

OPTICAL LINEAR ALGEBRA FOR COMPUTATIONAL LIGHT TRANSPORT

by

Matthew O'Toole

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

© Copyright 2016 by Matthew O'Toole

Abstract

Optical Linear Algebra for Computational Light Transport

Matthew O’Toole

Doctor of Philosophy

Graduate Department of Computer Science

University of Toronto

2016

Active illumination refers to optical techniques that use controllable lights and cameras to analyze the way light propagates through the world. These techniques confer many unique imaging capabilities (*e.g.* high-precision 3D scanning, image-based relighting, imaging through scattering media), but at a significant cost; they often require long acquisition and processing times, rely on predictive models for light transport, and cease to function when exposed to bright ambient sunlight.

We develop a mathematical framework for describing and analyzing such imaging techniques. This framework is deeply rooted in numerical linear algebra, and models the transfer of radiant energy through an unknown environment with the so-called light transport matrix. Performing active illumination on a scene equates to applying a numerical operator on this unknown matrix.

The brute-force approach to active illumination follows a two-step procedure: (1) optically measure the light transport matrix and (2) evaluate the matrix operator numerically. This approach is infeasible in general, because the light transport matrix is often much too large to measure, store, and analyze directly.

Using principles from optical linear algebra, we evaluate these matrix operators in the optical domain, without ever measuring the light transport matrix in the first place. Specifically, we explore numerical algorithms that can be implemented partially or fully with programmable optics. These optical algorithms provide solutions to many longstanding problems in computer vision and graphics, including the ability to (1) photo-realistically change the illumination conditions of a given photo with only a handful of measurements, (2) accurately capture the 3D shape of objects in the presence of complex transport properties and strong ambient illumination, and (3) overcome the multipath interference problem associated with time-of-flight cameras. Most importantly, we introduce an all-new imaging regime—optical probing—that provides unprecedented control over which light paths contribute to a photo.

Acknowledgements

First and foremost, I am forever indebted to my advisor Kyros Kutulakos for his guidance and unwavering support throughout my time at the University of Toronto. Our research sessions together have been highly engaging, challenging, and thought provoking; I could not have asked for a better teacher or mentor. This work would have never been possible without him.

This work is also the product of several fruitful collaborations. Chapter 5 was done in collaboration with Supreeth Achar and Srinivasa Narasimhan. Supreeth's engineering prowess and dedication exceeded all expectations, and I'm very fortunate to have worked with both him and Srinivas. Chapter 6 was also the result of working closely with Felix Heide, Lei Xiao, Matthias Hullin, and my undergraduate advisor Wolfgang Heidrich. I also thank John Mather and Ramesh Raskar for their contributions to this work.

Thanks to Allan Jepson and Eugene Fiume for their timely feedback, advice over the years, and willingness to meet on such short notice. I also thank my non-supervisory committee members, David Fleet and Ravi Ramamoorthi, for taking the time to be part of my examination committee.

I also want to acknowledge the many students, postdocs, and friends from DGP that have made an impact in my life: Fanny Chevalier, John Hancock, Sam Hasinoff, Christian Lessig, Nigel Morris, Derek Nowrouzezahrai, Peter O'Donovan, Xavier Snelgrove, Huixuan Tang, Michael Tao, Jack Wang, and far too many others to name here (but they know who they are). Special thanks to my friends from back home: Daniel Allard, David Chow, Vincent Lam, Bardia and Ida Noohi, and Bijan Shahrokhi. Thank you to Abhishek Ranjan for the climbing sessions, Hanieh Bastani for the snowboard trips to Blue Mountain, and all those from the ice rink: Anand Agarawala, Paul Blacklock, Patrick Dubroy, Cosmin and Ramona Munteanu, and Karen Reich.

Thank you to all those who made my stay at the MIT Media Lab so very special: Matt Hirsch, Douglas Lanman, Kshitij Marwah, Belen Masia, Ramesh Raskar, Andreas Velten, Gordon Wetzstein, and Di Wu.

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the CGS-M, PGS-D, and GRAND NCE programs, the Robert E. Lansdale / Okino Computer Graphics Graduate Fellowship, and the Queen Elizabeth II Computer Science Scholarship in Science and Technology (QEII-GSST) program.

Enfin, je souhaite remercier ma famille et ma copine: je ne serais pas ce que je suis aujourd'hui sans votre soutien, encouragement, patience, et amour. Merci.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	3
1.1 The Light Transport Matrix	5
1.1.1 Radiant Energy	5
1.1.2 Ambient Light	6
1.2 Light Sources and Cameras	6
1.2.1 Cameras	6
1.2.2 Light sources	7
1.2.3 Noise	7
1.3 Light Paths	9
1.3.1 Elements of the Light Transport Matrix	11
1.4 Light Transport Analysis	11
1.4.1 Image-based Relighting	11
1.4.2 Triangulation-based 3D Scanning	12
1.4.3 Dual Photography	13
1.4.4 Limitations	13
1.5 Overview	14
2 Optical Computing for Fast Light Transport Analysis	15
2.1 Computing with Light	16
2.1.1 A Simple Example: Optical Power Iteration	16
2.1.2 Optical Krylov Subspace Methods	18
2.2 Optical Arnoldi for Transport Acquisition	21
2.2.1 Relation to Prior Work on Transport Acquisition	23
2.3 Optical GMRES for Inverse Transport	25
2.3.1 Relation to Prior Work on Inverse Transport	25
2.4 Implementation	27
2.5 Results	29
2.5.1 Acquiring Transport Matrices with Optical Arnoldi	29

2.5.2	Inverting Light Transport with Optical GMRES	33
2.6	Relation to Work in Optical Computing	33
2.7	Inexact Krylov Subspace Methods	35
2.8	Summary and Contributions	35
3	Optical Probing by Primal-Dual Coding	36
3.1	Probing Light Transport	38
3.2	Photography by Optical Probing	40
3.2.1	Naive Approach: Path Isolation	41
3.2.2	Optical Matrix Probing	43
3.3	Implementation	45
3.4	Results	48
3.5	Discussion	53
3.6	Summary and Contributions	54
4	Structured Light Transport	55
4.1	The Stereo Transport Matrix	56
4.2	Dominance of Non-Epipolar Transport	58
4.3	Imaging by Structured Light Transport	60
4.4	Live Structured-Light-Transport Imaging	63
4.5	Implementation	67
4.6	Results	68
4.7	Summary and Contributions	72
5	Energy-Efficient Probing by Homogeneous Matrix Factorization	73
5.1	Lights, Masks and Energy Efficiency	75
5.1.1	Redistributive Projection and Sensor Masking	75
5.1.2	Optical Probing	77
5.1.3	Homogeneous Factorization	78
5.2	Homogeneous Low-Rank Factorization	79
5.2.1	Implications of the Theory	80
5.3	DMDs for Live Energy-Efficient Imaging	81
5.4	Lasers for Live Energy-Efficient Imaging	86
5.4.1	Epipolar and Non-Epipolar Imaging	87
5.4.2	Epipolar Structured Light	89
5.4.3	Live Dual Videography	91
5.5	Summary and Contributions	92
6	Probing Transient Light Transport	94
6.1	Related Work on Transient Imaging	96
6.2	Light Transport in Space and Time	97
6.3	Analysis by Temporal Frequency Probing	100

