a) System schematic. (b) Chief ray and marginal ray path in x-z and y-z plane.

The image reconstruction pipeline is shown in Fig. 2. At each Dove prism rotation angle, we capture a dispersed projection (x, λ). To minimize the correlations between projections and maximize information content for reconstruction, projections are sampled uniformly in an angular range [0, π]. Upon completion of acquisition of a (x, λ, θ) datacube at selected angles, the

datacube is rearranged to a wavelength-dependent sinogram stack (x, θ)λ. Next, we construct a 2D spatial image (x, y) using the sinogram (x, θ) at each wavelength through filtered inverse Radon transformation. Other algorithms such as fast iterative shrinkage-thresholding algorithm (FISTA) [68] or deep learning can be used to further increase the image quality at the expense of an increased computation cost [69]. The final spectral datacube (x, y, λ) is computed by converting sinograms at all wavelengths to the corresponding spatial images.
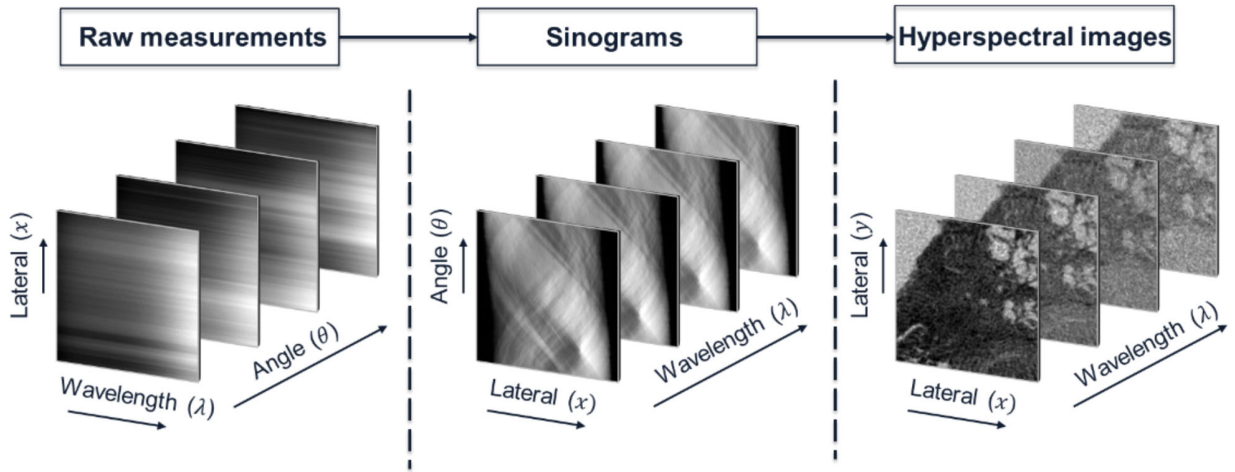
Figure 26. Image reconstruction pipeline.

3.6 Imaging results using TIPS

To measure spatial resolution, we directly imaged a negative USAF resolution target at the object plane (i.e., the front focal plane of L1) under monochromatic illumination at 532 nm. Like the sparse-view computed tomography (CT), TIPS measures only a subset of projections, whose total number is smaller than the pixel resolution of the input image [70]. Therefore, the corresponding inverse problem is under-determined and, generally, the more projections the system acquires, the higher the image quality. Reconstruction results with the total number of projections = 90, 45, and 15 are shown in Fig. 3(a), (b), and (c), respectively. In each panel, the

sinogram is shown at upper-left, and the reconstructed image is shown at the bottom, and a boxed zoom-in area from the reconstructed image is shown at upper-right. We define the data compression ratio γ as:

$$\gamma = \frac{N_y}{N_\theta},\tag{11}$$

where N_y is the number of pixels along the y axis in the recovered image, and N_θ is the number of projections. The reconstructed image has a resolution of 300×300 pixels, corresponding compression ratios 3.3, 6.6, and 20 in Fig. 3(a), (b) and (c), respectively. The spatial resolutions were measured by calculating the visibility of bar features in each reconstructed image with a threshold 0.2, and the results are 7 μm, 11 μm, and 16 μm, respectively. It is worth noting that a USAF target is generally considered a non-sparse object. Therefore, the image quality is more sensitive to the projection number than a sparse object, evidenced by the degraded image quality as the compression ratio increases. By contrast, when imaging a sparse object like a butterfly pattern, a much fewer number of projections are sufficient to recover a high-fidelity image (detailed in section 3.7). Therefore, the tunable compression ratio allows an effective measurement for a given object.

The spatial resolution of TIPS is fundamentally limited by diffraction like in a conventional optical system. However, because TIPS captures only a limited number of projections, the image quality is practically determined by the sparsity of the object like that in sparse-view CT. In TIPS, we can tune the number of projection measurements based on the prior knowledge of object sparsity and, thereby, provide balanced spatiotemporal resolutions.

When imaging the USAF target, we set the exposure time to $\sim$ 15 ms for each frame. With an additional 200 ms for rotating the dove prism between adjacent steps, the total acquisition time is $\sim$ 20 seconds.
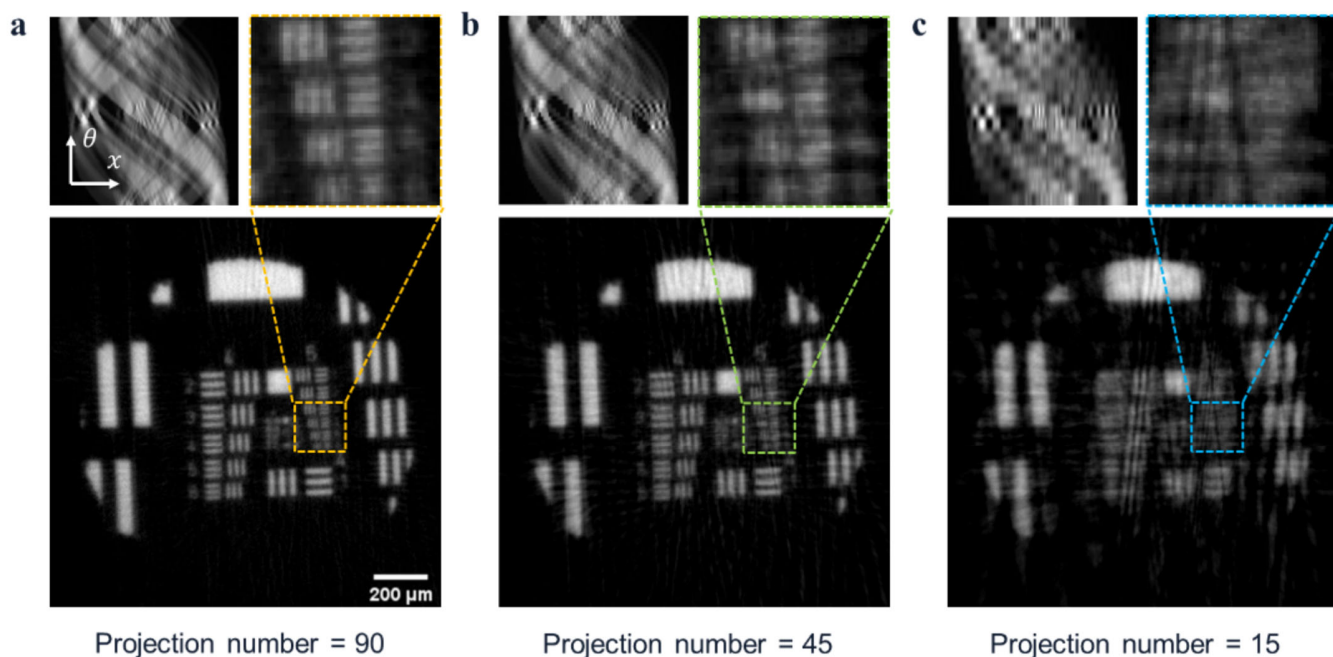
Figure 27. Reconstructed USAF resolution target with (a) 90 projections, (b) 45 projections, and (c) 15 projections. In each panel, upper-left: sinogram; bottom: reconstructed image; upper-right: a boxed area in the reconstructed image.

To evaluate spectral resolution, we placed a pinhole (Thorlabs, P20D) at the object plane of TIPS, and illuminated it with monochromatic light of different wavelengths. The corresponding pixel locations of the projections were recorded. The results were fitted with a linear polynomial, as shown in Fig. 28(a). The slope of the line indicates the spectral sampling. Because 1 nm bandwidth in wavelength occupied 5.5 pixels on the camera, the spectral sampling of the system is 0.18 nm. The spectral resolution was measured as the full-width half-maximum (FWHM) of the spectral response when the illumination wavelength is 532 nm. Here, we limited the source wavelength using a 1 nm bandpass filter (Thorlabs, FL532-1). The raw measurement result is shown in Fig. 28(b). The raw measurement has a FWHM of ~9 pixels. However, this width is a convolution of the geometric image of the pinhole on the camera (~3 pixels), the bandwidth of the light source (corresponding to ~6 pixels on the camera), and the system spectral resolution.

Because the width of a convoluted function can be computed as [71]:

$$w(f_1 * f_2 * f_3) = w(f_1) + w(f_2) + w(f_3) - 2 \qquad (12)$$

where w denotes the width of the function, $*$ denotes the convolution operator, and fi (i = 1:3) denotes the individual function in a discrete form, the width of the spectral resolution on the camera is derived to be 2 pixels. Given a 0.18 nm spectral sampling, the spectral resolution is, therefore, 0.36 nm.
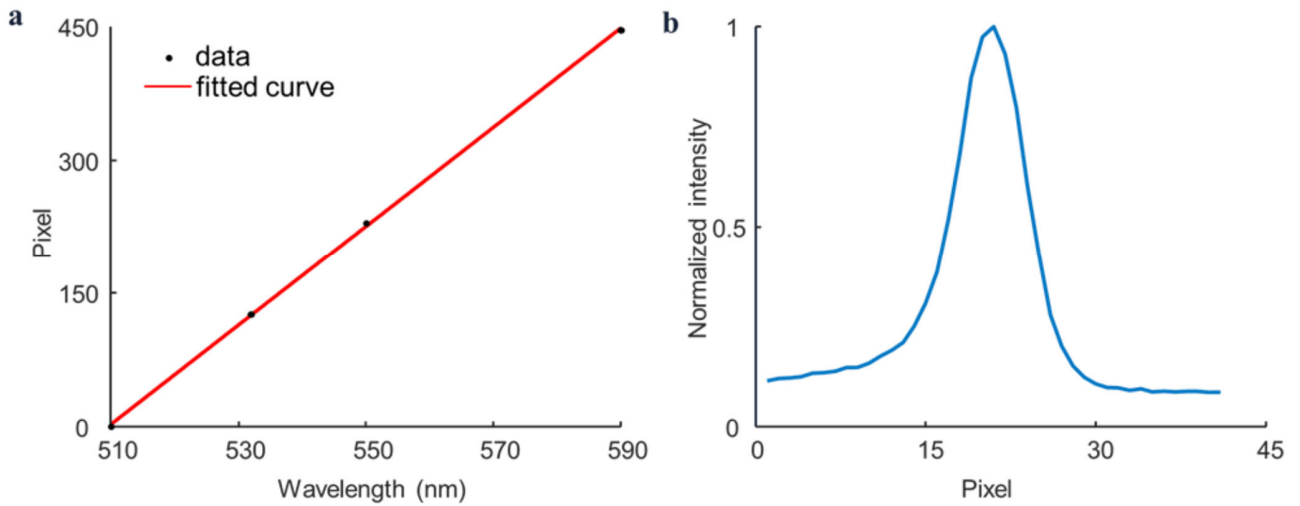
Figure 28. Spectral calibration. (a) Spectral response locations in pixel of different wavelengths on the camera. (b) Raw response under 532 nm illumination (bandwidth = 1 nm).

Next, we tested the system on a lung cancer hematoxylin and eosin (H&E) stained pathology slide illuminated by a broadband halogen lamp (Amscope, HL250-AS) in the transmission mode. Four represented reconstructed images at 530 nm, 550 nm, 570 nm, and 590 nm are shown in Fig. 29(a). The color of each image was rendered according to the corresponding wavelength based on CIE 1931 observer. As shown in the figure, hematoxylin (area 1) and eosin (area 2) have a significant spectral discrepancy in the 540-590 nm spectral range. Relative absorption

spectra of area 1 and area 2 are shown in Fig. 29(b). The relative absorption at each wavelength is calculated by subtracting the transmission intensity from the background intensity.
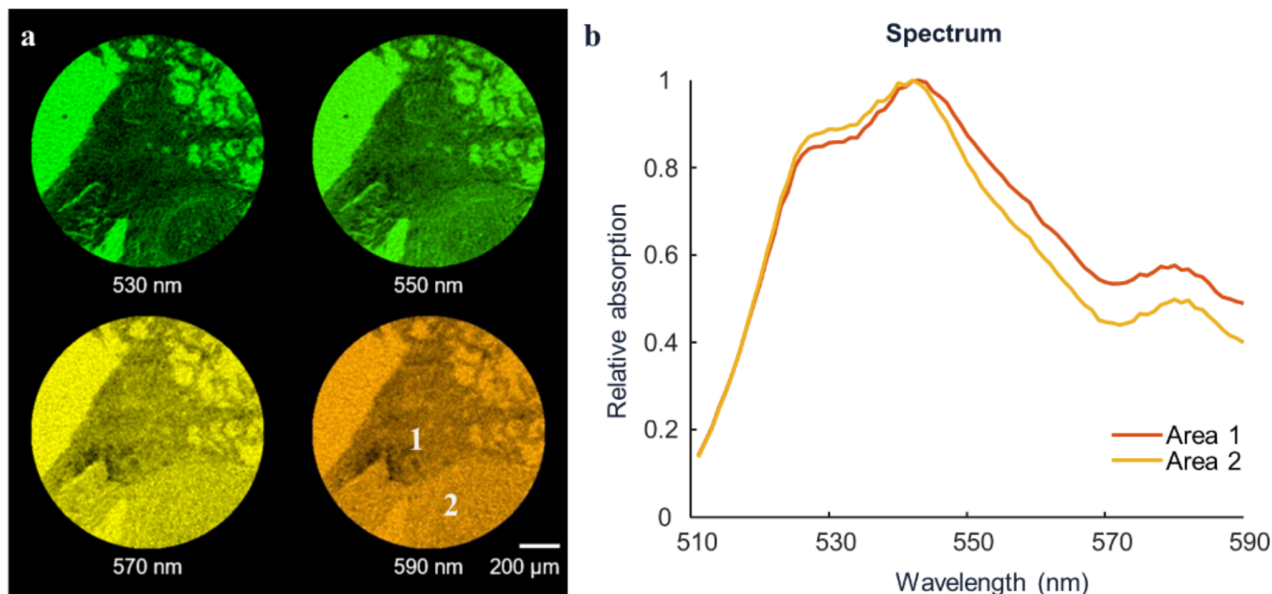
Figure 28. Spectral imaging of lung cancer hematoxylin and eosin (H&E) slide in transmission mode. (a) Spectral images at 530 nm, 550 nm, 570 nm, and 590 nm. (b) Relative absorption spectra of area 1 and area 2. The relative absorption at each wavelength is calculated by subtracting the transmission intensity from the background intensity.

We further demonstrated the combination of TIPS with a low-coherence full-field spectral domain interferometer [72,73]. The system layout is shown in Fig. 29(a), which consists of a Michelson interferometer and TIPS. We illuminate the sample with a low-coherent light source (a halogen lamp; Amscope, HL250-AS) filtered with a 40 nm bandpass filter centered at 550 nm (Thorlabs, FB550-40), yielding a 3.8 μm depth resolution and a 210 μm depth range in the final reconstructed volumetric image. The focal length of L1 and L2 are 50 mm and 150 mm, respectively. The scattered light from the sample combines with the light reflected from the reference mirror at a beam splitter, forming an interferogram at the input plane of TIPS. After acquiring the corresponding spectral datacube, we compute the depths by applying a Fourier transform to the spectrum acquired at each spatial location. To evaluate the system, we imaged

two crossed hairs extending along orthogonal directions at two different depths. The results are

shown in Fig. 29(b), where the reconstructed en face images are shown in the first row, B-scan

(i.e., cross-sectional) image is shown in the middle, and a 3D volumetric image is shown at the

bottom. The measured depths of hair 1 and 2 are 26.6 μm and 98.8 μm, respectively.
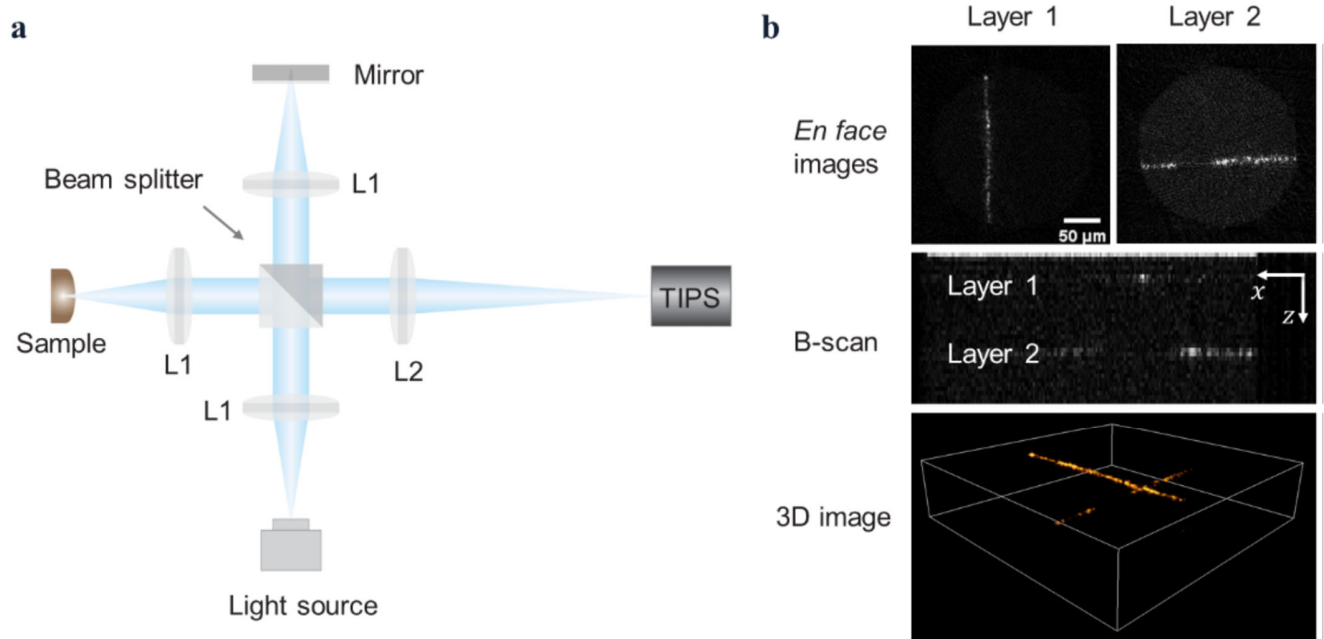
Figure 29. Depth imaging with full-field spectral domain interferometer.
(a) System layout. (b) Reconstructed results of two intersecting hairs at
two depths. First row, en face reconstructed results; middle, B-scan cross-
sectional image; bottom, a 3D volumetric image of two hairs.

