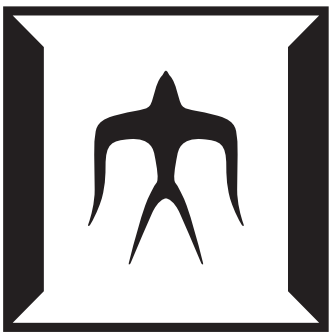
論文 / 著書情報
Article / Book Information

題目(和文)	薄型で機能的なコンピュテーショナルイメージングに基づくカメラシステム
Title(English)	A thin and functional camera system based on computational imaging
著者(和文)	CHENXiao
Author(English)	Xiao Chen
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12543号, 授与年月日:2023年9月22日, 学位の種別:課程博士, 審査員:山口 雅浩,熊澤 逸夫,金子 寛彦,SLAVAKIS KONSTANTINO,渡辺 義浩
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12543号, Conferred date:2023/9/22, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

A Thin and Functional Camera System Based on Computational Imaging



Xiao Chen

Supervisor: Prof. Masahiro Yamaguchi

Department of Information and Communications Engineering
School of Engineering
Tokyo Institute of Technology

This dissertation is submitted for the degree of
Doctor of Engineering

August 2023

Acknowledgements

I am sincerely grateful to my supervisor, Professor Masahiro Yamaguchi, whose guidance, assistance, and support have been instrumental in the completion of this thesis. His dedication as a pure researcher and his diligent work ethic have deeply influenced my own approach to research. I would also like to express my gratitude to Associate Professor Tomoya Nakamura from Osaka University and Assistant Professor Saori Takeyama from Yamaguchi Lab. They have consistently provided invaluable assistance with my research, tackling technical details, and mathematical problems, and even offering guidance on job searches.

I extend my heartfelt thanks to Dr. Shunsuke Igarashi and Mr. Takuto Watanabe, who were my student tutors when I first arrived in Japan. They played a crucial role in introducing me to the study of lensless cameras and provided valuable support during my initial stages of learning. I am truly grateful for the support and camaraderie of Dr. Xiuxi Pan, Dr. Alexis Sánchez Salazar, Mr. Cunyuan Ji, Dr. Isamu Nakao, and Dr. Hossain Md Shakhawat. We spent countless days and nights together in the laboratory, sharing experiences not only in research but also in life. Special thanks to Mr. Zhenghao Chen from Inagi Lab and Mr. Zhenpeng Zuo from Boston University for their assistance, both academically and personally.

I will always cherish the memories of the joyful and exciting times spent with the members of the Yamaguchi laboratory.

My deepest appreciation goes to my father, Mr. Weidong Zhou, my mother, Ms. Feng Chen, and all my family members for their unwavering support, help, and understanding throughout my journey. Especially for my late grandfather, who watched over me since my infancy, and even in his ailing moments, offered unwavering encouragement during my difficult times. Lastly, I want to express my love and gratitude to Miss. Deng. We have experienced so much together, and without your encouragement, I would not have achieved what I have. I am excited to spend the rest of my life with you, forever by my side.

Finally, I am very grateful for the support from the Tokyo Tech Academy for Convergence of Materials and Informatics (TAC-MI) and the Tokyo Tech Pioneer Doctoral Research Project (Cross the border!), not only financially, but also academically and professionally. I also wish to extend my heartfelt gratitude to Dr. Eiji Shidoji and Mr. Tsuguhide Isemura of AGC Inc. Their roles as my industry mentors have been pivotal, offering both sage advice on my career trajectory and profound insights into life

Xiao Chen in Yokohama

Abstract

To overcome the limitations of traditional cameras, which rely on lenses to form images on image sensors, computational imaging technology offers promising solutions for developing thin and functional camera systems. Computational imaging combines optical modulation with digital computation, shifting the burden of the optical element to post-digital processing. In terms of the camera itself, a lensless camera that uses a coded mask instead of an optical lens is discussed. Compared to traditional lens-based cameras, a lensless camera is more compact and less expensive, and its depth-sensing capability allows the image to be refocused after capture. By considering the diffraction properties, I mathematically analyze the resolution limit of a Fresnel zone aperture (FZA)-based lensless camera. In addition, I propose two methods for super-resolution in the FZA lensless camera using image synthesis techniques. Furthermore, I discuss an AI-based refocusing model for lensless cameras. Considering the versatile functions of an imaging system that provides not only photography function but also other media functions. A novel see-through screen (STS) camera based on a holographic waveguide device (HWD) is proposed to realize a thin and versatile media system. The STS camera can be applied to portable devices and achieve frontal imaging for eye contact. I analyze the imaging process of the STS camera based on ray tracing and propose a noise reduction method. I also discuss the possibility of combining lensless cameras with the STS camera for future applications. This study contributes to improving the resolution of lensless cameras under diffraction influences, discusses a learning-based refocusing reconstruction model for the lensless camera, and provides an integrated design for combining imaging and other media functions based on HWD, thus promoting the miniaturization and multifunctionality of cameras.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

$\mathscr{F}$	Fourier transform
$\mathscr{F}^{-1}$	inverse Fourier transform
j	unit imaginary number $\sqrt{-1}$

Greek Symbols

β	parameter controlling the pitch size of the FZA
π	$\simeq 3.14\ldots$
φ	initial phase of the FZA

Acronyms / Abbreviations

AI	Artificial Intelligence
CCD	Charge Coupled Device
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Network
CS	Compressed Sensing
FS	Fringe Scanning
FZA	Fresnel Zone Aperture
HOE	Holographic Optical Elements

IoT Internet of Things

MLS Maximum Length Sequences

MURA Modified Uniformly Redundant Array

PSF Point Spread Function

STS See Through Screen

URA Uniformly Redundant Array

vHOE volume Holographic Optical Elements

Chapter 1

Introduction

1.1 Background and Motivation

Images serve as valuable sources of information in our IoT-driven society and the demand for cameras, as important imaging devices, is continuously increasing. Traditional camera systems typically consist of optical lenses and image sensors (CMOS or CCD), where light from the object is captured by the lens and recorded by the sensor as intensity values. However, the reliance on optical lenses poses limitations on camera size, cost, and compactness due to the need for sufficient space and high-precision manufacturing. Moreover, depth information is now essential for various applications, but a single camera is unable to provide this information. A stereo camera[1, 2] or light-field technology [3]is needed to extract the depth information, which needs multiple sensors and lenses, leading to increased costs. In addition, the camera is commonly integrated into various visual media systems. However, due to traditional design approaches, the camera is typically positioned separately from other media devices, such as displays. This disjointed setup leads to larger overall system sizes, making it challenging to deploy them on portable devices for increased flexibility in applications. In certain interaction scenarios, the user's viewing angle towards the display may not align with the camera's angle of view. This mismatch can result in suboptimal user experiences, particularly in online meetings where the lack of eye contact during communication can be problematic.

With the rise of the digital society, image-sensing technology finds applications in diverse industries. For example, small robots are used to explore or detect objects in a gap

environment, in which case the size and weight of the camera embedded in the robot should be controlled. However, the thickness and focusing length requirements of the optical lens limit the development of miniaturization. This is also a problem with smartphone design, where the camera is an issue that limits the thickness of the phone. Another problem with smartphones is that the camera affects the design of the screen. The coordinated design of screens and cameras is a hot topic of research. As mentioned earlier, the split design means that the camera cannot capture the frontal view of the user looking at the screen. Moreover, in a trend of full-screen smartphone design, the camera destroys the true full-screen design. Some smartphone companies are proposing a concept of an under-display camera, where the camera is placed under the pixels of the screen and the distance between the pixels is increased in a certain area to let the camera take pictures, which reduces the display quality. Hence, alternative solutions are sought.

Computational imaging is cross-disciplinary research involving optical modulation and digital processing. It shifts the burden from optics to post-processing. By designing the optical encoding and digital decoding pipeline, the imaging system could break the optical limits and recover higher dimensions of information. Based on the concept of computational imaging, this study proposes a thin and functional camera system. This study is divided into two main parts. The first part focuses on leveraging lensless imaging technology to achieve a thin camera structure that possesses depth-sensing capabilities. The second part of the study involves integrating the camera function with other media functionalities in a media terminal and proposing a holographic waveguide device-based see-through screen camera. In addition, a combination of lensless imaging with a see-through screen camera is discussed for a more advanced camera system.

1.2 Thesis Structure

This thesis is structured into 6 chapters as follows,

Chapter 1. Introduction

- This section describes the background of this thesis including the demand from the society development and the motivation based on this.

Chapter 2. Computational imaging for thin and functional camera system

- This section reviews the basic knowledge of the technology used for solving issues including general computational imaging, a lensless imaging system, and a Holographical optical elements-based imaging system.

Chapter 3. Superresolution in FZA lensless camera

- This section first presents the related research on FZA lensless cameras, including the original geometric optics-based analysis. Then, the issues affecting the resolution are defined. A theoretical introduction to the proposed method is presented with computer simulations and optical experiments. Advantages and limitations are also discussed.

Chapter 4. AI-based refocusing for lensless camera

- This section provides an overview of the relevant AI-based methods in lensless camera research and their limitations. Then the idea of an AI-based refocusing reconstruction model is proposed. Details of the experiment are presented in the section.

Chapter 5. See-through-screen camera and its application with lensless camera

- This section introduces the limitations of the traditional split design of camera and display, and reviews some published work. The theoretical explanation with ray tracing based analysis of the proposed see-through screen camera is presented. Details of experimental settings, results and discussion are also presented. Finally, future work on the combined design of lensless camera with STS camera is also presented.

Chapter 6. Conclusion

- This section concludes the thesis and explores future work and possible applications.

Chapter 2

Computational imaging for thin and functional camera system

2.1 Background

Over the past few decades, cameras have developed rapidly in a wide range of applications, including smartphones, SLR cameras, and drones. However, the structure of cameras hasn't changed much, based on a set of optical lenses and an image sensor. The light reflected from the object passes through the optical lens and is then sampled by the pixels in the sensor, creating an inverted image of the object on the sensor. In addition, most research studies optical lens optimization or sensor signal processing algorithms independently of each other. To overcome the limitations in resolution, information dimension, function, and size caused by the conventional camera structure, computational imaging is a promising technology that considers the parallel design and joint optimization of optical systems and image processing algorithms to further overcome the limitations of traditional imaging systems, rather than a collection of individual components[4–8].

Computational Imaging is a novel imaging design consisting of front-end optical encoding and back-end digital decoding. Based on the application motivation, the vision system can be designed in a flexible way [9, 10]. Computational Imaging technology can be divided into three classes based on different objectives[11]. The first is for high performance, mostly in terms of traditional camera evaluation indices, e.g. super-resolution[12], wider field of view[13], and deeper depth of field[14]; the second is for obtaining new information

that is impossible for the traditional camera, such as 3D information[15, 16], spectral information[17], and phase information[18]. The third is for special size requirements or special design requirements for integration with other digital devices, e.g. light size[19, 20]. There is also some research that could promote a thin and functional camera system. *Coded aperture imaging* is based on an intensity modulator at the pupil plane. It can be used for defocus deblurring and depth estimation[21–23]. Some studies report the possibility of using coded aperture for spectral imaging[24]. This technology could bring depth information from the 3D world alongside the 2D image projection, but it is still based on the optical lens, and the size and cost are still limited by the lens. *Light field technology* [25–27] uses a lens array or camera array to capture 3D information from the world by building a multi-dimensional model called a plenoptic function that includes wavelength, location, angle and time. This technology can be used for 3D reconstruction[28, 29], post-capture refocusing[30], and depth estimation[31, 32]. However, some challenges such as high computational complexity and high hardware cost still need to be urgently addressed, and because of the lens, the true thin can not realize.

We argue that the optical lens is the factor that limits the size and function of conventional cameras. A lensless imaging technology is proposed for a thin camera system with depth-sensing capability. Since the camera is usually integrated with a display in the visual media terminals, the convergence of the camera with the display is a promising idea for a thin imaging system. Based on this, another idea is proposed to make a display as a camera by using the HOE. A brief introduction to lensless imaging (Section.2.2) and HOE-based imaging systems (Section.2.3) is presented in the following.

2.2 Lensless imaging system

The development of the camera is rapid, but the basic structure of the camera remains the same. The optical lens is the main problem limiting the development of the camera. Recently, with the popularity of computational imaging - a new imaging technology that combines the design of front-end hardware and back-end algorithm together, rather than modifying each separately. A type of computational imaging technology known as a lensless camera, which replaces the optical lens with a phase or amplitude mask placed directly in front of a bare image sensor, has recently attracted a great deal of attention. In the normal optical lens-based camera system, the impulse response or PSF is very small and almost a point, so a point-to-point mapping function is established between the object scene and the measurement

on the sensor. However, in a lensless camera, the PSF is actually the shadow of the mask pattern, which is very large and takes up most of the sensor area. So the measurement on the sensor is the superposition of these shadows and a reconstruction or decoding process is required [33–35]. As shown in Fig.2.1, (a) describes the normal imaging process of a lens-based camera, where the captured image is an inverted image of the object, then goes through the ISP pipeline before being displayed to humans, the digital process here is just some quality enhancement of the captured image; (b) shows the encoding and decoding process of lensless cameras, the captured image is meaningless to the human eye and a digital image reconstruction process is required.

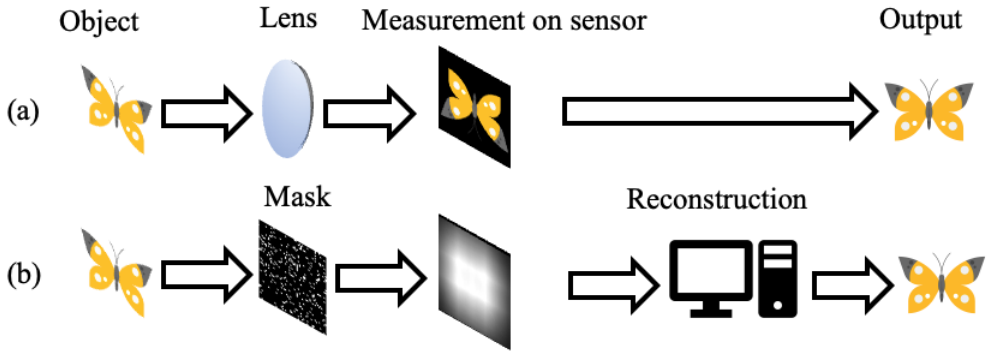


Fig. 2.1 Imaging pipelines of the (a) conventional camera and (b) lensless camera.

This type of special imaging structure brings a number of potential advantages to the measurement of vision in terms of size, cost, and performance. The detailed explanation is as follows,

- **Thin Structure.** Because there is no optical lens, there is no need to reserve space for a focusing lens. This means that the thickness of the lensless camera could be similar to the thickness of the image sensor, as the mask could act as a film cover for the image sensor at a small distance. This feature allows a lensless camera to be a thin camera system.
- **Low Cost.** The lens manufacturing process requires high-precision engineering to ensure optimum image quality. This involves intricate designs, precise alignment of lens elements and advanced optical coatings, all of which add to the complexity and cost of production. In addition to the cost of the lens itself, the assembly of optical lenses into cameras requires high precision and high cost. In a lensless camera, the cost of the lens itself and its assembly could be eliminated, reducing the cost of the entire camera.

- **Depth Aware.** In the shift-invariant imaging system, the image captured on the sensor is the convolution between the PSF and the intensity distribution of the object. The size of the PSF on the sensor changes according to the distance of the object. The PSF in the traditional camera is very small and the change in size could be neglected in the depth of field. However, the PSF in the lensless camera is huge and the size of the PSF is sensitive to changes in distance. So this unique capability could be used for depth information processing, including 3D reconstruction[36], depth estimation[37], and refocusing[38].

2.2.1 Coded mask

The traditional camera uses a focusing optical lens to converge the light from the object onto the image sensor. In a lensless camera, instead of an optical lens, a coded mask is placed in front of the image sensor to encode the incoming object light. The design of the mask is an important issue in the study of lensless cameras. The following explanation will focus on mask type and mask pattern.

Mask Type

An amplitude mask has different open and closed areas on the aperture. Traditional binary masks have only two distinct areas: transparent and opaque. However, there are masks that use greyscale amplitude, allowing the intensity of transmitted light to be modulated. The manufacturing process of amplitude masks is very simple. However, the light efficiency of the amplitude mask is low because the mask's dark area is large. And the low light efficiency affects the image quality. To achieve high resolution, the pitch size of the mask should be small enough, but diffraction is more serious when the pitch size is small. To increase the light efficiency of the mask, a phase mask is used in the lensless camera. Phase masks modulate the input light by changing the thickness of the mask. It can provide higher image quality than an amplitude mask. In addition, the PSF on the sensor from the phase mask can be sharper and more contrasty, which helps to improve the reconstruction quality.

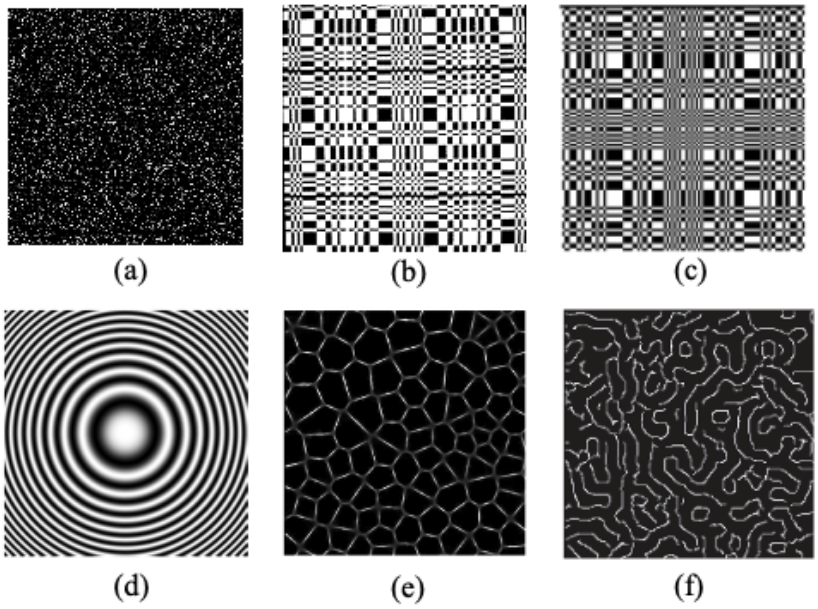


Fig. 2.2 Mask patterns from (a) random mask; (b) URA [39]; (c) MURA [40]; (d) FZA [43]; (e) DiffuserCam [36]; (f) PhlatCam [38].

Mask Pattern

Discovered by a Chinese philosopher called Mo Zi around the 4th century BC, the pinhole camera is one of the earliest and simplest forms of a lensless camera. However, a single hole severely limits light efficiency. To increase light efficiency, the random mask is proposed, where the multiple pinholes are arranged randomly on the mask, [39] pointed out that the number of N holes on the mask can increase the SNR to roughly $\sqrt{N}$ times. To achieve a near-flat frequency response and a near-delta autocorrelation function of mask, URA[39] and MURA[40] are proposed. In addition, the computational complexity of the reconstruction process has led some researchers to consider the design of mask patterns. [41, 42] proposes a FlatCam, which is the outer product of the MLS, to reduce the computational complexity. A lensless camera based on FZA has recently been proposed[43], whose mathematically modelable pattern facilitates deconvolution reconstruction and 3D processing. The aforementioned are amplitude mask-based lensless cameras; In the research on phase mask-based lensless cameras, [36, 44] uses a diffuser film as a phase mask, which is very simple and cheap. However, as the pattern of the diffuser is random, the distance between the mask and the sensor should be varied to achieve the best imaging performance. [38] proposes a PhlatCam using a Perlin noise-based contour pattern which produces high-contrast and

spatially-sparse PSF. Some of the typical mask patterns mentioned above are shown in Fig2.2, (a) is random masks, (b) is URA, (c) is MURA, (d) is FZA, (e) is PSF from diffuser mask, and (f) is contour PSF.

2.2.2 reconstruction methods

Due to the unique structure of lensless cameras, signal reconstruction is necessary. The image reconstruction methods in lensless cameras can be roughly divided into two types. The first is the traditional method, which is model-based. The second is the learning-based method, which is data-driven.

Model-based method

In the model-based methods, the PSF is needed in the calculation. There are two types, the first is the deconvolution method[45], which is fast but not robust to noise; the other is the iterative method based on the theory of compressive sensing[36], which is more robust to noise but the regularisation term needs careful selection and the interactive calculation takes time to converge. In addition, the accuracy of the PSF measurement and system calibration affects the quality of image reconstruction in the model-based method.

Learning-based method

Learning-based method can provide reconstruction images with fewer artifacts and higher quality[46]. The PSF measurement is not mandatory in the learning-based method[47]. For this method, the image data and the AI model design are more important. Unlike a lens-based camera, there is no off-the-shelf database for a lensless camera because the image data is different for each mask-based camera, and the different calibration or optical settings cause the data to be unique. For lensless cameras, there are usually two ways of collecting data. The first is to use a beam splitter to capture the same scene with a normal camera and a lensless camera at the same time[48], this requires careful calibration, the other is to do the simulation based on the imaging forward model, which is not accurate[49]. Typically, the AI model used in a lensless camera is a CNN-based architecture[50, 51], [52, 53] argues that the CNN architecture focuses on local features, but the lensless camera imaging system is more suitable for global feature extraction, so the Transforemr-based [54]architecture is

more appropriate. The inherent problem of AI-based image reconstruction is that it is not explainable. So, [48] proposes an unrolling ADMM with learning based together to use AI to optimize the parameters in traditional methods, which can provide the explainable high-quality result.

2.2.3 Possible applications

Due to their low cost, lensless cameras can replace normal cameras in some cases, such as infrared monitors[55] and depth cameras[37]. For wide-field imaging, a fisheye camera with optical distortion calibration is usually used, while a lensless camera can be used for ultra-widefield imaging without distortion[56]. Also, due to the thin structure, lensless cameras can be used in IoT devices for vision tasks such as text recognition[57], face detection[58], and QR code detection[59].

2.3 HOE-based imaging system

2.3.1 Holographic Optical Elements

Cameras are increasingly being integrated into IoT devices to enable various multimedia applications, often in combination with other functions like displays. However, the traditional approaches to integrating cameras in IoT devices often involve the use of multiple optical elements. This approach introduces complexity, adds weight to the system, requires calibration, and increases the overall cost. Moreover, each optical element typically serves a different function and operates independently, further complicating the system. Additionally, the field of view (FOV) and depth of field (DOF) limitations of traditional camera systems, caused by the use of optical lenses, may result in an unsatisfactory user experience or restrict the overall flexibility of the device. To address these challenges, alternative approaches are needed to provide compact, lightweight, and versatile camera systems for IoT devices.

The Holographic Optical Elements (HOEs) offer an effective solution to these challenges by leveraging holography technology. Holography involves recording the complete wavefront of interfered coherent beams onto a medium, which can then be used to reproduce the original wavefield (Fig.2.3) [60, 61].

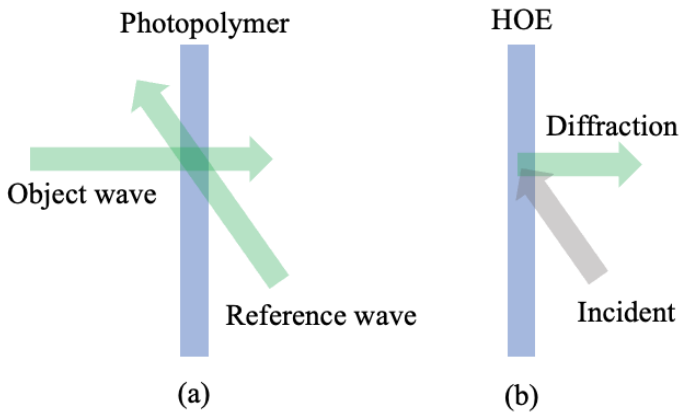


Fig. 2.3 (a) Recording process of a HOE; (b) Reconstruction process.

HOEs are produced by recording the interference pattern between reference and object beams on a photosensitive material. The process typically includes: illuminating a photosensitive polymer, with a reference beam and an object beam. The object beam carries the desired optical function to be encoded in the hologram. When the reference and object beams interfere with each other in the photosensitive material, they create an interference pattern. This interference pattern consists of variations in light intensity and phase that represent the desired optical function. The photosensitive material captures and stores this pattern, essentially recording the wavefront information. The recorded interference pattern acts as a diffraction grating capable of diffracting incident light according to the stored information. [62–64].

In general, HOEs can be classified into two types based on the selected parameters during the manufacturing process: transmission type and reflection type[65](Fig.2.4). By carefully designing the interference pattern and controlling the imaging process parameters, such as the angle and intensity of the reference and object beams, the photosensitive material can be customized to create HOEs that replicate the functions of conventional optical components, such as mirror and lenses, through the diffraction of light.[66]

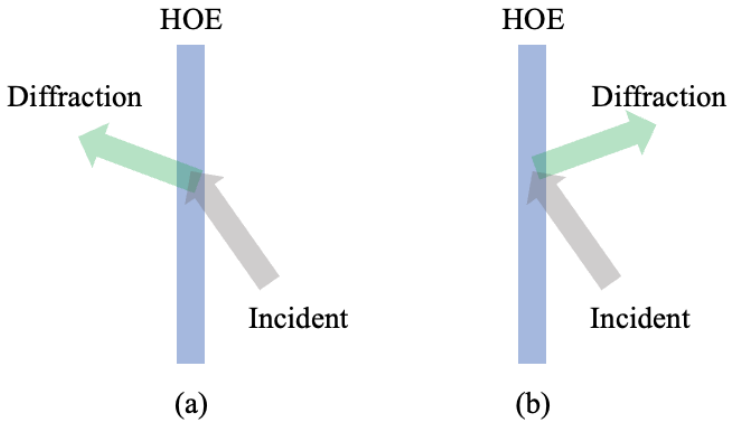


Fig. 2.4 HOE of (a) Transmission type; (b) Reconstruction type.

2.3.2 Applications

Holographic optical elements (HOEs) offer a versatile and compact solution for various optical functions. They are thin and see-through films that can simultaneously perform multiple optical functions based on their selective angular and wavelength characteristics[67]. HOEs have found extensive applications in visual display systems. In head-mounted displays (HMDs)[68–71], HOEs can be used to create immersive and realistic visual experiences. Similarly, in near-eye displays(Fig.2.5(b)[72] and Fig.2.5(c)[73]), HOEs can enhance the viewing experience by enabling compact and lightweight optical systems that provide high-resolution images and comfortable viewing angles. The function of three-dimensional (3D) displays[74, 75] can also be provided by HOE. In addition to display systems, HOEs can also be employed in image-capturing optical systems[76]. In thin illumination devices[77, 78], HOEs provides an efficient and compact solution for illuminating, the light can be directed and controlled to achieve desired illumination patterns or effects. This is particularly useful in applications such as signage, and automotive lighting. Augmented reality (AR) displays[79] benefit from the unique capabilities of HWDs, which can capture and guide light based on a transparent film, enabling the overlay of digital information onto the real-world view. By integrating HOE, AR displays can achieve high-quality image projection with wide field-of-view and minimal distortion. Eye-gaze detection systems utilize HOEs to accurately track a person's eye movements and determine their gaze direction[80]. So that the display can selectively redirect light toward the user. This technology is particularly valuable in

applications such as human-computer interaction and virtual reality. Furthermore, HOEs are being considered for heads-up display (HUD) systems[81–83] applications. HUDs project relevant information, such as navigation instructions or vehicle status, onto a transparent surface in the driver's field of view(Fig.2.5(a)[84]). HOEs offer advantages in HUD systems by providing a compact and lightweight solution with high optical efficiency, enabling clear and readable information without obstructing the driver's view.

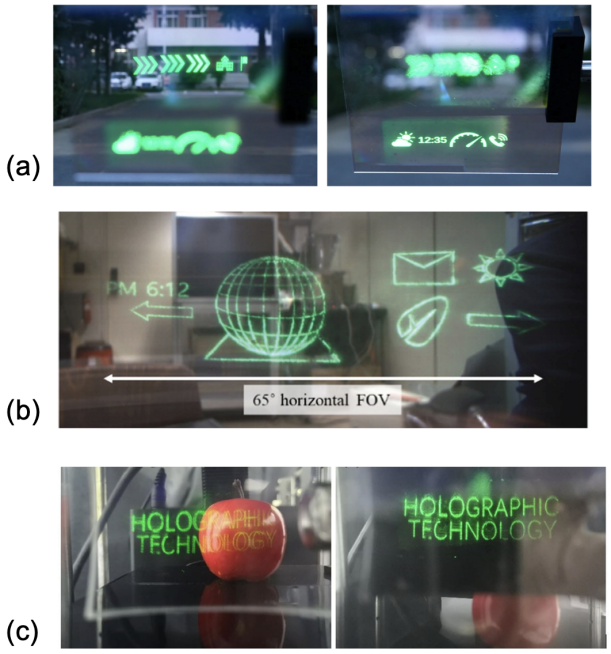


Fig. 2.5 Applications of HOE in display systems. Source: (a) from [84], (b) from [72], and (c) from [73].