6.4	Implementation	102
6.5	Results	106
6.6	Summary and Contributions	111
7	Conclusions and Future Work	112
7.1	Limitations	114
7.2	Future Research	114
	Appendices	116
A	Extended Discussion	117
A.1	One-shot optical probing versus mask-less multi-image acquisition	117
B	Proofs and Derivations	119
B.1	Expanded derivations	119
B.1.1	Derivation of Direct-Enhanced and Indirect-Only Probing Matrices	119
B.1.2	Derivation of Eq. (3.9)	121
B.1.3	Derivation of Eq. (4.8)	122
B.1.4	Derivation of Eq. (4.13)	122
B.1.5	Derivation of Eq. (5.15)	123
B.2	Proofs of Propositions 1 and 2	126
B.2.1	Proof of Proposition 1	126
B.2.2	Proof Sketch of Proposition 2	128
B.3	Fourier Transform and Convolution for Matrix-valued Functions	129
C	Hardware	130
	Bibliography	131

Notation

Scalars, Vectors, and Matrices

a	Scalar
\mathbf{a}	Vector
$\mathbf{a}[m]$	m^{th} element of the vector \mathbf{a}
\mathbf{a}_i	i^{th} vector of a sequence, or block i of a block partitioned vector \mathbf{a}
\mathbf{A}	Matrix
$\mathbf{A}[n]$	n^{th} column of the matrix \mathbf{A}
$\mathbf{A}[m, n]$	Element at row m and column n of the matrix \mathbf{A}
\mathbf{A}_{ij}	Block i, j of a block partitioned matrix \mathbf{A}

Vector Operations

$ \mathbf{a} $	Absolute value of vector \mathbf{a}
$(\mathbf{a})_+$	Positive component of vector \mathbf{a} , <i>i.e.</i> $\max(\mathbf{a}, 0)$
$(\mathbf{a})_-$	Positive component of negated vector \mathbf{a} , <i>i.e.</i> $\max(-\mathbf{a}, 0)$
$\bar{\mathbf{a}}$	Complement of binary vector \mathbf{a} , <i>i.e.</i> $1 - \mathbf{a}$
$\arg \mathbf{a}$	Argument vector \mathbf{z} of a complex vector $\mathbf{a} = \mathbf{a} e^{i\mathbf{z}}$
$\mathbf{a} \circ \mathbf{b}$	Hadamard (element-wise) product between vectors \mathbf{a} and \mathbf{b}

Matrix Operations

$\mathbf{A}\mathbf{b}$	Matrix-vector product, <i>i.e.</i> $\sum_n \mathbf{A}[n]\mathbf{b}[n]$
\mathbf{A}^T	Transpose of matrix \mathbf{A}
\mathbf{A}^k	k^{th} power of square matrix \mathbf{A}
\mathbf{A}^{-1}	Inverse of matrix \mathbf{A}
$\mathbf{A} \odot \mathbf{B}$	Hadamard (element-wise) product between matrices \mathbf{A} and \mathbf{B}
\mathbf{A}^\dagger	Moore-Penrose pseudoinverse of matrix \mathbf{A} , <i>i.e.</i> $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ (where \mathbf{A} has full column rank)

Norms

$\ \mathbf{a}\ _p$	ℓ_p norm of vector \mathbf{a} , <i>i.e.</i> $(\sum_m \mathbf{a}[m] ^p)^{\frac{1}{p}}$
$\ \mathbf{a}\ _1$	ℓ_1 norm of vector \mathbf{a} , <i>i.e.</i> $\sum_m \mathbf{a}[m] $
$\ \mathbf{a}\ _2$	ℓ_2 norm of vector \mathbf{a} , <i>i.e.</i> $\sqrt{\sum_m \mathbf{a}[m] ^2}$
$\ \mathbf{a}\ _\infty$	ℓ_∞ norm of vector \mathbf{a} , <i>i.e.</i> $\max_m \mathbf{a}[m] $
$\ \mathbf{A}\ _F$	Frobenius norm of matrix \mathbf{A} , <i>i.e.</i> $\sqrt{\sum_{m,n} \mathbf{A}[m, n] ^2}$
$\ \mathbf{A}\ _W$	Weighted Frobenius norm of matrix \mathbf{A} , <i>i.e.</i> $\sqrt{\sum_{m,n} \mathbf{W}[m, n] \mathbf{A}[m, n] ^2}$
$\ \mathbf{A}\ _{\max}$	Max norm of matrix \mathbf{A} , <i>i.e.</i> $\max_{m,n} \mathbf{A}[m, n] $
$\ \mathbf{A}\ _{p,q}$	$\ell_{p,q}$ norm of matrix \mathbf{A} , <i>i.e.</i> $(\sum_m (\sum_n \mathbf{A}[m, n] ^p)^{\frac{q}{p}})^{\frac{1}{q}}$

Additive and Multiplicative Identities

1	Vector of ones
0	Vector of zeros
I	The identity matrix

Units

J	Joules
m	Meters
s	Seconds
Hz	Hertz, <i>i.e.</i> s^{-1}
W	Watts, <i>i.e.</i> $J \cdot s^{-1}$
dB	Decibels (logarithmic unit of measurement)
lm	Lumens
1	No Units

Common Scalar Definitions

E	Number of epipolar lines (1)
I	Number of camera pixels (1)
K	Number of iterations (1)
P	Number of projector pixels (1)
T	Total exposure time (s)
g	Sensor gain (1)
σ	Standard deviation of read noise (1)
Φ	Power of light source (W)
Ψ	Number of photoelectrons per unit of radiant energy (J^{-1})
ρ	A projector's redistribution ratio (1)
τ	Transient travel time (s)
ω	Transient frequency (s^{-1})

Common Vector Definitions

a	The ambient vector of length I (J)
i	The photo vector of length I (J)
m	The mask vector of length I (1)
n	The noise vector of length I (1)
p	The illumination vector of length P (J)

Common Matrix Definitions

M	Matrix of size $I \times K$ representing K mask vectors (1)
P	Matrix of size $P \times K$ representing K illumination vectors (J)
T	The light transport matrix of size $I \times P$ (1)
T$^\omega$	The transient frequency transport matrix of size $I \times P$ for frequency ω (1)
\Pi	The probing matrix of size $I \times P$ (1)

Chapter 1

Introduction

The interaction between light and matter produces a wide array of complex optical phenomena. On a clear day, the particles within the atmosphere scatter light of shorter (blue) wavelengths much more strongly than longer (red) wavelengths to produce a blue sky. At sunset, sunlight must travel a much longer distance through the atmosphere to reach an observer; thus, light at shorter wavelengths scatters away before reaching the observer to produce the appearance of an orange or red sun. The visible spectrum of light appears in rainbows after a rainy day by the way white light disperses through raindrops. Wavy water surfaces refract and focus light to form caustics along lake and sea beds. The scattering of light through participating media produces clouds and fog. Despite its complex nature, this interaction between light and matter is easy to appreciate for its beauty and is basic to our experience of the world.

The aim of *computational light transport* is to analyze such interaction with controllable lights and cameras. The term *active imaging* refers to sensing devices that use controllable light sources to probe an environment, by sending a light signal into a scene and detecting the light reflected in response. This is in contrast with the term *passive imaging*, reserved for devices that simply detect the naturally occurring light from an environment. For example, a camera is a device that can act either passively or actively; modern cameras are commonly built with flash bulbs that can actively discharge a burst of light when operating under low-light conditions.

