All Photons Imaging

Time-Resolved Computational Imaging Through Scattering for Vehicles and Medical Applications with Probabilistic and Data-Driven Algorithms

by

Guy Satat

B.Sc., Technion, Israel Institute of Technology (2013)
S.M., Massachusetts Institute of Technology (2015)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Author .....
Program in Media Arts and Sciences,
School of Architecture and Planning,
May 3, 2019

Certified by .....
Ramesh Raskar
Associate Professor
Thesis Supervisor

Accepted by .....
Tod Machover
Academic Head
Program in Media Arts and Sciences

All Photons Imaging

Time-Resolved Computational Imaging Through Scattering for Vehicles and Medical Applications with Probabilistic and Data-Driven Algorithms

by

Guy Satat

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 3, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

One of the greatest challenges in computational imaging is scaling it to work outside the lab. The main reasons for that challenge are the strong dependency on precise calibration, accurate physical models, and long acquisition times. These prevent practical progress towards medical imaging and seeing through occlusions such as fog in the wild. This dissertation demonstrates that with data-driven and probabilistic modeling we can alleviate these dependencies, and pave the way towards real-world time-resolved computational imaging through extreme scattering conditions using visible light.

The ability to image through scattering media in the visible part of the electromagnetic spectrum holds many applications in various industries. For example, seeing through fog would enable autonomous robots to operate in challenging weather conditions; augment human driving; and allow airplanes, helicopters, and drones to take off and land in dense fog conditions. In medical imaging, the ability to see into the body with near-infrared light would reduce the exposure to ionizing radiation and provide more clinically meaningful data.

In order to image in diverse and extreme scattering conditions, we develop novel algorithms inspired by techniques in signal processing, optimization, statistical analysis, compressive sensing, and machine learning that leverage time-resolved sensing. More specifically, we demonstrate techniques that computationally leverage all of the optical signal, including scattered light, as opposed to locking onto a specific part of the optical signal. Furthermore, we show that by introducing probabilistic formulation to the imaging problem, the resulting system does not require user input for calibration and priors; this makes our systems more practical for real-world scenarios and enables them to operate in a wide range of scattering conditions.

We consider four cases of imaging through scattering media with increasing complexity:

1. A theoretical analysis of time-resolved single pixel imaging, which demonstrates scene reconstruction even when the entire scene is measured with a single pixel, an equivalent of simple scattering or a blur that is easy to model.

2. A data-driven calibration invariant technique for imaging through simple scattering (a sheet of paper).

3. Imaging through a thick tissue phantom by utilizing all of the optical signal with minimal assumptions on the tissue properties.

4. Imaging through a wide range of dense, dynamic, and heterogeneous fog conditions. In that case, we introduce a probabilistic model that is able to recover the occluded target reflectance and depth without any assumption about the fog.

Thesis Supervisor: Ramesh Raskar
Title: Associate Professor

All Photons Imaging

Time-Resolved Computational Imaging Through Scattering for Vehicles and Medical Applications with Probabilistic and Data-Driven Algorithms

by

Guy Satat

Thesis Advisor .....

Ramesh Raskar

Associate Professor in Media Arts and Sciences

MIT Media Arts and Sciences

Thesis Reader .....

Joseph A. Paradiso

Alexander W. Dreyfoos (1954) Professor in Media Arts and Sciences

MIT Media Arts and Sciences

Thesis Reader .....

Fadel Adib

Assistant Professor in Media Arts and Sciences

MIT Media Arts and Sciences

Acknowledgments

I first joined Professor Ramesh Raskar's group as a visiting student seven years ago. Before the visit, I did not imagine what a tremendous effect it will have on my life. During my short visit, I learned so much from Ramesh and it was clear to me that it wasn't enough. The visit led to me joining Ramesh's group for a Masters followed by a PhD. Ramesh always inspired me to work on the important and hard problems first. I am grateful for his willingness to teach, mentor and guide me during the last seven years.

I thank my committee. Joe Paradiso has been a part of my academic journey in the Media Lab since I joined. He has always inspired me to think about interesting and new ways to leverage physical interpretation and time-resolved sensing in my work. I have learned a lot from countless interactions and conversations with Fadel Adib. My work is inspired by Fadel's rigorous work in imaging with RF.

Special gratitude to Moungi Bawendi. Moungi opened his lab for me, and almost all the experimental work presented in this thesis was done using hardware from the Bawendi Lab. Moungi's generosity allowed me to pursue my research goals.

I thank Petros Boufounos. Petros was my professor in a class on information theory at MIT which was the inspiration and basis to the FemtoPixel work presented here (Chapter 3).

Many mentors have been instrumental to my work and growth (most of them from the Camera Culture Group). Christopher Barsi and Barmak Heshmat taught me all I know about optics and working with ultrafast optics. Dan Raviv taught me the basics of optimization and machine learning which were the foundation for my research. Pratik Shah has been a valuable mentor who inspired and encouraged me to step out of my comfort zone. Micha Feigin was a reliable mentor on anything related to electronics. Gili Bisker has helped me shape my ideas on imaging through scattering. John Werner was an amazing resource and a source of inspiration in countless ways.

I thank my collaborators at the University of Glasgow (previously at Heriot-Watt

University) — Daniele Faccio, Gabriella Musarra, Ashley Lyons, and Alessandro Boccolini, for inviting and hosting me at Heriot-Watt University, and for helping with the experimental demonstration of FemtoPixel.

I am profoundly thankful to Matthew Tancik. Matt joined the Camera Culture group as an undergraduate student and quickly became an integral part of the group and my research. I was fortunate to have Matt as a collaborator and am grateful for everything I learned from him. Matt was a key contributor to many of the results presented here.

I thank all of my co-authors: Christopher Barsi, Moungi Bawendi, Ou Chen, Daniele Faccio, Otkrist Gupta, Barmak Heshmat, Sachin Katti, Manikanta Kotaru, Ashley Lyons, Gabriella Musarra, Nikhil Naik, Krithika Ramchander, Ramesh Raskar, Dan Raviv, Albert Redo-Sanchez, Tristan Swedish, and Matthew Tancik.

~

I am grateful for my family. My grandparents David, Raizi, Feri, and Monzi have dedicated their lives so that their children and grandchildren will be able to pursue a better education. My parents, Aviva and Kobi, have been an incredible role model for me. I am grateful for their love, support, and education. I can only hope to be at least as good as a parent as they were and are to me. My sister, Gili, and brother-in-law, Ido, have always been wonderful friends, I am thankful for their love and support.

Most of all I am thankful to my wife, Talia. Words cannot express my gratitude for her support, encouragement, friendship, and partnership during our mutual journey in the last few years. Talia has been an active part of my academic career and reviewed every paper I submitted. I also have to thank my son, Ben, who never fails to remind me that there is always time to play.

Previous Published Materials

This thesis revises the following publications:

Chapter 3 revises:

[149] Guy Satat, Matthew Tancik, and Ramesh Raskar, "Lensless imaging with compressive ultrafast sensing," *IEEE Transactions on Computational Imaging*, 2017.

[147] Guy Satat, Gabriella Musarra, Ashley Lyons, Barmak Heshmat, Ramesh Raskar, and Daniele Faccio, "Compressive Ultrafast Single Pixel Camera," *OSA Computational Optical Sensing and Imaging*, 2018.

Chapter 4 revises:

[148] Guy Satat, Matthew Tancik, Otkrist Gupta, Barmak Heshmat, and Ramesh Raskar, "Object classification through scattering media with deep learning on time resolved measurement," *Optics express*, 2017.

Chapter 5 revises:

[145] Guy Satat, Barmak Heshmat, Dan Raviv, and Ramesh Raskar, "All photons imaging through volumetric scattering," *Nature Scientific Reports*, 2016.

[144] Guy Satat, Barmak Heshmat, and Ramesh Raskar, "All Photons Imaging Through Layered Scattering Materials," *OSA Propagation Through and Characterization of Atmospheric and Oceanic Phenomena*, 2017.

Chapter 6 revises:

[151] Guy Satat, Matthew Tancik, and Ramesh Raskar, "Towards Photography Through Realistic Fog," *IEEE International Conference on Computational Photography*, 2018.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this dissertation we develop probabilistic and data-driven algorithms that leverage statistics of scattered photons. These algorithms tackle, by design, the dependency of computational imaging on highly calibrated and accurate physical models as well as long acquisition times.

The main crutches of computational imaging are the dependency on an accurate and calibrated physical model, and long acquisition times. This is a result of the sensitivity of inverse problems to model mismatch and poor signal to noise ratio (SNR). This prevents computational imaging through scattering with visible light from scaling to real-world applications.

Imaging through scattering media with visible light is a great challenge with many potential applications. Fundamentally, the scattering invalidates basic imaging conditions, and as a result, our eyes and cameras cannot focus light from objects that are occluded by the scattering media. Some examples of materials that are mostly scattering visible light include fog and tissue (Fig. 1-1). While simply seeing through such highly scattering materials with our bare eyes is impossible, we show that it is possible using computational imaging.

Here, we demonstrate, in four different regimes of scattering, that probabilistic and data-driven algorithms along with time-resolved sensing can alleviate these challenges. Table 1.1 provides a brief overview of the different scattering conditions con-

	Scattering Complexity	Number of Scattering Events	Model and Assumptions	Geometry	Approach	Main Contributions
No lens (Chapter 3)	Low	1	Simple blur due to lensless imaging	Reflection	Compressive sensing	• Single pixel time-resolved imaging. • Optimized compressive sensing. • Efficient acquisition.
Paper (Chapter 4)	Medium	2	Single scatter event	Reflection	Data-driven	• Data-driven computational imaging. • Calibration invariant imaging. • Train on simulation, test in reality.
Tissue (Chapter 5)	High	Hundreds	Volumetric scattering, layered material	Transmission	Blind deconvolution	• Imaging through scattering with all photons. • Scattering estimation, independent of target. • Support layered materials.
Fog (Chapter 6)	Extremely high	Hundreds	Arbitrary volumetric scattering (dynamic, dense, heterogeneous)	Reflection	Probabilistic modeling	• Probabilistic model and inversion. • Separation between fog background and hidden scene signal. • Prior knowledge and calibration aren't needed.

Table 1.1: Scattering conditions considered in this dissertation. Color indicates the complexity of the considered scenario. The reflection geometries considered in the top two rows are easier compared to the bottom row, since there is no back reflectance due to the scattering.

sidered here, along with the main challenges and the contributions in these regimes. Specifically, we consider (in increasing complexity) the following conditions:

1. Imaging with a single pixel and without a lens. We provide theoretical analysis for the case of a single pixel that accumulates contributions from all scene points, an equivalent to simple scattering. The presented framework enables acquisition times that are as $50\times$ faster compared to prior techniques.

2. Imaging through paper. We demonstrate a data-driven solution for imaging through simple scattering (a single scatter event). The technique provides the ability to recover the pose of an object hidden behind the scattering layer without needing precise calibration or an accurate model and works in real-time.

3. Imaging through tissue phantom. Here we introduce the concept of using the entire optical signal (All Photons Imaging) to see through scattering. The presented technique resolves a target through 15 mm tissue phantom at 5.9 mm resolution with minimal assumptions about the scattering material.

4. Imaging through fog. We tackle the challenge of back reflectance from scattering media in optical reflection mode. In this case, we need to separate between the

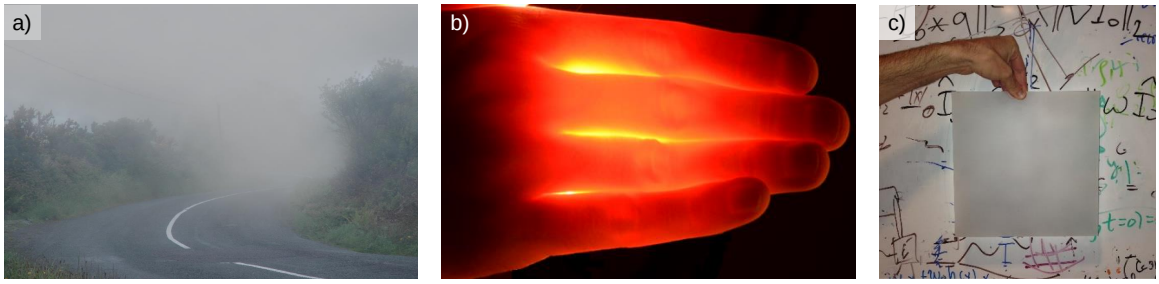

Figure 1-1: Examples of scattering media in visible light: a) fog, b) tissue, and c) optical diffuser.

photons that back-reflect from the fog, and those that hit the target. The solution is based on a probabilistic imaging framework and is demonstrated in realistic fog scenarios (dense, dynamic, and heterogeneous) with visibility as low as 30 cm.

1.1 Why Imaging Through Scattering With Visible Light?

The ability to see through scattering media holds many applications:

- Seeing through fog enables driver augmentation and robust autonomous navigation in degraded weather conditions. It also improves the safety and robustness of flying platforms (airplanes, helicopters, and drones) in reduced visibility at low-level flight.

- Imaging through tissue with visible light allows for non-invasive sensing inside the body with non-ionizing radiation, and potentially better functional imaging when compared to other modalities such as ultrasound and MRI.

- Underwater imaging is commonly challenging due to water turbulence as well as the presence of various particles in the water such as sand and plankton. The ability to image underwater would improve the safety and monitoring of underwater installations such as offshore wind farms, gas, oil, and communication infrastructure.

- Many human–computer interaction systems require a camera that observes the human, which brings up many privacy issues. It is beneficial to capture the essential information for the interaction while corrupting the input to preserve user privacy. This is possible with scattering media such as a sheet of paper or an optical diffuser to obstruct the camera's field of view.

- Non-line of sight (NLOS) imaging is the general ability to capture information that is not directly observable by the camera. Examples of this include lensless imaging and seeing around corners. Lensless imaging has many important applications in general purpose sensing when the use of a high-quality lens with a large aperture is not possible.

These capabilities extend the concept of a camera to a general purpose sensor coupled with an algorithm to reconstruct the desired information.

Traditionally, overcoming optical scattering is achieved by moving away from the visible part of the spectrum. Examples include radio frequency (RF) for imaging through fog, and x-ray and ultrasound for seeing into the body. However, the use of visible light for imaging holds many advantages:

- Optical contrast – the ability to distinguish among different materials. With visible light, a small change in wavelength results in a significant change in the interaction with different atoms and molecules. Thus, imaging in visible light enables many applications in medical imaging; for example, distinguishing between different types of tissue that is very hard or impossible to achieve with other imaging modalities. Another application is the ability to read road signs in fog since the color on road signs has significant different spectral signatures in the visible spectrum.

- Resolution – the optical wavelengths ($400\,\text{nm} - 700\,\text{nm}$) are much smaller when compared to RF ($1\,\text{cm} - 1\,\text{m}$), mm-wave ($1\,\text{mm} - 1\,\text{cm}$), and THz ($100\,\mu\text{m} - 1\,\text{mm}$). As a result it is easier to capture higher resolution images with smaller apertures.

- Non-ionizing radiation – this advantage is primarily important for medical applications. While there is always a tendency to avoid unneeded x-ray exposure, some patients like pregnant women and children cannot be exposed to x-ray. Medical imaging with visible light does not suffer from these limitations.

- Availability of fluorophores – these are commonly used in functional medical imaging to improve the contrast with different target tissues. Such fluorophores are more commonly available and are easier to design in the visible spectrum, as opposed to x-ray. Thus functional medical imaging is potentially more informative in the visible spectrum.

1.2 Why Computational Imaging Through Scattering Is Sensitive to Accurate Modeling and Calibration?

Traditionally, computational imaging through scattering media requires solving an inverse problem. In most cases, it is relatively easy to formalize a forward model which is a mapping from scene to measurement. However, the required task is a mapping from measurement to scene, hence the name "inverse problem".

The challenge in solving the inverse problem is the assumption that the forward model is accurate. An inaccurate forward model can be a result of:

- **Model Mismatch:** In that case the forward model does not properly account for the physics of the problem; therefore it is impossible to explain the raw measurement by the forward model, for example – assuming a specific physical model for the scattering, that does not capture the full complexity of the scattering process.

- **Calibration:** The forward model usually includes many physical constants that describe the imaging system and physics of the problem (for example, the camera field-of-view or scattering media thickness). Calibrating a forward

model usually requires extensive effort to accurately measure all parameters and properties.

It is important to note that model mismatch and insufficient calibration can happen simultaneously and to different extents. Calibrating an imaging system also assumes we have full control of the specific parameters that are measured. This assumption may be extremely prohibitive in real-world scenarios.

1.3 Imaging through Scattering — Techniques and Regimes Presented Here

We present techniques that aim to tackle fundamental issues with prior methods to image through scattering. Fundamentally, our goal is to computationally maximize the information extracted from the optical signal. This allows our techniques to operate in a wide range of challenging scattering conditions without calibration. To that end, we develop different algorithms that are inspired by modern statistical analysis, optimization, signal processing, compressive sensing, and machine learning.

A common thread among the techniques presented here is ultrafast time-resolved sensing. By ultrafast we consider a sensor with picosecond time resolution. Since light propagates 0.299 mm in 1 ps, at picosecond time resolution the speed of light is non-negligible. Throughout this dissertation, we show that time-resolved sensing is essential for imaging through scattering.

The scattering regimes considered here can be broadly classified as sparse and volumetric scattering. In sparse scattering, the number of scattering events photons undergo is relatively small. This includes the general problem of lensless imaging and imaging through a sheet of paper. In volumetric scattering, the number of scattering events photons undergo is very large. This includes tissue phantoms and fog.

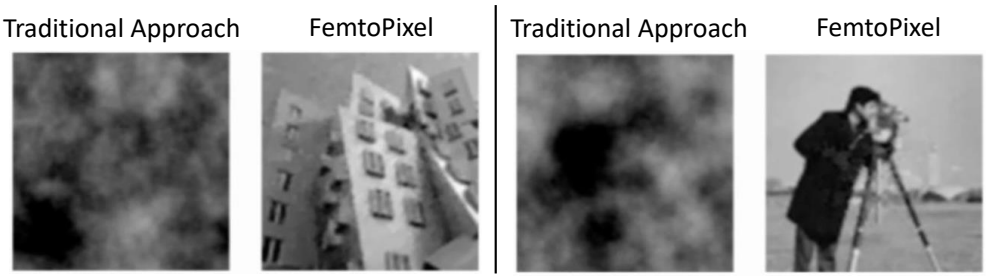

Figure 1-2: FemtoPixel – Lensless single pixel imaging with compressive ultrafast sensing. Two examples with the same acquisition time comparing traditional techniques to FemtoPixel. To achieve similar reconstruction quality to FemtoPixel, the traditional approach (not time-resolved) requires a 50× longer acquisition time.

Lensless Imaging with a FemtoPixel[1]

Chapter 3 presents a compressive sensing framework for time-resolved imaging without a lens, an equivalent to a single scattering event. The presented framework provides design tools for imaging systems. These include analysis tools for the trade-off between system complexity, imaging quality, and acquisition time. Most importantly, by leveraging the physics of time-resolved sensing, we develop an optimization framework for the ideal sensors' position and imaging patterns with compressive sensing. The suggested approach demonstrates 50× faster acquisition time when compared to traditional (not time-resolved) single-pixel cameras (see Fig. 1-2).

Data-Driven Calibration Invariant Imaging Through Paper[2]

One of the biggest challenges in computational imaging is the need to solve an inverse problem due to the sensitivity to accurate physical modeling and calibration. Chapter 4 presents an alternative to solving an inverse problem by using a data-driven approach. In this case, a data-driven algorithm is used to solve a computational imaging problem without calibration. The neural network directly learns a mapping

[1] Abridged versions of this work appeared as [149] "Lensless Imaging with Compressive Ultrafast Sensing," G Satat, M Tancik, and R Raskar, *IEEE Transactions on Computational Imaging*, 2017. And, [147] "Compressive Ultrafast Single Pixel Camera," G Satat, G Musarra, A Lyons, B Heshmat, R Raskar, and D Faccio, *OSA Computational Optical Sensing and Imaging*, 2018.

[2] An abridged version of this work appeared as [148] "Object Classification through Scattering Media with Deep Learning on Time-Resolved Measurement," G Satat, M Tancik, O Gupta, B Heshmat, and R Raskar, *Optics express*, 2017.

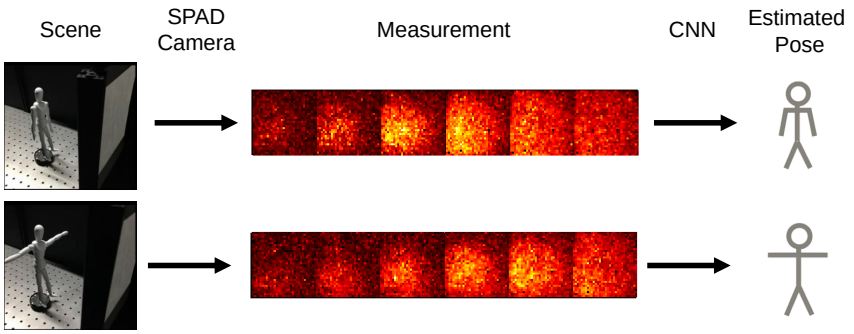

Figure 1-3: Calibration invariant imaging through scattering. A CNN learns to estimate the pose of a mannequin that is occluded by a sheet of paper. Our training technique results in a CNN that is invariant to variations in the system calibration. The network was trained before the experiments took place, and the optical system was built without calibration.

from the measurement to the hidden scene. Thus, there is no need to model the physics of the problem and then inverting the mathematical model. Furthermore, the presented technique is based on a neural network training strategy, which results in a model that is invariant to variations in calibration parameters. We demonstrate this approach on the problem of classifying the pose of a mannequin hidden behind a sheet of paper. In lab experiments, the measurement from a time-resolved single photon avalanche diode (SPAD) camera is fed into a trained convolutional neural network (CNN) which correctly classifies the mannequin poses without any calibration (Fig. 1-3).

Imaging Through Thick Tissue Phantom[3]

Chapter 5 introduces "All Photons Imaging" (API) — a computational technique for imaging through a thick tissue phantom (Fig. 1-4). We introduce a probabilistic interpretation for the imaging system point spread function (PSF). This simplifies the PSF estimation from the measurement itself, and avoids the need to solve an ill-posed blind deconvolution problem. The technique utilizes time-resolved measurements for imaging through volumetric scattering. By using all of the optical signal, including

[3]Abridged versions of this work appeared as [145] "All Photons Imaging Through Volumetric Scattering," G Satat, B Heshmat, D Raviv, and R Raskar, *Nature Scientific Reports*, 2016.
And, [144] "All Photons Imaging through Layered Scattering Materials," G Satat, B Heshmat, and R Raskar, *OSA Computational Optical Sensing and Imaging*, 2017.

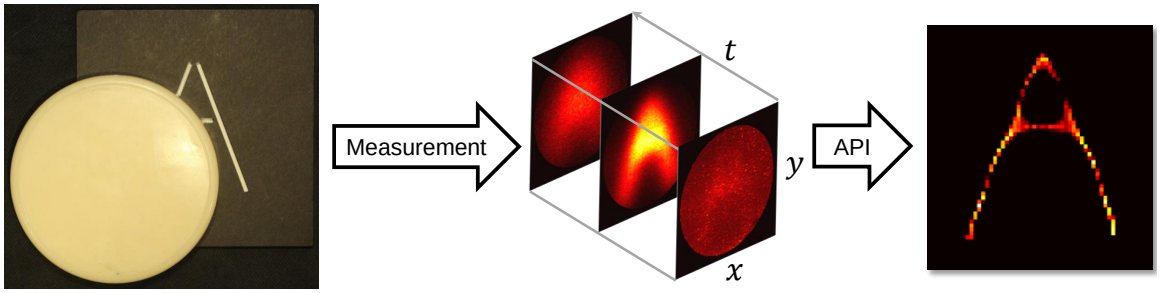

Figure 1-4: All photons imaging through 15 mm thick tissue. API leverages time-resolved measurement to invert the scattering and recover the hidden target.

early arrived (non-scattered) and diffused photons, it achieves better spatial resolution at greater depths. Compared to conventional early photon measurements for imaging through a 15 mm tissue phantom, our method shows a 2× improvement in spatial resolution (4 dB increase in PSNR). Unlike other methods, which aim to lock at a specific part of the optical signal (coherent, ballistic, acoustically modulated, etc.), our framework aims to use all of the optical signal. This results in an all optical, calibration-free imaging system that enables wide field imaging through volumetric scattering at poor SNR and does not require to raster-scan the scene.

Imaging Through Realistic Fog[4]

The technique presented in Chapter 5 was limited to optical transmission mode and to materials that are homogeneous in the $x - y$ plane (perpendicular to the optical axis). Thus it is primarily applicable to medical imaging applications. Chapter 6 introduces a technique that alleviates these constraints and allows imaging through dense, dynamic and heterogeneous fog in optical reflection mode (Fig. 1-5). The technique is demonstrated in a wide range of realistic fog conditions at up to 30 cm visibility. The main challenge that arises in optical reflection mode is the need to separate between the photons that back-reflect from the fog (background) and the photons that hit the target and are then detected by the camera. This is unlike optical transmission mode, in which all of the measured photons interact with the

[4]An abridged version of this work appeared as [151] "Towards Photography Through Realistic Fog," G Satat, M Tancik, and R Raskar, *IEEE International Conference on Computational Photography (ICCP)*, 2018.

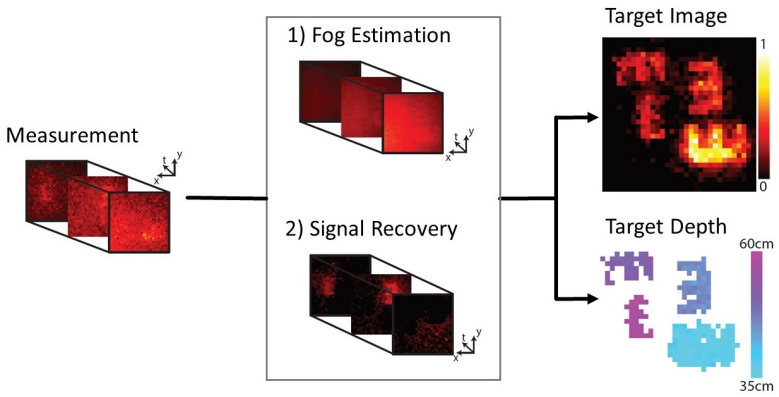

Figure 1-5: Imaging through realistic fog. With a time-resolved single photon detector we separate background photons that back-reflect due to the fog from the signal of the occluded target. We recover the target reflectance and depth in a wide range of realistic, dense, heterogeneous, and dynamic fog conditions.

target. The solution is based on a probabilistic framework that separates, in real time, between background and signal photons, without any assumption about the fog. The extracted signal photons are used to reconstruct the occluded scene reflectance and depth map.

1.4 Main Contributions

This thesis's contributions to the computational imaging literature include:

- **Probabilistic modeling of scattering:**
  - A probabilistic interpretation of the scattering PSF which extends to both modeling and estimation. This interpretation demonstrates, for the first time, probabilistic imaging through thick tissue phantom.
  - A probabilistic model and algorithm for imaging through extremely dense, dynamic, and heterogeneous fog conditions, and recovering both target reflectance and depth.
- **Data-driven computational imaging:**
  - The first data-driven calibration-invariant computational imaging technique. Demonstrated in imaging through sparse scattering.

- A method for training deep learning algorithms with synthetic data for NLOS imaging. Experimentally demonstrated on real data.

- **Time-resolved and single photon sensing for imaging through scattering:**

  - The first framework for lensless imaging with compressive ultrafast sensing. Includes an optimization procedure for computing the optimal sensors' positions and compressive masks.

Our contributions extend beyond the above algorithms and theoretical frameworks to the design, implementation, and *practical evaluation* of imaging through various scattering media including sheet of paper, thick tissue phantoms, and fog, with streak and SPAD cameras.

Chapter 2

Background and Related Works

Imaging through scattering media and occlusions is a long lasting problem. Traditionally it is accomplished in non-visible light modalities such as RF, x-ray, ultrasound, etc. As mentioned in Chapter 1, the advantages of visible light have energized an effort towards imaging through scattering media with visible light. In this chapter, we introduce the basic physics of scattering, including widely accepted models and notations, and introduce different modalities for imaging through scattering media.

2.1 Light-Matter Interaction

All the advantages of imaging with visible light described in the Introduction are due to the interaction between light and matter. This interaction can be classified as a combination of:

- **Absorption:** the photons are absorbed in the material and the energy is converted to heat or photons of lower energy (fluorescence).
- **Scattering:** the photons change their propagation direction as a result of the interaction, but maintain their energy (wavelength).

The case of imaging through highly absorbing materials (for example imaging through a wooden wall with visible light) is somewhat hopeless. This is because the sensor would not capture many, or any, photons. Imaging through scattering media, on the

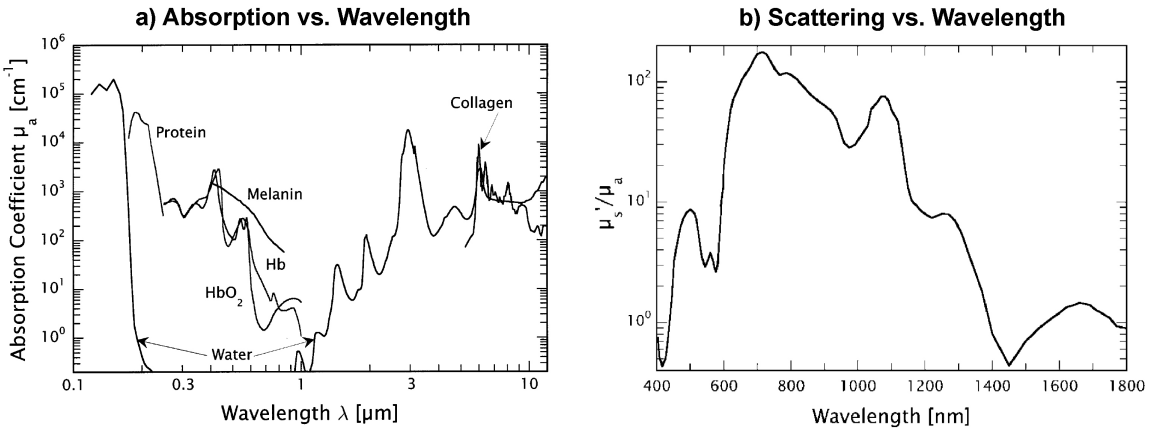

Figure 2-1: Absorption and scattering in biological tissue (figure from Vogel and Venugopalan 2003 [180]). (a) Absorption coefficient and (b) Scattering coefficient as a function of optical wavelength. The scattering coefficient is normalized by the absorption coefficient. Note that the x-axis in the plots is different. Scattering is only characterized in the $400 - 1800\,\text{nm}$ regime where absorption is low.

other hand, is hard yet possible. In the case of scattering, the signal is corrupted, but it is detected by the sensor. This opens the door to computational imaging techniques to de-scatter the signal and reconstruct the occluded scene.

Imaging in scattering conditions is preferred compared to absorbing media, as demonstrated in medical imaging. Fig. 2-1 shows the scattering and absorption coefficient as a function of wavelength in human tissue. The dip in absorption at $800\,\text{nm}$ is considered as the "window into the body" (despite the significant scattering in this regime).

2.2 Scattering and the Effect on Imaging

To evaluate the scattering effect on imaging we use a model inspired by the Plenoptic function [2]. Consider a photon entering into scattering media at position r, angle of propagation ν and time t. The photon will emerge at position r', angle of propagation ν' and time t'. Note that if there was no scattering (i.e. the media occupies zero space) we would have $r = r'$, $\nu = \nu'$, and $t = t'$. On the other hand, if there is scattering r', ν', and t' are different.

In this thesis, we broadly model scattering as sparse and volumetric scattering.

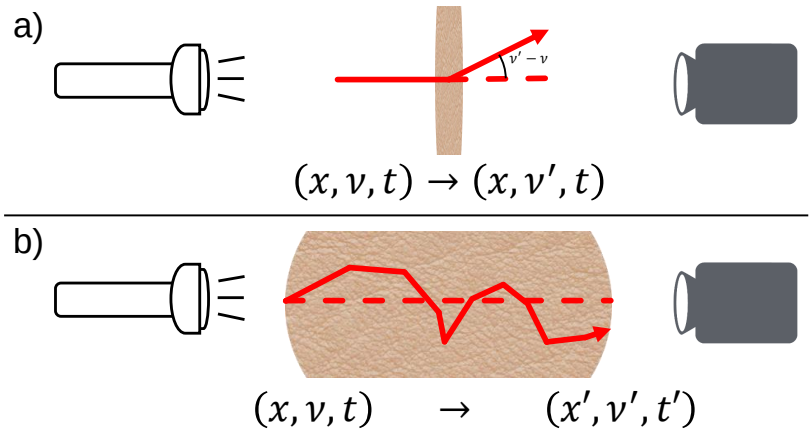

Figure 2-2: a) Sparse and b) Volumetric scattering. The red arrow demonstrates an example of a photon path in these conditions, compared to the dashed line which demonstrates the path without scattering.

- **Sparse scattering:** The scattering process can be modeled as a single scatter event. Because scattering is just a change of angle, we get: $r' = r$, $\nu' \neq \nu$, and $t' = t$. This model is also applicable to multiple scattering events if they are adjacent such that they can effectively be modeled as a single event. Sparse scattering is a good model for a thin scattering layer such as a milky glass or a sheet of paper. Other related problems include looking around corners and lensless imaging. Fig. 2-2a demonstrates sparse scattering.

- **Volumetric scattering:** The photons undergo multiple scattering events in the media. These scattering events are not adjacent, and cannot be modeled as an effective single event. Here, $r' \neq r$, $\nu' \neq \nu$, and $t' \neq t$. This is, of course, a much more challenging situation for imaging since the signal is substantially more corrupt. Volumetric scattering is a good model for scattering media such as fog, tissue, and turbid water. Fig. 2-2b demonstrates volumetric scattering.

2.3 Physics of Scattering

First, we introduce basic terminology in scattering theory. Propagation of photons in scattering media can be considered as a random walk in 3D where the distance a photon propagates between interaction events with the material is random. At each

interaction, the photon can be absorbed or scatter at a random angle.

Phase Function

The phase function describes the relationship between the angle of incoming light to the outgoing light angle. Broadly speaking, it is a function of the wavelength of light and the geometry of the scattering particle. Below we provide a few examples of such phase functions:

- **Mie theory** — takes a very general approach that starts with Maxwell's equations, and assumes scattering spheres. Does not have a closed form solution.

- **Rayleigh scattering** — an approximation of Mie theory when the scattering particles size d is much smaller than the light wavelength λ. In that case, the phase function takes the form:

$$P(\theta) = \frac{3}{16\pi}\left(1 + \cos^2\theta\right) \tag{2.1}$$

  Rayleigh scattering has a strong wavelength dependency of $\sim \lambda^{-4}$ which is commonly known as the reason for blue skies (since shorter wavelengths tend to scatter more). Note that Rayleigh scattering is symmetric with respect to forward and backward scattering.

- **Henyey-Greenstein (HG)** — is a commonly used phase function with a single parameter for anisotropy $g \in [-1, 1]$:

$$P(\theta) = \frac{1 - g^2}{4\pi\left(1 + g^2 - 2g\cos\theta\right)^{1.5}} \tag{2.2}$$

  Here, positive g results in forward scattering and negative g results in backward scattering. $g = 0$ is isotropic scattering (i.e. the photon scatters in all angles with equal probability). When $g \to 1$ there is effectively no scattering since the photons simply continue to propagate in the same angle. It is common to use the HG function in Monte Carlo photon tracers.

Scattering Length Scales

Photons propagate along straight lines between interaction events in the medium. These interactions are characterized by different length scales defined by:

- **Scattering mean free path**, defined as the average distance between two consecutive scattering events in the medium. The scattering coefficient μ_s is defined as the inverse of the scattering mean free path, and has units of $\frac{1}{\text{length}}$.
- **Mean absorption path**, defined as the average distance a photon travels before it is absorbed in the medium. The absorption coefficient μ_a is defined as the inverse of the mean absorption path, and has units of $\frac{1}{\text{length}}$.
- **Mean free path (MFP)**, defined as the distance a photon travels before it is scattered or absorbed. The MFP is defined by its inverse — the transport coefficient μ_t, such that $\mu_t = \mu_a + \mu_s$. When $\mu_s >> \mu_a$ we get $\mu_t \sim \mu_s$, and the mean free path is simply the scattering mean free path.
- **Transport mean free path (TMFP)**, defined as the distance a photon propagates while it undergoes several scattering events and is still correlated to the original direction. It is defined by the inverse μ_s' such that $\mu_s' = \mu_s(1-g)$, where g is the HG anisotropy parameter.

For illustrative purposes we provide some examples for these quantities in fog and tissue:

- **Tissue** is mostly forward scattering with $g \sim 0.85$, MFP $\sim 0.1\,\text{mm}$, and mean absorption path in the range $10 - 100\,\text{mm}$ [120, 184].
- **Fog** describes a wide range of conditions; in general the water droplets geometry determines the phase function. Fog is known to be forward scattering (when modeled with HG function, $g = 0.85$ is a common selection). The MFP in fog varies a lot and is a function of the water droplets' density. For example, in visibility of 100 m we get MFP $\sim 25\,\text{m}$, and in visibility of 20 m we get MFP $\sim 5\,\text{m}$. Similarly to tissue, the absorption in fog is usually negligible [34, 61, 112].

Optical Thickness

Optical thickness (also known as optical depth) is the ratio between the intensity of light incoming to the medium and outgoing from the medium:

$$\mathrm{OT} = -\log \frac{P_{\mathrm{out}}}{P_{\mathrm{in}}} \tag{2.3}$$

The optical thickness is related to the MFP with the Beer-Lambert law:

$$\mathrm{OT} = e^{-\mu_t w} \tag{2.4}$$

where w is the thickness of the medium. This also shows us an easy way to measure μ_t: by taking optical power measurement through different material thicknesses and fitting a line to their log.

In the context of atmospheric scattering such as fog, the visibility is defined by:

$$V = \frac{3.912}{\mu_t} \tag{2.5}$$

The 3.912 factor arises from the visibility definition as a target with a contrast of 0.02.

Thouless Time

Thouless time defines the mean time it takes a photon to propagate from the source to the detector:

$$\tau_T = \frac{3}{c}\frac{w^2}{\mu_s{}'} \tag{2.6}$$

where w is the distance between source and target, and c is the speed of light in the medium. We note that different formulations result in different prefactors. Here the assumptions are a non dispersive medium with with refractive index of 1 [116].

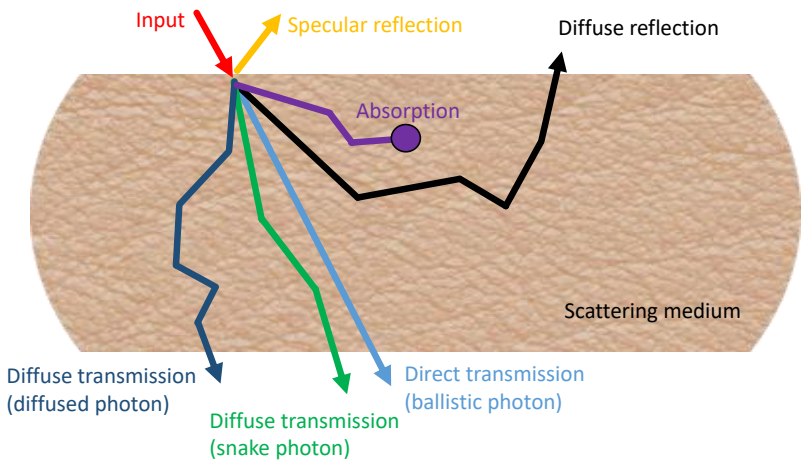

Figure 2-3: Photon transport modes in scattering media.

2.3.1 Photon Transport Modes Through Scattering Media

When a photon propagates through a scattering medium, it can undergo several types of interactions [18] (see figure 2-3):

- Absorption — within the medium.
- Specular reflection — from the surface of the medium.
- Diffuse reflection — after one or more scattering events within the medium.
- Direct transmission — no interaction with the medium (ballistic photons).
- Diffuse transmission — transmission through the medium after one or more scattering events. These photons are usually divided into:
  - Snake photons, which undergo a small number of scattering events.
  - Diffused photons, which undergo a large number of scattering events.

2.3.2 Radiative Transport Equation

It is common to write a conservation equation to describe the propagation of energy in scattering media. The equation is written for infinitesimal voxels and ignores

interaction between the photons (interference). This is known as the Boltzmann transport equation [18]:

$$\frac{1}{c}\frac{\partial L\left(r,\Omega,t\right)}{\partial t}+\nabla\cdot L\left(r,\Omega,t\right)\Omega+\mu_t L\left(r,\Omega,t\right)=\mu_s\int_{4\pi} f\left(\Omega,\Omega'\right)L\left(r,\Omega',t\right)d\Omega'+Q\left(r,\Omega,t\right) \tag{2.7}$$

where c is the speed of light in the medium, $L\left(r,\Omega,t\right)$ is the radiance at position r with direction Ω at time t, $f\left(\Omega,\Omega'\right)$ is the scattering phase function from angle Ω' into Ω, and $Q\left(r,\Omega,t\right)$ are sources. As in a standard conservation equation, the left hand side represents outgoing radiance, and the right hand side represents incoming radiance. Each term in Eq. 2.7 has a physical meaning, which is described below (from left to right):

1. The time derivative of the radiance, i.e. the net change of radiance in a small voxel at a given time point.
2. The radiance flux at direction Ω.
3. The absorption in the medium and scattering out to the angle Ω.
4. Scattering from all angles to angle Ω; this is the balance to term no. 3.
5. Sources in the voxel.

The main problem with Eq. 2.7 is its complexity. A common simplification to the Boltzmann transport equation is the photon diffusion equation, which is a zero order approximation [38]. The diffusion equation is expressed with the photon fluence rate $\Phi(r,t)$ such that $\Phi(r,t)=\int L(r,\Omega,t)d\Omega$. This approximation is valid when the radiance is almost uniform with respect to the angle parameter and results in:

$$-\nabla\cdot D\nabla\Phi(r,t)+c\mu_a\Phi(r,t)+\frac{\partial\Phi(r,t)}{\partial t}=\nu S(r,t) \tag{2.8}$$

where $S(r,t)=\int Q(r,\Omega,t)d\Omega$ (isotropic source), and $D=c/3\mu_s'$ is the diffusion coefficient [18]. Note that the diffusion coefficient may contain an absorption term [48, 49].

Assuming infinite, homogeneous, isotropic media without sources and absorption, the diffusion equation reduces to:

$$\frac{\partial \Phi(r,t)}{\partial t} = D\nabla^2 \Phi(r,t) \tag{2.9}$$

with the following solution in 3D:

$$\Phi(r,t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left\{-\frac{\|r\|_2^2}{4Dt}\right\} \tag{2.10}$$

See more details in Appendix A.

When the Diffusion Equation Fails

One of the challenges of the diffusion equation is its inability to describe ballistic and snake photons. It is common to see model mismatch when $w < 10/\mu_s$ [193]. In that case, the photons arrive earlier than predicted by the diffusion equation.

Several models try to address these challenges, for example by adding a delta term to the phase function [195] such that:

$$f\left(\Omega, \Omega'\right) \sim \alpha\delta\left(1 - \Omega \cdot \Omega'\right) + \left(1 - \alpha\right)\tilde{f}\left(\Omega, \Omega'\right) \tag{2.11}$$

where α is the fraction of ballistic photons.

2.3.3 Photon Transport as a Random Walk

It is also possible to consider photon transport as a random walk. In that case, the step size (propagation) is a random number sampled from an Exponential distribution with the mean equal to μ_s. When assuming isotropic scattering (similar to the assumption of diffusion equation), it is possible to derive the distribution density function of finding a photon at a certain location and time. Appendix A provides the derivation of such a random walk that results in Brownian motion.

Monte Carlo Simulation of Scattering

Closed form solutions to the photon transport equation only exist for simple cases, such as isotropic media, and are not applicable to the scattering problems considered here. Thus, it is common to simulate scattering with a Monte Carlo simulation [184]. In that case, photons are traced individually according to the following main steps:

1. **Propagation** — randomly sample the distance the photon propagates until the next scattering event (based on μ_s).

2. **Scatter** — randomly sample the scattering angle of the photon (based on the phase function).

3. **Termination conditions** — the photon can be absorbed, detected, or terminated by other conditions.

These steps are repeated until one of the termination conditions is met.

In our work, we extensively use Monte Carlo simulation. Since the computation of such a simulation is very costly, we implemented a GPU-accelerated version of the simulator, which can be found in [146].

2.4 Imaging through Scattering with Visible Light

Imaging through obstructions is commonly performed in non-visible parts of the electromagnetic spectrum, such as x-ray, THz [132], mm-wave [7], microwave [29], and RF [4, 3]. Here we focus on techniques based on visible light. These can be broadly categorized as pure light and multi-modal techniques. Table 2.1 summarizes the trade-offs among these techniques and compares them to All Photons Imaging.

Pure visible light techniques:

- **Wavefront Shaping [177]** — These techniques are based on coherent light and leverage the memory effect. They try to find the wavefront to illuminate the scene such that it conjugates the scattering. This results in a tight focus at a

specific location on the other side of the scattering media. The wavefront can be modulated with a spatial light modulator (SLM) or a digital micromirror device (DMD). The main advantage of this approach is the theoretical ability to achieve diffraction-limited imaging through scattering. However, at its simplest form, it requires access to both sides of the media, which limits its actual applicability. To overcome that challenge, it is commonly coupled with guidestars (see below).

- **Diffuse Optical Tomography (DOT) [18]** — Here the goal is to recover the 3D distribution of the scattering and absorption coefficients in the medium. This is commonly achieved by solving the radiative transfer equation or the diffusion equation. The measurement hardware is usually composed of many probes that surround the medium for illumination and detection. Many variants of DOT have been demonstrated such as time-domain and frequency-domain. Functional applications such as recovering fluorescence lifetime of markers in the medium [75] have also been demonstrated.

- **Optical Coherence Tomography (OCT) [76]** — This technique is based on interference between the signal reflecting from the scene and a reference beam. Changing the reference beam allows scanning through different depths and rejecting reflections from other layers. Standard methods are capable of achieving resolution in the order of micrometers. OCT is commonly used in retinal imaging and lesion analysis. OCT handles limited scattering situations and can penetrate a few mm in skin [99].

- **2/3 Photon Microscopy [41]** — This technique leverages non-linear effects in the medium. In a medium that supports non-linear optical interactions, two (or three) photons can combine to a single photon with double (or triple) energy. Because the effect is so weak it happens primarily in the beam's focus. Thus, with a controlled source it is possible to raster scan the source and detect only the higher energy photons; these photons are guaranteed to arrive from the focal spot regardless of scattering. This technique is common in microscopy.

- **Time Gating** — Since ballistic photons do not scatter, they can be used to recover the absorption coefficient within a medium in transmission mode (this is the fundamental concept behind x-ray imaging). Because the ballistic photons are the first to arrive, it is possible to time gate the imaging sensor such that only ballistic photons are used for imaging. Hardware used for time gating includes e.g. a streak camera [194] and Kerr gate [194]. The main limitation of time gating in transmission mode is SNR since only a few photons propagate without scattering. In optical reflection mode time gating is even more limited. In that case, beyond poor SNR, the measurement will always include back-reflected photons (photons that reflect from the scattering media without interacting with the target) which reduce the contrast.