2.4 Summary

This chapter provides an introduction to computational imaging. Computational imaging is an emerging field of research in which optical systems and signal processing are jointly optimized to achieve specific imaging functions and characteristics. It is not a simple complement to optical imaging and digital image processing, but an organic combination of optical modulation at the front end (physical domain) and information processing at the back end (digital domain) through optical coding and mathematical modeling of illumination and imaging systems to obtain images and information in a computationally reconstructed

manner. This new imaging method is expected to overcome the limitations of traditional optical imaging technology in terms of operating conditions, power consumption, and cost, and enable the acquisition of images and information in terms of function, and performance.

Since my research is about a thin and functional camera system, to break the limitations caused by the conventional camera, especially caused by the optical lens, I choose lensless imaging and HOE-based imaging, these two computational imaging technology to do the research. Lensless imaging eliminates the need for traditional optical lenses and instead relies on computational algorithms to reconstruct images. This approach offers advantages such as compactness, simplicity, and the ability to record 3D information. Consider the need to integrate the camera into the IoT device, especially with the display function. I suggest using HOE-based technology. It replicates the functions of conventional optical components like mirrors and lenses using a thin and transparent film. A brief introduction on lensless imaging and HOE-based imaging is presented in the subsections.

Chapter 3

Superresolution in FZA Lensless Camera

3.1 Related work

Among the various mask patterns available for lensless cameras, I choose the Fresnel Zone Aperture (FZA) as the coded mask for my lensless camera, named the FZA-lensless camera. The FZA pattern was first proposed by Mertz and Young[85, 86] in their research on incoherent coded imaging, where they extended the concept and application of holography[87]. Recently, Hitachi has made advancements in lensless camera research by employing the FZA pattern for the first time at visible wavelengths (shown in Fig.3.1(b)). Nakamura[88]conducted an analysis of the entire imaging process of the FZA imaging system from the perspective of wave optics. Through this analysis, it was discovered that diffraction in the system leads to null-frequency values in the Modulation Transfer Function (MTF), which consequently degrades the image resolution. To address this issue and improve resolution, Nakamura proposed an image synthesis method based on wavelength. While this method does increase resolution, it is important to note that it can have an impact on color quality. The synthesis method may introduce color artifacts or distortions due to wavelength-dependent processing.

Hitachi's research takes advantage of the FZA pattern, which is suitable for mathematical modeling, and introduces a fast image reconstruction method based on moire fringes[43, 89, 90]. However, it is important to note that their forward model is based on geometric optics, while in real situations, diffraction occurs as light passes through the mask. This discrepancy can affect the accuracy of the reconstructed image.

Another group from Tsinghua University[91] (shown in Fig.3.1(c)) proposes a method based on compressed sensing for lensless FZA camera. This method performs well when appropriate hyperparameters and regularization terms are chosen. However, it has certain limitations. Firstly, the compressed sensing-based method is not exclusively designed for FZA lensless cameras, which means it doesn't fully exploit the advantages offered by the FZA pattern. Secondly, the compressed sensing-based reconstruction process can be time-consuming, making it less suitable for practical applications.

Additionally, some researchers have explored the use of deep learning-based methods for image reconstruction in FZA lensless cameras[92, 49]. These approaches have shown promise in achieving reconstructed images with fewer artifacts. However, it is worth noting that the current work in this area is primarily based on simulated databases, which may not fully represent real-world scenarios. The mismatch between simulated and actual measurements can lead to incorrect mapping from observed measurements to the desired image.

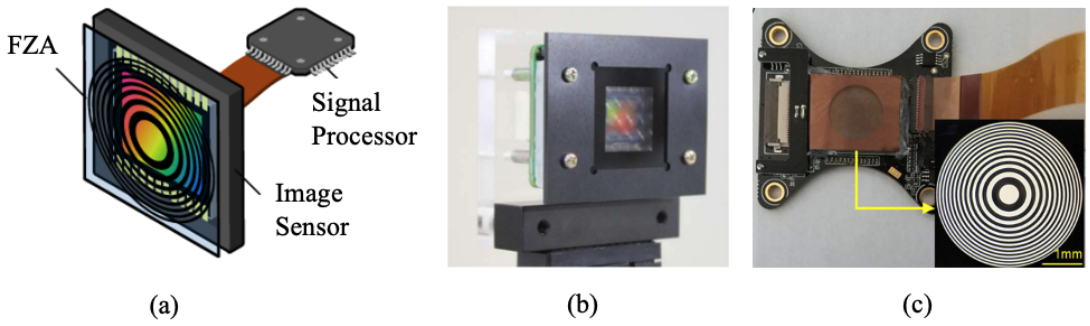


Fig. 3.1 (a) Scheme of FZA lensless camera [43]; prototype FZA lensless camera from (b) Hitachi [90] and (c) Tsinghua University [91].

3.1.1 Geometrical optics based conventional method with Fringe Scanning (FS)

Geometrical optics-based imaging analysis

The intensity transmittance T of the Fresnel Zone Aperture (FZA) mask can be mathematically modeled as follows:

$$T(x_{\mathrm{p}}, y_{\mathrm{p}}; \beta, \varphi) = \frac{1}{2}\left[1 + \cos\left\{\beta\left({x_{\mathrm{p}}}^2 + {y_{\mathrm{p}}}^2\right) + \varphi\right\}\right], \tag{3.1}$$

in the model, the Cartesian coordinates on the FZA plane are denoted as $(x_{\mathrm{p}}, y_{\mathrm{p}})$. The parameter which determines the pitch size of FZA mask, is controlled by β. The initial phase of the mask is represented by φ.

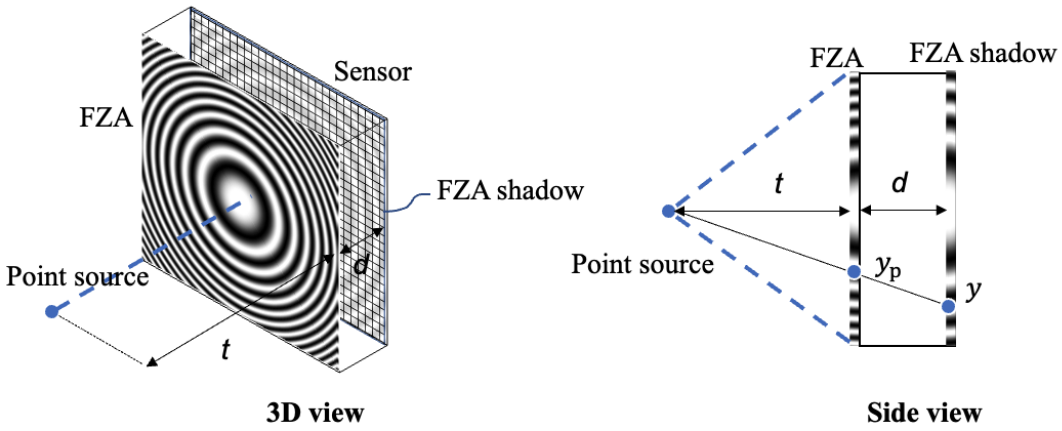


Fig. 3.2 Geometrical imaging system of FZA lensless camera with point light source.

As illustrated in Fig.3.2, the setup of the FZA lensless camera involves a point light source located at a distance t in front of the camera along the axis that crosses through the center of the FZA mask. The FZA mask is positioned in front of the image sensor at a distance d. The point light source serves as the object being imaged by the FZA lensless

camera. It emits light rays that propagate toward the FZA mask and subsequently toward the image sensor. The captured pattern, denoted as h, on the image sensor of the FZA lensless camera, can be considered as the impulse response or the PSF of the imaging system. In the geometrical optics model, the magnification coefficient, denoted as M, represents the ratio of the size of the image formed by the lensless camera system to the size of the object being imaged. It can be calculated using the following equation:

$$\mathrm{M} = \frac{y}{y_\mathrm{p}} = \frac{t+d}{t}. \tag{3.2}$$

The PSF pattern h, in the FZA lensless camera system, can be modeled as follows:

$$h(x,y;\beta,\varphi) = \frac{1}{2}\left[1+\cos\left\{\beta\left\{\left(\frac{x_\mathrm{p}}{\mathrm{M}}\right)^2+\left(\frac{y_\mathrm{p}}{\mathrm{M}}\right)^2\right\}+\varphi\right\}\right], \tag{3.3}$$

to simplify the function, let's consider,

$$x = \frac{x_\mathrm{p}}{\mathrm{M}},\ y = \frac{y_\mathrm{p}}{\mathrm{M}}, \tag{3.4}$$

where (x,y) are the coordinates on the sensor plane, and,

$$h(x,y;\beta,\varphi) = \frac{1}{2}\left[1+\cos\left\{\beta(x^2+y^2\right\}+\varphi)\right]. \tag{3.5}$$

Let's assume the object space is represented by coordinates (m,n), which correspond to a flat plane parallel to the FZA mask and located at a distance t from the lensless camera. The intensity distribution of a two-dimensional target on the object plane is denoted as $f(m,n)$. So that the magnified intensity pattern of the object on the sensor, $f_d(x',y')$, can be modeled as:

$$f_d(x',y') = f(\frac{t}{d}m, \frac{t}{d}n), \tag{3.6}$$

the subscript d indicates the mask-sensor distance.

Considering that the imaging process is shift-invariant, the captured image on the sensor, denoted as $g(x,y)$, can be expressed as a 2D convolution between the PSF and the intensity

distribution of the object on the sensor. Thus, the captured image can be written as:

$$g(x,y;\beta,\varphi) = \iint \frac{1}{2}\left[1+\cos\left\{\beta\left(\left(x-x^{'}\right)^2+\left(y-y^{'}\right)^2\right)+\varphi\right\}\right]\cdot f_d\left(x',y'\right)\mathrm{d}x'\mathrm{d}y', \tag{3.7}$$

where $\cdot$ means multiply operation and the above function could be simplified as:

$$g(x,y;\beta,\varphi) = h(x,y;\beta,\varphi) * f_d(x,y), \tag{3.8}$$

where $*$ is the 2D convolution operator.

Based on the Convolution Theorem, which states that convolution in the spatial domain is equivalent to multiplication in the frequency domain, I can analyze the imaging process conveniently by calculating the Fourier transform of the spatial convolutional process. Let's denote the Fourier transform of the captured image, $g(x,y)$, as $G(u,v)$, where (u,v) represents the spatial frequencies in the frequency domain. Similarly, I denote the Fourier transform of the PSF, $h(x,y)$, as $H(u,v)$, and the Fourier transform of the object's intensity distribution on the sensor, $f_d(x,y)$, as $F_d(u,v)$. So that the Eq.3.8 can be rewritten in the frequency domain as:

$$G(u,v;\beta,\varphi) = H(u,v;\beta,\varphi)\cdot F_d(u,v). \tag{3.9}$$

Based on Euler's theorem and integral of exponential function, which states,

$$e^{j\theta} = \cos(\theta) + j\sin(\theta) \tag{3.10}$$

$$\int e^{-(ax^2+bx)}dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}} \tag{3.11}$$

the frequency response H can be calculated as,

$$\begin{aligned}
H(u,v;\beta,\varphi) &= \mathscr{F}\{h(x,y)\} \\
&= \mathscr{F}\left\{\frac{1}{2}\left[1+\cos\left\{\beta(x^2+y^2\right\}+\varphi)\right]\right\} \\
&= \frac{1}{2}\mathscr{F}\left\{1+\frac{1}{2}\exp\left\{j(\beta(x^2+y^2)+\varphi)\right\}+\frac{1}{2}\exp\left\{-j(\beta(x^2+y^2)+\varphi)\right\}\right\} \\
&= \pi\delta(u,v)+\frac{1}{4}\mathscr{F}\left[\exp\left\{j(\beta(x^2+y^2)+\varphi)\right\}+\exp\left\{-j(\beta(x^2+y^2)+\varphi)\right\}\right] \\
&= \pi\delta(u,v)+\frac{1}{4}\left[\iint\exp\left\{j(\beta(x^2+y^2)+\varphi)\right\}\exp(-2j\pi(ux+vy))d\mathrm{x}d\mathrm{y}\right. \\
&\left.+\iint\exp\left\{-j(\beta(x^2+y^2)+\varphi)\right\}\exp(-2j\pi(ux+vy))d\mathrm{x}d\mathrm{y}\right] \\
&= \pi\delta(u,v)+\frac{\pi}{2\beta}\sin\left\{\frac{\pi^2}{\beta}(u^2+v^2)-\varphi\right\}.
\end{aligned} \tag{3.12}$$

Thus, in the imaging process of a single FZA lensless camera, the spatial domain model is represented by Eq.3.8, while the frequency domain model is described by Eq.3.9. These models are based on the geometrical optics approximation. Additionally, the frequency response, denoted as H, is expressed in Eq.3.12.

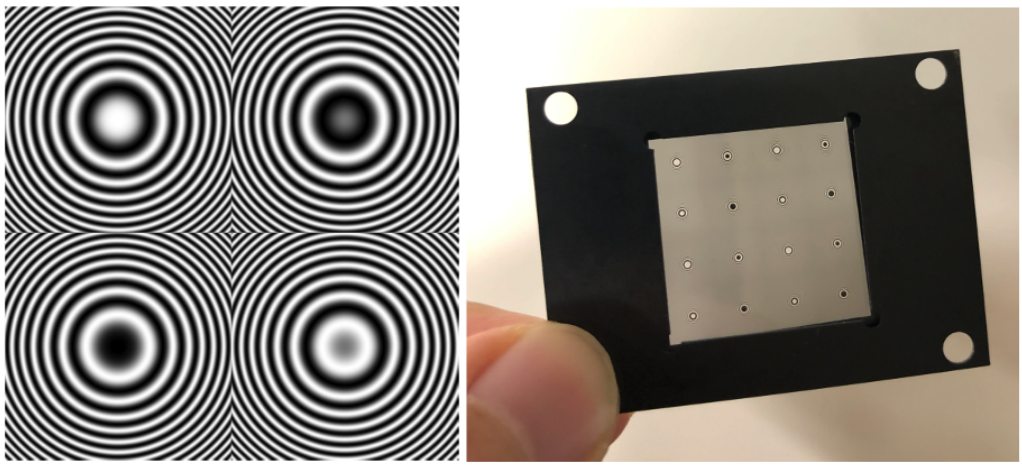
FS-based image reconstruction method

By bringing the calculation formula of H into Eq.3.9, the imaging process without considering noise in the frequency domain can be described as:

$$G(u,v;\beta,\varphi)=\left[\pi\delta(u,v)+\frac{\pi}{2\beta}\sin\left\{\frac{\pi^2}{\beta}(u^2+v^2)-\varphi\right\}\right]\cdot F_d(u,v). \tag{3.13}$$

A reconstructed estimate $\hat{F}_d(u,v)$ is expected from the recovery process using fast deconvolution since the frequency response is already known. However, there are certain factors that can degrade the quality of the reconstruction. The presence of a strong Direct Current (DC) component and several zero-crossings in the frequency response. The DC component arises due to the lensless imaging principle, where the captured image is an incoherent

superposition of multiple coded intensity patterns. This principle leads to a large bias in the captured image, resulting in a strong DC component in the frequency response. The zero-crossings in the frequency response are a consequence of the special design of the FZA mask itself. These zero-crossings can amplify noise and introduce artifacts, leading to a degradation in the reconstruction quality if directly used in a deconvolution filter.



Simulation mask pattern Picture of real mask

Fig. 3.3 Mask synthesize FZAs with different initial phases.

To address these issues in FZA lensless camera, an additional pre-processing called Fringe Scanning (FS) or phase shifting is proposed by [93, 43]. In order to remove the DC component and compensate for the loss of information at the zero points, four FZA with different initial phases are synthesized ($(\varphi = 0, \frac{1}{2}\pi, \frac{3}{2}\pi, 2\pi)$) as a mask distributed over the four different areas (shown in Fig.3.3), then the image captured on the sensor is divided into four areas according to the different masks and synthesized together using the follows:

$$\begin{aligned} H_{\mathrm{FS}}(u,v;\beta) &= j\left[H(u,v;\beta,0) - H(u,v;\beta,\pi)\right] + \left[H\left(u,v;\beta,\frac{3}{2}\pi\right) - H\left(u,v;\beta,\frac{1}{2}\pi\right)\right] \\ &= \exp\left(j\frac{\pi^2}{\beta}(u^2+v^2)\right). \end{aligned} \tag{3.14}$$

H_{FS} is the new synthesized frequency response which, without the DC component and zero crossing, maintains a constant amplitude value of the OTF in the multi-FZA lensless camera. So that the new imaging process in the frequency domain can be modeled as follows:

$$G_{\mathrm{FS}}(u,v) = H_{\mathrm{FS}}(u,v;\beta) \cdot F_d(u,v), \tag{3.15}$$

In this situation, a direct inverse filter,$H_{\mathrm{FS}}^{\mathrm{inv}}$, can be used for deconvolution reconstruction, as:

$$\begin{aligned} &\hat{F}_d(u,v) = G_{\mathrm{FS}}(u,v) \cdot H_{\mathrm{FS}}^{\mathrm{inv}}, \\ &s.t.\ H_{\mathrm{FS}}^{\mathrm{inv}} = \exp\left(-j\frac{\pi^2}{\beta}(u^2+v^2)\right), \end{aligned} \tag{3.16}$$

inverse Fourier transform $\mathscr{F}^{-1}$ can be applied on it to obtain the reconstruction image $\hat{f}_d$ in the spatial domain. Mathematically, it can be expressed as follows:

$$\hat{f}_d(x,y) = \mathscr{F}^{-1}\left\{\hat{F}_d(u,v)\right\}. \tag{3.17}$$

The flowchart of the imaging and processing in the conventional geometrical optics-based method is shown in Fig.3.4.

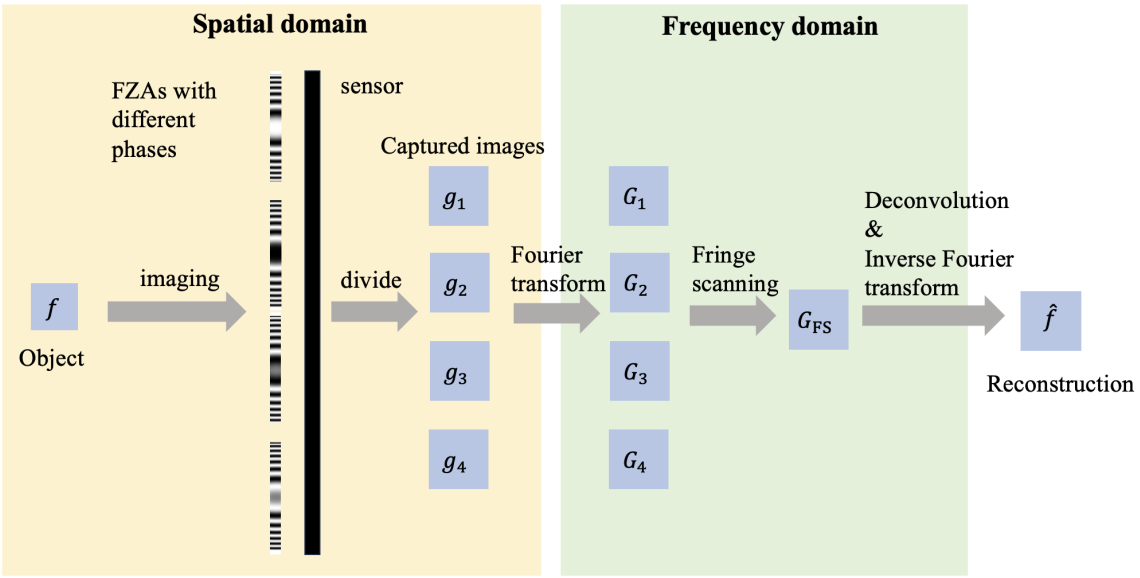


Fig. 3.4 Flowchart of the conventional method which based on the geometrical optics model.

3.2 Wave-optics based imaging process

The FZA lensless camera boasts a rapid deconvolution reconstruction, attributed to its mathematically modelable pattern. Thus, the precision of the PSF becomes crucial. Yet, real-world diffraction introduces discrepancies between the designed forward model and the actual forward process. This misalignment impacts the quality and accuracy of model-based image reconstructions for FZA lensless cameras, leading to artifacts and skewed estimates in the restored images. In a study by [88] utilizing the FS method, it was revealed that diffraction creates zero points in the modulation transfer function (MTF) of the synthesized frequency response in multi-FZA lensless cameras. During deconvolution, this synthesized frequency response acts as an inverse filter to regenerate the original image. Zero points in the response, however, serve as null frequencies, muting certain spatial frequencies tied to object details. This research affirmed that these zeros are influenced by the light's wavelength and the gap between the FZA mask and the image sensor. It designated the first zero's frequency as the resolution boundary. This insight suggests that measurements across varying wavelengths can be amalgamated to offset information loss at zero points and to push the first zero point to a higher frequency slot. As a result, a wavelength-based synthesis approach is proposed for superresolution in FZA lensless cameras, though it might compromise the color integrity of the reconstructed image.

In this segment, I dissect the diffraction propagation inherent to a single-FZA lensless camera system, employing the angle-spectrum propagation framework. Delving into the diffraction propagation offers pivotal insights into the origins and nature of the zeros. Such scrutiny elucidates the genesis of zero points within the synthesized frequency response.

To calculate the wavefront $U(u,v;\beta,\varphi)$ on the sensor plane after diffraction propagation, here the magnification coefficient is omitted for simplicity, the given calculation function is:

$$U(u,v;\beta,\varphi) = \mathscr{F}^{-1}\left\{\mathscr{F}\left\{T(x_{\mathrm{p}},y_{\mathrm{p}};\beta,\varphi)\right\}\cdot H_{\mathrm{F}}(u,v)\right\}, \tag{3.18}$$

$H_{\mathrm{F}}(u,v)$in this section represents the Fresnel-region diffraction term, which accounts for the diffraction effects in the imaging process, shown as:

$$H_{\mathrm{F}}(u,v) = e^{j\frac{2\pi}{\lambda}d}\cdot\exp\left\{-j\pi\lambda d(u^2+v^2)\right\}, \tag{3.19}$$

where λ is the wavelength and d is the distance between FZA mask and sensor. Based on the information provided, and the Fourier transform of the FZA pattern given by Eq.3.12. Using

this information, I can achieve the following expression for the wavefront on the sensor plane:

$$\begin{aligned}\mathscr{F}\{U(u,v;\beta,\varphi)\} &= \mathscr{F}\left\{T(x_{\mathrm{p}},y_{\mathrm{p}};\beta,\varphi)\right\}\cdot H(u,v)\\ &= \left[\pi\delta(u,v)+\frac{\pi}{2\beta}\sin\left\{\frac{\pi^2}{\beta}(u^2+v^2)-\varphi\right\}\right]\\ &\cdot e^{\left(j\frac{2\pi}{\lambda}d\right)}\exp\left\{-j\pi\lambda d\left(u^2+v^2\right)\right\}.\end{aligned} \tag{3.20}$$

Based on Euler's theorem, the sine term can be decomposed into its complex exponential form to give the following equation:

$$\begin{aligned}\mathscr{F}\{U(u,v;\beta,\varphi)\} = &\left[\pi\delta(u,v)+\frac{\pi}{4\beta j}\exp\left\{j\left(\frac{\pi^2}{\beta}\left(u^2+v^2\right)-\varphi\right)\right\}\right.\\ &\left.-\frac{\pi}{4\beta j}\exp\left\{-j\left(\frac{\pi^2}{\beta}\left(u^2+v^2\right)-\varphi\right)\right\}\right]\\ &\cdot e^{\left(j\frac{2\pi}{\lambda}d\right)}\exp\left\{-j\pi\lambda d\left(u^2+v^2\right)\right\}\end{aligned} \tag{3.21}$$

After performing the inverse Fourier transform on the expression, the wavefront on the sensor can be modeled as follows:

$$\begin{aligned}U(u,v;\beta,\varphi) = &\iint\left[\pi\delta(u,v)+\frac{\pi}{4\beta j}\exp\left\{j\left(\frac{\pi^2}{\beta}\left(u^2+v^2\right)-\varphi\right)\right\}\right.\\ &\left.-\frac{\pi}{4\beta j}\exp\left\{-j\left(\frac{\pi^2}{\beta}\left(u^2+v^2\right)-\varphi\right)\right\}\right]\cdot e^{\left(j\frac{2\pi}{\lambda}d\right)}e^{\left\{-j\pi\lambda d\left(u^2+v^2\right)\right\}}\mathrm{d}u\mathrm{d}v\\ &= \exp\left(j\frac{2\pi}{\lambda}d\right)\cdot\left[\frac{1}{2}\right.\\ &+\frac{\pi^2}{4\left(\pi^2-\beta\pi\lambda d\right)}\exp\left(\frac{-j\beta\pi^2\left(x^2+y^2\right)}{\pi^2-\beta\pi\lambda d}-j\varphi\right)\\ &\left.+\frac{\pi^2}{4\left(\pi^2+\beta\pi\lambda d\right)}\exp\left(\frac{-j\beta\pi^2\left(x^2+y^2\right)}{\pi^2-\beta\pi\lambda d}+j\varphi\right)\right].\end{aligned} \tag{3.22}$$

The impulse response $\mathrm{PSF}(x,y;\beta,\varphi)$ can be calculated as:

$$
\begin{aligned}
\mathrm{PSF}(x,y;\beta,\varphi) &= U(u,v;\beta,\varphi)U^{*}(u,v;\beta,\varphi) \\
&= \frac{1}{4} + \frac{\pi^4}{16(\pi^2-\pi\lambda\beta d)^2} + \frac{\pi^4}{16(\pi^2+\pi\lambda\beta d)^2} \\
&+ \cos\left(\frac{2\pi}{\lambda}d\right)\cdot\left[\frac{\pi^2}{4(\pi^2-\pi\lambda\beta d)}\cos\left(\frac{-\beta\pi^2(x^2+y^2)}{4(\pi^2-\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d-\varphi\right)\right. \\
&\left.+\frac{\pi^2}{4(\pi^2+\pi\lambda\beta d)}\cos\left(\frac{\beta\pi^2(x^2+y^2)}{4(\pi^2+\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d+\varphi\right)\right] \\
&+ \sin\left(\frac{2\pi}{\lambda}d\right)\left[\frac{\pi^2}{4(\pi^2-\pi\lambda\beta d)}\sin\left(\frac{-\beta\pi^2(x^2+y^2)}{4(\pi^2-\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d-\varphi\right)\right. \\
&\left.+\frac{\pi^2}{4(\pi^2+\pi\lambda\beta d)}\sin\left(\frac{\beta\pi^2(x^2+y^2)}{4(\pi^2+\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d+\varphi\right)\right] \\
&+ \frac{\pi^4}{8(\pi^4-(\pi\lambda\beta d)^2)}\cos\left(\frac{-\beta\pi^2(x^2+y^2)}{4(\pi^2-\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d-\varphi\right) \\
&\cdot\cos\left(\frac{\beta\pi^2(x^2+y^2)}{4(\pi^2+\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d+\varphi\right) \\
&+ \frac{\pi^4}{8(\pi^4-(\pi\lambda\beta d)^2)}\sin\left(\frac{-\beta\pi^2(x^2+y^2)}{4(\pi^2-\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d-\varphi\right) \\
&\cdot\sin\left(\frac{\beta\pi^2(x^2+y^2)}{4(\pi^2+\pi\lambda\beta d)}+\frac{2\pi}{\lambda}d+\varphi\right),
\end{aligned}
\tag{3.23}
$$

where $\cdot^*$ is the complex conjugate. With the approximation $\frac{\lambda\beta d}{\pi} \approx 0$, and use the replacement that $T_1 = \pi^2 - \pi\lambda\beta d$ and $T_2 = \pi^2 + \pi\lambda\beta d$, and angle sum identity, we have:

$$
\begin{aligned}
\mathrm{PSF}(x,y;\beta,\varphi) = &\frac{3}{8} \\
&+ \frac{1}{4}\cos\left(\frac{2\pi}{\lambda}d\right)\cdot\left[\cos\left(\frac{-\beta\pi^2(x^2+y^2)}{4T_1}+\frac{2\pi}{\lambda}d-\varphi\right)\right. \\
&\left.+\cos\left(\frac{\beta\pi^2(x^2+y^2)}{4T_2}+\frac{2\pi}{\lambda}d+\varphi\right)\right] \\
&+ \frac{1}{4}\sin\left(\frac{2\pi}{\lambda}d\right)\left[\sin\left(\frac{-\beta\pi^2(x^2+y^2)}{4T_1}+\frac{2\pi}{\lambda}d-\varphi\right)\right. \\
&\left.+\sin\left(\frac{\beta\pi^2(x^2+y^2)}{4T_2}+\frac{2\pi}{\lambda}d+\varphi\right)\right] \\
&+ \frac{1}{8}\cos\left(\frac{-\beta\pi^2(x^2+y^2)}{4T_1}+\frac{2\pi}{\lambda}d-\varphi\right)\cdot\cos\left(\frac{\beta\pi^2(x^2+y^2)}{4T_2}+\frac{2\pi}{\lambda}d+\varphi\right) \\
&+ \frac{1}{8}\sin\left(\frac{-\beta\pi^2(x^2+y^2)}{4T_1}+\frac{2\pi}{\lambda}d-\varphi\right)\cdot\sin\left(\frac{\beta\pi^2(x^2+y^2)}{4T_2}+\frac{2\pi}{\lambda}d+\varphi\right) \\
&= \frac{3}{8} \\
&+ \frac{1}{2}\cos\left(\frac{2\pi}{\lambda}d\right)\cdot\cos\left(\frac{-\beta^2\lambda d(x^2+y^2)}{\pi}+\frac{2\pi}{\lambda}d\right)\cdot\cos(\beta(x^2+y^2)+\varphi) \\
&+ \frac{1}{2}\sin\left(\frac{2\pi}{\lambda}d\right)\cdot\sin\left(\frac{-\beta^2\lambda d(x^2+y^2)}{\pi}+\frac{2\pi}{\lambda}d\right)\cdot\cos(\beta(x^2+y^2)+\varphi) \\
&+ \frac{1}{16}\left[\cos\left(\frac{-2\beta^2\lambda d(x^2+y^2)}{\pi}+\frac{4\pi}{\lambda}d\right)+\cos(2\beta(x^2+y^2)+2\varphi)\right] \\
&- \frac{1}{16}\left[\cos\left(\frac{-2\beta^2\lambda d(x^2+y^2)}{\pi}+\frac{4\pi}{\lambda}d\right)-\cos(2\beta(x^2+y^2)+2\varphi)\right]
\end{aligned}
\tag{3.24}
$$