Besides flash photography, there is a long history of devices that rely on active illumination. In the mid-1920s, photoelectric cells captured some of the first “moving photos” by raster scanning a spot of light across a subject at fast rates [113]. In the 1950s, the introduction of confocal microscopes increased the resolution of microscopes by means of rejecting all scattered light emanating from the out-of-focus regions of a volumetric specimen [85]. The finite time required for light to travel from a source, to a target, and back to a detector—its “time-of-flight”—was used in combination with the constant speed of light to calculate the distance of objects in the early 1960s [93]. In 1980, photometric stereo was introduced to compute surface orientation from the appearance of objects lit under different lighting conditions [136].

Unfortunately, there are several pitfalls when it comes to using active illumination. Many techniques require hundreds or even thousands of measurements [112, 114, 132], producing bandwidth,

storage, and processing issues. Moreover, a common assumption is that these measurements are repeatable over time (*i.e.* the scene is static, or the acquisition period is short enough that the scene appears to be static). These techniques can be inaccurate for complex scenes, particularly when light “misbehaves” [40]. And outdoors, simply detecting the active illumination in the presence of overwhelmingly strong sunlight is a challenge [44, 83].

This thesis presents a mathematical framework for describing and analyzing active illumination. The underpinning principle is the linear relationship between emitted and received light signals, represented by a scene’s *light transport matrix* [22, 91]. Analyzing a scene with active illumination is tantamount to applying a mathematical operator on this light transport matrix. Though the study of algorithms for performing matrix operations—*numerical linear algebra*—is remarkably mature [123], evaluating a matrix operator to analyze the light transport matrix remains a significant challenge for two simple reasons: the light transport matrix is unknown and is often too large to measure, store, or analyze directly.

To overcome these practical issues, this thesis draws inspiration from work on analog optical computing: processing data with light sources, modulators, lenses, and detectors. More than forty years ago, analog optical computing sought to revolutionize the field of information processing [35, 75]. Its key tenet was that the speed and parallelism of light enables very high data rates, especially for signal processing and pattern recognition tasks [4]. Although the versatility and rapid advance of microprocessors ultimately overshadowed optical computing’s early ambitions, the designs studied were very diverse; examples range from general-purpose optical computers [38] to highly specialized ones for matrix calculations [5, 72]. This thesis is a first attempt to apply optical computing principles to computational light transport. Its aim is to demonstrate that one can efficiently analyze the light transport matrix of real-world scenes by applying concepts from *numerical linear algebra* to the *optical domain* with controllable lights and cameras.

The main contribution of this thesis is a framework for analyzing the transport matrix without capturing many photos or relying on expensive processing requirements. The thesis proposes solving longstanding problems in vision and graphics by designing new imaging systems, these systems represent optical-domain implementations of matrix operations and are thus well founded mathematically, and these are experimentally shown to push the state of the art. We achieve the ability to

- photo-realistically manipulate the illumination conditions of a given photo with only a handful of measurements;
- directly acquire photos previously thought impossible to capture in a single shot, in which specific light paths have been blocked, attenuated, or enhanced;
- accurately capture the 3D geometry of scenes having complex light transport properties;
- efficiently perform active 3D scanning in brightly-lit environments (*e.g.* under sunlight); and
- overcome the multipath interference problem associated with time-of-flight cameras.

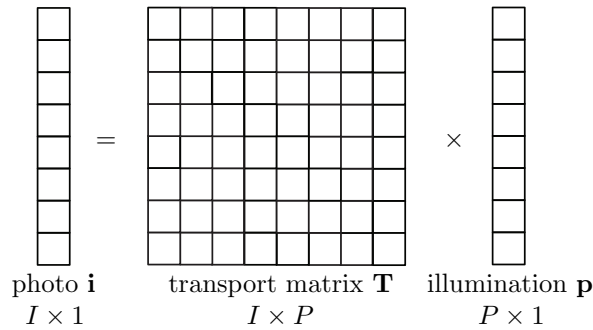


Figure 1.1: The light transport equation.

1.1 The Light Transport Matrix

A common assumption made throughout computer graphics and vision is the linearity of light transport. The interaction of light with a scene can be described mathematically in terms of a simple linear relation known as the *light transport equation* (Figure 1.1) [22, 91]:

$$\mathbf{i} = \mathbf{T}\mathbf{p} \quad (1.1)$$

where the vector \mathbf{p} of length P represents the radiant energy emitted by P controllable light sources during an exposure period of length T seconds (s), the vector \mathbf{i} of length I represents the radiant energy incident on each of I sensors during the same exposure period, and \mathbf{T} is the scene’s $I \times P$ light transport matrix. Element $\mathbf{T}[m, n]$ of the transport matrix is a unitless value describing the fraction of radiant energy transmitted from source n that reaches sensor m . If the matrix \mathbf{T} is known, the light transport equation predicts the appearance of the scene under any given illumination conditions—with light paths representing various transport phenomena (*e.g.* caustics, inter-reflections) all included. Sections 1.2 and 1.3 discuss the components of the light transport equation in further detail.

But first, we start by describing one of the basic radiometric quantities associated with the light transport equation: *radiant energy*. Radiometric quantities that depend on differential areas and angles (*e.g.* radiance and irradiance [125]) are omitted from this discussion to highlight a key point: the light transport equation is an abstract model for the transfer of radiant energy and is completely independent from the geometric configuration of light sources, sensors, and scenes.

1.1.1 Radiant Energy

Light is composed of *photons*, elementary particles that carry discrete bundles of energy along light paths at the speed of light. The energy of a photon E , measured in joules (J), is inversely proportional to its wavelength (*i.e.* its “color”):

$$E = \frac{hc}{\lambda} \quad (1.2)$$

where the wavelength λ is measured in meters (m), $c \approx 2.998 \times 10^8$ m/s is the speed of light, and $h \approx 6.626 \times 10^{-34}$ J · s is the Planck constant. Because shorter wavelengths correspond to higher

energy, “blue” photons ($\lambda \approx 450$ nm) have higher energy than “red” photons ($\lambda \approx 650$ nm). *Radiant energy* is the energy of N photons, measured by summing the energy of the N individual photons.

The value of element $\mathbf{p}[n]$ in the light transport equation describes the radiant energy emitted by light source n , and the value of element $\mathbf{i}[m]$ describes the radiant energy incident on sensor m . The light transport equation assumes that the distribution of energy radiated at different wavelengths is fixed for each light source $\mathbf{p}[n]$.

1.1.2 Ambient Light

In addition to the light emitted by *controllable* sources, *uncontrollable* light sources (*e.g.* the sun or room lighting) can also contribute radiant energy to sensors. This ambient light can be modelled by adding a vector \mathbf{a} of length I to the right-hand side of the light transport equation, where each element $\mathbf{a}[m]$ is the total amount of ambient radiant energy incident on sensor m over the exposure period T .

1.2 Light Sources and Cameras

1.2.1 Cameras

Most modern digital cameras use CCD (charge-coupled device) or CMOS (complementary metal-oxide-semiconductor) chips to record radiant energy. Their role is to convert incident radiant energy into readable digital numbers.

These chips typically contain millions of photosensitive pixels arranged in a regular grid, each of which acts as a light sensor. A photon striking a pixel can generate an electron through the physical process known as the photoelectric effect, and the electron emitted in this manner is known as a photoelectron [51]. This effect can be attributed to a transfer of energy from the absorbed photon to the emitted photoelectron. The quantum efficiency (QE) of a chip is the ratio of the number of emitted photoelectrons to the number of absorbed photons, a quantity that depends on the wavelength of light.

Each pixel accumulates photoelectrons over an exposure time T . Once this time elapses, the number of photoelectrons are read from each pixel as an amplified analog signal. An analog-to-digital converter (ADC) quantizes the analog signal into a digital number known as a pixel value.