3.7 TIPS discussion

Measuring a line projection at a time, TIPS is a counterpart of conventional pushbroom

spectrometry in the Fourier domain. However, TIPS outperforms pushbroom spectrometers in

both light throughput and signal-to-noise ratio (SNR).

First, Light throughput (Jacquinot advantage). In conventional pushbroom imaging

spectrometers, a slit is positioned at a conjugate image plane to measure only one line of the

image at a time. The line image is then spectrally dispersed by a diffraction grating or a prism. Given widefield illumination, the light throughput equals the ratio of the slit width to the conjugate imaging field of view (FOV) at the slit plane, i.e., η = dslit/FOV. On the other hand, the slit width determines the image resolution and the space-bandwidth-product (SBP), $SBP = (FOV/dslit)^2$. Therefore, there is a trade-off between the light throughput and the image resolution and SBP. In contrast, in TIPS, the slit is located at the pupil plane at the x-z plane. The light throughput equals the ratio of the slit width to the conjugate pupil diameter (D) at the slit plane, i.e., η = dslit/D, where D relates the stop size (D0) through D = D0/M. Here M is pupil demagnification ratio in the x-z plane, and $M = f_{L2}/f_{CL}$, where f_{L2} and f_{CL} are the focal lengths of lens L2 and the cylindrical lens, respectively. Because f_{CL} does not contribute to the image magnification in the y-z plane, we can use a cylindrical lens with a short focal length to increase the pupil demagnification ratio, reducing the conjugate pupil diameter (D) at the slit plane and, thereby, increasing the light throughput. Moreover, since TIPS scans in the Fourier domain, the slit width is decoupled with the image resolution, eliminating the trade-off between the light throughput and image resolution/SBP.

Second, SNR (Fellgett advantage). TIPS performs measurement through multiplexing signals in the spatial domain. If the measurement contains only signal-independent noise sources, such as random thermal noise or readout noise in the cameras, the multiplexing approach can enhance the SNR of the recovered signal, which is known as the Fellgett advantage [74,75]. Like that in Fourier transform infrared spectroscopy (FTIR), TIPS can improve the SNR by a factor of N$\sqrt{P}$ compared with its spatial-scanning counterpart. Here N denotes the image dimension in pixels, and P is the number of projections. Details derivation is provided in Refs [76].

Furthermore, we quantitatively compared the spectral responses of TIPS to a standard fiber spectrometer (STSIS-L-25-400-SMA, Ocean Optics, Inc.). We first captured an image of a color

checker target (ColorChecker Passport Photo 2, X-Rite) using TIPS and recovered the spectrum. Next, we measured the spectrum of the same checker using the fiber spectrometer. The normalized spectra obtained by TIPS and the fiber spectrometer are shown in Fig. 30(a), matching well with each other. The full spectral range of TIPS is ~150 nm (450-600 nm), which is limited by the bandpass filter in the system. In Fig. 28(b), we showed the results only in a selected spectral range (510-590 nm) because this range shows the most significant difference between the spectra of hematoxylin and eosin. On the camera, we read out only part of the chip that corresponds to the spectral range of the system.

To validate the depth accuracy in the combined system (a spectral-domain interferometer and TIPS), we first positioned a planar specular-reflection object (a mirror) at a plane conjugated with the reference mirror in the interferometer. Next, we translated the reference mirror along the optical axis from 25.4 μm to 203.2 μm with a step size of 25.4 μm. At each step, the recovered depth from interferometry matches well with the ground truth (Fig. 30(b)). The mean squared error is 1 μm.
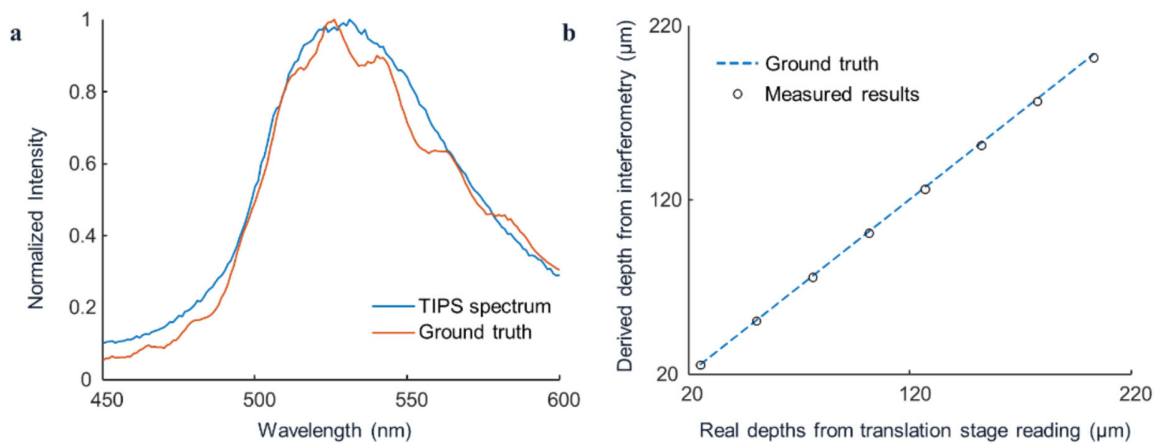
Figure 30. Accuracy of spectral and depth measurements. (a) Comparison between normalized spectra measured by TIPS and a standard fiber spectrometer (ground truth). (b) Comparison between derived depths from interferometry and ground truth.

In summary, we developed a new category of spectral imaging methods, TIPS, which can provide a tailored data compression ratio for a given target. TIPS also outperforms conventional pushbroom imaging spectrometers in both light throughput and SNR. Seeing its advantages, we expect TIPS can find broad applications in various disciplines.

Chapter 4 Hyperspectral light field imaging

Although light field and hyperspectral imaging techniques continue to mature, combining them on a single platform is non-trivial. Multi-camera systems are usually plagued by misalignment errors or a bulky setup and low-light efficiency [20–22]. To overcome the above limitations, herein, we present a snapshot hyperspectral light field imaging system that integrates light field imaging with the IMS. The resultant system can capture the spatial, angular, and spectral information of light rays in a single snapshot and with a single camera, measuring a multi-dimensional datacube of size 66×66×5×5×40 (x, y, u, v, λ).

Furthermore, measuring a high-dimensional plenoptic function faces two main challenges: dimension reduction and measurement efficiency. To address this problem, we developed a compressed-sensing five-dimensional imager, termed snapshot hyperspectral light field tomography (Hyper-LIFT).

4.1 Light field IMS

The optical schematic and a photograph of the system are shown in Figs. 31(a) and 31(b), respectively. A broad-band light source (Thorlabs, SLS201L) illuminates the object. The reflected light is imaged by a 1.0X telecentric lens (Edmund, 58430) onto a microlens array (RPC photonics, MLA-S100-f8). Here, we use an unfocused light field imaging scheme where the microlens array is placed at the image plane of the imaging system, while the detector is placed at the back focal plane of the microlens array [2]. The telecentric lens has a field of view of 8.8mm×6.6 mm, with an adjustable f number ranging from 6 to 22. The working distance of the telecentric lens is 83 mm. The microlens array has a 100 μm lens pitch and a 0.8 mm focal length. An intermediate pupil array image is formed at the back focal plane of the microlens array, co-locating with the conjugated image plane of the IMS. The image of a resolution cell of

the hypercube at this intermediate image plane is 15.7 µm. Next, the pupil array image is relayed to a mapping mirror in the IMS. The mapping mirror splits an image into 480 slices within 5×8 subfields. Each subfield contains 12 image slices and has a blank area between adjacent image slices. After passing through the dispersive prism, the spectrum of an image slice occupies the blank area between two adjacent image slices. The dispersion prisms are designed for visible light (400 nm–700 nm). A final image is formed on the IMS detector. Two representative raw images under monochromatic and white-light illumination with a zoomed subfield are shown in Figs. 31(c) and 31(d), respectively. Under monochromatic illumination, the width of each image slice is one pixel, and the spacing between two adjacent image slices within a subfield is 40 pixels. The IMS was assembled by Rebellion Photonics Inc., and the optical parameters of the IMS are detailed in Refs. [58,59].
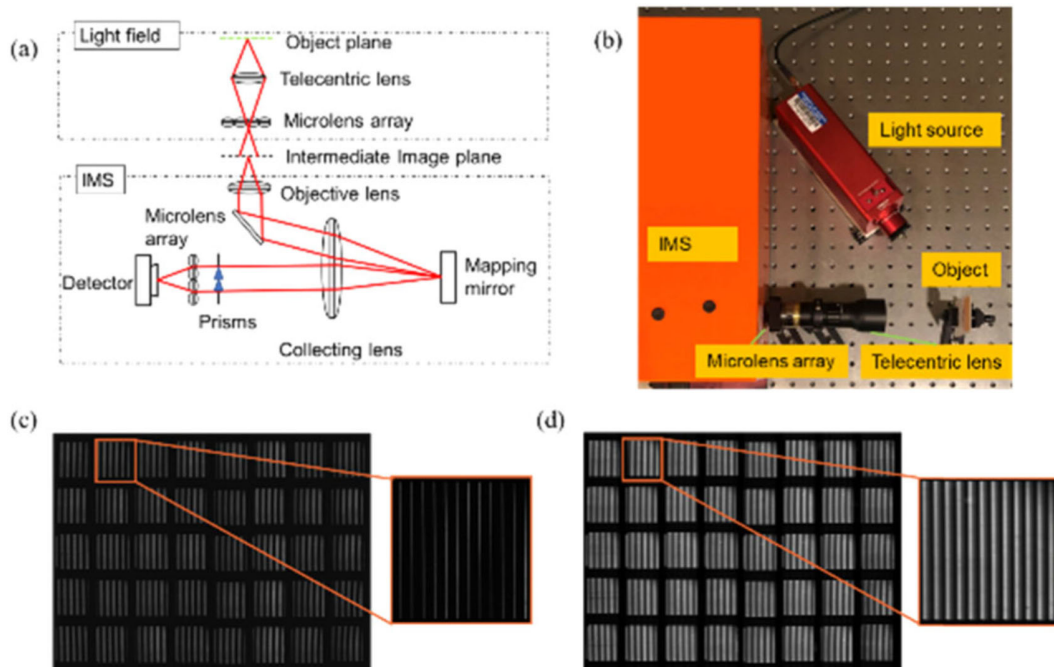
Figure 31. Optical setup. (a) Schematic. (b) Photograph of the system. (c) Raw image captured by the IMS under monochromatic illumination. (d) Raw image captured by the IMS under white-light illumination

53

Figure 2 shows the image processing pipeline. A raw image captured by the system was first mapped to a hypercube, which contains light field data at different wavelengths. At each wavelength, we further processed the light field data to generate a 4D light field datacube (x ; y ; u; v) and the correspondent epipolar plane images [77]. Finally, we calculated the depth map through the scale-depth space transform [26].
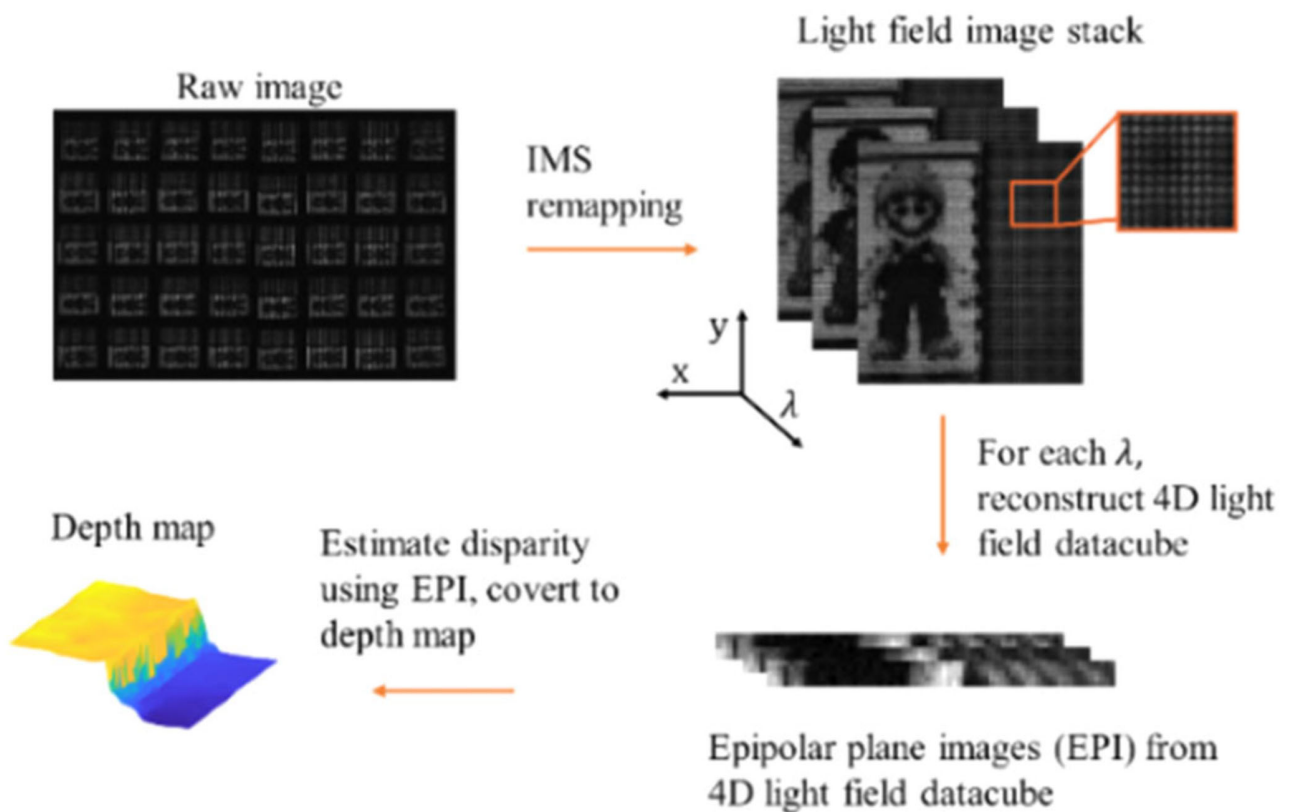
Figure 32. Image processing pipeline.

To assess the lateral resolution of the entire system, we imaged a USAF 1951 resolution target (Thorlabs, R1DS1P) at the nominal working distance of the telecentric lens. The reconstructed result is shown in Fig. 33(a). Figure 33(b) shows the inverted intensity plot along the red line in Fig. 33(a). The bars of group 2, element 3 were clearly resolved, indicating that the resolution of our system is 5 lp/mm. Because the IMS has a resolution cell (15.7 μm) smaller than the microlens pitch (100 μm), the lateral resolution is limited by the light field imaging module

rather than the IMS. The theoretical lateral resolution of an unfocused light field imaging module is determined by the microlens pitch [36],

$$r = \frac{1}{2d} \tag{13}$$

where r is the lateral resolution [line pairs per millimeter (lp/mm)], and d is the microlens pitch. In our case, the microlens pitch is 100 μm; therefore, the theoretical lateral resolution is 5 lp/mm. Our experimental result matches with theoretical expectation.
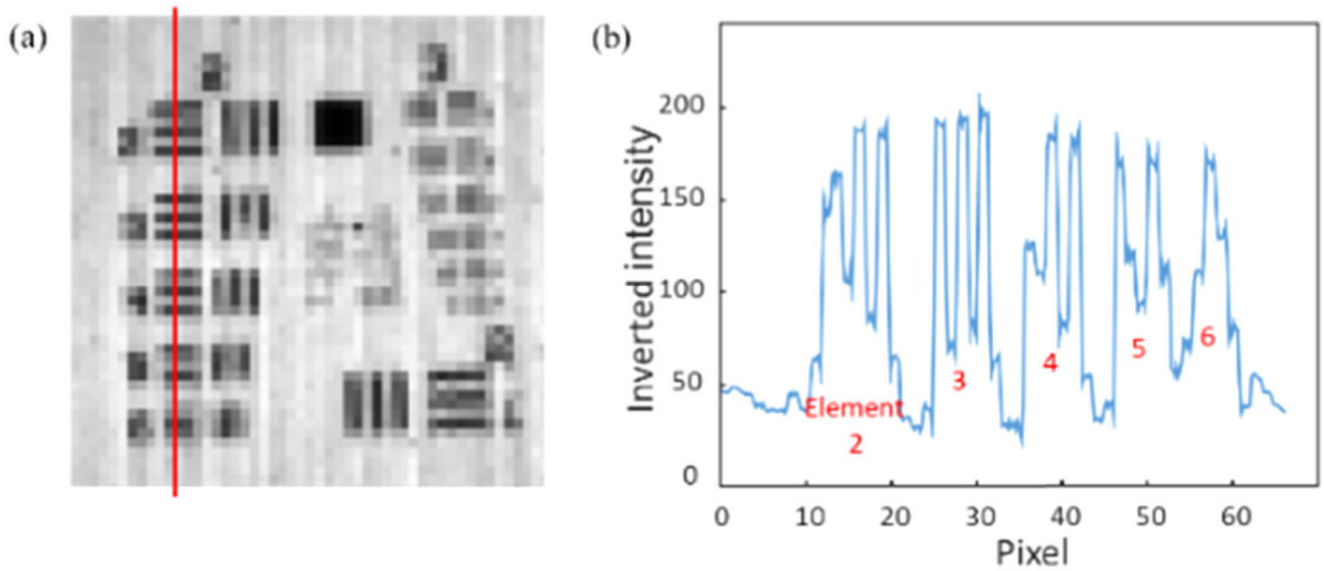
Figure 33. Lateral resolution. (a) Image of the USAF 1951 resolution target. (b) Inverted intensity plot of the red line in (a) through the elements of group 2.