Multi-modal techniques:

- **Acousto-Optics (Guidestar) [190]** — This is a complementary method to wavefront shaping. It enables to raster scan the imaging spot of wavefront shaping in the medium without physical access to the spot location. The most common technique uses the acousto-optic effect. Since ultrasound effectively does not scatter in tissue, it can easily form an ultrasound focal spot. Due to the acousto-optic effect, photons that propagate through the ultrasound focal spot will slightly shift their optical frequency. These photons can be easily detected outside the medium. Traditional acousto-optic techniques are limited to the ultrasound resolution.

- **Photo-Acoustics [185]** — When light is absorbed in tissue, it creates a thermally induced pressure wave which can be detected using ultrasound. Thus, it can form an image of light absorption in the medium. The resolution of this technique decreases with depth at a rate of 1/200 and can reach several cm deep. The main advantage of using photoacoustic methods is the ability to resolve optical contrast in ultrasound depths (few cm).

Method	Photoacoustic	Wavefront Shaping with Guidestar	Diffuse Optical Tomography	Optical Coherence Tomography	2/3 Photon Microscopy	Time Gating	All Photons Imaging
Spatial Resolution	50 μm	0.5 μm	1 cm	10 μm	1 μm	1 cm	5 mm
Photon Utilization	Only absorbed photons	Only acoustically modulated (efficiency is ~1%)	Only diffused (other photons are noise)	Only ballistic	Only through second harmonic generation	Only ballistic	All photons (with additional time tagging)
Hardware Complexity	High	High	Medium	Low-Medium	Medium	Medium	Medium
Dynamic Scenes	No	No	No	No	No	No	Yes
Raster Scan	Yes	Yes	Yes	Yes (1 dimension)	Yes	No	No
Remote Sensing	No	No	No	Partial	No	Yes	Yes
Field of View	Small	Small	Small	Small	Small	Large	Large
Requires Ultrasound	Yes	Yes	No	No	No	No	No

Table 2.1: Comparison of different imaging through scattering techniques with visible light.

2.5 Time-Resolved Sensing

Time-resolved sensing is key to the techniques presented in this dissertation. This is the ability to resolve an optical signal with picosecond time resolution. Here we provide a brief introduction to time-resolved sensing. Time-resolved measurement techniques can be broadly divided to: 1) systems that manipulate the signal and map the time response to another domain, and then perform "regular" analog to digital measurement; and, 2) circuit techniques for fast time to digital (TDC) conversion.

2.5.1 Signal Manipulation

The goal of the methods described in this section is to manipulate the target signal and reduce the burden from the TDC hardware. In some cases, there is no need for a TDC since the manipulation is mapping the time domain signal to another domain like spectrum or a spatial axis. Many techniques fall under this category, such as Sequentially Timed All-Optical Mapping Photography (STAMP) [118], Kerr gate [183], and Time Expansion [181]. We focus here on streak camera since it is the sensor used in Chapter 5.

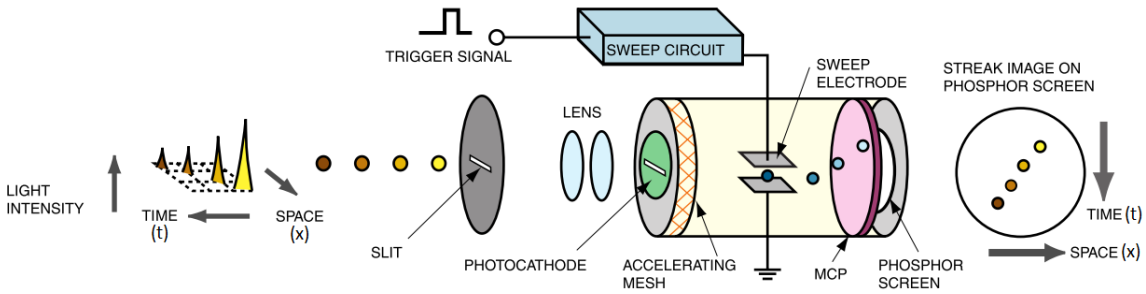

Figure 2-4: Streak camera operation principle (figure taken from [64]).

Streak Camera

Streak cameras are the fastest optical measurement device, with up to 0.2 ps [64] (not considering pump-probe approaches). The technique is based on a mapping of the temporal profile to spatial measurement using a CCD [154]. A streak camera measures an $x - t$ slice from the $x - y - t$ data cube. To slice the y axis the streak camera uses a slit on the focal plane. Photons that enter the slit are converted to electrons with a photo-cathode. A fast oscillating electric field is sweeping the electrons in the y direction such that the earlier electrons (photons) are mapped up and the later electrons (photons) are mapped down. At the edge of the cathode, a microchannel plate (MCP) is used to amplify the electrons' signal before they hit a phosphor screen to convert them back into photons. The photons are finally measured with a CCD. See Fig. 2-4 for system schematic. Streak camera characteristics:

- The streak camera is synchronized with a trigger (usually connected to a pulsed laser source that is used to illuminate the scene), such that each emitted pulse triggers a new sweep of the electric field in the cathode.

- The streak camera has a shutter with exposure time in the order of ms. Thus, the camera usually integrates over many laser pulses. The number of integrated pulses equals the product of the laser repetition rate (can be an order of MHz or kHz) and the streak camera exposure time. A pulse picker can be used to measure events from a single pulse; this is essential in case the phenomena is not repetitive.

- The time resolution is controlled by the speed of the electric field sweep (faster sweep results in better time resolution).

- Since the time profile is mapped to the y axis of the CCD, the time window that can be measured is a result of the selected time resolution and the number of pixels in the y axis of the CCD. For example, in case of 2 ps and 512 pixels we can measure a time window of 1.024 ns.

Various techniques have been applied to measure the full $x - y - t$ data cube, including:

- Rotating mirrors (periscope) [178, 145]: a set of two mirrors is used in a periscope-like configuration to scan through the y axis (each y slice is measured independently). This is the technique used in Chapter 5.

- Lenslet array [69]: a titled lenslet array is placed in front of the main camera lens and used to multiplex the y axis onto the x axis. The number of y pixels equals the number of lenslets, and the number of x pixels is the ratio between the number of CCD pixels and the number of lenslets. Thus, this approach sacrifices spatial x resolution for fast acquisition. This method enables a single shot measurement of non-repetitive phenomena like plasma discharge.

- Compressive sensing techniques [51]: compressive sensing is used to multiplex the full $x - y - t$ data cube onto the $x - t$ measurement. To that end, the slit is opened to the maximum width, and the scene is encoded with spatial pseudo-random masks. This method also enables real-time operation although there is a loss of optical signal due to the masks.

The main advantage of the streak camera is time resolution. It is the only device that is capable of measuring full spatiotemporal profile at low ps time scales. However, due to the complicated measurement pipeline, it suffers from poor sensitivity.

2.5.2 Time to Digital Conversion

These are techniques that directly time-tag the signal. Usually, the photons are converted to an electric current using a photodiode. The electric signal is then time-tagged with a TDC. Single photon avalanche diode (SPAD) detectors, such as the one used in Chapters 4, 6, use different variants of such TDC architectures. Here we provide a brief description of several TDC architectures; for a more extensive review see [26].

Time to voltage

This architecture first maps time to voltage (using a charge-pump for example), followed by traditional ADC quantization and digitization [72]. This approach usually achieves order of 30 ps time resolution. The main limitation is the non-linear process of time to voltage — this results in non-uniform quantization of the time profile. Another challenge is the technology scaling; as the feature size gets smaller the non-linearity of the charge-pump is worsened.

Flash TDC

The main challenge with directly digitizing a signal with ps resolution is the need of a $\sim$ 100 GHz clock. The flash TDC provides an alternative that uses multiple phase-aligned slower clocks which drive a set of flip-flops. The flip-flops effectively digitize the time between start and end signals [8]. A delay gate (usually an inverter or inverters) on the start line controls the propagation of the signal between the flip-flops. The final digitized time is simply the binary code of the flip-flops. The time resolution of this approach is defined by the delay of the inverter (usually order of $\sim$ 20 ps).

Vernier-line

Vernier-line [47] is similar to the flash TDC, but it provides better time resolution by using dual delay lines (on both start and stop signal). The delay lines are designed

such that the stop signal propagates slightly faster compared to the start signal. The process ends when the end signal reaches the start signal. The digitized time is simply the time difference between the two paths. This approach amplifies any mismatch between the lines and requires careful calibration.

Pipeline TDC

Pipeline TDC [26] uses a similar concept to pipeline ADC. First, a coarse TDC digitizes the signal; the residual is then amplified (time-stretched) and digitized by a fine TDC. The main reason for this split approach is the limited range of time amplifiers. This approach requires substantial area and power (all the time amplifications are calculated, but only one is used for the fine digitization). Another disadvantage is the latency due to the multiple steps.

Successive approximation TDC

The successive approximation TDC [26] recursively evaluates the time delay between the start and stop signals one bit at a time (essentially this is a binary search). The start and stop signals loop independently, and the goal of the TDC is to find the delay between them, such that they arrive at a phase detector (PD) within a least significant bit (LSB).

Noise shaping TDC

The noise shaping TDC [26] works in a similar way to the noise shaping ADC (like the sigma-delta digitization approach). In this approach, a gated ring oscillator is enabled by the time signal, and its status remains the same for the next signal to digitize. Thus the digitization error is a function of the previous events and assuming they are independent, results in a white measurement noise instead of a systematic error.

2.5.3 Time-Resolved Sensors Used Here

Here we use two types of time-resolved sensing systems:

1. **Streak Camera** — as described above, this is a line scanning system with a time resolution of 2 ps. We used a streak camera to image through a 1.5 cm thick tissue phantom as described in Chapter 5.

2. **SPAD Camera** — recently, time-resolved sensors have been demonstrated as part of single-photon counters (SPAD cameras). Such systems are single photon sensitive, and each detected photon is time-tagged with a time resolution in the order of a few tens-of-picoseconds. Beyond the time resolution, these sensors provide access to the statistical nature of light that is also beneficial for the purpose of overcoming scattering. The key advantage of SPAD detectors is that they are silicon-based and can be produced with CMOS process. As a result, they have the potential to scale to commodity sensors. Our SPAD camera is a 32×32 SPAD array. Each pixel is single-photon sensitive with 56 ps time resolution. We used the SPAD camera to image through a sheet of paper in Chapter 4, and through fog in Chapter 6.

Chapter 3

Lensless Imaging with FemtoPixel

Imaging through scattering media has an interesting analogy to lensless imaging (or bare sensor imaging). In lensless imaging, the goal is to capture a photo of the scene without a lens (a bare detector). Before explaining the connections between lensless imaging and imaging through scattering media, it is worth focusing on the applications of lensless imaging as it is a fundamental problem in its own.

In the visible part of the spectrum we have good manufacturing capabilities of lenses and dense array of pixels (these are commonly available in smartphone cameras). In other parts of the spectrum, these capabilities are not as common. For example, there are no good and low-cost lenses and dense pixel arrays in the UV, IR, THz, and RF spectra. While some camera systems exist in these spectra, they are usually very expensive. Eliminating the need for a lens, and relaxing the need for a dense array of detectors, opens the door for affordable imaging in these challenging spectra.

This chapter describes FemtoPixel, a framework for lensless imaging with a single (or few) detectors. The main novelty in the approach is the use of ultrafast detectors (with picosecond time resolution) for compressive lensless and single-pixel imaging. Picosecond time resolution allows distinguishing between photons that arrive from different parts of the scene with mm resolution. Thus, time-resolved detectors provide more information per measurement when compared to a regular detector, and allow us to substantially accelerate the acquisition process when compared to vanilla

single pixel imaging (50× faster in some of the configurations considered here). Moreover, the time-resolved sensor is characterized by a measurement matrix that enables us to optimize the active illumination patterns and reduce the required number of masks even further. The first principle analysis of time-resolved sensing presented in this chapter will serve as a foundation for imaging in more challenging scattering conditions in the following chapters.

The main technical contributions presented in this chapter include:

1. A computational imaging framework for lensless imaging with a compressive time-resolved measurement.

2. An analysis of a time-resolved sensor as an imaging pixel.

3. An algorithm for ideal sensor placement in a defined region.

4. An algorithm for optimized illumination patterns.

3.1 Connection Between Imaging Through Scattering and Lensless Imaging

To better understand the similarities between imaging through scattering media and lensless imaging, we first consider the function of a lens in an imaging system. Imaging is defined as the mapping of a scene onto a sensor. The lens in an imaging system is responsible for a one-to-one mapping of every scene point to points on the detector (Fig. 3-1a). The last statement is true just for scene points that are in focus, which is the scenario considered here. Since a scene point can be thought of as a point source (emitting light in many directions), the lens should map all light rays emitted from a specific point to a dedicated point on the detector and be invariant to the rays' angle. Removing the lens from the imaging system eliminates the one-to-one mapping between the scene and detector, such that each scene point is mapped to all detector points (Fig. 3-1b).

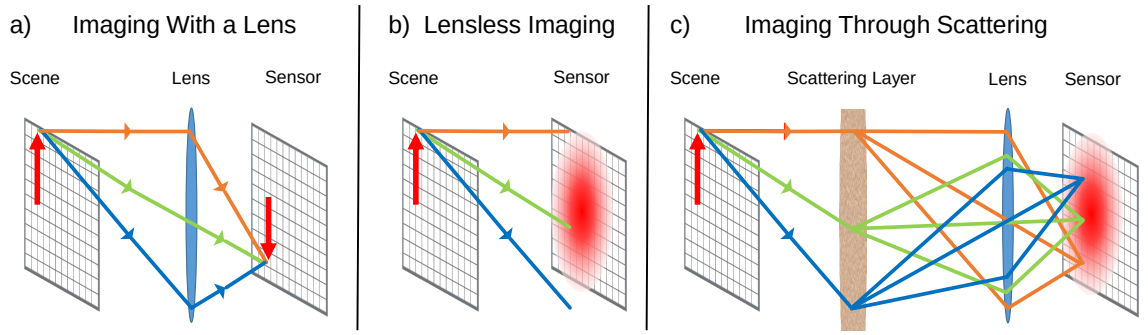

Figure 3-1: Lensless imaging is equivalent to imaging through scattering. a) A lens creates a one-to-one mapping of scene points to points on the sensor. The lens is focusing light from each scene point to the detector and eliminates the angular information in the mapping. b) In lensless imaging, the one-to-one mapping is gone, and each scene point is mapped to all detector points. c) When imaging through a sparse scattering layer, the lens is focused on the scattering layer. This results in a measurement that is equivalent to lensless imaging.

The connection to imaging through scattering media should now be more apparent. When we image through a sparse scattering material such as thin milky glass, the camera is focused on the milky glass (since the object is beyond the line of sight). Thus each detector point is uniquely mapped to a point on the scattering layer. This can be thought of as a remote sensor that integrates light over all angles (up to the aperture of the camera) at a particular diffuser position (Fig. 3-1c). That is, it is equivalent to replacing the scattering layer with a bare sensor.

3.2 Compressive Sensing and Single Pixel Camera Background

As mentioned before, imaging is traditionally performed with a lens. This is a physics-centric approach that puts the entire burden of imaging on the hardware, i.e. no algorithms are required. Thus, this approach is limited by the availability of a high quality lens, a dense detector array, and other parameters such as the aperture size, pixel pitch, and number of pixels. Recently, this physics centric approach has been challenged by computational photography that aimed to produce high quality photos with lower quality hardware. This has later evolved to computational imaging which

includes ideas on forming images based on more general measurements and sensors. In computational imaging, the measurement process encodes the target during the measurement. The measurement is transformed to an image in a computational reconstruction process.

One of the most notable examples of computational imaging is the single pixel camera [45] that demonstrates the recovery of a photo using a single pixel. It is useful to note that imaging with a single pixel can simply be done with a projector. In this case, the projector raster scans the scene, and each illuminated position is measured with the detector and stored to computationally form the complete image; this is common in microscopy and THz imaging.

A significant limitation of the raster scan approach is light sensitivity, since only a single point in the scene is illuminated. This can be alleviated by spatially multiplexing the illuminated points, for example using a Hadamard basis. In this case, instead of illuminating a single point on the scene, multiple points are illuminated based on some basis. If we denote the j-th pattern as $\mathbf{g}_j$ (a vectorized version of the pattern) and the pixel's measurement as m_j we get:

$$\mathbf{m} = \bar{\mathbf{G}}\mathbf{f} \tag{3.1}$$

where $\mathbf{f}$ is the vectorized scene, and the columns of $\bar{\mathbf{G}}$ are the $\mathbf{g}_j$ patterns. When $\bar{\mathbf{G}}$ is a basis matrix, such as when Hadamard patterns are used, it can easily be inverted to recover the scene $\mathbf{f}$.

The main constraint of the last approach is that the number of measurements has to be equal to the number of recovered pixels i.e. $\bar{\mathbf{G}}$ is a square, full rank matrix. Compressive sensing [25, 44] alleviates this constraint by allowing $\bar{\mathbf{G}}$ to be under-determined. That is, the number of measurements is less than the number of recovered pixels. This is achieved by introducing priors on the recovered scene. More specifically, if we know that the hidden scene can be sparsely represented under some known basis than we can recover the target even when the number of measurements is much smaller than the number of recovered pixels.

The main advantage of this compressive approach is the reduction in the total acquisition time of the system. It is important to note that in the case of a regular camera, the measurement of the scene pixels is done in parallel by multiple detectors (in one shot). In case of a single-pixel camera (in all its forms), this measurement is serialized in the time domain. The differences among the different approaches are in how this serialization process is performed: simple serialization, simple multiplexing, or compressive multiplexing.

The field of compressive sensing is very broad; here we briefly review the key concepts required for completeness of this chapter. Many extensive reviews and books have been written on compressive sensing, for example [135, 166] which are dedicated to optical applications.

For our purpose here, the application of compressive sensing is in solving Eq. 3.1 when $\bar{\mathbf{G}}$ is under-determined. Compressive sensing provides recoverable guarantees on $\mathbf{f}$ when: 1) $\mathbf{f}$ can be sparsely represented in some known basis $\mathbf{B}$ such that $\mathbf{f} = \mathbf{B}\mathbf{x}$ where $\mathbf{x}$ is a sparse vector, and 2) $\bar{\mathbf{G}}$ satisfies the restricted isometry property (RIP) [24]. The RIP requirement effectively requires that $\bar{\mathbf{G}}$ is nearly orthonormal when it operates on sparse vectors. The second requirement also defines the required number of measurements. It is well known that Hadamard matrices, random matrices sampled from Bernoulli and Gaussian distributions satisfy the RIP requirement [12]. Indeed, the first single pixel camera demonstration [170] used Bernoulli random matrix as a sensing operator. That is, compressive sensing allows us to solve the following equation:

$$\hat{\mathbf{x}} = \arg\min_{x} \|\mathbf{x}\|_0 \text{ such that } \mathbf{m} = \mathbf{G}\mathbf{B}\mathbf{x} \tag{3.2}$$

where $\|\cdot\|_0$ is the ℓ_0 "norm". Compressive sensing goes one important step further and relaxes the need for the ℓ_0 "norm" [44]. It has been shown that the convex ℓ_1 version can replace the ℓ_0 "norm" in Eq. 3.2, such that:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{G}\mathbf{B}\mathbf{x} - \mathbf{m}\|_2^2 + \lambda\|\mathbf{x}\|_1 \tag{3.3}$$

Here, we wrote the optimization in a form that allows slack for noise and model mis-

match, as well as a control of the regularization strength (λ). For imaging problems, it is common to assume that the image is sparse in gradient domain (under a total variation prior [137]), and solve the following optimization problem:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\mathbf{G}\mathbf{f} - \mathbf{m}\|_2^2 + \lambda \|\mathbf{f}\|_{TV} \tag{3.4}$$

For the purpose of the discussion here, it is important to note that the patterns used to encode spatial information can be practically realized in two ways:

- **Passive illumination**. In this case, a lens is used to focus light from the scene onto a digital micromirror array (DMD) or spatial light modulator (SLM). The encoded light is then captured by the detector. While this approach *requires* an imaging lens it can operate with ambient light. This is the common implementation of a single pixel camera [45].

- **Encode with active illumination**. In this case, a projector is used to project the patterns onto the scene. The detector can simply be placed to integrate light reflected from the scene. A lens may be added to improve light collection efficiency but is not required. An imaging lens may or may not be required inside the projector. For example, laser projectors do not require an imaging lens.

From a mathematical point of view, these two options are equivalent. In our work, we consider the second approach with an active illumination which is more relevant for time-resolved sensing.

Our goal in this chapter is to consider the case when the detector is ultrafast. In this case, the detector captures a time-resolved measurement per projected pattern. This extra information will help reduce the number of required patterns and will shorten the overall acquisition time.

3.3 Related Works

Compressive sensing has inspired many novel imaging modalities. Examples include: ultra-spectral imaging [9], subwavelength imaging [169], wavefront sensing [125], holography [20], imaging through scattering media [105], terahertz imaging [187], methane gas sensing [54], and ultrafast imaging [19, 51].

As discussed earlier, the single pixel camera [45] is one of the most notable applications of compressive sensing in imaging. It was later extended to general imaging with masks [11]. Single pixel imaging also extends to multiple sensors. For example, multiple sensors were used for 3D reconstruction of a scene by using stereo reconstruction [167]. Multiple sensors were also incorporated with optical filters to create color images [188].

3.3.1 Compressive Time-Resolved Sensing for Imaging

Time-resolved imaging has been first suggested by Raskar and Davis [129]. Time-resolved sensing has been mostly used to recover scene geometry. This is known as LIDAR [157]. LIDAR was demonstrated with a compressive single pixel approach [87, 33]. Time-resolved sensing has also been suggested to recover scene reflectance [88, 189] for lensless imaging, but without the use of structured illumination and compressive sensing.

Other examples of compressive time-resolved sensing include non-line of sight imaging, for example imaging around a corner [63] and through scattering [140]. Imaging around corners with sparsity priors was also demonstrated with low-cost time-of-flight sensors [66]. A review of these and other methods can be found in [143].

Here, we use compressive deconvolution with time-resolved sensing for lensless imaging to recover target reflectance.

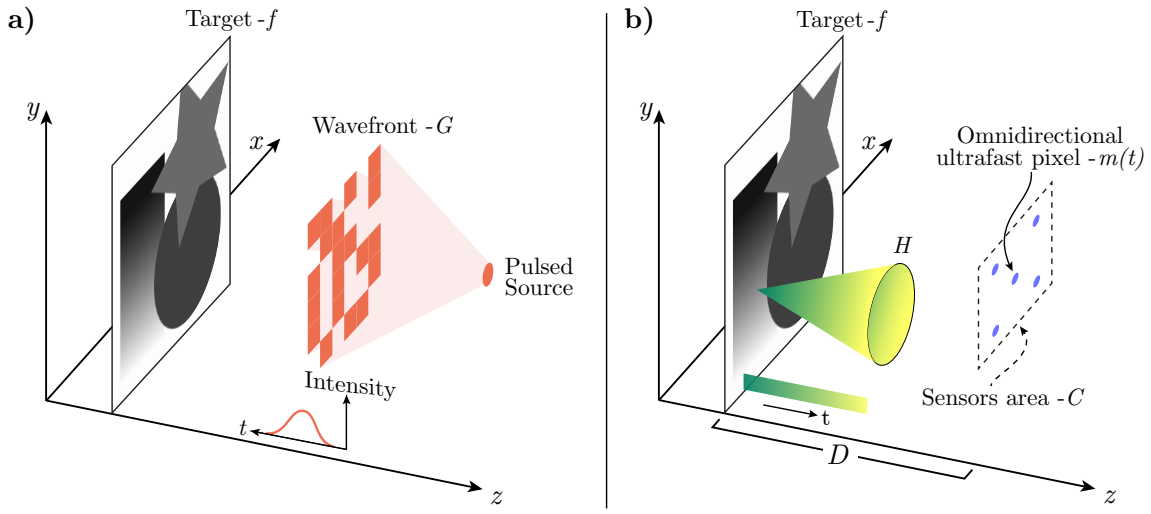

Figure 3-2: FemtoPixel – Lensless imaging with compressive ultrafast sensing. a) Illumination, a time pulsed source is wavefront modulated (**G**) and illuminates a target with reflectance **f**. b) Measurement, omnidirectional ultrafast sensor (or sensors) measures the time dependent response of the scene $m(t)$. **H** is a physics-based operator that maps scene pixels onto the time-resolved measurement.

3.3.2 Single Pixel Camera, Ghost Imaging, and Dual Photography

Other communities have also discussed the use of indirect measurements for imaging. In the physics community, the concept of using a single pixel (bucket) detector to perform imaging is known as ghost imaging and was initially thought of as a quantum phenomenon [124]. It was later realized that computational techniques can achieve similar results [159]. Ghost imaging was also incorporated with compressive sensing [83, 82]. In the computational imaging community, this is known as dual photography [158].

3.4 FemtoPixel Framework

Our goal is to develop a framework for compressive imaging with time-resolved sensing. The system overview is shown in Fig. 3-2. We start by defining the problem statement.

A target $\mathbf{f} \in \mathbb{R}^L$ with L pixels (we assume a vectorized representation to simplify the notation) is illuminated by a spatially modulated pulsed plane wave $\mathbf{g} \in \mathbb{R}^L$. For example, $\mathbf{g}$ can be produced by spatially modulating a time pulsed source such as a pulsed laser projector. The reflected light from the scene is measured by an omnidirectional ultrafast detector positioned on the sensors plane. The detector's time resolution T, and the time-resolved measurement is denoted by $\mathbf{m} \in \mathbb{R}^N$, where N is the number of time bins in the measurement. Better time resolution (smaller T) increases N. $\mathbf{H} \in \mathbb{R}^{N \times L}$ is the measurement operator defined by the space to time mapping that is enforced by special relativity (finite speed of light). At the limit of no time resolution, $N = 1$ (bucket detector), $\mathbf{H}$ is just a single row of ones, and the process is reduced to a regular single-pixel camera.

Under the FemtoPixel framework we consider K sensors ($i = 1..K$) with N time samples. The scene is illuminated by M patterns ($j = 1..M$) so the time-resolved measurement of the i-th sensor, when the target is illuminated by the j-th illumination pattern, is defined by: $\mathbf{m}_{i,j} = \mathbf{H}_i \mathbf{G}_j \mathbf{f}$. Where $\mathbf{G}_j \in \mathbb{R}^{L \times L}$ such that $\mathbf{G}_j = \text{diag}\{\mathbf{g}_j\}$ (a diagonal matrix with the pattern $\mathbf{g}_j$ on the diagonal). Concatenating all measurement vectors results in the total measurement vector $\vec{\mathbf{m}} \in \mathbb{R}^{NKM}$, such that the total measurement process is:

$$\vec{\mathbf{m}} = \begin{bmatrix} \vdots \\ \mathbf{m}_{i,j} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{H}_i \mathbf{G}_j \\ \vdots \end{bmatrix} \mathbf{f} = \mathbf{Q}\mathbf{f} \tag{3.5}$$

where, $\mathbf{Q}$ is an $NKM \times L$ matrix which defines the total measurement operator.

Here we invert the system defined in Eq. 3.5 using a compressive sensing approach. To that end, we analyze and physically modify $\mathbf{Q}$ to make the inversion robust. More specifically, we analyze and optimize the following fundamental components of $\mathbf{Q}$:

- **Physics-based time-resolved light transport matrix H**. $\mathbf{H}$ is a mapping from the spatial coordinates of the scene to the time-resolved measurement ($\mathbf{H} : \mathbf{r} \rightarrow t$). In section 3.4.1 we derive a physical model of $\mathbf{H}$ and discuss

its structure and properties. $\mathbf{H}$ can be modified by changing the sensor time resolution and position.

- **Combination of multiple sensors**. Multiple sensors can be placed in the sensor plane. Each sensor results in a different time-resolved light transport matrix $\mathbf{H}_i$. Section 3.4.2 discusses the considerations of sensor placement and presents an algorithm for optimized sensor placement in the sensor's plane.

- **Illumination (probing) matrix G**. This matrix is similar to the sensing matrix in the single pixel camera. In our analysis, we assume the modulation is performed on the illumination side. Section 3.4.3 presents an algorithm for generating optimized illumination patterns for compressive ultrafast imaging.

Broadly speaking, inverting Eq. 3.5 is robust if there is little linear dependence among the columns of $\mathbf{Q}$ (so that it has sufficient numerical rank). This is evaluated by the mutual coherence [50] which is a measure for the worst similarity of the matrix columns and is defined by:

$$\mu = \max_{1\leq a,b\leq L, a\neq b} \frac{\left|\mathbf{Q}_a{}^T \mathbf{Q}_b\right|}{\left\|\mathbf{Q}_a\right\|_2 \left\|\mathbf{Q}_b\right\|_2} \tag{3.6}$$

From here on, as suggested in [46], we will use an alternative way to target the mutual coherence which is computationally tractable and defined by:

$$\mu = \frac{1}{L}\left\|\mathbf{I}_L - \tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}}\right\|_F^2 \tag{3.7}$$

where $\mathbf{I}_L$ is the identity matrix of size L, $\tilde{\mathbf{Q}}$ is $\mathbf{Q}$ with columns normalized to unity, and $\|\cdot\|_F$ is the Frobenius norm. This definition also directly targets the RIP [46] and provides guarantees for using compressive sensing. We use Eq. 3.7 as a quantitative measure for evaluating and optimizing $\mathbf{Q}$.

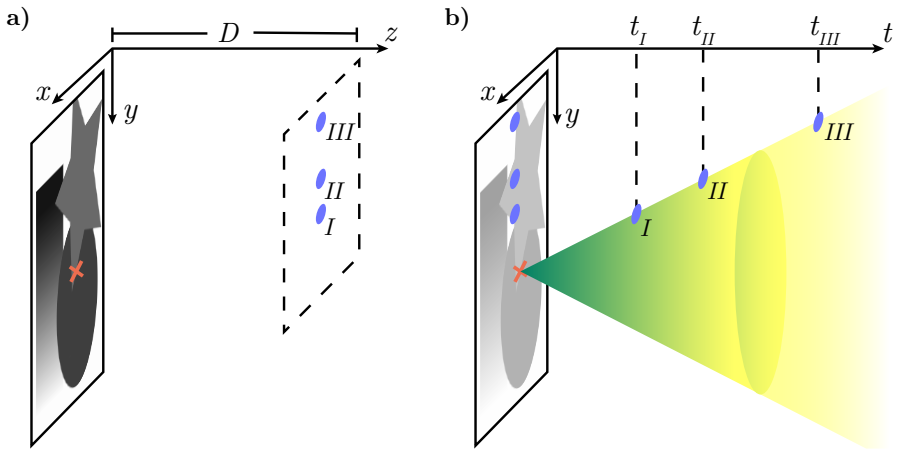

Figure 3-3: Light cone schematic for a planar stationary target. a) Scene geometry, the target plane and sensor plane are separated by a distance $z = D$. Three detectors (marked with blue circles and numbered I, II, III) are positioned at different y positions in the detector plane. b) The light-like part of the light cone emanating from the target point marked with a red 'X' defines the event measurement times at the different detectors. Due to the light cone geometry, the light will arrive at the detectors at different times. First it will be measured by detector I which is closest to the source, followed by detectors II and III. These times are defined by Eq. 3.8. Note that the horizontal axis describes in a) the z-axis, and in b) the time-axis.

3.4.1 Time-Resolved Light Transport

Here, we develop a generic light transport model for time-resolved imaging. Information propagation is governed by the finite speed of light. This provides geometrical constraints to an image formation model. Information propagation is conveniently described in a Minkowski space with the space-time four-vector $(\mathbf{r}, t) = (x, y, z, t)$. For example, for a point source at position $\mathbf{r}'$ pulsing at time t', and a sensor at position $\mathbf{r}$, the space-time interval between the source $(\mathbf{r}', t')$ and the sensor $(\mathbf{r}, t)$ is defined by:

$$s^2 = \|\mathbf{r} - \mathbf{r}'\|_2^2 - c^2(t - t')^2 \tag{3.8}$$

where c is the speed of light. Enforcing causality and light-like behavior is equivalent to setting $s^2 = 0$ which defines the light cone. Fig. 3-3 shows a schematic of the light cone and demonstrates how the same event is measured in various positions at different times.

As a result of the light cone geometry, the time-resolved measurement of a sen-

sor positioned at $\mathbf{r}$ and time t of an arbitrary three-dimensional time dependent scene $f(\mathbf{r}', t')$ is the integral over all $(\mathbf{r}', t')$ points on the manifold M defined by $s^2 = \|\mathbf{r} - \mathbf{r}'\|_2^2 - c^2(t - t')^2 = 0$:

$$m(\mathbf{r}, t) = \int_M \frac{1}{\|\mathbf{r} - \mathbf{r}'\|_2^2} f(\mathbf{r}', t') dM \tag{3.9}$$

where the $1/\|\mathbf{r} - \mathbf{r}'\|_2^2$ term accounts for the intensity drop-off.

Since the time scales of light transport are in the order of $0.3 \frac{\text{mm}}{\text{ps}}$ we assume that the scene is stationary compared to these time scales. That is t' is fixed and assumed $t' = 0$ without loss of generality. Next, we assume a planar scene at $z' = D$ and sensor positioned at x, y and $z = 0$. Due to the circular symmetry of the light cone we get:

$$m(x, y, t) = \int_0^{2\pi} \frac{1}{c^2 t^2} f\left(x + \rho\cos(\theta'), y + \rho\sin(\theta')\right) \rho d\theta' \tag{3.10}$$

with $\rho = \sqrt{c^2 t^2 - D^2}$. The intensity drop-off is written as a function of time since: $\|\mathbf{r} - \mathbf{r}'\|_2^2 = c^2 (t - t')^2 = c^2 t^2$.

The sensor's finite time resolution T corresponds to a time sampling of $m(x, y, t)$ that is denoted by $\mathbf{m}$. A sensor positioned at location $\mathbf{r}_i = (x_i, y_i)$ will produce a measurement $\mathbf{m}_i = \mathbf{H}_i \mathbf{f}$, where $\mathbf{H}_i$ is defined by the kernel in Eq. 3.10, and $\mathbf{f}$ is a discretized, lexicographically ordered representation of the target reflectance map $f(x, y)$. $\mathbf{H}_i$ is a mapping from a two-dimensional spatial space to a time measurement, that is dependent on the detector position and its time resolution. Below we discuss the properties of this kernel.

One-Dimensional Analysis

It is informative to analyze $\mathbf{H}$ in a planar world ($y = 0$) with the sensor at the origin (Fig. 3-4). In that case Eq. 3.10 is simplified to:

$$m(t) = \frac{1}{c^2 t^2} \left[f\left(-\sqrt{c^2 t^2 - D^2}\right) + f\left(\sqrt{c^2 t^2 - D^2}\right) \right] \tag{3.11}$$

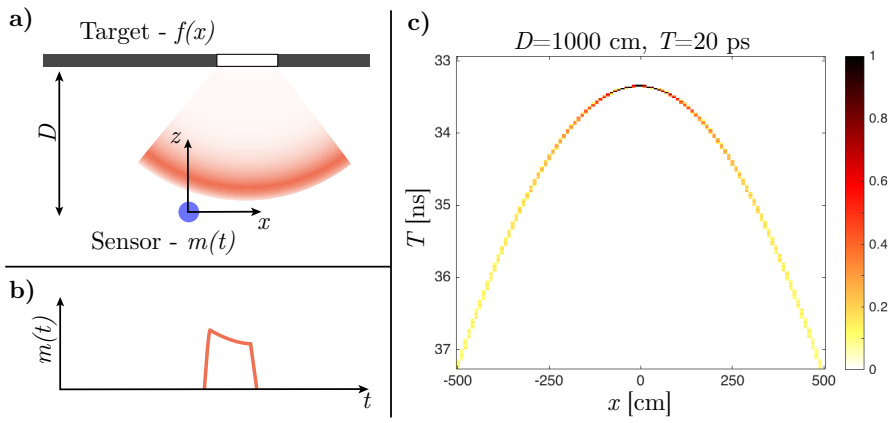

Figure 3-4: Time-resolved light transport in a one-dimensional world. a) Geometry, the target is a black line with a white patch, at a distance D from the time-resolved sensor. b) The time-resolved measurement produced by the sensor. The signal start time corresponds to the patch distance, and the time duration to the patch width. c) The measurement matrix $\mathbf{H}$, generated by Eq. 3.11. Here the distance to the target is $D = 1000\,\text{cm}$ and the sensor has a time resolution of $T = 20\,\text{ps}$.

Fig. 3-4c shows an example of $\mathbf{H}$ matrix in that case. This simple example demonstrates the key properties of the time-resolved measurement:

1. It is a nonlinear mapping of space to time.

2. The mapping is not unique (two opposite space points are mapped to the same time slot).

3. Spatial points that are close to the sensor are undersampled (adjacent pixels mapped to the same time slot).

4. Spatial points that are far from the sensor are oversampled but the signal is weaker.

These properties affect imaging parameters as described next:

1. **Resolution limit.** Due to the undersampling closer to the detector, the minimum recoverable spatial resolution is defined by the closest point to the sensor: $ct_1 = D$, and the next time slot: $c(t_1 + T) = \sqrt{D^2 + \Delta x^2}$, which results in:

$$\Delta x = cT\sqrt{1 + 2\frac{D}{cT}} \tag{3.12}$$

Fig. 3-5 shows a few cross sections of Eq. 3.12 for relevant distances D and time resolution T. Better time resolution is required for scenes that are farther (in order to achieve the same recoverable resolution).

2. **Signal to noise ratio and dynamic range limitation.** The closest scene point to the sensor defines the measurement gain. In order to avoid saturation that is defined by the intensity I_{sat}, we require that $I_{sat} > AD^{-2}$, where A accounts for all measurement constants. The farthest measurable point from the sensor (x_{max}) should result in a measurement above the noise floor: $I_n < A\left(D^2 + x_{max}^2\right)^{-1}$.

   Since the noise is usually amplified by higher gain, I_{sat} is proportional to I_n. If we choose $I_{sat} = BI_n$ (for some constant $B > 1$) we get: $x_{max} < \sqrt{B-1}\ D$. This demonstrated that as the scene gets closer, the coverage area gets smaller. These dynamics are true even if there is a more complicated relationship between I_{sat} and I_n.

The combined effect of these phenomena is demonstrated in Fig. 3-6. In this example, we consider a 'half plane' ($x > 0$), where the target reflectance is $f(x) = sin(x)$. The detector has a time resolution of $T = 20\,\text{ps}$, with additive white Gaussian noise that results in $\text{SNR} = 35\,\text{dB}$. For this simple demonstration, we use the Moore-Penrose pseudoinverse to invert the system, such that $\hat{\mathbf{f}} = \mathbf{H}^{\dagger}\mathbf{m}$. The inversion demonstrates that close to the origin ($x < 50\,\text{cm}$) the reconstruction suffers from an undersampled measurement; this area is not sensitive to the measurement noise, and looks identical with zero noise. The noise has an obvious effect on the reconstruction farther from the origin ($x > 700\,\text{cm}$).

Analysis of a planar scene

All the properties discussed in the context of one-dimensional scene extend to the case of a planar scene. Eq. 3.10 shows that the measurement process integrates over circles centered around the sensor. Due to the finite time resolution, the circles extend to rings. The rings are thinner for further points, according to $\rho_n = \sqrt{c^2(nT + t_0)^2 - D^2}$,

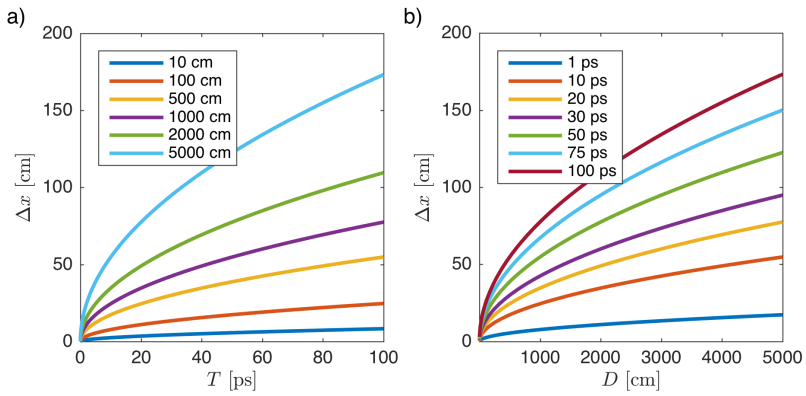

Figure 3-5: Recoverable resolution with time-resolved sensing. a) Plots for various scene distances D as a function of sensor time resolution T. b) Plots for various sensor time resolutions as a function of target distance D.

where n is the time sample number and $t_0 = D/c$ is the time of arrival from the closest point. Fig. 3-7 shows the ring structure for a few cases of time resolution and target distance.

In this chapter, we consider the case of a planar fronto-parallel scene to the sensor plane. In that case, Eq. 3.10 provides the structure of the $\mathbf{H}_i$ matrix. The kernel $\mathbf{H}_i$ maps rings with varying thicknesses from the scene plane to specific time bins in the measurement. At this point, we understand and have quantitative tools to predict the effects of changing the sensor's time resolution and position on the measurement matrix $\mathbf{Q}$. Naturally, better time resolution will reduce the mutual coherence. An alternative to improved time resolution is to add more sensors as discussed next.

3.4.2 Sensors Positioning

The use of multiple sensors is a natural extension to the single pixel camera. The sensors' position affects the measurement matrix $\mathbf{Q}$ and so can be optimized. Here we derive an algorithm for sensors placement in an array in order to reduce the mutual coherence of $\mathbf{Q}$. To simplify the array structure, we constrain the sensors to a single plane $z = 0$ and to an allowed physical area. The algorithm accepts two parameters: the number of sensors K and the allowed physical area $\mathcal{C}$, and provides the ideal positions under these constraints.

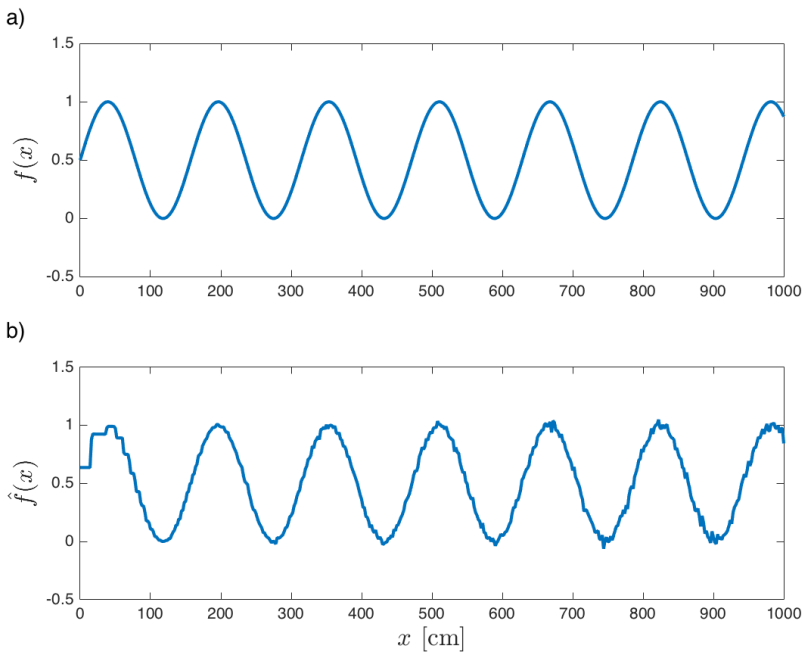

Figure 3-6: Effects of averaging and noise on time-resolved sensing. a) $f(x)$ is a sinusoid on the positive half plane, at a distance $D = 1000\,\text{cm}$ from a sensor with time resolution $T = 20\,\text{ps}$ and measurement noise of $\text{SNR} = 35\,\text{dB}$. b) $\hat{f}(x)$ is the result of inverting the system using the Moore-Penrose pseudoinverse, which demonstrates the undersampled measurement close to the sensor, and sensitivity to noise further away from the sensor.

First, we consider the case of $K = 2$. Starting with Eq. 3.10, the goal is to maximize the difference between $m(x_1, y_1, t)$ and $m(x_2, y_2, t)$. This is achieved by choosing $\mathbf{r}_1 = (x_1, y_1)$, $\mathbf{r}_2 = (x_2, y_2)$ which are furthest apart (to minimize overlap of the rings as shown in Fig. 3-7).

In case of arbitrary K, the goal is to select $i = 1..K$ positions $\mathbf{r}_i$ within an area $\mathcal{C}$ such that the distance between the sensors is maximized. This can be achieved by solving:

$$\{\mathbf{r}_i\}_{i=1..K} = \underset{\{\mathbf{r}_i \in \mathcal{C}\}_{i=1..K}}{\arg\max} \left\{ \sum_{k=1}^{K} \min_{k \neq k'} \|\mathbf{r}_k - \mathbf{r}_{k'}\|_2 \right\} \tag{3.13}$$

Eq. 3.13 can be solved by a grid search for a small number of sensors. A more general solution is to relax the problem and follow the equivalent of a Max-Lloyd quantizer [123]. The steps are as follows:

1. Initialize K random positions in the allowed area $\mathcal{C}$.

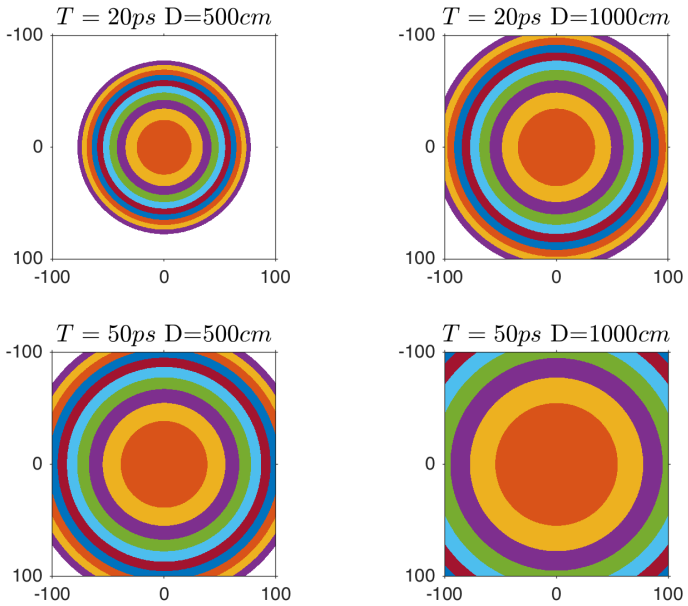

Figure 3-7: Time-resolved sensing maps rings with varying thickness to different time bins. The color represents time samples indexes (for the first 10 samples). As the time resolution worsens or the target is further away, the rings become thicker. The images show various sensor time resolutions T and target distances D for a subset area of 100 cm $\times$ 100 cm.

2. Repeat until convergence:

    - Calculate the Voronoi diagram of the point set.
    - Move each sensor position to the center of its cell.