We describe the measured pixel values of RAW (*i.e.* unprocessed) images as follows [51]:

$$\hat{\mathbf{i}} = \min \{ \Psi \mathbf{i} / g + i_{\text{offset}} + \mathbf{n}, i_{\text{max}} / g \} \quad (1.3)$$

where the vector \mathbf{i} of length I represents the radiant energy generated for each of the I pixels, and the vector $\hat{\mathbf{i}}$ contains the corresponding pixel values. The vector $\Psi \mathbf{i}$ represents the number of generated photoelectrons, where the scalar Ψ is the ratio of the quantum efficiency over the energy of a single photon (assuming a specific wavelength λ). The controllable scalar g represents sensor gain (the amplification factor applied to the analog signal); this is controlled by selecting an ISO value (the standardized industry scale for gain) where larger ISO values correspond to smaller g values. The

constant i_{offset} describes a sensor’s black level (the pixel value associated with pure black), and the scalar i_{max} represents the full-well capacity (the amount of photoelectrons a pixel can hold). The random vector \mathbf{n} represents sensor noise accumulated throughout the sensing pipeline, and is described in more detail in Section 1.2.3.

Imaging techniques exist to accurately convert pixel values back to radiant energy. Our approach is to invert Eq. (1.3) by (1) appropriately controlling gain g and exposure time T to avoid pixel saturation, (2) capturing a single RAW photo, (3) subtracting the scalar i_{offset} from the photo’s pixel values, and (4) multiplying the result with scalar g/Ψ . Merging multiple photos, each captured with different exposure and gain settings, is another well-established approach to accurately measure radiant energy [23, 51], especially for scenes with a high dynamic range where the ratio between the largest and smallest radiant energy values is large.

1.2.2 Light sources

Throughout this thesis, we use projectors as our controllable source of illumination, where every projector pixel is itself a controllable light source.

We frequently use *radiant power* to describe the rate at which energy is emitted by these light sources. Radiant power is defined as the rate of emitted (or received) radiant energy per unit time, measured in watts ($\text{W} = \text{J} \cdot \text{s}^{-1}$):

$$\Phi = \frac{dE}{dt} \quad (1.4)$$

where time t is measured in seconds. A light source produces total energy ΦT measured in joules after time T . *Luminous power*, measured in lumens (lm), is the photometric equivalent to radiant power [122]. Lumens represent the perceived power of light, accounting for the varying sensitivity of the human eye to different wavelengths of light.

We rely on three different projector technologies for illumination: DLP (digital light processing), LCD (liquid crystal display), and MEMS (microelectromechanical systems) projectors. In a DLP projector, the image is created by selectively reflecting light with an array of mirrors, known as a Digital Micromirror Device (DMD); each mirror represents a projector pixel that can be rapidly toggled on or off. A LCD projector passes light through an array of liquid crystal cells, whose transmittance can be controlled electronically. A MEMS laser projector uses a mirror that can pan and tilt to scan a laser beam across a scene; as the beam moves across the scene, the projector modulates the beam intensity to produce the desired image.

1.2.3 Noise

Random fluctuations or uncertainty in physical events (*e.g.* the detection of photoelectrons) is known as *noise*. Camera sensors have three primary sources of noise: read noise, photon noise, and quantization noise.

$$\mathbf{n} = \mathbf{n}_{\text{read}} + \mathbf{n}_{\text{photon}} + \mathbf{n}_{\text{quantization}} \quad (1.5)$$

Each element $\mathbf{n}[m]$ is a randomly sampled value that impacts the measured pixel value $\hat{\mathbf{i}}[m]$, as explained in Eq. (1.3).

Read noise, \mathbf{n}_{read} , is the electronic noise introduced by the sensor prior to analog-to-digital conversion. The random values of vector \mathbf{n}_{read} are typically modelled with a mean-zero Gaussian distribution having standard deviation σ . This noise is independent at each pixel, and independent of the signal intensity. The value of the standard deviation σ depends on the gain setting g and the specifications of the sensor itself.

Photon noise, $\mathbf{n}_{\text{photon}}$, exists because of the discrete nature of electron and photon particles, and the assumption that the emission or detection of individual photons can be treated as independent events. Although the expected number of photoelectrons detected by a pixel, $\Psi\mathbf{i}[m]$, can be any non-negative real value, the actual number of photoelectrons detected, $\mathbf{x}[m]$, must be integer. The probability of a pixel observing exactly k photoelectrons is described by the Poisson distribution:

$$\Pr(\mathbf{x}[m] = k) = \frac{(\Psi\mathbf{i}[m])^k e^{-\Psi\mathbf{i}[m]}}{k!} \quad (1.6)$$

where $e \approx 2.718$ is Euler's number, $k!$ is the factorial of k , and $\mathbf{x} = \Psi\mathbf{i} + g\mathbf{n}_{\text{photon}}$. The mean and variance of the Poisson distribution is $\Psi\mathbf{i}$. It follows that the mean value of the noise vector $\mathbf{n}_{\text{photon}}$ is zero, and its variance is $\Psi\mathbf{i}/g^2$. When the number of expected photoelectrons is large, the probability distribution function can be closely approximated with a Gaussian distribution.

Quantization noise, $\mathbf{n}_{\text{quantization}}$, represents the round-off error introduced by the analog-to-digital conversion step. Specifically, there exists a range of analog signals that produce the same digital number after quantization, and the vector $\mathbf{n}_{\text{quantization}}$ is the difference between the original and quantized signals. Quantization noise can be modelled as a uniformly distributed random value between $-b/2$ and $b/2$, where the scalar b is the least significant bit of the digital pixel value. The variance of a continuous uniform distribution with spread b is $\frac{1}{12}b^2$.

The *signal-to-noise ratio* (SNR) of $\hat{\mathbf{i}}$ characterizes image quality as the ratio between its digital pixel value and the standard deviation of its noise:

$$\text{SNR}(\hat{\mathbf{i}}) = 20 \log_{10} \left[\frac{\mathbb{E}[\hat{\mathbf{i}}]}{\sqrt{\text{Var}(\mathbf{n})}} \right] \quad (1.7)$$

Here, we express this ratio in decibels (dB), a logarithmic unit of measurement typically associated with SNR. Given the image formation model in Eq. (1.3) and the noise terms in Eq. (1.5), the expression for the SNR of digital sensors can be expanded to

$$\text{SNR}(\hat{\mathbf{i}}) = 20 \log_{10} \left[\frac{\Psi\mathbf{i}/g}{\sqrt{\sigma^2 + \Psi\mathbf{i}/g^2 + \frac{1}{12}b^2}} \right] \quad (1.8)$$

where we assume that the signal $\hat{\mathbf{i}}$ is not saturated, and the constant i_{offset} representing the sensor's black level is subtracted from the measured pixel value. An image with mean SNR 20 dB (*i.e.* a 10:1 ratio between signal and noise) is considered to be of “acceptable image quality”, and an image with mean SNR 32.04 dB (*i.e.* a 40:1 ratio) has “excellent image quality” [100].

1.3 Light Paths

We adopt the geometric optics model for light transport, where light travels along straight lines known as *rays*. When light strikes a surface, light may be reflected or refracted along a new ray. A *light path* is a sequence of rays connecting any two points (*e.g.* source and sensor).

Figure 1.2 illustrates two important light paths in a projector-camera system: *direct paths*, and *indirect paths*. Direct paths are light paths that interact with at most one scene point, whereas indirect paths interact with two or more scene points. Note that, in practice, a camera pixel records light from a bundle of direct (and indirect) paths, because the physical area of a pixel is non-zero. Similarly, a projector pixel emits a bundle of direct (and indirect) paths.