After simplification, the formula of PSF is described as

$$\begin{aligned}\mathrm{PSF}(x,y;\beta,\varphi) = &\frac{3}{8}\\ &+\frac{1}{2}\cos(\beta(x^2+y^2)+\varphi)\left[\cos\left(\frac{2\pi}{\lambda}d\right)\cdot\cos\left(\frac{-\beta^2\lambda d(x^2+y^2)}{\pi}+\frac{2\pi}{\lambda}d\right)\right.\\ &\left.+\sin\left(\frac{2\pi}{\lambda}d\right)\cdot\sin\left(\frac{-\beta^2\lambda d(x^2+y^2)}{\pi}+\frac{2\pi}{\lambda}d\right)\right]\\ &+\frac{1}{8}\left[2\cos^2(\beta(x^2+y^2)+\varphi)-1\right],\end{aligned} \tag{3.25}$$

finally, the PSF formula can be expressed as follows:

$$\mathrm{PSF}(x,y;\beta,\varphi) = \frac{1}{4}+\frac{1}{2}\cos\left\{\frac{\beta^2\lambda d}{\pi}\left(x^2+y^2\right)\right\}\cos\left\{\beta\left(x^2+y^2\right)-\varphi\right\}+\frac{1}{4}\cos^2\left\{\beta\left(x^2+y^2\right)-\varphi\right\}. \tag{3.26}$$

The Optical Transfer Function (OTF) of the single FZA imaging system is the Fourier transform of Eq.3.27. Before performing the Fourier transform, let's rearrange the equation, shown as:

$$\begin{aligned}\mathrm{PSF}(x,y;\beta,\varphi) = &\frac{1}{4}+\frac{1}{2}\cos\left\{\frac{\beta^2\lambda d}{\pi}\left(x^2+y^2\right)\right\}\cos\left\{\beta\left(x^2+y^2\right)-\varphi\right\}\\ &+\frac{1}{4}\left[\frac{1}{2}+\frac{1}{2}\cos(2\beta(x^2+y^2)-2\varphi)\right]\\ &=\frac{3}{8}\\ &+\frac{1}{4}\cos\left(\frac{\beta^2\lambda d+\beta\pi}{\pi}(x^2+y^2)-\varphi\right)\\ &+\frac{1}{4}\cos\left(\frac{\beta^2\lambda d-\beta\pi}{\pi}(x^2+y^2)+\varphi\right)\\ &+\frac{1}{8}\cos\left\{2\beta\left(x^2+y^2\right)-2\varphi\right\}.\end{aligned} \tag{3.27}$$

After expanding the equation using Euler's theorem, I obtain:

$$\begin{aligned}\mathrm{PSF}(x,y;\beta,\varphi) = \frac{3}{8} &\\ &+\frac{1}{8}\exp\left(j\frac{\beta^2\lambda d+\beta\pi}{\pi}(x^2+y^2)-j\varphi\right)\\ &+\frac{1}{8}\exp\left(-j\frac{\beta^2\lambda d+\beta\pi}{\pi}(x^2+y^2)+j\varphi\right)\\ &+\frac{1}{8}\exp\left(j\frac{\beta^2\lambda d-\beta\pi}{\pi}(x^2+y^2)+j\varphi\right)\\ &+\frac{1}{8}\exp\left(-j\frac{\beta^2\lambda d+\beta\pi}{\pi}(x^2+y^2)-j\varphi\right)\\ &+\frac{1}{16}\exp\left(j2\beta(x^2+y^2)-2j\varphi\right)\\ &+\frac{1}{16}\exp\left(-j2\beta(x^2+y^2)+2j\varphi\right).\end{aligned} \tag{3.28}$$

So that the OTF can be calculated as the Fourier transform of PSF as:

$$\begin{aligned}\mathrm{OTF}(u,v;\beta,\varphi) = \frac{3}{4}\pi\delta(u,v) &+\frac{j\pi^2}{8\beta(\beta\lambda d+\pi)}\exp\left(\frac{-j\pi^3}{\beta(\beta\lambda d+\pi)}(u^2+v^2)-j\varphi\right)\\ &-\frac{j\pi^2}{8\beta(\beta\lambda d+\pi)}\exp\left(\frac{j\pi^3}{\beta(\beta\lambda d+\pi)}(u^2+v^2)+j\varphi\right)\\ &+\frac{j\pi^2}{8\beta(\beta\lambda d-\pi)}\exp\left(\frac{-j\pi^3}{\beta(\beta\lambda d-\pi)}(u^2+v^2)+j\varphi\right)\\ &-\frac{j\pi^2}{8\beta(\beta\lambda d-\pi)}\exp\left(\frac{j\pi^3}{\beta(\beta\lambda d-\pi)}(u^2+v^2)-j\varphi\right)\\ &+\frac{j\pi}{32\beta}\exp\left(\frac{-j\pi^2}{2\beta}(u^2+v^2)-2j\varphi\right)\\ &-\frac{j\pi}{32\beta}\exp\left(\frac{j\pi^2}{2\beta}(u^2+v^2)+2j\varphi\right).\end{aligned} \tag{3.29}$$

After using approximation of $\frac{\beta\lambda d}{\pi} \approx 0$ and angle sum identity, we can obtain:

$$\begin{aligned}\mathrm{OTF}(u,v;\beta,\varphi) = \frac{3}{4}\pi\delta(u,v) &+\frac{\pi}{2\beta}\sin\left\{\frac{\pi^2}{\beta}\left(u^2+v^2\right)+\varphi\right\}\cos\left\{\lambda d\pi\left(u^2+v^2\right)\right\}\\ &+\frac{1}{16}\frac{\pi}{\beta}\sin\left\{\frac{\pi^2}{2\beta}\left(u^2+v^2\right)+2\varphi\right\}.\end{aligned} \tag{3.30}$$

The derived point spread function (PSF) and optical transfer function (OTF) of a single FZA lensless camera system, are calculated based on wave optics theory. By analyzing the PSF and OTF, we gain a deeper understanding of the imaging characteristics and limitations of the FZA lensless camera. This knowledge can guide us in developing techniques and algorithms to improve the resolution and image quality of the camera system.

3.3 Distance based method

Previous research analyzing the multi-FZA lensless camera system based on FS and wave optics theory has revealed the presence of zero crossings in the frequency response. These zero crossings significantly impact the image resolution and are attributed to the diffraction phenomenon. Understanding the origin and effects of these zero crossings is crucial to developing strategies to mitigate their impact on image resolution. Therefore, I use the frequency response of a single FZA lensless camera with different FZA phases and then apply the FS method to simulate the frequency response of a multi-FZA lensless camera.

Using four different phases ($\varphi = 0, \frac{\pi}{2}, \pi$, and $\frac{3\pi}{2}$) of the OTF (Eq.3.30), then synthesizing them by FS, a new synthesized frequency response taking diffraction into account, $H_{\mathrm{FS-diff}}$, is obtained as follows,

$$\begin{aligned} H_{\mathrm{FS-diff}}(u,v;\lambda,d) &= \left\{ \mathrm{OTF}\left(u,v;\beta,\frac{\pi}{2}\right) - \mathrm{OTF}\left(u,v;\beta,\frac{3}{2}\pi\right) \right\} \\ &\quad + j\left\{\mathrm{OTF}\,(u,v;\beta,0) - \mathrm{OTF}\,(u,v;\beta,\pi)\right\} \\ &= \exp\left(j\frac{\pi^2}{\beta}(u^2+v^2)\right) \cdot \cos(\pi\lambda d(u^2+v^2)). \end{aligned} \tag{3.31}$$

This equation consists of two parts. The first part is the previously derived synthesized frequency response, Eq.3.14, using the phase shifting method based on geometric optics. The second part is a cosine term, which can be inferred to arise from considering the effects of diffraction. This equation has the same form as the formula presented by [88]. It also confirms his observation regarding the zero points in the frequency response and their dependence on wavelength and the distance between the mask and sensor. By comparing Eq.3.14 and Eq.3.31, I observe that the effect of phase shifting is to remove the first term (DC component) and the third term, while transforming the sinusoidal part of the second term into an exponential function. The combined frequency response is then solely dependent on distance and wavelength. As shown in Eq.3.31, the zero-crossing in the frequency response

after using phase shifting is caused by the cosine part. The frequency position of the zero crossing u_{zero} can be modeled as follows, here I only consider the u-axis for instance:

$$u_{\text{zero}} = \sqrt{\frac{1+2n}{2\lambda d}}, n = 0, 1, 2... \tag{3.32}$$

From the aforementioned analysis, with a fixed wavelength λ, the position of the zero-crossing is exclusively determined by the mask-sensor distance d. Consequently, OTFs measured at varying d values exhibit zero-crossings at distinct frequencies. Here, the frequency of the initial zero point dictates the resolution limit. Through the synthesis of these OTFs, we can compensate for the information deficit, producing a refined OTF that places the resolution boundary at an elevated frequency point.

For ease of calculation and representation purposes, Eq.3.31 can be rewritten as follows:

$$\begin{aligned} &H_{\text{FS}-\text{diff}}(u,v;\lambda,d) = H_{\text{FS}}(u,v) \cdot H_{\text{diff}}(u,v;\lambda,d) \\ &s.t.\ H_{\text{diff}}(u,v;\lambda,d) = \cos(\pi\lambda d(u^2+v^2)). \end{aligned} \tag{3.33}$$

To eliminate the zero points without relying on the previous color-channel synthesis technique, my focus is on the mask-sensor distance d. In order to achieve this, I propose an image synthesis method that involves capturing multiple images at different mask-sensor distances[94, 95].

An introduction to the method is given as:

$$\begin{aligned} G_{\text{FS\_1}}(u,v) &= H_{\text{FS}-\text{diff}}(u,v;d_1) \cdot F_d(u,v) + N_1(u,v), \\ G_{\text{FS\_2}}(u,v) &= H_{\text{FS}-\text{diff}}(u,v;d_2) \cdot F_d(u,v) + N_2(u,v). \end{aligned} \tag{3.34}$$

$G_{\text{FS\_1}}$ and $G_{\text{FS\_2}}$ are two captured images in the frequency domain after FS with two mask-sensor distances d_1 and d_2, respectively, assume that $d_1 > d_2$. F_d is the intensity distribution of the original object on the sensor. For the sake of simplicity and ease of explanation, the scaling factor required for image resizing due to varying distances has been neglected in the following. N_1 and N_2 are the 2D Fourier transform of the noise component in each imaging system. Taking into account the presence of noise in the imaging process, I have opted to utilize a least mean square error (LMSE) based method for image reconstruction[96–99]. This approach aims to minimize the average squared difference between the reconstructed image and the original image, effectively reducing the impact of noise on the final result.

Firstly, the inverse filter (Eq.3.16) is applied to remove the exponential function as,

$$\begin{aligned} G^{'}_{\text{FS\_1}}(u,v) &= H_{\text{diff}}(u,v;d_1)\cdot F_d(u,v)+N_1(u,v),\\ G^{'}_{\text{FS\_2}}(u,v) &= H_{\text{diff}}(u,v;d_2)\cdot F_d(u,v)+N_2(u,v).\\ s.t.\ G^{'}_{\text{FS\_1}}(u,v) &= G_{\text{FS\_1}}(u,v)\cdot H^{\text{inv}}_{\text{FS}},\quad G^{'}_{\text{FS\_2}}(u,v) = G_{\text{FS\_2}}(u,v)\cdot H^{\text{inv}}_{\text{FS}}, \end{aligned} \tag{3.35}$$

Assume the synthesized estimation is $\hat{F}_d$, shown as:

$$\hat{F}_d = M_1\cdot G^{'}_{\text{FS\_1}}(u,v)+M_2\cdot G^{'}_{\text{FS\_2}}(u,v), \tag{3.36}$$

where the variables of spatial frequency are omitted for simplicity, M_1 and M_2 are the inverse filters. The estimation minimizes the mean square error given as:

$$e^2 = E\left\{|\hat{F}_d - F_d|^2\right\}, \tag{3.37}$$

The calculation can be transferred to the spatial domain, as,

$$\begin{aligned} e^2 &= E\left\{|\hat{f}_d(x)-f_d(x)|^2\right\}\\ &= E\left\{|\sum_i m_i(x) * g_{\text{FS}_i}(x)^{'} - f_d(x)|^2\right\}, \end{aligned} \tag{3.38}$$

Calculate the deviation of the above function, get,

$$\begin{aligned} \frac{\partial e^2}{\partial m_j(t)} &= 2E\left\{|\sum_i\sum_r m_i(r)g_{\text{FS}_i}(x-r)-f_d(x)|g_{\text{FS}_j}(x-t)\right\}\\ &= 2E\left\{\sum_i\sum_r m_i(r)g_{\text{FS}_i}(x-r)g_{\text{FS}_j}(x-t)\right\} - 2E\left\{f_d(x)g_{\text{FS}_j}(x-t)\right\}\\ &= 0, \end{aligned} \tag{3.39}$$

Therefore, the following formula can be derived,

$$\begin{aligned} E\left\{\sum_i\sum_r m_i(r)g_{\text{FS}_i}(x-r)g_{\text{FS}_j}(x-t)\right\} &= E\left\{f_d(x)g_{\text{FS}_j}(x-t)\right\}\\ \sum_i\sum_r m_i(r)S_{g_{\text{FS}_i},g_{\text{FS}_j}}(t-r) &= S_{f_d,g_{\text{FS}_j}}(t)\\ \sum_i m_i(t) * S_{g_{\text{FS}_i},g_{\text{FS}_j}}(t) &= S_{f_d,g_{\text{FS}_j}}(t), \end{aligned} \tag{3.40}$$

where $S_{f,g}(x)$ means the cross correlation between $f(x)$ and $g(x)$.After applying the Fourier transform to the above formulas, I will obtain the following,

$$\sum_i M_i(u) R_{g_{\mathrm{FS}_i}, g_{\mathrm{FS}_j}}(u) = R_{f_d, g_{\mathrm{FS}_j}}(u), \tag{3.41}$$

where $R_{f,g}(u)$ means the cross spectrum density between $f(x)$ and $g(x)$. Considering the 2 different distance case and Expressing Eq.3.41 in the matrix form, I can obtain,

$$\begin{bmatrix} R_{g_{\mathrm{FS}_1}, g_{\mathrm{FS}_1}} & R_{g_{\mathrm{FS}_2}, g_{\mathrm{FS}_1}} \\ R_{g_{\mathrm{FS}_1}, g_{\mathrm{FS}_2}} & R_{g_{\mathrm{FS}_2}, g_{\mathrm{FS}_2}} \end{bmatrix} \cdot \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} R_{f_d, g_{\mathrm{FS}_1}} \\ R_{f_d, g_{\mathrm{FS}_2}} \end{bmatrix} \tag{3.42}$$

So that we can get,

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} R_{g_{\mathrm{FS}_1}, g_{\mathrm{FS}_1}} & R_{g_{\mathrm{FS}_2}, g_{\mathrm{FS}_1}} \\ R_{g_{\mathrm{FS}_1}, g_{\mathrm{FS}_2}} & R_{g_{\mathrm{FS}_2}, g_{\mathrm{FS}_2}} \end{bmatrix}^{-1} \cdot \begin{bmatrix} R_{f_d, g_{\mathrm{FS}_1}} \\ R_{f_d, g_{\mathrm{FS}_2}} \end{bmatrix} \tag{3.43}$$

Based on the factor that $R_{g_{\mathrm{FS}_i}, g_{\mathrm{FS}_j}} = E\left\{ G_{\mathrm{FS}_i} G^*_{\mathrm{FS}_j} \right\}$ and the noise is independent and has no relates with other terms, I can get the final expression of the inverse filter in the frequency domain as,

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} \frac{H^*_{\mathrm{w}}(d_1)}{|H_{\mathrm{w}}(d_1)|^2 + \frac{\eta_1}{\eta_2}|H_{\mathrm{w}}(d_2)|^2 + \frac{\eta_1}{\eta_f}} \\ \frac{H^*_{\mathrm{w}}(d_2)}{|H_{\mathrm{w}}(d_2)|^2 + \frac{\eta_2}{\eta_1}|H_{\mathrm{w}}(d_1)|^2 + \frac{\eta_2}{\eta_f}} \end{bmatrix} \tag{3.44}$$

where η_f, η_1, and η_2 represent the spectral densities of the original image and the noises in the images captured at distances d_1 and d_2, respectively. By applying the inverse Fourier transform, the final reconstructed image $\hat{f}_d$ can be obtained, as,

$$\hat{f}_d = \mathscr{F}^{-1}\{\hat{F}_d\}. \tag{3.45}$$

The flowchart of the proposed method can be checked in Fig.3.5.

In addition to the aforementioned factors, assessing the noise sensitivity is crucial for any method that employs inverse filtering, as it fundamentally impacts the reliability and quality of the restored image. In this study, I quantify the noise robustness of our method by computing the Mean Squared Error (MSE) e^2 for the reconstructed images. The MSE serves as an indicator of the deviation introduced due to noise and any inaccuracies in the

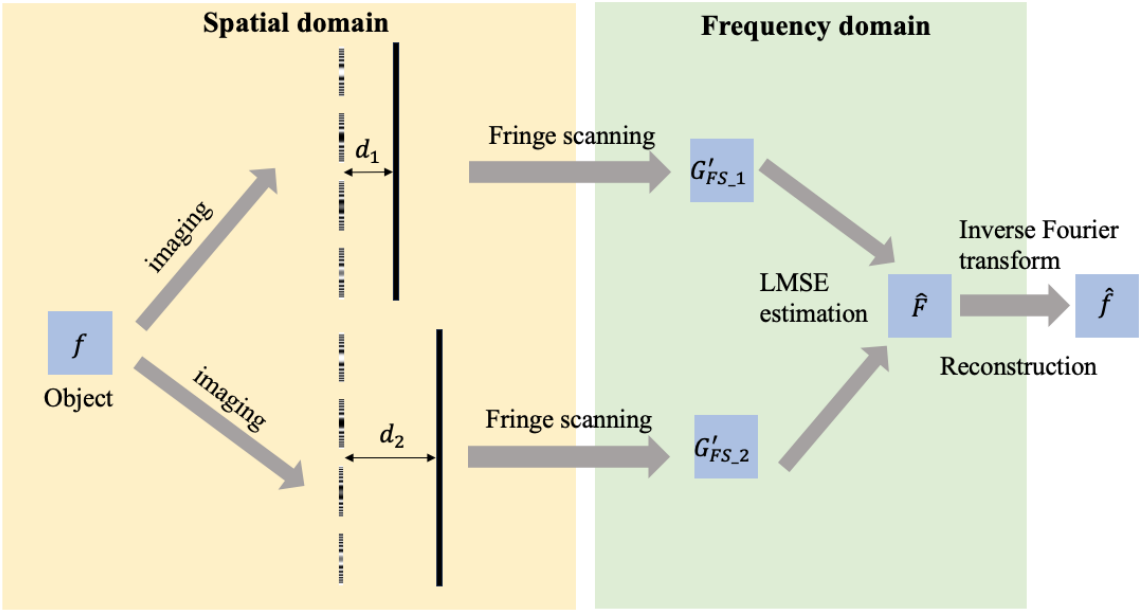


Fig. 3.5 Flowchart of the proposed mask-sensor distance-based method.

reconstruction process. As follows,

$$\begin{aligned} e^2 &= E\left\{|\hat{F}_d - F_d|^2\right\} \\ &= E\left\{|M_1 \cdot G'_{\mathrm{FS\_1}}(u,v) + M_2 \cdot G'_{\mathrm{FS\_2}}(u,v) - F_d|^2\right\} \\ &= E\left\{\left|\frac{-\eta F}{|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2 + \eta}\right|^2\right\} + E\left\{\frac{|H_{\mathrm{w}}(d_1)|^2\sigma^2 + |H_{\mathrm{w}}(d_2)|^2\sigma^2}{(|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2 + \eta)^2}\right\}, \end{aligned} \tag{3.46}$$

where,

$$\eta^2 = |H_{\mathrm{w}}(d_1)|^2 = |H_{\mathrm{w}}(d_2)|^2. \tag{3.47}$$

The right term is the noise part Ns, where,

$$\begin{aligned} Ns &= E\left\{\left|\frac{-\eta F}{|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2 + \eta}\right|^2\right\} + E\left\{\frac{|H_{\mathrm{w}}(d_1)|^2\sigma^2 + |H_{\mathrm{w}}(d_2)|^2\sigma^2}{(|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2 + \eta)^2}\right\} \\ &= E\left\{\frac{\sigma^2}{\left[1 + \frac{\eta}{|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2}\right]\left[|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2 + \eta\right]}\right\}. \end{aligned} \tag{3.48}$$

From the provided equation, the factor $|H_{\mathrm{w}}(d_1)|^2 + |H_{\mathrm{w}}(d_2)|^2$ plays a pivotal role in determining how noise from the captured image is either amplified or suppressed. Noise amplification occurs at frequencies where this factor assumes a lower value. Therefore, to ensure a stable system that's resilient to noise, it's imperative to judiciously select the mask-sensor distance values. This careful selection aims to achieve a flat amplitude distribution, enhancing the system's robustness against noise.

3.4 FZA pitch size based method

In the distance-based method, adjusting the mask-sensor distances introduces several challenges. Firstly, it requires an additional control system to accurately change the distances. Secondly, minor variations in the imaging system can lead to errors in the system, affecting image registration and calibration. These errors reduce the robustness and practicality of the method in real-world applications. Additionally, the distance-based method relies on the Fringe Scanning method, which requires a minimum of 8 shots to gather the necessary measurements for super-resolution processing. This process can be time-consuming and resource-intensive, further limiting its practicality in real-world applications. [100] propose a method that utilizes higher harmonics of a binarized FZA pattern to achieve super-resolution without synthesizing information from changing λ or d. However, one limitation of this method is that the weak higher harmonics signals can have an impact on the signal-to-noise ratio (SNR). This means that the quality of the reconstructed image may be affected by the presence of noise in the higher harmonics.

To address these challenges, I have shifted my focus to the single FZA imaging system, as depicted in Eq.3.30. As shown in the formula, if the wavelength and mask-sensor distance are fixed, the position of the zero-crossing is solely determined by the mask pitch size, value of β. Former methods use the Fringe Scanning (FS) technique, eliminate the DC component in the Optical Transfer Function (OTF) and remove the third term in Eq.3.30. In this method, I use a pair of masks with the same β but different phase φ to eliminate the DC term while retaining the third term. This allows the zero-crossings to depend on the value of β, as

expressed in the following equation:

$$\begin{aligned} H_i &= \mathrm{OTF}\left(\beta_i; 0\pi\right) - \mathrm{OTF}\left(\beta_i; \frac{3}{2}\pi\right) \\ &= \frac{\pi}{2\beta_i}\cos\left(\pi\lambda d\left(u^2+v^2\right)\right)\left[\sin\left(\frac{\pi^2}{\beta_i}\left(u^2+v^2\right)\right)\right. \\ &\left.+\cos\left(\frac{\pi^2}{\beta_i}\left(u^2+v^2\right)\right)\right] + \frac{\pi}{8\beta_i}\sin\left(\frac{\pi^2}{2\beta_i}\left(u^2+v^2\right)\right), \end{aligned} \tag{3.49}$$

where the subscript i represents different β values. Therefore, by synthesizing the images captured from mask patterns with different β values, I can offset the zero-crossings and attain enhanced image quality, positioning the system's initial zero point at a higher frequency.

Similar to before, I consider using measurements from two different β values. Therefore, the two different imaging processes can be modeled as follows:

$$\begin{aligned} G_1(u,v;\beta_1) &= H_1(u,v;\beta_1)\cdot F(u,v) + N_1(u,v), \\ G_2(u,v;\beta_2) &= H_2(u,v;\beta_2)\cdot F(u,v) + N_2(u,v). \end{aligned} \tag{3.50}$$

where $G(u,v;\beta_i)$ represents the captured image in the frequency domain from the pair of masks with β_i after processing. F denotes the original object in the frequency domain, and N_1 and N_2 represent the noise present in each imaging system. To calculate the reconstruction image, I utilize the same LMSE framework as presented in the distance-based method. The framework is depicted below:

$$\hat{F} = P_1\cdot G_1(\beta_1) + P_2\cdot G_2(\beta_2), \tag{3.51}$$

where P_1 and P_2 are the deconvolution filters, as:

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} \frac{H_1^*(\beta_1)}{|H_1(\beta_1)|^2 + \frac{\eta_1}{\eta_2}|H_2(\beta_2)|^2 + \frac{\eta_1}{\eta_f}} \\ \frac{H_2^*(\beta_2)}{|H_2(\beta_2)|^2 + \frac{\eta_2}{\eta_1}|H_1(\beta_1)|^2 + \frac{\eta_2}{\eta_f}} \end{bmatrix} \tag{3.52}$$

where η_f, η_1, and η_2 represent the spectral density of the original image, noise in the images captured with masks of β_1 and β_2 respectively. The reconstruction image in the spatial domain can be expressed as follows:

$$\hat{f}(x,y) = \mathscr{F}^{-1}\left\{\hat{F}(u,v)\right\}. \tag{3.53}$$

The flowchart of the proposed method can be checked in Fig.3.6.

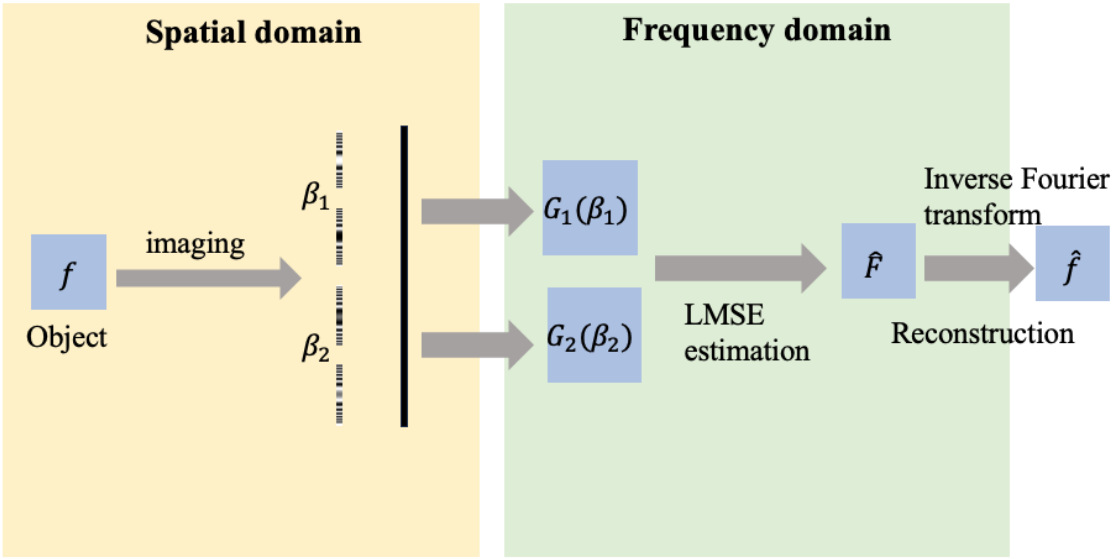


Fig. 3.6 Flowchart of the proposed beta-based method.

3.5 Simulation and Experiments

In this section, the effectiveness of the proposed method is evaluated through computer simulations and optical experiments. The experiments are conducted in various environments, including the use of 2D resolution target chart images displayed on an LCD monitor, single natural 3D objects, and multiple 3D objects. By testing the method in these different scenarios, I can assess its performance and applicability.

3.5.1 Simulation

In the simulation, I design a time-division multi-pattern FZA mask, which allows for the sequential switching of mask patterns. The computational camera setup consists of a stack of FZA masks and a color image sensor. The sensor is placed on a linear translation stage, enabling the adjustment of the mask-sensor distance d. A two-dimensional object is

positioned in front of the mask at a distance t, with the axis passing through their centers perpendicular to the mask plane. The entire system configuration is depicted in Fig.3.7. For the simulation, I assume that the sensor has a resolution of 1024×1024 pixels, with a pixel size of 5.5 μm. The dimensions of an individual FZA mask are 5.63 mm $\times$ 5.63 mm.