This positioning algorithm is evaluated in Fig. 3-8 for various system parameters by assessing the effect of the sensor time resolution, number of sensors and array size (square of varying area) on the mutual coherence cost objective (Eq. 3.7). Several key features of the system are observed:

1. Improving time resolution reduces the number of required sensors non-linearly.
2. It is always beneficial to improve the sensors' time resolution.
3. The sensor area defines a maximum number of useful sensors, beyond which there is no significant decrease in the mutual coherence (increasing the array size linearly reduces the mutual coherence for a fixed number of sensors).

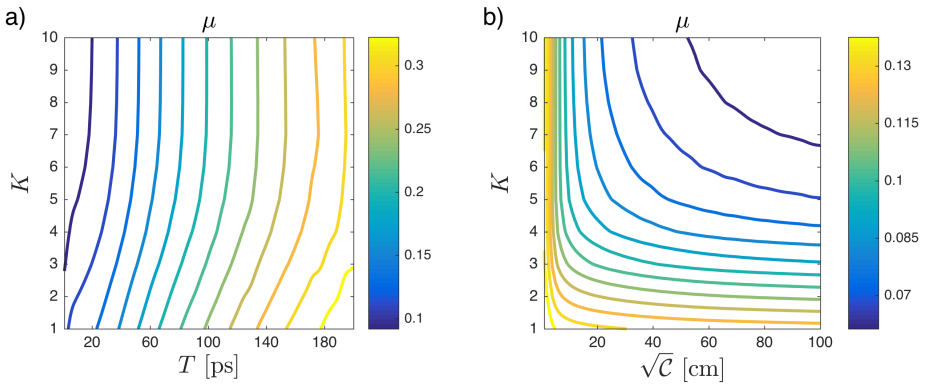

Figure 3-8: Mutual coherence as function of sensor number K, time resolution T, and array size $\mathcal{C}$. The target size is $5\,\text{m} \times 5\,\text{m}$, composed of 80×80 pixels, and at a distance of $D = 10\,\text{m}$ from the sensor plane. a) Mutual coherence contours for a varying number of sensors and their time resolution (for fixed array size $\mathcal{C} = 10\,\text{cm} \times 10\,\text{cm}$). b) Similar to (a) with varying array size constraint (for fixed time resolution $T = 20\,\text{ps}$).

4. It is possible to trade off between different aspects of the system's hardware by traveling on mutual coherence contours. For example, a decrease in the sensor time resolution can be balanced by adding more sensors. This can be useful for realizing an imaging system as sensors with lower time resolution are less expensive and easier to manufacture.

3.4.3 Ideal Compressive Patterns Optimization

We now make the leap to compressive sensing. Previous sections discussed single sensor considerations and sensors placement in an array under an assumption of uniform illumination. This section covers ideal active illumination patterns. We assume the illumination wavefront is amplitude-modulated; this can be physically achieved by a digital micromirror device (DMD) or liquid crystal display (LCD).

When considering different illumination patterns, Hadamard patterns and random patterns sampled from a Bernoulli distribution are a standard choice. Instead, we suggest patterns that directly aim to minimize the mutual coherence of the measurement matrix. The mathematical patterns may have negative values, which can be represented by taking a measurement with an "all on" pattern and subtracting it from the other measurements (due to the linearity of the system) [11].

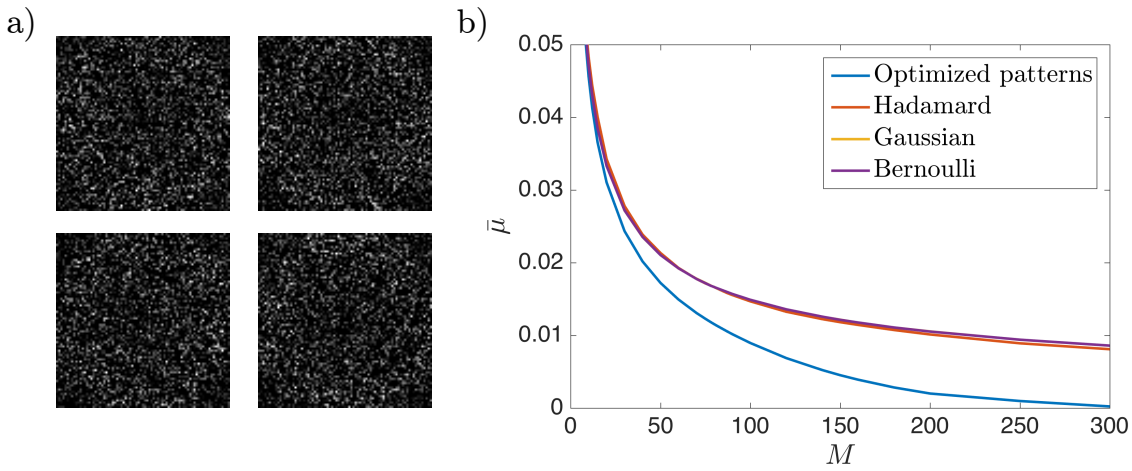

Figure 3-9: The value of optimized active illumination patterns. The patterns are optimized for a $5\,\mathrm{m} \times 5\,\mathrm{m}$ target composed of 80×80 pixels at a distance of $D = 10\,\mathrm{m}$ from the sensor plane. The measurement is simulated with $K = 1$ sensors and $T = 20\,\mathrm{ps}$. a) Examples of several patterns computed for $M = 50$. b) Comparison of different active illumination methods and their effect on the mutual coherence for varying M. The optimized patterns outperform Hadamard and random patterns sampled from Gaussian and Bernoulli (in $\{-1, 1\}$) distributions.

Our goal is to find a set of M illumination patterns that would minimize Eq. 3.7 (given the sensors' properties). Remember that the illumination patterns in the system appear in $\mathbf{G}_j$. Since the illumination matrix $\mathbf{G}_j$ is performing pixel-wise modulation of the target, it is a diagonal matrix with the pattern values on the diagonal $\mathbf{G}_j = \mathrm{diag}\{\mathbf{g}_j\}$, where $\mathbf{g}_j$ is a vector containing the j-th pattern values. Taking a closer look at Eq. 3.5, we stack all the sensor matrices $\mathbf{H}_i$ into $\overline{\mathbf{H}}$ such that:

$$\mathbf{Q} = \begin{bmatrix} \vdots \\ \mathbf{H}_i\mathbf{G}_j \\ \vdots \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{H}}\mathbf{G}_1 \\ \vdots \\ \overline{\mathbf{H}}\mathbf{G}_M \end{bmatrix} \tag{3.14}$$

which helps since now each mask appears just once. Based on Eq. 3.7, the ideal patterns are the solution to:

$$\{\mathbf{g}_j\}_{j=1..M} = \underset{\{\mathbf{g}_j \in [-1,1]^L\}_{j=1..M}}{\arg\min} \left\{ \left\| \mathbf{I}_L - \tilde{\mathbf{Q}}^T\tilde{\mathbf{Q}} \right\|_F^2 \right\} \tag{3.15}$$

This can be solved with standard constrained optimization solvers. To simplify the problem and make it differentiable we replace the domain requirement with a simple regularizer as follows:

$$\{\mathbf{g}_j\}_{j=1..M} = \underset{\mathbf{g}_{j=1..M}}{\arg\min} \left\{ \left\| \mathbf{I}_L - \tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}} \right\|_F^2 + \lambda \sum_{j=1}^{M} \|\mathbf{g}_j\|_2^2 \right\} \tag{3.16}$$

where λ is chosen such that by the end of the optimization procedure all values are within $[-1, 1]$. Appendix B provides the derivation for the cost function and its gradient that help to efficiently solve Eq. 3.16 with gradient descent.

Fig. 3-9a shows several examples of the patterns computed by solving Eq. 3.15. Fig. 3-9b demonstrates the value of the optimized patterns compared to Hadamard and random patterns sampled from Gaussian and Bernoulli (in $\{-1, 1\}$) distributions. For very few illumination patterns (below ten) all patterns are comparable. However, when more illumination patterns are allowed, the optimized patterns are performing better by reducing the mutual coherence faster compared to the other approaches. As predicted, the performances of Hadamard, Gaussian and Bernoulli patterns are nearly identical. Effectively this means that for a given mutual coherence target, using the optimized patterns would require fewer patterns (shorter acquisition time) when compared to random patterns. For example, 100 optimized patterns or 300 traditional patterns would result in comparable performance.

Figure 3-10 shows an analysis of the effect of the number of allowed illumination patterns, the sensor time resolution, and the number of sensors on the overall mutual coherence. As predicted by CS theory, increasing the number of patterns has a strong effect on the mutual coherence. This strong effect enables to easily relax the demands on the hardware requirements when needed. However, as more patterns are allowed, there are increasingly more dependencies on the sensors' parameters. This demonstrates the synergy between compressive sensing and time-resolved sensing. In this case, traveling on mutual coherence contours allows one to trade off system complexity (cost, size, power) with acquisition time (increased when more patterns are required).

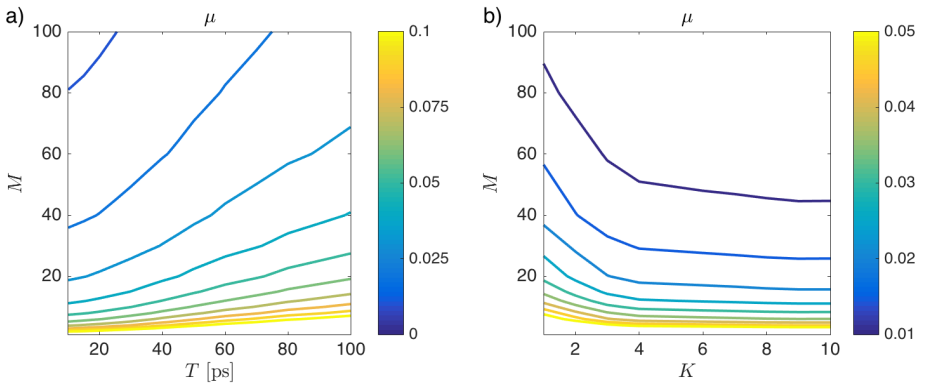

Figure 3-10: Mutual coherence as function of number of illumination patterns M, sensor time resolution T, and sensor number K. The target size is $5\,\text{m}\times 5\,\text{m}$, composed of 80×80 pixels, and at a distance of $D = 10\,\text{m}$ from the sensor plane. The sensor area $\mathcal{C}$ is a square of size $10\,\text{cm}\times 10\,\text{cm}$. a) Mutual coherence contours for a varying number of illumination patterns and sensors' time resolution (for a fixed number of sensors $K = 1$). b) Similar to (a) with a varying number of sensors (for fixed time resolution $T = 20\,\text{ps}$).

3.5 Simulation Results

We now demonstrate target reconstruction using the FemtoPixel framework. The following scenario is considered: the target dimensions are $5\,\text{m}\times 5\,\text{m}$ with 80×80 pixels ($L = 6400$) and it is placed $10\,\text{m}$ away from the detector plane. The detector array is limited to a square area of $\mathcal{C} = 10\,\text{cm}\times 10\,\text{cm}$. The detector placement method used is described in section 3.4.2 and the illumination patterns are computed using the algorithm suggested in section 3.4.3. The measurement operator is simulated as described in section 3.4.1 to produce the total measurement vector. White Gaussian noise is added to the total measurement vector to produce a measurement SNR of 60 dB. The targets simulated here are natural scenes (sparse in the gradient domain). To invert the measurement model in Eq. 3.5 we use TVAL3 [102] (with TVL2 and a regularization parameter of 2^{13} for all targets). The reconstruction quality is evaluated with both Peak Signal to Noise Ratio (PSNR — higher is better, performs pointwise comparison) and Structural Similarity index (SSIM — ranges in $[0, 1]$, higher is better, takes into account the spatial structure of the image [186]).

So far, the discussion focused on reducing the mutual coherence of the measurement matrix $\mathbf{Q}$. Fig. 3-11 demonstrates the effect of various system parameters on

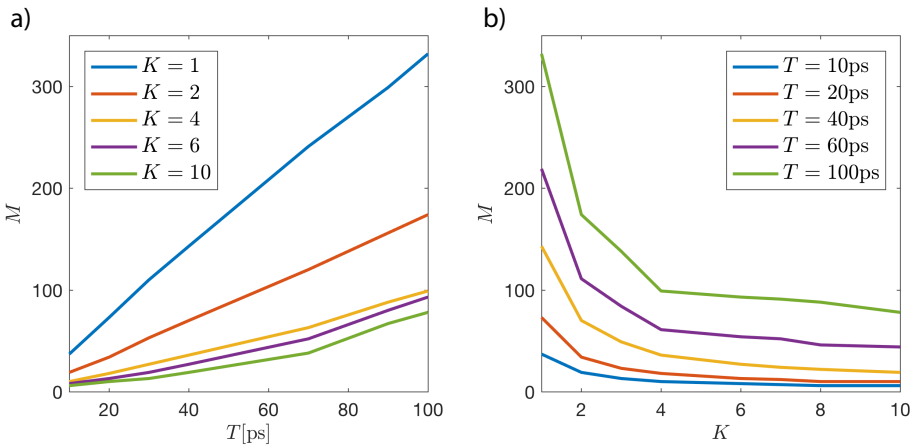

Figure 3-11: System parameters effect on reconstruction quality. Various design points (different number of sensors K and time resolution T) are simulated. The number of optimized illumination patterns M is set as the minimal number of patterns required to achieve reconstruction quality with SSIM ≥ 0.95 and PSNR ≥ 40 dB. The target is the cameraman image (see Fig. 3-12 right). a) Demonstrates the trends of various numbers of detectors K as a function of the time resolution T. b) Shows the trends of different detector time resolutions as a function of the number of detectors.

the full reconstruction process. The target used is the cameraman image (Fig. 3-12 right). The goal is to find the minimal number of illumination patterns in order to produce a reconstruction quality defined by SSIM ≥ 0.95 and PSNR ≥ 40 dB. This is repeated for various number of detectors with different time resolutions. The trends demonstrate a linear relationship between the number of illumination patterns and the detector time resolution needed for the specified reconstruction quality. Another notable effect is the significant gain in the transition from one to two detectors followed by a diminishing return for additional detectors. This gain decreases as the detector time resolution improves. These trends can be useful to trade off different design constraints. For example, for the specified reconstruction quality the user can choose one detector with a time resolution of 20 ps and 80 patterns. The same acquisition time can be maintained with two simpler detectors of 40 ps. Alternatively, two detectors with 20 ps require only 40 patterns (shorter acquisition time) for equal reconstruction quality.

We now compare the suggested design framework to a traditional (non-time aware) single pixel camera. This is simulated with an $\mathbf{H}$ matrix with just one row of

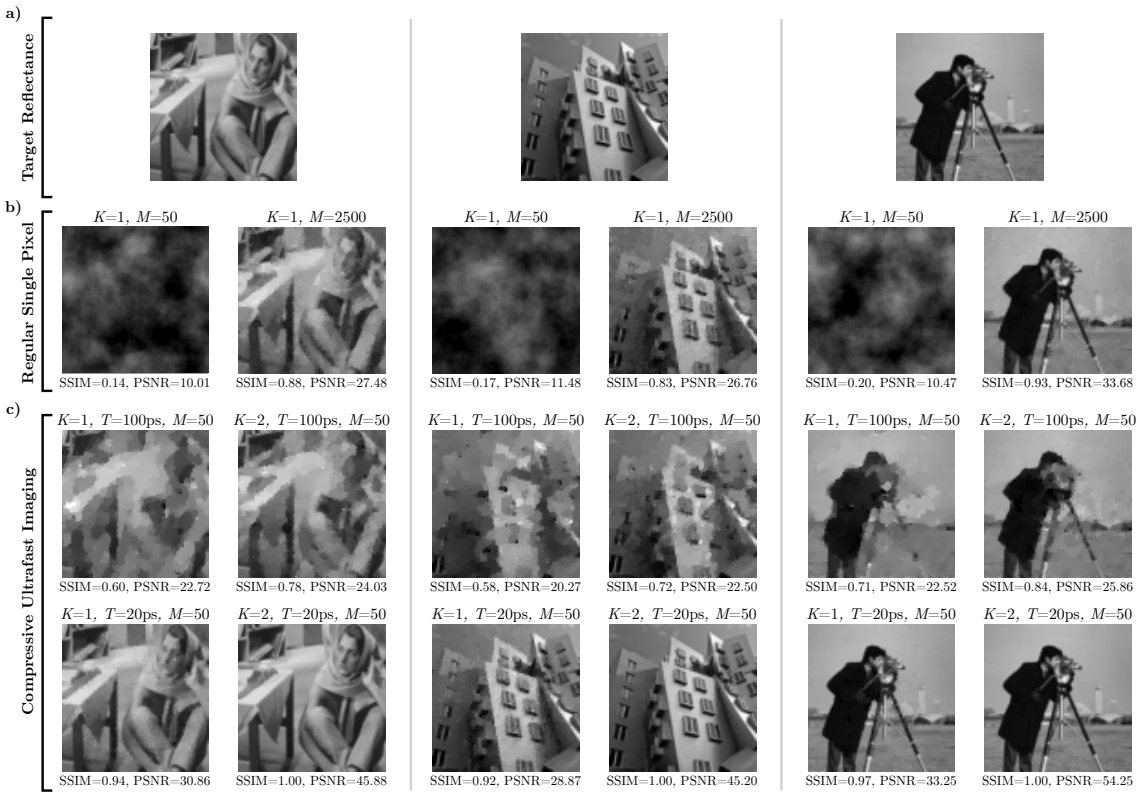

Figure 3-12: FemtoPixel simulation results. a) The target image. b) Result with a regular single pixel camera with $M = 50$ and $M = 2500$ patterns. c) Results with compressive ultrafast sensing with $M = 50$ for four design points with time resolution of $T = 100\,\text{ps}$ and $T = 20\,\text{ps}$, and $K = 1$ and $K = 2$. All reconstructions are evaluated with SSIM and PSNR. The results demonstrate the strong dependency on time resolution. Recovery with $K = 2$ and $T = 20\,\text{ps}$ shows perfect reconstruction on all targets based on SSIM. All measurements were added with white Gaussian noise such that the measurement SNR is 60 dB.

ones. The illumination patterns are sampled from a Bernoulli random distribution in $\{-1, 1\}$ in a similar way to the original single pixel camera experiments [45]. Fig. 3-12 shows the results for three different targets. Reconstructions with a traditional single pixel camera are shown in Fig. 3-12b for $M = 50$ and $M = 2500$ patterns. Four different design points of compressive ultrafast imaging are demonstrated in Fig. 3-12c: $\{K = 1, T = 100\,\text{ps}\}$, $\{K = 2, T = 100\,\text{ps}\}$, $\{K = 1, T = 20\,\text{ps}\}$, and $\{K = 2, T = 20\,\text{ps}\}$, all with $M = 50$ patterns (such that the acquisition time is equal). Several results are notable:

- Reconstruction with $K = 2$, $T = 20\,\text{ps}$, and $M = 50$ achieves perfect quality

based on SSIM for all targets.

- Reconstruction with $K = 1$, $T = 20\,\text{ps}$, and $M = 50$ outperforms the traditional single pixel camera approach with $50\times$ fewer illumination patterns and demonstrates the potential gain of the FemtoPixel framework.

- A traditional single pixel reconstruction with $M = 50$ patterns (same acquisition time as the compressive ultrafast imaging design points discussed) fails to recover the scene information.

- There is a significant gain in performance when improving the sensor time resolution.

3.6 Experimental Results

In this section, we describe an experimental demonstration of the FemtoPixel framework [147]. For illumination, a femtosecond Ti:Sapphire pulsed source (80 MHz repetition rate, 120 fs pulse duration, 800 nm wavelength) is incident on a Digital Micromirror Device (DMD, Texas Instruments). The DMD is projecting the desired compressive patterns (with a resolution of 32×32) on the target and is controlled by a computer. The target is imaged with a Becker&Hickl photomultiplier tube (PMT) detector and sampled with Time-Correlated Single Photon Counter (TCSPC) electronics with a total impulse response time of 27 ps. A sketch of the experimental setup is shown in Fig. 3-13a.

The first step in the experiments is the calibration of the measurement matrix $\mathbf{H}$. This step is accomplished by imaging a white wall with full rank Hadamard patterns ($32 \times 32 = 1024$ in this case) that are used to recover $\mathbf{H}$. Fig. 3-13b shows the structure of the measurement operator, with a ring structure that is similar to that found in Fig. 3-7. To construct the result in Fig. 3-13b we plot the time bin that corresponds to the maximum value of the time response in each pixel.

Figure 3-14a shows experimental recovery results using compressive ultrafast sensing. The target is a white circle. The figure shows recovery results using the sug-

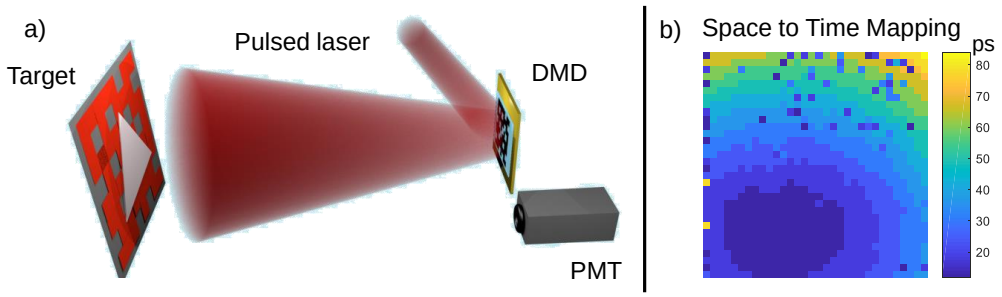

Figure 3-13: FemtoPixel experimental setup. a) Optical setup sketch. b) The measured **H** operator shows the expected ring structure.

gested approach and compares it to a traditional single pixel camera (without any time sensitivity) for a different number of compressive masks. The reconstructions are evaluated with SSIM and PSNR. Fig. 3-14b also presents plots for the complete SSIM and PSNR trends as the number of used masks is reduced (more compression). We also evaluate the mutual coherence for the measured operator with a different number of patterns, and compare the results to a regular single-pixel camera. As predicted the FemtoPixel framework provides superior mutual coherence for any given number of masks. The lower mutual coherence is the reason for better reconstruction quality (higher SSIM) at all levels of compression.

3.7 Discussion and Summary

This chapter presented FemtoPixel, a framework for time-resolved compressive lensless imaging. This framework provides the user with design tools for situations in which lensless imaging is essential. It allows the user to effectively balance available resources. Furthermore, the FemtoPixel framework can be thought of as bridging the gap between traditional cameras on one end (pure hardware solution) and regular single-pixel cameras on the other (minimal hardware). This is demonstrated in Fig. 3-15. As noted in the figure, this continuum of camera design points provides a unique trade-off between dependency on software vs. hardware, and its effect on the acquisition time.

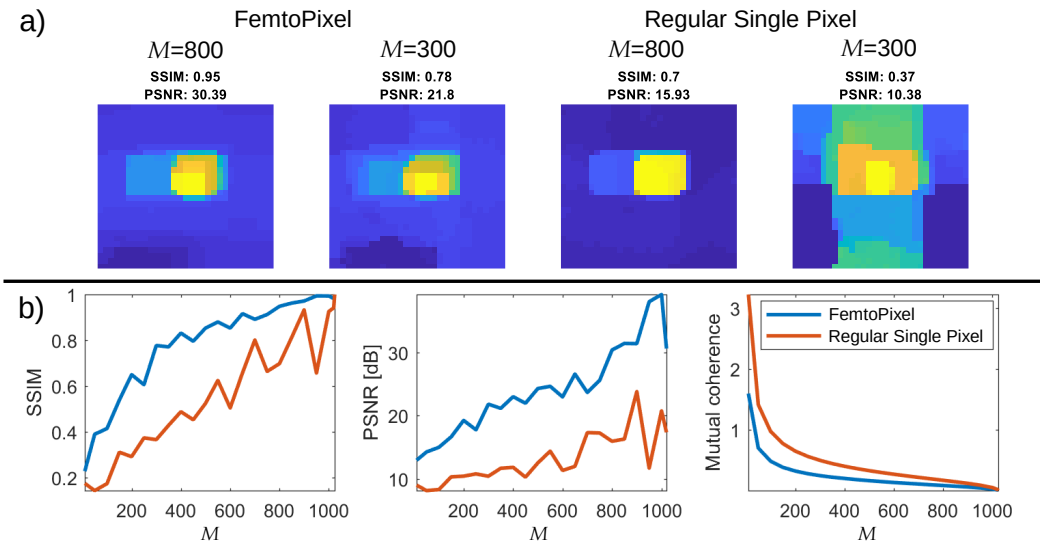

Figure 3-14: FemtoPixel experimental results. a) Recovery of a circle shape using FemtoPixel and compared to a regular single pixel camera with 300 and 800 masks. b) Mutual coherence vs. number of patterns for FemtoPixel framework and a regular single pixel camera, demonstrating the superior results of FemtoPixel.

3.7.1 Implementation Considerations

Section 3.4.3 analyzed only wavefront amplitude modulation. There are many other ways to use coded active illumination in order to minimize the mutual coherence. For example, we assumed the wavefront is just a pulse in time, but we can perform coding in the time domain as well. This will cause different pixels on the target image to be illuminated at different times. Physical implementation of such delays is possible with, for example, tilted illumination and fiber bundles (notice that while phase SLM induces varying time delays on the wavefront, these time scales are shorter than current time-resolved sensor resolutions). Analysis of such implementation requires detailed care with the interplay between the $\mathbf{H}$ and $\mathbf{G}$ matrices (since $\mathbf{G}$ becomes time-dependent). More specifically, in this case, $\mathbf{G}$ is creating a mapping from space to space-time, and $\mathbf{H}$ is a mapping from space-time to time i.e. from the $\mathbf{H}$ perspective, the scene is no longer stationary.

The forward model (Eq. 3.9) assumes the wave nature of light is negligible. This assumption is valid if: 1) Diffraction is negligible: the scene's spatial features are significantly greater compared to the illumination wavelength. 2) Interference is neg-

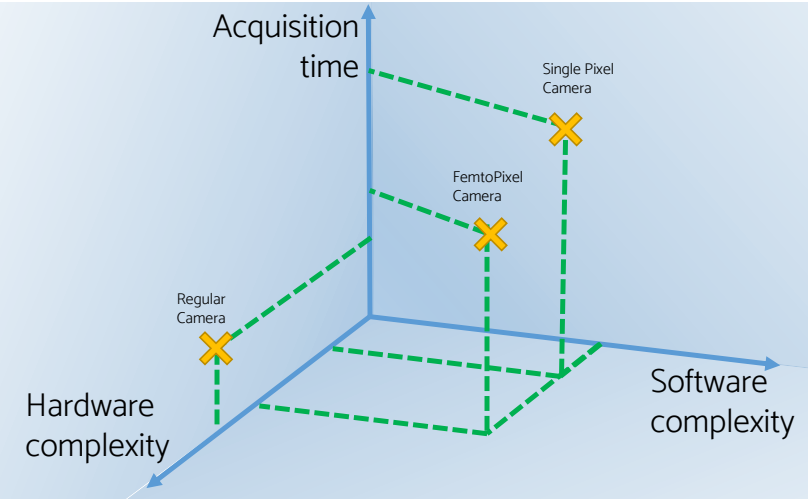

Figure 3-15: FemtoPixel provides a manifold of camera design options between traditional cameras and single pixel cameras.

ligible: the coherence length of the illumination source is significantly smaller compared to the geometrical features. For pulsed lasers, the coherence length is inversely proportional to the pulse bandwidth; this usually results in sub-cm coherence lengths.

Currently available time-resolved sensors allow a wide range of potential implementations. For example, streak cameras provide picosecond or even sub-picosecond time resolution [154]; however, they suffer from poor sensitivity. Alternatively, SPADs are compatible with standard CMOS technology [134] and allow time tagging with resolutions on the order of tens of picoseconds. These devices are available as a single pixel or in pixel arrays. It is interesting to note that adding a time-resolved capability to a single detector is usually divided into two processes: 1) creating a fast detector (fast rise and fall times), and 2) time to digital conversion. The second step (as described in Sec. 2.5.2) is independent of the detector, and can be a completely separate hardware. This is important for the manufacturing of novel sensors across the spectrum.

The illumination pattern may be constrained to binary values based on the implementation (DMD allows only binary values while an SLM and LCD allow grayscale values). The illumination masks optimization algorithm from Sec. 3.4.3 can be easily extended to produce binary values if required. For example, the regularizer can be

replaced such that the cost function is:

$$\{\mathbf{g}_j\}_{j=1..M} = \underset{\mathbf{g}_{j=1..M}}{\arg\min} \left\{ \left\| \mathbf{I}_L - \tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}} \right\|_F^2 + \lambda \sum_{j=1}^{M} \left\| (\mathbf{g} - 1)(\mathbf{g} + 1) \right\|_2^2 \right\} \tag{3.17}$$

This regularizer tries to produce values that are close to either -1 or 1. The regularization parameter can be adapted to produce masks with values that are arbitrarily close to these desired values. In our experiments, we did not observe any performance degradation when generating such binary masks.

3.7.2 Limitations

The main limitations of using our suggested approach are:

- We assume a linear imaging model (linear modeling in imaging is common, for example [45, 11]).
- We assume a planar scene (if the plane is not fronto-parallel, the rings are replaced by ellipses, which only affects the structure of $\mathbf{H}$). Furthermore, we require a known geometry.
- Time-resolved sensing requires an active pulsed illumination source and a time-resolved sensor. These can be expensive and complicated to set up. However, as we have demonstrated here, they provide a different set of trade-offs for lensless imaging, and reduce the overall acquisition time.

3.7.3 Conclusions and Future Work

Our work here considered known geometry and recovered scene reflectance. Notably, other works assume known reflectance and recover scene geometry with compressive sensing [87, 33]. These depth recovery works use time-resolved sensors to recover the depth in a similar way to LIDAR. It would be a natural next step to fuse the two approaches and iterate between geometry and reflectance recovery to recover both.

In summary, we demonstrated a novel compressive imaging architecture for using ultrafast sensors with active illumination for lensless imaging. We discussed analysis tools for hardware design, as well as algorithms for ideal sensor placement and illumination patterns, which directly target the RIP for robust inversion with compressive deconvolution. The presented approach allows lensless imaging with a single pixel and dramatically better acquisition times compared to previous results. This enables novel lensless single pixel imaging in challenging environments. The approach and analysis presented here open new avenues for other areas with potential tight coupling between novel sensors and compressive sensing algorithms. Furthermore, the time-resolved analysis for this simple blur case provides a foundation for the next chapters, in which we discuss more challenging scattering conditions.

Chapter 4

Data-Driven Computational Imaging Through Occlusions

One of the greatest challenges in inverse problems such as non-line-of-sight (NLOS) imaging is the need for a highly calibrated forward model. An uncalibrated forward model results in a model mismatch that poses a significant challenge to producing a reliable solution to the inverse problem. In fact, model mismatch can be harder to tackle than poor measurement SNR. Here we present a data-driven approach that results in an algorithm that is robust to model mismatch and poor calibration.

To demonstrate the advantages of data-driven computational imaging we demonstrate a calibration-free imaging technique that allows identification and classification of objects hidden behind a scattering layer (regular printer paper). We use a data-driven approach instead of tuning a forward model and directly inverting the optical scattering. With a Monte Carlo rendering model, we synthesize a large dataset that contains random realizations of all optical parameters. This allows us to use a deep neural network for classification of objects that are hidden from the camera.

We show that by training the network with data that includes variations in all model parameters, the network learns a representation that is not only invariant under traditional transformations like translation, but also invariant to variations in all the other model parameters. This effectively allows calibration-free imaging through scattering conditions.

The main technical contributions presented in this chapter include:

- Introducing a data-driven technique for NLOS computational imaging through scattering.
- A data-driven technique for imaging through scattering that is invariant to perturbations in calibration.
- A method to train data-driven techniques for NLOS imaging on synthetic data that is capable of generalizing to real-world experiments without any transfer learning or fine-tuning.
- A method that allows real-time classification through scattering medium and beyond the line of sight.
- A system capable of human pose identification while preserving user privacy.

4.1 Why Data-Driven Computational Imaging?

Figure 4-1 motivates the problem. In traditional NLOS imaging, the model is physics-driven. In that case a few target observations are used to build a forward model. This forward model is calibrated to the specific optical system. The next step in this offline stage (equivalent to training) is to build an inverse problem algorithm, that is an algorithm that takes a measurement as an input and produces the hidden scene as an output. At test time, the measurement is fed into the inverse problem algorithm to produce the target. The forward model can be used as part of an iterative solver, that together with the inverse problem tries to find the best target that explains the measurement. There are four important properties to this approach:

1. Almost all aspects of the problem are engineered, with a few or no "black boxes".
2. The physical model tends to have very few parameters.
3. The inverse model is built based on the forward model, and tends to be very sensitive to model mismatch and calibration.

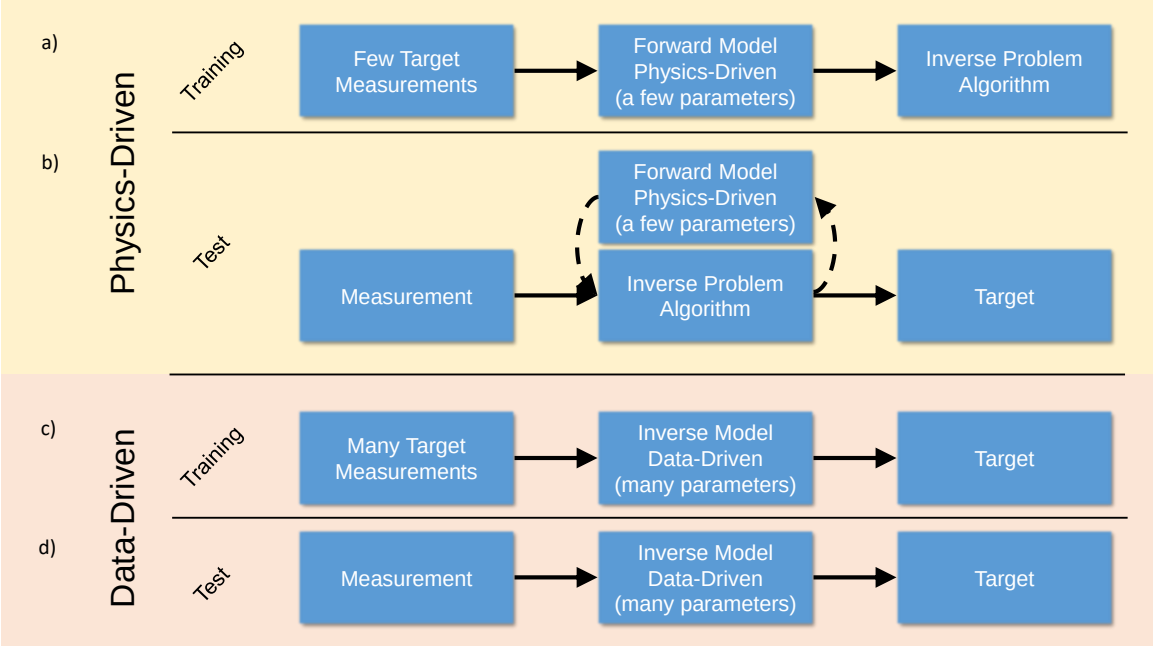

Figure 4-1: Data-driven vs. physics-driven NLOS imaging. Data-driven model directly learns the mapping from measurement to target, while a physics-driven approach requires an intermediate forward model.

An alternative to the physics-driven approach is a data-driven technique. In that case, many training examples are required. These are used to train a "black box" algorithm that directly learns the mapping from measurement to target. That is, there is no intermediate forward model representation, and there is no need to calibrate a forward model.

In this chapter, we show that a properly trained data-driven model is extremely robust to model mismatch when compared to a traditional physics based approach.

4.2 Related Works

Imaging through sparse (or thin) scattering layers has been demonstrated with various techniques as discussed in Chapter 2. For completeness of this chapter, we briefly provide the related works that are based on ToF. We note that imaging around a corner and through a thin scattering layer are both cases of sparse scattering and are very similar from both physics (modeling) and inverse problem perspectives.

Active ToF methods are commonly used for NLOS imaging [129]. Active NLOS imaging's most notable advantage is the large field of view and the ability for remote sensing. Different technologies have been demonstrated as impulse [63, 178] or phase [66, 67, 80, 79] based systems. Various aspects of imaging have been studied [17] and demonstrated, for example full scene reconstruction [178, 77], reflectance recovery [117], pose estimation [130], and tracking [53]. NLOS imaging has been demonstrated with impulse based systems like streak camera [178] and SPAD [22, 98]. All of these approaches are based on a forward model with many physical parameters that require calibration. Here we demonstrate a calibration-free approach for imaging through scattering.

Calibrating system parameters is required in many vision [197] and imaging systems. Having a calibration-free system relaxes many requirements on system design and usage scenarios, but it is usually very challenging. Some examples of calibration-free systems have been demonstrated in specific domains such as augmented reality [95, 192] and gaze tracking [161].

To achieve calibration invariant NLOS imaging we leverage convolutional neural networks (CNN), which have become the main workhorse in many computer vision tasks. Their power to perform dimensionality reduction from noisy data [179], object classification [160], segmentation [30], super resolution [43], classification of spatiotemporal data in videos [81, 173], and to capture invariants [14] makes them appealing for this application. The input data to the CNN here is a time-resolved measurement. Spatiotemporal convolutions have been shown to outperform single image neural networks [81] and long short-term memory (LSTM) networks [175].

In the context of imaging, neural networks have also been applied successfully in microscopy [182], compressive imaging [91], ToF [57], medical imaging [1], classification with coherent light [5, 138], and synthetic aperture radar domains [126]. CNN have been shown to increase image registration accuracy by learning more robust features [127] in SAR applications. Remote sensing with data-driven techniques has been suggested [31, 73] and demonstrated in dehazing with a CNN [23].

Here, we leverage the empirical observation that neural networks learn invariants.

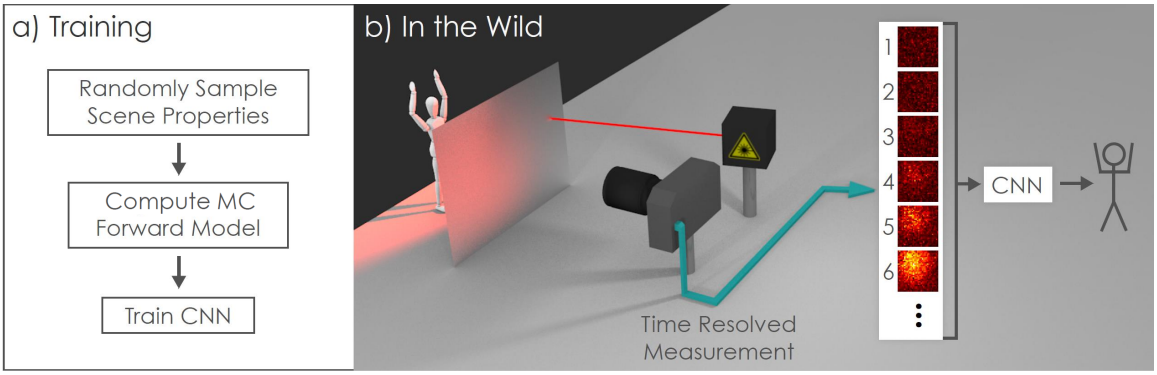

Figure 4-2: Calibration-free object classification through scattering. a) The training phase is an offline process in which synthetic data that includes variations in all physical parameters is used to train a CNN for classification. b) Once the CNN is trained the user can simply set up the optical system in the scene (SPAD camera and pulsed laser), capture measurements (six examples of time-resolved frames are shown), and classify the hidden object with the CNN without having to precisely calibrate the system.

The study of the invariants learned by deep neural networks is an active research topic [6, 14, 21, 58, 109]. In our work, we use synthesized data to train a CNN that is both invariant to changes in forward model physical parameters and is able to correctly classify hidden objects behind scattering media.

4.3 Calibration Invariant Target Classification Through Scattering Layer

The requirement for calibration when performing imaging through scattering is directly related to the need for a physical model that explains the measurements. Such physical models simulate light transport from the target to the sensor and depend on the geometry and other physical parameters of the system. Since inverting scattering is ill-posed, any mismatch between the physical model and the actual measurement will degrade performance. As a result, accurate measurement of physical parameters like illumination position, camera orientation, etc. is needed. This prohibits many inversion-based techniques to scale to real-world applications. The approach presented here allows calibration-free imaging for classification of objects hidden be-

hind scattering media or beyond the line of sight of the camera. More specifically, we demonstrate an active pulsed illumination with a single photon time-resolved sensitive camera, used to classify handwritten digits and the pose of a mannequin occluded by a sheet of paper (an example of sparse scattering).

The imaging pipeline is shown in Fig. 4-2. A ToF camera is used to increase measurement diversity. The hardware used here is a SPAD camera that time tags each detected photon. The imaging pipeline can be partitioned into two halves:

1. An offline process. A generic forward model is used to synthesize a large dataset of potential ToF measurements. The calibration parameters in the forward model are randomly sampled. The synthesized dataset is used to train a deep neural network. The resulting network is invariant to perturbation in calibration parameters, effectively allowing calibration-free imaging through scattering.

2. An online phase. During test time, the user can set up the optical system without calibration, and classify, in real-time, hidden objects occluded by the scattering layer.

4.3.1 Synthetic Data Generation

The synthetic data is generated with a Monte Carlo (MC) renderer. The imaging system is based on a 32×32 SPAD camera. Since the camera is single photon sensitive, an MC rendering system is very efficient as it directly simulates the measurement process with individual photons. MC is a very generic forward-modeling technique that can be easily modified to simulate various system geometries like looking around corners and seeing through scattering medium.

An MC renderer requires many parameters to properly produce a realistic result. We divide these model parameters into two categories: calibration and target parameters. In calibration parameters, we consider all of the geometry parameters, including illumination, camera calibration, scattering parameters, time jitter, and noise. The target parameters are specific to the target and include its position and scale (and of course the target label or instance). Table 4.1 lists the model parameters.

Calibration parameters	
Laser	
- Incident position	$L_P \sim U(-4, 4)$ cm
Diffuser	
- Scattering profile	$D_D \sim N(0, \sigma)$, $\sigma \sim U(0.8, 1.2)$ rad
Camera	
- Position	$C_P \sim U(-1.5, 1.5)$ cm
- Time resolution	$C_{TR} \sim N(0, \sigma)$, $\sigma \sim 56 + U(-5, 5)$ ps
- Time jitter	$C_{TS} \sim U(0, 3 * 56)$ ps
- Field of view	$C_{FV} \sim U(0.1, 0.2)$ rad
- Homography	Normal distributions
Noise	
- Dark count	$N_{DC} \sim U(3000, 9000)$ photons
Target parameters	
- Position	$T_{P_{x,y,z}} \sim U(-4, 4)$ cm
- Scale	$T_S \sim U(18, 30)$ cm

Table 4.1: List of parameters and distributions for calibration and target parameters used in mannequin dataset.

To achieve calibration invariant imaging, we must introduce significant variation in calibration parameters as part of the training data. This helps the network learn a representation that is robust to such variations. Such diversity in the training data is accomplished by randomly sampling the parameters. Thus for each label in the training data we have many examples that are a result of rendering different scenes generated by random model parameters. Table 4.1 provides the specific distributions from which the parameters are sampled.

Given the samples of calibration and target parameters, the scene is completely defined and can now be simulated with a MC single photon tracer. The ray tracer is used to simulate the propagation of individual photons from the illumination source, through the diffuser, onto the target, back to the diffuser and finally to the camera (see Algorithm 1). This process takes into account the propagation time of the individual photons. We note that each photon undergoes a random process generated by the scattering (once towards the target and again on the way back to the camera), and by the time jitter noise profile.

Algorithm 1 MC forward model

```
Initialize scene by randomly sampling:
    Target: label, instance, position, size
    Laser: incident position
    Diffuser: scattering profile
    Camera: position, time resolution, time jitter, field of view, homography parameters
for All photons do
    Calculate initial intersection point with diffuser
    Randomly sample diffuser local scattering profile
    Randomly sample photon's angle after diffuser
    Calculate photon's intersection point with target
    if does not hit target then
        continue to next photon
    end if
    Randomly sample angle after reflection from target
    Calculate photon's intersection point with diffuser
    if does not hit diffuser then
        continue to next photon
    end if
    Randomly sample diffuser local scattering profile
    Randomly sample photon's angle after diffuser
    Map photon to camera sensor using homography
    Randomly sample photon's arrival time jitter
    Store photon's arrival time (with jitter) and location
end for
Bin recorded photons into discrete time frames.
Add dark count noise to measurement
```

While a sheet of paper is a strongly scattering media with multiple scattering events, it can be modeled as a single scatter event due to: 1) the propagation time through the paper ($\sim$10 ps) [27] is much smaller compared to the time resolution of the SPAD camera. 2) The scene size (target feature size and scene length scales) are much larger compared to the scatterer thickness, so we can approximate the photon exit coordinate to be equal to the entrance coordinate.

The scattering paper is assumed to be heterogeneous; that is achieved by randomly sampling different scattering profiles for different positions on the paper. When a photon hits the paper, we first randomly sample the scattering profile parameter σ (here we assume σ is sampled from a uniform distribution). Given the scattering

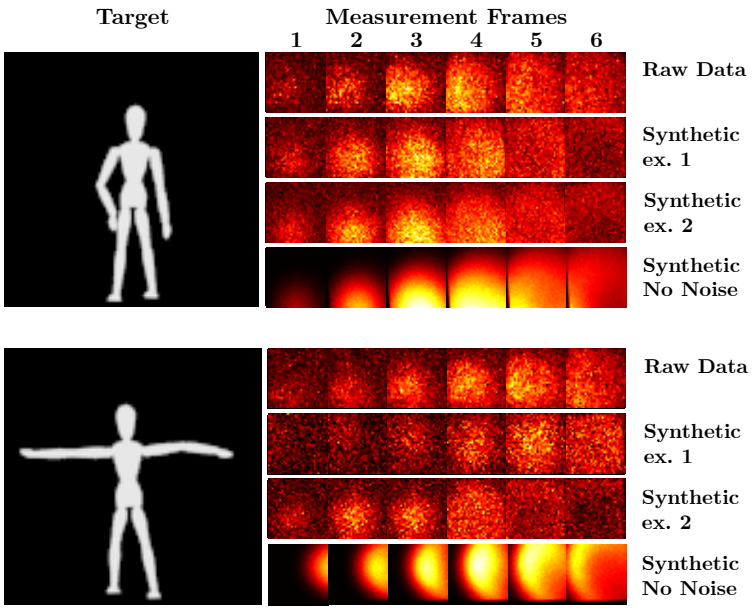

Figure 4-3: Comparison of SPAD measurement and MC model. The targets are two poses of a mannequin placed behind a sheet of paper. The data shows six frames (each frame is 32×32 pixels) of raw SPAD measurements, examples of two synthetic results generated by the MC model with similar measurement quality, and a synthetic result with high photon count and no additive noise. Note that differences between synthetic ex. 1, 2 and the raw measurement are due to the fact that the MC model was never calibrated to this specific setup. The synthetic images represent different instances chosen randomly from the dataset. The synthetic example with high photon count helps to distinguish between measurement (or simulated) noise and the actual signal as well as to observe the full signal wavefront.

parameter σ the photon's outgoing angle is deviated by an angle that is sampled from a Gaussian distribution $N(0, \sigma)$. Thus, every time a photon hits the paper, it encounters a different scattering profile, which simulates a heterogeneous medium.