There are two general ways to classify direct paths. A *specular direct path* occurs when light is specularly redirected by a scene point (*e.g.* a mirror reflection). A *diffuse direct path* occurs when the light incident on a scene point is reflected at many angles, *i.e.*, it scatters light diffusely.

We also define two new categories to classify different indirect light paths: *specular indirect paths*, and *diffuse indirect paths*. Specular indirect paths are reflected or refracted specularly multiple times, and scatter diffusely no more than once (*e.g.* caustics). On the other hand, diffuse indirect paths scatter two or more times within a scene before reaching a sensor (*e.g.* diffuse inter-reflections, sub-surface scattering).

Note that a regular photo of a scene records all radiant energy, regardless of the paths it follows to reach the sensor. However, this photo can also be viewed as being the sum of multiple latent images, each representing a different *component* of light transport (Figure 1.3). For example, a photo is the sum of a direct component (Figure 1.3(b)) and an indirect component (Figure 1.3(c)).

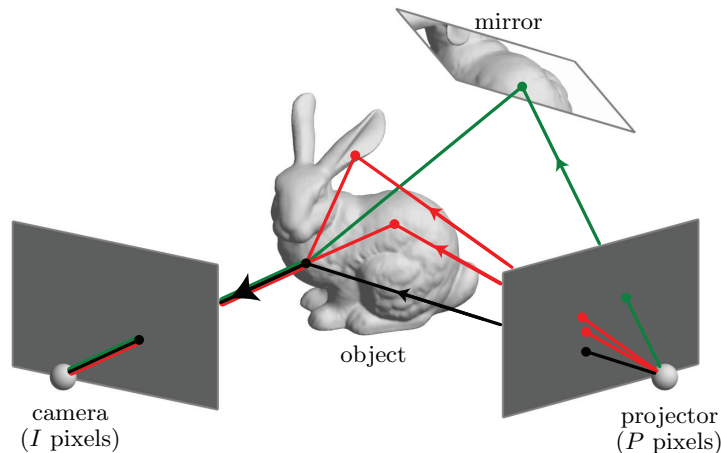


Figure 1.2: **Light paths for a projector-camera system.** There are many ways for light to reach the camera. There are direct paths (black), where light travels from the projector, to a single scene point, and back to the camera. There are also indirect paths (red, green), where light interacts with multiple scene points before reaching the camera. We also classify indirect paths as being either specular or diffuse. A diffuse indirect path (red) scatters at two or more diffuse points, and a specular indirect path (green) interacts with at most one diffuse scene point. Note that the mirror specularly redirects—rather than scatters—the green light path.

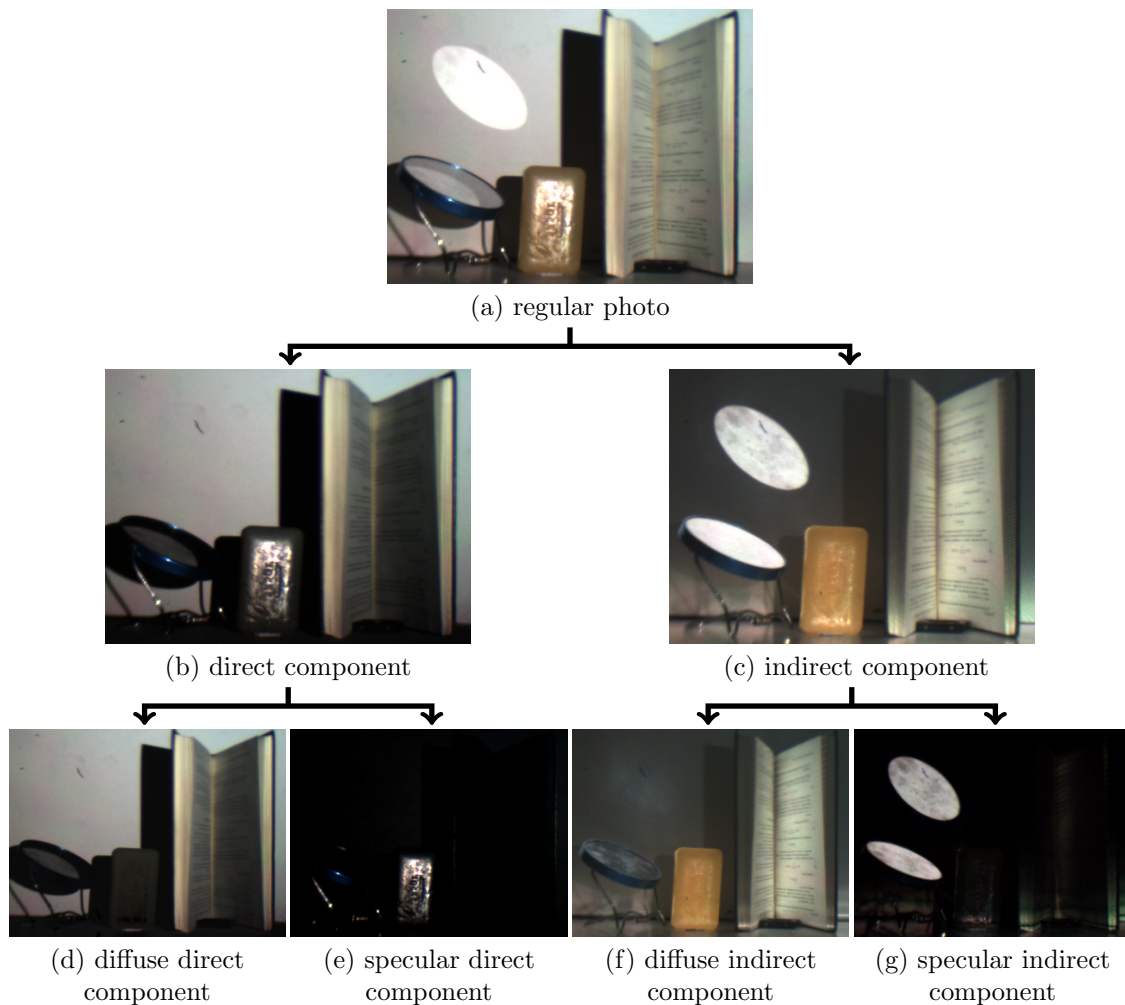


Figure 1.3: **Components of light transport.** We use a projector and a camera to approximate different components of light transport for a real scene; more details in Chapter 3 and Section 3.4. **(a)** A regular photo of the scene records all light, regardless of how it reaches the sensor. From left to right, the scene contains a mirror, a bar of soap, and an open book. The mirror specularly reflects light onto the wall behind the scene; light penetrates the surface of the bar of soap, scatters multiple times, and exits the surface at another point; some light specularly reflects off the surface of the soap; and the open book diffusely inter-reflects light between its pages. **(b)** The direct component represents all light going from the source, to a point in the scene, and directly back to the camera. Note that the light reflected by the mirror no longer appears on the wall. **(c)** The indirect component includes the light reflected by the mirror, the sub-surface scattering induced by the soap, and the light inter-reflecting within the open book. **(d)** The majority of direct light is a result of diffuse scattering. Here, the bar of soap appears very dark, because its appearance in (a) is due to specular highlights and sub-surface scattering. **(e)** The specular direct component captures the soap’s specular highlight. **(f)** Diffuse inter-reflections and sub-surface scattering are examples of diffuse indirect transport. **(g)** The specular indirect component includes light travelling from the source, to the mirror, to the wall, and back to the sensor; it also includes light going from the source, to the wall, to the mirror, and back to the sensor. Because the indirect components are dim relative to the direct components, we brighten (c), (f), and (g) by a factor of 2.