To perform three-dimensional (3D) imaging, we must evaluate depth accuracy within the depth range. The depth range of an unfocused light field imaging camera is given by [36],

$$R = \left[-\frac{df}{s_e}, \frac{df}{s_e}\right] \tag{14}$$

where R is the depth range, d is the microlens pitch, f is the focal length of the microlens, and Se is the pixel size on the intermediate image plane. In our experiments, the depth range is from -5mm to 5 mm with respect to the nominal working distance of the telecentric lens, and the depth

55

axis points towards the system. We used a flat printed grid pattern as an object, mounted it on a motorized translation stage (Thorlabs, MTS50-Z8), and scanned the object within the depth range along the axial direction. The scanning step is 0.25 mm. At each scanning step, we captured an image and reconstructed the corresponding disparity map [26]. The result is shown in Fig. 34. The red dots denote the experimentally measured disparities, and the blue line is a third-order polynomial curve fitted to the experimental results. We then calculated the root-mean-square (RMS) error at each scanning step by comparing the reconstructed depth map with the ground truth. The average RMS error within the depth range is 0.36mm.
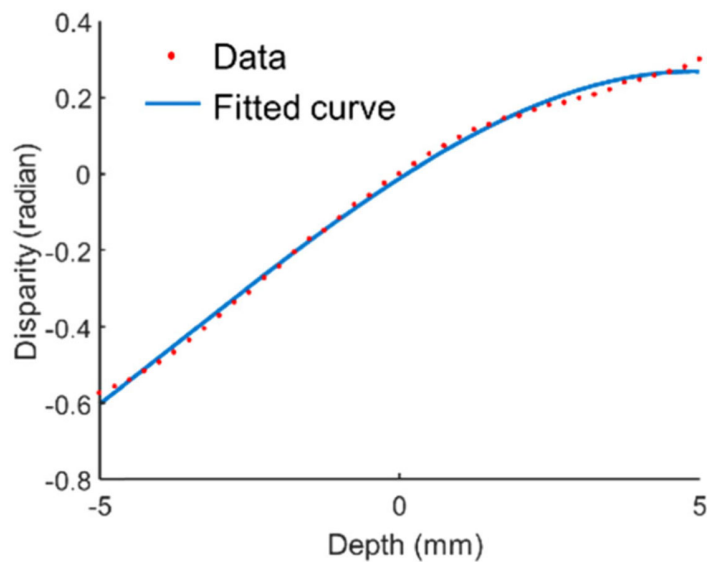
Figure 34. Disparity versus depth curve.

Next, we characterized the spectral resolution of the system. As previously mentioned, within a subfield, an image slice passes through a dispersive prism, and the spectrum of the image slice fills the blank area between two adjacent image slices. The width of the blank area is 40 pixels on the IMS detector, and these 40 pixels are used to sample the wavelengths from 400 nm to 700 nm. To evaluate the relationship between pixel positions within a blank area and the

56

corresponding wavelengths, a chromatic dispersion curve was measured, as shown in Fig. 35(a). In this step, we operated the system in the transmission mode, i.e., putting the light source on the optical axis and facing it towards the telecentric lens. Then, we put a 10 μm diameter pinhole (Thorlabs, P10D) at the nominal working distance of the telecentric lens and scanned the pinhole at nine random lateral locations. At each location, six color filters (Thorlabs, FB470-10, FB510-10, FB550-10, FB590-10, FB630-10, FB670-10) were used sequentially, and the corresponding pixel locations were recorded. The final curve is an average result of the data taken at nine pinhole locations and was fitted with a second-order polynomial. The spectral resolution is defined as the bandwidth of the spectrum occupied by a pixel, and it is a function of the wavelength. The result is shown in Fig. 35(b). The spectral resolution at the blue light end is higher than that at the red light end because of the nonlinear spectral dispersion of the prism.
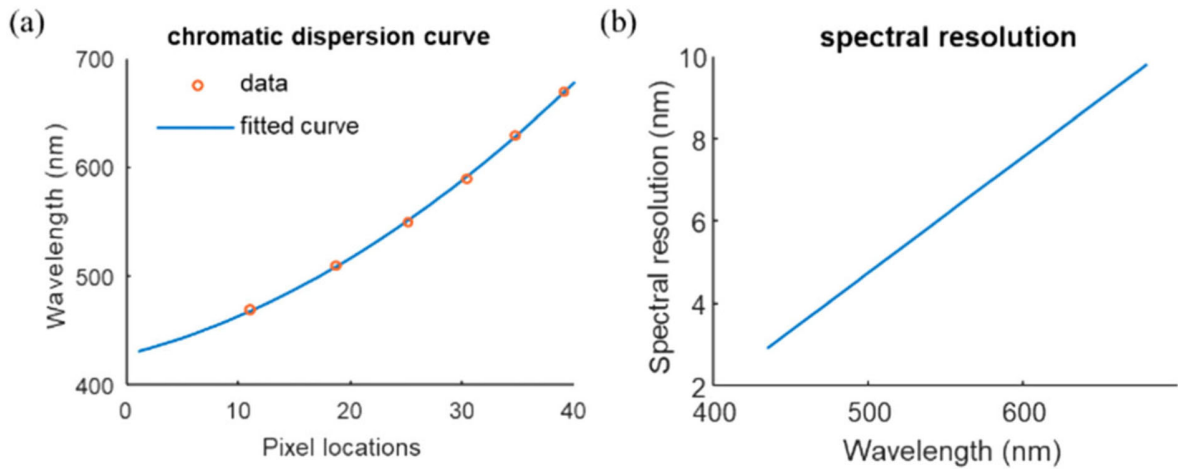
Figure 35. Spectral dispersion of the system. (a) Chromatic dispersion curve of the system. (b) Spectral resolution.

Finally, we demonstrated the imaging performance of the system by imaging dynamic 3D spectral scenes. The object contains two parts: a picture of a toy on the left and a grid pattern on the right. The toy picture is mounted on the motorized translation stage and continuously moving axially back and forward within the depth range. The grid pattern is located at a depth near -5

mm, and it is static. By using the previously measured chromatic dispersion curve and the disparity-depth mapping curve, we generated 40 color channel images and a depth map. Figure 36(a) shows the reconstructed images at three representative wavelengths (480 nm, 540 nm, 600 nm) when the toy picture is at the nominal working distance of the telecentric lens. The color of the image was rendered according to the corresponding wavelength based on the International Commission on Illumination (CIE) 1931 observer. Figure 36(b) shows the reconstructed depth maps of the toy image at -3 mm, 0 mm, and 3 mm at the 600 nm wavelength and depth profile images along the red line in Fig. 6(a). This experiment demonstrates the real-time imaging capability of the system in performing 4D (x, y, z, λ) measurement.
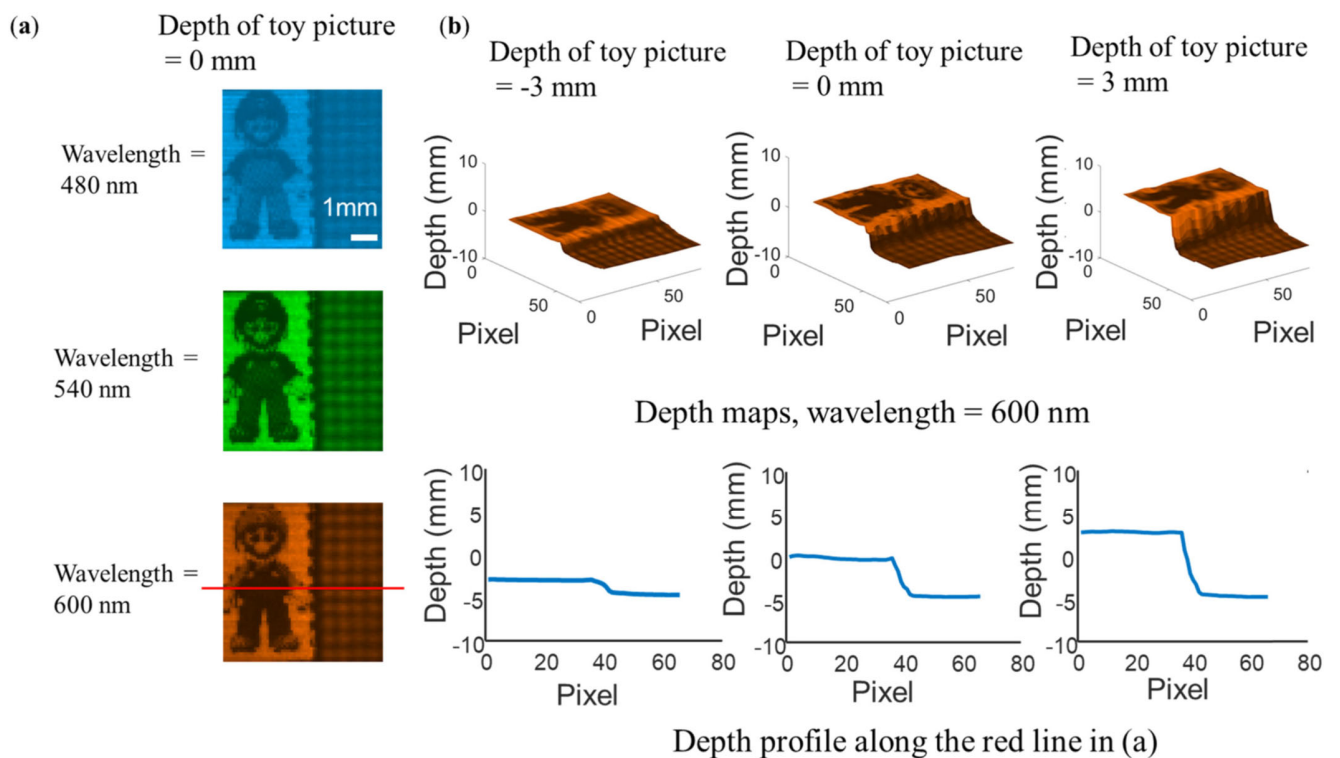
Figure 36. Imaging dynamic 3D spectral scenes. (a)Reconstructed images at 480 nm, 540 nm, and 600nmwavelength. The depth of the toy picture is 0 mm. (b) Reconstructed images at depths of the toy picture = -3 mm, 0 mm, and 3mmat 600 nm wavelength. The first row shows the depth maps, and the second row shows depth profile images of the red line in (a).

58

4.2 Snapshot hyperspectral light field tomography (Hyper-LIFT)

We present snapshot hyperspectral light field tomography (Hyper-LIFT), a highly efficient method in recording a 5D (x, y, spatial coordinates; u, v, angular coordinates; λ, wavelength) plenoptic function. Using a Dove prism array and a cylindrical lens array, we simultaneously acquire multi-angled 1D en face projections of the object like those in standard sparse-view computed tomography. We further disperse those projections and measure the spectra in parallel using a 2D image sensor. Within a single snapshot, the resultant system can capture a 5D data cube with 270×270×4×4×360 voxels. We demonstrated the performance of Hyper-LIFT in imaging spectral volumetric scenes.

An optical field can be characterized by a seven-dimensional (7D) plenoptic function, $P(x, y, z, u, v, \lambda, t)$ (x, y, z, spatial coordinates; u, v, emittance angles; λ, wavelength; t, time) [78,79]. Measuring a high-dimensional plenoptic function faces two main challenges: dimension reduction and measurement efficiency. On the one hand, because most photon detectors are in two dimensions (2D) (i.e., image sensors), one dimension (1D) (i.e., line sensors), or zero dimensions (0D) (i.e., single pixel sensors), capturing a high-dimensional plenoptic function with a low-dimensional detector usually requires extensive scanning along another dimension. To characterize this ability, we define a dimension reduction factor as $\epsilon = N_p/N_D$, where N_p and N_D are the dimensionalities of the plenoptic function to be measured and the detector deployed, respectively. Because low-dimensional detectors usually image faster and cost less than their high-dimensional counterparts, the greater the ϵ, the higher the frame rate, and the more economical the system. On the other hand, under the conventional Nyquist sampling condition, measuring a high-dimensional plenoptic function usually requires the detector array to have a large number of elements, posing challenges with the data transfer and storage. An effective

approach to break this limitation is compressed sensing [65–67], which allows using much fewer measurements to recover a scene, provided that the object can be considered sparse in a specific domain. To quantify the sampling efficiency, we define a compression ratio as $r = S_N/S_C$, where S_N and S_C are the sampling number determined by the Nyquist–Shannon theorem and compressed sensing, respectively. The higher the r, the more efficient the measurement.

Although techniques have advanced significantly in reducing the dimensionality of a plenoptic function, it is nontrivial to build an imager with a large compression ratio while maintaining high image quality. For example, in hyperspectral light field imaging, to acquire a 5D data cube (x, y, u, v, λ), most current imagers [20-24] are built on Nyquist sampling and provide only a unity compression ratio $r = 1$. For a given detector array, this leads to a trade-off between the samplings along the spatial, spectral, and angular axes. For example, in an IMS-based hyperspectral light field camera [80]. the total number of plenoptic data cube voxels is limited to 66×66×5×5×40 (*x, y, u, v, λ*), restricting its application in high-resolution imaging.

To overcome this problem, we present a new multi-dimensional imaging architecture and demonstrate it in 5D (x, y, u, v, λ) plenoptic imaging. The resultant method, referred to as snapshot hyperspectral light field tomography (Hyper-LIFT), can capture a 270×270×4×4×360 (x, y, u, v, λ) data cube using a 2D detector array in a single snapshot. Based on the same conceptual thread as light field tomography (LIFT) [69], Hyper-LIFT is highly efficient in acquiring the light field data through simultaneously recording en face parallel beam projections of the input scene along sparsely spaced angles, enabling a compression ratio of 16.8. Moreover, Hyper-LIFT captures additional spectral information by further dispersing the en face beam projections in the spectral domain. By converting the angular information to depths, we demonstrated Hyper-LIFT in hyperspectral volumetric imaging.

4.3 Image formation and optical setup

The core idea of Hyper-LIFT is to reformulate the light field acquisition as a sparse-view computed tomography (CT) problem. To create multi-angled en face parallel beam projections of an object, we put an array of rotated Dove prisms and cylindrical lenses at the pupil plane of a main objective lens and image the object from different perspectives. For a given perspective, the image formation model through a rotated Dove prism and a cylindrical lens (i.e., a subfield) is shown in Fig. 36. The Dove prism rotates the input perspective image by an angle of 2θ, where θ is the rotation angle of the Dove prism itself. The rotated perspective image is then imaged by a cylindrical lens. The resultant image is essentially a convolution of the geometric image of the rotated object and a line spread function provided by the cylindrical lens. We put a slit at the back focal plane of the cylindrical lens and sampled the image along the horizontal axis. The 1D signals obtained is a "projection" of the object at the angle of 2θ, resembling the projection measurement in traditional x-ray CT. The image formation can be formulated as:

$$b^{\theta} = \boldsymbol{TR^{\theta}}g \tag{15}$$

where g is the vectorized two-dimensional (2D) perspective image, $\boldsymbol{R^{\theta}}$ is the rotation operator representing the function of the Dove prism at the angle θ, T denotes the signal integration by the cylindrical lens at the 1D slit, and b^{θ} is the 1D signal sampled by the slit. After passing through a diffraction grating, the 1D projection is dispersed along the vertical axis. The final image is measured by a 2D detector array, and the spectral channels of the projection are obtained simultaneously.
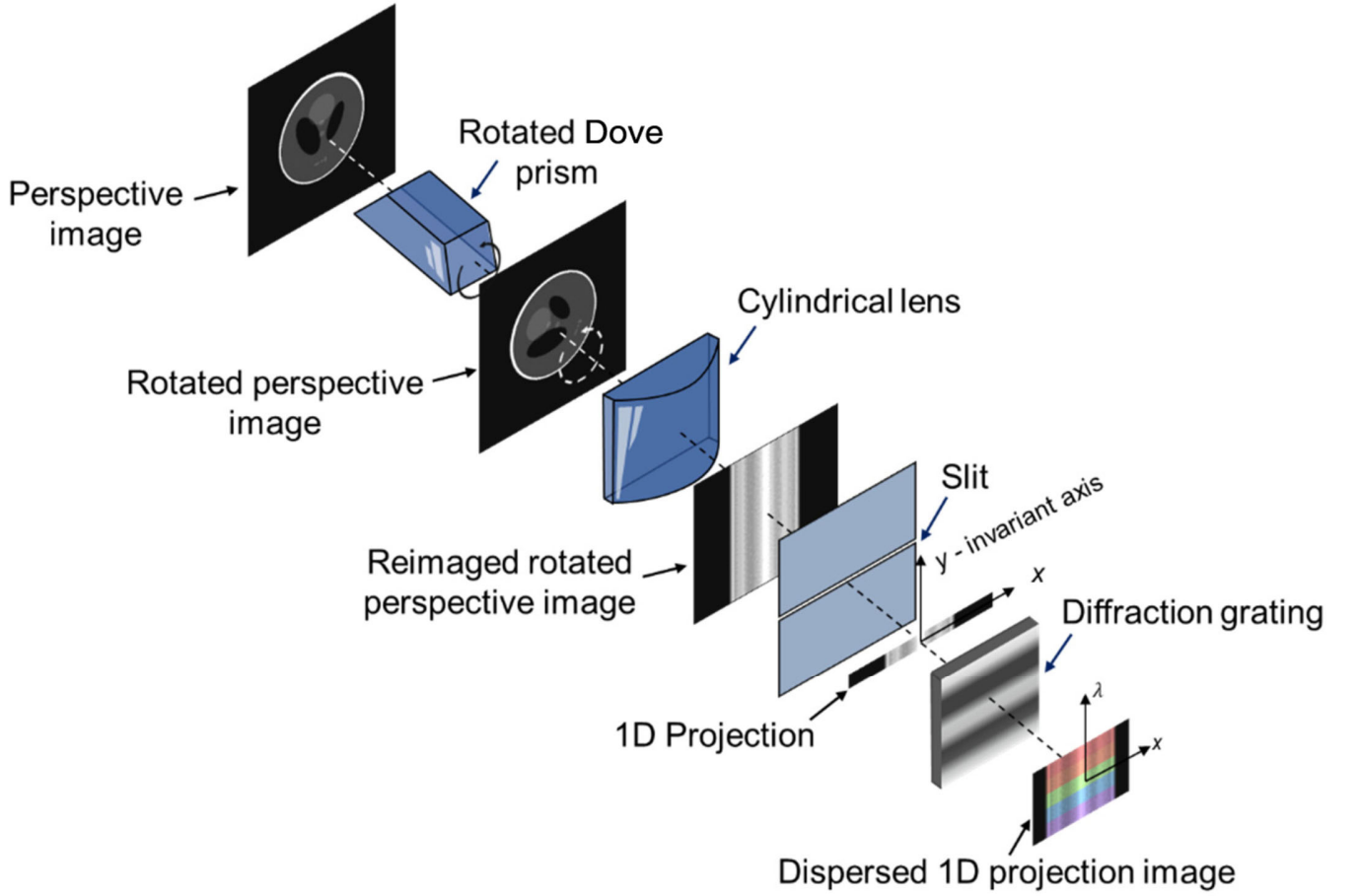
Figure 37. Image formation model.