SPAD array measurements are only based on thousands of detected photons. Since there is no need to render the full time-dependent scene irradiance, the computational burden of photon tracing with MC is low (we simulate 10^6 photons for each data point, which takes $\sim$1 second on a regular desktop computer). Figure 4-3 compares raw measurements taken with the SPAD camera and instances of the forward model (chosen randomly from the synthetic dataset).

Each data point in the training dataset corresponds to a specific target label measured by a system that is defined by random target and calibration parameters:

- The target is defined by a label and an instance (for example, a specific mannequin pose); these are simply selected from the dataset. For improved robustness it is preferred to increase diversity in all parameters. This is achieved by scaling the target with parameters that are sampled from distributions of plausible target size. Finally, the target is placed at a random 3D location behind the diffuser. The location is sampled from a uniform distribution which defines the NLOS volume of interest.

- The imaging system is defined by a realization of various calibration parameters that are sampled from random distributions. User input is involved only in determining the random distributions, which are defined based on approximate measurements, for example observation of the system geometry by the naked eye. If a parameter is easy to evaluate (for example, the laser position on the diffuser), it can be modeled with a Gaussian distribution with the known mean and small variance. Otherwise, it can be modeled with a uniform distribution.

Varying calibration parameters in the training data allows the CNN to be invariant to changes in those parameters within the training range (see below).

4.3.2 Model Training

The synthetic random dataset generated with the MC forward model is used to train a CNN for classification of hidden objects behind a diffuser. CNNs are a natural fit for this task since: 1) they have been shown to perform well in classification tasks, 2) they are designed to be invariant to translations, and 3) they learn to be invariant to other data transformations like scaling, rotation and, as demonstrated here, variations in the system calibration parameters.

Several neural network architectures were considered. The data structure in our case is composed of several frames, which is similar to the case of action recognition and gesture classification from short videos. Works such as [81, 163] indicated that convolutional architectures produce robust classification in that task. Thus, multiple convolutional architectures were evaluated including VGG [164], ResNet [65], and sev-

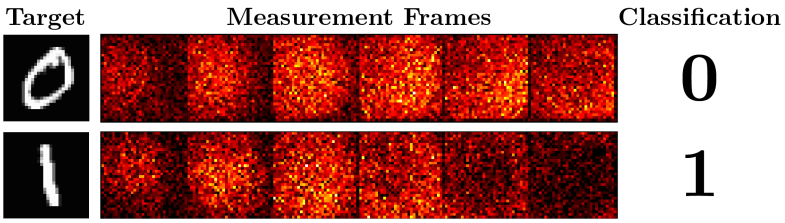

Figure 4-4: Successful classification of handwritten digits through scattering. The '0' and '1' digits from the MNIST dataset are placed behind a sheet of paper. Raw SPAD measurements are input into the CNN and correctly classified.

eral custom shallower networks with various combinations of layers. All architectures performed similarly on the classification task with marginally better performance for VGG. The VGG topology was selected and modified by extension of convolution filters into time domain (3D space-time filters). Filters were resized to $3 \times 3 \times 10$ where the last index denotes the time dimension. The training time on 60,000 data points is approximately two hours on an Nvidia Titan XP GPU.

4.3.3 Calibration Invariance Analysis

To evaluate our approach, we used the well-known MNIST dataset of handwritten digits. The goal is to evaluate the CNN's ability to classify hidden objects while being invariant to changes in calibration parameters. To that end, 60,000 training samples (6000 per label) and 10,000 test samples are synthesized with the MC model. Each data point is a realization of a different set of target and calibration parameters. The classification result on the test set is an overall classification accuracy of 74% (compared to a 10% random guess accuracy). These simulations demonstrate the ability to classify objects hidden behind a scattering layer without calibration.

As a proof of concept lab experiment, we cut two targets from cardboard shaped like zero and one digits, placed them behind a sheet of paper, and measured the response with the SPAD camera. The two time-resolved measurements were correctly classified as zero and one using the above network (Fig. 4-4). The training dataset generation and network training were performed prior to this data acquisition. This demonstrates that our method is robust to variations in calibration parameters on

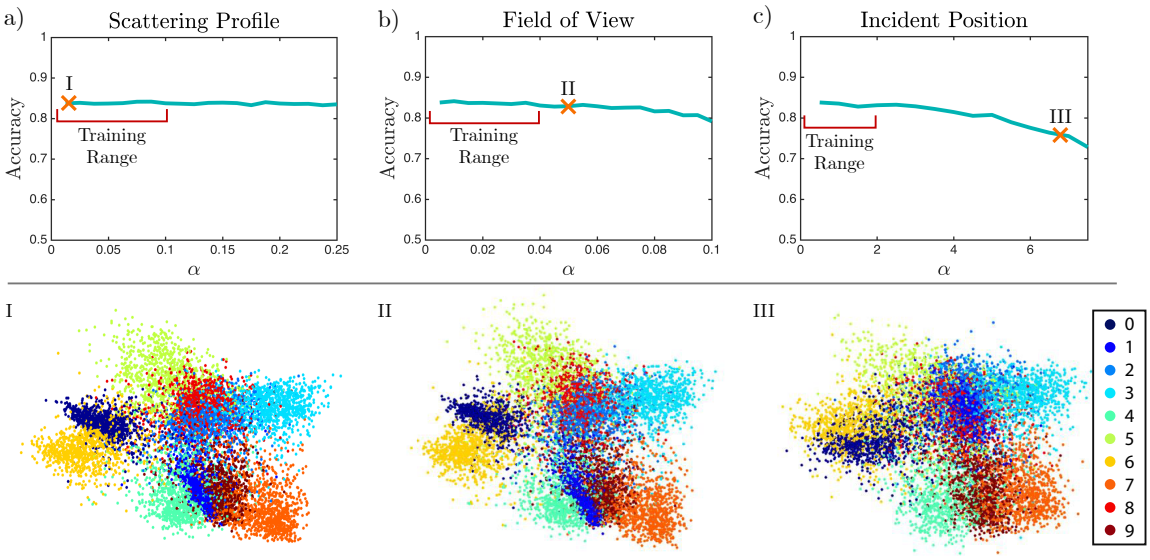

Figure 4-5: CNN learns to be calibration invariant. The CNN is trained with the complete random training set (based on the MNIST dataset), and evaluated with test sets in which all model parameters are fixed except for one that is randomly sampled from distributions with growing variance. Three parameters are demonstrated (other parameters show similar behavior): a) diffuser scattering profile variance $D_D \sim N(0, \sigma)$, $\sigma \sim U(1-\alpha, 1+\alpha)$rad; b) camera field of view $C_{FV} \sim U(0.15-\alpha, 0.15+\alpha)$rad; and c) illumination source position $L_P \sim U(-\alpha, \alpha)$ cm. The top plots show the classification accuracy as a function of the parameter distribution variance in the test set. Red lines show the ranges used for training. The 'X' symbols point to specific locations sampled for PCA projections in the bottom part of the figure. PCA projections show a color map where each digit has a different color. Performance is maintained beyond the training range and starts to slowly degrade further from it, as can be observed in PCA projection III, where more mixing is apparent at a test range 2.5× larger compared to the training set.

raw data. Sec. 4.4 provides more challenging experimental results.

In order to evaluate the extent of the network's ability to handle variations in calibration parameters, a set of controlled synthetic experiments were performed. We used the trained network with the MNIST dataset, and tested it with multiple test sets that were generated for the purpose of this evaluation. In each test set, all calibration parameters are held fixed (on the mean), except for one parameter that is randomly sampled from distributions with different variances. Thus, the CNN's sensitivity to variations in individual parameters is probed independently. Specifically, for each calibration parameter to be investigated, multiple test sets are generated, each one

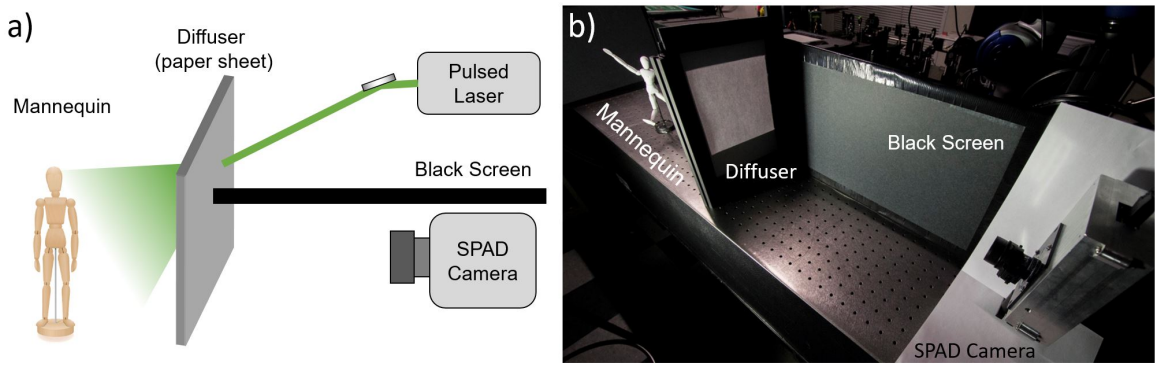

Figure 4-6: Optical setup. The setup was built after the network was trained and without calibration. a) Sketch of the optical setup. A pulsed laser incident on a diffuser (a sheet of paper) and illuminates the hidden mannequin. A SPAD camera is focused on the diffuser. A black screen is used to block direct reflection from the incident laser on the diffuser to the camera. b) Photograph of optical setup, the pulsed laser is hidden behind the black screen.

with a different distribution variance. The variance is scanned starting from zero (i.e. just the mean) throughout the range that was used for training, and then continues to grow beyond the training range, up to at least 2.5× of the training range variance.

Figure 4-5 demonstrates results for three calibration parameters (other parameters demonstrate similar behavior). As can be seen from the test accuracies, performance is maintained within the variance range used for training, and extended well beyond that range. This demonstrates the network's ability to learn a model that is invariant to changes in the calibration parameters within the training range and nearly invariant beyond that range. For example, in Fig. 4-5(c) the network was trained with data that had the illumination position distributed uniformly within 5 cm from the mean. Yet, the test performance starts to slightly drop only after the illumination position is found within 10 cm of the mean. Qualitative evaluations of these results are also presented in the bottom part of Fig. 4-5, with PCA projections of the activations from the penultimate layer of the CNN; these demonstrate sustained performance well beyond the training range.

This analysis shows that the network performance is maintained when the calibration parameters deviate from the mean within the training range. Furthermore, even if the network was trained under an assumption of certain ranges for system

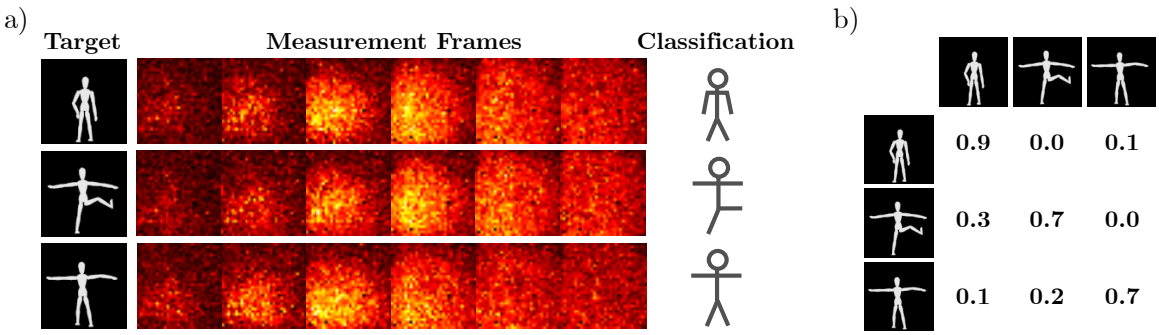

Figure 4-7: Calibration invariant classification of mannequin occluded by paper. a) Three examples (rows) demonstrate target pose, raw SPAD measurement (first six frames), and the successful classification. b) Confusion matrix for classification of raw test set (10 samples per pose).

parameters, the performance degrades slowly if the actual calibration parameters are outside the training range. Thus the network learns to generalize and extrapolate beyond the ranges used for training.

4.4 Experimental Results

The optical setup is shown in Fig. 4-6. A pulsed source (NKT photonics SuperK) with a repetition rate of 80 MHz and pulse duration of 5 ps is spectrally filtered to a band of 580 ± 10 nm. The camera is a single photon avalanche diode (SPAD) array (Photon Force PF32) with 32×32 pixels, and a time resolution of 56 ps. The laser is incident on the diffuser at $\sim 45°$. The camera is focused on the diffuser (a regular sheet of paper which presents non-uniform scattering properties). A black screen separates the camera from the incident position of the laser on the diffuser (to prevent direct reflection from the diffuser to the camera). The optical setup demonstrates a reflection mode geometry. The first 64 time bins of the SPAD measurement are used, such that the data structure is of size $32 \times 32 \times 64$ (the large number of frames guarantees consistency and flexibility of the data structure). Several examples of the measurement frames are provided in Fig. 4-3.

In this experiment, the occluded target is a flexible mannequin (20 cm head to toe). We define three different poses for the mannequin using various positions of

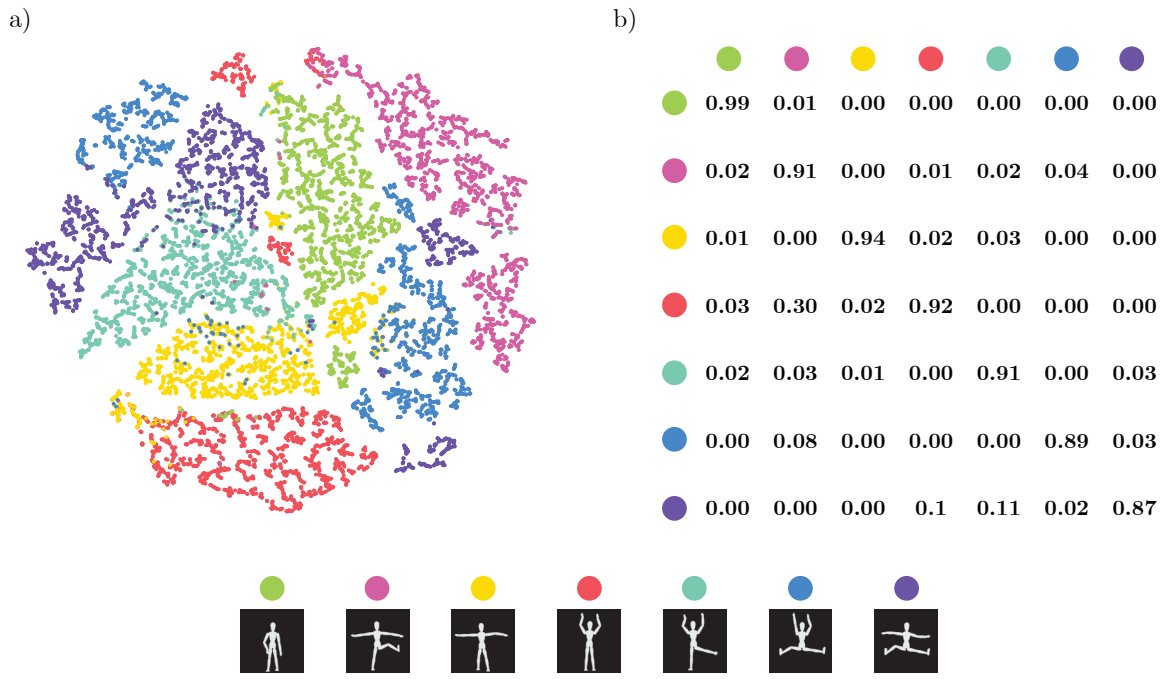

Figure 4-8: Calibration invariant classification among seven poses on synthetic test dataset. a) t-SNE visualization demonstrates the CNN ability to classify among the seven poses. b) Confusion matrix for classification on the synthetic test dataset.

hands and legs (Fig. 4-7). CNN training is accomplished by synthesizing 24,000 samples for training and 6000 samples for validation. Translations and perturbations to the mannequin's head and limbs are applied to create multiple instances of each pose. The test set is composed of 30 raw SPAD measurements, 10 per pose. For each measurement, the mannequin is moved around and the position of the hands, legs, and head are adjusted. The CNN classifies correctly 23 out of the 30 tests (76.6% overall accuracy, compared to 33.3% random guess accuracy). Fig. 4-7(a) shows examples of mannequin poses, SPAD measurements, and classification. Fig. 4-7(b) shows the confusion matrix of this raw test set.

Here training is performed on one dataset (synthetic) and tested on another dataset (gathered by lab experiments). In general, it is challenging to train and test on different datasets and it is common to note performance degradation in such cases. The degradation in performance can potentially be mitigated with domain adaptation methods (e.g. [94]). It is important to note that the training set was generated and the CNN trained prior to the experimental setup construction, which

further demonstrates the robustness of the proposed method.

To further explore the sensitivity to the number of poses we expanded the training set to include seven different poses (Fig. 4-8 bottom shows illustrations of the poses). The poses include a diverse combination of limb positions. For each label 8000 training examples and 2000 test examples were generated (total training set of 56,000 examples and 14,000 test set examples). Figure 4-8(a) shows a two dimensional student's t-distributed stochastic neighbor embedding (t-SNE) [107] visualization of activations from the CNN penultimate layer generated on the test set. This visualization demonstrates that the network correctly separates the classes. Fig. 4-8(b) shows the confusion matrix for this synthetic test set. The network is able to classify the seven classes with 91.86% accuracy (compared to 14.29% random accuracy). The synthetic test accuracy for the network trained only on the three poses (Fig. 4-7) achieved 96.7% (compared to 33.33% random accuracy). This indicates the ability to experimentally classify among more poses without a significant decrease in accuracy.

4.5 Evaluation

To evaluate our approach we compare its classification performance to several other classification techniques. The classification task is based on the three mannequin poses. We create two datasets for evaluation, each one consists of $24,000$ training examples and $6,000$ test examples. The clean dataset demonstrates the algorithms' sensitivity just to variation in calibration parameters (decoupling the sensitivity to measurement quality). The realistic dataset probes the algorithms' ability to classify on the actual lab experiments.

1. *Clean dataset*: This dataset aims to probe the ability to classify under extreme variation in calibration parameters in a noiseless measurement case. It is based on synthetic measurements with calibration parameters varying in ranges that are twice as large compared to the realistic dataset, and with 10^8 photons without any additive noise (Fig. 4-3 shows two noiseless examples from this dataset). In this case, both training and testing datasets are synthetic.

Training set	Clean dataset	Realistic dataset
Mean Example	33.3	33.3
KNN	53.0	30.0
SVM	57.1	20.0
Random forest	68.8	30.0
Single layer network	68.2	23.8
Our CNN	84.0	76.6

Table 4.2: The proposed approach outperforms other techniques on clean and realistic datasets. The CNN outperforms all methods in the clean dataset, and is the only method that achieves results that are better than random accuracy on the realistic dataset.

2. *Realistic dataset*: This is the dataset used for training the network described in section 4.4. It is based on renderings with 10^6 photons with an additive noise to approximate our SPAD measurements (see Fig. 4-3 synthetic examples 1 and 2). In this case the training is performed on the synthetic data and testing is based on the 30 lab measurements.

The results are summarized in Table 4.2. While some of the traditional algorithms perform reasonably well on the clean dataset, they fail on the realistic dataset. Our approach significantly outperforms the traditional algorithms on the clean dataset, and as demonstrated previously, it performs well on the realistic lab measurements, while the other methods fail (achieve random accuracy or below).

The different classification approaches that were used for comparison are:

1. *Mean example*: For each label we take the mean of the training data, such that we have one representative sample per label. Classification is performed based on the nearest neighbor (closest sample in the dictionary to the measurement). This approach fails on both datasets.

2. *K-nearest neighbors*: Since this method may be sensitive to dictionary size, it is first evaluated on the clean dataset. We randomly choose a varying number of samples from the training set to form different dictionary sizes. We consider two approaches here: a) nearest neighbor — for each test point the chosen label is the label of the closest dictionary element. b) K-nearest neighbors

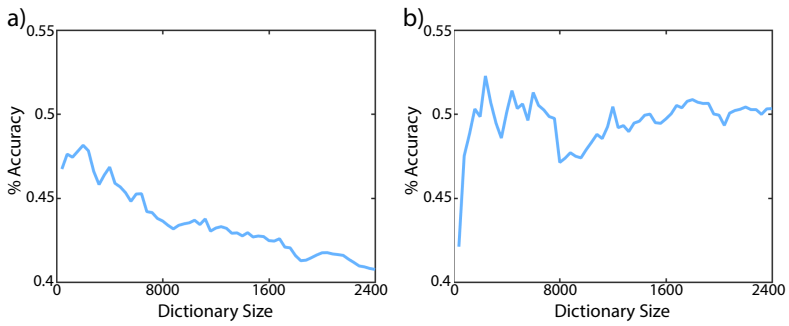

Figure 4-9: Performance of the K-nearest neighbor approach on the clean dataset. Classification accuracy with varying dictionary size for a) nearest neighbor classifier, and b) K-nearest neighbors classifier.

(KNN) — for each test point, the chosen label is the label of the majority of the K-nearest neighbors. K is chosen for each dictionary size with a validation set (taken from the training set). These results are presented in Fig. 4-9. To account for sensitivity to the order of data selection, the experiment is repeated and averaged over eight instances, and each time the training set is shuffled.

The nearest neighbor approach shows decreased performance with an increase in data size due to the increased ambiguity between dictionary elements. The K-nearest approach is able to overcome this limitation and provides classification accuracy in the range of 50% on the clean dataset. However, it fails on the realistic dataset.

3. *Support vector machine (SVM)*: The SVM is evaluated with different kernels, and achieved the best performance with a linear kernel. After hyper-parameters optimization, the SVM achieves 57.1% classification accuracy on the clean dataset and fail on the realistic dataset.

4. *Random forest*: A random forest is trained with 100 trees. The random forest achieves 68.2% accuracy on the clean dataset, and fails on the realistic dataset.

5. *Single layer network*: A neural network composed of one hidden layer. This network achieves 68.2% accuracy on the clean dataset, and like the previous methods, it fails on the realistic dataset.

This analysis presents the key difficulty and the requirement for both calibration invariance and robustness to noise. While some of the traditional approaches perform reasonably well on the clean data, they fail on the realistic dataset. Our approach not only outperforms the other techniques on the clean dataset, but it is also the only one that achieves results that are better than random accuracy on the realistic dataset.

4.6 Discussion

4.6.1 Limitations

The main limitations of the proposed technique are:

1. While our approach is invariant to variations of calibration parameters within the training range, it still requires some approximate measurements or knowledge of system parameters and geometry. This limitation is somewhat mitigated by the fact that the network can operate well beyond its training regime (as demonstrated in Fig. 4-5).

2. Another limitation is the need to synthesize a dataset and train the CNN on different types of geometries, which might slow down the process when arriving at a completely new setting. Faster hardware for data generation and CNN training can potentially address this in the future.

3. Active acquisition systems like the ones used here may suffer from interference with ambient illumination. This can be more challenging with single photon counting sensors. One possible solution is the use of narrow-band spectral filters to pass only the source's wavelength. These filters are already used in systems such as LIDARs.

4.6.2 The Importance of Time Resolution

The measurement system suggested here uses time-resolved measurements with few spatial pixels (32×32). The importance of temporal resolution for classification when

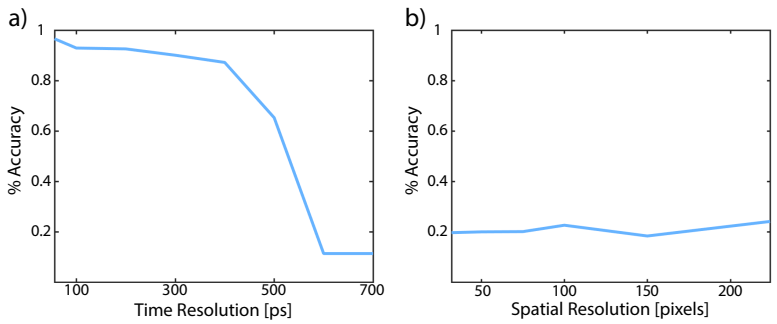

Figure 4-10: Time resolution is more important than pixel count for imaging through scattering. a) Classification accuracy vs. time resolution (for 32x32 pixels). b) Classification accuracy vs. number of pixels (for a non time-resolved system).

imaging through scattering media is evaluated with the suggested imaging pipeline. The MC model is used to create training and test sets (based on the MNIST dataset) with different detector time resolutions. The results are plotted in Fig. 4-10(a), where we note that the performance degrades slowly until the time resolution nears 400 ps and then degrades rapidly. In the scenes discussed and analyzed here, the time between the first and last signal photons spans roughly 500 ps, so any time resolution better than that provides at least two frames with a signal which allows the network to learn temporal filters.

As seen from the measurements provided in Fig. 4-3, the spatial features have very little high-frequency content, and therefore, unsurprisingly, low pixel count is sufficient for classification. To quantitatively evaluate this, we use the same pipeline to simulate no time dependency, while varying the pixel count. Fig. 4-10(b) demonstrates that simply adding more pixels doesn't improve the classification accuracy.

This analysis is limited to the particular scene considered here and evaluates two extremes: low pixel count with varying time resolution and no time resolution with varying spatial resolution. This demonstrates theoretical performance of commercially available hardware variants.

4.6.3 What Does The CNN Learn?

The importance of time-resolved data for classification with CNN can be observed from the filters the network learns (Fig. 4-11). Inspection of these indicates that the

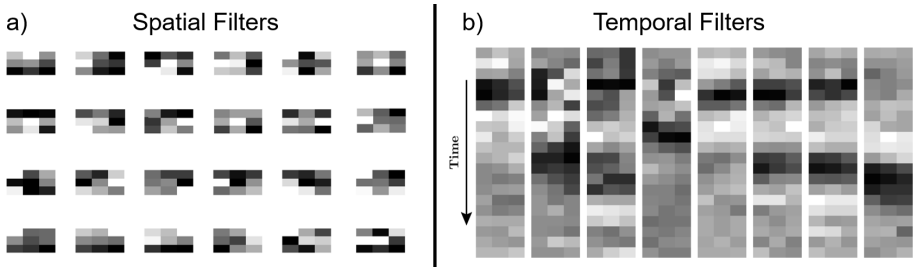

Figure 4-11: Examples of spatiotemporal filters learned by the CNN. The network generates both a) spatial and b) temporal filters for inference.

network performs derivatives in the time domain. Similar spatiotemporal features have been demonstrated when using CNNs for action recognition in videos [81]. The temporal features learned by our network combined with the strong dependency of classification accuracy on the SPAD's time resolution demonstrates the significance of time resolution to data-driven imaging through scattering.

4.6.4 Scaling To Real-World Scenes

Several aspects can be taken into account when considering the potential of this approach to scale into real-world applications:

- **Hardware:** Our hardware is a SPAD camera. Since SPAD cameras are manufactured with scalable semiconductor processes, they can be commoditized. Other approaches, like phase based ToF systems, are also a possibility (probably with significantly lower time resolution, which would impact its ability to classify). This is another demonstration that SPAD cameras are especially useful for imaging through scattering since:

  - they are single-photon sensitive, which is extremely useful in NLOS geometries where the optical signal is very weak;
  - the time resolution of $\sim$50 ps corresponds to 1.5 cm of spatial resolution, which is reasonable for room-sized scenes;
  - the low spatial resolution is not necessarily a drawback (as discussed above).

- **Real-time operation:** Since classification requires only a forward pass through the trained neural network, it can be performed in real-time using specialized hardware (such as GPUs). The only caveat is the case of a completely new scene that requires rendering new synthetic dataset and training a CNN. This requires anticipatory preparation before the real-time operation.

- **Flexibility:** The suggested forward model is based on an MC photon tracer. The MC model is very flexible and can render a wide range of optical geometries and materials.

4.7 Conclusions and Future Work

We introduced a data-driven method for object classification through a sparse scattering layer that is effectively invariant to variations in calibration parameters. This approach leverages the ability of neural networks to learn invariants to perturbations that are introduced during training. We demonstrated that the network is invariant to changes in the forward model parameters within the training range (and nearly invariant beyond that range) for the purposes of classification. An important cornerstone of our approach is its ability to generate synthetic data based on a generic forward model that is used to train and evaluate the neural network. This data-driven approach can alleviate lengthy experimental calibrations that are traditionally required in computational imaging systems.

The future of data-driven computational imaging is very promising. As discussed at length in the introduction to this chapter, data-driven techniques directly solve the computational imaging task, instead of the traditional approach which starts with a forward model that is later inverted. Chapter 7 provides specific examples of future data-driven computational imaging directions.

Chapter 5

Imaging Through Tissue

One of the main motivations for imaging through scattering media with near-visible light is medical imaging. Medical imaging has revolutionized medical diagnosis, as was recognized by several Nobel prizes (e.g. X-ray - 1901, CT - 1979, and MRI - 2003). Pictorially, it provides physicians with the ability to non-invasively see inside the body. Notably, these common imaging techniques are not based on visible light. X-ray, Ultrasound, and MRI were chosen since the energy in this spectra scatters very little as it interacts with tissue. Thus, it is relatively easy to ignore or reject scattered energy for imaging purposes. For example, in x-ray imaging, part of the sensor architecture rejects scattered photons based on their angle of arrival. So the advantage of non-visible energy is clear. However, as mentioned in the Introduction (Chapter 1), there are several key motivators for medical imaging with visible light, below we review these advantages specifically for medical imaging applications:

1. Non-ionizing radiation: some populations (e.g. pregnant women, children, cancer patients) cannot be exposed to x-ray radiation. This limits the availability of this imaging modality to a significant part of the population that actually requires it the most (e.g. ultrasound is routinely used to examine the fetus during pregnancy). It would be beneficial to have more non-ionizing imaging modalities beyond these specific populations as well.

2. Blood oxygenation measurement is a useful marker for tissue activity. Oxy-

genated and deoxygenated blood have different absorption spectra in the NIR; this property is the basic mechanism enabling pulse oximeter devices. Other imaging modalities (X-ray, MRI, Ultrasound) do not allow such measurement. Currently, pulse oximeter-like devices do not provide spatial information and only measure bulk or averaged properties. Imaging through tissue would enable measuring the 3D distribution of oxygenated blood in tissue or brain.

3. Functional imaging: it is common to use fluorescent markers in medical imaging. These markers are engineered to attach to specific types of tissue (like malignant tumors). Furthermore, the markers are engineered to fluoresce in different wavelengths. It is common to use such markers and contrast enhancements in CT and MRI tests. This enables creating maps that show geometry (like regular imaging) overlaid with functional information. Having this ability in visible light would be very beneficial since it is easier to engineer such markers to respond in visible light. The main challenge in this case is the scattering of light which we tackled in [142].

4. Optical contrast: different types of tissue have different optical properties that make their appearance different in visible light. However, in x-ray there is less variability, thus it is hard to distinguish among different types of tissues in x-ray (and similarly in Ultrasound). Imaging with visible light can create maps of different types of tissues in the body.

5. Resolution: imaging resolution is often limited by the wavelength. The wavelength of NIR ($\sim 800\,\text{nm}$) is much smaller compared to ultrasound wavelengths ($\sim 0.1\,\text{mm}$). This property allows high-resolution imaging with visible light.

Our previous works [139, 141, 142] tackled some specific examples of medical imaging (skin perfusion measurement, and locating and classifying fluorescent tags through scattering media), but focused on sparse scattering. This chapter demonstrates a technique for seeing through thick tissue.

As described in the related works (Chapter 2), many imaging techniques to see through scattering media perform different types of gating on the non-scattered light.

This is highly inefficient since the majority of the photons do scatter. In transmission optical mode, the probability to measure a photon that didn't interact with the media drops exponentially with the medium thickness and the inverse transport mean free path. This is modeled by the Beer-Lambert law (Eq. 2.4). As a result, such gating techniques must have a long integration time to measure enough un-scattered light. Thus, they are not applicable to non-stationary targets. In the context of medical imaging, the integration time is limited by the time dynamics of tissue such as blood flow speed. Beyond the long integration time, such gating techniques are limited by the finite size of the gate; as a result, they will always measure some photons that did scatter, which degrades the measurement.

Our motivation is to eliminate these two limitations. We note that the scattered light contains substantial information about the target and the scattering medium. Thus, instead of rejecting the scattered light in the measurement process we propose to measure it and computationally invert the scattering process. This approach allows us to improve our measurement SNR since we measure all of the optical signal, and to avoid the limitations of a finite gate. Because our technique is based on a measurement of the entire optical signal we call it "All Photons Imaging" (API).

The main technical contribution of this chapter is the introduction of API, a new technique for imaging through scattering based on the entire optical signal. API has several key advantages:

- It allows imaging through volumetric and highly scattering media (mean free path $\ll$ medium thickness). We demonstrate API in imaging through a 1.5 cm thick tissue phantom, and achieve a spatial resolution of 5.9 mm.
- API doesn't require prior knowledge of the material thickness and its optical properties – making it calibration free.
- API is invariant to variations in the optical properties of the material along the optical axis. Thus it naturally supports layered structures such as skin tissue.
- API is wide-field and non-invasive. It is capable of remotely imaging tissue and resolving the target occluded on the other side.

- API does not require any form of raster scan. Other imaging modalities with visible light such as diffuse optical tomography require consecutive measurements through a raster scan of the target.

5.1 Related Works

As discussed in Chapter 2, imaging through scattering media is a widely explored problem. For completeness, we provide relevant related works to medical applications.

For sparse scattering materials, such as very thin tissue, many pure hardware solutions have been suggested that are based on coherence [60, 76], time-reversal [191, 177, 115, 86, 74] , time-of-flight [117, 142], and speckle correlations [16, 84]. In the case of volumetric scattering which is more common in medical imaging, more sophisticated techniques are required. These commonly leverage a combination of acoustics and optics (acousto-optics [190] and photo-acoustics [185]), nonlinear effects (two-photon microscopy [68]), and time gating [183]. These methods utilize a physical separating parameter to lock onto the small set of ballistic photons that have not scattered. Unlike these methods, API is an all-optical technique that does not rely on intrinsic properties of the optical signal (coherence, polarization, etc.). Since API is calibration-free and wide-field, it is appealing for full organ imaging [121].

One common technique to image through scattering that goes beyond a pure hardware solution is Diffuse Optical Tomography (DOT) [37, 40]. DOT usually assumes the diffusion model, and aims to computationally invert it [18]. DOT is commonly performed in time [96] and frequency [36] domain. DOT has been used to image small animals [93], and the human cortex [196, 90] and breast [35, 32]. DOT has also been demonstrated along with fluorescence markers [133, 75]. We compare API to DOT in Table 5.1. DOT performs a sparse sample of the space-time profile, unlike our demonstration of dense sampling for API. DOT requires some form of raster scanning of the illumination source, unlike API's flood illumination which allows single shot measurement. Finally, API does not assume the specifics of the diffusion model, which allows it to operate in a wide range of conditions and avoid model mismatch.

	API	DOT
Measurement and use of full-dense spatio-temporal profile	Yes	No, DOT performs a sparse sampling of space. Some methods use full field cameras (but even if they perform time resolved measurement, it is of low time resolution).
Entire scene illuminated simultaneously	Yes	No, The illumination source is raster scanned or multiple sources are illuminated sequentially.
Field of view	Variable, Illumination is flood illumination so it doesn't pose a restriction. Measurement is done with a remote camera, so wider/smaller field of view is a simple function of camera lens.	Fixed, Usually based on rigid systems. Some methods use a standoff camera but still require raster scanning (i.e. limited flexibility in illumination field of view).
Contact with target	No, Applicable to remote sensing.	Yes, Usually requires contact.
Requires rigid structure around target	No	Yes, A rigid structure of illumination and sensing probes.
Potential for model mismatch	Limited (model is flexible)	High (assumes the diffusion model)
Works with fluorescence markers	Potentially (Not demonstrated)	Yes
Works with heterogeneous materials	Partial (only layered materials)	Yes
Recovered Information	Absorption coefficient of the target, Scattering coefficient of the equivalent uniform material.	Absorption and scattering coefficient of the medium and target

Table 5.1: Comparison of API and DOT.

5.2 All Photons Imaging Algorithm

A high-level overview of API is presented in Fig. 5-1. A pulsed laser flood illuminated the target in optical transmission mode. The target is adjacent to the tissue phantom. The other side of the tissue phantom is imaged by a streak camera. A scanning mechanism with the streak camera provides the space-time-resolved measurement. Section 5.3 provides additional experimental setup details.

The space-time measurement of the streak camera is effectively an ultrafast video (x, y, t). The duration of each frame is 2 ps. Different frames capture photons that traveled different optical paths inside the tissue phantom. The earlier frames show photons that scattered less, but are very noisy since there is a small number of such photons. Later frames show photons that traveled longer paths inside the tissue (scattered more), but are not noisy since there are many scattered photons. Another perspective is that the earlier frames encode more information about the target while the later frames encode more information about the tissue phantom.

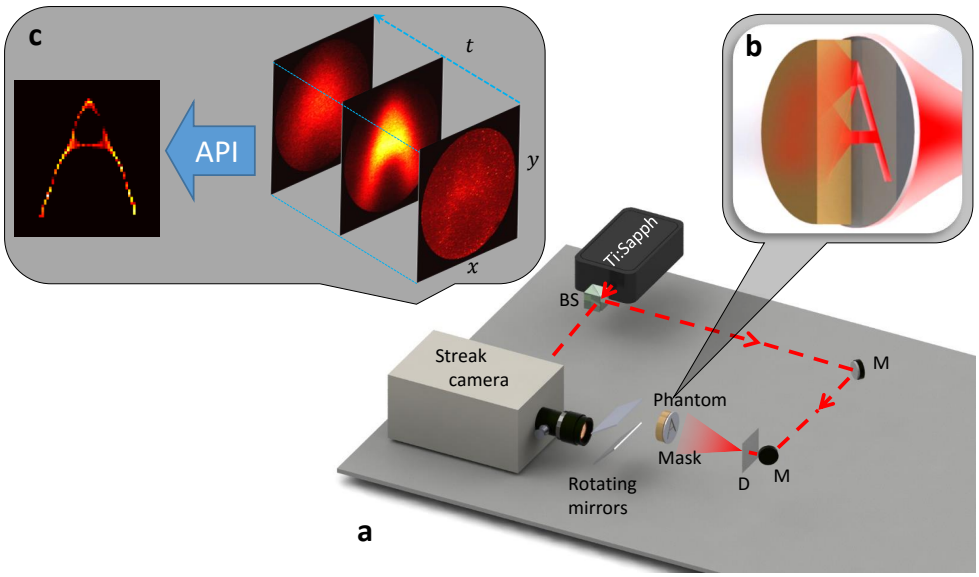

Figure 5-1: All Photons Imaging through tissue. a) Optical setup. A pulsed laser is scattered by a diffuser sheet (D) to flood illuminate the target mask. The mask is adjacent to the 1.5 cm tissue phantom which is imaged with a streak camera. b) Schematic of the scattering process inside the tissue phantom. c) The optical setup captures the time-resolved measurement. Each frame corresponds to a different arrival time of the distorted signal from the mask. Using API the mask is recovered.

Since API is designed to be calibration-free (doesn't assume prior knowledge of the tissue phantom), it is effectively a blind deconvolution problem where our goal is to recover both the scattering and target from the measurement. API avoids the need to solve this blind deconvolution problem by recovering the blur (convolution) kernel from the measurement without any assumptions about the target. This is possible due to the time-resolved measurement. We note that the target is not a time-dependent object (at least not with picosecond time scales). The rich time dynamics that are noticeable in Fig. 5-1c are purely due to the scattering. Thus the scattering is both blurring the target, and also increasing the measurement dimensionality (introducing more information). This provides yet another perspective of our measurement process, where each frame is a measurement of the target corrupted by a different blur kernel. The different blur kernels along the different frames are related to one another by the physics of the scattering process.

5.2.1 Forward Model

We denote the unknown target as $s(x, y)$ and the measurement as $m(x, y, t)$. These are related to each other with a convolution kernel $K(x, y, t)$ such that:

$$m(x, y, t) = K(x, y, t) * s(x, y) \tag{5.1}$$

where $*$ is the convolution operator performed on the x, y coordinates. Note that it is clear from Eq. 5.1 that $K(x, y, t)$ is performing a mapping from space to space-time. This is similar to the space-time mapping performed by the **H** operator in Chapter 3.

Equation 5.1 presents our blind deconvolution problem where both $K(x, y, t)$ and $s(x, y)$ are unknown.

5.2.2 Signal Independent Scattering Kernel Recovery

Here we describe our solution to recover $K(x, y, t)$ from $m(x, y, t)$ without any assumption of $s(x, y)$. Thus it allows us to avoid the blind deconvolution problem.

First, we note that we can consider $K(x, y, t)$ from a probabilistic perspective. Consider a point source that sends a single photon. $K(x, y, t)$ can be thought of as the probability density function to measure that photon at (x, y, t). With this observation we can rewrite $K(x, y, t)$ with the Bayes rule:

$$K(x, y, t) = f_T(t)W(x, y|t) \tag{5.2}$$

Here, $f_T(t)$ is the probability density function to measure the photon at time t, and given that time $W(x, y|t)$ is the probability density function to measure the photon at location (x, y). It is important to note that effectively we didn't make any assumptions in the transition to Eq. 5.2.

Equation 5.2 is very powerful since it reduces the search space over $K(x, y, t)$. More specifically, since $f_T(t)$ is not a function of space it is clearly independent of $s(x, y)$ and should be easy to estimate. Eq. 5.2 helps us to interpret the measurement process. $f_T(t)$ defines the overall probability to measure photons at different times,

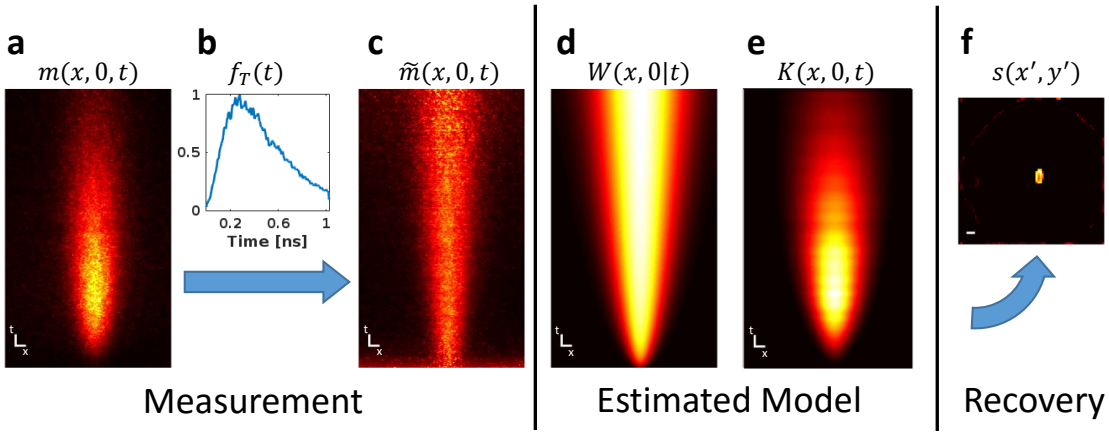

Figure 5-2: All Photons Imaging model estimation. Demonstrated on a point source example. a) Raw measurement. b) Recovery of time-only function $f_T(t)$. c) The normalized measurement $\tilde{m}(x,y,t)$. d) Recovered $W(x,y|t)$ after estimating its parameters. e) The final estimated kernel $K(x,y,t)$ after multiplying the estimated $f_T(t)$ and $W(x,y|t)$. f) Recovery of the point source - the result of the deconvolution procedure. Panels a,c,d,e show cross sections of the $x-y-t$ functions for $y=0$.

for example it shows low probability to measure photons at an early time (ballistic photons), and higher probability to measure photons at later times (scattered photons). Furthermore, $W(x,y|t)$ is exactly our earlier interpretation where each frame is a result of the target distorted by a different blur kernel.

We note that since $f_T(t)$ and $W(x,y|t)$ are probability functions, they should be normalized to 1. Thus it is likely that $W(x,y|t)$ will have a time-dependent normalization coefficient. For simplicity in our recovery, we absorb this coefficient into $f_T(t)$, and estimate an un-normalized version of $W(x,y|t)$. Since we only care about recovering $K(x,y,t)$ this has no effect on the overall recovery process.

The next two sections describe how we estimate $f_T(t)$ and $W(x,y|t)$ from the measurement. We experimentally demonstrate this estimation step-by-step in Fig. 5-2 with an example of a 4 mm point source at the center of the mask plane. It is important to mention that we only use a point source for illustrative purposes; our results of recovering more complicated targets use exactly the same reconstruction procedure. Fig. 5-2 shows cross sections of the measurement and estimated functions for $y=0$ i.e. $m(x,y=0,t)$. The time coordinate is presented on the vertical axis.

Estimating the time-only function $f_T(t)$

Plugging Eq. 5.2 into Eq. 5.1 we get:

$$m(x, y, t) = f_T(t) \left[W\left(x, y|t\right) * s\left(x, y\right) \right] \tag{5.3}$$

Since $f_T(t)$ is independent of space coordinates, we can simply sum over the $x - y$ coordinates and get $f_T(t)$ up to a scalar. We found that searching for the coordinate (x_0, y_0) in $m(x, y, t)$ with the strongest signal performs better. Thus we set: $f_T(t) = m(x_0, y_0, t)$. Fig. 5-2b shows the estimated $f_T(t)$ in the considered point source example. We note that the curve captures information about all photon transmission modes (ballistic, snake, diffuse, etc.) without any enforced physical model. Fig. 5-2b demonstrates the low probability to measure ballistic photons (early time bins), high probability to measure scattered photons, and diminished probability to measure very late photons due to the finite optical signal.

Because $f_T(t)$ does not carry any spatial information, it does not help us in the reconstruction process. Thus, we normalize the measurement by it before proceeding to the next step:

$$\tilde{m}(x, y, t) = \frac{1}{f_T(t)} m(x, y, t) \tag{5.4}$$

Fig. 5-2c shows the normalized measurement.