1.3.1 Elements of the Light Transport Matrix

The *scattering throughput function* $f(x)$ represents the radiant energy carried along a light path x [125]. A property of this function is that it depends only on the geometry of the scene itself, and not on the power of the light source. For brevity, we omit a discussion on the geometry of these scattering throughput functions, and instead refer to the in-depth discussion provided by Veach [125].

Each element of the transport matrix $\mathbf{T}[m, n]$ describes the ratio of radiant energy transmitted along a bundle of light paths, and is the solution to an integral of scattering throughput functions:

$$\mathbf{T}[m, n] = \int_{\Omega_{m,n}} f(x) d\mu(x) \quad (1.9)$$

where $\Omega_{m,n}$ is the space of light paths connecting source n to sensor m , and $\mu(x)$ is the corresponding measure on the space of light paths. In computer graphics, rendering is the process of generating an image from the description of a scene by solving an integral of a similar form [125].

A physically-realistic scattering throughput function has several properties. The function must be non-negative (because radiant energy is a non-negative quantity) and must conserve energy:

$$0 \leq f(x), \quad 0 \leq \mathbf{T}[m, n] = \int_{\Omega_{m,n}} f(x) d\mu(x) \leq 1 \quad (1.10)$$

The scattering throughput function is also thought to obey the *Helmholtz reciprocity* principle [125]. Helmholtz reciprocity states that the value of measure $f(x)$ for a light path x does not depend on the direction light travels along said path (*i.e.* swapping the position of a projector pixel n and camera pixel m produces the same value $\mathbf{T}[m, n]$) [114].

1.4 Light Transport Analysis

After more than a decade of graphics and vision research, there is now a large repertoire of numerical methods available to analyze the matrix \mathbf{T} of real-world scenes. Examples include methods for decomposing [7], transposing [114], approximating [30], or inverting [112] the matrix; methods that use its properties for image-based rendering [22] or transport-robust shape acquisition [42]; and imaging techniques that enhance the contribution of specific light transport components [90, 106].

As an introduction to the area, this section focuses on three of the more established works in computational light transport: image-based relighting [46], triangulation-based 3D scanning [116], and dual photography [114].

1.4.1 Image-based Relighting

Image-based relighting uses the matrix \mathbf{T} and the light transport equation (Eq. (1.1)) to photo-realistically synthesize images under novel lighting conditions [46]. For synthetic lighting conditions \mathbf{p} , the light transport equation computes a photo of the scene lit under the given lighting conditions. Figure 1.4 shows an example of image-based relighting for a scene containing only two light sources.

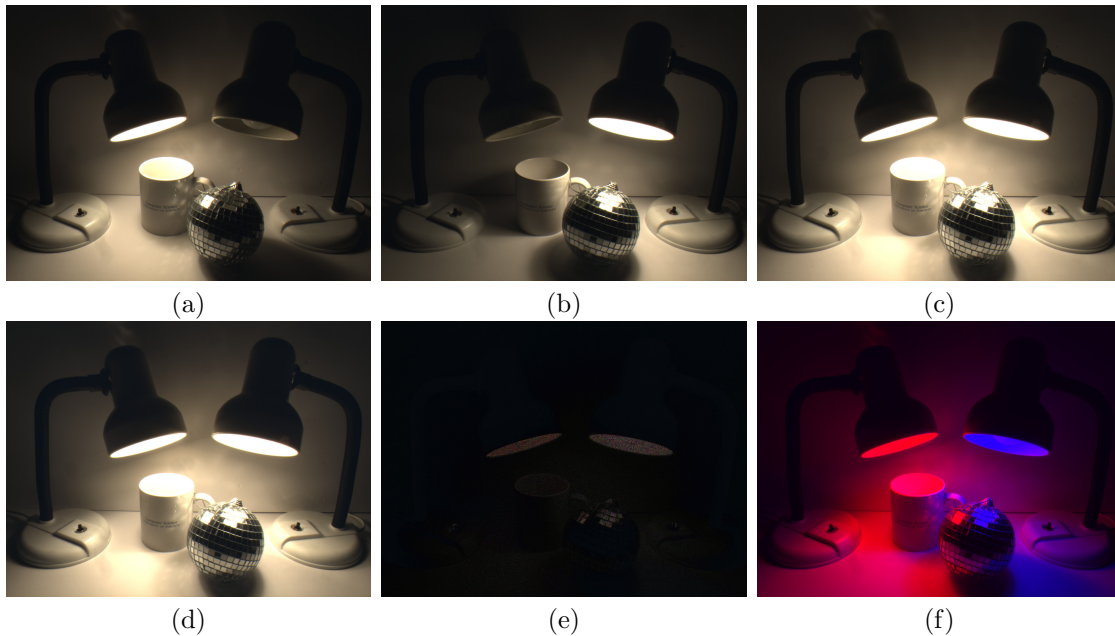


Figure 1.4: **Image-based relighting of a photo**, as first proposed by Haeberli [46]. The light transport equation can be used to synthesize photos of a scene under novel illumination conditions. The light transport matrix \mathbf{T} used here has size $I \times 2$, representing the I pixels in each photo and the two lamps within the scene. **(a)** RAW photo corresponding to column $\mathbf{T}[1]$ captured using illumination $\mathbf{p} = [1, 0]$ (left lamp “on”). **(b)** RAW photo corresponding to column $\mathbf{T}[2]$ captured using illumination $\mathbf{p} = [0, 1]$ (right lamp “on”). **(c)** RAW photo of the scene captured for $\mathbf{p} = [1, 1]$ (both lamps “on”). **(d)** A photo under synthetic lighting, by adding photos (a) and (b). **(e)** The difference between the RAW photo (c) and synthetic photo (d). The residual error is dominated by photon noise, which is largest in bright regions of the scenes. **(f)** A weighted combination of photos (a) and (b), created with the red channel of photo (a) with the blue channel of photo (b).

The visual effects industry uses image-based relighting to seamlessly place actors into virtual environments [21, 22, 52], by ensuring that the image of the actor is radiometrically consistent with the lighting within the virtual environment. To achieve this, a rig of cameras and light sources captures how an actor appears under different lighting directions to construct the light transport matrix. The image of the actor is then photorealistically relit with Eq. (1.1), before being superimposed onto a computer-generated scene.

1.4.2 Triangulation-based 3D Scanning

A triangulation-based 3D scanner uses a projector-camera system (Figure 1.2) to infer the shape of real-world objects [18, 34, 42, 110, 116]. To find the 3D position of a point on the surface of an object, the method first establishes a one-to-one correspondence between the projector pixel illuminating the point and the camera pixel observing the same point. Provided that the projector and camera have known calibration and pose, the 3D rays corresponding to these 2D pixels can be easily computed. Triangulation is the process of finding the intersection of these two rays to recover

the 3D point [48].

While triangulation-based 3D scanning has not previously been considered a light transport analysis technique, it does in fact represent a rudimentary operation on the transport matrix. This is because the correspondence problem requires identifying the camera and projector pixels that form direct light paths.

To establish correspondences, the standard approach is to identify the projector pixel n that transfers the most radiant energy to a particular camera pixel m [116]:

$$n = \arg \max_k \mathbf{T}[m, k] \quad (1.11)$$

This approach assumes direct light paths transfer more radiant energy than indirect light paths, so that matrix element $\mathbf{T}[m, n]$ represents a direct light path.