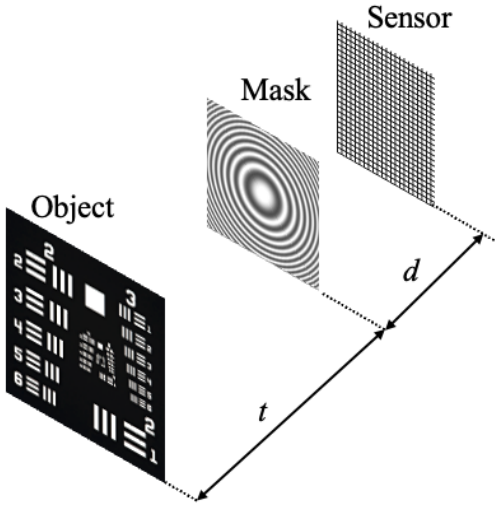


Fig. 3.7 Structure of the simulation system.

To simplify the discussion, I will refer to the mask-sensor distance-based method as Method A. In the simulation, I have chosen two different mask-sensor distances: d_1 = 5.0 mm and d_2 = 6.5 mm. Four masks with the same β = 25 rad/mm^2 but different initial phases ($\varphi = 0, \frac{\pi}{2}, \pi$, and $\frac{3\pi}{2}$) are used to apply the FS technique before implementing method A.

I will also refer to the FZA pitch size (β) based method as Method B. In Method B, I have used four images captured from masks with different β values and synthesized them. In this simulation, the mask-sensor distance d is fixed at 6.5 mm. For Method B, I have utilized two pairs of masks, where each pair shares the same β value but different initial phases ($\varphi = 0$ and $\frac{3\pi}{2}$). Specifically, I have chosen β_1 = 25 rad/mm^2 and β_2 = 21 rad/mm^2 for the two pairs of masks.

Resolution chart target

In the simulation, a resolution target chart is used to evaluate the performance of the proposed method. Specifically, the USAF-1951 resolution chart (Fig.3.8(a)) is used as the 2D test target. The object-mask distance is set to t = 250 mm, and the object size is 180 mm $\times$ 180 mm. To simulate the sensing process, we apply the wave-optics theory, where the input image is convolved with the incoherent point spread function (PSF) calculated from the wave propagation simulation. This process is equivalent to Eq.3.23 and is implemented using MATLAB software. In addition, additive white Gaussian noise is introduced to achieve a signal-to-noise ratio (SNR) of 30 dB.

The reconstruction images are shown in Fig.3.8, and the visually assessed resolution limitation is indicated by the red box in each image.

In Fig.3.8, the reconstruction images (b) and (c) are obtained from the conventional method based on FS and direct deconvolution, using a forward model based on geometric optics theory. The reconstruction performance of the conventional method is evaluated by analyzing the numerical resolution limit. In (b), group 3/element 2 is clear, while in (c), group 3/element 3 is the last clear pattern. These patterns correspond to object plane resolutions of 4.6 mm and 3.2 mm, respectively. The resolution in (c) is slightly higher than that in (b) because the mask-sensor distance in (c) is greater, resulting in the same pattern being sampled with more sensor pixels and achieving a higher resolution.

After applying the proposed method A, the resolution is improved to 1.6 mm, as shown in Fig.3.8(d), where group 4/element 3 is clear. The numerical analysis reveals a twofold improvement in spatial resolution compared to the conventional method. Furthermore, after applying the proposed method B, the resolution is further improved to 0.71 mm, as shown in Fig.3.8(e), where group 5/element 2 is clear. The numerical simulation demonstrates a significant enhancement in the spatial resolution after applying the proposed method.

Overall, both proposed methods (A and B) show substantial improvements in resolution compared to the conventional method, indicating their effectiveness in achieving higher image quality.

To clearly demonstrate the performance improvement of the proposed methods on the imaging system, the Modulation Transfer Function (MTF) line profiles of the different imaging systems are shown in Fig.3.9 and Fig.3.10. The MTF provides a quantitative measure of the imaging system's ability to reproduce spatial details.

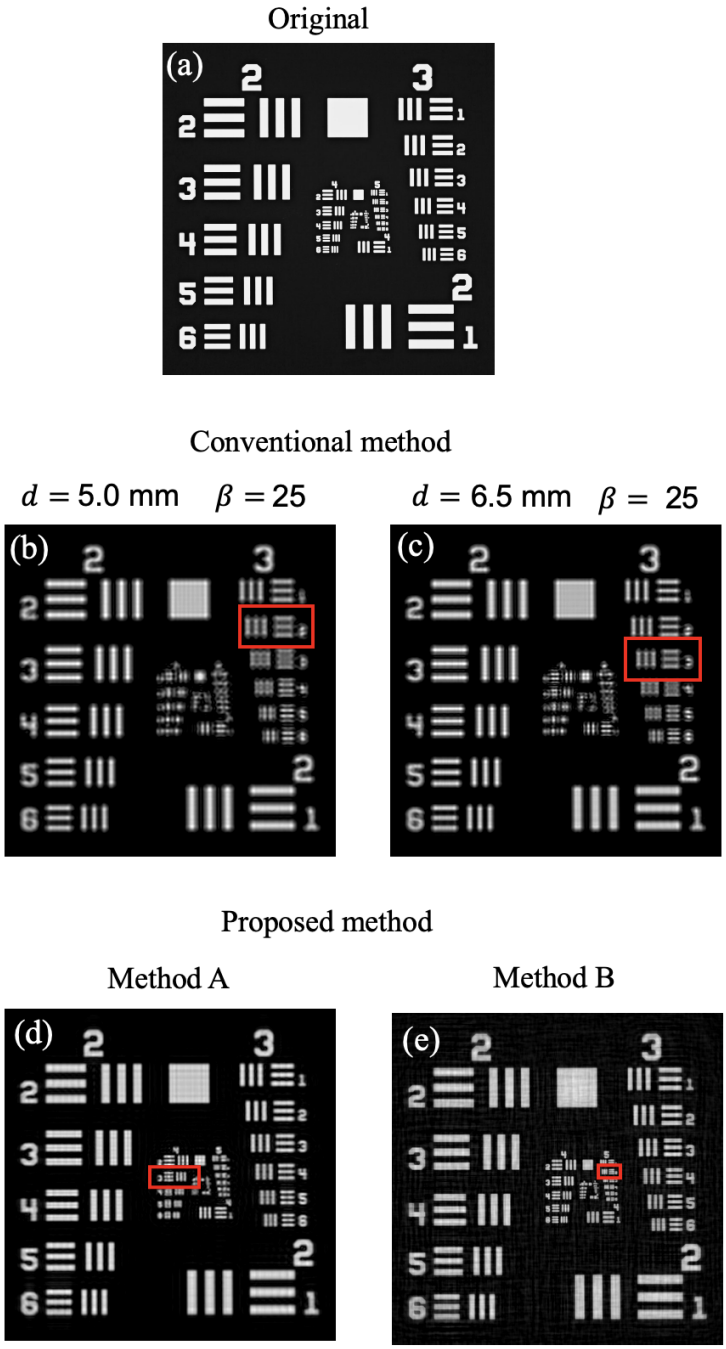


Fig. 3.8 (a) Original image; Reconstruction images from the conventional method at mask sensor distance (b) 5.0 mm and (c) 6.5 mm, both using mask β = 25; Reconstruction image from (d) proposed method A and (e) proposed method B; The red box indicates the resolution limitation.

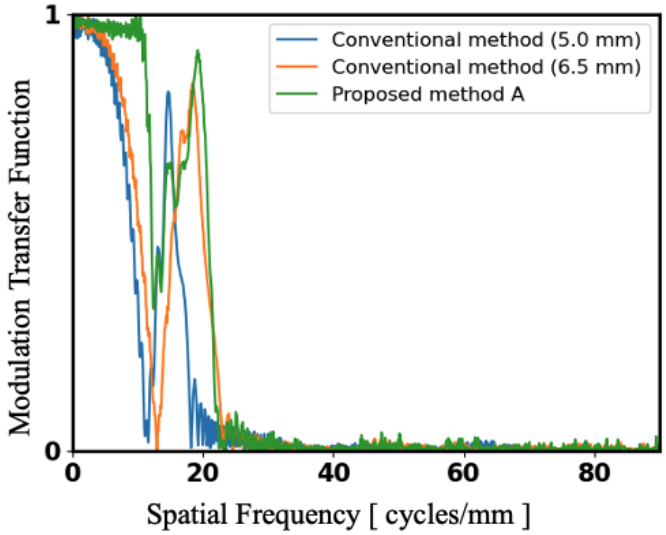


Fig. 3.9 Line profile of the MTF of imaging system using conventional method at mask sensor distance 5.0 mm and 6.5 mm; and the MTF from proposed method A. The spatial frequency is measured on sensor plane up to its Nyquist limitation.

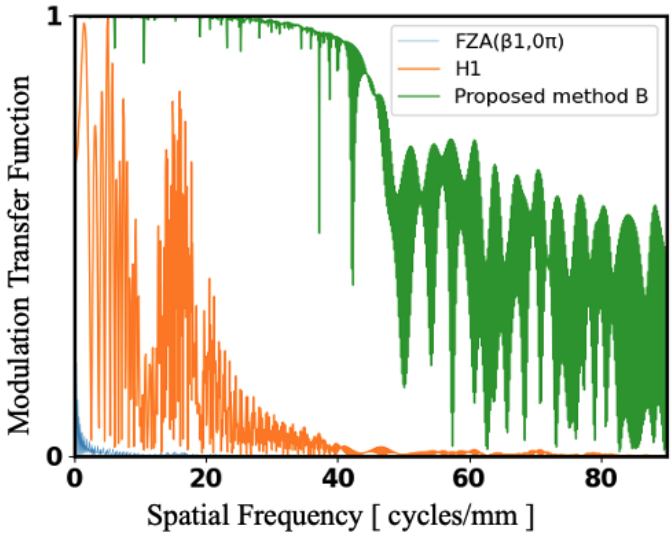


Fig. 3.10 Line profile of the MTF of imaging system using single FZA (β_1,0π); the MTF from synthesized FZA pair H_1; and from the proposed method B. The spatial frequency is measured on sensor plane up to its Nyquist limitation.

In Fig.3.9, The blue line represents the conventional imaging system with a mask-sensor distance of 5.0 mm, the orange line represents the conventional imaging system with a mask-sensor distance of 6.5 mm, and the green line represents the synthesized system of method A. In the Modulation Transfer Function (MTF) analysis, the first zero point of each line is considered as the resolution limitation. From the MTF analysis, we can observe that the resolution limitations, represented by u_{max}, for the conventional methods are 10.8 cycles/mm and 12.4 cycles/mm on the image sensor plane for mask-sensor distances of 5.0 mm and 6.5 mm, respectively. Considering the magnification factor $M = t/d$, where t is the object-mask distance and d is the mask-sensor distance, the resolution limits on the object plane, denoted as L_{obj}, are given by M/u_{max}. For the conventional methods with mask-sensor distances of 5.0 mm and 6.5 mm, the resolution limits on the object plane are approximately 4.63 mm and 3.10 mm, respectively. After applying the proposed method A, the resolution limitation increases to about 22.9 cycles/mm, corresponding to $M/22.9 = 1.68$ mm on the object plane. This demonstrates the effectiveness of the proposed method A in improving the resolution compared to the conventional methods.

In Fig.3.10, the blue line represents the MTF of the imaging system using a single FZA mask ($\beta_1,\ \varphi_1 = 0$), the orange line represents the MTF of the synthesized FZA pair H_1 calculated using Eq.3.49, and the green line shows the MTF of the proposed method B. The MTF of the conventional method is the same as the case with d = 6.5 mm shown in Fig.3.9. The resolution limitation (u_{max}) of the conventional imaging system is 12.4 cycles/mm at the sensor plane. After applying the proposed method B, where the loss of information at zero points is compensated, the resolution limitation has been expanded to 57.5 cycles/mm on the sensor plane. The resolution limits on the image plane can be obtained by M/u_{max}, which gives 3.13 mm and 0.67 mm, respectively. This analysis demonstrates the effectiveness of the proposed method B in significantly improving the resolution capability of the imaging system, allowing for finer details to be captured compared to the conventional methods.

The MTF analysis further supports the findings from the previous image reconstructions, confirming the superior performance of the proposed methods in enhancing the imaging system's resolution and capturing finer details.

RGB image

The color image "Parrot" (Fig.3.11(a)) is used to evaluate the performance of the proposed methods. The parameter settings for this simulation are the same as in the previous simulation, and Gaussian white noise is added to achieve an SNR of 30 dB in the captured image.

Fig.3.11(b) and (c) represent the reconstructed images using the conventional method with mask sensor distances of 5.0 mm and 6.5 mm, respectively. Fig.3.11(d) shows the reconstructed image using method A, which incorporates the synthesis of images captured with different mask-sensor distances. Fig.3.11(e) is the reconstruction from the proposed method B, based on the synthesis of images captured by masks with different pitch sizes. To further evaluate the quality of the reconstruction images, the Peak Signal-to-Noise Ratio (PSNR) is calculated. Additionally, the zoomed area of the red box is provided for visual evaluation.

As seen in (b) and (c), there is a blurred pattern observed in the zoomed area. This indicates a lower resolution due to the mismatch of the forward model used for reconstruction. After applying method A, the zoomed area becomes clearer but still slightly blurred. However, after applying method B, the zoomed area becomes much clearer.

Furthermore, by comparing the PSNR values, a more objective numerical comparison can be made. In Fig.3.11, the PSNR is measured to be 20.3 dB and 20.4 dB in the conventional methods. However, after performing the proposed methods, the PSNR improves to 23.4 dB for method A and 23.8 dB for method B. These higher PSNR values indicate the better performance of the proposed methods in terms of image quality.

3.5.2 Experiment

In the optical experiment, the Holoeye LC 2012 Spatial Light Modulator (SLM) is utilized to display the FZA pattern (256 pixels$\times$256 pixels). The SLM has a resolution of 1024 pixels $\times$ 768 pixels, with a pixel pitch of 36 μm. The SLM is shown in Fig.3.12(a). The image sensor used in the experiment is the CMOSIS CMV4000, which has a resolution of 2048 pixels $\times$ 2048 pixels. The pixel pitch of the image sensor is 5.5 μm. The image sensor is depicted in Fig.3.12(b). To implement Method A, the selection of mask-sensor distance combinations should aim to achieve a high-frequency first zero-crossing in the synthesized system. However, there are physical constraints related to the structure of the Spatial Light

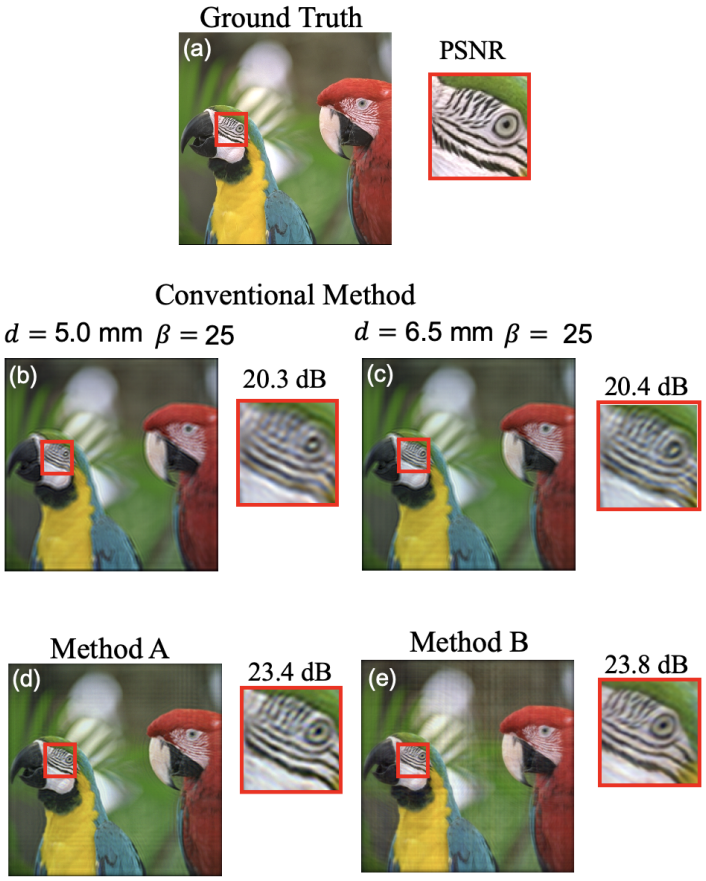


Fig. 3.11 (a) Original image; Reconstruction images from the conventional method at mask sensor distance (b) 5.0 mm and (c) 6.5 mm with $\beta = 25$; Reconstruction image from (d) proposed method A and proposed method B (e); The red box shows the zoom area.

Modulator (SLM) and the polarizer, which impose a minimum mask-sensor distance of 7.0 mm. In order to achieve a thinner camera design, I experimentally selected the combination of mask-sensor distances d equal to 7.0 mm and 7.5 mm for the implementation of Method A, and the β of FZA mask equal to 12 rad/mm^2. To carry out Method B, I have selected two pairs of masks with differentβ values. The chosen beta values are 12 rad/mm^2 and 9 rad/mm^2. These values are determined based on the limitations imposed by the pixel width of the Spatial Light Modulator (SLM) and the size of the FZA pattern. Each pair of masks has a phase difference of 0 and $\frac{3}{2}\pi$, respectively. The mask-sensor distance is set to 7.5 mm for this particular experiment.

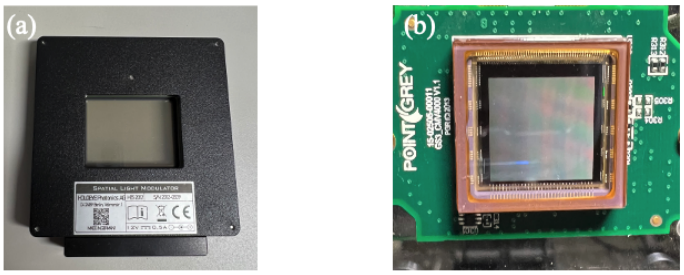


Fig. 3.12 (a) Spatial Light Modulator (SLM) and (b) Image Sensor used in the experiment.

Resolution chart target

For the quantitative analysis of the proposed methods using the SLM imaging system, I conducted an experiment using the 1951 USAF resolution chart. The resolution chart was displayed on an LCD monitor (ASUS VA32AQ) with a pixel pitch of 0.27 mm. The size of the displayed chart (Fig.3.13) was 222 mm $\times$ 227 mm. The object-mask distance was set to t = 250 mm. The reconstruction results obtained from the experiment are presented in Fig.3.13. The reconstructed results are shown in Fig.3.14 and Fig.3.15.

In Fig.3.14, (a) represents the result obtained from proposed method A, while images (b) and (c) are obtained from the conventional method with mask-sensor distances of 7.0 mm and 7.5 mm, respectively. The red box highlights the region where the resolution limitation can be visually assessed. In the image (b), the resolution limitation in the object plane is approximately 3.5 mm, while in image (c), it is around 3.0 mm. By applying method A, the resolution is significantly improved, and the resolution limitation in the object plane is

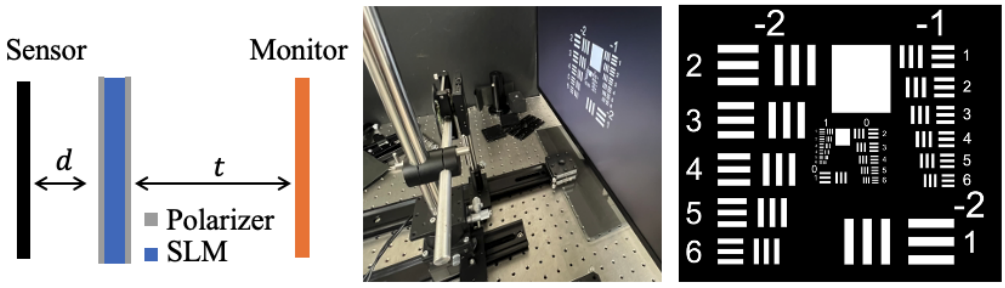


Fig. 3.13 Experiment setup and the resolution chart target.

enlarged to 1.4 mm, as shown in image (a). These results provide evidence that method A successfully increases the resolution of the captured images.

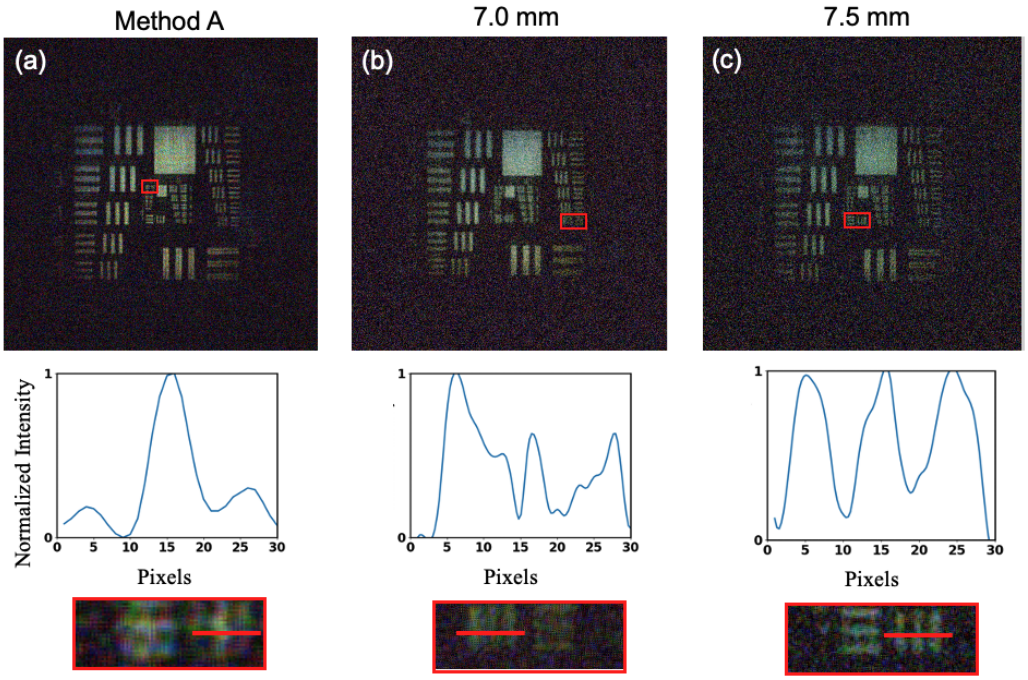


Fig. 3.14 The first row in (a), (b) and (c) is the reconstruction images from the proposed method A, the conventional method at mask sensor distance 7.0 mm and 7.5 mm, respectively. Red box indicates the visually assessed resolution limitation, the zoomed area of the red box and the according line profile are in the third and second row of (a), (b) and (c), respectively. The line profiles are drawn after the min-max normalization of the image intensity.

The experiment results of method B are presented in Fig.3.15. Image (b) is reconstructed using the conventional method with a β value of 9, while image (c) is obtained with a β value of 12. In contrast, image (a) represents the result obtained from the proposed method B. By visually evaluating the images, the red box indicates the region where the resolution limitation can be observed. In image (b), the resolution limitation in the object plane is approximately 3.5 mm, while in image (c), it is around 3.0 mm. However, after applying the proposed method B, the resolution limitation is significantly improved, and it is reduced to 1.6 mm in the object plane, as shown in image (a). This outcome provides further evidence to support the effectiveness of method B in enhancing the resolution of the captured images.

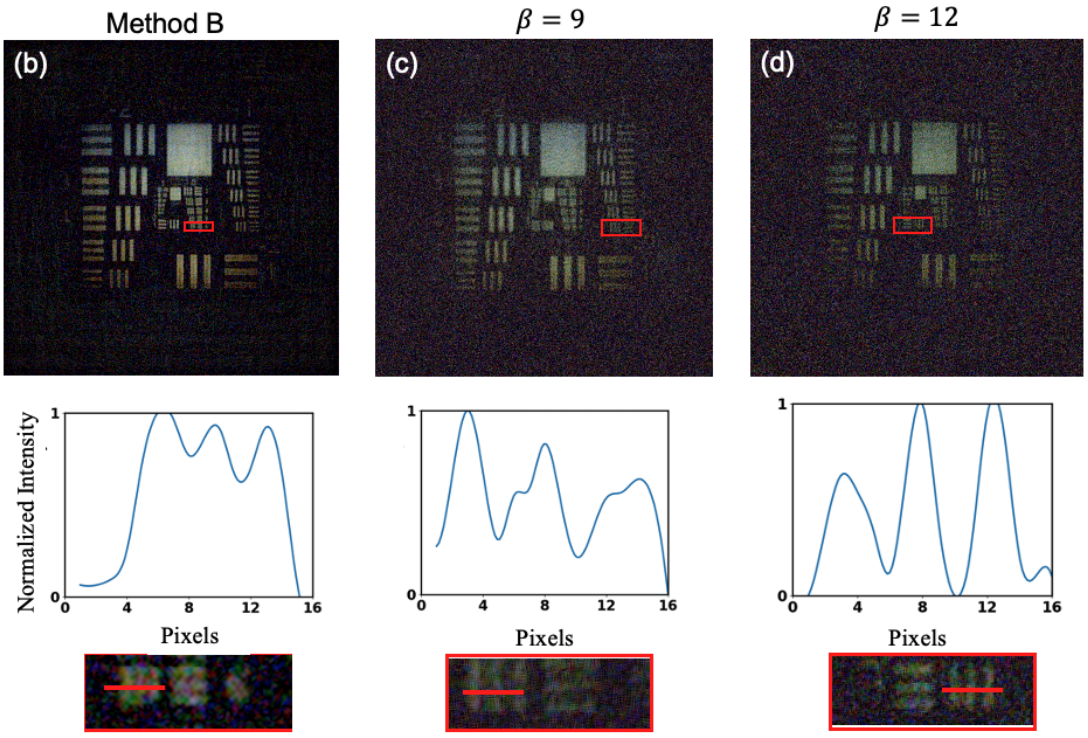


Fig. 3.15 Reconstruction images from (a) proposed method B; conventional method using (b) β=9 and (c)β=12. Red box indicates the visually assessed resolution limitation. The zoomed area of the red box, and the according line profile are in the third and second rows of (a), (b), and (c), respectively. The line profiles are drawn after the min-max normalization of the image intensity.

3D object

To verify the performance of the proposed methods on natural 3D objects, another experiment is set up as shown in Fig.3.16: (a) shows object A and object B used in the experiment and (b) shows the diagram of the experimental setup. The reconstruction results are shown in Fig.3.17 and Fig.3.18 . For comparison, I consider the Alternating Direction Method of Multipliers (ADMM) with TV. which is a popular method for image reconstruction. For comparison purposes, the Alternating Direction Method of Multipliers (ADMM) with Total Variation (TV) regularization is considered in this experiment. ADMM-TV is a popular method used for image reconstruction. In this experiment, ADMM-TV is utilized for refocusing reconstruction. The Point Spread Functions (PSFs) corresponding to two different distances are employed to estimate two images focused at different depths. Each ADMM reconstruction is performed for 500 iterations, with a tuning parameter τ set to 0.5.

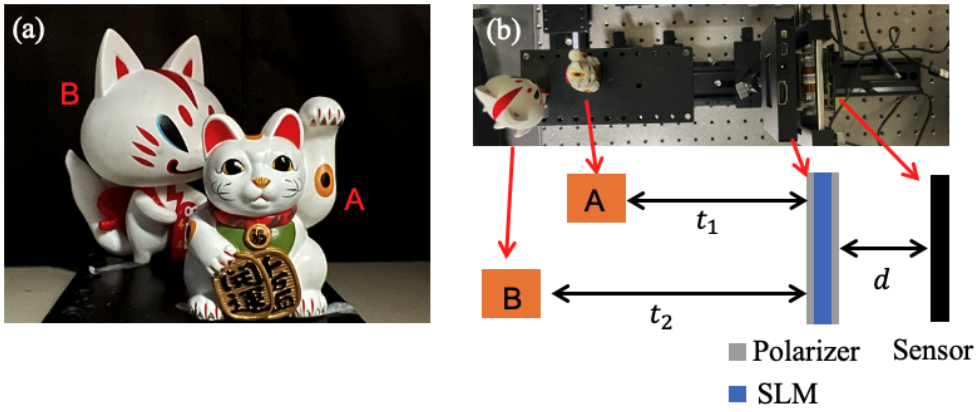


Fig. 3.16 (a) Object A and B; (b) experiment diagram.

The reconstruction results at focused distances of 130 mm and 170 mm are presented in In Fig.3.17 . In these images, the red box represents the in-focus object, while the blue box indicates the out-of-focus object. When examining the conventional methods (Fig.3.17(a), (b), (c), and (d)), we can observe a blurred pattern around the edges of the image. This blurring is a result of the mismatch between the geometric optics-based forward model and the diffraction propagation. After applying the proposed method A, which is based on wave optics theory, the blurred pattern is effectively removed, and the edges appear sharper, as depicted in Fig.3.17(e) and (g). However, it should be noted that the proposed method A may exhibit more noticeable noise due to the manual adjustment of the mask sensor distance.