Estimating the time-dependent scattering kernel $W(x, y|t)$

To model $W\left(x, y|t\right)$ we use a probabilistic interpretation again and use the central limit theorem. Since a photon performs a random walk (or Brownian motion) inside the tissue, its final measurement (x, y, t) is a result of the sum of many scattering events. The central limit theorem then predicts that the final measurement would be distributed with a Normal distribution. In this case, it results in a Normal distribution with a time-dependent variance:

$$W\left(x, y|t\right) = \exp\left\{ -\frac{x^2 + y^2}{\sigma^2(t)} \right\} \tag{5.5}$$

To estimate $\sigma^2(t)$ we note the following connection between different frames. Combining Eqs. 5.3, 5.4 and selecting a specific frame at time t_i we get:

$$\tilde{m}(x,y,t_i) = \exp\left\{-\frac{x^2+y^2}{\sigma^2(t_i)}\right\} * s\,(x,y) \tag{5.6}$$

Taking the Fourier transform with respect to the spatial coordinates results in:

$$\mathcal{F}\left\{\tilde{m}(x,y,t_i)\right\} = \beta_i \exp\left\{-\sigma^2(t_i)\left(\omega_x^2+\omega_y^2\right)\right\}\mathcal{F}\left\{s(x,y)\right\} \tag{5.7}$$

where $\mathcal{F}$ is the Fourier transform along the x, y coordinates, and β_i captures the transform coefficients. Now considering two specific times t_2 and t_1 we get:

$$\begin{aligned}\mathcal{F}\left\{\tilde{m}(x,y,t_2)\right\} &= \beta_2 \exp\left\{-\sigma^2(t_2)\left(\omega_x^2+\omega_y^2\right)\right\}\mathcal{F}\left\{s(x,y)\right\} \\ &= \frac{\beta_2}{\beta_1}\exp\left\{-\left(\sigma^2(t_2)-\sigma^2(t_1)\right)\left(\omega_x^2+\omega_y^2\right)\right\}\mathcal{F}\left\{\tilde{m}(x,y,t_1)\right\}\end{aligned} \tag{5.8}$$

Taking the inverse Fourier transform results in:

$$\tilde{m}(x,y,t_2) = \tilde{\beta}\exp\left\{-\frac{x^2+y^2}{\sigma^2(t_2)-\sigma^2(t_1)}\right\} * \tilde{m}(x,y,t_1) \tag{5.9}$$

where $\tilde{\beta}$ is a normalization coefficient. Thus, up to a normalization factor, we have a simple transformation between selected time frames. To estimate $\sigma^2(t)$, we expand it with a Taylor expansion: $\sigma^2(t) = a_0 + a_1 t + a_2 t^2 + \cdots$. To estimate the different coefficients we note the following relationship between specific time bins:

$$\tilde{m}\,(x,y,t_2) = \tilde{\beta}\left[\exp\left\{-\frac{x^2+y^2}{a_1(t_2-t_1)+a_2(t_2-t_1)^2+\cdots}\right\}\right] * \tilde{m}\,(x,y,t_1) \tag{5.10}$$

Since we have many time frames (512 in our case), we can derive 511! different equations to estimate the unknown coefficients in Eq. 5.10. In practice we find that a linear model is sufficient such that:

$$W\,(x,y|t) = \exp\left\{-\frac{x^2+y^2}{a_0+a_1 t}\right\} \tag{5.11}$$

We estimate a_1 using a simple line search with a sample of the potential 511! equations. Lastly, to estimate a_0 we note that rewriting the variance as: $a_1\left(a_0/a_1+t\right)$ reveals the interpretation of $t_0=-a_0/a_1$. More specifically, t_0 is the ballistic time that can be estimated according to the first time bin where $\tilde{m}(x,y,t)$ is above the noise floor. Note in Fig. 5-2c that the very bottom (origin of time axis) is very noisy until it abruptly cleans up. We choose this time as t_0, and set $a_0=-t_0a_1$. This results in the estimated $W\left(x,y|t\right)$ as can be seen in Fig. 5-2d. Note that Eq. 5.11 resembles the diffusion equation, and if we choose $a_1=4D$ we get:

$$W\left(x,y|t\right)=\exp\left\{-\frac{x^2+y^2}{4D(t-t_0)}\right\} \tag{5.12}$$

where D is the diffusion coefficient.

Finally to complete the estimation of $K(x,y,t)$ we multiply the estimated $f_T(t)$ by the estimated $W\left(x,y|t\right)$. The result in shown in Fig. 5-2e. Note that the estimated kernel $K(x,y,t)$ in Fig. 5-2e resembles the measured point source in Fig. 5-2a, which shows that our model captures the key physical aspects of the system.

5.2.3 Inverse Problem Formulation

With the estimated kernel $K(x,y,t)$ solving Eq. 5.1 is reduced to a much simpler deconvolution problem. This problem can be cast as the following optimization problem:

$$\hat{s}=\arg\min_{s}\left\{\|\mathbf{A}s-\tilde{m}\|_2^2+\lambda R\left(s\right)\right\} \tag{5.13}$$

Here, $\mathbf{A}$ is a matrix form of the kernel $W(x,y|t)$. Each column in $\mathbf{A}$ is the vectorized predicted measurement of a specific point source location. The predicted measurements are calculated by computing $W(x,y|t)*s_{i,j}(x,y)$ where $s_{i,j}(x,y)$ is a point source target at position (i,j). These computed vectors are lexicography ordered and stacked in $\mathbf{A}$. Since $\mathbf{A}$ is a blurring operator in space and time it is not invertible. Thus, we regularize the problem with $R(s)$ that is defined by prior statistical knowledge of the occluded scene. λ is the regularization strength.

In our experiments, our targets were composed of simple lines, thus we used the ℓ_1 norm: $R(s) = \|s\|_1 = \sum_i |\hat{s}_i|$. To solve Eq. 5.13 we used FISTA [13], which was initialized by the noisy ballistic photon estimate $\tilde{m}(x, y, t_0)$.

5.2.4 Algorithm Implementation Details

The measurement $m(x, y, t)$ is of size $305 \times 305 \times 512$ where each entry corresponds to $0.3\,\text{mm} \times 0.3\,\text{mm} \times 2\,\text{ps}$. The recovered scenes have a resolution of 70×70 pixels, where each pixel corresponds to 1 mm. We run FISTA with 10^4 iterations, and the regularization parameter is set to $\lambda = 0.004$ for all scenes.

Computational Issues

The specified resolutions above result in the size of $\mathbf{A}$ equal to $47{,}628{,}800 \times 4900$ (over 10^{11} elements) which is too large to save in memory. To overcome this challenge we note that the FISTA algorithm (similar to other first order methods) requires only $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\tilde{m}$. $\mathbf{A}^T\mathbf{A}$ has a modest size of 4900×4900 and $\mathbf{A}^T\tilde{m}$ is a vector of length 4900. These objects are calculated before FISTA starts.

To calculate $\mathbf{A}^T\mathbf{A}$ without explicitly storing $\mathbf{A}$ in memory, we recall that each column in $\mathbf{A}$ is the vectorized $W(x - x_i, y - y_i|t)$, which we denote by v_i. Next we use the fact that $\mathbf{A}^T\mathbf{A} = \sum_n \left(\bar{A}_n\right)^T \left(\bar{A}_n\right)$. Here $\left(\bar{A}_n\right)$ are parts of the full matrix $\mathbf{A}$. More specifically, we split $\mathbf{A}$ to smaller blocks, compute the Gram matrix of each block, and sum-up the results. Each of the blocks has all columns, and a user defined number of rows. The choice of the number of sub blocks is defined by the available memory.

To compute $\mathbf{A}^T\tilde{m}$, we note that each element in the resulting vector is the dot product between v_i and $\tilde{m}$. Thus, we sequentially generate the v_i-s, and compute the dot product, then store the result in the vector to be used by FISTA.

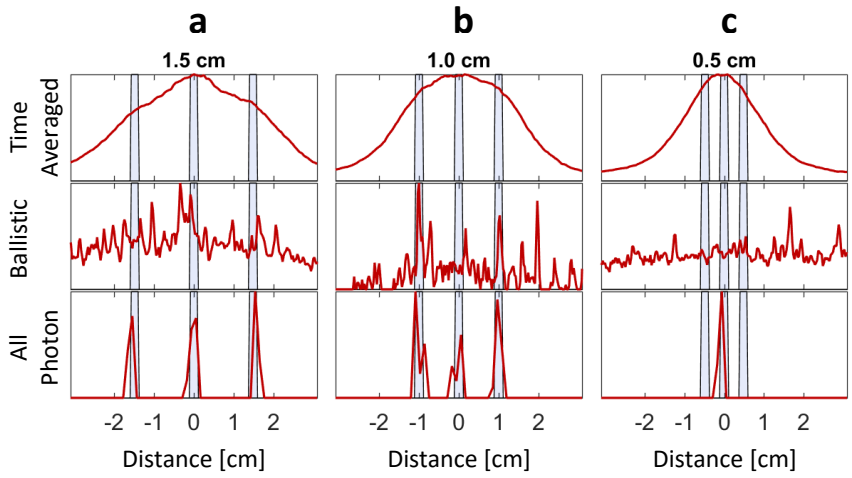

Figure 5-3: Recovery of 1D slits target with API. Three slits separated by 1.5, 1.0 and 0.5 cm (a, b, c respectively) and their recovery with Time Averaging and Ballistic photons compared to API. Blue shadings are the slits' location ground truth. API shows significant advantages in recovering the three slits when they are separated by up to 0.5cm, while the other methods fail in all cases.

5.3 Experimental Results

API is experimentally evaluated with a set of 1D and 2D targets. We compare API to:

- Time averaging: integrates over the time axis and does not leverage temporal information.
- Ballistic: selects the first time bin with signal above the noise floor.

As a 1D target we use a set of three targets composed of slits. The slits are separated by 1.5 cm, 1.0 cm, and 0.5 cm. The results are shown in Fig. 5-3. As predicted, the time averaged result is very blurry and does not reveal the locations of the slits. The ballistic photon measurement is very noisy and the exact location of the slits is embedded in the noise. On the other hand, API clearly recovers the exact locations of the slits for the case of 1.5 cm and 1.0 cm separation. As we show next, our system's resolution limit is 0.59 cm which explains the failure of API in the 0.5 cm case.

To demonstrate the recovery of 2D targets with API, we place masks shaped

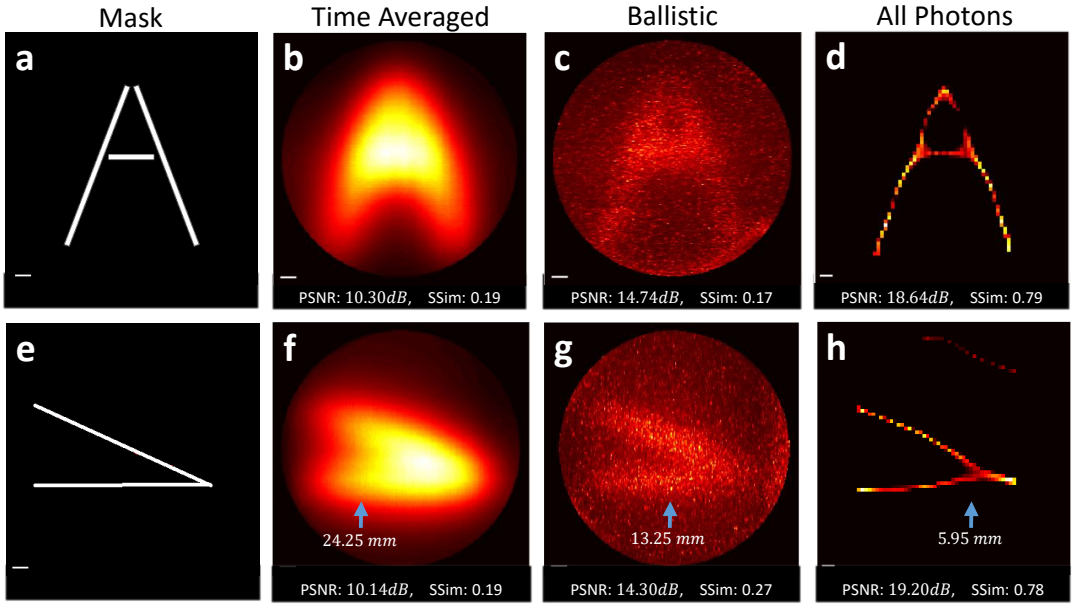

Figure 5-4: API recovers 2D scenes. a) 'A' shaped hidden mask. b) Recovered scene without using time-resolved data; the result is very blurry. c) Recovered scene using only ballistic photons; the signal is embedded in the noise level. d) Recovered scene using API; the result clearly recovers the hidden scene. e-h) Mask and results for a wedge-shaped scene. Blue arrows mark the points used to evaluate the best recoverable resolution and the corresponding resolution. All reconstructions are quantitatively evaluated with both PSNR and SSIM (ranges in $[0, 1]$, higher is better). Scale bar equals 5 mm.

like the letter 'A' and a wedge (Fig. 5-4). These results demonstrate again that the Time Averaging result is blurry and the information content of the scene is gone. Furthermore, the Ballistic photon measurement captures some of the information, but the signal is comparable to the measurement noise. API, on the other hand, is clearly able to capture the information content of the scene. It is interesting to note that while Ballistic photons use a single time gate of 2 ps, we still note substantial blur in the measurement. This demonstrates that pure hardware-based solutions will capture some of the scattered light, and should be augmented by computational techniques. We also note that deblurring the Ballistic photons measurement is very challenging due to the high noise level.

Beyond the qualitative results in Fig. 5-4, we also provide quantitative results:

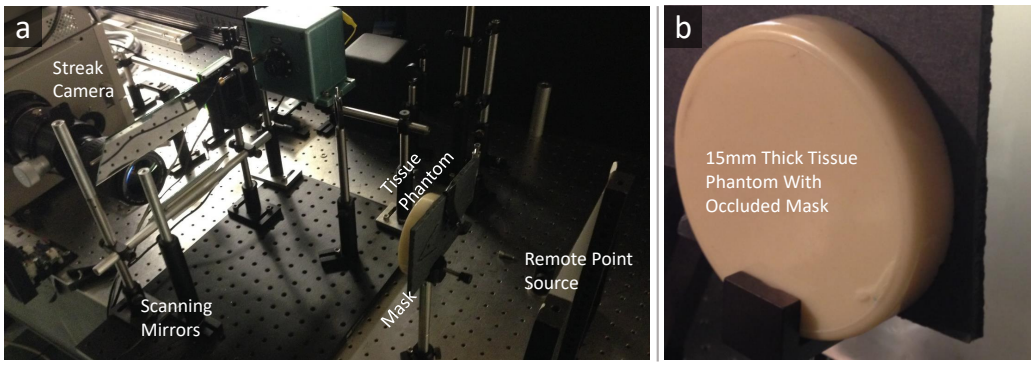

Figure 5-5: API optical setup. a) Photo of the API experimental setup. b) Photo of the tissue phantom with the occluded mask at the back.

Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index (SSIM) [186]. PSNR performs a pixel-wise comparison of the reconstruction to the ground truth (higher is better). SSIM takes into account the structure of the image and compares spatial information between the reconstruction and the ground truth. SSIM ranges in $[0, 1]$, higher is better. API outperforms the other methods in both metrics.

The wedge shape allows us to evaluate the resolution limits of the different techniques. We estimate the recoverable resolution limit, that is the ability to separate between two point sources. The recoverable resolution limit is estimated by the distance between the wedge lines at the point in which they merge. The best resolution is indicated by blue arrows in Fig. 5-4, and the corresponding resolution in mm is overlaid on the reconstructions. API demonstrates $2\times$ better resolution than the Ballistic photons result, and $4\times$ better than the Time Averaged result.

5.3.1 Experimental Implementation Details

Figure 5-5 shows a photo of the optical setup and the tissue phantom. A Ti:Sapph (795 nm, 0.4 W, 50 fs pulse duration, and 80 MHz repetition rate) is focused onto a polycarbonate thin diffuser (Edmund Optics, 55-444) 40 cm away from the sample to produce a remote point source. The point source results in an approximate pulsed plane wave that illuminates the Intralipid tissue phantom (reduced scattering coefficient of $100\,\mathrm{cm}^{-1}$). The sensor is a streak camera (Hamamatsu C5680) with a time resolution of 2 ps and a time window of 1 ns. The sensor has a 1D aperture

and records the time profile of a horizontal slice of the scene ($x-t$). A set of two motorized mirrors is scanning the y axis of the scene in a periscope configuration to measure the $305 \times 305 \times 512$ matrix. The exposure time of each $x-t$ slice is 100 ms (total acquisition time is 30 s per scene). The problem of measuring the full $x-y-t$ cube in a single shot has already been solved by e.g. time-space multiplexing [69, 118] and compressive techniques [51].

5.4 Sensitivity and Dynamics

5.4.1 Recoverable Resolution

To evaluate the suggested technique we leverage a Monte Carlo simulation. Here, our goal is to find the effect of sensor time resolution on the recoverable spatial resolution as a function of the tissue thickness.

The Monte Carlo simulation with 10^9 photons is performed on a scattering medium with a scattering coefficient of $200\,\mathrm{cm}^{-1}$, and an HG anisotropy coefficient of 0.85. The thickness of the medium is varied in the range of $10-50$ mm. The time resolution of the detector is varied in the range of $2-250$ ps. For each configuration, we perform the simulation with a target that is composed of two point sources separated by a distance d. The resulted measurement is then used as an input to the full procedure defined in Sec. 5.2 to recover the two point source targets. This process is repeated, each time with a smaller d until the recovered target shows a single point source. The last d in which two point sources were recovered defines the recoverable spatial resolution for the simulated medium thickness and time resolution.

The results of this analysis are shown in Fig. 5-6. As predicted, better temporal resolution of the sensor results in better recoverable resolution. To further demonstrate this trend, we plot several cross sections of different sensor temporal resolutions (Fig. 5-6b), and for several cross sections of different diffuser thicknesses (Fig. 5-6c). For time resolutions below 50 ps, we gain exponentially better recoverable resolution for increasing diffuser thickness. This is especially true for the range $12-30$ mm.

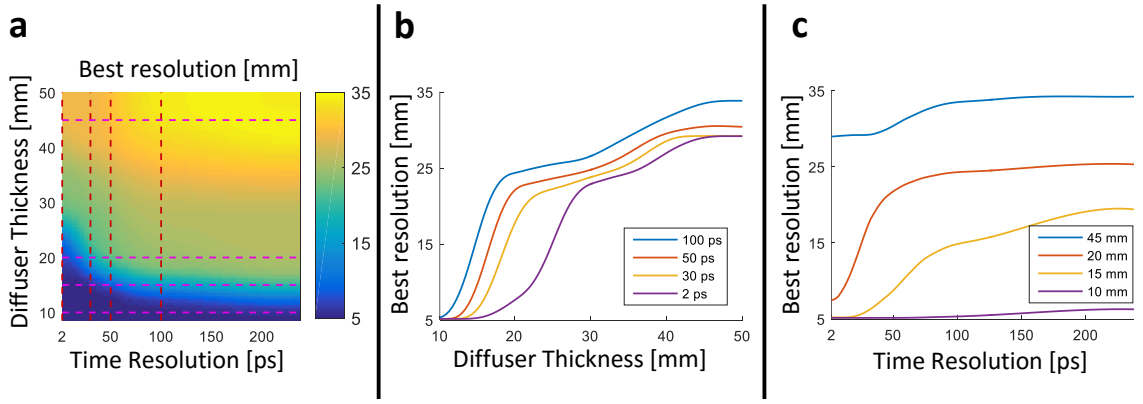

Figure 5-6: API recoverable resolution as a function of time resolution and medium thickness. a) Monte Carlo simulation results for varying sensor time resolution and diffuser thickness; colors represent the best recoverable resolution in mm. b) Vertical cross sections (for different time resolutions). c) Horizontal cross sections (for different diffuser thickness).

This demonstrates the benefit of ultrafast time-resolved sensing. The improved time resolution inputs the reconstruction framework with a more diverse set of measurements, which increases the robustness of the inversion process and allows better recoverable resolution. Better time resolution helps in two ways: 1) As the medium gets thicker, $W(x, y|t)$ becomes broader in $x-y$ and t. With better time resolution, the kernel estimation process is more robust. 2) More time bins provide more instances of the scattering kernel $W(x, y|t)$, which improves the deconvolution robustness.

5.4.2 Noise Sensitivity

One of the advantages of API is its robustness to measurement noise. This robustness is a result of the time-resolved measurement. Since each time frame captures the same target corrupted by a different scattering kernel, we effectively measure the target multiple times which reduces the sensitivity to measurement noise. To quantitatively demonstrate this we run Monte Carlo simulation with 10^9 photons through a 15 mm thick scattering medium with a scattering coefficient of $200\,\mathrm{cm}^{-1}$, and an HG anisotropy coefficient of 0.85. Similarly to the previous subsection, the target is composed of two pin holes separated by a distance d. The measurement is added with white Gaussian noise to simulate measurement PSNR in the range of $20-90$ dB. For

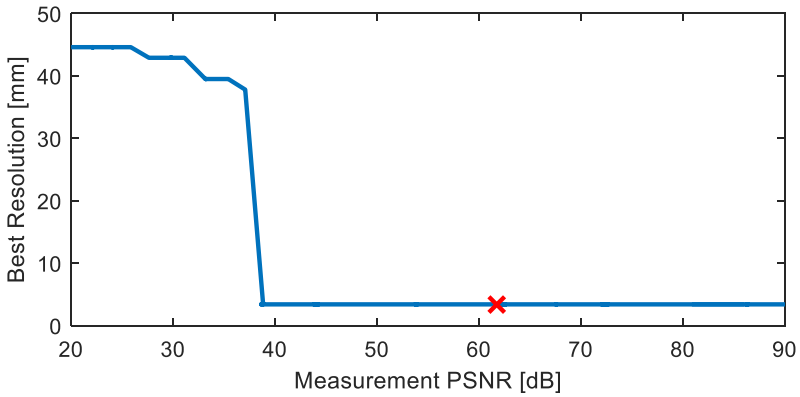

Figure 5-7: API sensitivity to measurement noise. Monte Carlo simulation results in different levels of measurement noise and its effect on recoverable resolution. API performs very robustly for PSNR above 39 dB. The experimental measurement PSNR is noted in red cross (61.7 dB).

each PSNR point, we run multiple simulations with decreasing d, and the simulated measurement is processed by the API algorithm to estimate the hidden scene. The recoverable resolution is defined as the smallest d for which the reconstruction result produces two point sources.

Figure 5-7 shows the results of this analysis, and demonstrates that API performs very robustly for measurement PSNR above 39 dB. For lower PSNR the first step of API (estimation of the scattering kernel) is less stable, which results in the rapid degradation of recoverable resolution.

To compare these simulation results to the experimental measurements, we used the method suggested by Xinhao et al. [104] to estimate the PSNR in our experimental measurements, and found it to be 61.7 dB (indicated by red 'X' in Fig. 5-7 which is well above the 39 dB failure point). This indicates that a shorter exposure time would result in comparable reconstruction quality.

Figure 5-8 demonstrates API's ability to recover complex targets in the presence of high measurement noise. The results are based on the above Monte Carlo simulation with added white Gaussian noise that results in measurement PSNR in the range of $42.5 - 44.2$ dB for the different targets. Similarly to the experimental results, we compare API to Time Averaging and Ballistic photons recovery. As predicted, the time averaged version doesn't suffer from the noise since it averages it over the

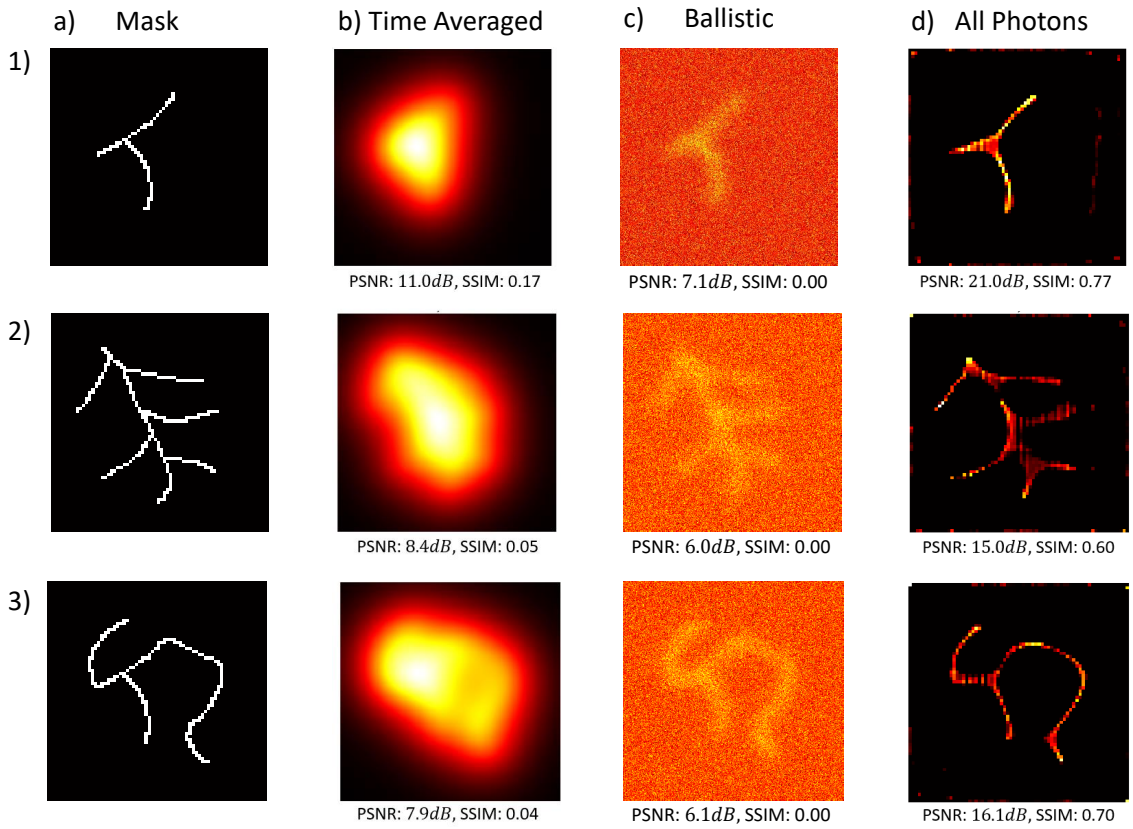

Figure 5-8: API successfully recovers complex targets with noisy measurements. a) The target mask. b) Time Averaged result. c) Ballistic photon measurement. d) API recovery. The three rows correspond to three different targets. The measurements PSNR in targets 1-3 are 42.5, 44.5, and 43.9 dB respectively. All reconstructions are evaluated with PSNR and SSIM.

different frames. The ballistic photons recovery suffers substantially from the noise indicated by the very poor PSNR and SSIM. API, on the other hand, performs similarly to the high PSNR experimental measurement and is able to recover the complicated structures.

5.5 Imaging Through Layered Materials

The main assumption in Eq. 5.1 is a homogeneous medium in the $x-y$ coordinates (perpendicular to the optical axis). This assumption allowed us to write the forward model as a convolution along the $x-y$ coordinates. However, we note that the z coordinate (the optical axis) does not appear in the model, and not in the scattering

kernel. This effectively means that this imaging geometry (and the model) is invariant to variations along the optical axis. Such variations include changes in the thickness of the medium, and changes in the optical properties of the material along the optical axis – for example a material composed of different layers. This is especially important since skin tissue is composed of different layers.

To demonstrate this invariance, we consider layered material composed of N layers. The thickness of the n-th layer is l_n, and it has a scattering coefficient of μ_{T_n}, such that $\sum_{n=1}^{N} l_n = L$, where L is the total thickness of the medium. Since our system is invariant to the composition of the different layers we consider the equivalent uniform material defined by a thickness L and a scattering coefficient:

$$\mu_{T_U} = \sum_{n=1}^{N} \frac{l_n}{L} \mu_{T_n} \tag{5.14}$$

That is, the equivalent uniform material's scattering coefficient is the mean of the different layers' scattering coefficient weighted by their thickness. Notably the order of the layers doesn't change the properties of the equivalent uniform material.

To demonstrate the validity of this model, we perform Monte Carlo simulations with 10^7 photons. We consider two scattering media and their equivalent uniform medium (total of four different materials):

1. Layered material composed of 15 layers with identical thickness $l_n = 0.2\,\text{cm}$, such that $L = 3\,\text{cm}$. μ_{T_n} was sampled from a uniform distribution in the range $[100, 300]\text{cm}^{-1}$. The scattering coefficient depth profile for the layered material and its equivalent uniform material are shown in Fig. 5-9a top.

2. Layered material composed of 8 layers with varying thicknesses such that the total thickness $L = 2\,\text{cm}$. μ_{T_n} was sampled from a uniform distribution in the range $[100, 300]\text{cm}^{-1}$. The scattering coefficient depth profile for the layered material and its equivalent uniform are shown in Fig. 5-9a bottom.

Figure 5-9 shows the result of the four Monte Carlo simulations for a point source target (columns b,d). We note that the PSFs in columns b and d are indeed nearly

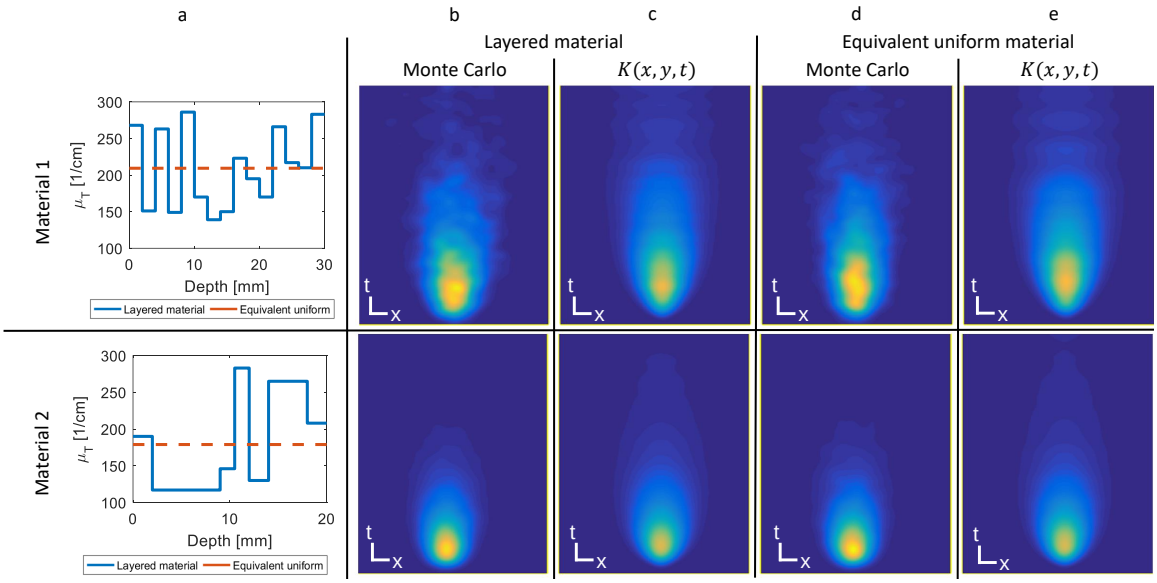

Figure 5-9: Layered materials and their equivalent uniform have the same PSF. The two rows show examples of different materials. a) Depth cross section for the scattering coefficient, and the equivalent uniform (dashed line). b) Monte Carlo simulation – PSF of the layered material. c) Estimated scattering kernel by API for the layered material. d) Monte Carlo simulation – PSF of the equivalent uniform material. e) Estimated scattering kernel by API for the equivalent uniform material. Panels b-e show an $x-t$ cross section for $y=0$. Columns b and e are roughly identical (up to sampling noise), demonstrating that the layered material and its equivalent uniform share the same PSF, and that API captures the PSF of these four different materials.

identical (the variations are due to sampling noise caused by the Monte Carlo simulation). As expected, the PSFs of the two rows (materials of different thicknesses and scattering coefficients) are noticeably different. For completeness, Fig. 5-9 shows the estimated PSF by API (columns c,e). As predicted, API is able to capture the PSF shape of the different materials.

So far, we have shown that a layered material has an equivalent uniform material, and both result in the same PSF. In the next step, we simulate 2D masks and use the Monte Carlo simulation to compute the time-resolved measurement through these four different materials. The API algorithm is then used to reconstruct the occluded mask. Fig. 5-10 shows the results for two masks. As can be appreciated from the figure, all reconstructions look roughly the same. This demonstrates that API is invariant to variations of thickness and scattering coefficient along the optical axis.

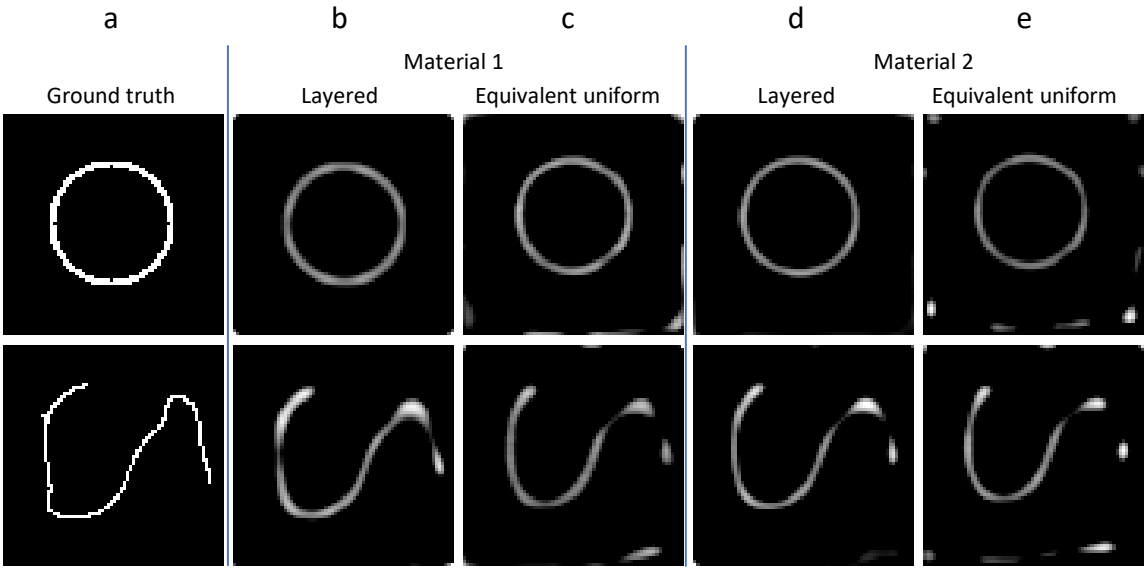

Figure 5-10: API is invariant to variations along the optical axis. a) Ground truth. b-e) Reconstruction results for the four materials defined in Fig. 5-9.

5.6 Discussion and Future Work

The approach presented in this chapter introduces imaging through volumetric scattering in optical transmission mode. In transmission mode, all the optical signal that arrives at the sensor interacts with the target during its propagation. API leverages this observation and captures all of the optical signal to computationally invert the scattering process, which results in improved results at shorter acquisition times compared to pure hardware-based solutions (time-gating). The inversion of the scattering processes is based on a probabilistic interpretation, that decouples the scattering kernel to a time-only function and a simple space-time function. These two functions are easy to estimate from the measurement itself, without prior knowledge or assumptions of the occluded scene. This interpretation also leads us to the observation that the time-resolved scattering can be viewed as a time-dependent blur-kernel, where each time bin in the measurement is a result of convolving the stationary target with a different blur kernel.

API is based on time-resolved measurements and demonstrated with a streak camera. Interestingly, a streak camera captures a time-resolved measurement within a single swipe – that is, it captures only the ballistic photons, and all photons at the

same cost. Thus, it is even more appealing to use API when a streak camera is the time-resolved measurement device. This is also similar to measurement with a SPAD array operating in photon-starved mode. In this regime, the detector has roughly the same probability to measure ballistic photons and scattered photons which, again, makes API an appealing solution when using a SPAD detector.

The main limitations of the presented approach include:

- **Transmission mode** – while this is the key observation that allows us to use all of the optical signal, it is also a key limitation. Certain applications cannot be performed in transmission mode (for example, remote sensing and depth sensing in degraded weather). In the context of medical imaging, certain applications such as mammography are usually performed in transmission mode. But even for those applications the current approach is limited since there is no tissue between the illumination source and the target. A potential solution to this limitation is to create a synchronized pulsed source inside the medium; this can be achieved with nonlinear effects, e.g. two-photon [85] or localized plasma discharges [52] which can be used for atmospheric studies [128].

- **Homogeneous assumption** – the suggested approach assumes the scattering material is homogeneous in the $x - y$ plane. This can potentially be relaxed by breaking the model to piecewise smooth areas, but would still be limited. Furthermore, the ability to image through layered materials assumes there is no inter-reflection between the layers (i.e. the entire medium has the same index of refraction). This may be a strong assumption in certain applications.

In summary, we presented API, a computational imaging technique to see through volumetric scattering. The approach is based on ultrafast time-resolved sensing (demonstrated with a streak camera), in optical transmission mode. API was demonstrated in resolving 5.9 mm features through a 1.5 cm thick tissue phantom. The sensitivity of API to measurement noise, and to the detector time resolution were analyzed. Lastly, the invariance of API to variations along the optical axis was demonstrated. API has several key advantages: it is wide-field, does not require raster scan,

and does not require any prior knowledge of the scattering media (calibration-free).

Chapter 6 addresses API's main limitations with a technique capable of imaging through extremely dense and heterogeneous fog in optical reflection mode.

Chapter 6

Imaging Through Extremely Dense Fog

The main challenges with the approach presented in Chapter 5 are the strict assumptions of optical transmission mode, and a convolution model that assumes layered scattering media. Here, we alleviate these challenges by developing a probabilistic computational imaging technique to overcome scattered light. The presented approach recovers the target reflectance and depth map, works in optical reflection mode, and does not require calibration or prior knowledge about the fog properties. We experimentally demonstrate it with a wide range of realistic fog conditions: dense (clear visibility to 30 cm visibility), dynamic, and heterogeneous. For example, we demonstrate recovering objects 57 cm away from the camera when the visibility is 37 cm. In that case the depth is recovered with a resolution of 1 cm and the target reflectance is recovered with an improvement of 4 dB in PSNR and 3.4× reconstruction quality in SSIM compared to time gating techniques.

Imaging in optical reflection mode is essential for long-range sensing. It is also fundamentally harder than optical transmission mode. In transmission mode, all of the measured photons interact with the target, while in reflection mode, most of the measured photons simply back-reflect from the fog before even reaching the target. Thus, in reflection mode we must separate between optical signal due to background and signal due to the target.

We show that time profiles of photons scattered by fog have a distribution (Gamma) that is different from photons reflected from objects occluded by fog (Normal). With this observation, we develop a computational imaging algorithm that separates between photons back-reflected from the fog and photons reflected from the occluded object, without prior knowledge about the fog properties. The imaging system is designed in optical reflection mode with a minimal footprint and is based on LIDAR hardware. Specifically, we use a single photon avalanche diode (SPAD) camera that time tags individual detected photons.

Dense fog is a significant limitation in many transportation systems such as self-driving cars, augmented driving, airplanes, helicopters, drones, and trains. The ability to see through fog may augment a driver with a heads-up display, showing obstructed objects on the road in front of the vehicle, or read a road sign that is not visible. Similar use cases are essential for autonomous vehicles where the future goal is to allow a car to drive in any weather. Other applications include identifying a clear flight path for drones, helicopters, and airplanes, and allowing trains to maintain speed in foggy weather.

The main industry solution for imaging through fog is based on radio waves, e.g. in 94 GHz [7] where fog is transparent. There are several challenges in using radar for imaging, including: 1) resolution – due to the long wavelength it is hard or even impossible to classify objects, and the use cases are limited to detection; and 2) optical contrast – at such long wavelengths it is only possible to measure bulk properties of materials [3], and impossible for example to identify road marks and read road signs. Techniques for long range and large field of view imaging through volumetric scattering in the visible range are usually limited to time gating, which requires long integration times and prior knowledge of the scene depth map. Furthermore, time gating and other techniques to image through scattering media, such as phase conjugation and acousto-optics, reject scattered photons during the measurement process. Because these methods reject a substantial amount of the optical signal, they operate at a low signal-to-noise ratio (SNR).

Instead of rejecting the scattered photons during the measurement process, we

measure all of the optical signal (both scattered and unscattered photons) and computationally use it to reject the fog. The motivation for our approach is the fact that the scattered photons hold substantial information about the target and the scattering medium, and that information is useful for improving imaging capabilities. Our suggested method uses the scattered photons to estimate the fog properties and computationally reject it from the measurement. Thus, it works at higher measurement SNR and does not depend on prior knowledge of the scene or scatterer. We experimentally demonstrate our recovery technique in a fog chamber with realistic fog conditions (dense, dynamic, and heterogeneous). In our demonstration, the camera and illumination are adjacent to the fog chamber to further emulate realistic imaging scenarios.

The main technical contributions of the approach presented in this chapter are:

1. A time-domain technique for seeing through fog, demonstrating the ability to recover the scene image and depth.

2. An experimental demonstration of the technique in dense, dynamic, and heterogeneous fog conditions in a fog chamber (as opposed to milky water or other phantoms) for a wide range of fog densities (visibilities) and ten different static scenes.

3. An experimental demonstration of a seamless integration between the suggested technique to an off-the-shelf computer vision algorithm (OCR) that effectively enables the OCR to read text occluded by the fog, without any modification to the OCR.

4. A probabilistic physical model that describes time domain statistics of photons reflected by fog, and photons reflected by an occluded target.

5. An algorithm that estimates pixel-wise fog parameters from the measurement itself without any calibration or prior knowledge, as well as an expectation maximization derivation to refine the parameters.

6.1 Related Works

Many works in computer vision tackle the problem of vision through the atmosphere in a regime that assumes a single scatter event [165]. In that case, the degradation model is an additive background from the fog or haze. Some techniques to overcome such atmospheric scattering include polarization [153, 176], patch base recurrence [10], haze line estimation [15], and lightfield imaging [174]. Data-driven techniques for dehazing have been explored [23], along with rendering techniques [103, 156]. While most work is dedicated to reflectance recovery, some recover depth of objects through scattering media [67, 62, 78, 174].

Works that explicitly handle the case of highly scattering materials similar to the one discussed here include depth sensing [67, 62, 80], cloud tomography [100, 71, 101], and scattering parameters estimation for computer graphics [55, 56]. Our goal is different since we aim to perform both reflectance and depth recovery of a scene occluded by highly scattering media.

Overcoming fog in LIDAR is known as Laser Imaging Through Obscurants (LITO) [155], where the primary solution to overcome scattering is time gating [111, 136, 39, 97]. These methods work when the object is far away (less coupling with back reflectance from the fog), but are limited by the signal-to-noise ratio, which requires long integration times and a stationary scene. Another limitation of time gating is the need to manually select the time gate bin or to rely on prior depth map knowledge.

In this work, we use single photon counts along with a probabilistic model to reject the fog. Other examples of using probabilistic models for imaging with a few photons [89, 162] did not consider imaging through scattering media.

In the closest work to ours, Srinivasa *et al.* [119] used structured light to spatially decouple the back reflectance and signal. Here we achieve this with a per-pixel time profile, which allows imaging through more challenging scattering conditions, and results in better quality.

6.2 Time-Resolved Statistics in Fog

Consider a pulsed light source emitting photons into a foggy scene. Adjacent to the light source is a time-resolved camera. For each measurement frame, each pixel holds the arrival time of the first detected photon during the frame acquisition time. A measured photon can be classified as one of the following:

- **Background photon** – a photon that did not interact with the target, thus it only holds information about the fog. Due to the scattering dynamics, background photons arrive at different times.
- **Signal photon** – a photon that interacted with the target, thus it holds information about the target reflectivity and depth.
- **Dark counts (noise)** – these false detections are uniformly distributed in time.

The dark count rate in our detector is below 30 Hz which is an order of magnitude less than the background and signal counts in our measurements, thus we neglect it from our model.

Next, we derive a probabilistic model that describes the statistics of these photon classes. This model is pixel-wise, which is essential in handling heterogeneous scattering media such as fog. We leverage adjacent pixel statistics to refine our estimates, as described in section 6.3.2.

Since our detector is single photon sensitive, our measurement per pixel is a list of photons' times of arrival. For each photon, we may ask what is the probability that it is a background photon or a signal photon. This information is encoded in the photon's time of arrival. But we start with a different, simpler, question: what is the probability density function $f_T(t)$ for photons time of arrival? As we'll show next, background and signal photons have different statistics in time. We combine these into a single model with the law of total probability:

$$f_T(t) = f_T(t|S)P_S + f_T(t|B)P_B \tag{6.1}$$

here, P_S, P_B are the priors to measure signal and background photons respectively. $f_T(t|S)$, $f_T(t|B)$ are the probability density functions to measure a photon at time t, given that it is a signal or background photon (respectively). The ratio between P_B and P_S captures the probability of measuring a background vs. signal photon. Next, we derive empirical models for $f_T(t|S)$, $f_T(t|B)$.

6.2.1 Background Statistics

Based on scattering theory, we know that the distance a photon propagates between consecutive scattering events is exponentially distributed with a mean of $1/\mu_s$ (see Sec. 2.3). Equivalently the time between scattering events is also exponentially distributed with a mean of $c\mu_s$ (c is the speed of light). In this chapter, for simpler notation, we set $c = 1$. In that case, the time between scattering events $k-1$ and k denoted by τ_k has the following probability density function:

$$f_{\tau_k}(t) = \mu_s e^{-\mu_s t} \tag{6.2}$$

Since the scattering events are independent, so are the different times τ_k. A detected photon undergoes multiple scattering events such that the detection time is $T = \sum_{k=1}^{K} \tau_k$. The sum of independent exponential random variables is Gamma distributed $T \sim GAMMA(K, \mu_s)$, where K and μ_s are the shape and rate parameters. Thus, we can model the probability density of measuring a background photon at time t, denoted as $f_T(t|B)$, with the parameters K and μ_s encoding the physical properties of the fog:

$$f_T(t|B) = \frac{\mu_s^K}{\Gamma(K)} t^{(K-1)} \exp\{-\mu_s t\} \tag{6.3}$$

where $\Gamma(K)$ is the Gamma function. Fig. 6-1 shows time-resolved measurements at different concentrations of fog and the corresponding Gamma distribution fits. We measure fog densities with optical thickness (OT) where OT $= 0$ is clear visibility. As can be seen in the figure, the Gamma distribution matches the raw measurements

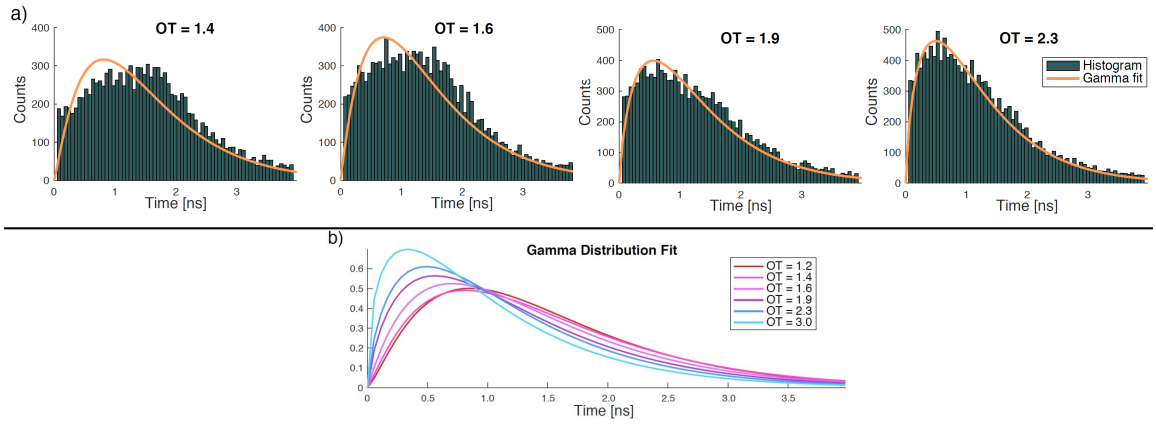

Figure 6-1: Fog background model. a) Experimental time-resolved measured histograms along with fitted Gamma distributions. The panels correspond to different optical thicknesses (OT) of fog. The plots show that a Gamma distribution captures well the dynamics of time-resolved scattering in fog, especially at high densities. b) Fitted Gamma distributions for a wide range of fog densities. The plots show that different fog densities (optical thicknesses) result in different time profiles.

well, especially for higher levels of fog. As we show next, our complete pipeline naturally overcomes this potential model mismatch at lower levels of fog.