The diverse application space of 3D scanners includes digitally documenting works of art (*e.g.* Michelangelo’s David) [77], enabling users to interact with a computer through gestures and poses (*e.g.* Microsoft’s Kinect), replicating objects with 3D printers, and navigating autonomous vehicles and robots through treacherous environments.

1.4.3 Dual Photography

Dual photography is a method that photorealistically synthesizes an image from the light source’s point of view, by interchanging the roles of sensors and sources (*i.e.* lights sources behave like sensors, and sensors behave like light sources). This method by Sen et al. [114] exploits the Helmholtz principle, which states that the transfer of radiant energy along a light path does not depend on the direction of propagation.

According to Eq. (1.1), the radiant energy measured by each sensor m is a weighted sum of transport elements associated with sensor m (*i.e.* $\sum_n \mathbf{T}[m, n]\mathbf{p}[n]$). In dual photography, the radiant energy measured by a light source n is instead a weighted sum of transport elements associated with source n (*i.e.* $\sum_m \mathbf{T}[m, n]\mathbf{r}[m]$), as given by the dual equation:

$$\mathbf{s} = \mathbf{T}^T \mathbf{r} \quad (1.12)$$

where the matrix \mathbf{T}^T is the transpose of the light transport matrix, element $\mathbf{r}[m]$ is the radiant energy emitted by sensor m , and element $\mathbf{s}[n]$ is the radiant energy detected by light source n . For projector-camera systems (Figure 1.2), this dual equation computationally interchanges the position of the projector and camera within the scene. This signifies that the m^{th} row of transport matrix \mathbf{T} is an image of the scene lit by camera pixel m and captured from the projector’s point of view.

1.4.4 Limitations

The main challenge associated with these computational light transport methods (and others) is that they require the matrix \mathbf{T} of a scene to be known [92, 112, 114, 132], which is compounded by the fact that the transport matrix is often extremely large. For example, the transport matrix

of a one megapixel camera and a one megapixel projector already contains 10^{12} elements. Directly measuring such a matrix with a standard 30 Hz camera would require 10^6 measurements and over 9 hours of acquisition time, all while assuming that the scene remains perfectly static in the confines of a darkroom. Moreover, assuming the camera records 8-bit pixel values, the uncompressed transport matrix requires one terabyte of storage space.

To deal with these issues, a variety of techniques have been proposed to efficiently capture the matrix [22, 27, 30, 97, 99, 111, 114, 115, 130]. Unfortunately, these techniques often require hundreds or thousands of high dynamic range photos for complex scenes, can be computationally demanding, and their performance is hard to characterize in a scene-independent way.

1.5 Overview

One of the key questions addressed in this thesis is how to overcome the bottleneck associated with computational light transport. Motivated by large-scale matrix computations which have been the focus of numerical linear algebra since the 1970s [109], our solution involves optical algorithms that analyze the light transport matrix \mathbf{T} *without ever measuring the matrix directly*. To achieve this, we design algorithms that require nothing more than the ability to compute matrix-vector products, the same operation performed by the light transport equation (Eq. (1.1)).

Chapter 2 first introduces the concept of optical linear algebra for analyzing light transport, using a class of iterative methods known as Krylov subspace algorithms. In numerical linear algebra, these Krylov subspace algorithms are among the most successful to probe for eigenvectors of large matrices through matrix-vector products. Chapters 3 to 5 explore a new regime of imaging known as optical probing, a generalized imaging technique that provides fine-grain control over which light paths contribute to a single photo. Optical probing makes it possible to manipulate the direct and indirect light paths within a scene, dynamically scan 3D geometry of objects with complex optical properties, and even use a low-power light source to scan objects under bright sunlight. Chapter 6 establishes a fundamental link between computational light transport and time-of-flight imaging to analyze the time light takes to travel along specific light paths.

Chapter 2

Optical Computing for Fast Light Transport Analysis

We present a first attempt to apply optical computing principles to the analysis of light transport in a real-world scene. Specifically, we show how to efficiently perform numerical computations on a scene’s unknown transport matrix (eigenvector analysis, low-rank approximation, inversion) by doing part of those computations in the optical domain with projectors and cameras.

Recall that real-world transport matrices pose unique challenges for its analysis; they can be extremely large, making storing and analyzing the full matrix infeasible. Moreover, according to Eq. (1.1), the only way to get information about the matrix \mathbf{T} is by multiplying it with a vector, *i.e.*, by illuminating the scene with vector \mathbf{p} and capturing the product, $\mathbf{T}\mathbf{p}$, in a photo.

Fortunately, very large matrices have long been a subject of study in numerical linear algebra. In particular, the family of *Krylov subspace methods* [109] is designed for matrices just like \mathbf{T} , *i.e.*, very large and unobservable matrices that can only be accessed by computing their product with a vector. These iterative algorithms are well understood and come with explicit accuracy and convergence guarantees.

Here we leverage this body of work for light transport analysis by implementing Krylov subspace methods partially in optics. Our approach is based on a simple principle: treat the scene as a “black-box subroutine” that accepts any non-negative vector \mathbf{p} as “input” and returns as “output” the vector’s product, $\mathbf{T}\mathbf{p}$, with the unknown transport matrix. Thus, any efficient numerical method that relies exclusively on matrix-vector products can be readily implemented in optics and used to analyze \mathbf{T} . To do the conversion, we just replace all matrix-vector products with calls to a function that computes them *optically*, with illuminate-and-capture operations. This turns Krylov subspace methods into complete pipelines for analyzing \mathbf{T} —as they pursue their numerical objective, they fully specify how to illuminate the scene and how to process its photos.

Implementing Krylov subspace methods directly in the optical domain has several advantages. First, the convergence rate of these methods depends only on the distribution of \mathbf{T} ’s singular values, not its absolute size. This means that \mathbf{T} can be analyzed at full resolution by capturing very

few photos. Second, computations are efficient because the only computationally-expensive step is multiplying the full-resolution matrix \mathbf{T} with a vector—which we do optically. Third, optical implementations are straightforward because they differ from widely-available numerical software in just one step, *i.e.*, multiplication with \mathbf{T} . Last but not least, by moving this multiplication to the optical domain we make other computations feasible on the full-resolution \mathbf{T} , beyond mere acquisition—computing eigenvectors of \mathbf{T} , computing products with \mathbf{T} 's inverse—*without having to acquire the transport matrix first*.

We focus on optical versions of two Krylov subspace methods: *Arnoldi iteration* to acquire a low-rank approximation of the matrix for image-based relighting (Section 2.2); and *generalized minimal residual (GMRES)* to invert light transport (Section 2.3).

2.1 Computing with Light

2.1.1 A Simple Example: Optical Power Iteration

We begin with the toy problem of computing the principal eigenvector of a square transport matrix \mathbf{T} . A non-zero vector \mathbf{v} is an eigenvector of a square matrix \mathbf{T} if it satisfies the following equation:

$$\mathbf{T}\mathbf{v} = \lambda\mathbf{v} \quad (2.1)$$

where λ is a scalar known as the eigenvalue associated with eigenvector \mathbf{v} . The principal eigenvector of a matrix produces the eigenvalue with the greatest absolute value, $|\lambda|$.

Here, we show how to implement *power iteration* in optics. Power iteration is a simple numerical algorithm for estimating the principal eigenvector of a square matrix. The algorithm assumes that the absolute value of the principal eigenvalue is strictly larger than all other eigenvalues [123]. When implemented optically, it estimates the principal eigenvector of \mathbf{T} without advance knowledge of the matrix and without directly capturing any of its elements.