This manual adjustment can introduce calibration artifacts during the image registration process. By visual evaluation, it can be concluded that the results from the ADMM-TV method have a similar image quality to the proposed method. However, it is worth noting that the iteration-based ADMM-TV method requires a significantly longer reconstruction time compared to the proposed method. Moreover, the ADMM-TV method also necessitates careful parameter tuning.

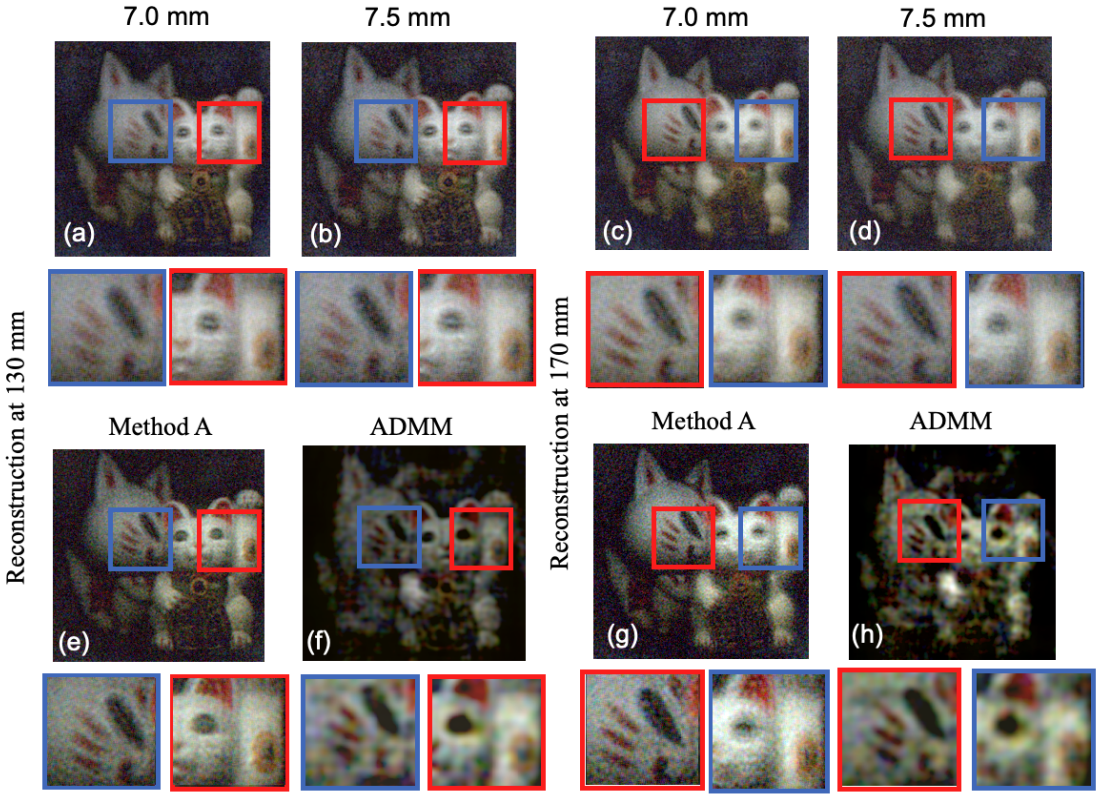


Fig. 3.17 Left: Under focused distance 130 mm: (a) and (b) are the results from conventional method at mask sensor distance 7.0 mm and 7.5 mm, (e) and (f) are the results from proposed method A and ADMM-TV; Right: Under focused distance 170 mm: (c) and (d) are the results from conventional method at mask sensor distance 7.0 mm and 7.5 mm, (g) and (h) are the results from proposed method A and ADMM-TV; Red box shows the zoomed area from in-focus object and blue box shows the zoomed area from out-of-focus object.

In Fig.3.18, (a) and (d) depict the reconstructed images obtained using FZA (β=9) with the conventional method at focus distances of 130 mm and 170 mm, respectively. Fig. 3.18 (b) and (e) show magnified versions of Fig. 3.17 (b) and (d) respectively. On the other hand, Fig. 3.18 (c) and (f) present the reconstructed images using method B at the focus distances of 130 mm and 170 mm, respectively. The red boxes highlight the zoomed areas from the

in-focused objects. By comparing the images, a significant improvement can be observed in terms of sharper and clearer details after applying method B. Moreover, method B offers the advantage of avoiding extra alignment errors and artifacts that may occur in method A. The blue boxes indicate the zoomed areas from the out-of-focus objects. It is easy to observe the differences between the in-focus and out-of-focus objects when the focus distance is changed. This observation confirms the effective refocusing reconstruction ability of the proposed method B.

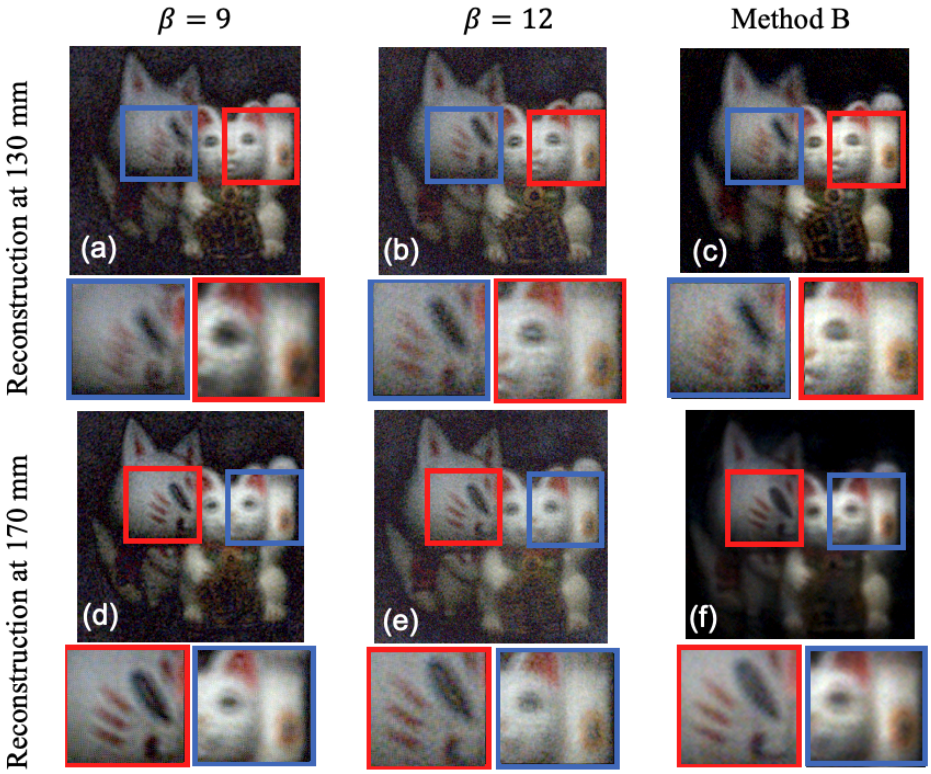


Fig. 3.18 Reconstruction images from conventional method at focus distance 130 mm using (a) β=9; (b)β=12; at focus distance 170 mm using (d) β=9; (e)β=12; images are reconstructed from proposed method B at distance (c) 130 mm and (f) 170 mm. Red box shows the zoomed area from in-focus object and blue box shows the zoomed area from out-of-focus object.

3.6 Summary

In this chapter, I provide a brief introduction to the published works on the study of FZA lensless cameras, highlighting their contributions and limitations. Following that, I present the geometric optics-based imaging model of the FZA lensless camera, which is based on the assumption of shift invariance and describes the imaging process. I discuss the existence of zero-crossings in the frequency response caused by the special design of the FZA pattern.

Next, I introduce the conventional method based on the fringe scanning technique, which effectively reconstructs images under ideal geometric conditions. This method synthesizes masks with different phases to compensate for the zero-crossings and produce a smooth frequency response. However, in real-world scenarios, diffraction occurs, leading to a mismatch between the proposed forward model and the actual imaging process. This mismatch introduces additional zero-crossings in the frequency response, thereby limiting the image resolution.

To address this limitation, I propose a wave optics-based analysis of the imaging process, identifying the relationship and causes of the zero-crossings resulting from diffraction. I then propose an LMSE-based synthetic method that utilizes information from different mask-sensor distances. While this method is effective, it is not practical due to certain limitations. Another method I propose involves synthesizing images using images captured from masks with different spacing. I present computer simulations and optical experiments to validate the effectiveness of these methods. Quantitative analysis is performed using resolution targets, and the refocusing ability of the FZA lensless camera is also evaluated using various natural 3D objects as targets.

Two primary challenges necessitate further exploration. First, the proposed approach, which involves adjusting the mask-sensor distance or altering the mask pattern to capture varying information, may fall short in efficiently capturing high-speed moving objects. Second, the imaging principles inherent to lens-based cameras often dictate specific lighting prerequisites, constraining the efficacy of lensless cameras in low-light settings. Moreover, determining the optimal mask-sensor distance combination or the ideal pitch size combination remains problematic. A potential solution could be a data-driven method, where the synthesized frequency response serves as the input, and the corresponding distance combination set or pitch size combination set is the output. Once trained, this model could suggest the optimal combination values tailored to a specific frequency response input.

Chapter 4

AI based refocusing for lensless camera

4.1 Related work

Post-capture refocusing stands as a notable feature for cameras, allowing for adjustments to the bokeh effect in photography or for targeted analysis in cell measurements. Conventional cameras struggle to achieve this without resorting to intricate and pricey light-field technologies. The core distinction between lensless and lens-based cameras is rooted in their respective point spread functions (PSFs) and the subsequent implications on depth perception. In lensless cameras, the omission of a tangible lens results in an expansive and diffused PSF. The configuration and magnitude of the PSF in such cameras are intricately tied to the distance of the captured subject. Consequently, the image accrued inherently houses depth-related information regarding the objects within the frame. This characteristic equips lensless cameras with refocusing aptitude and even the potential to reconstruct an object's 3D tableau[36]. When deploying the conventional model-based refocusing technique for lensless cameras, the pertinent PSF information at the envisioned focus distance becomes pivotal. Integrating this PSF into the reconstruction phase allows the image to undergo refocusing, rendering a lucid depiction of the subject at that precise range. This results in a revamped image wherein objects at the chosen distance manifest with enhanced clarity, while those beyond or before it transition into varying degrees of blur[95].

To achieve refocusing reconstruction in lensless cameras, [45]measures the point spread function (PSF) at different distances prior to the image reconstruction process. By capturing the PSF information at multiple distances, they are able to incorporate it into the reconstruc-

tion algorithm to achieve refocusing capabilities. The PSF real-measurement method can increase the accuracy of the PSF, but it is severely affected by the calibration accuracy and the multiple measurements are time-consuming and inflexible. I have decided to use FZA as the coded mask for the lensless camera. FZA offers several advantages, one of which is its special pattern that allows for the mathematical calculation of the imaging model. This makes refocusing in the FZA lensless camera relatively straightforward, as the PSF can be easily calculated at any desired distance[90, 95, 59]. However, the theoretical calculation is not accurate, especially in the case of lensless cameras, as there are too many variables that can affect the quality of the image reconstruction. Although the [36] does not explicitly mention refocusing, it is indeed possible to use the 3D reconstruction result to achieve refocusing capabilities. However, the computational cost of 3D reconstruction can be high. Furthermore, in the case of blurred images, the regularisation term used in the ADMM may consider the blurred regions as noise and attempt to reconstruct them, resulting in a loss of depth information. This can degrade the overall quality of the reconstructed image and affect the accuracy of the depth information[101].

In recent years, the application of learning-based methods has gained popularity in the lensless community. However, it is worth noting that most learning-based methods focus on generating an all-in-focus image rather than a selected focus distance image. A primary challenge in lensless camera refocusing reconstruction lies in the nature of its training data. The process of dataset accumulation in the lensless camera realm typically bifurcates into two methodologies. The first involves the numerical simulation of a 2D image based on a forward model. Conversely, the second displays a 2D image on a monitor, which is then captured simultaneously by both lens-based and lensless cameras. Both techniques sidestep the consideration of 3D objects, explaining the absence of reported learning-based refocusing reconstructions in lensless camera literature[48, 49, 51]. In their research, [37] unveiled a study showcasing a feed-forward deep neural network adept at deducing depth maps from a singular lensless camera snapshot. Their model leveraged a self-simulated dataset derived from an RGBD dataset for training purposes. Their findings underscored the latent capabilities of lensless cameras in parsing 3D information, chiefly by deriving depth insights from the captured patterns, with the aid of their uniquely curated dataset. While such depth maps hold promise for myriad applications, including scene comprehension or segmentation, the intensity images reconstructed by their methodology remain uniformly focused. Should the overarching aim gravitate towards exacting refocusing, augmenting their approach with supplementary techniques might be imperative.

4.2 Proposal and data preparation

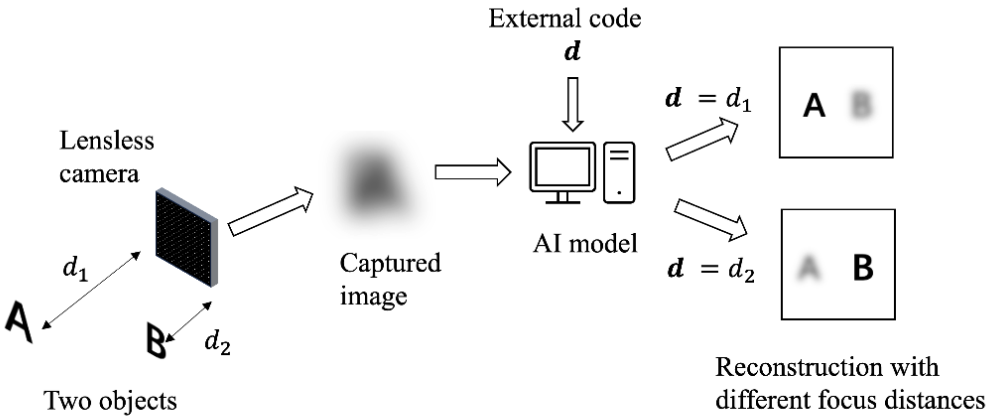


Fig. 4.1 Proposed learning-based refocusing reconstruction pipeline.

I would like to propose an arbitrary refocusing reconstruction model for lensless cameras (Fig.4.1), allowing the user to control the focusing distance. Drawing inspiration from speech synthesis research[102, 103], where external control codes are used to determine the target speaker or emotion, I consider using a similar approach in lensless cameras. Specifically, I propose using an input matrix of the same size as the image, filled with a constant value to represent the desired focusing distance for the output, as shown in Fig.4.2. By employing this approach, the model can take the user-defined digital propagation matrix (DPM) concept used in [104] and apply it to determine the target focusing plane. The input matrix, filled with consistent values, serves as a representation of the desired distance for refocusing. This allows the model to reconstruct the image accordingly, placing emphasis on the specified focusing distance.

Data preparation

In this research, I advocate for the use of genuine experimental data for data collection. When it comes to the coded pattern, my approach involves employing two monitors to display 2D images, positioning them at varying distances from the lensless camera. By manipulating the distances of these monitors, I can garner a dataset that is both rich and versatile. As

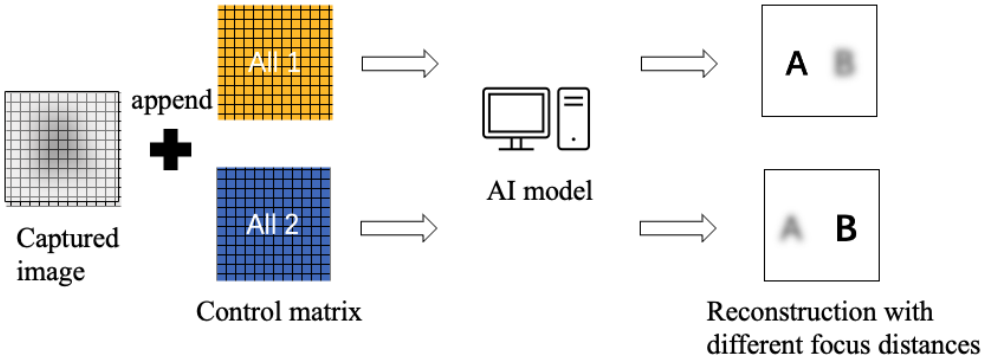


Fig. 4.2 Diagram of the external control signal.

for sourcing ground truth data, capturing it via a lens-based camera would be the ideal procedure. However, in light of logistical constraints such as experimental setup and the time-intensive nature of this method, I have elected to employ a simulation process. This involves simulating the Point Spread Function (PSF) of the lens-based camera to produce out-of-focus images, which can be conveniently adjusted. As a result, the model, once trained, can be operationalized in real-world contexts without the imperative to measure the PSF from the lensless camera, a process that often proves challenging in acquiring high-precision data.

Nonetheless, owing to the time constraints inherent to this study, I have opted to initially lean on numerical simulations during the data preparation phase. I selected RGB images from three open datasets (DIV2K dataset [105], Mir Flickr[106], and ImageNet[107]), resulting in a total of 22,600 images. In the simulation, I use the same lensless camera parameter settings as in [53], where the mask size is 2.15 mm and the pattern is shown in Fig.4.3(a); the mask-sensor distance is 2.5 mm and the image sensor pixel pitch is 0.0024 mm (2048 $\times$2048 pixels). Two rectangular 2D images are placed at 160 mm and 200 mm from the lensless camera, with sizes of 158 mm$\times$79 mm and 196 mm$\times$98 mm respectively as shown in Fig.4.3(b). The simulated in-focus image is obtained by reshaping and resizing the original image, while the simulated out-of-focus image is generated by adding Gaussian blur to the resized image. The captured image is obtained through a convolution process based on geometric optics, similar to the approach described in [45],

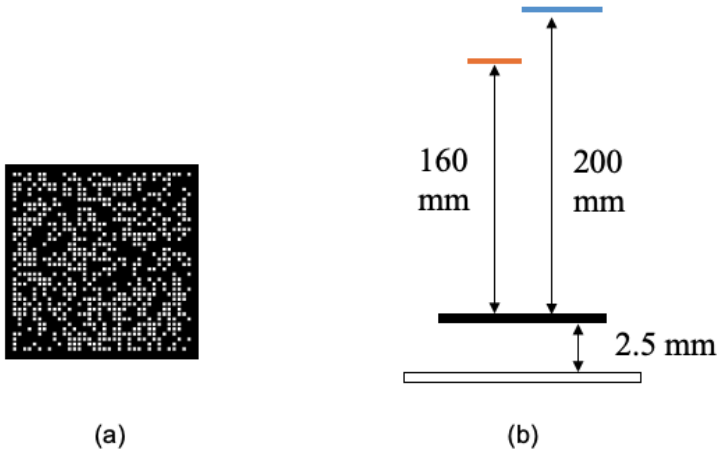


Fig. 4.3 (a) Mask pattern and (b) Simulation setting.

Fig. 4.4 Simulation of the (a) captured image; Reconstruction images (b) and (c) at different focusing distances.

$$g = \mathrm{PSF}_{d_1} * f_1 + \mathrm{PSF}_{d_2} * f_2, \tag{4.1}$$

where PSF_{d_i} and f_i means the PSF and object at distance d_i, respectively. g means the captured image. As shown in Fig.4.4, (a) is the simulated captured image and (b) and (c) are the corresponding ground truth with different focusing distances.

4.3 Experiment

4.3.1 Experiment settings

The network architecture utilized in this study is adopted from [53], as shown in Fig.4.5. However, certain modifications have been made. Firstly, the input channel structure has been altered to include not only a single channel but a two-channel structure. This structure consists of an original channel containing the raw captured image and an appended control matrix that represents the selection of focusing distance. Secondly, for the purpose of facilitating training and expediting result verification, the input size has been reduced to 256 $\times$ 256 pixels. Similarly, the output size has also been reduced to 256 $\times$ 256 pixels. These adjustments have been implemented as part of the exploration phase to ensure ease of training and efficient evaluation of outcomes. The details of the parameters in the model are also the same as the architecture in [53], as shown in the Table.4.1.

Table 4.1 Parameters setting.

	Patchify			Transformer		
Block No.	K	S	P	C	L	N
1	7	4	3	64	2	1
2	3	2	1	128	2	2
3	3	2	1	256	2	4
4	3	2	1	512	2	8

The training specifics are outlined as follows: the training process was conducted using an Intel Xeon E5-2698 v4 CPU (2.2GHz), along with two NVIDIA TESLA V100 GPUs (32GB) for enhanced performance. Python 3.6.5 and PyTorch 1.7.1 were utilized for the implementation. The Adam optimizer was employed with β_1 set to 0.9 and β_2 set to 0.999. A weight decay of 0.01 was applied, while the batch size was set to 8. The initial learning rate was set to 5e-4, and a learning rate warm-up phase of 15000 steps was incorporated.

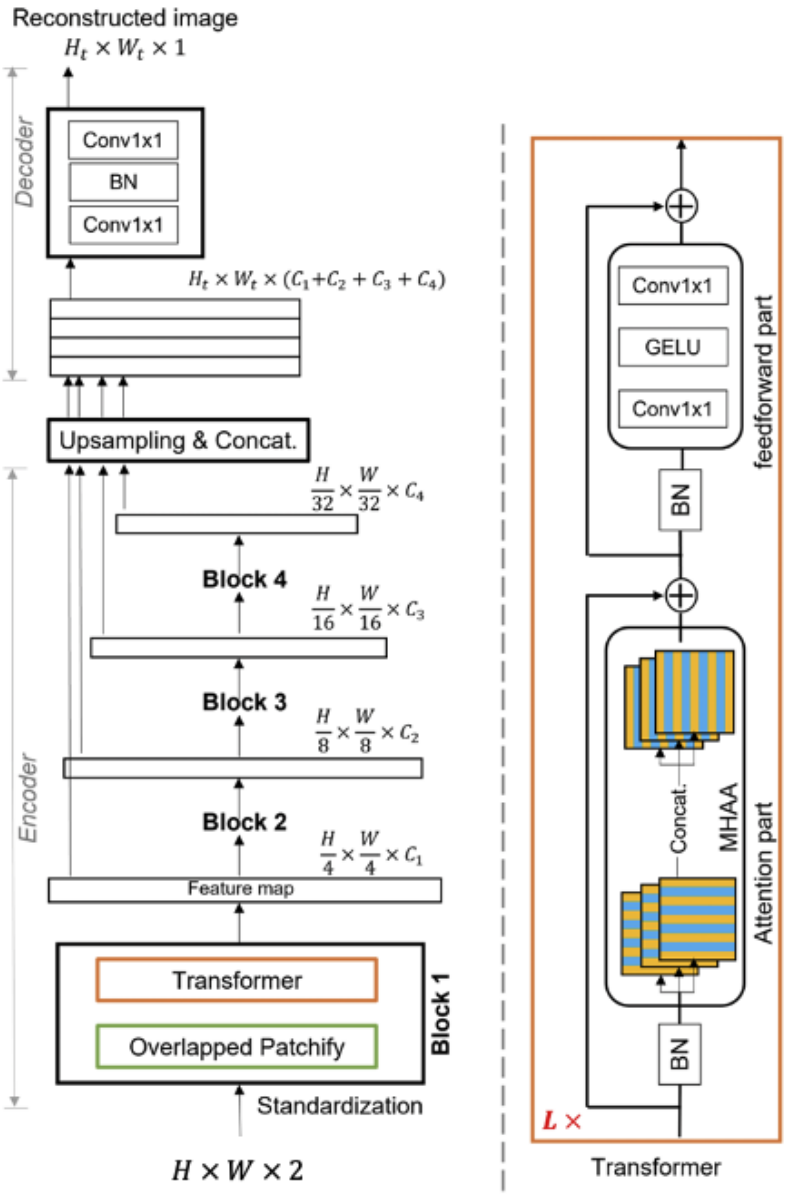


Fig. 4.5 Image reconstruction model. Revised from [53].

4.3.2 Experiment result

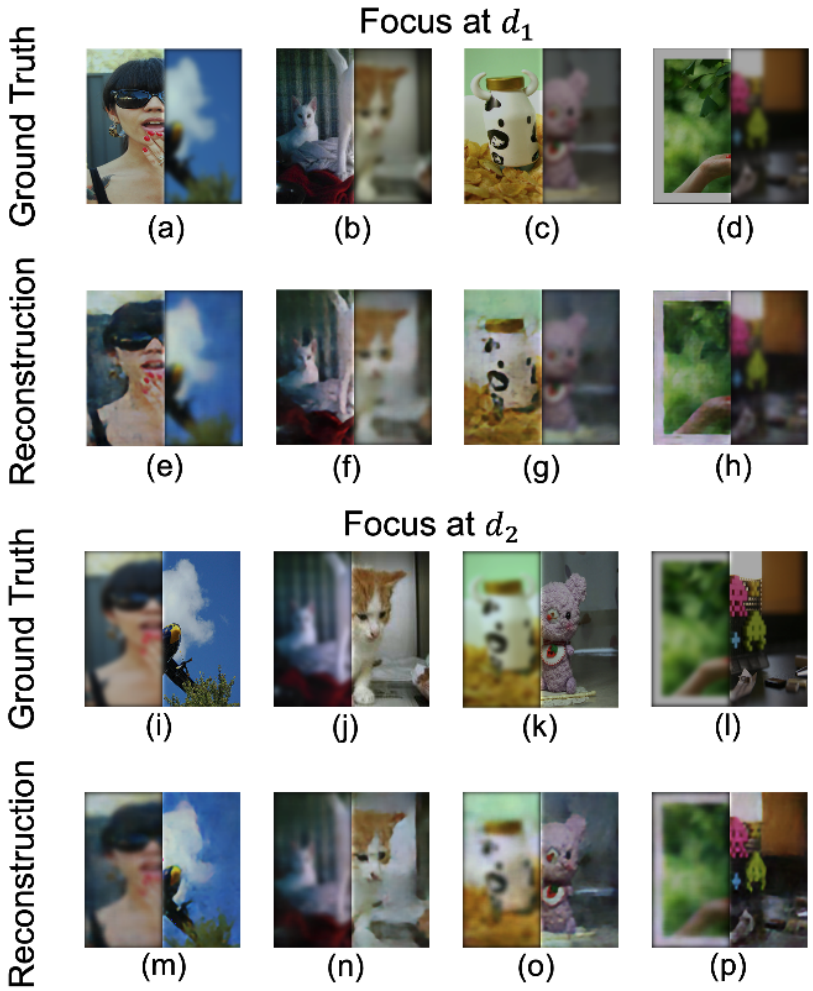


Fig. 4.6 Focus at distance d_1: (a)-(d) are ground truth, and (e)-(h) are reconstruction images. Focus at distance d_2: (i)-(l) are ground truth, and (m)-(p) are reconstruction images.

The reconstruction results at various focusing distances, along with their corresponding ground truth images, are depicted in Fig.4.6. By comparing the reconstructed images at different focusing distances, a distinct disparity can be observed: the image reconstructed at the in-focus distance appears clear and sharp, whereas the out-of-focus images exhibit a noticeable blur. This outcome further verifies the efficacy of the model incorporating an external control code, as it enables the reconstructed image to be focused at the desired distance as dictated by the external code.

By evaluating the PSNR (Table.4.2) between the reconstructed images and the corresponding ground truth, we can verify the high-quality performance of the proposed transformer-

based model in image reconstruction. This outcome further demonstrates the effectiveness of the transformer architecture in the task of reconstructing images from lensless camera data, highlighting its utility and value in this domain.

Table 4.2 PSNR value.

	Ground Truth No.							
	(a)	(b)	(c)	(d)	(i)	(j)	(k)	(l)
PSNR [dB]	25.0	24.1	24.1	24.2	23.5	25.3	23.6	23.5

To assess the adaptability of the reconstruction method for control codes beyond the defined parameters, I conducted tests over a range from 0.1 to 1.0. The results are presented in Fig.4.7. The top row displays images focused at distance d_2, while the bottom row reveals those focused at distance d_1. Each image is labeled with its control code value; those in red indicate results reconstructed from control codes provided during training. Below each image, the PSNR value for the reconstruction of the object at distance d_2 is displayed.

Upon visual inspection, it's evident that results for control code values ranging from 0.1 to 0.5 are nearly indistinguishable. Similarly, outputs for values between 0.6 and 1.0 exhibit comparable patterns. This suggests that the trained model struggles to focus at distances not explicitly defined during training. Furthermore, control codes not utilized during training appear to have negligible impact on the outcomes; the reconstructions align closely with the results from the nearest defined control code. The PSNR values further elucidate this phenomenon. A noticeable transition in image quality is observed between control code values 0.5 and 0.6, indicating that changes in the control code directly influence the shifting focus distance.