6.2.2 Signal Statistics

Next, we model the time of arrival of photons that interacted with the target as a Gamma-distributed random variable, given that the photon interacted with the target (with similar arguments to the background model). In practice, we find that we can use a Normal model for this distribution. This can be justified since in this case the number of scattering events is large, and when the shape parameter, K, of a Gamma distribution is large, it resembles a Normal distribution. Another, more practical reason is that our detector time resolution along with jitter obscures more complicated time dynamics.

The Normal distribution mean, μ, corresponds to the depth of the object. The variance, σ^2, encodes the time dynamics these photons undergo. Empirically (as shown below), the majority of the contribution to σ^2 is due to the system time-jitter. The probability density function of measuring a signal photon at time t is:

$$f_T(t|S) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} \tag{6.4}$$

6.3 Imaging Algorithm

6.3.1 Pixel-Wise Model Estimation

Equation 6.1 serves as our physical model. It is important to note that our input per pixel is a list of photons' arrival times within a fixed exposure window. In a SPAD detector, for each emitted pulse we may (or not) detect a photon per pixel. The arrival time is the time between pulse emission and photon detection. We use multiple arrival times per pixel to estimate the model (more implementation details are provided in Sec. 6.4).

Next, we describe our approach to estimate the five terms in Eq. 6.1 from raw measurements of photons' times of arrival. The estimation pipeline is composed of four steps: 1) complete time profile estimation, 2) background distribution estimation, 3) signal distribution estimation, and 4) priors estimation. Fig. 6-2 shows the estimation results for different levels of fog and for targets at different depths.

Estimating the complete time profile $f_T(t)$

The individual photon detection events are mapped to the distribution $f_T(t)$ using a kernel density estimator (KDE) and denoted by $\widehat{f_T}(t)$. KDE has two main advantages over a traditional histogram: 1) it performs well with a few sampling points, and 2) there is no need to specify a bin size. While there are techniques for automatic selection of the KDE bandwidth, here we select 80 ps. This value matches the full-width-half-max of our detector time response curve.

Estimating the background time dynamics $f_T(t|B)$

The physical model describing the background is a Gamma distribution. The distribution parameters are estimated using maximum likelihood. The estimated distribution

is denoted by $\widehat{f_T}(t|B)$.

Since the majority of the photons are background photons, we use all measured photons in this step and effectively treat the signal photons as noise. When this assumption is not valid (less fog), the remainder of the pipeline accounts for the errors introduced by this step.

The maximum likelihood estimator used here is based on the algorithm in [114]. Specifically, we use an iterative algorithm to estimate K, with the following update rule:

$$\frac{1}{K^{new}} = \frac{1}{K} + \frac{1}{K^2} \frac{\frac{1}{m}\sum_{i=1}^{m} \log x^i - \log\left(\frac{1}{m}\sum_{i=1}^{m} x^i\right) - \Psi(K) + \log(K)}{\frac{1}{K} - \Psi'(K)} \tag{6.5}$$

where m is the number of detected photons per pixel, and x^i is the i-th photon arrival time. We iterate over Eq. 6.5 five times, and initialize the iterations with:

$$K^0 = \frac{0.5}{\frac{1}{m}\sum_{i=1}^{m} \log x^i - \log\left(\frac{1}{m}\sum_{i=1}^{m} x^i\right)} \tag{6.6}$$

We then use the estimated K to estimate μ_s according to:

$$\hat{\mu}_s = \frac{1}{\hat{K}} \frac{1}{m} \sum_{i=1}^{m} x^i \tag{6.7}$$

Estimating the signal time dynamics $f_T(t|S)$

With the probability functions for the complete time profile and background, it is possible to subtract the two curves and isolate a proxy to the probability density function of the signal $f_T(t|S) \approx f_T(t) - f_T(t|B)$. To that end we fit $\widehat{f_T}(t) - \widehat{f_T}(t|B)$ to a Gaussian curve that estimates the signal $\widehat{f_T}(t|S)$. This assumes again $P_B \approx 1$.

Negative values in the subtraction above are set to zero. In this step we effectively stop thinking about these functions as probability densities. Our goal here is to develop a robust estimator. To account for this mathematical inaccuracy, we also suggest a refinement with an expectation maximization algorithm that is mathematically accurate (Sec. 6.7).

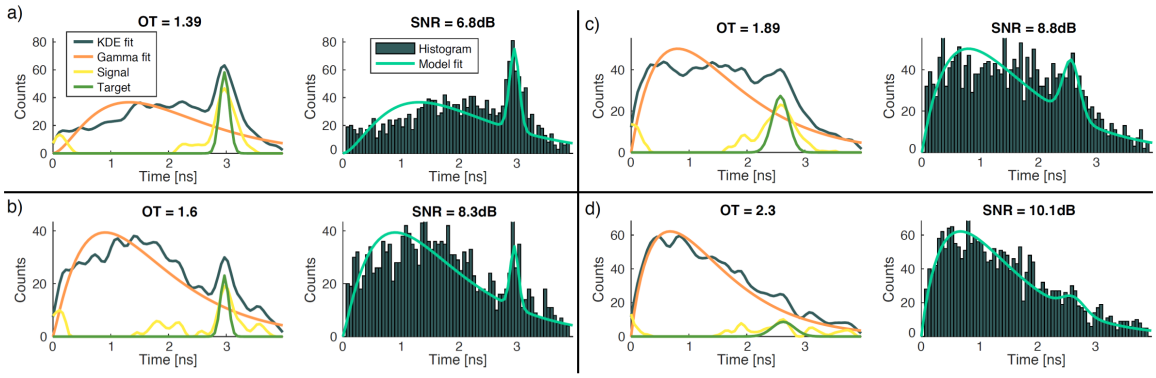

Figure 6-2: Rejecting back reflectance and signal recovery. Demonstrated on four different levels of fog: optical thicknesses of OT $= 1.39, 1.6, 1.89, 2.3$ for panels a-d respectively. In each panel, the left plot shows the recovered KDE, Gamma distribution, estimated signal, and estimated target distributions. The right plot shows the histogram generated by the raw photon counts and the fitted model (Eq. 6.9) including the SNR between the two. The target in panels a+b is at a depth that corresponds to 3.02 ns, and the target in panels c+d is at 2.58 ns. Note that in all cases there are substantially more background than signal photons.

Estimating the priors P_S, P_B

With the background and signal estimated distributions, the parameters P_S and P_B are estimated by solving:

$$\left[\hat{P}_S, \hat{P}_B\right] = \underset{0\le P_S, P_B \le 1}{\arg\min} \sum_t \left(P_S \widehat{f_T}(t|S) + P_B \widehat{f_T}(t|B) - \widehat{f_T}(t) \right)^2 \tag{6.8}$$

This is where edge cases in which there is no fog or no target are accounted for – the solution will be $\hat{P}_B \to 0$ or $\hat{P}_S \to 0$ accordingly. Note that the solution to Eq. 6.8 is a simple least squares.

Back to Photon Counts

So far, all estimators were probability density functions. These are mapped to actual photon counts $N(t)$, the number of photons measured at time bin t, by computing a normalization factor $\hat{\alpha}$ such that $\sum_t N(t) = \hat{\alpha} \sum_t \left(\widehat{P_S} \widehat{f_T}(t|S) + \widehat{P_B} \widehat{f_T}(t|B) \right)$. This step is necessary for consistent results across pixels that receive a different number of

photons. The final estimated model is:

$$\hat{N}(t) = \hat{\alpha}\left(\widehat{P_S f_T}(t|S) + \widehat{P_B f_T}(t|B)\right) \tag{6.9}$$

Figure 6-2 shows the steps in the recovery process for different cases of fog and targets. It also provides the SNR between the estimated model and the measured histogram. Fig. 6-3 shows similar results over a wider range of conditions, which demonstrates the success of the suggested technique to model the physical measurements.

6.3.2 Leveraging Spatial Correlations

The model presented so far was pixel-wise. This approach is beneficial in handling heterogeneous fog. However, it ignores the spatial correlations that obviously exist in the scene. For example, it is safe to assume that the fog properties K, μ_s are a smooth function of space (or at least piecewise smooth). It is also known that the scene depth and reflectance are piecewise smooth.

To leverage these properties we introduce a total variation denoiser. Such a denoiser operates on an input image I_{noisy} and solves the following optimization problem:

$$\hat{I} = \arg\min_{I} \sum_{m,n} \sqrt{{I_x}^2(m,n) + {I_y}^2(m,n)} + \lambda \|I_{\text{noisy}} - I\|_2^2 \tag{6.10}$$

We use the ℓ_1 formulation which is known to produce more piecewise smooth results [137, 28]. Here we adapt the implementation in [106], and apply the denoiser on the spatial recovery of K, μ_s, μ, and σ^2.

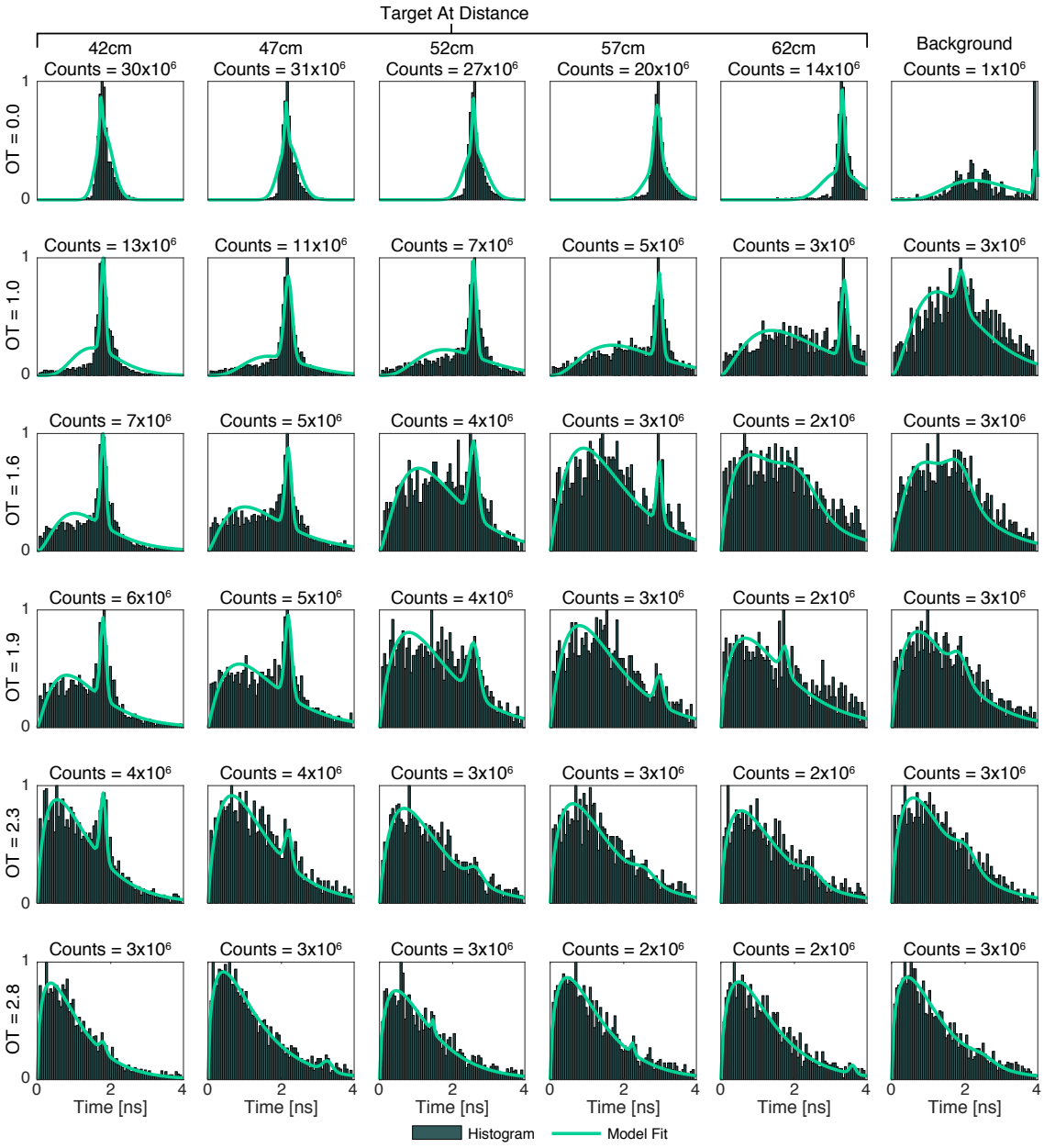

Figure 6-3: The probabilistic algorithm successfully models physical measurements in a wide range of fog conditions, and target distances. Different rows correspond to different optical thicknesses. Different columns correspond to targets at different distances, and a background case. The acquisition time is fixed and identical for all examples. Due to the significant difference in the number of detected photons in these conditions, all plots are normalized to 1, and the total number of photon counts is reported in the title of each plot. The model fails to estimate the target for higher levels of fog and when the target is farther away.

6.3.3 Target Depth and Reflectance Recovery

The properties of the target are encoded by the signal part of 6.9:

$$\widehat{N_S}(t) = \hat{\alpha}\widehat{P_S}\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\exp\left\{-\frac{(t-\hat{\mu})^2}{2\hat{\sigma}^2}\right\} \tag{6.11}$$

Depth Estimation:

The target depth is encoded by the ballistic photons. These would be captured by the early part of the Normal distribution. In practice, the estimated Normal variances in our experiments are in the order of $1-2$ time bins. This is due to both scattering and measurement jitter. We found that using $\hat{\mu}$ is a robust estimate of the target depth. Since the Normal distribution is fitted whether or not there is an object in the pixel, we use the reflectance estimation as a confidence map to reject pixels where no target exists (see below).

Reflectance Estimation:

The reflectance value per pixel is captured by the non time-dependent coefficients of Eq. 6.11. That is: $R = \hat{\alpha}\widehat{P_S}\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}$. We found that including the variance dependent normalization factor provided cleaner results. We further refine the reflectance estimation by multiplying each pixel by the square of its estimated depth (to account for the one-over-depth-square drop-off).

Further Refinement:

In case of low fog, the reflectance is governed by the illumination intensity profile. To account for that we utilize an intensity calibration measurement which was performed offline at the center of our fog chamber. The intensity profile is accounted for only when our estimated optical thickness ($\widehat{\mathrm{OT}}$) is less than 1 (the optical thickness estimation is described below). To create a smooth transition, we use a linear interpolation based on the estimated optical thickness between 0 to 1. When $\widehat{\mathrm{OT}} = 0$, the intensity profile is completely accounted for, and when $\widehat{\mathrm{OT}} \geq 1$ it is ignored.

6.3.4 Optical Thickness Estimation

The background model captures basic physical properties of the fog, as discussed in Sec. 6.2.1. Here, we show that the background model parameters are a strong predictor for the measured optical thickness.

In our experiments we measure the optical thickness as a function of time while fog is added to the fog chamber. This measurement is not used in the reconstruction procedure. We use a single measurement taken without a target in the fog chamber to develop a predictor for the optical thickness based on estimated Gamma model parameters. The results are shown in Fig. 6-4. We found that a model based on $\hat{K}$ is a robust model for the optical thickness. Our optical thickness predictor is:

$$\widehat{\mathrm{OT}} = \theta_1 e^{\theta_2 \hat{K}} \tag{6.12}$$

where the θ-s are the estimated model parameters. We perform a robust fit which results in a model with $R^2 = 0.9987$.

Note that the model slightly underestimates the OT for OT > 2.2. We attempted to add a dependency in μ_s which helped to fix this underestimation at high OT, but made the prediction noisier at lower OT.

This prediction model is powerful for three reasons:

1. It demonstrates that the estimated parameters of the Gamma distribution map to a physical quantity. This is another evidence for the validity of the background model.

2. For many computer vision applications, it is very beneficial to know what is the optical thickness. For example, it can serve as a predictor for the visibility, or the maximum depth, in which our algorithm works well. This may be beneficial to determine cars maximum driving speed.

3. As described in the previous section, this estimate helps to refine the recovered result.

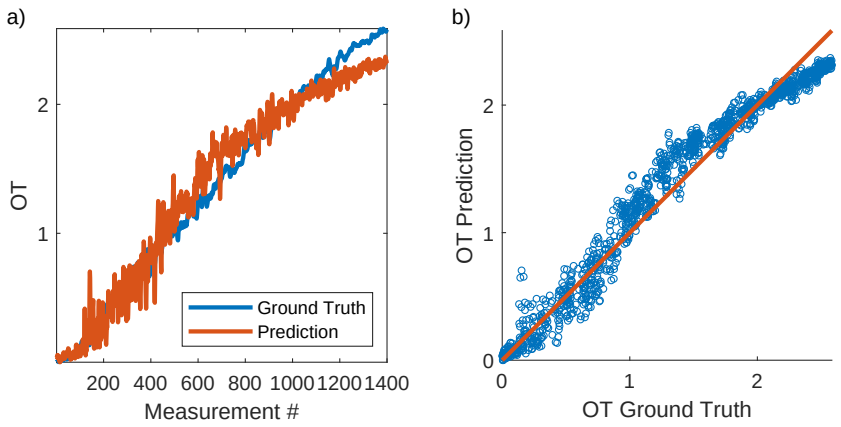

Figure 6-4: Fog background model predicts optical thickness. The estimated background model parameters are used to predict the optical thickness. a) Ground truth and the prediction as a function of time while fog is added to the chamber. b) The optical thickness prediction vs. ground truth, along with a straight line for reference.

It is important to note that the optical thickness in our experiment is measured in transmission mode, while our background model is for reflection mode. Thus, we don't expect to directly estimate the physical mean free path using this predictor. Instead, we find this to be a useful measure of fog density.

6.4 Experimental Setup

The experimental setup is shown in Fig. 6-5. The detector is a PhotonForce PF32 SPAD camera that is composed of 32×32 pixels. Each pixel is single-photon sensitive and time-tags measured photons with a nominal time resolution of 56 ps. The camera exposure time is set to 100 µs (the PF32 measures the arrival time of the first detected photon per pixel per exposure). Each reconstruction is composed of 20,000 frames. In these settings we produce a new reconstruction every 100 µs, while using a sliding window with a history of 2 s.

For illumination, a SuperK pulsed super-continuum laser is spectrally filtered to a narrow band around 580 nm (the camera is equipped with a similar filter to reject the background). The laser repetition rate is 80 MHz with a pulse duration of 5 ps and an average laser optical power of 0.15 W. The laser is diffused before entering

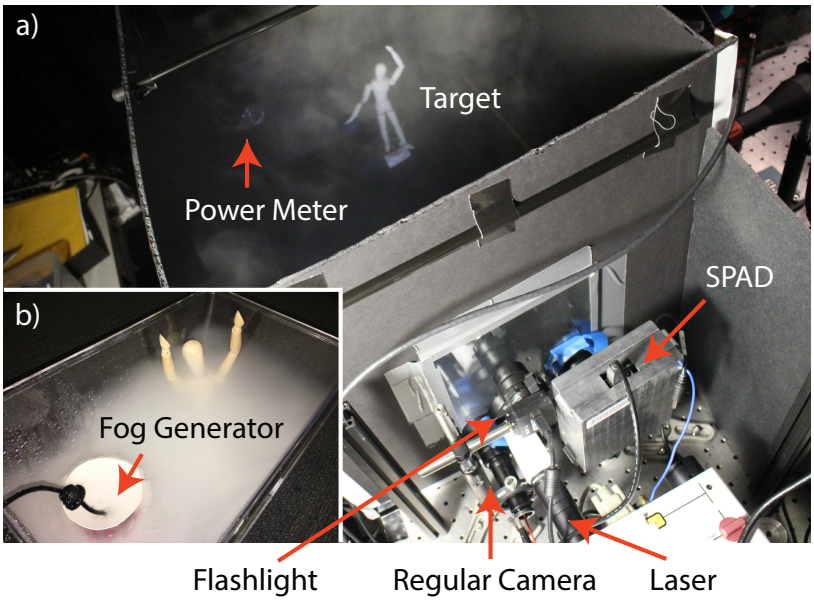

Figure 6-5: Experimental setup. a) The fog chamber with a mannequin inside. This photograph was taken with minimal fog density and shows the SPAD, pulsed laser, traditional camera, and flashlight. Illumination and measurement are performed through a glass window in the chamber. A power meter is placed inside the fog chamber to quantify the optical thickness. The fog generator is composed of an ultrasonic transducer in water and a fan placed on the far side of the chamber (not visible). b) Example of the fog generator inside a small open aquarium. In this case the fan is off, which results in low concentration.

the chamber, and it flood-illuminates the scene (without fog). The camera and laser are positioned in reflection mode.

To evaluate and compare our method we placed a regular monochromatic camera (Point Grey Chameleon) along with an independent CW flashlight at a wavelength of 850 nm. The flashlight average optical power is 1 W and the camera quantum efficiency at this wavelength is 15 %. The different wavelengths were chosen to make sure that the two imaging systems can operate simultaneously without affecting one another. Spectral optical filters are placed on the SPAD and regular cameras to ensure that each camera measures contributions only from its dedicated light source. The flashlight floodlit the tank and was positioned such that it did not illuminate the target directly to reduce the glare from the fog. The camera integration time is 100 ms. In all of the reported results, the background lights are turned off so that both imaging systems equally benefit from a directional illumination source.

The cameras and illumination sources are placed adjacent to the fog chamber. The chamber dimensions are $0.5 \times 0.5 \times 1\,\mathrm{m}^3$. To generate fog, a powerful ultrasonic transducer is placed in water along with a fan (similar to a cold mist humidifier). The fog generator is placed at the far side of the fog chamber. This configuration is capable of producing dense fog with visibility of a few centimeters into the chamber.

The far side of the chamber also holds an optical power meter to measure the fog's optical thickness. Optical thickness at time t is calculated by $-\log(P_t/P_0)$, where P_0 is the power measured when there is no fog, and P_t is the power at time t. This measurement is not used as part of the reconstruction process.

6.5 Experimental Results

The experimental system described above was used to evaluate the suggested approach. Ten different targets are placed in the fog chamber at different locations. The fog generator is turned on, and a continuous capture of SPAD frames is performed until the fog density in the chamber saturates (in the order of 15 min).

To evaluate the results, we compare our reflectance reconstruction to the measurement taken with the regular camera. Note that the NIR wavelength used for the regular camera undergoes less scattering which results in a sharper image, especially at low fog densities. Because of the different perspective and acquisition properties, the regular camera is considered a qualitative comparison. The second comparison is to photon counting with the SPAD camera. In this mode, the camera simply accumulates individual detected photon events (non time-aware). The third comparison is to time gating. In this mode the time bin was selected manually to be the first time bin that holds information about the target. We compute Peak SNR (PSNR) and structural similarity (SSIM, ranges in $[0, 1]$, higher is better) to quantitatively compare our reflectance recovery method to photon counting and time gating (the chosen ground truth is taken from a photon counting measurement without fog). Similarly, the regular camera results are compared to a regular photo taken without fog.

Figure 6-6 shows results for a target composed of a set of four 'E' shapes ($3 \times 5\,\mathrm{cm}^2$)

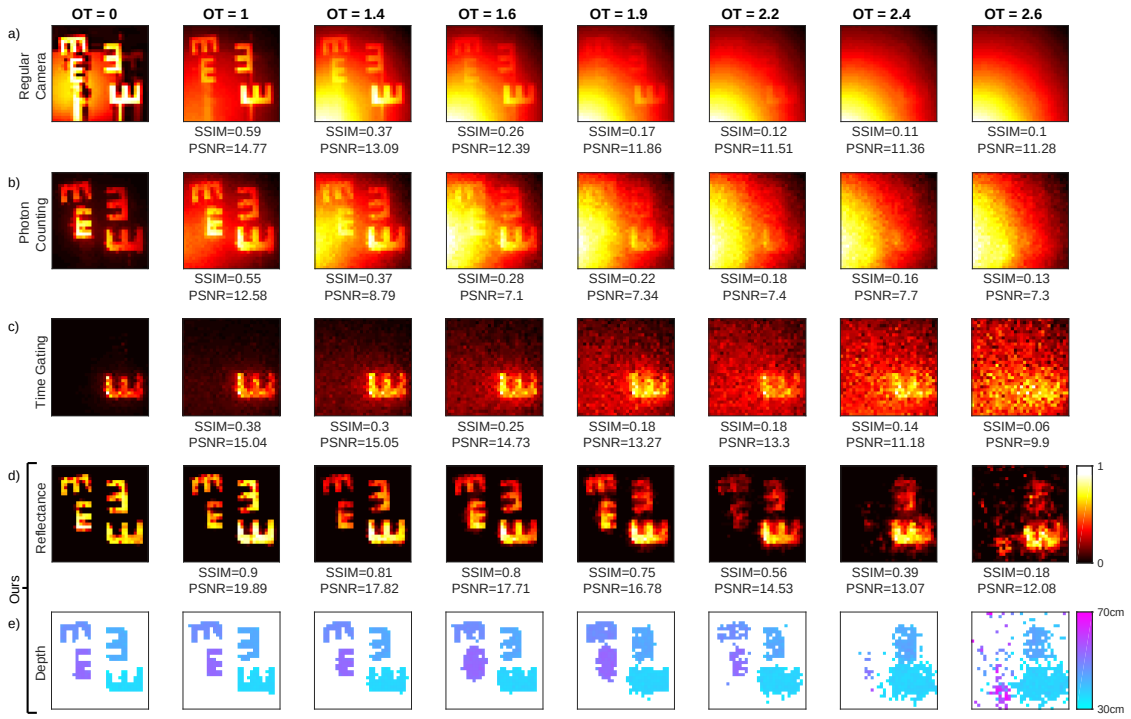

Figure 6-6: Recovery of a multi-depth target at realistic, dense, dynamic, and heterogeneous fog, with the 'E' shapes. Different columns demonstrate cases of different levels of fog. Rows show different reconstructions including: a) Image taken with a regular camera (the longer wavelength used for this measurement undergoes less scattering, which results in less challenging imaging conditions). b) Result with SPAD camera in photon counting mode. c) Result of time gating using the SPAD camera, where the time gate was selected manually to the first time bin with meaningful information. d) Reflectance reconstruction with our technique. e) Depth reconstruction with our technique. SSIM and PSNR metrics provide quantitative comparisons. The left column shows a measurement without fog (ground truth).

at different orientations and depths ($36, 43, 47, 53$ cm from the camera). As can be seen from the results, our method is able to reject the significant backscatter that governs the regular camera and photon counting results. In comparison to time gating we note that time gating is much noisier, requires one to manually select the correct time bin, and recovers only one depth. The suggested method outperforms these techniques in both SSIM and PSNR, and degrades much slower with increasing fog levels. Furthermore, the method accurately recovers the depth of the different targets up to OT $= 2.2$, after which it loses the farther target while recovering the closer ones.

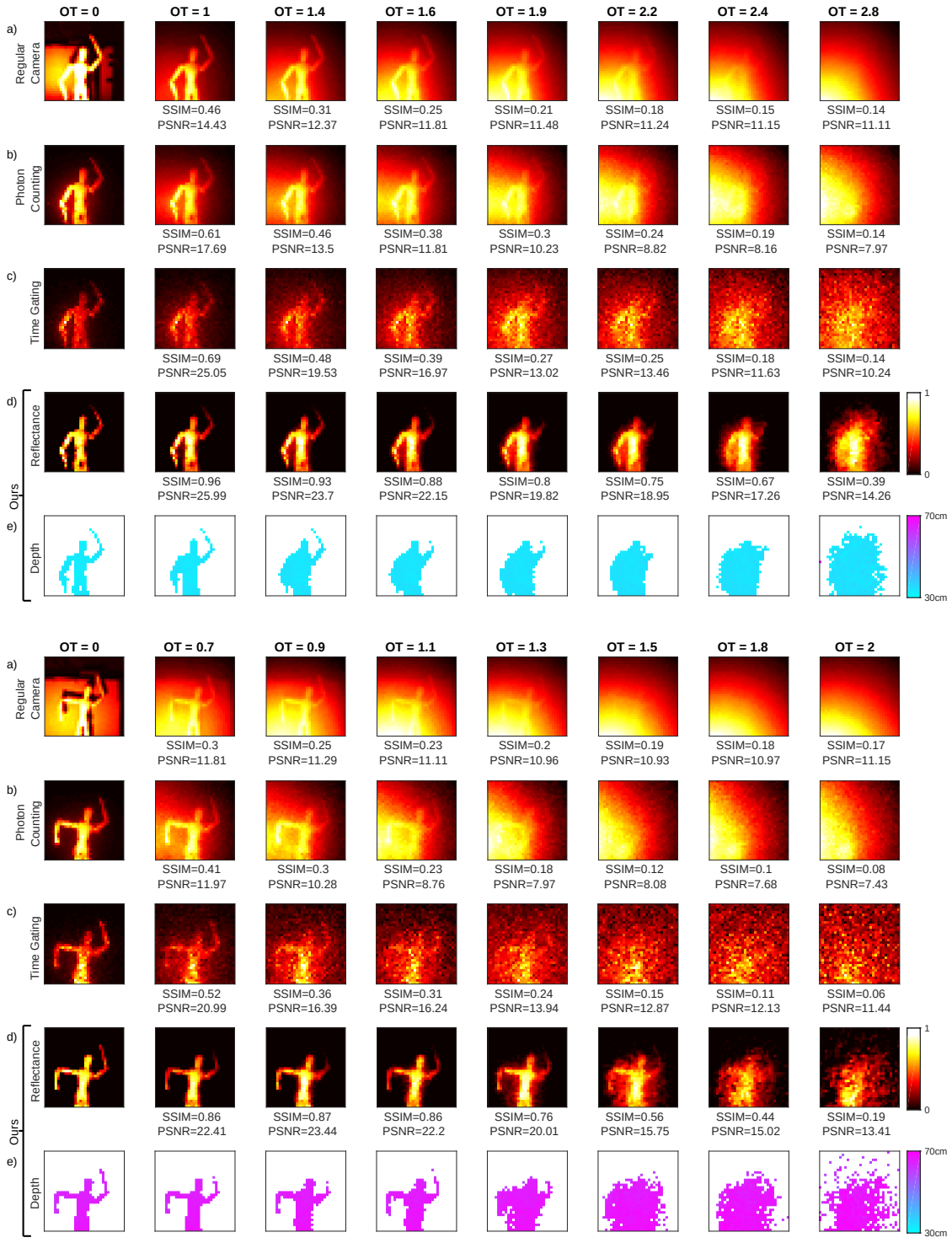

Figure 6-7: Recovery of complicated structures occluded by dense fog. Top – a Mannequin 35 cm away from the camera. Bottom – a Mannequin 70 cm away from the camera. See Fig. 6-6 for panels description.

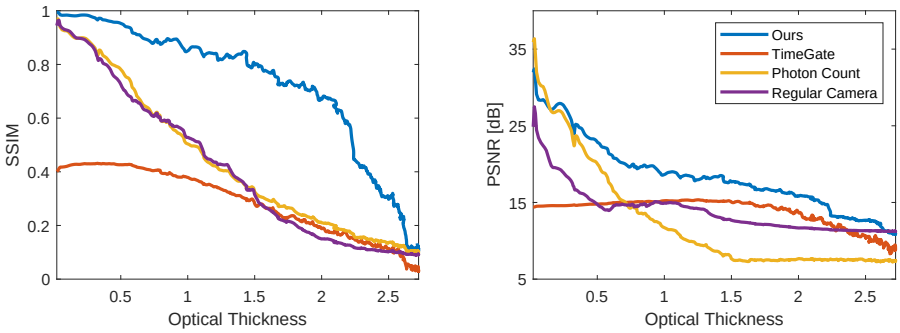

Figure 6-8: The suggested approach produces superior results over the entire range of fog levels. Showing image recovery accuracy vs. optical thickness. The accuracy is evaluated with a) SSIM, and b) PSNR on the 'E' shapes.

These results are demonstrated in more experiments as follows. Fig. 6-7 shows recovery of a mannequin (20 cm tall) placed at the depth of 35 cm and 70 cm away. In both cases, our technique is able to reject the background around the mannequin very effectively while the other results are saturated by the background. This is more apparent when the target is farther away. Note that when the target is farther away, we recover it up to OT = 1.5 after which we recover a coarse shape and depth. When the target is closer, we resolve the object features up to OT = 2.4. This result demonstrates the tradeoff between fog level and maximum detectable depth. When there is less fog, our technique recovers more distant targets.

6.6 Analysis

Unless otherwise noted, the analysis described is experimentally performed exactly as described above, with the same code and parameters.

6.6.1 Image Recovery Accuracy

Figures 6-6, 6-7 demonstrated the reconstruction quality at specific optical thicknesses. It is useful to quantitatively compare our technique to the references (regular camera, photon counting, and time gating) over the full range of fog densities considered here. This is demonstrated in Fig. 6-8 for the 'E'-shaped targets.

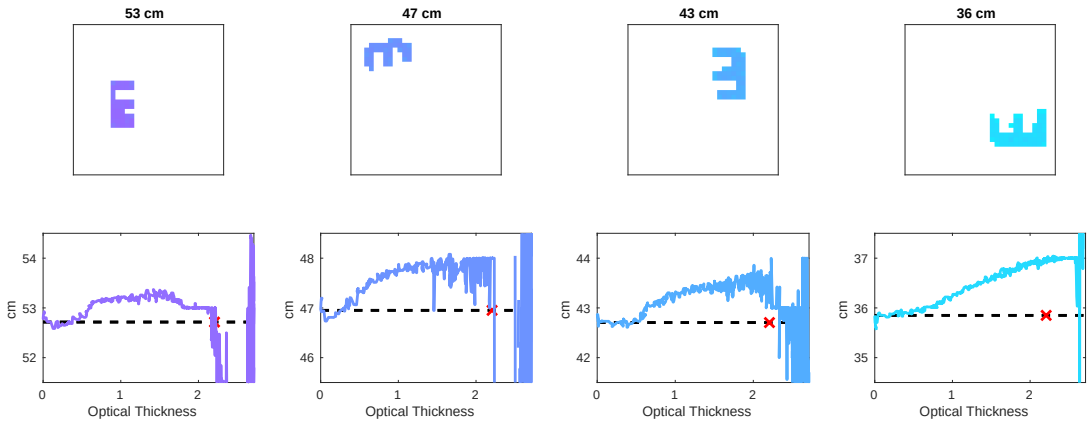

Figure 6-9: Depth recovery accuracy as a function of optical thickness. Demonstrated on four targets (columns). Top row shows the segmented mask used for each target. Bottom row shows the recovered depth for each target as a function of optical thickness. The dashed black line indicates the ground truth based on the first few frames without fog. The red cross indicates OT = 2.2 which is the optical thickness in which we lose the farther targets.

Target Distance	53 cm	47 cm	43 cm	36 cm
OT $= 0 \rightarrow 2.7$	-2.318 ± 5.054	-0.035 ± 3.31	0.067 ± 1.097	1.401 ± 2.002
OT $= 0 \rightarrow 2.2$	0.339 ± 0.189	0.725 ± 0.308	0.569 ± 0.285	0.627 ± 0.387

Table 6.1: Depth recovery error for the four targets in Fig. 6-9. The top row considers all captured data (up to OT = 2.7). The bottom row considers data up to OT = 2.2 (the optical thickness in which we lose the farther targets). All numbers are provided in cm.

We note that our approach provides substantially better quality over the full range of optical thickness, and is especially noticeable from the SSIM metric. Note that the time gating approach starts at a low quality since it recovers just one plane.

6.6.2 Depth Recovery Accuracy

We analyze the accuracy in depth recovery with the 'E' shapes. Each letter is segmented, such that we probe the accuracy in recovering the following depths: 53, 47, 43, 36 cm. For each target plane and each time point, we use the median of the estimated depth.

The depth estimation for the four depths is plotted as a function of optical thickness in Fig. 6-9. As a baseline for the recovery, we also plot the depth based on the

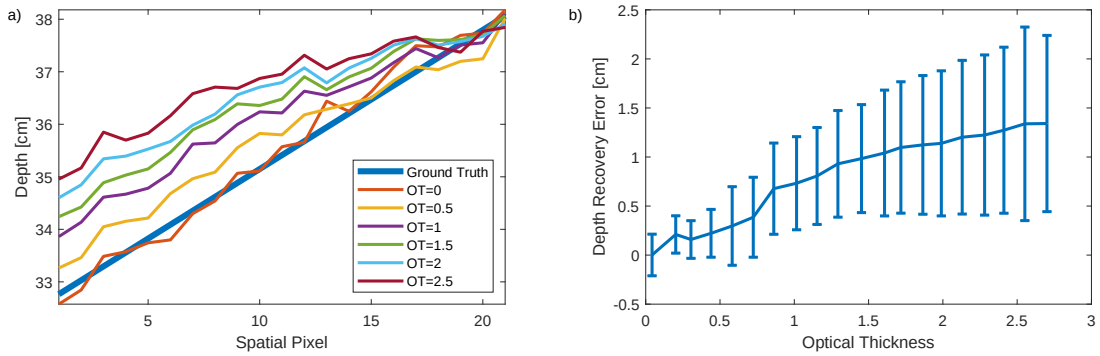

Figure 6-10: Depth recovery accuracy for a slanted plane. a) The estimated depth profile as a function of spatial pixel for different optical thicknesses. The ground truth depth profile is the thick blue line. b) Error and standard deviation bars for the deviation of the estimated depth profile from a line, as a function of optical thickness.

first frames (with no fog). As the visibility reduces, the depth accuracy drops (first to farther targets). As expected from our model, we tend to estimate the target to be slightly farther away (within a centimeter). That is because the depth estimation is based on the Normal distribution mean, which is biased by scattered light.

Another appreciable point in the figure is the point of failure. Once a target is lost, the accuracy drops dramatically because we have false detections at these pixels. See for example in Fig. 6-6 columns for OT $= 2.4, 2.6$. Table 6.1 provides the overall error for each depth. We provide two numbers for each target: 1) based on all data (up to OT $= 2.7$), and 2) up to OT $= 2.2$, where all targets are properly resolvable. When there is sufficient signal and we resolve the target, our approach has a sub-centimeter bias and error for targets in the range of $36 - 53$ cm.

To further evaluate the depth recovery accuracy, we place a slanted white wall inside the fog chamber. The wall covers most of the camera field of view (22 pixels), such that the closest point to the camera is at a distance of 32 cm, and the furthest point is at a distance of 38 cm. Similarly to previous experiments, we add fog to the chamber and report the depth reconstruction result as a function of optical thickness. Fig. 6-10 shows the results of this experiment. In Fig. 6-10a, we plot the recovered depth profile as a function of the spatial pixels for different optical thicknesses. In this result, we note that the estimate bias is a function of the target depth. For closer

targets, the bias is larger than for more distant targets, and at 38 cm it diminishes.

Another noticeable property is demonstrated in Fig. 6-10b – the deviation from a line. In that case, for each optical thickness we recover the depth profile and perform a simple linear fit. We then calculate the mean and standard deviation error of the depth profile from its fitted line. Effectively we estimate how far the depth profile is from a line. As can be seen in the figure, the error bias and standard deviation increase as the fog density increases. In both figures, we note that the error is bounded by 2 cm throughout.

6.6.3 Reflectance Recovery Accuracy

To measure the accuracy of our reflectance recovery, we place a checkerboard-like target, where each square is of a different shade of gray. In total, we have six different shades, ranging from black to white, spread on a 6×6 grid, such that each color appears six times (Fig. 6-11a). The target size is $10 \times 10\,\mathrm{cm}^2$. We place the target at a distance of 70 cm inside our fog chamber and repeat the experiment and reconstruction process. Since we recover relative reflectance, the reconstruction is rescaled to $[0, 1]$. Then, the mean value across the six different squares for each color is calculated. In the end, we have six reflectance values per recovered frame. The results are shown in Fig. 6-11, and our recovery is compared to time gating and photon counting.

We note that this is a particularly challenging task. First, the target is far away. Second, the photon flux that illuminates each square changes dramatically as the fog is added – initially, it is dominated by the illumination spatial profile, and later mostly affected by the heterogeneous scattering. From Fig. 6-11, we note that our technique correctly sorts all colors up to OT $= 1.7$, where we start to observe mixing. At lower fog densities, our technique recovers the correct color with marginal error (and performing better on the brighter colors). On the other hand, the time gating and photon counting techniques present significant mixing for OT ≥ 1.1. The photon counting technique fails for OT ≥ 1.2 when the recovery is dominated by the fog.

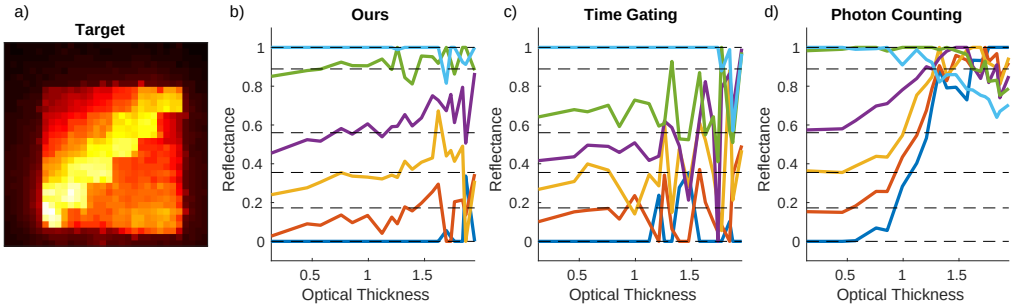

Figure 6-11: Reflectance recovery accuracy. a) Photon count measurement of the reflectance target without fog. b) Recovery of the reflectance values for the six different values with our technique, as a function of optical thickness. Ground truth marked by black horizontal dashed lines. c-d) similar to b, for time gating and photon counting respectively, demonstrating mixing of colors at lower levels of fog.

6.6.4 How Many Photons Are Measured?

One of the challenges in imaging through fog is that as the fog level increases, we measure less overall photons, and the ratio between background and signal photons increases. Both are making the problem more challenging. To demonstrate this trend we plot the total photon counts measured at a SPAD pixel while fog is being added. We use the 'E' targets measurement, and compare the number of photons measured at a pixel selected on one of the target planes, as well as a background pixel. These results are plotted in Fig. 6-12.

For this comparison, we used the same settings as before, specifically 80 MHz laser repetition rate with a frame exposure time of 100 µs window and 20,000 frames per data point. In this case, each pixel can record up to 1 photon out of 8000 pulses per exposure time, and we aggregate 20,000 such frames for analysis. As demonstrated in Fig. 6-12a, for OT < 1 we get more photons on a target pixel as opposed to a background pixel. For OT > 1 the number of photon counts on pixels with and without a target is comparable. This is also apparent in the photon count reconstructions shown in Fig. 6-6. Another noticeable property is that as the fog level increases, we get less photons on pixels with a target, and more photons on pixels without a target. We note that our photon acquisition efficiency for OT > 1 is on the order of $1 : 60{,}000$ (for every 60,000 pulses we capture 1 photon per pixel).

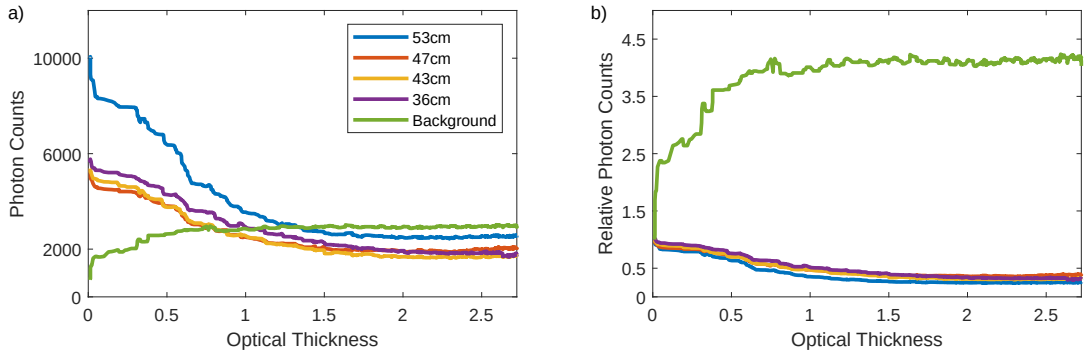

Figure 6-12: Photon counts drop as fog is added. Photon counts per pixel on targets at four different depths, as well as a pixel without a target (background). a) Photon counts vs. optical thickness. b) Same as a) where the curves are normalized by the photon counts at OT = 0.

Lastly, the results in Fig. 6-12a are also a function of the illumination spatial profile (which is very noticeable for the lower levels of fog). To account for that, we normalize each curve based on the photon counts at OT = 0. These results are shown in Fig. 6-12b. As can be seen, the number of photons measured on pixels with a target drops by an order of 2×, while the number of photons measured on pixels without a target increases by an order of 4×. Together, this indicates that the contrast is reduced by an order of 8× between the no fog case, and fog levels above OT = 1 (and 10× for OT > 1.5). This clearly demonstrates the challenge in imaging with photon counting mode or a regular camera.

6.6.5 How Many Photons Are Needed For Reconstruction?

In all of our reconstructions we used a fixed value of 20,000 frames. We now analyze the sensitivity of our results to this parameter. The sensitivity to acquisition window duration is similar to measurement SNR sensitivity analysis in a classic acquisition system. We use the 'E' dataset and perform reconstructions (for our technique, time gating, and photon counting), while varying the number of available frames. These results are presented in Fig. 6-13. For this analysis, we use the SSIM quality measure to compare the results where the ground truth is the recovery with no fog and maximum frames (30,000 in this case).

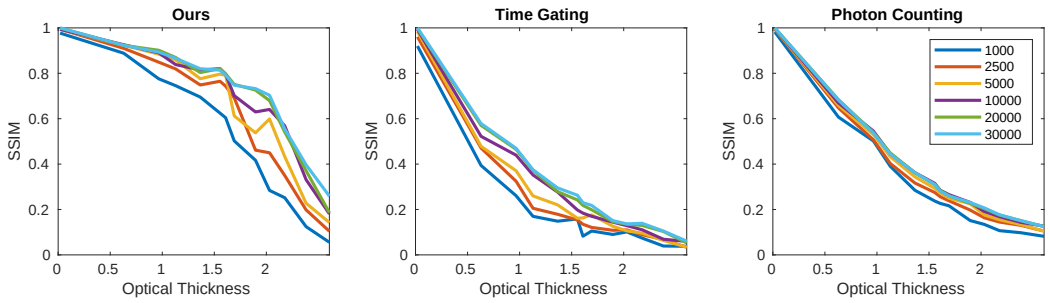

Figure 6-13: The effect of exposure window on recovery quality. SSIM metric for the recovery of the 'E' shapes as a function of optical thickness, each curve is the result of a different number of frames. a-c) Our technique, time gating, and photon counting respectively. Ours performs equally well with fewer photons at low fog.