In Hyper-LIFT, we use an array of Dove prisms with varied rotation angles to acquire the dispersed en face projections in parallel. The forward model with n Dove prisms can be formulated as

$$b(\lambda) = T \begin{bmatrix} R^{\theta 1} \\ ... \\ R^{\theta n} \end{bmatrix} g(\lambda) = Ag(\lambda) \tag{16}$$

where $\mathbf{A}$ is the forward operator representing the parallel beam projections at different angles, and $\mathbf{b}(\lambda)$ is the sinogram at wavelength λ. Because each Dove prism observes the same scene from different perspectives, the light field is naturally sampled in the projection data with an

62

angular resolution equal to the number of Dove prisms.
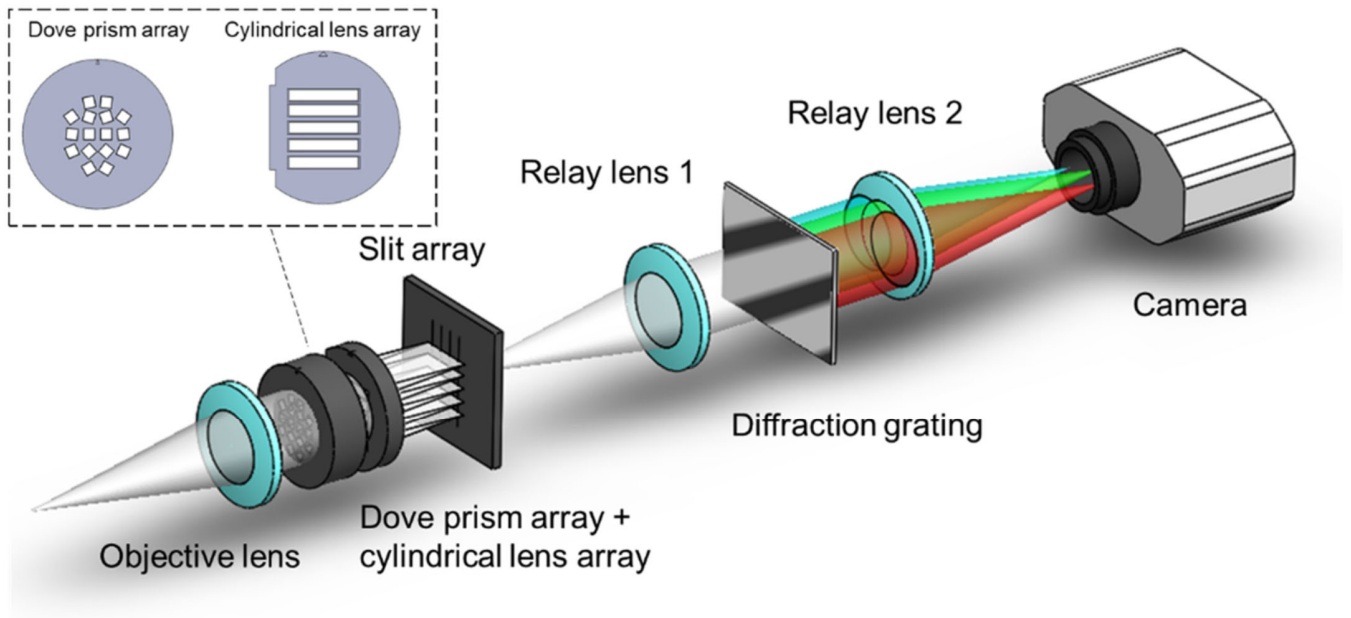
Figure 38. Optical schematic of a Hyper-LIFT camera.

The schematic of the optical setup is shown in Fig. 38. The object is located at the front focal plane of an objective lens (f = 50 mm, f/# = 1.4). The field of view of the system is ~3 mm. A Dove prism array that comprises 16 multi-angled Dove prisms (height=2 mm, length=8.4 mm, and spacing between adjacent prisms=2.9 mm) is located at the back focal plane of the objective lens. After the Dove prism array, a 5×1 cylindrical lens array (height=2 mm, length=12 mm, focal length=20 mm) focuses the collimated beam onto a 4×1 slit array (width = 10 μm), which slices the input image. The resultant 1D projections are reimaged to a CCD sensor (Lumenera, Lt16059H; pixel pitch = 7.4 μm) using a 4f relay system (f = 100 mm, f/# = 2). A transmission diffraction grating (300 groves/mm) is located at the Fourier plane of the relay system to disperse the projections.

The pictures of the mounts that hold the dove prisms and cylindrical lenses are shown in Fig.39,

where the numbers indicate the rotation angles in degrees of dove prisms. Instead of arranging the dove prisms in a rectangular array, we chose a near circular pattern to fulfill the aperture of the main lens. Each dove prism rotates the perspective image at the corresponding pupil location, which is further imaged by a cylindrical lens.
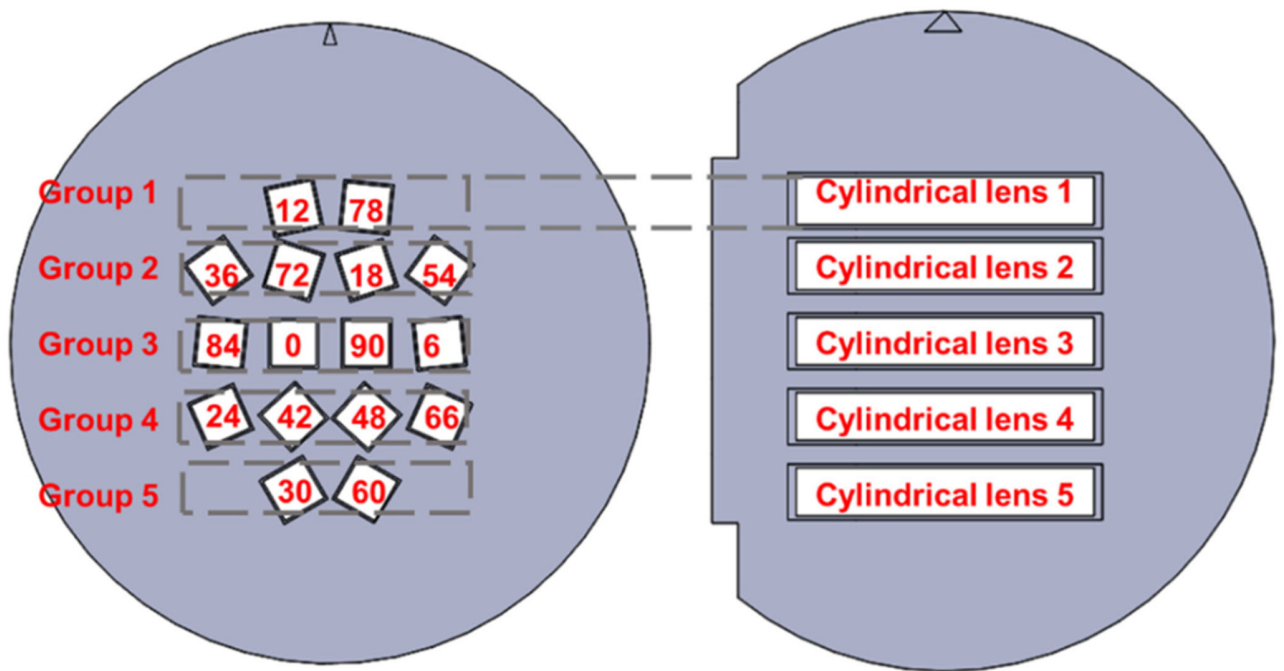
Figure 39. Mounting plates that hold the dove prisms and cylindrical lenses. The numbers in the left panel indicate the rotation angles of dove prims in degrees. The dove prisms are divided into five group, each passing the light to the same cylindrical lens.

In Hyper-LIFT, the spatial sampling of the projections can be calculated as the height of a Dove prism divided by the camera pixel pitch. The spectral sampling of the reconstructed data cube is limited by the spacing between adjacent Dove prisms. The angular sampling is determined by the number of Dove prisms and their locations at the pupil plane. In our setup, the reconstructed data cube has a dimension of $270\times270\times4\times4\times360$ (x, y, u, v, λ). Noteworthily, in Hyper-LIFT, only

the voxels along the spectral only the voxels along the spectral dimension are directly mapped onto the 2D sensor, while the spatial and angular voxels are multiplexed in projection measurement. Additionally, instead of acquiring the entire set of projections under the Nyquist sampling condition, Hyper-LIFT measures only a sparse subset of the projections based on compressed sensing. By contrast, previous multi-dimensional imagers [80-82] directly map plenoptic data cube voxels to a 2D image sensor. As a result, the size of the reconstructed data cube is limited by the available sensor pixels, leading to a trade-off between spatial, angular, and spectral sampling. Using compressed sensing, Hyper-LIFT alleviates this trade-off, thereby enabling the measurement of a large format plenoptic function.

To calibrate the spectral response, we placed a pinhole (Thorlabs, P50D) at the nominal object plane and illuminated it with monochromatic light of different wavelengths. The corresponding pixel locations of the projections were recorded. Because the dispersion curve of a diffraction grating is linear, five wavelengths provide a fitting with negligible localization errors (Fig. 40(b), top). Because 1 nm bandwidth in wavelength occupied four pixels on the camera, the system provides a spectral resolution of 0.25 nm.

To calibrate the depth, we put a pinhole (Thorlabs, P50D) at the front focal plane of the objective lens, scanned the pinhole from -3.75 mm to 3.75 mm along the depth axis with a 0.625 mm step, and captured an image at each depth. Next, we digitally refocused each pinhole image by tuning the shearing parameter. For each pinhole image, we identified the shearing parameter that best brings the image back in focus and recorded corresponding physical depth. The best focus image can be found by maximizing a focus measure (e.g., sum of modified Laplacian) for each pixel of the image. We fitted the curve with a linear model. The resultant shearing to depth curve is shown in the bottom figure in Fig. 40(b). Using this curve, we can digitally refocus a 3D objective, and the corresponding depth can be deduced based on the shearing parameter.
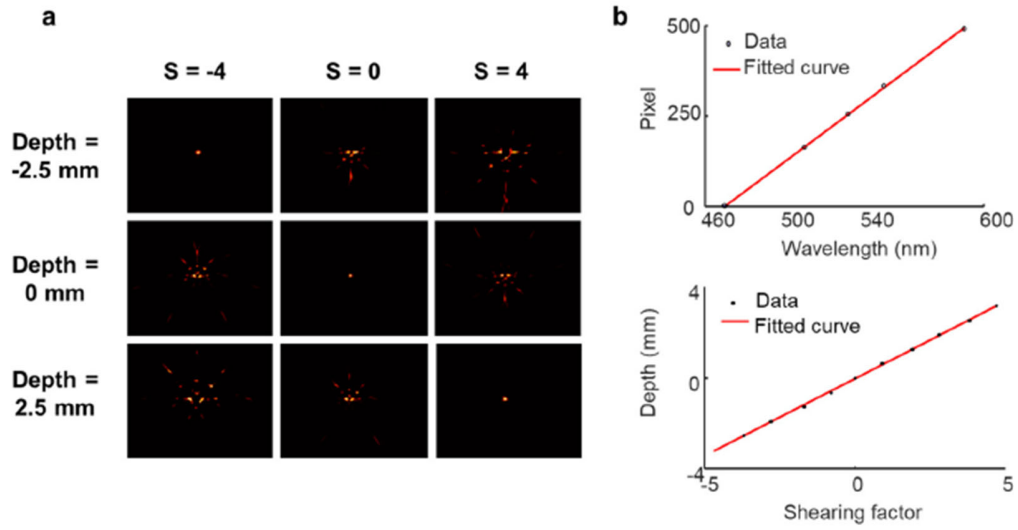
Figure 40. Spectral and depth calibration. (a) Three example pinhole images refocused with different shearing factors. The pinhole images were captured at different depths. (b) Measured chromatic dispersion curve and shearing to depth relation.

Like conventional sparse-view CT, Hyper-LIFT requires the object to be sparse in a specific domain. We quantified the resolution under such a condition by imaging a point object (a 10 μm pinhole). The lateral and axial resolution resolutions are measured as the full width half-maximum (FWHM) of the impulse response along the lateral and axial direction, respectively. The measured lateral and axial resolutions are 22 μm and 1mm, respectively.
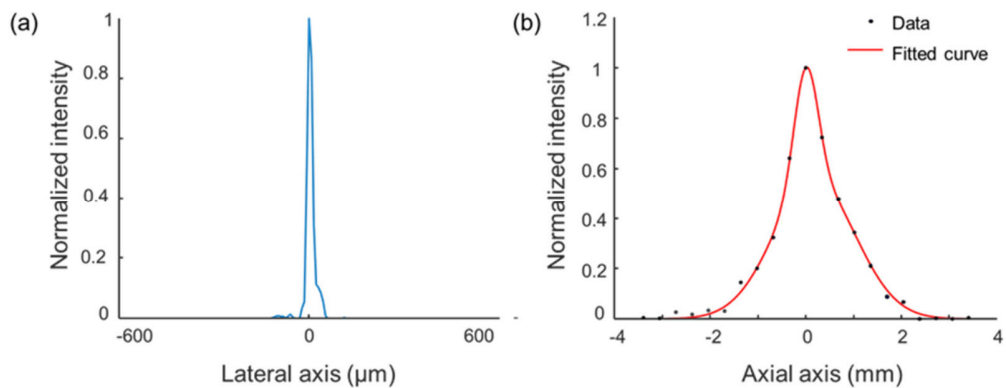
Figure 41. Characterization of spatial and axial resolutions. (a) Lateral intensity distribution of a reconstructed 10 μm pinhole. (b) Axial intensity distribution of a reconstructed 10 μm pinhole, fitted with a Gaussian model.

4.4 Image reconstruction

In Hyper-LIFT, because the number of projections n is smaller than the pixel resolution of the input perspective image, Eq. (16) is under-determined and hence can be considered as a sparse-view CT problem [70]. The image reconstruction of a monochromatic scene can be achieved by iteratively solving the optimization problem:

$$\underset{\hat{g}}{\mathrm{argmin}} \|\boldsymbol{b} - \boldsymbol{A}g\|_2^2 + \rho\varphi(g)_1, \tag{17}$$

where $\varphi(g)$ is a transform function that sparsifies the image, $\cdot 1$ is the L1 norm, and ρ is the hyperparameter that weights the regularization term. More details of solving this inverse problem and the sparsity requirements can be found in [69]. A reconstruction example of a planar monochromatic scene is shown in Fig. 42. To limit the illumination bandwidth, a 3 nm bandpass filter centered at 532 nm was placed right after a broadband light source (Amscope, HL250-AS). The ground truth and a raw image captured by the Hyper-LIFT camera are shown in Figs. 42(a) and 42(b), respectively. The sliced projections are along the vertical axis, and the spectra of the projections are dispersed along the horizontal axis. To perform image reconstruction, we first divide the raw image into 16 subfields, each containing a projection. Next, the projections are extracted based on their center coordinates, which can be calibrated by imaging an optical pinhole. Finally, the object is reconstructed using Eq. (17), and the reconstruction result is illustrated in Fig. 42(c).

To reconstruct images at different wavelengths, we apply Eq. (3) to the corresponding sinogram $\boldsymbol{b}(\lambda)$. Because the spectrum is dispersed along one axis, the projection at wavelength λ can be directly obtained by extracting the line image at the corresponding position, which can be deduced from the chromatic dispersion curve of the diffraction grating. The spectral range and resolution can be deduced from the dispersion curve, and they are 90 nm and 0.25 nm,

respectively, in the current setup. It is worth noting that the spectral range and resolution are scalable by using a different dispersion element or a different focal length relay system.
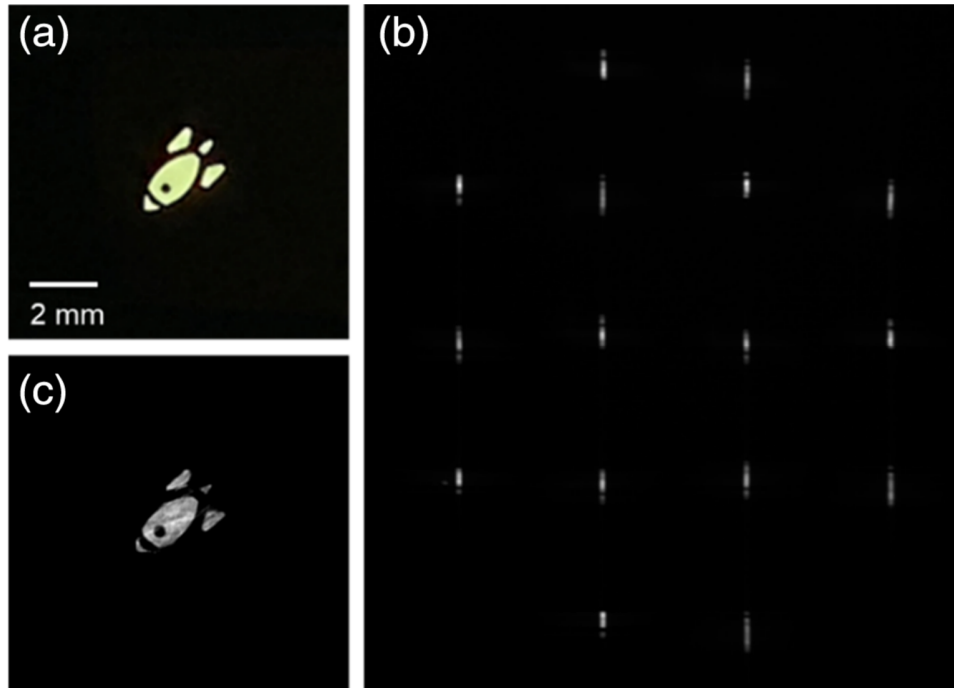
Figure 42. Reconstruction of a planar object illuminated by monochromatic light (wavelength, 532 nm). (a)Ground-truth photographic image. (b) Raw Hyper-LIFT image. (c) Reconstructed image.