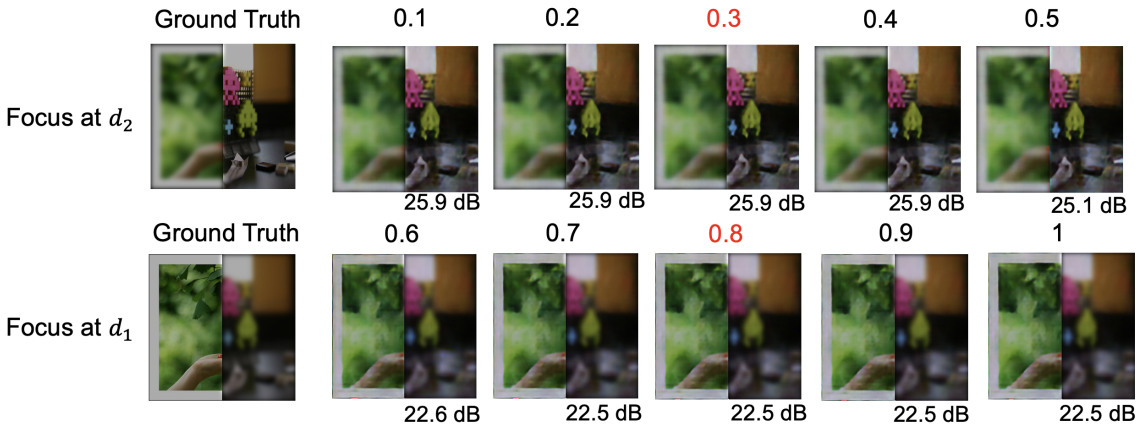


Fig. 4.7 Predictions from different control codes.

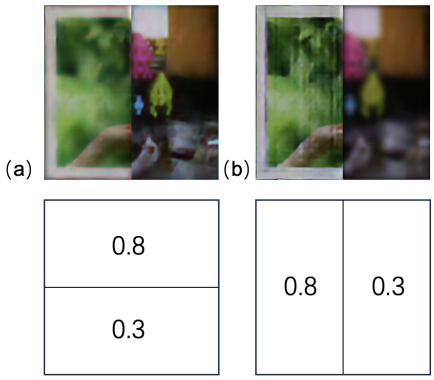


Fig. 4.8 Predictions from control codes with different spatial structures.

In addition to experimenting with control codes featuring various fill values, I delved into testing the model using control codes with distinct spatial structures. The results are depicted in Fig.4.8. The bottom row showcases the spatial structure of the control codes, while the top row displays the corresponding reconstruction outcomes.

Examining Fig.4.8(a), the left segment is reconstructed, presenting an in-focus image, the right segment, albeit reconstructed to some extent, still appears slightly out of focus. Contrastingly, in the outcome of Fig.4.8(b), only the left segment is distinctly reconstructed, leaving the right segment blurred. Furthermore, the reconstruction of the left segment exhibits certain vertical-pattern artifacts. These observations suggest that the current design of the control code cannot achieve selective or partial control refocusing. Considering the inherent imaging characteristics of the lensless camera, coupled with the mechanism of the transformer model, there seems to be a predominant emphasis on global information. As a result, the spatial structure of the control code exerts minimal influence on the ultimate reconstruction. For enhanced flexibility and to attain partial control refocusing capabilities, I recommend future experiments that involve integrating the control code at varied positions. This might provide deeper insights into the model's responsiveness and potential improvements.

4.4 Summary

This chapter delves into the exploration of a learning-based refocusing reconstruction model specifically designed for lensless cameras. One of the advantages of lensless cameras is their ability to encode depth information from the scene, surpassing traditional lens-based cameras in this aspect. However, due to the absence of a 3D dataset tailored for lensless cameras, the

typical output of a lensless camera reconstruction is an all-in-focus 2D image. In this study, I address this limitation by constructing a simulation dataset containing two objects positioned at different distances from the lensless camera. The aim is to utilize this dataset to develop a learning-based refocusing model specifically for lensless cameras.

To achieve control over the focusing distance of the reconstructed image, I drew inspiration from the field of speech synthesis research. By incorporating an external control code as an additional input alongside the raw captured image, the model generates a reconstruction image focused at the specified distance. The effectiveness of this design was verified through simulation experiments, which yielded positive results.

However, there are some limitations to consider. Firstly, the control range is currently restricted to only two discrete distances, whereas the ultimate goal is to enable the selection of a focus distance within a broader and smoother range. Secondly, the simulation data used in this study is overly idealized, lacking the presence of noise, and the model itself relies on a geometric optical model that does not fully account for complex scenarios. Lastly, the model's reliance on a transformer architecture introduces challenges in training, necessitating the need for enhanced training capabilities to streamline the process.

Chapter 5

See-through-screen Camera (STS) and its application with lensless camera

5.1 Related work

Camera systems are commonly integrated into IoT devices to provide photographic functionality and work in conjunction with other functions such as display or to support the sensing function. However, the traditional camera system, which includes a set of optical lenses and an image sensor, limits the overall size of IoT devices. In addition, the separate design of cameras and displays and the camera's field of view could lead to an unsatisfactory user experience.

With the current trend of full-screen smartphones, the presence of front cameras poses a challenge to achieving a seamless design (shown in Fig.5.1). Some companies have proposed a solution by incorporating an elevating front camera that uses a microelectronic motor to control the camera's pop-up mechanism. While this design allows for a full-screen experience, the additional control system can potentially impact the reliability of the smartphone and add weight. Moreover, this approach still maintains the split design of the camera and screen. To achieve a more robust solution for full-screen smartphones, the concept of under-display cameras has garnered significant attention[108, 109]. Under-display cameras involve hiding the camera beneath the display to capture images. There are two primary approaches to achieving this: enlarging the pixel distance above the camera to allow light to pass through[110, 111] or making certain pixels transparent to enable light

transmission[112, 113]. Both methods can enable a see-through screen camera and a full-screen display. However, it's important to note that these approaches may decrease display quality due to lower resolution in the area above the camera. Additionally, the camera's performance can be affected by diffraction influences[114].

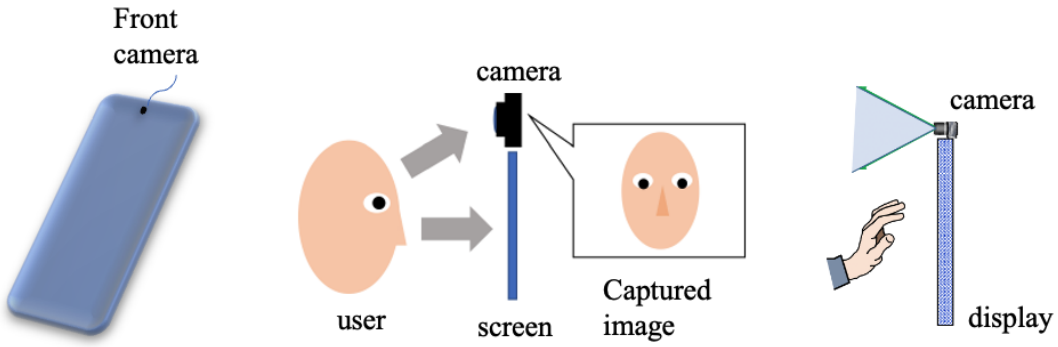


Fig. 5.1 Limitations of IoT devices (Left) full-screen design of smartphone; (Middle) Conventional online-meeting device (Right) Touchless screen.

The split design of the camera and screen not only impacts the full-screen experience but also hinders the user's interaction between the two. One common function in IoT devices is online meetings or video calls. However, due to the limited field of view of the camera and the separate design of the camera and display, the camera cannot capture the user's frontal image accurately, resulting in a strange image being displayed to others shown in Fig.5.1). This lack of accurate eye contact and other emotional expressions during the meeting can diminish the overall meeting quality, as these non-verbal cues are crucial for effective communication [115, 116]. Additionally, cameras are sometimes utilized to support touchless screens for gesture recognition or face detection. While this design enables the main functionality, it imposes limitations on user gestures. If a user's finger is outside the detection area specified by the device, it cannot be captured and detected, thus affecting the overall user interaction (shown in Fig.5.1).

To address the aforementioned issues, a software solution[117] has been proposed to correct the captured frontal image of the user. However, this solution is still dependent on the split design of the camera and display, and it cannot be applied to other scenarios. This is primarily due to the fixed field of view of the camera, which limits the captured information. Additionally, relying solely on computer vision solutions may introduce an unrealistic feeling to the corrected images. There are also hardware-based solutions that have been proposed to address these issues. These solutions [118–121]involve the use of a half-mirror and/or a separately aligned camera to capture the frontal image of the user. However, one limitation of these solutions is their bulky size, which restricts their application to mobile devices and other compact devices. A thin 'see-through' lens-less camera[122] has been proposed as a possible solution for capturing frontal images, enabling features such as eye-contact video meetings and near-screen gesture detection in a portable form factor. However, this camera faces challenges due to its low light efficiency, which negatively impacts its performance in practical applications. Furthermore, the resolution and image quality of this camera are severely limited, further restricting its usability in real-world scenarios. Based on the advantages of HOE, previous research[123, 124] has explored solutions to correct the photographic angle during online meetings. One proposed solution involves an off-axis virtual imaging system with a volume HOE (vHOE), which is capable of capturing the frontal image of the user's face. However, this system requires the separate placement of the holographic mirror and camera, and its size and calibration pose limitations on its extended functionality.

5.2 Ray-tracing analysis of STS camera

5.2.1 Proposed STS camera system

In this thesis, I propose a holographic waveguide device (HWD)-based see-through screen (STS) camera. The optical system of the proposed HWD-based STS camera is shown in Fig.5.2. It consists of two volume reflection HOEs attached to different surfaces of transparent glass plates. These volume HOEs are designed to have a large diffraction angle, exceeding the critical angle of the glass plate. The incident light from the object, which is Bragg-matched, is diffracted by the lower rear volume HOE. It propagates within the screen through total internal reflection and reaches the upper volume HOE. Finally, it is diffracted onto the image sensor by the upper front surface volume HOE. The screen acts as a waveguide device and

enables the proposed HWD-based STS camera to capture the frontal image of a user looking at the screen.

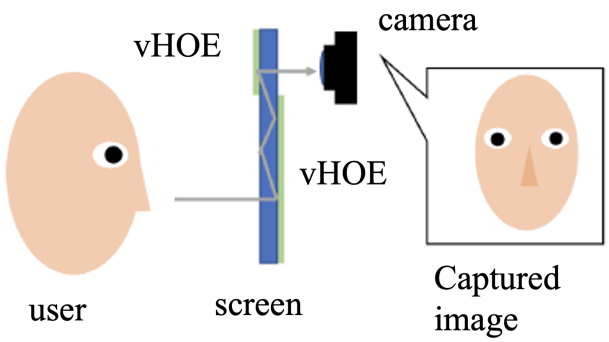


Fig. 5.2 Optical system of the proposed HWD-based STS camera.

The STS camera offers a potential solution for eye-contact telecommunication systems by incorporating a display behind the screen. This design is convenient and eliminates the need for precise optical calibration. In addition to its application in video communication systems, the proposed STS camera holds great potential for various visual media applications due to its compact and slim form factor, showcasing its versatile capabilities. One notable application of the proposed STS camera is its integration as an under-display camera for full-screen smartphones (Fig.5.3(a)), enabling a seamless and uninterrupted display experience without any spacing between pixels. Furthermore, the system finds utility in near-screen gesture recognition and touchless user interface displays. Traditional touchless user interface systems rely on infrared or capacitive sensing, limiting the types of gestures they can detect. Conversely, camera-based gesture recognition systems face challenges in detecting near-screen gestures due to their narrow field of view, as depicted in Fig.5.1. However, the proposed system overcomes this limitation, as illustrated in Fig.5.3(b).

One issue encountered with the HWD-based STS camera is the need for a larger vHOE to achieve a wider field of view. However, this leads to image superimposition and shifts on the image sensor, resulting in a significant vertical discrete blur that adversely affects the image quality. In the optical system depicted in Fig.5.4(a), dashed and solid lines represent incident light rays from different object points with varying numbers of total reflections

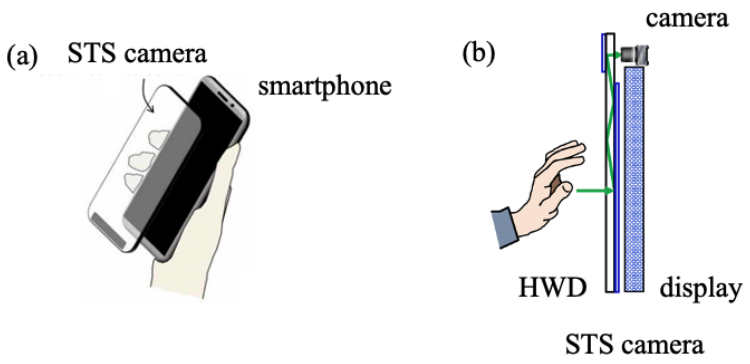


Fig. 5.3 Applications of HWD-based STS camera (a) Under-display-camera for a full-screen smartphone (b) near-screen touchless display.

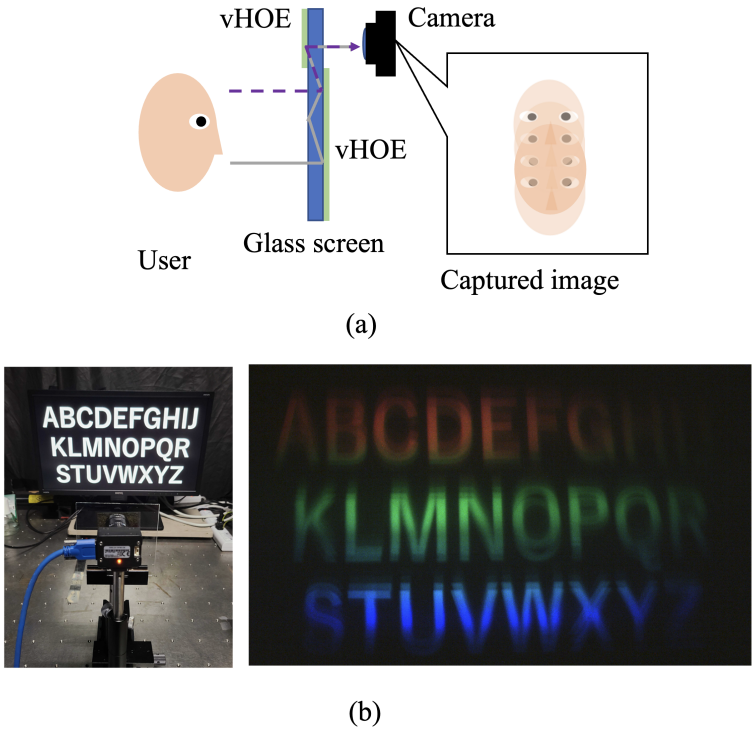


Fig. 5.4 (a) Schematic of the overlapping in the HWD-based STS camera; (b) Captured image from the optical experiment using Pointgray color camera with RGB module. The alphabet shown on the monitor is used as the object.

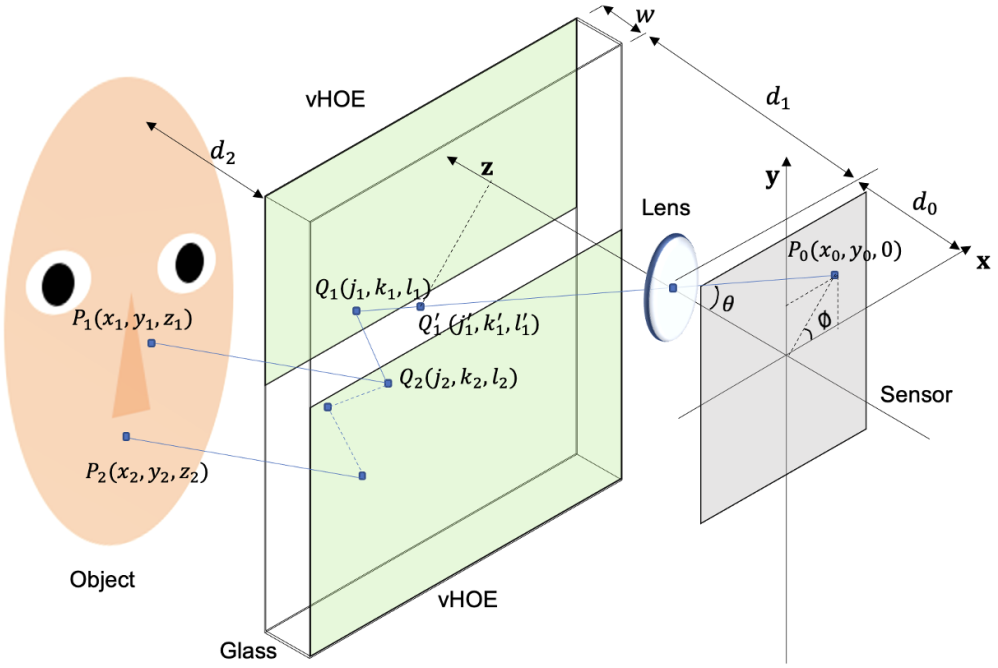
received by the same sensor pixel after diffraction by the top vHOE. The captured image in Fig.5.4(b) exhibits noticeable blurring caused by the multiplexing of images with different total reflection numbers. Additionally, color variations are observed where the top appears red, the middle is green, and the bottom is blue, despite using a vHOE recorded at a single wavelength, as shown in the image in Fig.5.4(b). This occurs because the received wavelength varies vertically due to the angle of incidence from the object, as governed by the Bragg diffraction law [125].

5.2.2 Model of blurred imaging by ray-tracing analysis

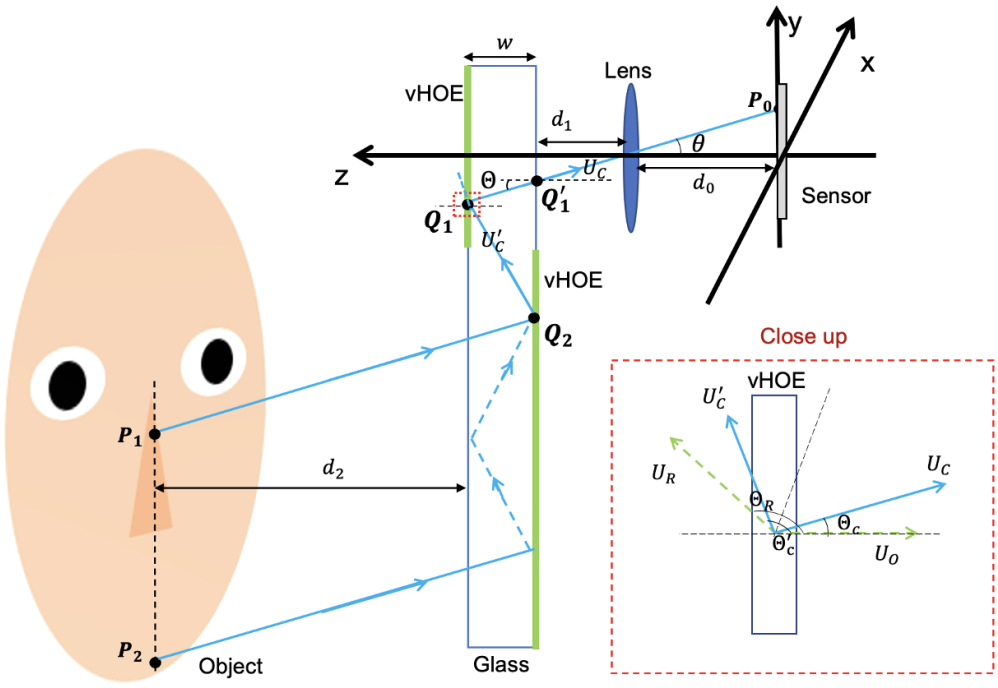
The blur observed in the system is a result of the superimposition of images from different counts of total reflections. To quantitatively analyze the optical transfer function of the proposed system, a ray-tracing method is employed. By tracing the light rays obtained by the sensor pixel based on the Bragg condition[126, 127], the relationship between the object and captured image can be calculated. As depicted in Fig.5.5, I assume a Cartesian coordinate system centered in the middle of the image sensor, with the sensor situated in the x-y plane and the z-axis perpendicular to the see-through screen (STS) plane. The camera lens is positioned at a distance of d_1 mm from the top volume holographic optical element (vHOE) and has a focal length of f mm. The distance from the lens to the sensor is d_0 mm. The object is located d_2 mm away from the STS, which has a width of w mm. The propagation of light is described by the azimuth angle ϕ and elevation angle θ. The horizontal (x) direction follows the law of reflection, while the vertical (y) direction involves Bragg diffraction and total reflection. To elucidate the vertical light propagation, a side view of the imaging system is presented in Fig.5.5(b).

Starting from a pixel point $\mathrm{P}_0(x_0, y_0, 0)$ on the sensor, it receives the incident light ray U_C originating from the point $\mathrm{Q}_1(j_1, k_1, l_1)$ on the top volume holographic optical element (vHOE). This ray passes through the point $\mathrm{Q}_1^{'}(j_1^{'}, k_1^{'}, l_1^{'})$ on the glass surface of the holographic waveguide device (HWD). The elevation angle θ and azimuth angle ϕ determine the direction of the incident light ray U_C, calculated as,

$$\tan\theta = \frac{\sqrt{x_0^2 + y_0^2}}{d_0}, \tan\phi = \frac{y_0}{x_0}. \tag{5.1}$$



(a)



(b)

Fig. 5.5 Imaging model based on ray-tracing (a) 3D view and (b) side view.

The geometric relationship between $\mathrm{Q}_1^{'}(j_1^{'},k_1^{'},l_1^{'})$ and $\mathrm{P}_0(x_0,y_0,0)$ can be modelled as,

$$\begin{bmatrix} j_1^{'} \\ k_1^{'} \\ l_1^{'} \end{bmatrix} = \begin{bmatrix} -\frac{d_1}{d_0}x_0 \\ -\frac{d_1}{d_0}y_0 \\ d_0+d_1 \end{bmatrix}, \text{and} \tag{5.2}$$

$$\begin{bmatrix} j_1 \\ k_1 \\ l_1 \end{bmatrix} = \begin{bmatrix} j_1^{'} \\ k_1^{'} \\ l_1^{'} \end{bmatrix} + \begin{bmatrix} w\tan\Theta\cos(\phi+\pi) \\ w\tan\Theta\sin(\phi+\pi) \\ w \end{bmatrix}, \tag{5.3}$$

The elevation angle Θ inside the glass medium can be related to the elevation angle θ outside the medium using the equation $\sin\theta = n\sin\Theta$, where n represents the refractive index of the glass medium. In this equation, the refractive index of air is assumed to be 1 for simplicity. The azimuth angles remain the same both inside and outside the medium. Additionally, for simplicity, the refractive indices of the glass and hologram medium are assumed to be equal.

The close-up view in Fig.5.5(b) illustrates the recording process of the volume holographic optical element (vHOE). The green dashed lines represent the recording waves of the vHOE. The object wave, denoted as U_O, enters the photopolymer vertically with an incident angle $\Theta_\mathrm{O} = 0$. On the other hand, the reference wave, denoted as U_R, enters the photopolymer at an angle Θ_R. Both U_O and U_R propagate perpendicular to the x-axis. Angles inside the medium will be represented by capital Θ with suffixes, and the grating vector $\mathbf{K} = [K_x, K_y, K_z]^\mathrm{T}$ (Fig.5.6) can be expressed as follows:

$$\begin{bmatrix} K_x \\ K_y \\ K_z \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2\pi n}{\lambda}(\sin\Theta_\mathrm{o} - \sin\Theta_\mathrm{R}) \\ \frac{2\pi n}{\lambda}(\cos\Theta_\mathrm{o} - \cos\Theta_\mathrm{R}) \end{bmatrix}, \tag{5.4}$$

where λ is the wavelength of the recording beams. From the diffraction formula of a thick grating, the relationship between the diffracted and incident beams can be expressed by the Bragg condition, which is given by:

$$\frac{1}{2\pi n}\begin{bmatrix} K_x \\ K_y \\ K_z \end{bmatrix} = \frac{1}{\lambda_C}\begin{bmatrix} \sin\Theta_\mathrm{C}\cos\phi_\mathrm{C} - \sin\Theta_\mathrm{C}^{'}\cos\phi_\mathrm{C}^{'} \\ \sin\Theta_\mathrm{C}\sin\phi_\mathrm{C} - \sin\Theta_\mathrm{C}^{'}\sin\phi_\mathrm{C}^{'} \\ \cos\Theta_\mathrm{C} - \cos\Theta_\mathrm{C}^{'} \end{bmatrix} = \frac{1}{\lambda}\begin{bmatrix} 0 \\ \sin\Theta_\mathrm{O} - \sin\Theta_\mathrm{R} \\ \cos\Theta_\mathrm{O} - \cos\Theta_\mathrm{R} \end{bmatrix} \tag{5.5}$$

where Θ_C, ϕ_C, $\Theta_C^{'}$, and $\phi_C^{'}$ represent the elevation and azimuth angles inside the medium for the incident and diffracted beams, respectively. The incident angles can be determined using Eq.5.3 as $\Theta_C = \Theta$ and $\phi_C = \phi + \pi$.

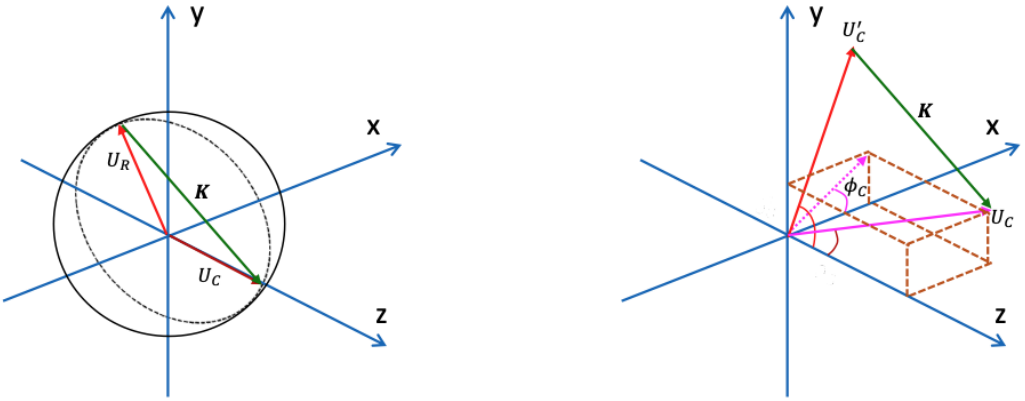


Fig. 5.6 K-space.

Eq.5.4 represents the Bragg condition, as illustrated in Fig.5.6. This relationship allows us to determine the wavelength and angles of the diffracted beam at point Q_1. Building upon the previous explanation, I can calculate the position of the diffraction point Q_2 on the bottom vHOE using the following equation:

$$\begin{bmatrix} j_2 \\ k_2 \\ l_2 \end{bmatrix} = \begin{bmatrix} j_1 + w\tan\Theta_C^{'}\cos\phi_C^{'} \\ k_1 + w\tan\Theta_C^{'}\sin\phi_C^{'} \\ l_1 - w \end{bmatrix}. \tag{5.6}$$

The top and bottom vHOEs are produced using the same fabrication process and adhere to the same Bragg condition. Consequently, the diffraction angles of the bottom vHOE at Q_2 are $\Theta_C^{'}$ and $\phi_C^{'}$. This implies that Q_2 receives incident light from the object with angles Θ_C and ϕ_C. The location of the object point P_1 can be determined using the following calculation:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} j_2 + w\tan\Theta_C\cos\phi_C + d_2\tan\theta\cos\phi_C \\ k_2 + w\tan\Theta_C\sin\phi_C + d_2\tan\theta\sin\phi_C \\ l_2 + w + d_2 \end{bmatrix}. \tag{5.7}$$

In the scenario where total reflection is involved, as depicted in Fig.5.5, Q_2 will also receive light after a single total reflection from object point P_2. Utilizing the previous analysis, the

position of P_2 can be computed using the following equation:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} j_2 + 2w\tan\Theta_C^{'}\cos\phi_C^{'} + w\tan\Theta_C\cos\phi_C + d_2\tan\theta\cos\phi_C \\ k_2 + 2w\tan\Theta_C^{'}\sin\phi_C^{'} + w\tan\Theta_C\sin\phi_C + d_2\tan\theta\sin\phi_C \\ l_2 + w + d_2 \end{bmatrix}. \tag{5.8}$$

In general, the pixel point $P_0(x_0, y_0, 0)$ on the sensor receives the light ray from the object point $P(x_i, y_i, z_i)$, and this relationship can be modeled as follows:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} -\frac{d_1+d_2}{d_0}x_0 + (2i-1)w\tan\Theta_C^{'}\cos\phi_C^{'} + 2w\tan\Theta_C\cos\phi_C \\ -\frac{d_1+d_2}{d_0}y_0 + (2i-1)w\tan\Theta_C^{'}\sin\phi_C^{'} + 2w\tan\Theta_C\sin\phi_C \\ d_0 + d_1 + d_2 + w \end{bmatrix}, \tag{5.9}$$

where the index i indicates the number of total reflections. The propagation will stop when the diffraction point exceeds the boundary of the bottom vHOE screen. Based on the previous analysis, one sensor pixel will receive multiple light beams from different object points after a varying number of reflections by ray-tracing. As a result, various images with vertical shifts are multiplexed on the sensor. Furthermore, it is important to note that the wavelength of the received light varies depending on the position of the object point due to the Bragg condition given by Eq.5.5.