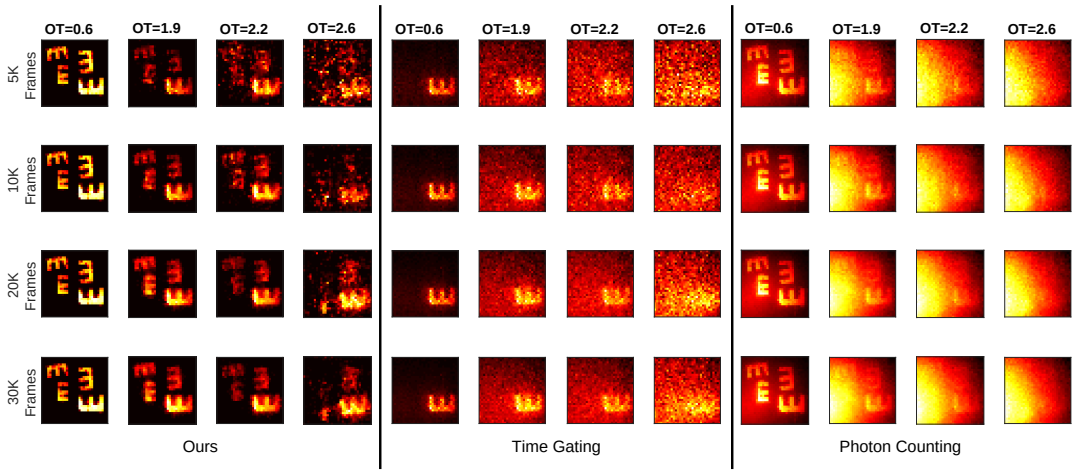

Figure 6-14: Qualitative results of recovery with varying exposure window. Columns show different optical thicknesses, rows are different allowed number of frames. This demonstrates the adaptive property of our approach. Time gating marginally gains from having more frames (even at lower levels of fog). Photon counting does not gain from having more frames regardless of the fog level.

As expected, when the fog level is low, we require fewer frames. As the fog level rises we gain from having more frames. More specifically, up to OT = 0.5, 1000 frames are sufficient. Then, up to OT = 1.5, 10,000 frames are enough, followed by a gain from having 20,000 frames as used in our other experiments.

Comparing time gating and photon counting, we note that time gating always gains from more photon counts, while photon counting performs roughly the same for all options. We also provide qualitative results in Fig. 6-14 that clearly demonstrate

that our approach requires fewer photons when the level of fog is lower.

The decision on the number of frames to use is done in software and does not affect the measurement process. Thus, it is possible to adaptively change the window duration as a function of the estimated optical thickness. The challenge in this approach is that the recovery quality is a function of the optical thickness and the target distance. Thus, by reducing the number of photons we may potentially miss farther away targets at denser fog.

6.7 EM Algorithm

The imaging algorithm included several assumptions, most notably the following subtraction: $f_T(t|S) \approx f_T(t) - f_T(t|B)$ (which also approximated $P_B \approx 1$). We can consider the estimated model as an initializer to an expectation maximization (EM) algorithm to refine the estimates and fix any steps that may have been mathematically inaccurate. EM is a natural fit to the problem we have here — splitting data into two distributions, while having to estimate the parameters of these distributions. The problem is, of course, ill-posed without a guarantee to converge to a global minimum. But in our case we already have a good initial guess, which makes EM appealing. By using EM we can also select the photons that hit the target from the data, as it is naturally estimated as part of the algorithm. We note that EM algorithms are in general very hard to initialize, and it'll be challenging to use an EM algorithm to completely replace our algorithm.

An EM algorithm is composed of two steps:

- **Expectation Step:** given the parameters of the distributions, it calculates the probabilities for a photon to belong to each class (membership probabilities).
- **Maximization Step:** with the membership probabilities, it calculates the distributions' parameters while taking into account the membership probabilities as weights.

Appendix C provides the derivation for the update rules of these steps for a mixture

Figure 6-15: Recovery with the additional expectation maximization step. Different columns show different levels of fog. The rows compare the recovery with and without the additional EM step for reflectance and depth. We note that up to OT = 1.4, including the EM step improves the results, and after that the results degrade.

of Gamma and Normal probabilities.

In our implementation we start with the initial parameters as described above, and then perform 500 iterations of the EM algorithm. Due to numerical instability of the algorithm we perform 5 different restarts, and in each restart we slightly perturb the initial values. In total we perform 2500 iterations. The selected parameters are chosen based on the iteration that minimized the log-likelihood of the EM algorithm.

At this point we can separate between background and signal photons. To that end, we evaluate the probability that each photon belongs to the signal or background class:

$$\widehat{P_S f_T}(x^i|S) \underset{\text{Signal}}{\overset{\text{Background}}{\lessgtr}} \widehat{P_B f_T}(x^i|B) \tag{6.13}$$

If the expression to the left (right) is larger we classify this photon as a signal (background) photon.

The number of classified signal photons N_S corresponds to the target reflectivity at that pixel. The mean of the signal photons corresponds to the target distance at

that pixel. Figure 6-15 shows recovery results with the EM algorithm.

In practice, we find that the EM algorithm performs better at low levels of fog. When the fog level is higher, it actually degrades the result. Further analysis of these cases revealed the following: Because photons at the tail of the Gamma distribution are associated with the Normal distribution, the Gamma distribution shifts to the left, which results in a mismatch of the tail that is accounted for by shifting the Normal distribution. Obviously, this solution will have a lower log-likelihood, but it is nevertheless a wrong solution (it matches well more background photons at the expense of fewer signal photons). This is a result of model mismatch (the physics is more complicated than our Gamma and Normal model).

Another strong limitation of the EM algorithm is the computational cost. Without the EM step, our algorithm runs at about $\sim 5\,\mathrm{s}$ per frame (this time is governed by the total variation denoising steps), and with the EM addition, this number increases to $\sim 200\,\mathrm{s}$. While these numbers are based on an unoptimized Matlab code on a standard desktop computer, they reflect the significant run-time penalty of this step.

Due to these two limitations we prefer our basic model.

6.8 Reading in Dense Fog

Many computer vision and robotics problems, such as recognition and identification, are solved today with machine learning. Here, we consider such tasks in the case of fog. One potential solution is to train data-driven models on measurements with fog. This may be very challenging, as some of these tasks are particularly hard, even without obstructions such as fog. Furthermore, as discussed here extensively, fog is a continuum of densities. This would require gathering data in a wide range of fog conditions, which is very challenging. The alternative we suggest here is to use our approach to produce a photo and a depth map as if the fog were not there. These reconstructions can be fed into the computer vision algorithms that were trained with fog-less data as is. Thus the data-sensitive part of the algorithm can be trained without fog, and when integrated with our approach, it can work in foggy scenarios.

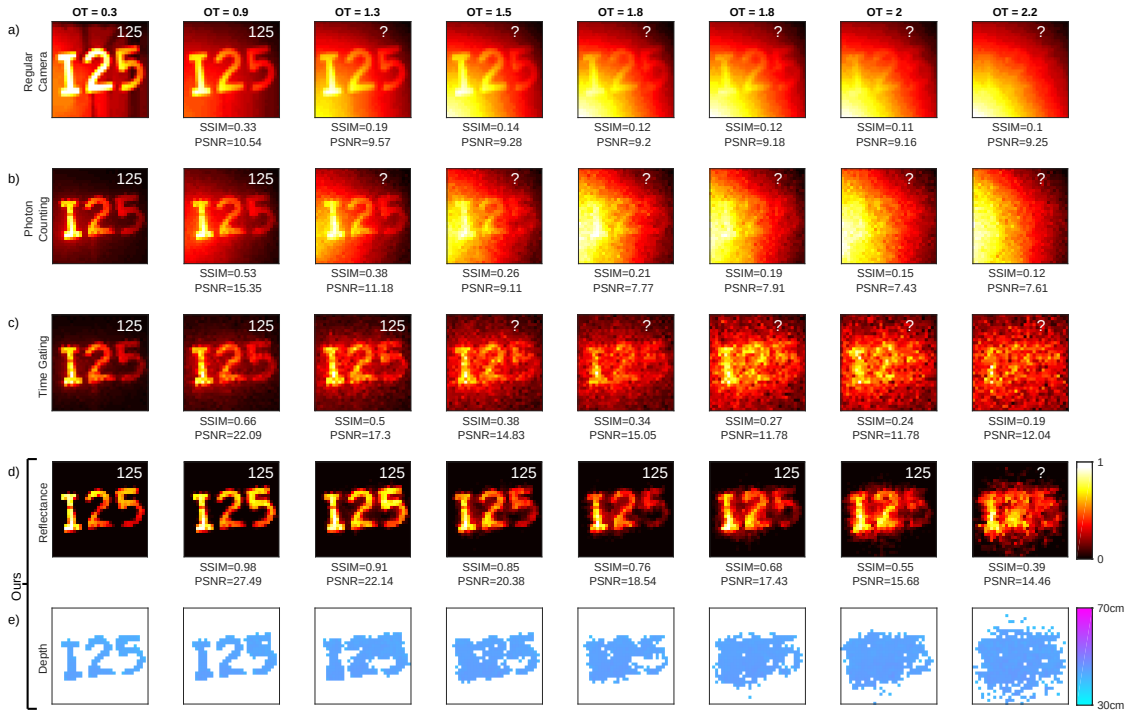

Figure 6-16: Reading in dense fog – recovery of the '125' target. The white text at the top left corner of each panel shows the OCR result on that image. '?' indicates no text found. See Fig. 6-6 for panels description.

We demonstrate this approach in the task of optical character recognition (OCR). To that end, we use an off-the-shelf OCR by the Google Vision API. To the best of our knowledge, this OCR was not trained specifically to work with low contrast inputs that characterize fog. We place several letters and digits in the fog chamber and repeat the experiment and recovery procedures. The OCR performance on our reflectance recovery is compared to regular camera, photon counting, and time gating.

Figure 6-16 shows recovery examples for a target composed of the digits '125' placed at a distance of 45 cm from the camera. As can be seen, the OCR performs well on our reconstruction, while it fails on the noisy time gating, as well as the low contrast measurements that characterize the photon count and regular camera recovery. Interestingly, the OCR fails to recover the digits even when it is still relatively easy for humans. This further demonstrates the challenge and need for descattering solutions, as demonstrated in this dissertation. Fig. 6-17 shows another example with

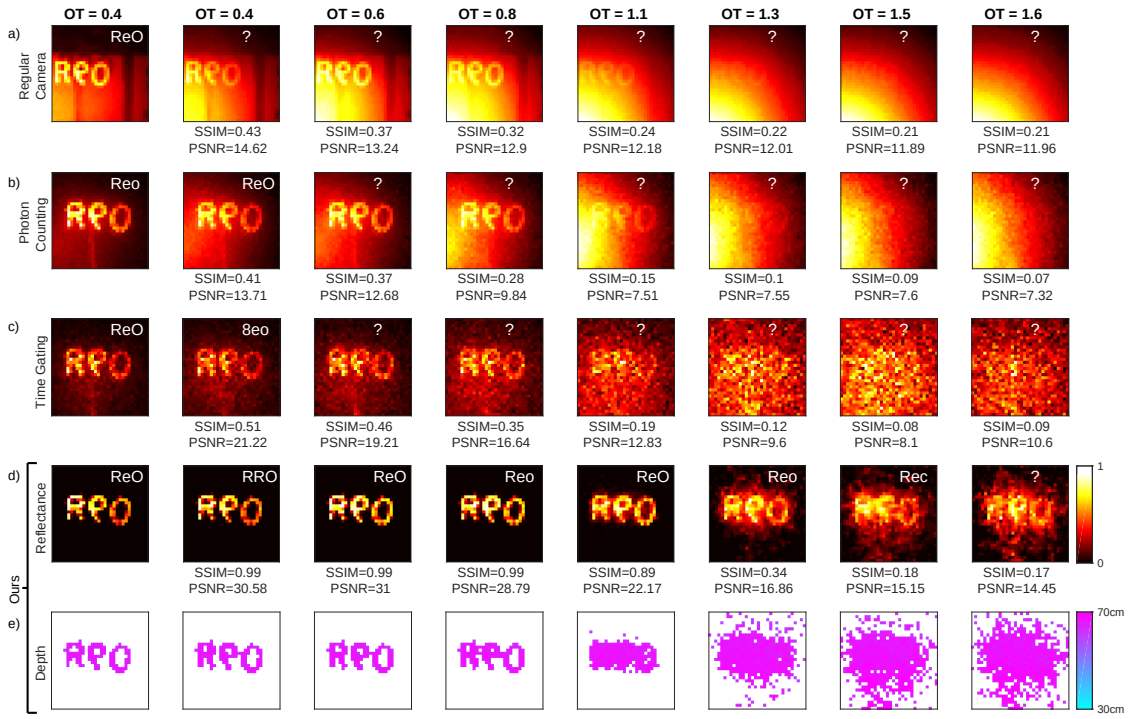

Figure 6-17: Reading in dense fog – Recovery of the 'Re0' target. The white text at the top left corner of each panel shows the OCR result on that image. '?' indicates no text found. See Fig. 6-6 for panels description.

a target composed of the letters 'ReO', which was placed 70 cm away from the camera. Since this target is farther away, it is even more challenging.

In total we performed four experiments with the following text targets:

- The digits '125' were placed 45 cm away from the camera. Each digit was 4 cm tall, and the entire target was 10.5 cm wide.

- The letters 'ReO' were placed 70 cm away from the camera. Each letter was 3.5 cm tall, and the entire target was 10 cm wide.

- The letters 'THRU' were placed 35 cm away from the camera. Each letter was 4 cm tall, and the entire target was 13 cm wide.

- The letters 'STOP' were placed 34 cm away from the camera. Each letter was 3 cm tall, and the entire target was 11 cm wide.

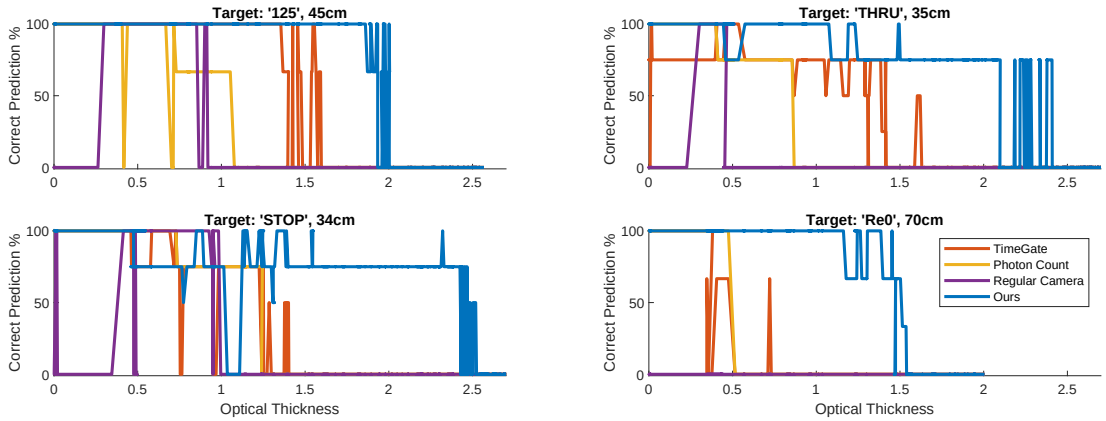

Figure 6-18: Reading in dense fog – Accuracy as a function of optical thickness. Our approach allows the off-the-shelf OCR to correctly classify text over a wide range of fog conditions. Four different targets are demonstrated, the panels' titles indicate the text of the target and its distance from the camera. The y-axis is the percentage of correctly classified characters.

We evaluate the OCR performance on each of these targets with our method and compared to regular camera, photon counting, and time gating. To get quantitative results we define a correct prediction metric as the percentage of correctly classified letters or digits. We allow confusions between similar uppercase and lowercase letters, as well as similar characters (O and 0, P and F, etc.). The correct prediction metric is reported for a wide range of optical thicknesses in Fig. 6-18.

This result demonstrates the significant advantage of our approach. As can be seen, with our approach the off-the-shelf OCR is able to recover the text, while with the other approaches it fails rather quickly. The most apparent result is for the 'ReO' target, which was the most challenging in terms of target size and distance. In this case, the other techniques fail very quickly at OT = 0.5, while our approach can classify correctly all characters up to OT = 1.2, and at least two correct characters up to OT = 1.5. Similar trends are observed with the other targets. We note that the four-letter targets ('THRU' and 'STOP') were not illuminated uniformly, which results in a loss of the rightmost letter rather quickly in all cases.

6.9 Discussion

6.9.1 Limitations

The key limitations of the suggested approach are:

1. Our approach is pixel-wise and neglects the spatial nature of scattering. While this is enough to reject the background from the measurement, it is not able to spatially deblur the signal. Spatial blurring may potentially be more prominent in large scale scenes and high-resolution sensors.

2. The demonstrated results are produced with a history window of 2 s (with a mean of 2440 photons per recovery). While this is enough for dynamic fog with a stationary scene, it is not enough for dynamic scenes. This may be alleviated with dynamic window techniques as described above.

6.9.2 Sensitivity to Sensor Spatial Resolution

The sensor used here is composed of only 32×32 pixels. SPAD cameras with megapixel spatial resolution and nanosecond time resolution have already been demonstrated [110]. Such sensors would also be useful as part of an imaging framework that accounts for the complete space-time scattering profile similar to the one suggested in Chapter 5. This would potentially sharpen the results further.

6.9.3 Sensitivity to Sensor Time Resolution

The suggested imaging method is based on the notion that background and signal photons have different statistics in time. This allows one to distinguish between them and reject the back reflectance from the fog. As the detector time resolution reduces, this ability diminishes. The relevant time scales to consider are the standard deviation of the background Gamma distribution and Normal target distribution. The sensor time resolution should be smaller than both. Another aspect of time resolution is its mapping to depth resolution and accuracy, which is likely to be stricter in large scale scenes.

6.9.4 Sources for Model Mismatch

Several aspects were neglected from our model:

Noise

As mentioned before, the dark counts in our experiments are negligible compared to the signal. This may be an issue when attempting to recover with fewer frames.

Absorption

The model essentially treats absorption as any other optical loss in the system. Since the model only takes into account measured photons, it is invariant to the number of actual photons sent to the scene. Thus, absorption and other losses are irrelevant to the reconstruction procedure and would only affect the total acquisition time.

Background Model

The background model used here assumes that K, the number of scattering events before detection, is a constant parameter similar to μ_s. In practice K is a random variable that is realized per photon detection. We also note that the mean of the background Gamma model is $\mathbb{E}T|B = K/\mu_s$. We found this relationship to be extremely accurate with Monte Carlo simulations. Given the photon time of arrival x, the number of scattering events it underwent is $K = x\mu_s$. We found this predictor to work well above $K = 10$ for a wide range of scattering conditions. Thus we conclude that while K should be an independent variable from $T|B$, they are in fact highly correlated. As a result, this assumption is not a strong source for model mismatch, particularly at high densities of fog.

We also explored other modeling alternatives, for example, Stable distributions which are common in Brownian motion modeling. We found that a Stable distribution is usually a very good fit for our background; however it was not as robust as a Gamma distribution. This can be explained by the noisy nature of our measurement, and the challenge of fitting a Stable distribution that does not have a closed form probability

density function.

It is worth noting that fitting heavy-tailed distributions, such as the one observed in fog, is usually challenging. In our case, a practical choice was a Gamma distribution which is physically grounded, and relatively easy to estimate. Furthermore, based on our measurements, we note that the Gamma distributions are an extremely good fit when the fog level is high. For low levels of fog, we observe distributions that are more uniform and exhibit less dynamics in time. As discussed in the algorithm description, we naturally overcome this model mismatch when calculating the prior probabilities, since in low fog situations the signal component is dominant anyway.

Signal Model

The signal model used here is a Normal distribution. We found this to be a practical and efficient approximation. However, there is one obvious flaw with this approximation – it results in biased depth estimates (up to 2 cm for 35 cm away targets, and dropping as the target is farther away). An unbiased depth estimate would be based on the ballistic photons (or estimating their time of arrival).

A potentially useful model that we explored was based on an Inverse Gaussian or Lévy distributions. These distributions are the solution for the first arrival time of a Brownian motion particle to a specific location in the volume. The Lévy model is the solution for the case without a drift, and the Inverse Gaussian is the solution for the case with a drift. The distributions parameters directly model the distance to the source – which is exactly the depth we want to estimate here. We found several fundamental issues with this approach:

1. Estimating a mixture of two heavy tail distributions (background and signal) is numerically challenging. One of the biggest advantages of using the short tail Normal distribution model is that it is sufficiently different from the background model and is easier to estimate (including available robust estimators).

2. The underlying Brownian motion assumes: 1) isotropic scattering, and 2) the source is a delta function in time. Both are not accurate in our case because:

a) Fog is mostly forward scattering and not isotropic; and, b) the scattering on the way to the target illuminates the target with a non-delta like function.

3. Challenging fit: Some of these models require estimating at what time the target starts to reflect light. This estimate is challenging, as it is dependent on the background heavy tail distribution, which is unstable in our case.

While one of these models would potentially be more theoretically accurate, and provide an unbiased estimate of the depth, in practice we found it to be highly unstable, and choose to use the Normal distribution instead. It is possible that sensors with better time resolution would provide better sampling of the time profile, with less jitter, in which case these models would be easier to estimate and their results would be more meaningful.

6.9.5 Real World Considerations

Two points should be taken into account when considering using the suggested approach for other scenarios and real-world applications:

- **Scene scale:** We expect our technique to work as-is for larger scale scenes with similar optical thicknesses. Consider a specific concentration of fog in our chamber. When the chamber volume increases, the fog concentration is reduced. The number of scattering events remains the same, while the mean free path increases, and the effective scattering is similar. Since the mean free path increases, we expect to relax the demand on the sensor time resolution. However, as the distance between scattering events increases, the spatial nature of the scattering may become more significant. Tackling such cases would require detectors with higher spatial resolution and the consideration of the space-time scattering profile discussed above.

- **Use of existing hardware:** The development of LIDAR systems for self-driving cars is an ongoing process. Some of these systems are based on pulsed illumination with time gating for depth mapping and are usually based on raster

scanning. This is very similar to the SPAD camera used here. And, as mentioned above, we expect to relax the demand for high time resolution. Thus, in general, we don't expect hardware to be a significant constraint.

6.10 Conclusions and Future Works

We experimentally demonstrated a probabilistic computational imaging technique for seeing through dense, dynamic, and heterogeneous fog. The technique provides the image and depth map of an occluded scene and was demonstrated in a wide range of optical thicknesses, and with ten different targets and geometries. The technique does not require any prior knowledge about the fog, which enables it to operate, by design, in a wide range of conditions. Furthermore, the hardware requirements of the system are similar to LIDAR, making it suitable for various imaging applications.

One of the notable examples presented in this chapter is the introduction of new capabilities to an off-the-shelf OCR for detecting and classifying text in dense fog, without modifying the OCR. We showed that we can simply feed the output of our technique to the OCR and seamlessly identify text occluded by dense fog. This demonstrates the potential to extend many other computer vision tasks to operate in degraded weather conditions without (or with minimal) modifications to such classification, detection, or recognition algorithms.

Potential future work includes:

- Better coupling between the suggested technique and more demanding computer vision algorithms. For example, recognition tasks that involve the photo and depth map of the scene in degraded weather.

- Further refinement of the algorithm. For example, by limiting the EM algorithm to refine the input without significant deviation from the initial condition, or an alternative algorithm that would jointly estimate the six model parameters.

- Accounting for spatial blur. As mentioned in the limitations discussion, our model ignores the spatial scattering that signal photons undergo between the

target and detector. This can potentially be solved by combining the approach suggested in Chapter 5 to overcome such blur. Our initial studies found that, at least in the cases presented here, this spatial blur has limited effect. However it may be more noticeable with higher resolution sensors, or larger scale scenes.

- Coupling with structured light or raster scan. Most LIDAR systems raster scan the scene. Thus raster scanning, instead of flood illumination, is a natural extension. In general, raster scan of the scene will result in an easier estimation problem. The current approach naturally extends to such cases, but specifically accounting for such structured light in the image formation model may result in better results.

- Scattering beyond fog. While the presented approach was demonstrated in fog, it is relevant to other scattering media. For example, we have already performed initial tests in turbid water [150]. It may also be beneficial in other occluding regimes, such as smoke or dust.

Chapter 7

Conclusions

By developing probabilistic and data-driven algorithms that leverage statistics of scattered photons, we tackled the dependency of computational imaging on highly calibrated and accurate physical models as well as long acquisition times. These are fundamental limitations in computational imaging, which prohibit scaling solutions outside the lab. This challenge is even worse in imaging through scattering with visible light, which is a highly ill-posed problem. The techniques presented here aim to tackle these limitations by design. With **shorter acquisition times**, **calibration invariant** imaging, **data-driven** computational imaging, **optimized compressive imaging**, and robust **probabilistic frameworks**, we hope computational imaging through scattering media will have a tangible real-world impact. We covered four different scattering regimes and applications as summarized next.

Chapter 3 introduced the case of lensless imaging as a simple example of sparse scattering. In that case, the lack of a lens results in a single scatter event. To overcome that challenge we introduced FemtoPixel — a compressive time-resolved single pixel imaging framework. As part of this framework, we studied, from first principles, the space-to-time mapping performed by time-resolved measurements. This analysis provided the foundation for time-resolved computational imaging. This chapter also introduced an algorithm for ideal sensor placement and an algorithm for generating ideal compressive patterns. Both algorithms demonstrated the interesting interplay

between algorithms and optical system design. With these algorithms we demonstrated that the FemtoPixel framework can result in acquisition times that are 50× faster compared to traditional single-pixel cameras.

Chapter 4 introduced data-driven computational imaging. We showed that with proper training data, the resulting algorithm can be invariant to perturbations in system calibration. Furthermore, the algorithm generalizes to perturbations that are outside of the training set (for example the illumination incident position is outside the range considered during training). With this approach, we trained a data-driven model on synthetic data, and only then tested it in a lab experiment. The model was able to perform without ever being calibrated to the imaging system. Another advantage of the technique is the ability to run in real-time (due to the non-iterative nature of the algorithm). These contributions were experimentally demonstrated by classifying the pose of a mannequin hidden by a sheet of paper, an example for sparse scattering.

Chapter 5 considered the challenging case of volumetric scattering in imaging through a tissue phantom. To tackle that challenge, we introduced the concept of All Photons Imaging — the idea that scattered light holds useful information about the hidden target, and that it shouldn't be rejected in the measurement process (which is what many other techniques try to accomplish). We showed that with time-resolved measurement, each measured frame is a realization of the target corrupted by a different blur kernel. Furthermore, we developed an algorithm to estimate these blur kernels from the measurement and independently of the target. This allows our approach to operate without a need for prior knowledge about the scattering (calibration free). We experimentally demonstrated API in imaging through a 1.5 cm tissue phantom and achieved a spatial resolution of 5.9 mm. Lastly, we showed that API is invariant to variations along the optical axis, such as the thickness of the medium. More specifically, we showed that a layered material has an equivalent uniform material. The layered and equivalent uniform materials share the same PSF and result in the same reconstruction capabilities.

Finally, Chapter 6 tackled the most challenging scattering problem. In this case

we experimentally demonstrate seeing through fog. The fog density varies all the way between clear conditions (no fog) to visibilities as low as 30 cm. Furthermore, in these conditions, the fog is heterogeneous and dynamic. Beyond the challenging scattering conditions, the demonstrated geometry is reflection mode. In that case, most of the measured photons back-reflect from the fog, and only a small portion actually hit the target before detection. To overcome these challenges, we introduced a probabilistic imaging framework that effectively separates between background and signal photons. This technique recovers the occluded scene reflectance and depth maps. For example, we demonstrated the recovery of objects that are 57 cm away from the sensor when the visibility was only 37 cm.

The techniques described here relaxed fundamental requirements in computational imaging for long acquisition times and highly calibrated models. Furthermore, the presented techniques are suitable for long range and wide-field sensing, making them suitable for real-world applications. To achieve these goals our solutions were inspired by different research fields including signal processing, machine learning, optimization, statistics, and computer vision. These algorithms were grounded by physics and optics, and enriched by time-resolved sensing. We believe that computational imaging through scattering with visible light will enable a wide range of novel applications in vehicles and medical imaging.

7.1 Future Outlook

Finally, we discuss our contributions with a broader perspective and potential future follow ups and applications:

- **A probabilistic interpretation of scattering and its use in computational imaging.** A probabilistic interpretation of scattering helped us derive the model in Chapter 5 and, it was an essential component of the algorithm in Chapter 6. We find such probabilistic interpretations useful for two main reasons: 1) Scattering is naturally modeled as a stochastic phenomenon, thus probabilistic modeling is often easier, and matches well to the physics of the

problem. 2) Probabilistic modeling opens the door to tools from probability theory, which again help the model, and recovery algorithms.

- **Single-photon-sensitive detectors for computational imaging.** We expect to see broader adoption of such techniques driven by the cost reduction of single-photon sensitive sensors, and recent performance improvements (increased spatial resolution, improved temporal resolution, better sensitivity). These devices expose the stochastic nature of light, making it even more appealing to use probabilistic modeling. We foresee several computational imaging applications that would benefit from single photon sensitivity: 1) Overcoming scattering — as discussed extensively in this dissertation. 2) Operating at extremely low photon flux. Because of the favorable noise model of SPADs (no signal-dependent noise) they are a better choice for imaging in challenging situations. We have already shown some of these benefits in the context of scattering [150]. 3) SPADs would also be useful outside the visible part of the spectrum, for example in x-ray imaging.

- **Data-driven computational imaging.** While data-driven techniques have revolutionized computer vision, there are still many open opportunities in computational imaging. Our work on calibration invariant imaging was one of the first demonstrations of data-driven computational imaging in the field, and the first to leverage it for calibration invariant imaging. There are many more problems that data-driven techniques can help to solve in computational imaging, as described next:

  - **Beyond classification**. In Chapter 4, we demonstrated a classification task that is considered easier than regression. However, classification can also be used for target localization (by discretizing the target volume). More novel architectures, such as an autoencoder, can be used for full scene reconstruction. In that case, the network is trained with image pairs, such that the input is the measurement from the camera point of view, and the output is a photo taken from a virtual point of view where the target

volume is visible. We have recently demonstrated these in [171]. Another promising architecture is based on generative models (e.g. [59]). In this case the network can be composed of two parts. The first is an encoder that encodes the underlying scene information from the measurement (this part is similar to the classification task we demonstrated). The second part is a decoder that would generate the full scene based on the recovered information (similar to a graphics rendering program). While each part has been demonstrated separately, the combination is challenging.

– **What does the network learn?** While this question is broadly considered open in the context of machine learning, we can try to answer parts of it in the context of computational imaging. More specifically, we can try to estimate the network sensitivity to different inputs. This can help in sensor selection, and scene optimization. We have recently taken the first steps in this direction in [172].

– **Data generation techniques.** In Chapter 4 we used an accurate and costly (in terms of run time) MC forward model. As a result, data generation was the bottleneck of our approach. New promising generative models for data generation (e.g. [92] and GANs) can alleviate this challenge. This can help accelerate the preparation of training data that is required when encountering a new scene.

– **Different types of scattering.** As discussed in the introduction, the problem of imaging through sparse scattering is equivalent to the problem of seeing around corners. Recently, we demonstrated data-driven imaging around corners [168] and beyond line of sight [172]. Data-driven approaches can also be used to alleviate problems of volumetric scattering. A direction we recently explored in the context of imaging through fog [152]. The main challenge in the case of volumetric scattering is the long rendering process which makes the problem extremely challenging when trying to address a wide range of scattering conditions, as done in Chapter 6.

– **Physically explainable data-driven computational imaging.** In Chapter 4 we introduced an end-to-end data-driven approach, where it is hard to introduce any physical or mathematical intuition to the solution. However, there are two more alternatives that leverage data-driven techniques which are more physically explainable. To better understand these approaches, we note that a classic inverse problem is usually written as:

$$\hat{x} = \arg\min_{x} \left\{ \|\mathbf{A}x - b\|_2^2 + R(x) \right\} \tag{7.1}$$

where $\mathbf{A}$ is the forward model, b is the measurement, $R(x)$ is some regularizer, and x is the target. When we consider this equation, we note that the end-to-end data-driven version is a mapping $b \rightarrow x$, such that the entire model is replaced with a data-driven algorithm. However, we can also replace just a part of the solution with a data-driven approach:

1. Learning the regularizer — instead of using some broad statistics such as sparsity, we can learn the target distribution from data. This can make the regularizer more specific to the considered problem. In that case, the solution can be decomposed into two steps: first solving for the measurement (also known as the data term) - $\|\mathbf{A}x - b\|_2^2$, and second, replacing the regularizer with a data-driven denoiser. These steps can be done iteratively until convergence. We note that the data-driven regularizer can be different for each step in the iteration (i.e. different models trained separately [42]). Here, we have the freedom to introduce physics into the construction of $\mathbf{A}$ (similar to the approach presented in Chapter 5).

2. Learning the forward model — this is similar to the approach suggested in Chapter 4, but instead the algorithm can specifically enforce a known regularizer. Another option is to split the forward model into multiple stages where each stage is learned with data, and the physics of the problem defines the connections between the different modules.

- **All Photons Imaging.** One of the key ideas presented in this thesis is the use of scattered light for imaging (while many other techniques aim to reject the scattered light and lock onto the ballistic component). Fusing computational imaging with time-resolved sensing of scattered light enabled us to image through scattering media at shorter acquisition times and with better quality. This was demonstrated in Chapters 5, 6.

  The All Photons Imaging concept can be extended to other domains and spectra. For example, in x-ray imaging it is common to reject scattered light in the measurement process. With All Photons Imaging it would be possible to leverage the scattered light to: 1) Reduce radiation dosage (since the scattered photons are used for reconstruction instead of rejected before the measurement). 2) Extract further clinically meaningful data, for example, the scattering properties of the tissue (current x-ray systems recover the absorption properties of the tissue). Similarly, this concept can be useful in other regimes such as THz, and RF.

- **Imaging through challenging scattering in optical reflection mode.** The technique to image through dense fog in Chapter 6 ignored the scattering of photons on the way back from the target to the camera. While we have not found this to be a significant issue in our optical setup, it may become a prominent effect in other cases. A potential solution to this limitation is to leverage the approach presented in Chapter 5. In that case we use the signal and background separation technique suggested in Chapter 6, followed by a descattering step as suggested in Chapter 5. It is worth noting that the background photons should provide useful information for the descattering step since both are a result of the same physical process. In fact, we can model the problem as follows:

$$m(x, y, t) = K(x, y, t; z) * [s(x, y, z) + \lambda] \tag{7.2}$$

  where $K(x, y, t; z)$ is a depth dependent scattering kernel, $s(x, y, z)$ is the oc-

cluded scene, and λ is the fog reflectivity. Note that $K(x, y, t; z) * s(x, y, z)$ is similar to the model in Chapter 5, and that the two terms in parentheses represent the signal and background term similarly to Chapter 6.

- **Time-resolved sensing for computational imaging through scattering.** In all chapters we showed, independently, that time-resolved computational imaging is superior to non-time aware techniques. Time-resolved sensing enabled the presented systems to operate at faster acquisition times, lower measurement SNR, and resulted in better reconstruction quality. We hope that TCSPC SPAD systems will become a commodity in the near future and allow for cost-effective time-resolved sensing solutions. The use of time-resolved sensing also opens the door for radically new imaging modalities as we recently demonstrated [70].

  As discussed in Sec. 6.9.5, when considering imaging through scattering with time-resolved measurement, the time response due to the scattering has some variance σ^2. This variance is a function of the scattering mean free path μ_s, and the number of scattering events before detection. To overcome the scattering, the detector must have a time resolution T such that $T \ll \sigma$. Alternatively (and potentially more demanding) is that $T \sim \mu_s$ (where we consider μ_s as the mean time between scattering events). As the time resolution improves due to hardware advances, we foresee new applications with visible light, such as medical imaging (where the mean free path is shorter compared to fog), or even reading through a closed book [132] with visible light.

- **Optimized compressive imaging**. Chapters 5 and 6 demonstrated imaging with flood illumination which is substantially harder compared to raster scanning. However, it is obvious that using structured light would provide even more information that will result in better recovery. The structured light optimization algorithm presented in Chapter 3 would be useful for that application. In that case, the scattering will act as the physical operator to be "conjugated" by the structured light. Furthermore, as demonstrated in Chapter 3 there is a

clear trade-off between structured light and time resolution, which can be leveraged for more complicated scattering conditions. This is a direction we started pursuing in the context of imaging through tissue [108].

This approach can also benefit other problems in computational imaging. For example, when the measurement is of the form $y = \mathbf{G}\mathbf{H}x$ where $\mathbf{H}$ is the imaging forward model and $\mathbf{G}$ is the compressive measurement. Then it is possible to optimize $\mathbf{G}$ to minimize the number of required compressive patterns by following a similar procedure to the one in Appendix B. There are many applications where this approach can be beneficial, for example MRI, CT, DOT, and also navigation and estimation problems such as SLAM.

- **Compressive imaging for reflectance and depth recovery.** The technique presented in Chapter 3 assumed a known geometry and recovered the target reflectance. Other techniques assumed known reflectance and recovered target depth. Since the measurement hardware in these cases is very similar, it would be possible to combine the two techniques. The algorithm can be based, for example, on iterating between depth to reflectance recovery. It would also be possible to optimize the compressive masks, as suggested above.

~

I believe that computational imaging will play a fundamental role in shaping the future of imaging through scattering media. This thesis has laid the foundations by demonstrating practical use cases of computational imaging through occlusions. These include calibration invariant algorithms for robust imaging systems; imaging algorithms based on all of the optical signal for better reconstruction quality and faster acquisition times; and probabilistic modeling that makes scattering problems easier to estimate without prior knowledge on the scattering media. I believe these will be essential with the growing demand for practical solutions to image through occlusions.

Appendix A

Photon Transport in Scattering Media as a Random Walk

Here we derive the diffusion equation from a random walk perspective which results in Brownian motion. For simplicity, the derivation is done in 1D and can be easily extended to higher dimensions.

Consider the probability density function $\Phi(x,t)$ to find a photon at position x and time t. The photon makes a step of length Δ with some probability density function $\mu(\Delta)$. We start by developing a Taylor expansion of the photon at time $t+\tau$ (after the step):

$$\Phi(x,t+\tau) = \Phi(x,t) + \tau\frac{\partial\Phi}{\partial t} + \cdots \tag{A.1}$$

We note that that $\Phi(x,t+\tau)$ is a result of the transition $\Phi(x-\Delta,t) \rightarrow \Phi(x,t+\tau)$, and we must average over all possible Δ:

$$\Phi(x,t+\tau) = \int_{-\infty}^{\infty} \Phi(x-\Delta,t)\mu(\Delta)d\Delta \tag{A.2}$$

We can develop a Taylor expansion of $\Phi(x-\Delta,t)$:

$$\Phi(x-\Delta,t) = \Phi(x,t) - \Delta\frac{\partial\Phi}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2\Phi}{\partial x^2} + \cdots \tag{A.3}$$

Such that:

$$\int_{-\infty}^{\infty} \Phi(x-\Delta,t)\mu(\Delta)d\Delta =$$

$$= \Phi(x,t)\int_{-\infty}^{\infty} \mu(\Delta)d\Delta - \frac{\partial\Phi}{\partial x}\int_{-\infty}^{\infty} \Delta\mu(\Delta)d\Delta + \frac{\partial^2\Phi}{\partial x^2}\int_{-\infty}^{\infty} \frac{\Delta^2}{2}\mu(\Delta)d\Delta + \cdots \quad \text{(A.4)}$$

The first integral on the right is equal to 1 by the definition of $\mu(\Delta)$, so:

$$\int_{-\infty}^{\infty} \Phi(x-\Delta,t)\mu(\Delta)d\Delta =$$

$$= \Phi(x,t) - \frac{\partial\Phi}{\partial x}\int_{-\infty}^{\infty} \Delta\mu(\Delta)d\Delta + \frac{\partial^2\Phi}{\partial x^2}\int_{-\infty}^{\infty} \frac{\Delta^2}{2}\mu(\Delta)d\Delta + \cdots \quad \text{(A.5)}$$

Combining Eqs A.1, A.2, A.5 we get:

$$\tau\frac{\partial\Phi}{\partial t} = -\frac{\partial\Phi}{\partial x}\int_{-\infty}^{\infty} \Delta\mu(\Delta)d\Delta + \frac{\partial^2\Phi}{\partial x^2}\int_{-\infty}^{\infty} \frac{\Delta^2}{2}\mu(\Delta)d\Delta \quad \text{(A.6)}$$

Next, we define:

$$v = \frac{1}{\tau}\int_{-\infty}^{\infty} \Delta\mu(\Delta)d\Delta \quad \text{(A.7)}$$

$$D = \frac{1}{2\tau}\int_{-\infty}^{\infty} \Delta^2\mu(\Delta)d\Delta \quad \text{(A.8)}$$

Such that:

$$\frac{\partial\Phi}{\partial t} + v\frac{\partial\Phi}{\partial x} = D\frac{\partial^2\Phi}{\partial x^2} \quad \text{(A.9)}$$

Which is the Brownian motion equation with drift.

The main assumptions here are:

1. Homogeneous and isotropic medium.

2. First and second moments of $\mu(\Delta)$ are finite (there is no explicit need to define

a specific distribution such as exponential etc.).

3. Removing second order terms in time, and third order terms in space.

A.1 Solving the Brownian Motion PDE

The derivation here follows [131]. To solve Eq. A.9 we assume infinite medium with a source at the origin. We start with Fourier transform of $\Phi(x,t)$:

$$\tilde{\Phi}(k,t) = \int \Phi(x,t) e^{jkx} dx \tag{A.10}$$

Which simplifies Eq. A.9 to:

$$\dot{\tilde{\Phi}}(k,t) = (jkv - Dk^2)\Phi(k,t) \tag{A.11}$$

With the simple solution:

$$\Phi(k,t) = \Phi(k,0) e^{\left(jkv - Dk^2\right)t} = e^{\left(jkv - Dk^2\right)t} \tag{A.12}$$

Taking the inverse Fourier transform results in:

$$\Phi(x,t) = \frac{1}{2\pi} \int e^{\left(jkv - Dk^2\right)t - jkx} dk \tag{A.13}$$

with the solution:

$$\Phi(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left\{ -\frac{(x - vt)^2}{4Dt} \right\} \tag{A.14}$$

This is the well known Gaussian distribution with a time dependent variance as in Eq. 2.10. Solving the Brownian motion for higher dimensions is similar and results in different normalization factors.

Appendix B

Derivation of Illumination Patterns Optimization Algorithm for FemtoPixel

Here we provide a derivation for efficiently calculating the cost function in Eq. 3.15 and its gradient. We focus on the data term since the derivatives of the suggested regularizes are straightforward.

The data term of the cost function to minimize:

$$\gamma = \left\| \mathbf{I}_L - \tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}} \right\|_F^2 \tag{B.1}$$

We define $\mathbf{\Lambda}$ such that its j-th row is $\mathbf{g}_j^T$ ($\mathbf{\Lambda}$ is an $M \times L$ matrix). Our goal is to find $\mathbf{\Lambda}$ which minimizes γ.

First, we define $\mathbf{O} = \tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}}$ such that:

$$\begin{aligned} \gamma &= \|\mathbf{I}_L - \mathbf{O}\|_F^2 = Tr\left\{ (\mathbf{I}_L - \mathbf{O}) (\mathbf{I}_L - \mathbf{O})^T \right\} \\ &= Tr\left\{\mathbf{I}_L\right\} - 2Tr\left\{\mathbf{O}\right\} + Tr\left\{\mathbf{O}\mathbf{O}^T\right\} = \|\mathbf{O}\|_F^2 - L \end{aligned} \tag{B.2}$$

since $Tr\left\{\mathbf{O}\right\} = L$. Next, we define $\tilde{\mathbf{Q}} = \mathbf{Q}\mathbf{P}$ where $\mathbf{P}$ is a diagonal matrix with the

inverse of the columns norm:

$$\mathbf{P}_{n,n} = \frac{1}{\sqrt{\sum_{a=1}^{MKN} \mathbf{Q}_{a,n}{}^2}} = \frac{1}{\sqrt{\left(\sum_{j=1}^{M} \mathbf{\Lambda}_{j,n}{}^2\right)\left(\sum_{i=1}^{KN} \bar{\mathbf{H}}_{i,n}^2\right)}} \tag{B.3}$$

This allows us to write:

$$\mathbf{O} = \tilde{\mathbf{Q}}^T\tilde{\mathbf{Q}} = \mathbf{P}^T\mathbf{Q}^T\mathbf{Q}\mathbf{P} = \mathbf{P}\sum_{j=1}^{M}\left(\mathbf{D}_{\mathbf{\Lambda}_j}\bar{\mathbf{H}}^T\bar{\mathbf{H}}\mathbf{D}_{\mathbf{\Lambda}_j}\right)\mathbf{P} \tag{B.4}$$

where $\mathbf{D}_{\mathbf{\Lambda}_j}$ is a diagonal matrix with the j-th row of $\mathbf{\Lambda}$ on the diagonal. Next we note that:

$$\left[\bar{\mathbf{H}}\mathbf{D}_{\mathbf{\Lambda_j}}\right]_{n,m} = \sum_{a=1}^{L}\bar{\mathbf{H}}_{n,a}\left[\mathbf{D}_{\mathbf{\Lambda_j}}\right]_{a,m} = \bar{\mathbf{H}}_{n,m}\mathbf{\Lambda}_{j,m} \tag{B.5}$$

and:

$$\mathbf{O}_{n,m} = \mathbf{P}_{n,n}\sum_{j=1}^{M}\left(\mathbf{\Lambda}_{j,n}\left[\bar{\mathbf{H}}^{\mathbf{T}}\bar{\mathbf{H}}\right]_{n,m}\mathbf{\Lambda}_{j,m}\right)\mathbf{P}_{m,m} \tag{B.6}$$

which can be simplified to:

$$\mathbf{O}_{n,m} = \mathbf{W}_{n,m}\frac{\sum_{j=1}^{M}\mathbf{\Lambda}_{j,n}\mathbf{\Lambda}_{j,m}}{\sqrt{\left(\sum_{j=1}^{M}\mathbf{\Lambda}_{j,n}{}^2\right)\left(\sum_{j=1}^{M}\mathbf{\Lambda}_{j,m}{}^2\right)}} \tag{B.7}$$

with

$$\mathbf{W}_{n,m} = \frac{\left(\bar{\mathbf{H}}^T\bar{\mathbf{H}}\right)_{n,m}}{\sqrt{\left(\sum_{i=1}^{K}\mathbf{H}_{i,n}{}^2\right)\left(\sum_{i=1}^{K}\mathbf{H}_{i,m}{}^2\right)}} \tag{B.8}$$

Lastly, we define:

$$\mathbf{\Phi} = \mathbf{\Lambda}^T\mathbf{\Lambda} \tag{B.9}$$

such that:

$$\mathbf{O}_{n,m} = \mathbf{W}_{n,m}\frac{\mathbf{\Phi}_{n,m}}{\sqrt{\mathbf{\Phi}_{n,n}\mathbf{\Phi}_{m,m}}} \tag{B.10}$$

which allows us to write:

$$\mathbf{O} = \left(\sqrt{\mathbf{S}_{\mathbf{\Phi}}}\right)^{-1} (\mathbf{W} \odot \mathbf{\Phi}) \left(\sqrt{\mathbf{S}_{\mathbf{\Phi}}}\right)^{-1} \tag{B.11}$$

where $\mathbf{S}_{\mathbf{\Phi}}$ is a diagonal matrix with the diagonal entries of $\mathbf{\Phi}$, and $\odot$ denotes element wise multiplication. Finally:

$$\gamma = \left\| \left(\sqrt{\mathbf{S}_{\mathbf{\Phi}}}\right)^{-1} (\mathbf{W} \odot \mathbf{\Phi}) \left(\sqrt{\mathbf{S}_{\mathbf{\Phi}}}\right)^{-1} \right\|_F^2 - L \tag{B.12}$$

We note that $\mathbf{W}$ (Eq. B.8) is a constant matrix for the illumination pattern optimization and can be calculated a priori. Eq. B.9 and B.12 provide the final expression for $\gamma(\mathbf{\Lambda})$.