The whole reconstruction process can be divided into two steps: wavelength selection and CT reconstruction. Because the spectrum is directly dispersed on the detector using a diffraction grating, projections at different wavelength can be extracted using the dispersion curve, where the computational cost is negligible. In the CT reconstruction process, the image reconstruction speed is mainly dependent on the reconstruction algorithm applied. For example, when simply applying the inverse Radon transform through fast Fourier transform, the object can be reconstructed instantaneously. However, when using an iterative algorithm to improve image quality, the reconstruction takes ~5s for each wavelength on a personal computer (CPU, Intel Core i7-5700HQ).

68

4.5 Digital Refocusing

In a conventional light field camera, the acquired light field L can be parameterized by the aperture plane (u, v) and the image plane (x, y) as shown in Fig. 43(a). For a synthetic image plane (x', y') defined by a shifting parameter, the irradiance can be calculated as [2]:

$$E(x', y') = \iint L(u, v, \frac{x\prime+(\alpha-1)u}{\alpha}, \frac{y\prime+(\alpha-1)v}{\alpha}) du dv, \qquad (18)$$

Digital refocusing is achieved by shifting and adding up subaperture images. Unlike conventional light field cameras, Hyper-LIFT employs a Dove prism array to rotate the input scene, and their effect must be accounted for when calculating the shifting vector. Here, we denote the original shifting vector as s·(u, v), where (u, v) is image translation direction vector and s is the translation distance (i.e. shearing parameter). As illustrated in Fig. 43(b), the vector (u, v), is determined by the pupil coordinate and the rotation angle of the Dove prism. If the Dove prism is rotated by $\theta/2$, the resultant shifting vector can be derived as follows:

$$s \cdot \begin{pmatrix} u' \\ v' \end{pmatrix} = s \cdot \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{pmatrix} u \\ v \end{pmatrix} = s \cdot \begin{pmatrix} u \cdot \cos(\theta) + v \cdot \sin(\theta) \\ u \cdot \sin(\theta) - v \cdot \cos(\theta) \end{pmatrix}. \qquad (19)$$

Because Dove prisms rotate the input scene based on total internal reflection (TIR), an extra flipping matrix is added to reflect the redirection of light rays. The updated shifting vector is then applied to translating subaperture projections to obtain an updated sinogram. The final refocused image can be reconstructed using a classic computational tomography algorithm, such as analytic filtered backprojection.
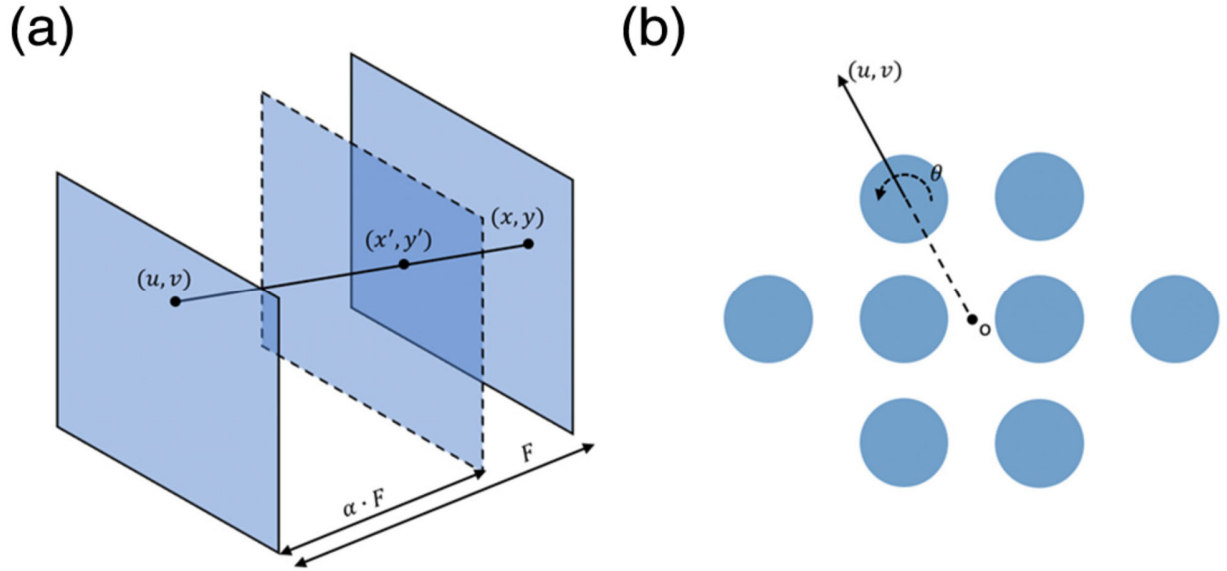
Figure 43. Digital refocusing in Hyper-LIFT. (a) Refocusing a synthetic image plane in a conventional light field camera. (b) In Hyper-LIFT, the shifting vector is determined by the pupil coordinate and the rotation angle of the dove prism. Each blue circle denotes a dove prism.

To demonstrate digital refocusing ability of Hyper-LIFT, we placed an object at depth = -3.75 mm, 0 mm, 3.75 mm with respect to the nominal focal plane of the Hyper-LIFT camera. At each object position, we captured a snapshot and generated the focal stack images. Fig. 44 (a), (b) and (c) show the sweeping of the focal plane images at the corresponding object setting. As expected, the object appears in focus only at its designated location. Noteworthily, the out-of-focus image in LIFT cameras appear as ghost images rather than uniformly radial blur as in conventional widefield imaging. This effect has been discussed in our previous publication [69].
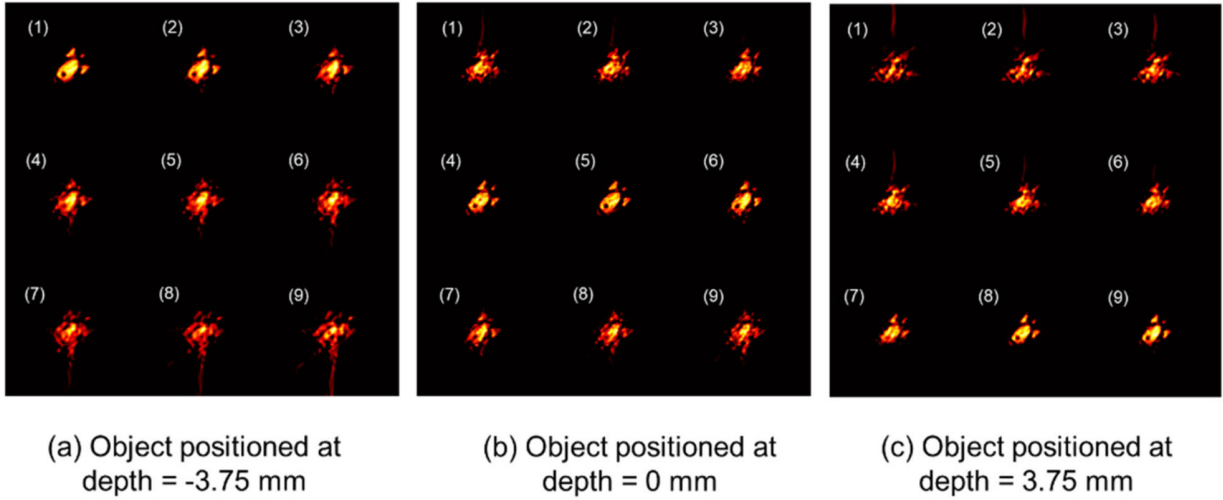
Figure 44. Sweeping of focal stack images for an object positioned at three depths. The numbers in the subpanels denote the indices of the shearing parameter. From (1) to (9), refocusing depth is from – 4 mm to 4 mm, with step size = 1 mm.

4.6 Results

We first demonstrated our system in imaging a rainbow planar object [Fig. 42(a)]. The illumination setup is shown in Fig. 45(a). A linear variable visible range bandpass filter (Edmund optics, 88365) is located at the conjugate plane of the planar object, illuminated by a broadband light source. To fit more wavelengths into the field of view, a 3.3:1 lens pair (Thorlabs, MAP1030100-A) is used to demagnify the linear filter. In this way, each lateral coordinate of the planar object is encoded with a unique color. At a given location, the spectral resolution of the linear variable filter is 7–20 nm. The raw image captured by the Hyper-LIFT camera is shown in Fig. 45(b). Compared to the monochromatic data in Fig. 42(b), the spectrum is dispersed along the horizontal axis. The projections at different wavelengths are extracted based on the chromatic dispersion curve. The pseudo-colored reconstructed panchromatic image and four representative color channels are shown in Fig. 45(c). This experiment demonstrates the hyperspectral imaging capability of our system.

71
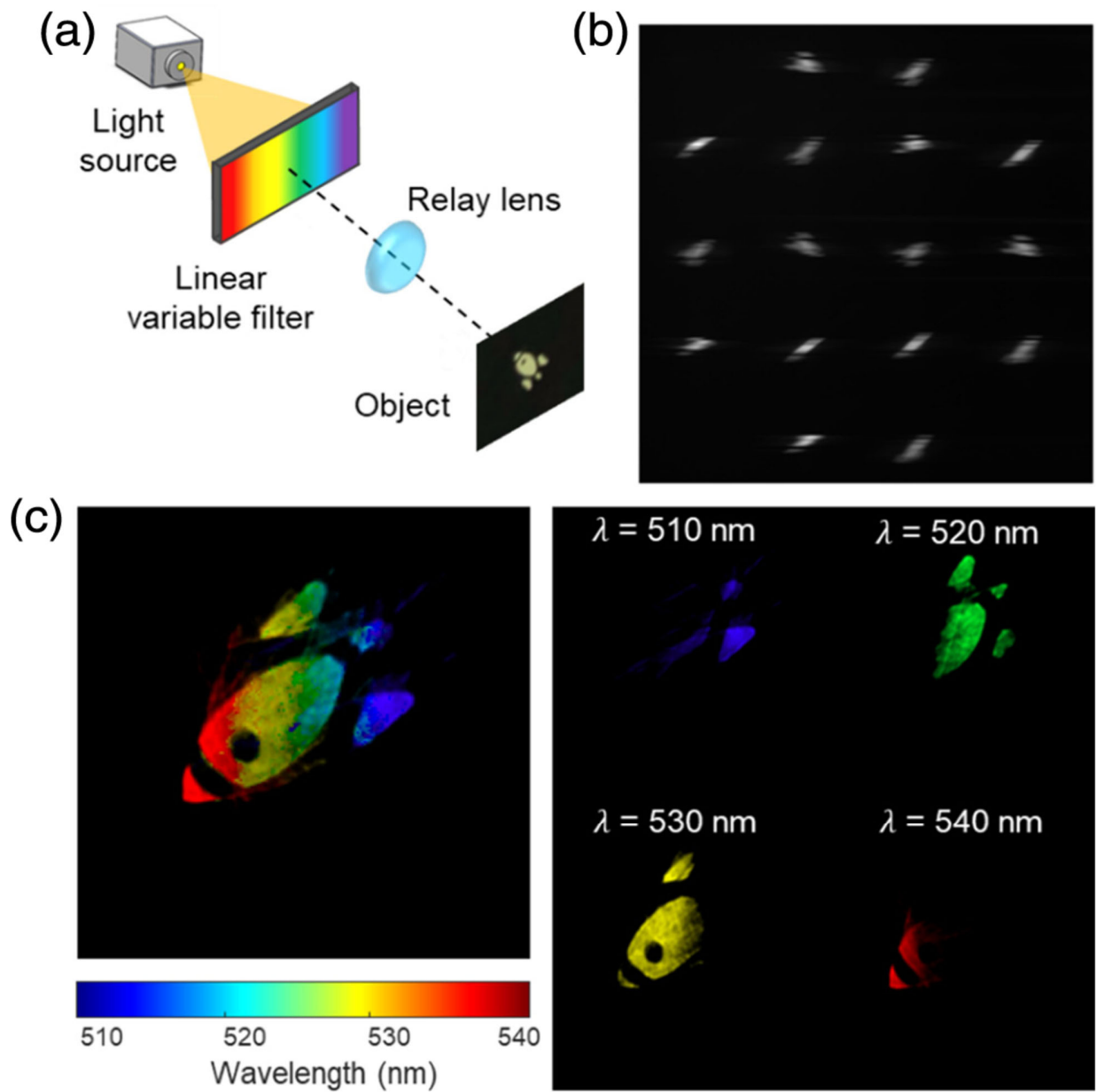
Figure 45. Hyperspectral imaging of a planar object. (a) Illumination setup. (b) Raw Hyper-LIFT image. (c) Pseudo-colored reconstructed panchromatic image and four representative color channels.

To demonstrate hyperspectral volumetric imaging enabled by light field capture, we imaged a 3D-printed object illuminated by a broadband light source. To avoid cross talk between adjacent subaperture images, a 40 nm bandpass filter at 550 nm is used to limit the illumination bandwidth. An en face photograph of the object is shown in Fig. 46(a). The letters "U" and "C"

were 3D printed at different depths (depth difference = 6 mm). The object is imaged to the fron

focal plane of the objective lens by a demagnifying relay lens pair (Thorlabs, MAP1030100-A).

To render a focal image stack, digital refocusing is performed by calculating the shifting

direction vector (u, v) and then changing the shearing parameter s in Eq. (19). In practice, the

shearing parameter s can be converted to a real depth. The reconstructed 3D image and two

representative numerically refocused images at the two letters' depths are shown in Fig. 46(b).

The pixel intensities within the boxed area in Fig. 46(b) are averaged, and a spectrum is shown in

Fig. 46(c). Finally, we quantified the level of defocus by calculating the sharpness of the object

at each refocused depth [Fig. 46(d)]. Here sharpness is defined by $(I\_max - I\_min)/(I\_max +$

$I\_min)$, where I (x, y) is the pixel value at the letter location. The sharpness of letters "U" and

"C" reaches a maximum at 0.8 mm and 0 mm, respectively, which gives the physical depth of

two letters in the object space of Hyper-LIFT.
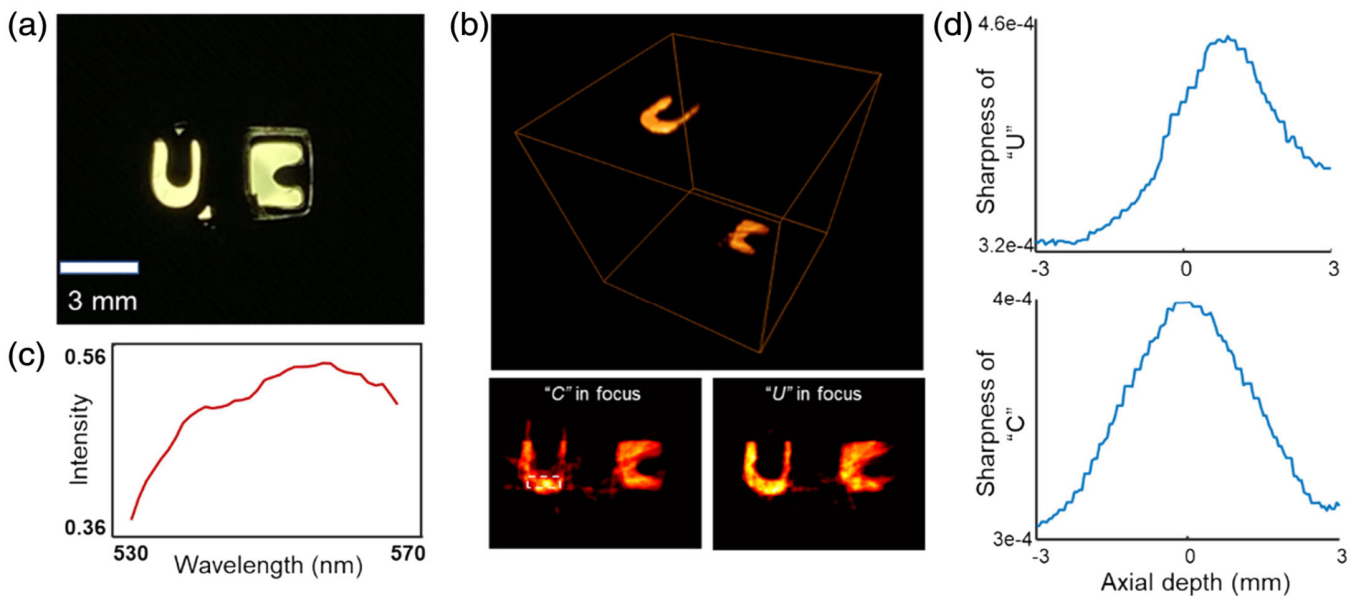
Figure 46. Hyperspectral volumetric imaging of a 3D object. (a) Ground truth en face image of the object. "U" and "C" were printed at two different depths. (b) Reconstructed 3D image and two numerically refocused images at the letters' depths. (c) Spectrum averaged within the dashed boxed area in the low left panel of (c). (d) Sharpness versus depth at two letter locations.

4.7 Discussion

The application of hyperspectral light field imaging continues to expand, with a variety of technologies sprouting in the past five years. Despite impressive functionality, these techniques have not been discussed in a common framework. Multi-dimensional imagers are typically evaluated in terms of the snapshot factor [83], which describes the portion of the plenoptic data cube voxels that can be seen by the imager at a time. The greater the snapshot factor, the higher the light throughput. However, the snapshot factor does not reflect the easiness of acquiring high-dimensional data with practical photodetectors. Therefore, we added two additional metrics—dimension reduction factor ϵ and data compression ratio r—to fully characterize the performance of multi-dimensional imagers.

It is noteworthy that, when designing the system, the light efficiency is collectively determined by the desired spectral resolution and the field of view (FOV). On the one hand, a wider slit increases the light throughput, but it also increases the cross talk between adjacent color channels, resulting in a reduced spectral resolution. On the other hand, imaging a large FOV can reduce the light efficiency because an off-axis field point requires the cylindrical lens to provide an elongated line spread function (LSF) in order to extend the signals to the slit location, which, in turn, decreases the irradiance at the slit. Therefore, there is a trade-off between the light throughput and the spectral resolution and the FOV.