5.2.3 Deblurred method

From the ray-tracing analysis of the relationship between the object and captured image, the imaging system can be described by a linear model:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{5.10}$$

In this equation, $\mathbf{g} = [g_1, \cdots g_{M^2}]^{\mathrm{T}}$ is an M^2-dimensional column vector representing the captured image of size $M \times M$. $\mathbf{f} = [f_1, \cdots f_{N^2}]^{\mathrm{T}}$ is an N^2-dimensional column vector representing the intensity of the original object of size $N \times N$. $\mathbf{H}$ is an $M^2 \times N^2$ matrix that describes the system transform model, and $\mathbf{n}$ is an N^2-dimensional column vector representing the noise.

To reconstruct the image $\hat{f}$, this study utilized the alternating direction method of multiplier (ADMM) with total variation (TV) regularization [38]. The reconstruction is formulated as:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \ \frac{1}{2}\|\mathbf{Hf} - \mathbf{g}\|_2^2 + \tau\|\Psi(\mathbf{f})\|_1, \tag{5.11}$$

In this equation, $\Psi(\cdot)$ represents the 2D gradient operator, and $\|\cdot\|_1$ denotes the l_1 norm. The regularization parameter τ controls the trade-off between fidelity to the measurements and the total variation regularization term.

5.3 Experiment

5.3.1 Fabrication of the vHOE

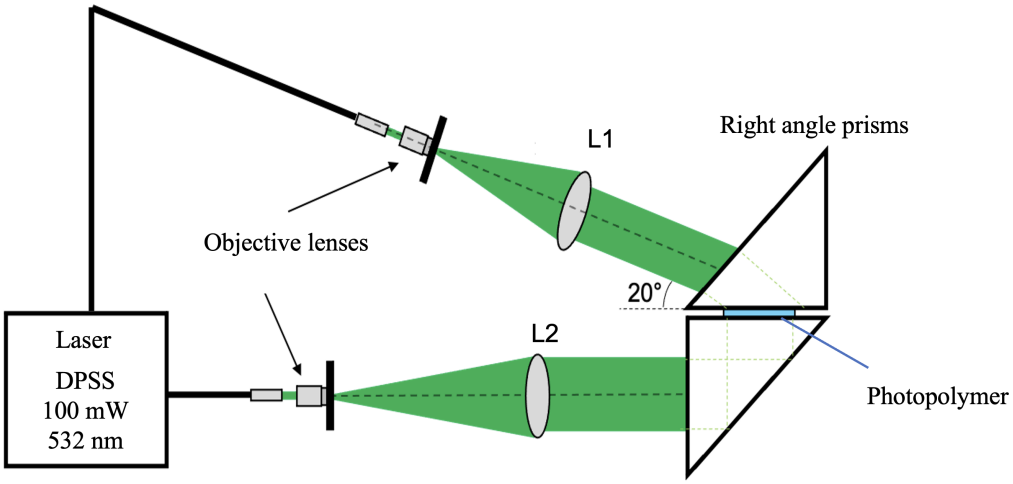


Fig. 5.7 Optical systems for recording the vHOE used in the HWD.

The optical system used in this study for fabricating the reflection-type volume holographic optical elements (vHOEs) is illustrated in Fig.5.7. A diode-pumped, solid-state continuous-wave laser (Samba 100 mW, Cobolt) operating at a wavelength of 532 nm served as the light source. The hologram recording was performed using a photosensitive material (Bayfol HX200, Covestro). To expose the HOE, parallel light was injected into the photopolymer perpendicular to its surface using a right-angled prism. The reference light was directed through the oblique side of the prism, entering the photopolymer at an angle of 61.2°, which exceeded the critical angle of the glass. Interference fringes were then recorded in the photopolymer to form the vHOEs. The dimensions of the vHOEs were 20 mm × 50 mm and 50 mm × 50 mm. The system's field of view (FOV) was 34.3°.

5.3.2 Experiment preparation

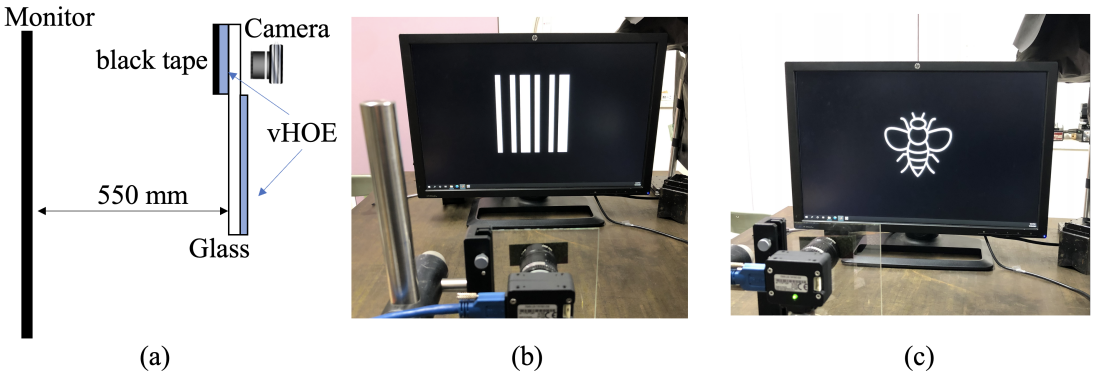


Fig. 5.8 (a) Configuration of measuring the system matrix; Optical experiment of (b) measuring the system matrix (c) capturing the target.

In the optical experiment, a liquid crystal display (LCD) monitor (HP ZR2440W) was employed to display the 2D targets. The monitor had a size of 1920 × 1200 pixels and a pixel pitch of 0.27 mm. A Pointgrey camera (CM3-U3-13Y3C-CS) operating in monochrome mode was utilized for capturing the images. The camera had a sensor size of 1280 × 1024 pixels, a pixel pitch of 4.8 μm, and a focal length of 6 mm. As depicted in Fig.5.8(a), the display was positioned 550 mm in front of the see-through screen camera. To prevent the camera from capturing the environment scene, black tape was employed to cover the top vHOE. The images captured on the camera sensor were cropped to a size of 440 × 440 pixels for further processing. The system's transformation matrix was measured using the

Hadamard basis patterns, which will be described in detail in the subsequent subsection. The actual implementation of the experiment is illustrated in Fig.5.8(b) and (c).

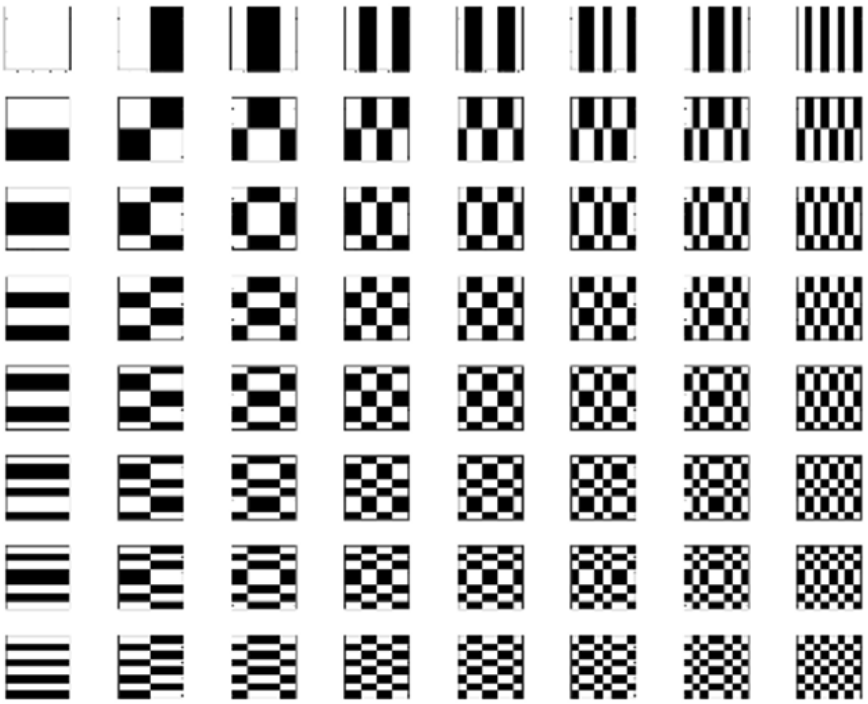


Fig. 5.9 Portion of the 64×64 size Hadamard basis patterns.

In order to measure the transformation matrix of the real imaging system, the following experimental preparation is carried out. In the optical experiment, a set of $N \times N$ Hadamard basis patterns (Fig.5.9) were used as input images to measure the system matrix $\mathbf{H}$ of the imaging system. The imaging process of the Hadamard patterns can be expressed as:

$$[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] = \mathbf{H}[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n], \tag{5.12}$$

Here, $[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ is a combination of column vectors representing the Hadamard patterns, and $[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ is the column vector of the captured images on the sensor.

To obtain the original mathematical expression, a modification process was performed. Let $\mathbf{u}_0$ represent the all-one Hadamard pattern shown on the monitor, and $\mathbf{v}_0$ be the captured image. The following transformations were applied:

$$\begin{aligned} \mathbf{u}_i^{'} &= 2\mathbf{u}_i - \mathbf{u}_0 \\ \mathbf{v}_i^{'} &= 2\mathbf{v}_i - \mathbf{v}_0 \\ i &= 0, 1, \ldots, n, \end{aligned} \tag{5.13}$$

Using these modified vectors, Eq.5.12 can be rewritten as:

$$\mathbf{V}^{'} = \mathbf{H}\mathbf{U}^{'}, \tag{5.14}$$

where $\mathbf{V}^{'} = [\mathbf{v}_1^{'}, \ldots, \mathbf{v}_n^{'}]$ and $\mathbf{U}^{'} = [\mathbf{u}_1^{'}, \ldots, \mathbf{u}_n^{'}]$. The matrix $\mathbf{U}^{'}$ satisfies the Hadamard property that its rows are mutually orthogonal, which can be expressed as:

$$\mathbf{U}^{'}(\mathbf{U}^{'})^{\mathrm{T}} = \mathbf{I}, \tag{5.15}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operator and $\mathbf{I}$ is the identity matrix. By multiplying both sides of Eq.5.14 by $(\mathbf{U}^{'})^{\mathrm{T}}$, we obtain the following equation:

$$\mathbf{V}^{'}(\mathbf{U}^{'})^{\mathrm{T}} = \mathbf{H}\mathbf{U}^{'}(\mathbf{U}^{'})^{\mathrm{T}}, \tag{5.16}$$

Simplifying the equation, we find:

$$\mathbf{H} = \mathbf{V}^{'}(\mathbf{U}^{'})^{\mathrm{T}}. \tag{5.17}$$

The system matrix $\mathbf{H}$ is then calculated and applied to the ADMM-TV reconstruction method in the subsequent steps. In this experiment, 4096 Hadamard basis patterns with a size of 64×64 pixels were displayed on the monitor at 640×640 pixels. The images captured on the sensor were cropped to 440×440 pixels for processing. $\mathbf{u}_i$ is a 4096×1 column vector, $\mathbf{v}_i$ is a 193600×1 column vector, and the calculated system transform matrix, $\mathbf{H}$, is a 193600×4096 vector.

Table 5.1 Evaluation indicators of the experiment's reconstruction images.

Object	CGM		ADMM	
	PSNR[dB]	Time[s]	PSNR[dB]	Time[s]
Apple	16.8	72.6	17.1	71.2
Bee	15.2	72.9	15.3	71.7
Face	18.7	73.1	18.7	71.4
Cameraman	12.0	76.7	15.1	75.8
Clock	9.6	77.5	11.9	75.1
Text	13.2	77.3	14.2	75.6

5.3.3 Result and analysis

In the optical experiment, three binary images (apple, bee, and face) and three grayscale images (cameraman, clock, and text) were used as targets. These images were resized to 64×64 pixels before being displayed on the monitor, which had a resolution of 640×640 pixels, to match the size of the Hadamard basis patterns.

The captured images exhibited significant vertical shift blurring, especially in the horizontal patterns, consistent with the theoretical analysis described in Section5.2.2. To reconstruct the images, the ADMM-TV method with a regularization parameter $\tau = 5$ was employed with 50 iterations, resulting in convergence to a threshold value. For comparison, the conjugate gradient method (CGM) was also used for image reconstruction.

The reconstruction results, as shown in Fig.5.10, effectively removed the blurred parts of the captured images, allowing for the reconstruction of the original objects using both the CGM and ADMM methods. However, some noise was still present at the top and bottom of the images in the binary image reconstruction using CGM. On the other hand, the ADMM method exhibited better noise elimination in the grayscale image reconstruction, resulting in smoother and clearer reconstructed images compared to CGM. The image quality was evaluated using the peak signal-to-noise ratio (PSNR), and the results are presented in Table.5.1. The numerical comparison demonstrated that the ADMM method possesses a more powerful denoising ability compared to CGM.

The results of the optical experiment verified the feasibility of the proposed frontal imaging system using the deblurring method. However, it is worth noting that the CGM and ADMM methods required a long reconstruction time, making them unsuitable for real-time communication systems. Future work will focus on developing more efficient image reconstruction algorithms to reduce the reconstruction time.

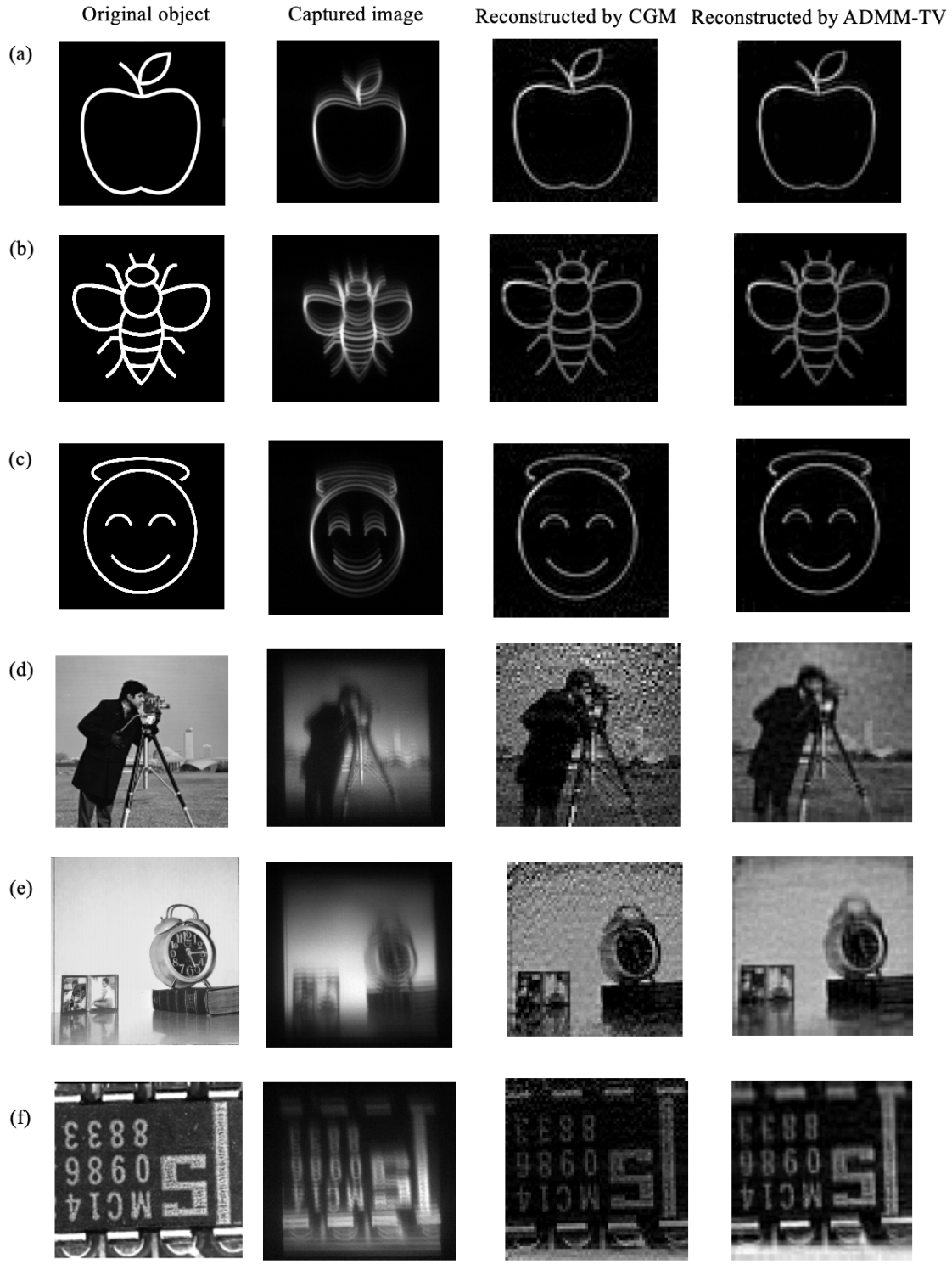


Fig. 5.10 Original objects (1st column), captured images (2nd column), and reconstructed images by the CGM (3rd column) and the ADMM-TV (4th column) of (a)Apple, (b)Bee, (c)Face, (d)Cameraman, (e)Clock and (f)Text in the experiment.

5.4 Screen-to-Sensor Imaging System (STIS)-a combination of STS camera and Lensless camera

In this study, an HWD-based STS camera is proposed that takes advantage of HOE technology to transform the screen into a camera. This design offers several advantages, including seamless integration of the camera with the display or other vision tasks, enabling the realization of a frontal imaging system in a compact and portable form factor. The HWD-based STS camera represents a new computational camera system that combines imaging capabilities with the functionality of a display, opening up possibilities for various applications. However, it still uses a conventional camera unit to capture the image, which adds weight and thickness to the overall computational imaging system and limits further application.

To further enhance the compactness and functionality of the imaging system, the combination of the HWD-based STS camera and lensless imaging can be explored. Lensless imaging eliminates the need for bulky lenses, enabling a more compact and lightweight system. It relies on computational algorithms to reconstruct images from the captured raw data. By integrating the lensless imaging approach with the STS camera, the overall size and weight of the system can be significantly reduced. This combination offers the advantages of both techniques, providing a compact and functional imaging system, named a Screen-to-Sensor Imaging System (STIS) as shown in Fig.5.11.

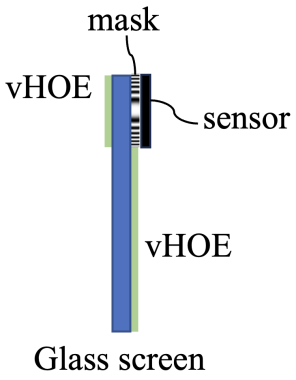


Fig. 5.11 Design of Screen-to-Sensor Imaging System.

The proposed STIS (Screen-to-Sensor Imaging System) offers versatile integration possibilities with various IoT devices, expanding its applications beyond online meeting terminals. Here are a few examples:

1. ADAS in Car Cabin: The STIS can be integrated into the modern car cabin as a screen for the electronic dashboard. Simultaneously, it can capture the driver's image with a wider field of view to monitor their safety. This integration provides a compact solution that combines display functionality with driver monitoring, enhancing the overall driving experience and safety.

2. Ultra-Thin Fingerprint Recognition Device: Traditional fingerprint recognition devices often have a bulky optical structure and limited functionality. By utilizing the STIS, an ultra-thin fingerprint recognition device with a built-in display function can be achieved. This integration improves user experience and allows for additional features, such as displaying notifications or interactive prompts.

3. Smart Mirror with Camera Function: By incorporating the STIS, a smart mirror can be designed with the capability to capture pictures of users. This integration adds a camera function to the mirror, enabling users to take selfies or capture images conveniently. It enhances the functionality of the smart mirror, making it a multi-purpose IoT device.

In each of these applications, the STIS offers the advantage of integrating imaging capabilities with a display function, resulting in compact and multifunctional IoT devices. By eliminating the need for separate camera units and leveraging the HOE-based technology, these devices can achieve thin form factors, improved functionality, and seamless integration of imaging and display capabilities.

5.5 Summary

In this chapter, the limitations of conventional camera systems or camera unit designs in IoT devices are discussed. These limitations include issues with full-screen smartphones, unsatisfactory online meeting experiences, and limited detection areas for touchless screens. To address these challenges, a solution is proposed using a holographic optical element (HOE) based waveguide device integrated into the imaging system, resulting in the development of the See-through-Screen (STS) camera.

The proposed imaging system faces a dilemma. To achieve a wider field of view, a larger volume holographic optical element (vHOE) is required. However, increasing the size of the vHOE causes diffracted light rays from the same object point to be diffracted at different Bragg conditions and reflect different numbers of times, resulting in blurring of the captured

image. To overcome this issue, the imaging process of the system is analyzed using ray tracing techniques to mitigate the vertical multiplexed blur caused by the larger size of the vHOE.

Subsequently, the Alternating Direction Method of Multipliers (ADMM) with Total Variation (TV) regularization is employed for image reconstruction based on the system transform matrix measured from the Hadamard matrix. Compared to the see-through lensless camera, the proposed STS camera offers higher light efficiency, a simpler image reconstruction method, and the potential for better image quality.

A prototype camera is developed and tested in an optical experiment using binary and grayscale images as targets. The evaluation results validate the effectiveness and practicality of the proposed theory, although the reconstruction quality is not yet optimal. Future advancements in reconstruction methods are expected to further enhance image quality.

The vHOE functions as a large-angle off-axis mirror, diffracting the Bragg-matched object light to the waveguide, which then carries the diffracted light to the camera positioned behind the screen. This setup enables the capture of the user's frontal image while they gaze at the screen, facilitating eye contact in telecommunication systems when combined with a display behind the screen. The compact size and integrated design of the proposed computational camera make it suitable for various visual devices. Potential applications include near-window gesture recognition and transparent communication systems when combined with a transparent display.

To realize a truly super-thin and functional camera system, a discussion on the combination of the STS camera with lensless imaging technology is provided in the thesis, exploring the potential for further advancements and improvements in the field.

There are several limitations of this study. The image reconstruction process using CGM and ADMM methods was time-consuming, and not suitable for real-time communication systems. Faster image reconstruction techniques should be explored in future work. Additionally, the experiments only used grayscale images, and future work should consider the use of full-color volume holographic optical elements (vHOEs) for reconstructing full-color images. Furthermore, the system matrix used for image reconstruction was measured at a specific distance, requiring remeasurement if optical conditions change or when reconstructing objects at multiple depths.

Chapter 6

Conclusion

Over the past few decades, cameras have undergone tremendous advancements in terms of performance and functionality. However, the fundamental structure of cameras has remained unchanged, relying on optical lenses and image sensors. This conventional lens-based camera system imposes limitations on the size and capabilities of cameras. Moreover, integrating cameras into media devices or IoT devices is an important application. The inclusion of lens-based cameras introduces weight and thickness restrictions to these devices and limits the user experience.

To address these challenges, I explored solutions in computational imaging, driven by the concept of eliminating optical lenses. As described in Chapter 2, computational imaging involves front-end optical encoding and back-end digital decoding, shifting the burden from optical lenses to digital processing. This approach opens up new possibilities and potential benefits for vision systems, including enhanced performance, access to additional information, and a more compact and lightweight structure.

In my research, I specifically chose to employ a lensless camera based on Fresnel zone aperture (FZA) to develop a thin, depth-sensitive camera system with fast image reconstruction. A common issue in amplitude mask-based lensless cameras is diffraction. In Chapter 3, leveraging the unique design of FZA, I proposed a wave-optics-based analysis that incorporates diffraction propagation. I conducted theoretical and qualitative analysis on the influence of diffraction on frequency response, followed by proposing an image synthesis method based on the mask-sensor distance to mitigate diffraction effects and enhance image resolution. To achieve more robust performance in practical applications, I also introduced

a mask-pattern-based image synthesis method. Both methods were proven effective not only through numerical simulations but also in optical experiments. It's noteworthy that model-based approaches tend to be more susceptible to noise and can be considerably time-intensive, especially when incorporating regularization terms. Conversely, learning-based methods, predominantly guided by data, typically exhibit fewer artifacts in their reconstructed results. Yet, a majority of these learning-based techniques predominantly produce all-in-focus images, even if they inherently assimilate depth information during the mapping process, because of the lack of a special dataset and control mechanism. In Chapter 4, I delved into a learning-based refocusing reconstruction model, introducing an external control code specifically designed to determine the focusing distance of the reconstructed output. While the experiments underscore the efficacy of this control code, there remains room for refining the model to achieve more flexible refocusing capabilities. This aspect is earmarked for future exploration.

Beyond the camera itself, it is crucial to consider integrating cameras with other digital devices due to the emerging trend of IoT. I argue that limitations in functional integration and user experience stem from the lens-based camera system and the traditional design that separates cameras from other media functions, such as displays. In Chapter 5, I proposed a see-through screen (STS) camera utilizing a holographic optical element (HOE)-based waveguide device. This STS camera employs the unique wavelength and angle sensitivity of HOE, allowing for frontal imaging with eye contact in online meetings and true near-screen gesture recognition in touchless displays. I also presented a ray tracing analysis, considering the Bragg law, for deblurring when using larger HOEs to achieve a wider field of view. A prototype camera was proposed and extensively tested in optical experiments. Additionally, combining lensless imaging with HOE-based imaging is a promising direction, but it poses new challenges that need to be addressed.

Overall, while this dissertation presents innovative approaches to address the limitations of traditional camera systems, it also highlights the need for continued research and development to overcome the remaining challenges and enhance the capabilities of thin and functional camera systems.

Future work

For lensless or STS cameras, as well as a combination of lensless and STS cameras, the accurate analysis of physical imaging models is crucial for image reconstruction and deblurring.

The quality of the reconstruction result heavily relies on the accuracy of the imaging model. Additionally, the imaging model is specific to objects at certain distances, meaning that any changes in distance require adjusting the model accordingly. This limitation restricts the flexibility of the device.

In the current landscape, AI is gaining popularity and playing a significant role in various research areas. However, most learning-based models are designed to learn a specific mapping based on a fixed object-camera distance. The learning-based refocusing reconstruction model presented in this study represents an initial exploration of a dynamic reconstruction model for lensless cameras. Moving forward, there is a need for more sophisticated and high dynamic range image reconstruction techniques that provide smoother results.

Furthermore, the proposed computational cameras should not be limited to capturing still scenes but should also consider moving objects. This presents a new challenge as fast reconstruction with depth localization becomes essential in such scenarios. The ability to quickly reconstruct images while accurately localizing the depth information is an important aspect to address in future developments.

References

[1] Banglei Guan, Yingjian Yu, Ang Su, Yang Shang, and Qifeng Yu. Self-calibration approach to stereo cameras with radial distortion based on epipolar constraint. *Appl. Opt.*, 58(31):8511–8521, Nov 2019.

[2] DooHyun Lee and InSo Kweon. A novel stereo camera system by a biprism. *IEEE Transactions on Robotics and Automation*, 16(5):528–541, 2000.

[3] Xiuxi Pan and Shinichi Komatsu. Light field reconstruction with randomly shot photographs. *Appl. Opt.*, 58(23):6414–6418, Aug 2019.

[4] Joseph N. Mait. A history of imaging: Revisiting the past to chart the future. *Opt. Photon. News*, 17(2):22–27, Feb 2006.

[5] Changyin Zhou and Shree K. Nayar. Computational cameras: Convergence of optics and processing. *IEEE Transactions on Image Processing*, 20(12):3322–3340, 2011.

[6] S.K. Nayar. Computational cameras: Redefining the image. *Computer*, 39(8):30–38, 2006.

[7] Kenneth Kubala, Edward Dowski, and W. Thomas Cathey. Reducing complexity in computational imaging systems. *Opt. Express*, 11(18):2102–2108, Sep 2003.

[8] Joseph N. Mait, Ravi Athale, and Joseph van der Gracht. Evolutionary paths in imaging and recent trends. *Opt. Express*, 11(18):2093–2101, Sep 2003.

[9] Joseph N. Mait, Gary W. Euliss, and Ravindra A. Athale. Computational imaging. *Adv. Opt. Photon.*, 10(2):409–483, Jun 2018.

[10] M. Levoy. Light fields and computational imaging. *Computer*, 39(8):46–55, 2006.

[11] Zuo Chao and Chen Qian. Computational optical imaging: An overview. *Infrared and Laser Engineering*, 51(2):20220110, 2022.

[12] Heejung Lee, JongWu Kim, JunWoo Kim, Philjun Jeon, Seung Ah Lee, and Dugyoung Kim. Noniterative sub-pixel shifting super-resolution lensless digital holography. *Opt. Express*, 29(19):29996–30006, Sep 2021.

[13] Zhi-Yong Hu, Yong-Lai Zhang, Chong Pan, Jian-Yu Dou, Zhen-Ze Li, Zhen-Nan Tian, Jiang-Wei Mao, Qi-Dai Chen, and Hong-Bo Sun. Miniature optoelectronic compound eye camera. *Nature Communications*, 13(1):5634, 2022.

[14] Edward R. Dowski and W. Thomas Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 34(11):1859–1866, Apr 1995.

[15] Ming-Jie Sun and Jia-Min Zhang. Single-pixel imaging and its application in three-dimensional reconstruction: A brief review. *Sensors*, 19(3), 2019.

[16] Jorge Ojeda-Casta neda and Cristina M. Gómez-Sarabia. Tuning field depth at high resolution by pupil engineering. *Adv. Opt. Photon.*, 7(4):814–880, Dec 2015.