We now develop an expression for the gradient of the cost function. Considering the chain rule for matrices [122]:

$$\left[\frac{d\gamma}{d\mathbf{\Lambda}}\right]_{n,m} = \sum_{a=1}^{L}\sum_{b=1}^{L} \frac{d\gamma(\mathbf{\Phi})}{d\mathbf{\Phi}_{a,b}} \frac{d\mathbf{\Phi}_{a,b}}{d\mathbf{\Lambda}_{n,m}} \tag{B.13}$$

Starting with the first term, if $a \neq b$:

$$\frac{d\gamma}{d\mathbf{\Phi}_{a,b}} = {\mathbf{W}_{a,b}}^2 \frac{2\mathbf{\Phi}_{a,b}}{\mathbf{\Phi}_{a,a}\mathbf{\Phi}_{b,b}} \tag{B.14}$$

and, if $a = b$:

$$\frac{d\gamma}{d\mathbf{\Phi}_{a,a}} = -\sum_{c=1}^{n} \frac{{\mathbf{W}_{a,c}}^2 {\mathbf{\Phi}_{a,c}}^2 + {\mathbf{W}_{c,a}}^2 {\mathbf{\Phi}_{c,a}}^2}{{\mathbf{\Phi}_{a,a}}^2 \mathbf{\Phi}_{c,c}} \delta_{c\neq a} \tag{B.15}$$

where δ is Kronecker delta. The second term in Eq. B.13 is given by [122]:

$$\frac{d\mathbf{\Phi}_{a,b}}{d\mathbf{\Lambda}_{n,m}} = \delta_{mb}\mathbf{\Lambda}_{n,a} + \delta_{ma}\mathbf{\Lambda}_{n,b} \tag{B.16}$$

Combining Eqs. B.13 through B.16 we get:

$$\begin{aligned}\left[\frac{d\gamma}{d\mathbf{\Lambda}}\right]_{n,m} &= \sum_{a=1}^{L}\sum_{b=1}^{L} E^1\delta_{ab}\left(\delta_{mb}\mathbf{\Lambda}_{n,a} + \delta_{ma}\mathbf{\Lambda}_{n,b}\right) \\ &+ \sum_{a=1}^{L}\sum_{b=1}^{L} E^2\delta_{a\neq b}\left(\delta_{mb}\mathbf{\Lambda}_{n,a} + \delta_{ma}\mathbf{\Lambda}_{n,b}\right)\end{aligned} \tag{B.17}$$

where E^1 and E^2 are the terms in Eqs. B.14 and B.15 respectively. After some algebra we get:

$$\begin{aligned}\left[\frac{d\gamma}{d\mathbf{\Lambda}}\right]_{nm} = \frac{2}{\mathbf{\Phi}_{mm}}\sum_{\substack{a=1\\a\neq m}}^{L}\Bigg[&\frac{\mathbf{\Lambda}_{na}}{\mathbf{\Phi}_{aa}}\left(\mathbf{W}_{am}{}^2\mathbf{\Phi}_{am} + \mathbf{W}_{ma}{}^2\mathbf{\Phi}_{ma}\right) - \\ &\frac{\mathbf{\Lambda}_{nm}}{\mathbf{\Phi}_{aa}\mathbf{\Phi}_{mm}}\left(\mathbf{W}_{am}{}^2\mathbf{\Phi}_{am}{}^2 + \mathbf{W}_{ma}{}^2\mathbf{\Phi}_{ma}{}^2\right)\Bigg]\end{aligned} \tag{B.18}$$

Lastly, we define:

$$\begin{aligned}\alpha_1 &= \mathbf{W}\odot\mathbf{W}\odot\mathbf{\Phi}\odot\mathbf{I}^- \\ \alpha_2 &= \mathbf{W}\odot\mathbf{W}\odot\mathbf{\Phi}\odot\mathbf{\Phi}\odot\mathbf{I}^-\end{aligned} \tag{B.19}$$

where, $\mathbf{I}^- = 1 - \mathbf{I}_L$ (matrix with all ones except for zeros on the diagonal), which allows us to write the final gradient as:

$$\begin{aligned}\mathbf{c}_1 &= \Lambda\mathbf{S}_{\mathbf{\Phi}}{}^{-1}\left(\alpha_1 + \alpha_1{}^T\right) \\ \mathbf{c}_2 &= \left(\mathbf{\Lambda}\odot\left(\mathbf{1}_{M\times L}\mathbf{S}_{\mathbf{\Phi}}{}^{-1}\left(\alpha_2 + \alpha_2{}^T\right)\mathbf{S}_{\mathbf{\Phi}}{}^{-1}\right)\right) \\ \frac{d\gamma}{d\mathbf{\Lambda}} &= 2\left(\mathbf{c}_1 - \mathbf{c}_2\right)\mathbf{S}_{\mathbf{\Phi}}{}^{-1}\end{aligned} \tag{B.20}$$

where $\mathbf{1}_{M\times L}$ is an $M \times L$ matrix with all ones.

Appendix C

Expectation Maximization Algorithm for Imaging Through Fog

Here we provide the derivation for the EM update rules used in the seeing through fog chapter 6.7.

We start with the log-likelihood function

$$\ell(\theta) = \sum_{i=1}^{m} \log P\left(x^i; \theta\right) \tag{C.1}$$

Here, $\theta = \{\mu, \sigma^2, K, \mu_s, P_S, P_B\}$ are the latent variables describing the model.

We consider $i = 1..m$ data points, such that x^i is the measured time of arrival of the i-th photon. The assignment of each photon to the background class (B) or signal class (S) is defined by $z^i \in \{B, S\}$, and we denote $j = B, S$. Which helps to define the membership probability:

$$Q^i\left(z^i\right) = P\left(Z^i | x^i; \theta\right) \tag{C.2}$$

For completeness, the distributions used here are:

$$P\left(x^i | z^i = S; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(x^i - \mu\right)^2}{2\sigma^2}\right\} \tag{C.3}$$

$$P\left(x^i|z^i = B; K, \mu_s\right) = \frac{1}{\Gamma(K){\mu_s}^K}\left(x^i\right)^{K-1}\exp\left\{-\frac{x^i}{\mu_s}\right\} \tag{C.4}$$

And, the prior probabilities are:

$$\begin{aligned} P_S &= P\left(z^i = S; \mu, \sigma^2\right) \\ P_B &= P\left(z^i = B; K, \mu_s\right) \end{aligned} \tag{C.5}$$

The EM algorithm is composed of two steps: Expectation and Maximization that are iterated.

C.1 Expectation Step

Our goal in the Expectation step (E step for short) is to estimate the probability that the $i-th$ sample belongs to the two potential probabilities: w_B^i for the Gamma distribution and w_S^i for the signal distribution.

$$w_j^i = Q\left(z^i = j\right) = P\left(z^i = j|\theta^j\right) \tag{C.6}$$

Here, $\theta^B = (K, \mu_s)$ and $\theta^S = (\mu, \sigma^2)$. The probability that the i-th photon belongs to the signal class is:

$$\begin{aligned} w_S^i &= P\left(z^i = S|x^i; P_S, P_B, \mu, \sigma^2, K, \mu_s\right) = \\ &= \frac{P\left(x^i|z^i = S; \mu, \sigma^2\right) P\left(z^i = S; \mu, \sigma^2\right)}{P\left(x^i|z^i = S; \mu, \sigma^2\right) P\left(z^i = S; \mu, \sigma^2\right) + P\left(x^i|z^i = B; K, \mu_s\right) P\left(z^i = B; K, \mu_s\right)} \end{aligned} \tag{C.7}$$

And, the probability that the i-th photon belongs to the background class is:

$$\begin{aligned} w_B^i &= P\left(z^i = B|x^i; P_S, P_B, \mu, \sigma^2, K, \mu_s\right) = \\ &= \frac{P\left(x^i|z^i = B; K, \mu_s\right) P\left(z^i = B; K, \mu_s\right)}{P\left(x^i|z^i = S; \mu, \sigma^2\right) P\left(z^i = S; \mu, \sigma^2\right) + P\left(x^i|z^i = B; K, \mu_s\right) P\left(z^i = B; K, \mu_s\right)} \end{aligned} \tag{C.8}$$

Equations C.7, C.8 are completely defined by Eqs. C.3, C.4, C.5.

C.2 Maximization Step

In the maximization step (M step for short) we use the membership probabilities from the E step as weights to update the parameters of the model. The log-likelihood function for the latent variables is:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{m} \sum_{z^i=S,B} Q\left(z^i\right) \log \frac{P\left(x^i, z^i|\theta\right)}{Q\left(z^i\right)} \tag{C.9}$$

Using the Bayes rule we get:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{m} \sum_{z^i=S,B} Q\left(z^i\right) \log \frac{P\left(x^i|z^i=j,\theta\right) P\left(z^i=j|\theta\right)}{Q\left(z^i\right)} \tag{C.10}$$

Which simplifies to:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{m} {w_S}^i \log \frac{P\left(x^i|z^i=S;\mu,\sigma^2\right) P_S}{{w_S}^i} + {w_B}^i \log \frac{P\left(x^i|z^i=B;K,\mu_s\right) P_B}{{w_B}^i} \tag{C.11}$$

Estimating μ

Taking the derivative of the argument in C.11 with respect to μ we get:

$$\frac{\partial}{\partial\mu}\left(\cdot\right) = \frac{\partial}{\partial\mu} \sum_{i=1}^{m} -{w_S}^i \frac{\left(x^i-\mu\right)^2}{2\sigma^2} = \sum_{i=1}^{m} {w_S}^i \frac{x^i-\mu}{\sigma^2} \tag{C.12}$$

And, setting it to zero we get:

$$\mu = \frac{\sum_{i=1}^{m} w_S^i x^i}{\sum_{i=1}^{m} w_S^i} \tag{C.13}$$

Estimating σ^2

Taking the derivative of the argument in C.11 with respect to σ^2 we get:

$$\frac{\partial}{\partial\sigma^2}\left(\cdot\right) = \frac{\partial}{\partial\sigma^2} \sum_{i=1}^{m} -\frac{1}{2}{w_S}^i \log\sigma^2 - w_S^i \frac{\left(x^i-\mu\right)^2}{2\sigma^2} = \sum_{i=1}^{m} {w_S}^i \frac{\left(x^i-\mu\right)^2}{2\sigma^2} - \frac{1}{2} w_S^i \tag{C.14}$$

And, setting it to zero we get:

$$\sigma^2 = \frac{\sum_{i=1}^{m} w_S^i (x^i - \mu)^2}{\sum_{i=1}^{m} w_S^i} \tag{C.15}$$

Estimating μ_s

First, to recover μ_s we take derivative of the argument in C.11 with respect to μ_s:

$$\frac{\partial}{\partial \mu_s}(\cdot) = \frac{\partial}{\partial \mu_s} \sum_{i=1}^{m} -w_B^i \log \mu_s - w_B^i \frac{x^i}{\mu_s} = -w_B^i K \frac{1}{\mu_s} + w_B^i \frac{x^i}{{\mu_s}^2} \tag{C.16}$$

Setting it to zero we get:

$$\mu_s = \frac{\sum_{i=1}^{m} w_B^i x^i}{K \sum_{i=1}^{m} w_B^i} \tag{C.17}$$

Which we update with the estimated K as described next.

Estimating K

Here, we derive a maximum likelihood estimator inspired by [114]. Taking the derivative of the argument in C.11 with respect to K:

$$\frac{\partial}{\partial K}(\cdot) = \frac{\partial}{\partial K} \sum_{i=1}^{m} -w_B^i K \log \mu_s - w_B^i \log \Gamma(K) + w_B^i (K-1) \log x^i - \frac{w_B^i x^i}{\mu_s} \tag{C.18}$$

Using Eqs. C.17, C.18 and combining terms we get:

$$\begin{aligned}
&\tfrac{\partial}{\partial K}(\cdot) = \\
&= \tfrac{\partial}{\partial K} \sum_{i=1}^{m} \left\{ -\log \Gamma(K) + K \left[\frac{\sum_{i=1}^{m} w_B^i \log x^i}{\sum_{i=1}^{m} w_B^i} - \log \left(\sum_{i=1}^{m} w_B^i \log x^i \right) - 1 \right] + K \log \left(K \sum_{i=1}^{m} w_B^i \right) \right\} \\
&= \frac{\sum_{i=1}^{m} w_B^i \log x^i}{\sum_{i=1}^{m} w_B^i} - \log \left(\sum_{i=1}^{m} w_B^i \log x^i \right) - \Psi(K) + \log \left(K \sum_{i=1}^{m} w_B^i \right)
\end{aligned} \tag{C.19}$$

here, $\Psi(K)$ is the digamma function. The maximum likelihood estimator developed in [113, 114] requires the second derivative as well:

$$\frac{\partial^2}{\partial K^2}(\cdot) = \frac{1}{K} - \Psi'(K) \tag{C.20}$$

Finally, the iterative solution to recover K is based on [113]:

$$\frac{1}{K^{new}} = \frac{1}{K} + \frac{1}{K^2}\frac{\frac{\partial}{\partial K}(\cdot)}{\frac{\partial^2}{\partial K^2}(\cdot)} \tag{C.21}$$

Which is:

$$\frac{1}{K^{new}} = \frac{1}{K} + \frac{1}{K^2}\frac{\frac{\sum_{i=1}^{m} w_B^i \log x^i}{\sum_{i=1}^{m} w_B^i} - \log\left(\sum_{i=1}^{m} w_B^i \log x^i\right) - \Psi(K) + \log\left(K\sum_{i=1}^{m} w_B^i\right)}{\frac{1}{K} - \Psi'(K)} \tag{C.22}$$

In our solution we iterate over Eq. C.22 5 times. Similarly to [114] we initialize the iterations with:

$$K^0 = \frac{0.5}{\log\left(\sum_{i=1}^{m} w_B^i \log x^i\right) - \frac{\sum_{i=1}^{m} w_B^i \log x^i}{\sum_{i=1}^{m} w_B^i}} \tag{C.23}$$

Estimating P_S, P_B

To recover the prior probabilities we take the derivative of the argument in C.11 with respect to P_S and P_B to get:

$$\begin{aligned}\frac{\partial}{\partial P_S}(\cdot) &= \frac{\partial}{\partial P_S}\sum_{i=1}^{m} w_S^i \log P_S = \frac{1}{P_S}\sum_{i=1}^{m} w_S^i \\ \frac{\partial}{\partial P_B}(\cdot) &= \frac{\partial}{\partial P_B}\sum_{i=1}^{m} w_B^i \log P_B = \frac{1}{P_B}\sum_{i=1}^{m} w_B^i\end{aligned} \tag{C.24}$$

To impose the constraint $P_S + P_B = 1$ we add a Lagrangian term (with a β coefficient):

$$\mathcal{L}(P_S, P_B) = \sum_{i=1}^{m} w_S^i \log P_S + w_B^i \log P_B + \beta\left(P_S + P_B - 1\right) \tag{C.25}$$

This adds a β term to Eqs. C.24. Setting the updated Eqs. C.24 to 0 we get:

$$\begin{aligned} P_S &= \tfrac{-1}{\beta} \sum_{i=1}^{m} w_S^i \\ P_B &= \tfrac{-1}{\beta} \sum_{i=1}^{m} w_B^i \end{aligned} \tag{C.26}$$

Enforcing $r + b = 1$ results in $\beta = -m$ such that:

$$\begin{aligned} P_S &= \tfrac{1}{m} \sum_{i=1}^{m} w_S^i \\ P_B &= \tfrac{1}{m} \sum_{i=1}^{m} w_B^i \end{aligned} \tag{C.27}$$

C.3 Summary

- The Expectation step is defined by Eqs. C.7, C.8.
- The Maximization step is defined by Eqs. C.13, C.15, C.17, C.27, and the iterative algorithm in Eqs. C.22, C.23.

Bibliography

[1] Atefeh Abdolmanafi, Luc Duong, Nagib Dahdah, and Farida Cheriet. Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. *Biomedical Optics Express*, 8, 2017.

[2] Edward H Adelson and James R Bergen. The plenoptic function and the elements of early vision. *Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology*, 1991.

[3] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34, 2015.

[4] Fadel Adib and Dina Katabi. See through walls with wifi! *ACM SIGCOMM*, 2013.

[5] Takamasa Ando, Ryoichi Horisaki, and Jun Tanida. Speckle-learning-based object recognition through scattering media. *Optics express*, 23, 2015.

[6] Fabio Anselmi, Joel Z Leibo, Lorenzo Rosasco, Jim Mutch, Andrea Tacchetti, and Tomaso Poggio. Unsupervised learning of invariant representations in hierarchical architectures. *arXiv preprint arXiv:1311.4158*, 2013.

[7] Roger Appleby and Rupert N Anderton. Millimeter-wave and submillimeter-wave imaging for security and surveillance. *Proceedings of the IEEE*, 95, 2007.

[8] Yasuo Arai and Masahiro Ikeno. A time digitizer cmos gate-array with a 250 ps time resolution. *IEEE Journal of Solid-State Circuits*, 31, 1996.

[9] Isaac August, Yaniv Oiknine, Marwan AbuLeil, Ibrahim Abdulhalim, and Adrian Stern. Miniature compressive ultra-spectral imaging system utilizing a single liquid crystal phase retarder. *Nature Scientific Reports*, 6, 2016.

[10] Yuval Bahat and Michal Irani. Blind dehazing using internal patch recurrence. In *IEEE International Conference on Computational Photography (ICCP)*, 2016.

[11] Sohail Bahmani and Justin Romberg. Compressive deconvolution in random mask imaging. *IEEE Transactions on Computational Imaging*, 1, 2015.

[12] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28, 2008.

[13] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2, 2009.

[14] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 2013.

[15] Dana Berman and Shai Avidan. Non-local image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[16] Jacopo Bertolotti, Elbert G Van Putten, Christian Blum, Ad Lagendijk, Willem L Vos, and Allard P Mosk. Non-invasive imaging through opaque scattering layers. *Nature*, 491, 2012.

[17] Ayush Bhandari and Ramesh Raskar. Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-d imaging. *IEEE Signal Processing Magazine*, 33, 2016.

[18] David A Boas, Dana H Brooks, Eric L Miller, Charles A DiMarzio, Misha Kilmer, Richard J Gaudette, and Quan Zhang. Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine*, 18, 2001.

[19] Bryan T Bosworth, Jasper R Stroud, Dung N Tran, Trac D Tran, Sang Chin, and Mark A Foster. High-speed compressed sensing measurement using spectrally-encoded ultrafast laser pulses. In *IEEE Information Sciences and Systems (CISS)*, 2015.

[20] David J. Brady, Kerkil Choi, Daniel L. Marks, Ryoichi Horisaki, and Sehoon Lim. Compressive holography. *Optics Express*, 17, 2009.

[21] Pratik Prabhanjan Brahma, Dapeng Wu, and Yiyuan She. Why deep learning works: A manifold disentanglement perspective. *IEEE Transactions on Neural Networks and Learning Systems*, 27, 2016.

[22] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics Express*, 23, 2015.

[23] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25, 2016.

[24] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346, 2008.

[25] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, 52, 2006.

[26] Ying Cao, Wouter De Cock, Michiel Steyaert, and Paul Leroux. Design and assessment of a 6 ps-resolution time-to-digital converter with 5 mgy gamma-dose tolerance for lidar application. *IEEE Transactions on Nuclear Science*, 59, 2012.

[27] J Carlsson, P Hellentin, L Malmqvist, Anders Persson, W Persson, and Claes-Göran Wahlström. Time-resolved studies of light propagation in paper. *Applied Optics*, 34, 1995.

[28] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical Foundations and Numerical Methods For Sparse Recovery*, 9, 2010.

[29] Gregory Charvat, Andrew Temme, Micha Feigin, and Ramesh Raskar. Time-of-flight microwave camera. *Nature Scientific Reports*, 5, 2015.

[30] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[31] Mingmin Chi, Antonio Plaza, Jón Atli Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104, 2016.

[32] Regine Choe, Soren D Konecky, Alper Corlu, Kijoon Lee, Turgut Durduran, David R Busch, Saurav Pathak, Brian J Czerniecki, Julia C Tchou, Douglas L Fraker, et al. Differentiation of benign and malignant breast tumors by in-vivo three-dimensional parallel-plate diffuse optical tomography. *Journal of Biomedical Optics*, 14, 2009.

[33] Andrea Colaço, Ahmed Kirmani, Gregory A Howland, John C Howell, and Vivek K Goyal. Compressive depth map acquisition using a single photon-counting detector: parametric signal processing meets sparsity. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2012.

[34] Joshua Colombi and Karim Louedec. Monte carlo simulation of light scattering in the atmosphere and effect of atmospheric aerosols on the point spread function. *Journal Optical Society America A*, 30, 2013.

[35] Alper Corlu, Regine Choe, Turgut Durduran, Mark A Rosen, Martin Schweiger, Simon R Arridge, Mitchell D Schnall, and Arjun G Yodh. Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans. *Optics Express*, 15, 2007.

[36] J. P. Culver, R. Choe, M. J. Holboke, L. Zubkov, T. Durduran, A. Slemp, V. Ntziachristos, B. Chance, and A. G. Yodh. Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: Evaluation of a hybrid frequency domain/continuous wave clinical system for breast imaging. *Medical Physics*, 30, 2003.

[37] Joseph P Culver, Andrew M Siegel, Jonathan J Stott, and David A Boas. Volumetric diffuse optical tomography of brain activity. *Optics Letters*, 28, 2003.

[38] B B Das, Feng Liu, and R R Alfano. Time-resolved fluorescence and photon migration studies in biomedical and model random media. *Reports on Progress in Physics*, 60, 1999.

[39] Ofer David, Norman S. Kopeika, and Boaz Weizer. Range gated active night vision system for automobiles. *Applied Optics*, 45, 2006.

[40] Hamid Dehghani, Subhadra Srinivasan, Brian W Pogue, and Adam Gibson. Numerical modelling and image reconstruction in diffuse optical tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367, 2009.

[41] Winfried Denk, James H Strickler, and Watt W Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248, 1990.

[42] Steven Diamond, Vincent Sitzmann, Felix Heide, and Gordon Wetzstein. Unrolled optimization with deep priors. *arXiv preprint arXiv:1705.08041*, 2017.

[43] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 2014.

[44] David L Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52, 2006.

[45] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25, 2008.

[46] Julio Martin Duarte-Carvajalino and Guillermo Sapiro. Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 18, 2009.

[47] Piotr Dudek, Stanislaw Szczepanski, and John V Hatfield. A high-resolution cmos time-to-digital converter utilizing a vernier delay line. *IEEE Journal of Solid-State Circuits*, 35, 2000.

[48] T. Durduran, A. G. Yodh, B. Chance, and D. A. Boas. Does the photon-diffusion coefficient depend on absorption? *Journal of the Optical Society of America A*, 14, 1997.

[49] D. J. Durian. The diffusion coefficient depends on absorption. *Optics Letters*, 23, 1998.

[50] Michael Elad. Optimized projections for compressed sensing. *IEEE Transactions on Signal Processing*, 55, 2007.

[51] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V. Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516, 2014.

[52] Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. Single-photon sensitive light-in-fight imaging. *Nature Communications*, 6, 2015.

[53] Genevieve Gariepy, Francesco Tonolini, Robert Henderson, Jonathan Leach, and Daniele Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10, 2015.

[54] Graham M Gibson, Baoqing Sun, Matthew P Edgar, David B Phillips, Nils Hempler, Gareth T Maker, Graeme PA Malcolm, and Miles J Padgett. Real-time imaging of methane gas leaks using a single-pixel camera. *Optics Express*, 25, 2017.

[55] Ioannis Gkioulekas, Anat Levin, and Todd Zickler. An evaluation of computational imaging techniques for heterogeneous inverse scattering. In *European Conference on Computer Vision (ECCV)*, 2016.

[56] Ioannis Gkioulekas, Shuang Zhao, Kavita Bala, Todd Zickler, and Anat Levin. Inverse volume rendering with material dictionaries. *ACM ToG*, 32, 2013.

[57] Maoguo Gong, Jiaojiao Zhao, Jia Liu, Qiguang Miao, and Licheng Jiao. Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 27, 2016.

[58] Ian Goodfellow, Honglak Lee, Quoc V. Le, Andrew Saxe, and Andrew Y. Ng. Measuring invariances in deep networks. *Advances in Neural Information Processing Systems 22*, 2009.

[59] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2014.

[60] J. W. Goodman, W. H. Huntley Jr, D. W. Jackson, and M Lehmann. Wavefront-reconstruction imaging through random media. *Applied Physics Letters*, 8, 1966.

[61] Martin Grabner and Vaclav Kvicera. Multiple scattering in rain and fog on free-space optical links. *Journal of Lightwave Technology*, 32, 2014.

[62] Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM ToG*, 34, 2015.

[63] Otkrist Gupta, Thomas Willwacher, Andreas Velten, Ashok Veeraraghavan, and Ramesh Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics Express*, 20, 2012.

[64] Hamamatsu. Guide to streak cameras. Accessed: 2019-03-09.

[65] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[66] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2014.

[67] Felix Heide, Lei Xiao, Andreas Kolb, Matthias B Hullin, and Wolfgang Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics Express*, 22, 2014.

[68] Fritjof Helmchen and Winfried Denk. Deep tissue two-photon microscopy. *Nature Methods*, 2, 2005.

[69] Barmak Heshmat, Guy Satat, Christopher Barsi, and Ramesh Raskar. Single-shot ultrafast imaging using parallax-free alignment with a tilted lenslet array. In *CLEO: Science and Innovations*, 2014.

[70] Barmak Heshmat, Matthew Tancik, Guy Satat, and Ramesh Raskar. Photography optics in the time dimension. *Nature Photonics*, 12, 2018.

[71] Vadim Holodovsky, Yoav Y Schechner, Anat Levin, Aviad Levis, and Amit Aides. In-situ multi-view multi-scattering stochastic tomography. In *IEEE International Conference on Computational Photography (ICCP)*, 2016.

[72] Sung-Min Hong. *A study of time amplifiers and time-to-voltage converters for data conversion applications*. PhD thesis, McGill University, 2007.

[73] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. Learning-based imaging through scattering media. *Optics Express*, 24, 2016.

[74] Roarke Horstmeyer, Haowen Ruan, and Changhuei Yang. Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue. *Nature Photonics*, 9, 2015.

[75] Steven S Hou, William L Rice, Brian J Bacskai, and Anand TN Kumar. Tomographic lifetime imaging using combined early-and late-arriving photons. *Optics Letters*, 39, 2014.

[76] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, and James G Fujimoto. Optical coherence tomography. *Science*, 254, 1991.

[77] Chenfei Jin, Zitong Song, Siqi Zhang, Jianhua Zhai, and Yuan Zhao. Recovering three-dimensional shape through a small hole using three laser scatterings. *Optics Letters*, 40, 2015.

[78] Achuta Kadambi, Jamie Schiel, and Ramesh Raskar. Macroscopic interferometry: Rethinking depth estimation with frequency-domain time-of-flight. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2016.

[79] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM ToG*, 32, 2013.

[80] Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics*, 35, 2016.

[81] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[82] Vladimir Katkovnik and Jaakko Astola. Compressive sensing computational ghost imaging. *Journal Optical Society America A*, 29, 2012.

[83] Ori Katz, Yaron Bromberg, and Yaron Silberberg. Compressive ghost imaging. *Applied Physics. Letters*, 95, 2009.

[84] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature Photonics*, 8, 2014.

[85] Ori Katz, Eran Small, Yefeng Guan, and Yaron Silberberg. Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers. *Optica*, 1, 2014.

[86] Ori Katz, Eran Small, and Yaron Silberberg. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature Photonics*, 6, 2012.

[87] Ahmed Kirmani, Andrea Colaço, Franco N C Wong, and Vivek K Goyal. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Optics Express*, 19, 2011.

[88] Ahmed Kirmani, Haris Jeelani, Vahid Montazerhodjat, and Vivek K. Goyal. Diffuse imaging: creating optical images with unfocused time-resolved illumination and sensing. *IEEE Signal Processing Letters*, 19, 2012.

[89] Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K Goyal. First-photon imaging. *Science*, 343, 2014.

[90] Stefan P Koch, Christina Habermehl, Jan Mehnert, Christoph H Schmitz, Susanne Holtze, Arno Villringer, Jens Steinbrink, and Hellmuth Obrig. High-resolution optical functional mapping of the human somatosensory cortex. *Frontiers in Neuroenergetics*, 2, 2010.

[91] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[92] Tejas D Kulkarni, William F Whitney, Pushmeet Kohli, and Josh Tenenbaum. Deep convolutional inverse graphics network. *Advances in Neural Information Processing Systems*, 2015.

[93] Anand TN Kumar, Scott B Raymond, Andrew K Dunn, Brian J Bacskai, and David A Boas. A time domain fluorescence tomography system for small animal imaging. *IEEE Transactions on Medical Imaging*, 27, 2008.

[94] Sriram Kumar and Andreas Savakis. Robust domain adaptation on the l1-grassmannian manifold. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2016.

[95] Kiriakos N. Kutulakos and James R. Vallino. Calibration-free augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 4, 1998.

[96] Eric Lapointe, Julien Pichette, and Yves Berube-Lauziere. A multi-view time-domain non-contact diffuse optical tomography scanner with dual wavelength detection for intrinsic and fluorescence small animal imaging. *Review of Scientific Instruments*, 83, 2012.

[97] Martin Laurenzis, Frank Christnacher, Emmanuel Bacher, Nicolas Metzger, Stéphane Schertzer, and Thomas Scholz. New approaches of three-dimensional range-gated imaging in scattering environments. In *Proc. SPIE*, 2011.

[98] Martin Laurenzis, Frank Christnacher, Jonathan Klein, Matthias B. Hullin, and Andreas Velten. Study of single photon counting for non-line-of-sight vision. In *Proc. SPIE*, 2015.

[99] Amanda Levine, Katie Wang, and Orit Markowitz. Optical coherence tomography in the diagnosis of skin cancer. *Dermatologic clinics*, 35, 2017.

[100] Aviad Levis, Yoav Y Schechner, Amit Aides, and Anthony B Davis. Airborne three-dimensional cloud tomography. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2015.

[101] Aviad Levis, Yoav Y Schechner, and Anthony B Davis. Multiple-scattering microphysics tomography. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2017.

[102] Chengbo Li, Wotao Yin, and Yin Zhang. User's guide for TVAL3: TV minimization by augmented Lagrangian and alternating direction algorithms. *CAAM Report*, 2009.

[103] Kunming Li, Yu Li, Shaodi You, and Nick Barnes. Photo-realistic simulation of road scene for data-driven methods in bad weather. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[104] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Single-image noise level estimation for blind denoising. *IEEE Transactions on Image Processing*, 22, 2013.

[105] Antoine Liutkus, David Martina, Sébastien Popoff, Gilles Chardon, Ori Katz, Geoffroy Lerosey, Sylvain Gigan, Laurent Daudet, and Igor Carron. Imaging with nature: compressive imaging using a multiply scattering medium. *Nature Scientific Reports*, 4, 2014.

[106] Manolis Lourakis. TV-L1 image denoising algorithm. https://www.mathworks.com/matlabcentral/fileexchange/57604-tv-l1-image-denoising-algorithm, 2016.

[107] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2008.

[108] Tomohiro Maeda, Guy Satat, and Ramesh Raskar. Time resolved imaging through tissue with structured light. *In Preparation*, 2019.

[109] Stéphane Mallat. Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374, 2016.

[110] Richard M. Marino and William R. Davis. Jigsaw: a foliage-penetrating 3d imaging laser radar system. *Lincoln Lab Journal*, 15, 2005.

[111] EA McLean, HR Burris, and MP Strand. Short-pulse range-gated optical imaging in turbid water. *Applied Optics*, 34, 1995.

[112] Samy Metari and François Deschenes. A new convolution kernel for atmospheric point spread function applied to computer vision. In *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[113] Thomas P Minka. Beyond newton's method, 2000.

[114] Thomas P Minka. Estimating a gamma distribution. *Microsoft Research, Cambridge, UK, Technical Report*, 2002.

[115] Allard Mosk, Ad Lagendijk, Geoffroy Lerosey, and Mathias Fink. Controlling waves in space and time for imaging and focusing in complex media. *Nature Photonics*, 2012.

[116] Allard Mosk, Yaron Silberberg, Kevin Webb, and Changhuei Yang. Imaging, sensing, and communication through highly scattering complex media. Technical report, Purdue University Lafayette IN, 2015.

[117] Nikhil Naik, Christopher Barsi, Andreas Velten, and Ramesh Raskar. Estimating wide-angle, spatially varying reflectance using time-resolved inversion of backscattered light. *Journal Optical Society of America A*, 2014.

[118] Keiichi Nakagawa, Atsushi Iwasaki, Yu Oishi, Ryoichi Horisaki, Akira Tsukamoto, Aoi Nakamura, Kenichi Hirosawa, Hongen Liao, Takashi Ushida, Keisuke Goda, et al. Sequentially timed all-optical mapping photography (stamp). *Nature Photonics*, 8, 2014.

[119] Srinivasa G. Narasimhan, Shree K. Nayar, Bo Sun, and Sanjeev J. Koppal. Structured light in scattering media. In *IEEE International Conference on Computer Vision (ICCV)*, 2005.

[120] Vasilis Ntziachristos. Going deeper than microscopy: the optical imaging frontier in biology. *Nature methods*, 7(8):603–14, August 2010.

[121] Vasilis Ntziachristos. Going deeper than microscopy: the optical imaging frontier in biology. *Nature Methods*, 7, 2010.

[122] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook. *Technical University of Denmark*, 7, 2008.

[123] Gabriel Peyré and Laurent D. Cohen. Geodesic remeshing using front propagation. *International Journal Computer Vision*, 69, 2006.

[124] T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko. Optical imaging by means of two-photon quantum entanglement. *Physics Review A*, 52, 1995.

[125] James Polans, Ryan P. McNabb, Joseph A. Izatt, and Sina Farsiu. Compressed wavefront sensing. *Optics Letters*, 39, 2014.

[126] Andrew Profeta, Andres Rodriguez, and H. Scott Clouse. Convolutional neural networks for synthetic aperture radar classification. *Proc. SPIE*, 9843, 2016.

[127] Dou Quan, Shuang Wang, Mengdan Ning, Tao Xiong, and Licheng Jiao. Using deep neural networks for synthetic aperture radar image registration. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016.

[128] P Rairoux, H Schillinger, S Niedermeier, M Rodriguez, F Ronneberger, R Sauerbrey, B Stein, D Waite, C Wedekind, H Wille, et al. Remote sensing of the atmosphere using ultrashort laser pulses. *Applied Physics B*, 71, 2000.

[129] Raskar Ramesh and James Davis. 5d time-light transport matrix: What can we reason about scene properties? Technical report, Massachusetts Institute of Technology, 2008.

[130] Dan Raviv, Christopher Barsi, Nikhil Naik, Micha Feigin, and Ramesh Raskar. Pose estimation using time-resolved inversion of diffuse light. *Optics Express*, 22, 2014.

[131] Sidney Redner. *A guide to first-passage processes*. Cambridge University Press, 2001.

[132] Albert Redo-Sanchez, Barmak Heshmat, Alireza Aghasi, Salman Naqvi, Mingjie Zhang, Justin Romberg, and Ramesh Raskar. Terahertz time-gated spectral imaging for content extraction through layered structures. *Nature Communications*, 7, 2016.

[133] William L Rice, Steven Hou, and Anand TN Kumar. Resolution below the point spread function for diffuse optical imaging using fluorescence lifetime multiplexing. *Optics Letters*, 38, 2013.

[134] J. A. Richardson, L. A. Grant, and R. K. Henderson. Low dark count single-photon avalanche diode structure compatible with standard nanometer scale cmos technology. *IEEE Photonics Technology Letters*, 21, 2009.

[135] Justin Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25, 2008.

[136] Brian E Roth, K Clint Slatton, and Matthew J Cohen. On the potential for high-resolution lidar to improve rainfall interception estimates in forest ecosystems. *Frontiers in Ecology and the Environment*, 5, 2007.

[137] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60, 1992.

[138] Alaa Saade, Francesco Caltagirone, Igor Carron, Laurent Daudet, Angélique Drémeau, Sylvain Gigan, and Florent Krzakala. Random projections through multiple optical scattering: Approximating kernels at the speed of light. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.

[139] Guy Satat. Imaging through scattering. Master's thesis, Massachusetts Institute of Technology, 2015.

[140] Guy Satat, Christopher Barsi, Barmak Heshmat, Dan Raviv, and Ramesh Raskar. Locating fluorescence lifetimes behind turbid layers non-invasively using sparse, time-resolved inversion. In *CLEO: Fundamental Science*, 2014.

[141] Guy Satat, Christopher Barsi, and Ramesh Raskar. Skin perfusion photography. In *IEEE International Conference on Computational Photography (ICCP)*, 2014.

[142] Guy Satat, Barmak Heshmat, Christopher Barsi, Dan Raviv, Ou Chen, Moungi G Bawendi, and Ramesh Raskar. Locating and classifying fluorescent tags behind turbid layers using time-resolved inversion. *Nature Communications*, 6, 2015.

[143] Guy Satat, Barmak Heshmat, Nikhil Naik, Albert Redo-Sanchez, and Ramesh Raskar. Advances in ultrafast optics and imaging applications. In *Ultrafast Bandgap Photonics*, volume 9835, 2016.

[144] Guy Satat, Barmak Heshmat, and Ramesh Raskar. All photons imaging through layered scattering materials. In *Propagation Through and Characterization of Atmospheric and Oceanic Phenomena*. Optical Society of America, 2017.

[145] Guy Satat, Barmak Heshmat, Dan Raviv, and Ramesh Raskar. All photons imaging through volumetric scattering. *Nature Scientific Reports*, 6, 2016.

[146] Guy Satat, Tomohiro Maeda, and Ramesh Raskar. GPU accelerated monte carlo renderer for scattering media. https://github.com/mitmedialab/MonteCarloRender, 2019.

[147] Guy Satat, Gabriella Musarra, Ashley Lyons, Barmak Heshmat, Ramesh Raskar, and Daniele Faccio. Compressive ultrafast single pixel camera. In *Computational Optical Sensing and Imaging*. Optical Society of America, 2018.

[148] Guy Satat, Matthew Tancik, Otkrist Gupta, Barmak Heshmat, and Ramesh Raskar. Object classification through scattering media with deep learning on time resolved measurement. *Optics Express*, 25, 2017.

[149] Guy Satat, Matthew Tancik, and Ramesh Raskar. Lensless imaging with compressive ultrafast sensing. *IEEE Transactions on Computational Imaging*, 3, 2017.

[150] Guy Satat, Matthew Tancik, and Ramesh Raskar. Imaging through volumetric scattering with a single photon sensitive camera. In *Mathematics in Imaging*. Optical Society of America, 2018.

[151] Guy Satat, Matthew Tancik, and Ramesh Raskar. Towards photography through realistic fog. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2018.

[152] Alisha Saxena. A data-driven approach to object classification through fog. Master's thesis, Massachusetts Institute of Technology, 2018.

[153] Yoav Y. Schechner, Srinivasa G. Narasimhan, and Shree K. Nayar. Instant dehazing of images using polarization. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2001.

[154] K Scheidt. Review of streak cameras for accelerators: features, applications and results. In *Proc. of EPAC*, 2000.

[155] Bradley W Schilling, Dallas N Barr, Glen C Templeton, Lawrence J Mizerka, and C Ward Trussell. Multiple-return laser radar for three-dimensional imaging through obscurations. *Applied Optics*, 41, 2002.

[156] Armin Schwartzman, Marina Alterman, Rotem Zamir, and Yoav Y Schechner. Turbulence-induced 2d correlated image distortion. In *IEEE International Conference on Computational Photography (ICCP)*, 2017.

[157] Brent Schwarz. Lidar: mapping the world in 3D. *Nature Photonics*, 4, 2010.

[158] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R. Marschner, Mark Horowitz, Marc Levoy, and Hendrik P. A. Lensch. Dual photography. *ACM Transactions on Graphics*, 24, 2005.

[159] Jeffrey Shapiro. Computational ghost imaging. *Physical Review A*, 78, 2008.

[160] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: An astounding baseline for recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[161] Sheng-Wen Shih, Yu-Te Wu, and Jin Liu. A calibration-free gaze tracking technique. In *IEEE International Conference on Pattern Recognition (ICPR)*, volume 4, 2000.

[162] Dongeek Shin, Feihu Xu, Dheera Venkatraman, Rudi Lussana, Federica Villa, Franco Zappa, Vivek K Goyal, Franco NC Wong, and Jeffrey H Shapiro. Photon-efficient imaging with a single-photon camera. *Nature Communications*, 7, 2016.

[163] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *Advances in Neural Information Processing Systems*, 2014.

[164] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[165] Ori Spier, Tali Treibitz, and Guy Gilboa. In situ target-less calibration of turbid media. *IEEE International Conference on Computational Photography (ICCP)*, 2017.

[166] Adrian Stern. *Optical compressive imaging*. CRC Press, 2016.

[167] B Sun, M P Edgar, R Bowman, L E Vittert, S Welsh, A Bowman, and M J Padgett. 3D computational imaging with single-pixel detectors. *Science*, 340, 2013.

[168] Tristan Swedish, Guy Satat, and Ramesh Raskar. Data-driven localization around corners with consumer cameras. *In Preparation*, 2019.

[169] A Szameit, Y Shechtman, E Osherovich, E Bullkich, P Sidorenko, H Dana, S Steiner, E B Kley, S Gazit, T Cohen-Hyams, S Shoham, M Zibulevsky, I Yavneh, Y C Eldar, O Cohen, and M Segev. Sparsity-based single-shot subwavelength coherent diffractive imaging. *Nature Materials*, 11, 2012.

[170] Dharmpal Takhar, Jason N Laska, Michael B Wakin, Marco F Duarte, Dror Baron, Shriram Sarvotham, Kevin F Kelly, and Richard G Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Computational Imaging IV*, 2006.

[171] Matthew Tancik, Guy Satat, and Ramesh Raskar. Flash photography for data-driven hidden scene recovery. *arXiv preprint arXiv:1810.11710*, 2018.

[172] Matthew Tancik, Tristan Swedish, Guy Satat, and Ramesh Raskar. Data-driven non-line-of-sight imaging with a traditional camera. In *Imaging Systems and Applications*. Optical Society of America, 2018.

[173] Graham W Taylor, Rob Fergus, Yann LeCun, and Christoph Bregler. Convolutional learning of spatio-temporal features. *European Conference on Computer Vision (ECCV)*, 2010.

[174] Jiandong Tian, Zak Murez, Tong Cui, Zhen Zhang, David Kriegman, and Ravi Ramamoorthi. Depth and image restoration from light field in a scattering medium. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[175] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[176] Tali Treibitz and Yoav Y. Schechner. Active polarization descattering. *IEEE Transactions Pattern Analysis Machine Intelligence*, 31, 2009.

[177] Ivo M Vellekoop, Aart Lagendijk, and AP Mosk. Exploiting disorder for perfect focusing. *Nature Photonics*, 4, 2010.

[178] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3, 2012.

[179] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 2010.

[180] Alfred Vogel and Vasan Venugopalan. Mechanisms of pulsed laser ablation of biological tissues. *Chemical Reviews*, 103, 2003.

[181] Albert H Walenta, J Fehlmann, H Hofer, J Paradiso, and G Viertel. The time expansion chamber as a high precision drift chamber. In *Proceedings, International Conference on Instrumentation for Colliding Beam Physics, SLAC*, 1982.

[182] Laura Waller and Lei Tian. Computational imaging: Machine learning for 3D microscopy. *Nature*, 523, 2015.

[183] L Wang, PP Ho, C Liu, G Zhang, and RR Alfano. Ballistic 2-d imaging through scattering walls using an ultrafast optical kerr gate. *Science*, 253, 1991.

[184] Lihong V. Wang and Hsin-I Wu. *Biomedical Optics: Principles and Imaging*. John Wiley & Sons, 2012.

[185] Lihong V Wang and Song Hu. Photoacoustic tomography: in vivo imaging from organelles to organs. *Science*, 335, 2012.

[186] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13, 2004.

[187] Claire M. Watts, David Shrekenhamer, John Montoya, Guy Lipworth, John Hunt, Timothy Sleasman, Sanjay Krishna, David R. Smith, and Willie J. Padilla. Terahertz compressive imaging with metamaterial spatial light modulators. *Nature Photonics*, 8, 2014.

[188] Stephen S. Welsh, Matthew P. Edgar, Richard Bowman, Phillip Jonathan, Baoqing Sun, and Miles J. Padgett. Fast full-color computational imaging with single-pixel detectors. *Optics Express*, 21, 2013.

[189] Di Wu, Gordon Wetzstein, Christopher Barsi, Thomas Willwacher, Qionghai Dai, and Ramesh Raskar. Ultra-fast lensless computational imaging through 5D frequency analysis of time-resolved light transport. *International Journal Computer Vision*, 110, 2014.

[190] Xiao Xu, Honglin Liu, and Lihong V. Wang. Time-reversed ultrasonically encoded optical focusing into scattering media. *Nature Photonics*, 5, 2011.

[191] Zahid Yaqoob, Demetri Psaltis, Michael S Feld, and Changhuei Yang. Optical phase conjugation for turbidity suppression in biological samples. *Nature Photonics*, 2, 2008.

[192] Yongduek Seo and Ki Sang Hong. Calibration-free augmented reality in perspective. *IEEE Transactions on Visualization and Computer Graphics*, 2000.

[193] K. M. Yoo, Feng Liu, and R. R. Alfano. When does the diffusion approximation fail to describe photon transport in random media? *Physical Review Letters*, 64, 1990.

[194] K. M. Yoo, Feng Liu, and R. R. Alfano. Imaging through a scattering wall using absorption. *Optics Letters*, 16, 1991.

[195] J. S. You, C. K. Hayakawa, and V. Venugopalan. Frequency domain photon migration in the δ- p 1 approximation: Analysis of ballistic, transport, and diffuse regimes. *Physical Review E*, 72, 2005.

[196] Benjamin W Zeff, Brian R White, Hamid Dehghani, Bradley L Schlaggar, and Joseph P Culver. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proceedings of the National Academy of Sciences*, 104, 2007.

[197] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.