Like standard CT, the spatial resolution of Hyper-LIFT is collectively determined by the number of projections and the angle distribution. By contrast, the spectral dimension is directly mapped to the camera without resolution loss. Figure 47(a) shows the peak signal-to-noise ratio (PSNR) versus the number of projections in the reconstructed image for a simulated circular object. The result indicates that the larger the number of projections, the higher the PSNR. Additionally,

74

Hyper-LIFT provides a uniform projection angle distribution from 0 to π, which minimizes the correlations in the projection data and maximizes information content for reconstruction. This is made possible by using a rotated Dove prism array. By contrast, our previous LIFT system created the projections using rotated cylindrical lenses, leading to a limited view problem. More specifically, the LIFT system only captures the projections with an angle in $[-\frac{\pi}{4}, \frac{\pi}{4}]$, rather than $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Therefore, the features along vertical dimension cannot be faithfully reconstructed. To demonstrate the capability of Hyper-LIFT in overcoming this problem, we imaged a group of bars of a USAF resolution target (Group 0 element 6) along both horizontal and vertical directions [Fig. 47(b)]. We further plot the intensity along the dashed line in both images. The average FWHM of a bar is 11 and 12 pixels along the vertical and horizontal direction, respectively. This result indicates that our system has the ability to equally resolve the object features along two orthogonal directions.
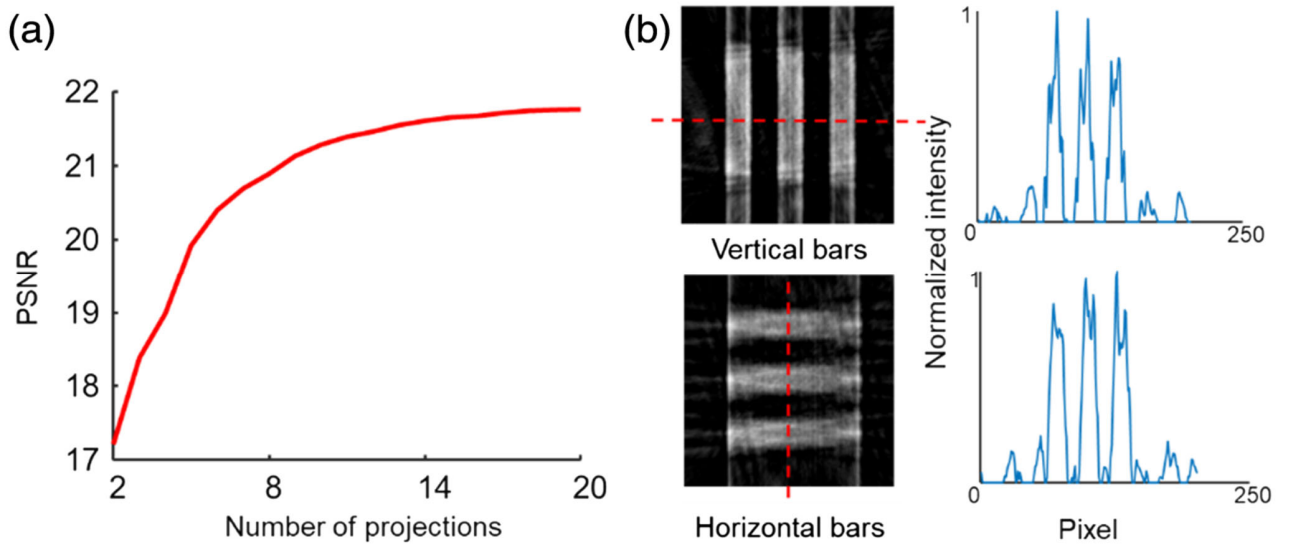

Figure 47. Reconstruction quality in Hyper-LIFT. (a) Peak signal-to noise ratio (PSNR) versus number of projections. (b) Resolving USAF resolution bars along the vertical and horizontal directions. The light intensities (right panel) were plot along the dashed lines in the reconstructed images.

75

In conclusion, we developed and experimentally demonstrated a versatile snapshot Hyper-LIFT imager, and it is the only imager that can perform a largescale hyperspectral light field measurement using a 2D detector array with a moderate format. Because the frame rate of cameras is generally proportional to the reciprocal of the total number of camera pixels in use, our compressed measurement scheme has an advantage in imaging speed. For example, when reading out only the regions of interest (ROIs) that receive light signals from the camera, our system can measure a 5D data cube (x, y, u, v, λ) with $270 \times 270 \times 4 \times 4 \times 360$ voxels at 30 Hz. By contrast, our previous uncompressed (i.e., direct mapping) hyperspectral light field camera [80] can operate at only 5 Hz, even when measuring a data cube of a much smaller size ($66 \times 66 \times 5 \times 5 \times 40$ voxels). This capability is highly desired in imaging applications that simultaneously require a high spatial, spectral, and temporal resolution, such as snapshot spectral-domain optical coherent tomography (SD-OCT) [73], where the previous system suffers from the trade-off between the spatial and spectral resolution. Therefore, we believe the Hyper-LIFT approach will be uniquely positioned in addressing the leading challenges in multi-dimensional optical imaging.

4.8 Light-sheet cascaded LIFT

We present light-sheet cascaded light field tomography (LSC-LIFT), a novel three-dimensional (3D) multispectral imaging microscopy technique. Building upon light field tomography (LIFT), LSC-LIFT incorporates a Dove prism array and a cylindrical lens array to transform a 3D scene into one-dimensional (1D) projections. Unlike the flood illumination employed in a conventional LIFT microscope, LSC-LIFT uses light-sheet illumination for optical scanning, suppressing the out-of-focus light and, thereby, enhancing image contrast. Our innovation also includes a two stage cascade design that significantly increases the number of projection angles for tomographic

reconstruction, thereby substantially improving image quality. Furthermore, we show that the 1D projections in LSC-LIFT can be spectrally dispersed, facilitating multispectral 3D imaging. We demonstrate the efficacy of LSC-LIFT in both fluorescence and scattering microscopy applications.

Light sheet microscopy (LSM) has become a pivotal tool for non-invasive, three-dimensional (3D) imaging of biological specimens [84-89]. Utilizing a scanning light sheet, LSM selectively illuminates a focal plane within the sample and captures images with a two-dimensional (2D) image sensor, enabling non-invasive imaging while minimizing photobleaching. Despite holding great promise, current LSM faces significant challenges, notably the reliance on mechanical scanning to acquire a large imaging volume and inability to image multiple biomarkers simultaneously.

One key challenge arises from the necessity of synchronizing the mechanical scanning of the detection objective lens with the axial movement of the light sheet. This synchronization is to ensure the layer of the object illuminated by the light sheet remains within the objective's depth of focus. However, mechanical scanning slows down the acquisition speed. To address this issue, several strategies have been explored, including using electrically tunable lenses [90, 91] and light field detection methods [92, 93]. Among these solutions, light field imaging stands out due to its snapshot acquisition capability. By capturing spatial (x, y) and angular information (θ, ϕ) of incoming light rays, light field imaging allows for computational refocusing at varying depths, thus achieving an extended depth of field (DOF). When applied to LSM, this ability allows a large imaging volume to be optically scanned while maintaining a high light-collection numerical aperture (NA).

On the other hand, traditional LSM systems provide only intensity images, constraining their ability to image multiple biomarkers and hindering their application in multiplex imaging. In

contrast, multispectral imaging can acquire the spectral information of an input scene [10, 94], allowing for simultaneous imaging of multiple chromophores. Nonetheless, multispectral imaging is typically limited to 2D imaging due to various restrictions on the image sensor.

While both multispectral and light field imaging hold great promise for LSM to expand its functionality, integrating them on the same platform presents considerable challenges. This difficulty arises from a fundamental trade-off involving the acquisition of information along various axes (x, y, z, λ), resulting in a limited resolution given a fixed number of camera pixels. Conventional single detector systems often resort to extensive scanning to mitigate this trade-off. In a previous effort, we developed a snapshot hyperspectral light field tomography (Hyper-LIFT) system to tackle this challenge [95]. Our method enables 3D spectral imaging within an extended DOF. However, Hyper-LIFT has two notable drawbacks. First, although it facilitates numerical refocusing, Hyper-LIFT lacks intrinsic optical sectioning capability, limiting its application in imaging dense tissue samples. Second, Hyper-LIFT is built upon a static optical architecture with a high data compression ratio. It is applicable to only samples exhibiting a high level of sparsity in specific domains.

To address these problems, herein we introduce light-sheet cascaded light field tomography (LSC-LIFT), which integrates light-sheet illumination with a novel two-stage cascaded LIFT on an open-top microscope platform [96]. This synergy remarkably improves image quality and enhances the system's capabilities in suppressing the out-of-focus light in 3D imaging. We demonstrate the effectiveness of our system in imaging both fluorescent and scattering 3D scenes.

The system schematic of the LSC-LIFT is shown in Fig. 1. The setup begins with a supercontinuum laser source (NKT photonics, FIU-15), whose beam is spatially expanded to twice its original size using a beam expander lens assembly (with focal lengths f1 = 50 mm and

f2 = 100 mm). An excitation spectral filter selects the wavelength band for illumination. After passing through a cylindrical lens (f = 150 mm, Edmund Optics, 34-655), the expanded beam forms a sharp line image at a Galvometer (Thorlabs, GVS 201), which adjusts the beam angles with different applied voltages. A scan lens (Thorlabs, LSM03-VIS) converts the varying angles into corresponding shifts in chief ray height, mapping angular changes to spatial coordinates. The beam is then focused by L1 (f = 50 mm) onto the pupil plane of an objective (Olympus, f = 45 mm, 0.1 NA). Consequently, a light sheet is generated, enabling selective illumination of a single plane within the sample with a scanning volume of 0.5×0.5×0.5 mm³.
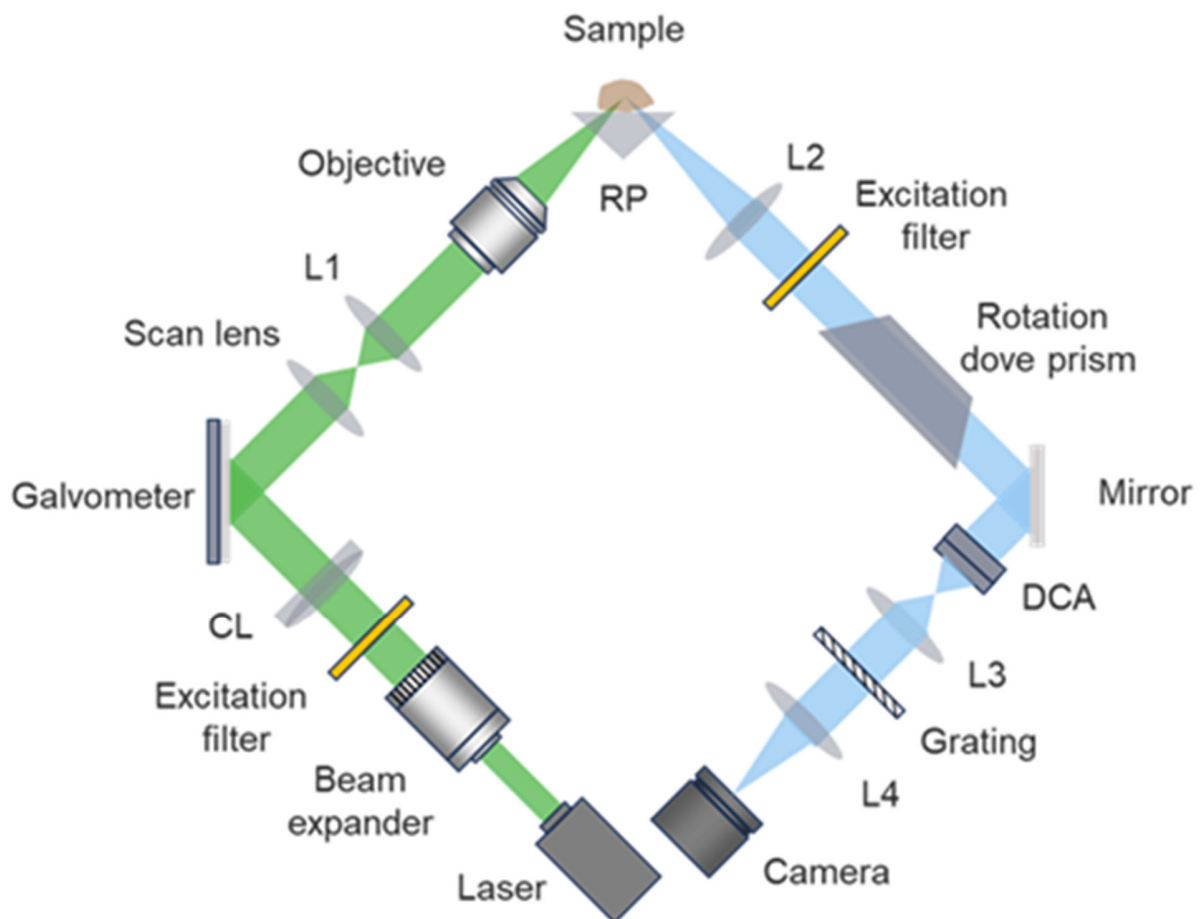

Figure 48. Schematic of light-sheet cascaded light field tomography (LSC-LIFT). CL: cylindrical lens, RP: right angle prism, DCA: Dove prism and cylindrical lens array.

The detection path (highlighted in blue in Fig. 1) involves two-stage cascaded image processing. In the first stage, the input scene undergoes rotation via a rotation Dove prism (RDP), while in the second stage, it undergoes additional rotation and transformation through an array of Dove prisms and cylindrical lenses (DCA). We uniformly spaced the angles of Dove prisms on the array in the range of [0, 180] degrees and rotated the RDP from zero to four degrees, incrementing by one degree per acquisition.

The image forward model for a single imaging channel on the array can be described as:

$$P^{(\theta+\phi)}(\lambda) = TR_\theta R_\phi g(\lambda) + \epsilon. \tag{20}$$

Here, $g(\lambda)$ denotes a monochromatic input scene at wavelength λ. R_θ, R_ϕ denote the image rotations applied by the first-stage and second-stage Dove prisms, respectively. T describes the combined function of the cylindrical lens and the slit. ϵ is the inherent noise of the imaging system. $P^{(\theta+\phi)}(\lambda)$ is the measured 1D projection at wavelength λ, associated with a perspective image rotated by a cumulative angle of $\theta + \phi$.

We rewrite the image forward model in an explicit form:

$$\boldsymbol{P}(\lambda) = T \begin{bmatrix} R^0 \\ ... \\ R^{180} \end{bmatrix} [R_0 \quad ... \quad R_4] g(\lambda) + \epsilon = \boldsymbol{A}g(\lambda) + \epsilon. \tag{21}$$

where $\mathbf{A}$ represents imaging operator of LSC-LIFT, and

$$\boldsymbol{A} = T \begin{bmatrix} R^0 & ... & R^{180} \\ ... & ... & ... \\ R^{180} & ... & R^{180+4} \end{bmatrix}. \tag{22}$$

Due to the relatively low number of projections compared to the pixel resolution of the input perspective image, the aforementioned forward model is under-determined. The solution of Eq. 21 essentially becomes a sparse-view Computed Tomography (CT) reconstruction problem. To estimate $g(\lambda)$, we use an iterative optimization algorithm, fast iterative shrinkage-thresholding

algorithm (FISTA) [68]:

$$argmin_{\hat{g}}\|\boldsymbol{P} - \boldsymbol{A}g\|_2^2 + \rho\|\varphi(g)\|_1. \qquad (23)$$

where $\varphi(g)$ is the regularization function that enforces sparsity in the spatial gradient domain. We use a hyperparameter, ρ, to adjust the weight of the regularization term.

We first calibrated the system's spectral response. In LSC-LIFT, we use a mask to extract the en-face 1D projections of perspective images after the second-stage array system. Each projection image then passes through a diffraction grating, with the dispersed spectrum spanning the space between adjacent subfields. This gap, measuring 2.9 mm, corresponds to 390 camera pixels. To establish the correlation between pixel positions and corresponding wavelengths, we measured a chromatic dispersion curve, as illustrated in Fig. 49(a). For calibration purposes, we used a spatial pinhole (Thorlabs, P20CB) as the object, positioned it at the nominal working distance of the detection system, and illuminated it with a halogen lamp (Amscope, HL250-AS). We then sequentially filtered the light source using five color filters (Thorlabs, FB470-10, FB510-10, FB550-10, FB590-10, FB630-10) and recorded the corresponding pixel locations. The dispersion curve was subsequently fitted using a first-order polynomial. Given that a wavelength bandwidth of 1 nm spans 4 pixels on the camera, the system's spectral sampling is 0.25 nm per pixel. The effective spectral resolution is the convolution of the spectral sampling of a single pixel with the geometrical image width of the slit, which approximates 10 nm. The spectral range is determined by the spacing between two adjacent subfields, and it is estimated to be 100 nm.

The depth calibration involves refocusing the cascaded LIFT system to the depth plane illuminated by the light sheet within its depth range. For this purpose, we measured a disparity curve. The procedure entailed axially adjusting the position of the pinhole, capturing the corresponding image, and then digitally refocusing each image. We recorded the disparity

81

values, s, when the image of the pinhole was optimally focused, as indicated by the highest sharpness. The relationship between the measured disparity and the axial location of the light sheet is illustrated in Fig. 49(b).
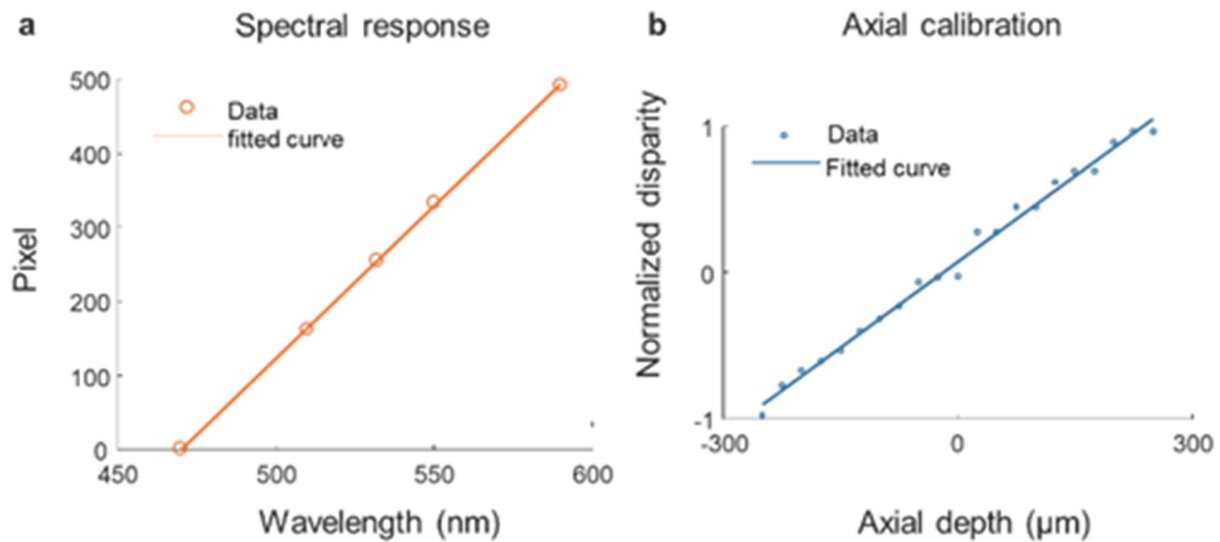
Figure 49. System calibration. (a) Chromatic dispersion. (b) Disparity versus axial depth.