[17] Jun Tanida, Tomoya Kumagai, Kenji Yamada, Shigehiro Miyatake, Kouichi Ishida, Takashi Morimoto, Noriyuki Kondou, Daisuke Miyazaki, and Yoshiki Ichioka. Thin observation module by bound optics (tombo): concept and experimentalverification. *Appl. Opt.*, 40(11):1806–1813, Apr 2001.

[18] Wonchan Lee, Daeseong Jung, Suho Ryu, and Chulmin Joo. Single-exposure quantitative phase imaging in color-coded led microscopy. *Opt. Express*, 25(7):8398–8411, Apr 2017.

[19] Yifan Peng, Qiang Fu, Hadi Amata, Shuochen Su, Felix Heide, and Wolfgang Heidrich. Computational imaging using lightweight diffractive-refractive optics. *Opt. Express*, 23(24):31393–31407, Nov 2015.

[20] Bingyun Qi, Wei Chen, Xiong Dun, Xiang Hao, Rui Wang, Xu Liu, Haifeng Li, and Yifan Peng. All-day thin-lens computational imaging with scene-specific learning recovery. *Appl. Opt.*, 61(4):1097–1105, Feb 2022.

[21] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26(3):70–es, jul 2007.

[22] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69–es, jul 2007.

[23] Changyin Zhou, Stephen Lin, and Shree K Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. *International journal of computer vision*, 93(1), 2011.

[24] Gonzalo R. Arce, David J. Brady, Lawrence Carin, Henry Arguello, and David S. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014.

[25] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, 2017.

[26] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996.

[27] Edward H Adelson, James R Bergen, et al. The plenoptic function and the elements of early vision. *Computational models of visual processing*, 1(2):3–20, 1991.

[28] Kaan Yücer, Alexander Sorkine-Hornung, Oliver Wang, and Olga Sorkine-Hornung. Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. *ACM Trans. Graph.*, 35(3), mar 2016.

[29] W. Matusik, A. Hornung, S. Avidan, and T. Basha. Structure and motion from scene registration. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1426–1433, Los Alamitos, CA, USA, jun 2012. IEEE Computer Society.

[30] Donald G. Dansereau, Oscar Pizarro, and Stefan B. Williams. Linear volumetric focus for light field cameras. *ACM Trans. Graph.*, 34(2), mar 2015.

[31] Can Chen, Haiting Lin, Zhan Yu, Sing Bing Kang, and Jingyi Yu. Light field stereo matching using bilateral statistics of surface cameras. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1518–1525, 2014.

[32] Tom E. Bishop and Paolo Favaro. The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):972–986, 2012.

[33] Vivek Boominathan, Jacob T. Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, 9(1):1–16, Jan 2022.

[34] Vivek Boominathan, Jesse K. Adams, M. Salman Asif, Benjamin W. Avants, Jacob T. Robinson, Richard G. Baraniuk, Aswin C. Sankaranarayanan, and Ashok Veeraraghavan. Lensless imaging: A computational renaissance. *IEEE Signal Processing Magazine*, 33(5):23–35, 2016.

[35] Sanjeev J. Koppal. A survey of computational photography in the small: Creating intelligent cameras for the next wave of miniature devices. *IEEE Signal Processing Magazine*, 33(5):16–22, 2016.

[36] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, Jan 2018.

[37] Dhruvjyoti Bagadthey, Sanjana Prabhu, Salman S. Khan, D Tony Fredrick, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. Flatnet3d: intensity and absolute depth from single-shot lensless capture. *J. Opt. Soc. Am. A*, 39(10):1903–1912, Oct 2022.

[38] Vivek Boominathan, Jesse K. Adams, Jacob T. Robinson, and Ashok Veeraraghavan. Phlatcam: Designed phase-mask based thin lensless camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1618–1629, 2020.

[39] E. E. Fenimore and T. M. Cannon. Coded aperture imaging with uniformly redundant arrays. *Appl. Opt.*, 17(3):337–347, Feb 1978.

[40] Stephen R. Gottesman and E. E. Fenimore. New family of binary arrays for coded aperture imaging. *Appl. Opt.*, 28(20):4344–4352, Oct 1989.

[41] M. Salman Asif, Ali Ayremlou, Ashok Veeraraghavan, Richard Baraniuk, and Aswin Sankaranarayanan. Flatcam: Replacing lenses with masks and computation. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 663–666, 2015.

[42] M. Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G. Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging*, 3(3):384–397, 2017.

[43] Kazuyuki Tajima, Takeshi Shimano, Yusuke Nakamura, Mayu Sao, and Taku Hoshizawa. Lensless light-field imaging with multi-phased fresnel zone aperture. *2017 IEEE International Conference on Computational Photography (ICCP)*, pages 1–7, 2017.

[44] Kyung Chul Lee, Junghyun Bae, Nakkyu Baek, Jaewoo Jung, Wook Park, and Seung Ah Lee. Design and single-shot fabrication of lensless cameras with arbitrary point spread functions. *Optica*, 10(1):72–80, Jan 2023.

[45] Yi Hua, Shigeki Nakamura, M. Salman Asif, and Aswin C. Sankaranarayanan. Sweepcam — depth-aware lensless imaging using programmable masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1606–1617, 2020.

[46] Oliver Kingshott, Nick Antipa, Emrah Bostan, and Kaan Akşit. Unrolled primal-dual networks for lensless cameras. *Opt. Express*, 30(26):46324–46335, Dec 2022.

[47] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, Sep 2017.

[48] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Learned reconstructions for practical mask-based lensless imaging. *Opt. Express*, 27(20):28075–28090, Sep 2019.

[49] Jiachen Wu, Liangcai Cao, and George Barbastathis. Dnn-fza camera: a deep learning approach toward broadband fza lensless imaging. *Opt. Lett.*, 46(1):130–133, Jan 2021.

[50] Xiuxi Pan, Tomoya Nakamura, Xiao Chen, and Masahiro Yamaguchi. Lensless inference camera: incoherent object recognition through a thin mask with lbp map generation. *Opt. Express*, 29(7):9758–9771, Mar 2021.

[51] Hao Zhou, Huajun Feng, Zengxin Hu, Zhihai Xu, Qi Li, and Yueting Chen. Lensless cameras using a mask based on almost perfect sequence through deep learning. *Opt. Express*, 28(20):30248–30262, Sep 2020.

[52] Xiuxi Pan, Xiao Chen, Tomoya Nakamura, and Masahiro Yamaguchi. Incoherent reconstruction-free object recognition with mask-based lensless optics and the transformer. *Opt. Express*, 29(23):37962–37978, Nov 2021.

[53] Xiuxi Pan, Xiao Chen, Saori Takeyama, and Masahiro Yamaguchi. Image reconstruction with transformer for mask-based lensless imaging. *Opt. Lett.*, 47(7):1843–1846, Apr 2022.

[54] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[55] Ilya Reshetouski, Hideki Oyaizu, Kenichiro Nakamura, Ryuta Satoh, Suguru Ushiki, Ryuichi Tadano, Atsushi Ito, and Jun Murayama. Lensless imaging with focusing sparse ura masks in long-wave infrared and its application for human detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, pages 237–253. Springer, 2020.

[56] Ilya Reshetouski, Ryuichi Tadano, Hideki Oyaizu, Kenichiro Nakamura, and Jun Murayama. Lensless mismatched aspect ratio imaging. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2021.

[57] Yinger Zhang, Zhouyi Wu, Peiying Lin, Yuting Wu, Lusong Wei, Zhengjie Huang, and Jiangtao Huangfu. Text detection and recognition based on a lensless imaging system. *Appl. Opt.*, 61(14):4177–4186, May 2022.

[58] Jasper Tan, Li Niu, Jesse K. Adams, Vivek Boominathan, Jacob T. Robinson, Richard G. Baraniuk, and Ashok Veeraraghavan. Face detection and verification using lensless cameras. *IEEE Transactions on Computational Imaging*, 5(2):180–194, 2019.

[59] Fangyu Liu, Jiachen Wu, and Liangcai Cao. Autofocusing of fresnel zone aperture lensless imaging for qr code recognition. *Opt. Express*, 31(10):15889–15903, May 2023.

[60] Gun-Yeal Lee, Jangwoon Sung, and Byoungho Lee. Recent advances in metasurface hologram technologies. *ETRI Journal*, 41(1):10–22, 2019.

[61] D. H. Close. Holographic Optical Elements. *Optical Engineering*, 14(5):145408, 1975.

[62] Yichen Wu, Yair Rivenson, Yibo Zhang, Zhensong Wei, Harun Günaydin, Xing Lin, and Aydogan Ozcan. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica*, 5(6):704–710, Jun 2018.

[63] William C. Sweatt. Describing holographic optical elements as lenses. *J. Opt. Soc. Am.*, 67(6):803–808, Jun 1977.

[64] Maria Antonietta Ferrara, Valerio Striano, and Giuseppe Coppola. Volume holographic optical elements as solar concentrators: An overview. *Applied Sciences*, 9(1):193, 2019.

[65] DH Close. Holographic optical elements. *Optical Engineering*, 14(5):408–419, 1975.

[66] Keehoon Hong, Jiwoon Yeom, Changwon Jang, Jisoo Hong, and Byoungho Lee. Full-color lens-array holographic optical element for three-dimensional optical see-through augmented reality. *Opt. Lett.*, 39(1):127–130, Jan 2014.

[67] Changwon Jang, Chang-Kun Lee, Jinsoo Jeong, Gang Li, Seungjae Lee, Jiwoon Yeom, Keehoon Hong, and Byoungho Lee. Recent progress in see-through three-dimensional displays using holographic optical elements. *Appl. Opt.*, 55(3):A71–A85, Jan 2016.

[68] Ichiro Kasai, Yasushi Tanijiri, Takeshi Endo, and Hiroaki Ueda. A practical see-through head mounted display using a holographic optical element. *Opt. Rev*, 8:241–244, 2001.

[69] Han-Ju Yeom, Hee-Jae Kim, Seong-Bok Kim, HuiJun Zhang, BoNi Li, Yeong-Min Ji, Sang-Hoo Kim, and Jae-Hyeung Park. 3D holographic head mounted display using holographic optical elements with astigmatism aberration compensation. *Opt. Express*, 23(25):32025–32034, 2015.

[70] Hee-Jae Kim, Sung-Keun Lee, Mei-Lan Piao, Nam Kim, and Jae-Hyeung Park. Three-dimensional holographic head mounted display using holographic optical element. In *Proc. International Conference on Consumer Electronics (ICCE)*, pages 132–133. IEEE, 2015.

[71] Yaakov Amitai. P-21: Extremely compact high-performance hmds based on substrate-guided optical element. In *SID Symposium Digest of Technical Papers*, volume 35, pages 310–313, 2004.

[72] Byoungho Lee, Kiseung Bang, Minseok Chae, and Chanhyung Yoo. Holographic optical elements for head-up displays and near-eye displays. *Proc. SPIE*, 11708:1170803, 2021.

[73] Jiasheng Xiao, Juan Liu, Zhenlv Lv, Xueliang Shi, and Jian Han. On-axis near-eye display system based on directional scattering holographic waveguide and curved goggle. *Opt. Express*, 27(2):1683–1692, Jan 2019.

[74] Koki Wakunami, Po-Yuan Hsieh, Ryutaro Oi, Takanori Senoh, Hisayuki Sasaki, Yasuyuki Ichihashi, Makoto Okui, Yi-Pai Huang, and Kenji Yamamoto. Projection-type see-through holographic three-dimensional display. *Nat. Commun.*, 7(12954), 2016.

[75] Hanle Zhang, Huan Deng, Minyang He, Dahai Li, and Qionghua Wang. Dual-view integral imaging 3d display based on multiplexed lens-array holographic optical element. *Applied Sciences*, 9(18), 2019.

[76] Changgeng Liu, Beatrice Pazzucconi, Juan Liu, Lei Liu, and Xincheng Yao. A holographic waveguide based eye tracker. *Proc. SPIE*, 10474:104741T, 2018.

[77] Qiang Huang and H. John Caulfield. Waveguide holography and its applications. *Proc. SPIE*, 1461:303 – 312, 1991.

[78] Marvin D. Drake, Mark L. Lidd, and Michael A. Fiddy. Waveguide hologram fingerprint entry device. *Optical Engineering*, 35(9):2499 – 2505, 1996.

[79] Hui-Ying Wu, Chang-Won Shin, and Nam Kim. Full-color holographic optical elements for augmented reality display. In *Holographic Materials and Applications*. IntechOpen, 2019.

[80] Minniu Zhou, Osamu Matoba, Yoichi Kitagawa, Yukako Takizawa, Tetsuya Matsumoto, Hideaki Ueda, Akio Mizuno, and Nobuyuki Kosaka. Fabrication of an integrated holographic imaging element for a three-dimensional eye-gaze detection system. *Appl. Opt.*, 49(19):3780–3785, Jul 2010.

[81] Colton M. Bigler, Micah S. Mann, and Pierre-Alexandre Blanche. Holographic waveguide hud with in-line pupil expansion and 2d fov expansion. *Appl. Opt.*, 58(34):G326–G331, Dec 2019.

[82] Craig T. Draper, Colton M. Bigler, Micah S. Mann, Kalluri Sarma, and Pierre-Alexandre Blanche. Holographic waveguide head-up display with 2-d pupil expansion and longitudinal image magnification. *Appl. Opt.*, 58(5):A251–A257, Feb 2019.

[83] Philippe Coni, Jean-Luc Bardon, Natacha Damamme, Seth Coe-Sullivan, and Fedor I Dimov. 55-3: Holographic grating to improve the efficiency of windshield hud. In *SID Symposium Digest of Technical Papers*, volume 49, pages 729–732. Wiley Online Library, 2018.

[84] Zhenlv Lv, Yuan Xu, Yan Yang, and Juan Liu. Multiplane holographic augmented reality head-up display with a real–virtual dual mode and large eyebox. *Appl. Opt.*, 61(33):9962–9971, Nov 2022.

[85] L Mertz and NO Young. Fresnel transformations of images. *SPIE milestone series ms*, 128:44–49, 1996.

[86] Harrison H Barrett. Fresnel zone plate imaging in nuclear medicine. *Journal of nuclear medicine*, 13(6):382–385, 1972.

[87] Gordon L Rogers. Gabor diffraction microscopy: the hologram as a generalized zone-plate. *Nature*, 166:237–237, 1950.

[88] Tomoya Nakamura, Takuto Watanabe, Shunsuke Igarashi, Xiao Chen, Kazuyuki Tajima, Keita Yamaguchi, Takeshi Shimano, and Masahiro Yamaguchi. Superresolved image reconstruction in fza lensless camera by color-channel synthesis. *Opt. Express*, 28(26):39137–39155, Dec 2020.

[89] Mayu Sao, Yusuke Nakamura, Kazuyuki Tajima, and Takeshi Shimano. Lensless close-up imaging with fresnel zone aperture. *Japanese Journal of Applied Physics*, 57(9S1):09SB05, 2018.

[90] Takeshi Shimano, Yusuke Nakamura, Kazuyuki Tajima, Mayu Sao, and Taku Hoshizawa. Lensless light-field imaging with fresnel zone aperture: quasi-coherent coding. *Applied optics*, 57(11):2841–2850, 2018.

[91] Jiachen Wu, Hua Zhang, Wenhui Zhang, Guofan Jin, Liangcai Cao, and George Barbastathis. Single-shot lensless imaging with fresnel zone aperture and incoherent illumination. *Light: Science & Applications*, 9(1):53, 2020.

[92] Wenbo Wan, Huihui Ma, Zijie Mei, Huilin Zhou, Yuhao Wang, and Qiegen Liu. Multiphase fza lensless imaging via diffusion model. *Opt. Express*, 31(12):20595–20615, Jun 2023.

[93] Ichirou Yamaguchi and Tong Zhang. Phase-shifting digital holography. *Opt. Lett.*, 22(16):1268–1270, Aug 1997.

[94] Xiao Chen, Tomoya Nakamura, Xiuxi Pan, Kazuyuki Tajima, Keita Yamaguchi, Takeshi Shimano, and Masahiro Yamaguchi. Resolution improvement in fza lens-less camera by synthesizing images captured with different mask-sensor distances. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2808–2812, 2021.

[95] Xiao Chen, Xiuxi Pan, Tomoya Nakamura, Saori Takeyama, Takeshi Shimano, Kazuyuki Tajima, and Masahiro Yamaguchi. Wave-optics-based image synthesis for super resolution reconstruction of a fza lensless camera. *Opt. Express*, 31(8):12739–12755, Apr 2023.

[96] Masuyoshi Yachida, Nagaaki Ohyama, and Toshio Honda. Image restoration using synthetic image processing method. *Optics Communications*, 74(1):5–9, 1989.

[97] S. Liliek Suryani, Masahiro Yamaguchi, Nagaaki Ohyama, Toshio Honda, and Kazuyoshi Tanaka. Estimation of an image sampled by a ccd sensor array using a color synthetic method. *Optics Communications*, 84(3):133–138, 1991.

[98] Nagaaki Ohyama, Masuyoshi Yachida, Eric Badique, Jumpei Tsujiuchi, and Toshio Honda. Least-squares filter for color-image restoration. *J. Opt. Soc. Am. A*, 5(1):19–24, Jan 1988.

[99] Masuyoshi Yachida, Nagaaki Ohyama, and Toshio Honda. Color image restoration using synthetic method. *Optics Communications*, 72(1):22–26, 1989.

[100] Kazuyuki Tajima, Yusuke Nakamura, Keita Yamaguchi, and Takeshi Shimano. Improving resolution of lensless imaging with higher harmonics of fresnel zone aperture. *Optical Review*, 29(2):153–158, 2022.

[101] Wenhui Zhang, Liangcai Cao, David J. Brady, Hua Zhang, Ji Cang, Hao Zhang, and Guofan Jin. Twin-image-free holography: A compressive sensing approach. *Phys. Rev. Lett.*, 121:093902, Aug 2018.

[102] Katsuki Inoue, Sunao Hara, Masanobu Abe, Nobukatsu Hojo, and Yusuke Ijima. An investigation to transplant emotional expressions in dnn-based tts synthesis. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1253–1258. IEEE, 2017.

[103] Hieu-Thi Luong and Junichi Yamagishi. Scaling and bias codes for modeling speaker-adaptive dnn-based speech synthesis systems. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 610–617. IEEE, 2018.

[104] Yichen Wu, Yair Rivenson, Hongda Wang, Yilin Luo, Eyal Ben-David, Laurent A Bentolila, Christian Pritz, and Aydogan Ozcan. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nature methods*, 16(12):1323–1331, 2019.

[105] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[106] Mark J. Huiskes and Michael S. Lew. The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, MIR '08, page 39–43, New York, NY, USA, 2008. Association for Computing Machinery.

[107] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[108] Marcos V. Conde, Florin Vasluianu, Sabari Nathan, and Radu Timofte. Real-time under-display cameras image restoration and hdr on mobile devices, 2022.

[109] Kinam Kwon, Eunhee Kang, Sangwon Lee, Su-Jin Lee, Hyong-Euk Lee, ByungIn Yoo, and Jae-Joon Han. Controllable image restoration for under-display camera in smartphones. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2073–2082, 2021.

[110] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9179–9188, 2021.

[111] Zhibin Wang, Yilu Chang, Qi Wang, Yingjie Zhang, Jacky Qiu, and Michael Helander. 55-1: Invited paper: Self-assembled cathode patterning in amoled for under-display camera. In *SID Symposium Digest of Technical Papers*, volume 51, pages 811–814. Wiley Online Library, 2020.

[112] Hailiang Wang, Yan Lin, Yaying Li, Yan Yang, Ting Zhou, Ruxing Chen, Yihua Zhu, Binbin Chen, and Junyi Li. P-132: An under-display camera optical structure for full-screen lcd. In *SID Symposium Digest of Technical Papers*, volume 51, pages 1881–1882. Wiley Online Library, 2020.

[113] Qian Yang, Zhiyong Yang, Yi-Fen Lan, and Shin-Tson Wu. Low-diffraction transparent micro light-emitting diode displays with optimized pixel structure. *Journal of the Society for Information Display*, 30(5):395–403, 2022.

[114] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 662–671, 2021.

[115] Shinji Kimura, Eriko Ooseki, Yuji Aburakawa, and Masahiro Yamaguchi. Evaluation and formulation of the sense of social telepresence in video-mediated communication systems: Contribution of eye contact to enhancing social telepresence. *Journal of the Society for Information Display*, 29(3):179–195, 2021.

[116] Fumino Matsui, Fumiaki Watanabe, Tomoya Nakamura, and Masahiro Yamaguchi. Unmixing of the background components in an off-axis holographic-mirror-based imaging system using spectral image processing. *Optics Express*, 28(26):39998–40012, 2020.

[117] Claudia Kuster, Tiberiu Popa, Jean-Charles Bazin, Craig Gotsman, and Markus Gross. Gaze correction for home video conferencing. *ACM Trans. Graph.*, 31(6), nov 2012.

[118] Fumiaki Watanabe, Tomoya Nakamura, Shiho Torashima, Shunsuke Igarashi, Shinji Kimura, Yuji Aburakawa, and Masahiro Yamaguchi. Dispersion compensation for full-color virtual-imaging systems with a holographic off-axis mirror. In Hans I. Bjelkhagen, editor, *Practical Holography XXXIV: Displays, Materials, and Applications*, volume 11306, page 1130604. International Society for Optics and Photonics, SPIE, 2020.

[119] Kazuhiro Otsuka. Mmspace: Kinetically-augmented telepresence for small group-to-group conversations. In *2016 IEEE Virtual Reality (VR)*, pages 19–28. IEEE, 2016.

[120] Philip V Harman. Autostereoscopic teleconferencing system. In *Stereoscopic Displays and Virtual Reality Systems VII*, volume 3957, pages 293–302. SPIE, 2000.

[121] Yevgenia Bondareva, Lydia Meesters, and Don Bouwhuis. Eye contact as a determinant of social presence in video communication. In *Proceedings of the 20th International Symposium on Human Factors in Telecommunication*, 2006.

[122] Ganghun Kim and Rajesh Menon. Computational imaging enables a “see-through” lens-less camera. *Opt. Express*, 26(18):22826–22836, Sep 2018.

[123] Tomoya Nakamura, Shinji Kimura, Kazuhiko Takahashi, Yuji Aburakawa, Shunsuke Takahashi, Shunsuke Igarashi, Shiho Torashima, and Masahiro Yamaguchi. Off-axis virtual-image display and camera by holographic mirror and blur compensation. *Opt. Express*, 26(19):24864–24880, Sep 2018.

[124] Shinji Kimura, Yuji Aburakawa, Fumiaki Watanabe, Shiho Torashima, Shunsuke Igarashi, Tomoya Nakamura, and Masahiro Yamaguchi. Holographic video communication system realizing virtual image projection and frontal image capture. *ITE Transactions on Media Technology and Applications*, 9(1):105–112, 2021.

[125] Howard Smith. *Principles of Holography*. John Wiley and Sons Publishing, 1969.

[126] John N. Latta. Computer-based analysis of holography using ray tracing. *Appl. Opt.*, 10(12):2698–2710, Dec 1971.

[127] Arno Klein. Relating vector ray-tracing equations for holograms of arbitrary shape and thickness. *J. Opt. Soc. Am. A*, 25(4):979–982, Apr 2008.

Appendix A

Achievement List

Directly related to this thesis

Journal Papers

1. **Xiao Chen**, Xiuxi Pan, Tomoya Nakamura, Saori Takeyama, Takeshi Shimano, Kazuyuki Tajima, and Masahiro Yamaguchi, "Wave-optics-based image synthesis for super resolution reconstruction of a FZA lensless camera," *Optics Express*, Vol. 31, No. 8, pp. 12739-12755 (2023).

2. **Xiao Chen**, Noriyuki Tagami, Hiroki Konno, Tomoya Nakamura, Saori Takeyama, Xiuxi Pan, and Masahiro Yamaguchi, "Computational see-through screen camera based on a holographic waveguide device," *Optics Express*, Vol. 30, No. 14, pp. 25006-25019 (2022).

Conference Presentation

1. **Xiao Chen**, Xiuxi Pan, Tomoya Nakamura, Kazuyuki Tajima, Keita Yamaguchi, Takeshi Shimano, and Masahiro Yamaguchi, "Resolution enhancement in FZA lensless camera using images captured with different mask-sensor distances," Optics & Photonics Japan 2020, 日本光学会年次学術講演会 予稿集 15aAJ4 , online (2020).

2. **Xiao Chen**, Tomoya Nakamura, Xiuxi Pan, Kazuyuki Tajima, Keita Yamaguchi, Takeshi Shimano, and Masahiro Yamaguchi, "Resolution Improvement in FZA Lens-Less Camera by Synthesizing Images Captured with Different Mask-Sensor Distances," *2021 IEEE International Conference on Image Processing (ICIP 2021)*, Anchorage, USA, pp.2808-2812 (2021).

3. **Xiao Chen**, Xiuxi Pan, Tomoya Nakamura, Saori Takeyama, and Masahiro Yamaguchi, "A wave-optics based superresolution method in FZA lensless camera by image synthesis from different mask patterns," Optics & Photonics Japan 2022, 日本光学会年次学術講演会 予稿集 14aBJ3 , Utsunomiya, Japan (2022).

Award

1. Optics & Photonics Japan 2022 第8回 OPJ 優秀講演賞
   https://opt-j.com/opj2022/award.html

Others

Journal Papers

1. Yuuki Sugawara, **Xiao Chen**, Ryuusei Higuchi, Takeo Yamaguchi, "Machine learning-aided unraveling of the importance of structural features for the electrocatalytic oxygen evolution reaction on multimetal oxides based on their A-site metal configurations," *Energy Advances* (2023)

2. Xiuxi Pan, **Xiao Chen**, Saori Takeyama, and Masahiro Yamaguchi, "Image reconstruction with transformer for mask-based lensless imaging," *Optics Letters*, Vol. 47, No. 7, pp. 1843-1846 (2022).

3. Xiuxi Pan, **Xiao Chen**, Tomoya Nakamura, and Masahiro Yamaguchi, "Incoherent reconstruction-free object recognition with mask-based lensless optics and the Transformer," *Optics Express*, Vol. 29, No. 23, pp. 37962-37978 (2021).

4. Xiuxi Pan, Tomoya Nakamura, **Xiao Chen**, and Masahiro Yamaguchi, "Lensless inference camera: object recognition through a thin mask with LBP map generation," *Optics Express*, Vol. 29, No. 7, pp. 9758-9771 (2021).

5. Tomoya Nakamura, Takuto Watanabe, Shunsuke Igarashi, **Xiao Chen**, Kazuyuki Tajima, Keita Yamaguchi, Takeshi Shimano, and Masahiro Yamaguchi, "Superresolved image reconstruction in FZA lensless camera by color-channel synthesis," *Optics Express*, Vol. 28, No. 26, pp. 39137-39155 (2020).

Conference Presentation

1. Xiuxi Pan, Tomoya Nakamura, **Xiao Chen**, and Masahiro Yamaguchi, "Resolution enhancement in FZA lens-less camera using images captured with different mask-sensor distances," Optics & Photonics Japan 2020, 日本光学会年次学術講演会 予稿集 15aAJ3 , online (2020).

2. Xiuxi Pan, **Xiao Chen**, Tomoya Nakamura, Saori Takeyama, and Masahiro Yamaguchi, "マスクを用いたレンズレス光学系のためのTransformerニューラルネットワークによる非干渉かつ再構成不要な物体認識," Optics & Photonics Japan 2021, 日本光学会年次学術講演会 予稿集 27pAS8 , online (2021).

3. Xiuxi Pan, **Xiao Chen**, Saori Takeyama, and Masahiro Yamaguchi. "Design of optically extended convolutional neural network," *The 13th International Conference on Optics-photonics Design & Fabrication* , ODF'22 Technical Digest, P-OTh-35, Sapporo, Japan (2022).

4. Xiuxi Pan, **Xiao Chen**, Saori Takeyama, and Masahiro Yamaguchi. "Lensless Imaging and Recognition with Transformer-based Neural Networks," *Information Photonics 2022 (IP2022)*, IP3-04, Yokohama, Japan (2022).

5. 山口雅浩，Xiuxi Pan, **Xiao Chen**，高橋京司, 武山彩織，"高度な計算アルゴリズムとの融合が可能にする新しい光画像センシング技術," レーザー学会学術講演会 第43回年次大会, レーザー学会学術講演会 第43回年次大会 予稿集, S04-19p-I-05 , Nagoya, Japan (2022).

6. 山口雅浩，**Xiao Chen**，田上典幸，小原一真, 高橋京司, 武山彩織，"ホログラフィック導波路デバイスを用いたディスプレイ表面透明カメラ," Optics & Photonics Japan 2022, 日本光学会年次学術講演会 予稿集 15aDS3 , Utsunomiya, Japan (2022).

7. Masahiro Yamaguchi, Noriyuki Tagami, **Xiao Chen**, Hiroki Konno, and Tomoya Nakamura, "Computational see-through screen camera using a holographic waveguide device," *OSA Imaging and Applied Optics Congress*, DF2F.6, online (2021).

8. 山口雅浩，田上典幸，**Xiao Chen**，中村友哉，武山彩織，“ホログラフィック導波路デバイスを用いたディスプレイ表面透明カメラ,” 電子情報通信学会研究会（画像技術，VR/AR，ヒューマンファクター関連一般), online (2021).