We evaluated the effect of the cascaded two-stage LIFT design on image quality. To emulate the performance of original Hyper-LIFT, we maintained the first Dove prism static in our setup. The imaging function was solely achieved through the second-stage Dove prism and cylindrical lens array. Utilizing 16 1D projections, each measuring 1×270 pixels, the system reconstructed an input scene of 270×270 pixels, achieving a compression ratio of 16.8. Nevertheless, this approach exhibits limitations when dealing with non-sparse scenes, resulting in degradation of image quality. For instance, in our experiment, a USAF resolution target was positioned at the sample plane and illuminated using a halogen lamp (Amscope, HL250-AS) in the transmission mode. The ground-truth image, captured by a reference camera (Thorlabs, CS235MU), is shown in Fig. 50(a). In the reconstructed image (Fig. 50(b)), fine features like bars in group 4 were not

fully resolved due to the limited number of projection angles acquired. In contrast, in the two-stage LIFT system, we augmented projection angles by incrementally rotating the first-stage Dove prism in five steps from zero to four degrees. This accordingly increased the set of image rotate angles obtained from the second-stage Dove prism array by a factor of five, leading to a total of 80 projection measurements. The resultant image quality was significantly improved, as evidenced in Fig. 50(c). To quantify the results, we plotted the intensities along the bars in group 4 (shown in Fig. 50(d)). The image visibility, defined as $(I\_max - I\_min)/(I\_max + I\_min)$, was improved from 0.09 to 0.25. The spatial resolution was determined to be 11 μm.

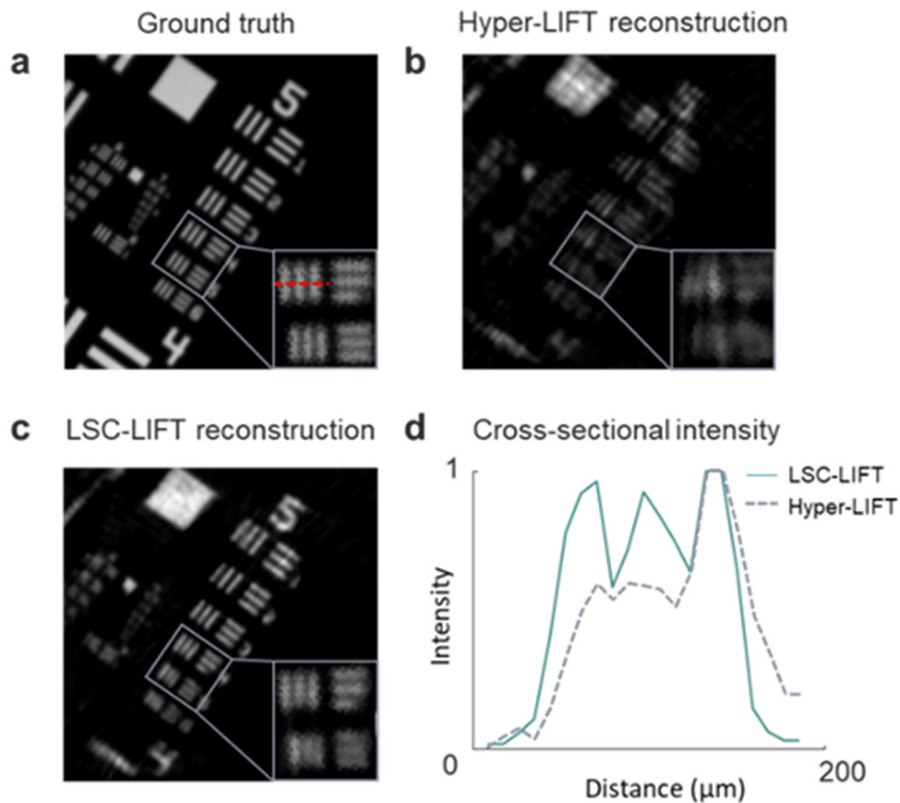Figure 50. Enhanced image quality and spatial resolution in cascaded two-stage LIFT imaging. (a) Ground truth. (b) Reconstruction of the USAF resolution target with a static Rotating Dove Prism (RDP). (c) Reconstruction of the same target with the RDP rotated five times, each increment by one degree. (d) Intensity comparison of bars in group 4 from both (b) and (c), demonstrating increased visibility and improved spatial resolution

We then showcased the system's spectral imaging capabilities in fluorescence microscopy. We mixed two types of fluorescent beads (Thermo Fisher F8844 and F8841) with emission peaks of 515 nm and 560 nm in an agarose gel at a ratio of 1:20 (bead solution to agarose). We used a short-pass (Thorlabs FESH0500) and a long-pass (FELH0500) filter for excitation and emission, respectively, and we scanned a volume of $0.5 \times 0.5 \times 0.5$ mm³. A representative reconstructed depth intensity image is shown in the left panel of Fig. 51(a), where we integrated the light signals across all wavelengths. Fig. 51(b) displays the recovered spectra of the two bead types, aligning well with their specified spectral characteristics. Based on their spectral signatures, we separated the beads using a linear unmixing algorithm and pseudo-colored them in Fig. 51(c).
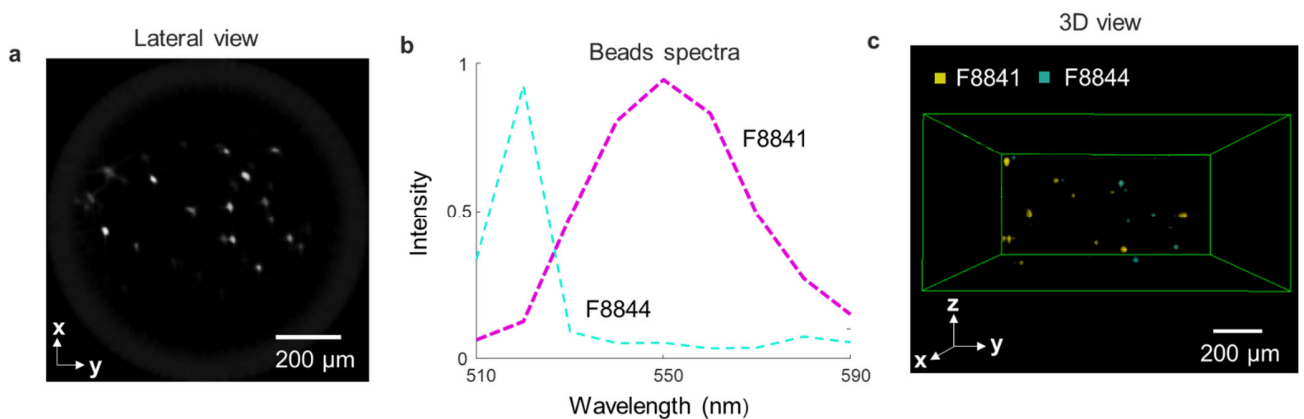
Figure 51. Multispectral 3D Imaging of Two Types of Fluorescent Beads. (a) Representative depth intensity image (b) Spectral reconstruction of two bead types. (c) 3D visualization.

Finally, we conducted 3D scattering imaging of a bovine tissue slice. We used a bandpass filter (Edmund optics, 65-744) to confine the illumination wavelength range to 510-590 nm. Figure 52(a) displays a representative depth layer captured by the reference camera, with an interface between fat (location A) and muscle (location B). The correspondent LSC-LIFT reconstruction is shown in Fig. 52(b), matching well with the ground truth. A 3D representation is depicted in Fig. 52(c), where the muscle's fibrous structures are distinctly visible.
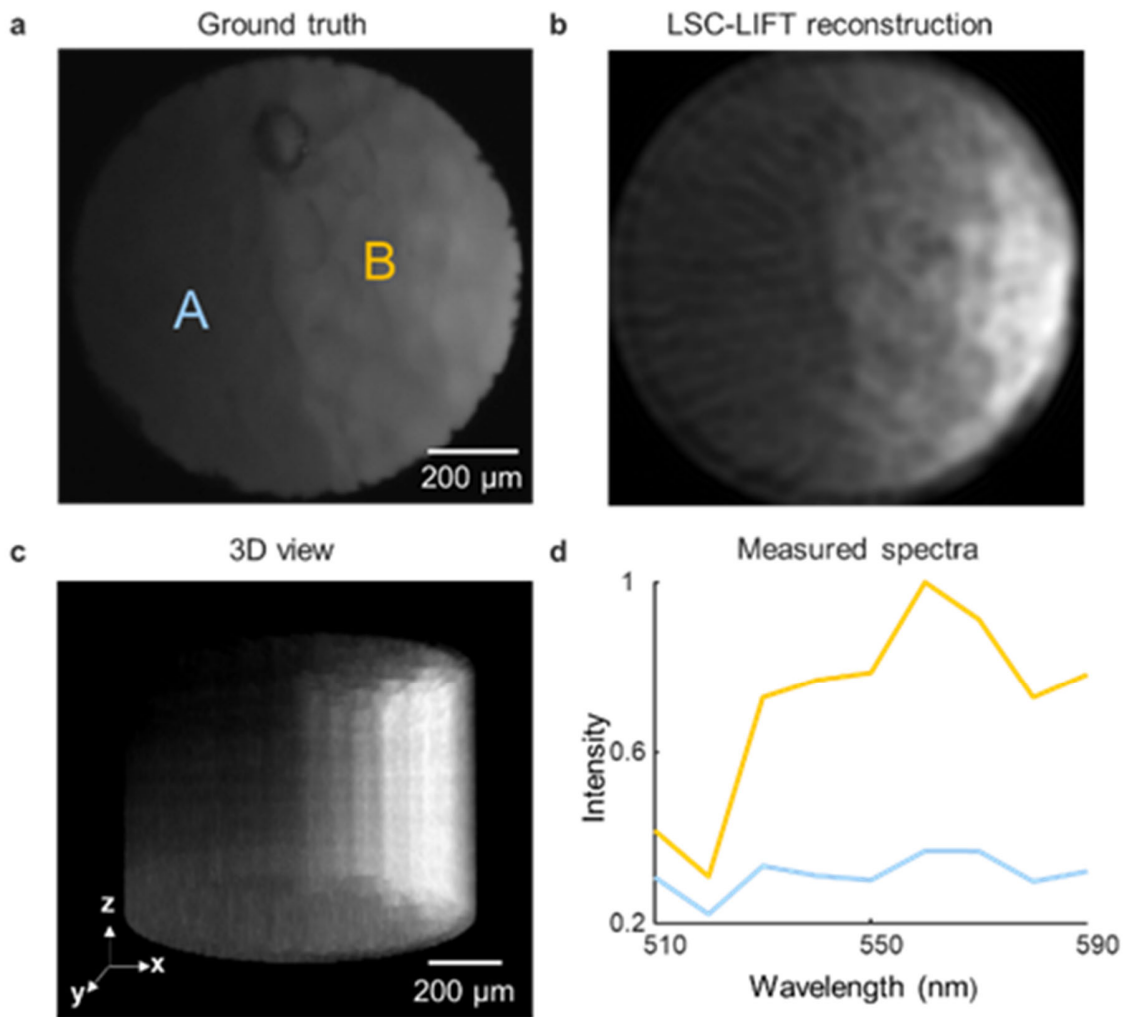
Figure 52. Multispectral 3D Imaging of a bovine tissue slice. (a) Ground truth of a single layer captured by the reference camera, illustrating fat (left side, labeled A) and muscle (right side, labeled B) which contains myoglobin. (b) LSC-LIFT reconstruction results. (c) 3D visualization with muscle fibers displayed on the right. (d) Spectral data representation for both fat and muscle regions.

The spectra recovered from the fat and muscle tissue locations are depicted in Fig. 52(d). These spectra represent a combination of optical absorption and scattering effects within the tissues. Notably, the spectrum obtained from muscle tissue, which is richly supplied with blood, displays a distinct valley at 555 nm, corresponding to the absorption peak of deoxygenated hemoglobin. On the other hand, the fat spectrum shows markedly higher intensities at wavelengths of 530 nm

and above, which is in line with the dependence of its diffuse reflectance on the wavelength within this spectral range.

In summary, LSC-LIFT represents a substantial leap forward compared to the original Hyper-LIFT in terms of image quality and optical sectioning capabilities. When paired with light-sheet illumination, it facilitates the acquisition of high-quality 3D images across an extended depth of field, a feat unattainable with traditional light-sheet microscopy methods. We envision that LSC-LIFT will enable a wide range of applications in biological sciences, particularly in scenarios requiring comprehensive volumetric and spectral data acquisition like in 3D digital pathology [97].

Chapter 5 Summary

While multi-dimensional imaging technologies continue to advance, there is still potential for system improvement. The key objectives for these systems include achieving high temporal resolution (snapshot), high spatial resolution, and high spectral/depth resolution. The main hurdle arises from the necessity to project a high-dimensional data cube onto a 2D detector, which inevitably leads to competition among information from different axes for the limited pixels available.

Hyperspectral imaging maps a 3D data cube (x, y, λ) onto a 2D detector array. Dispersion elements, such as gratings or prisms, are typically employed to separate wavelength information. This direct dispersion, however, introduces crosstalk, countless effort has been made to mitigate such issues. Scanning techniques, including pushbroom, point, and spectral scans [42-48], compromise temporal resolution to enhance spatial or spectral resolution. Brady's group has pioneered the coded aperture snapshot spectral imaging (CASSI) technique [49-51], which employs a random binary mask at the intermediate image plane to encode the incoming scene. Despite persistent crosstalk, exploiting the scene's sparsity allows for the recovery of both spatial and spectral details by addressing the inverse problem. Another innovative approach involves segmenting the input scene into several sub-elements, thereby generating vacant space on the detector for dispersion. Noteworthy in this category are the image mapper spectrometer (IMS) [54–57], utilizing an image mapper with angled mirror facets to segment the scene, and recent Hong et al.'s design of a glass lightguide, utilizing 3D printing technologies to divide the scene into pixels for dispersion [98].

An alternative strategy for wavelength selection involves the use of color filters to isolate single-wavelength scenes, with the principal drawback being reduced light throughput. A prevalent

design integrates a color microfilter array directly onto the sensor [99], typically in a 3×3 or 4×4 pixel grid, aligning each filter element with a camera pixel. Although sophisticated and widely commercialized, this method suffers from poor spatial sampling. While multi-camera setups can enhance pixel sampling, they tend to be cumbersome [100]. Kristina et al. have addressed this spatial limitation by employing a diffuser and spectral filter for wavelength demultiplexing [101].

Dai's group presents a novel method that employs colored sinusoidal illumination to distinguish various wavelength scenes within the Fourier domain, enabling snapshot multispectral imaging [17].

Light field imaging converts a 4D data cube (x, y, u, v) into a 2D detector array representation, often achieved by integrating a microlens array before the detector. Incorporating hyperspectral imaging, this technique progresses to encapsulating a 5D data cube (x, y, u, v, λ) within the detector array. Our prior implementation of a light field Image Mapper Spectrometer (IMS) is capable of capturing this 5D data cube in a single snapshot, ensuring complete light throughput and preserving a compact form factor. However, due to operating under the Nyquist sampling condition, the dimension of the data cube is constrained to be smaller than the total pixel count available on the detector array.

Compressed sensing emerges as a viable method for capturing expansive data cubes in a single snapshot, offering a strategic response to these imaging challenges. By leveraging the inherent sparsity of an input scene, compressed sensing enables the recovery of extensive data cubes using a reduced number of pixels. This approach not only diminishes the amount of data produced but also alleviates the burden on data storage.

In this dissertation, we introduce snapshot light field tomography (Hyper-LIFT) and tunable image projection spectrometry (TIPS), employing compressed sensing to mitigate the trade-offs

between various information dimensions. Despite the advances, the current designs exhibit notable limitations. Firstly, there's a challenge with low light throughput due to the indispensable use of a slit array in both systems, which results in significant light loss at the slit plane. Secondly, the quality of the images is a concern. The reconstruction of the input scene relies on an optical Radon transform, meaning image quality is inherently linked to the quantity of rotated Dove prisms utilized. Additionally, as the Dove prism array must be positioned at the Fourier domain, an increase in the number of Dove prisms enhances the reconstructed image quality but inversely affects spatial resolution. Thirdly, the fabrication of lens arrays, particularly Dove prism arrays, presents a significant challenge. This complexity hampers both the compactness and the cost-efficiency of our systems. Future work will focus on developing innovative and versatile imaging systems suited for a wide array of applications, such as surveillance, biomedical imaging, and autonomous driving. For example, combing coded aperture with LIFT system provides a better image quality than CASSI only system [102].

Other contributions are as follows: in the field of hyperspectral imaging, we have developed a rapid calibration method for Image Mapping Spectrometers (IMS) [103]. In light field imaging, we have established an optical design pipeline for light field cameras, which was demonstrated through the development of a light field endoscope [104].

Chapter 6 Reference

[1]   G. Lippmann, "Épreuves réversibles. Photographies intégrales," Comptes Rendus de l'Académie des Sciences, 1908.

[2]   Ng Ren, "Digital light field photography," *Stanford University*, 2006.

[3]   R. Prevedel, Y. G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, and A. Vaziri, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," Nat. Methods 11(7), 727–730 (2014).

[4]   N. Bedard, T. Shope, A. Hoberman, M. A. Haralam, N. Shaikh, J. Kovačević, N. Balram, and I. Tošić, "Light field otoscope design for 3D in vivo imaging of the middle ear," Biomed. Opt. Express 8(1), 260–272 (2017).

[5]   R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," IEEE Trans. Image Process. 24(3), 1060–1075 (2015).

[6]   K. Maeno, H. Nagahara, A. Shimada, and R. I. Taniguchi, "Light field distortion feature for transparent object recognition," in IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2013), pp. 122–135.

[7]   S. Zhu, X. Lv, X. Feng, J. Lin, P. Jin, and L. Gao, "Plenoptic Face Presentation Attack Detection," IEEE Access 8, 59007–59014 (2020).

[8]   K. Lynch, T. Fahringer, and B. Thurow, "Three-dimensional particle image velocimetry using a plenoptic camera," in 50th AIAA Aerospace Sciences Meeting (AIAA, 2012), pp. 1–14.

[9]   M. Z. Alam and B. K. Gunturk, "Hybrid Light Field Imaging for Improved Spatial Resolution and Depth Range," arXiv preprint arXiv:1611.05008 (2016).

[10] W. Jahr, B. Schmid, C. Schmied, F. O. Fahrbach, and J. Huisken, "Hyperspectral light sheet microscopy," Nat. Commun. **6**(1), 7990 (2015).

[11] S. Karpf, M. Eibl, W. Wieser, T. Klein, and R. Huber, "A time-encoded technique for fibre-based hyperspectral broadband stimulated Raman microscopy," Nat. Commun. 6(1), 6784 (2015).

[12] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," J. Biomed. Opt. 19(1), 010901 (2014).

[13] Q. Pian, R. Yao, N. Sinsuebphon, and X. Intes, "Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging," Nat. Photonics 11(7), 411–414 (2017).

[14] F. Yesilkoy, E. R. Arvelo, Y. Jahani, M. Liu, A. Tittl, V. Cevher, Y. Kivshar, and H. Altug, "Ultrasensitive hyperspectral imaging and biodetection enabled by dielectric metasurfaces," Nat. Photonics **13**(6), 390–396 (2019).

[15] K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array," Optica 7(10), 1298–1307 (2020).

[16] C. Ma, X. Cao, R. Wu, and Q. Dai, "Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system," Opt. Lett. 39(4), 937–940 (2014).

[17] C. Deng, X. Hu, J. Suo, Y. Zhang, Z. Zhang, and Q. Dai, "Snapshot hyperspectral imaging via spectral basis multiplexing in Fourier domain," Opt. Express 26(25), 32509–32521 (2018).

[18] Qi Cui and Rongguang Liang, "Chromatic confocal microscopy using liquid crystal display panels," Appl. Opt. 58, 2085-2090 (2019).

[19] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," Science 228(4704), 1147–1153 (1985).

[20] Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, 2017), pp. 3270–3278.

[21] K. Zhu, Y. Xue, Q. Fu, S. B. Kang, X. Chen, and J. Yu, IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 1131 (2019).

[22] S. Zhu, L. Gao, Y. Zhang, J. Lin, and P. Jin, Opt. Express 26, 26495 (2018).

[23] M. Levoy and P. Hanrahan, "light field rendering," in Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, (1996), P. 31–42.

[24] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," ACM Trans. Graphic 23(3), 294–302 (2004).

[25] C. Perwass and L. Wietzke, "Single-lens 3D camera with extended depth-of-field," in Human Vision and Electronic Imaging XVII, Proc. SPIE8291, 829108 (2012).

[26] I. Tosic and K. Berkner, "Light field scale-depth space transform for dense depth estimation," in IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2014), pp. 435–442.

[27] L. Gao, N. Bedard, and I. Tosic, "Disparity-to-depth calibration in light field imaging," in Imaging and Applied Optics, OSA Technical Digest, (Optical Society of America, 2016), paper CW3D.2.

[28] E. J. Tremblay, D. L. Marks, D. J. Brady, and J. E. Ford, "Design and scaling of monocentric multiscale imagers," Appl. Opt. 51(20), 4691–4702 (2012).

[29] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, and J. C. J. Fernández, "Light field geometry of a standard plenoptic camera," Opt. Express 22(22), 26659–26673 (2014).

[30] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing distance of a standard plenoptic camera," Opt. Express 24(19), 21521–21540 (2016).

[31] Y. Chen, X. Jin, and Q. Dai, "Distance measurement based on light field geometry and ray tracing," Opt. Express 25, 59-76 (2017).

[32] Y. Q. Chen, X Jin, and Q. H. Dai, "Distance estimation based on light field geometric modeling," 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, 2017, pp. 43–48.

[33] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Baseline and triangulation geometry in a standard plenoptic camera," Int. J. Comput. Vis. 126(1), 21–35 (2018).

[34] R. E. Fisher and B. Tadic-Galeb, Optical System Design (McGraw-Hill, 2000).

[35] R. Kingslake and R. B. Johnson, Lens design fundamentals (Academic, 2009).

[36] S. Zhu, A. Lai, K. Eaton, P. Jin, and L. Gao, "On the fundamental comparison between unfocused and focused light field cameras," Appl. Opt. 57(1), A1–A11 (2018).

[37] D. G. Dansereau, G. Schuster, J. Ford, and G. Wetzstein, "A wide-field-of-view monocentric light field camera," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), pp. 5048–5057.

[38] Tara M. Urner, Andrew Inman, Benjamin Lapid, and Shu Jia, "Three-dimensional light-field microendoscopy with a GRIN lens array," Biomed. Opt. Express 13, 590-607 (2022).

[39] E. Kwan, Y. Qin, and H. Hua, "Development of a Light Field Laparoscope for Depth Reconstruction," in Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP), OSA Technical Digest (online) (Optical Society of America, 2017), paper DW1F.2.

[40] S. Zhu, P. Jin, R. Liang, and L. Gao, "Optical design and development of a snapshot light-field laryngoscope," Opt. Eng. 57(2), 023110 (2018).

[41] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-D deconvolution for the light field microscope," Opt. Express 21(21), 25418–25439 (2013).

[42] K. C. Lawrence, W. R. Windham, B. Park, and R. J. Buhr, "A hyperspectral imaging system for identification of faecal and ingesta contamination on poultry carcasses," J. Near Infrared Spectrosc. 11, 269–281 (2003).

[43] M. Abdo, V. Badilita, and J. Korvink, "Spatial scanning hyperspectral imaging combining a rotating slit with a Dove prism," Opt. Express 27, 20290–20304 (2019).

[44] Y. J. Hsu, C.-C. Chen, C.-H. Huang, C.-H. Yeh, L.-Y. Liu, and S.-Y. Chen, "Line-scanning hyperspectral imaging based on structured illumination optical sectioning," Biomed. Opt. Express 8, 3005–3016 (2017).

[45] P.-H. Cu-Nguyen, A. Grewe, M. Hillenbrand, S. Sinzinger, A. Seifert, and H. Zappe, "Tunable hyperchromatic lens system for confocal hyperspectral sensing," Opt. Express 21, 27611–27621 (2013).

[46] N. Gat, "Imaging spectroscopy using tunable filters: a review," Proc. SPIE 4056, 50–64 (2000).

[47] M. C. Phillips and N. Hô, "Infrared hyperspectral imaging using a broadly tunable external cavity quantum cascade laser and microbolometer focal plane array," Opt. Express 16, 1836–1845 (2008).

[48] G. Di Caprio, D. Schaak, and E. Schonbrun, "Hyperspectral fluorescence microfluidic (HFM) microscopy," Biomed. Opt. Express 4, 1486–1493 (2013).

[49] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded

aperture snapshot spectral imaging," Appl. Opt. 47, B44–B51 (2008).

[50]   A. A.Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," Opt. Express 17, 6368–6388 (2009).

[51]   L.Wang, Z. Xiong, D. Gao, G. Shi, and F.Wu, "Dual-camera design for coded aperture snapshot spectral imaging," Appl. Opt. 54, 848–858 (2015).

[52]   M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," Appl. Opt. 34, 4817–4826 (1995).

[53]   B. K. Ford, M. R. Descour, and R. M. Lynch, "Large-image-format computed tomography imaging spectrometer for fluorescence microscopy," Opt. Express 9, 444–453 (2001).

[54]   L. Gao, R. T. Kester, and T. S. Tkaczyk, "Compact image slicing spectrometer (ISS) for hyperspectral fluorescence microscopy," Opt. Express 17, 12293–12308 (2009).

[55]   R. T. Kester, L. Gao, and T. S. Tkaczyk, "Development of image mappers for hyperspectral biomedical imaging applications," Appl. Opt. 49, 1886–1899 (2010).

[56]   L. Gao, R. T. Kester, N. Hagen, and T. S. Tkaczyk, "Snapshot image mapping spectrometer (IMS) with high sampling density for hyperspectral microscopy," Opt. Express 18, 14330–14344 (2010).

[57]   L. Gao, R. T. Smith, and T. S. Tkaczyk, "Snapshot hyperspectral retinal camera with the image mapping spectrometer (IMS)," Biomed Opt. Express 3, 48–54 (2012).

[58]   R. T. Kester, N. Bedard, L. S. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," J. Biomed. Opt. 16, 056005 (2011).

[59]   A. D. Elliott, L. Gao, A. Ustione, N. Bedard, R. Kester, D. W. Piston, and T. S. Tkaczyk, "Real-time hyperspectral fluorescence imaging of pancreatic β-cell dynamics

with the image mapping spectrometer," J. Cell Sci. 125, 4833–4840 (2012).

[60] M. E. Pawlowski, J. G. Dwight, T.-U. Nguyen, and T. S. Tkaczyk, "High performance image mapping spectrometer (IMS) for snapshot hyperspectral imaging applications," Opt. Express 27, 1597–1612 (2019).

[61] T.-U. Nguyen, M. C. Pierce, L. Higgins, and T. S. Tkaczyk, "Snapshot 3D optical coherence tomography system using image mapping spectrometry," Opt. Express 21, 13758–13772 (2013).

[62] N. Bedard, N. A. Hagen, L. Gao, and T. S. Tkaczyk, "Image mapping spectrometry: calibration and characterization," Opt. Eng. 51, 111711 (2012).

[63] L. Gao and T. S. Tkaczyk, "Correction of vignetting and distortion errors induced by two-axis light beam steering," Opt. Eng. 51, 043203 (2012).

[64] Y. Xue, K. Zhu, Q. Fu, X. Chen, and J. Yu, "Catadioptric hyperspectral light field imaging," in Proceedings of the IEEE International Conference on Computer Vision (2017), pp. 985–993.

[65] Y. Rivenson, A. Stern, and B. Javidi, "Overview of compressive sensing techniques applied in holography [Invited]," Appl. Opt. 52(1), A423–A432 (2013).

[66] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive Holography," Opt. Express 17(15), 13040–13049 (2009).

[67] R. Horisaki, J. Tanida, A. Stern, and B. Javidi, "Multidimensional imaging using compressive Fresnel holography," Opt. Lett. 37(11), 2013–2015 (2012).

[68] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. 2(1), 183–202 (2009).

[69] X. Feng and L. Gao, "Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging," Nat. Commun. 12(1), 2179 (2021).

[70] H. Kudo, T. Suzuki, and E. A. Rashed, "Image reconstruction for sparse-view CT and interior CT-introduction to compressed sensing and differentiated backprojection," Quantum Imaging Med. Surg. 3(3), 147–161 (2013).

[71] A. V. Oppenheim, J. Buck, M. Daniel, A. S. Willsky, S. H. Nawab, and A. Singer, Signals & Systems (Pearson Educación, 1997).

[72] Jaeyul Lee, Xiaoxi Du, Jongchan Park, Qi Cui, Rishyashring R. Iyer, Stephen A. Boppart, and Liang Gao, "Tunable image-mapping optical coherence tomography," Biomed. Opt. Express 14, 627-638 (2023).

[73] R. R. Iyer, M. Žurauskas, Q. Cui, L. Gao, R. Theodore Smith, and S. A. Boppart, "Full-field spectral-domain optical interferometry for snapshot three-dimensional microscopy," Biomed. Opt. Express 11(10), 5903–5919 (2020).

[74] P. R. Griffiths, "Fourier transform infrared spectrometry," Science 222(4621), 297–302 (1983).

[75] T. Hirschfeld, "Fellgett's advantage in UV-VIS multiplex spectroscopy," Appl. Spectrosc. 30(1), 68–69 (1976).

[76] Qi Cui, Jongchan Park, Jaeyul Lee, Zhaoqiang Wang, and Liang Gao, "Tunable image projection spectrometry," Biomed. Opt. Express 13, 6457-6469 (2022).

[77] L. Meng, L. Lu, N. Bedard, and K. Berkner, "Object space calibration of plenoptic imaging systems," U.S. patent 9544583 (Jan 10, 2017).

[78] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in Computational Models of Visual Processing (MIT, 1991), pp. 3–20.

[79] L. Gao and L. V. Wang, "A review of snapshot multidimensional optical imaging: measuring photon tags in parallel," Phys. Rep. 616, 1–37 (2016).

[80] Q. Cui, J Park, R. T. Smith, and L. Gao, "Snapshot hyperspectral light field imaging using

image mapping spectrometry," Opt. Lett. 45, 772-775 (2020).

[81] S. Zhu, L. Gao, Y. Zhang, J. Lin, and P. Jin, "Complete plenoptic imaging using a single detector," Opt. Express 26, 26495-26510 (2018).

[82] X. Lv, Y. Li, S. Zhu, X. Guo, J. Zhang, J. Lin, and P. Jin, "Snapshot spectral polarimetric light field imaging using a single detector," Opt. Lett. 45, 6522-6525 (2020).

[83] N. Hagen, L. Gao, T. Tkaczyk, and R. Kester, "Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems," Opt. Eng. 51, 111702 (2012).

[84] P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. Stelzer, "Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy," science 322, 1065-1069 (2008).

[85] P. A. Santi, "Light sheet fluorescence microscopy: a review," Journal of Histochemistry & Cytochemistry 59, 129-138 (2011).

[86] O. E. Olarte, J. Andilla, D. Artigas, and P. Loza-Alvarez, "Decoupled illumination detection in light sheet microscopy for fast volumetric imaging," Optica 2, 702-705 (2015).

[87] M. B. Ahrens, M. B. Orger, D. N. Robson, J. M. Li, and P. J. Keller, "Whole-brain functional imaging at cellular resolution using light-sheet microscopy," Nature methods 10, 413-420 (2013).

[88] E. H. Stelzer, F. Strobl, B.-J. Chang, F. Preusser, S. Preibisch, K. McDole, and R. Fiolka, "Light sheet fluorescence microscopy," Nature Reviews Methods Primers 1, 73 (2021).

[89] W. Shao, M. Chang, K. Emmerich, P. O. Kanold, J. S. Mumm, and J. Yi, "Mesoscopic oblique plane microscopy with a diffractive light-sheet for large-scale 4D cellular resolution imaging," Optica 9, 1374-1385 (2022).

[90] F. O. Fahrbach, F. F. Voigt, B. Schmid, F. Helmchen, and J. Huisken, "Rapid 3D light-sheet microscopy with a tunable lens," Optics express 21, 21010-21026 (2013).

[91] Y. Nakai, M. Ozeki, T. Hiraiwa, R. Tanimoto, A. Funahashi, N. Hiroi, A. Taniguchi, S. Nonaka, V. Boilot, and R. Shrestha, "High-speed microscopy with an electrically tunable lens to image the dynamics of in vivo molecular complexes," Review of Scientific Instruments 86 (2015).

[92] D. Wang, S. Xu, P. Pant, E. Redington, S. Soltanian-Zadeh, S. Farsiu, and Y. Gong, "Hybrid light-sheet and light-field microscope for high resolution and large volume neuroimaging," Biomedical optics express 10, 6595-6610 (2019).

[93] D. Liang, X. Peng, Y. Hu, F. Zhao, S. Zheng, G. Situ, and J. Liu, "Light-sheet light-field fluorescence microscopy," Optics and Lasers in Engineering 153, 107015 (2022).

[94] Z. Lavagnino, J. Dwight, A. Ustione, T.-U. Nguyen, T. S. Tkaczyk, and D. W. Piston, "Snapshot hyperspectral light-sheet imaging of signal transduction in live pancreatic islets," Biophysical journal 111, 409-417 (2016).

[95] Q. Cui, J. Park, Y. Ma, and L. Gao, "Snapshot hyperspectral light field tomography," Optica 8, 1552-1558 (2021).

[96] A. K. Glaser, N. P. Reder, Y. Chen, C. Yin, L. Wei, S. Kang, L. A. Barner, W. Xie, E. F. McCarty, C. Mao, A. R. Halpern, C. R. Stoltzfus, J. S. Daniels, M. Y. Gerner, P. R. Nicovich, J. C. Vaughan, L. D. True, and J. T. C. Liu, "Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues," Nature Communications 10, 2781 (2019).

[97] A. K. Glaser, N. P. Reder, Y. Chen, E. F. McCarty, C. Yin, L. Wei, Y. Wang, L. D. True, and J. T. C. Liu, "Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens," Nature Biomedical Engineering 1, 0084 (2017).

[98] Z. Hong, Y. Sun, P. Ye, D. A. Loy, and R. Liang, "Bio-Inspired Compact, High-Resolution Snapshot Hyperspectral Imaging System with 3D Printed Glass Lightguide Array," Adv. Opt. Mater. 11, 1–12 (2023).

[99] Yuri Murakami, Masahiro Yamaguchi, and Nagaaki Ohyama, "Hybrid-resolution multispectral imaging using color filter array," Opt. Express 20, 7173-7183 (2012).

[100] N. Genser, J. Seiler and A. Kaup, "Camera Array for Multi-Spectral Imaging," in IEEE Transactions on Image Processing, vol. 29, pp. 9234-9249, 2020.

[101] K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array," Optica 7, 1298-1307 (2020).

[102] Ruixuan Zhao, Qi Cui, Zhaoqiang Wang, and Liang Gao, "Coded aperture snapshot hyperspectral light field tomography," Opt. Express 31, 37336-37347 (2023).

[103] Qi Cui, Jongchan Park, Rishyashring R. Iyer, Mantas Žurauskas, Stephen A. Boppart, R. Theodore Smith, and Liang Gao, "Development of a fast calibration method for image mapping spectrometry," Appl. Opt. 59, 6062-6069 (2020).

[104] Qi Cui, Shuaishuai Zhu, and Liang Gao, "Developing an optical design pipeline for correcting lens aberrations and vignetting in light field cameras," Opt. Express 28, 33632-33643 (2020).