Power iteration uses the fact that the sequence $\mathbf{p}, \mathbf{T}\mathbf{p}, \mathbf{T}^2\mathbf{p}, \mathbf{T}^3\mathbf{p}, \dots$ converges to \mathbf{T} 's principal eigenvector for any initial vector \mathbf{p} not orthogonal to the principal eigenvector \mathbf{v} (*i.e.* $\mathbf{p}^T\mathbf{v} \neq 0$). The numerical algorithm simply generates this sequence for a fixed number of iterations using the boxed matrix-vector product shown in Algorithm 1:

Algorithm 1 *The power iteration algorithm.*

Numerical Implementation:

In: matrix \mathbf{T} , iterations K
Out: principal eigenvector of \mathbf{T}

- 1: $\mathbf{p}_1 =$ random vector
- 2: **for** $k = 1$ to K
- 3: $\mathbf{i}_k = \mathbf{T}\mathbf{p}_k$
- 4: $\mathbf{p}_{k+1} = \mathbf{i}_k / \|\mathbf{i}_k\|_2$
- 5: **return** \mathbf{p}_{k+1}

Optical Implementation:

In: iterations K
Out: principal eigenvector of \mathbf{T}

$\mathbf{p}_1 =$ positive vector
for $k = 1$ to K

illuminate scene with vector \mathbf{p}_k
 capture photo & store in \mathbf{i}_k

$\mathbf{p}_{k+1} = \mathbf{i}_k / \|\mathbf{i}_k\|_2$
return \mathbf{p}_{k+1}

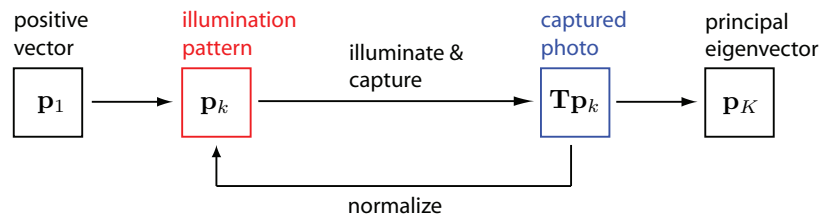


Figure 2.1: Power iteration with a projector and a camera.

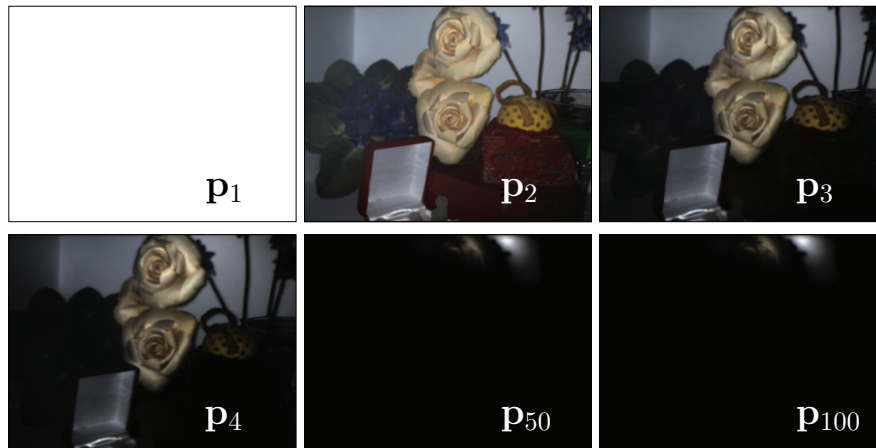


Figure 2.2: **Optical power iteration in action.** We first arranged the camera and projector to share the same viewpoint. We started with a constant illumination vector \mathbf{p}_1 , shown above, so the first photo of the scene was captured under constant illumination. That photo became the next illumination vector, \mathbf{p}_2 , also shown above. The illumination vectors change very little after about 50 captured photos, indicating that a good approximation of \mathbf{T} 's principal eigenvector was found.

Implementing power iteration in optics amounts to replacing this matrix-vector product shown in the numerical implementation of Algorithm 1 with the illuminate-and-capture operation shown in the corresponding optical implementation. This is only possible when the transport matrix is square, *i.e.*, when illumination vectors and captured photos are the same size.

The optical implementation turns the power iteration algorithm into an illumination procedure with a feedback loop. The procedure repeatedly captures a photo, converts it to a unit vector, and uses it to illuminate the scene. See Figure 2.1 for a diagram of this procedure. Figure 2.2 shows an example of using it to compute the principal eigenvector of the full-resolution transport matrix for a real scene.

Although very simple algorithmically, optical power iteration highlights an important point about our general approach: the efficiency of image acquisition is directly related to the convergence properties of the underlying numerical algorithm—the faster it converges, the fewer photos its optical implementation needs to capture.

From a numerical standpoint, power iteration is not an efficient algorithm for computing eigenvectors. It computes just one eigenvector, albeit the principal one, and the approximation error decreases by a factor of $|\lambda_2|/|\lambda_1|$ at each iteration, where λ_1 and λ_2 are the top two eigenvalues of

\mathbf{T} [123]. The algorithm may converge very slowly when \mathbf{T} 's top two eigenvalues are similar, and may not converge at all if they are identical. Naturally, these limitations are shared by its optical counterpart.

To analyze light transport efficiently, we focus on much more efficient numerical algorithms from the class of Krylov subspace methods.

2.1.2 Optical Krylov Subspace Methods

Krylov subspace methods represent some of the most important iterative algorithms for solving large linear systems [109]. Their relevance for light transport comes from the existence of powerful methods for analyzing large sparse matrices, like \mathbf{T} , be it square or rectangular.

Briefly, the Krylov subspace of dimension k is the span of vectors produced by power iteration after k steps:

$$\begin{array}{ccccccc} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \cdots & \mathbf{p}_{k+1} & & \\ & \Downarrow & \Downarrow & & \Downarrow & \cdot & \\ & \mathbf{T}\mathbf{p}_1 & \mathbf{T}^2\mathbf{p}_1 & \cdots & \mathbf{T}^k\mathbf{p}_1 & & \end{array} \quad (2.2)$$

While individual algorithms differ in their specifics, Krylov subspace methods take an initial vector \mathbf{p}_1 as input and, in their k^{th} iteration, compute a vector in the Krylov subspace of dimension k . The important characteristic of these methods is that they do not require direct access to the elements of \mathbf{T} ; all they need is the ability to multiply \mathbf{T} , and potentially its transpose, with a vector. This makes them readily implementable in optics.

Optical matrix-vector products for general vectors Unlike power iteration, general Krylov subspace methods require multiplying \mathbf{T} with vectors that may contain negative elements. Even though we cannot illuminate the scene with “negative light”, implementing such products optically is straightforward. We follow the approach outlined by Goodman [35], and express a general vector \mathbf{p} as the difference of two non-negative vectors $(\mathbf{p})_+$ and $(\mathbf{p})_-$:

$$\mathbf{p} = (\mathbf{p})_+ - (\mathbf{p})_- \quad (2.3)$$

$$\mathbf{T}\mathbf{p} = \mathbf{T}(\mathbf{p})_+ - \mathbf{T}(\mathbf{p})_- \quad (2.4)$$

To implement Eq. (2.4) optically, we use two illuminate-and-capture operations: one to compute $\mathbf{T}(\mathbf{p})_+$ and one to compute $\mathbf{T}(\mathbf{p})_-$. We subtract the two captured photos to get the product with \mathbf{p} .

This differencing operation has the additional benefit that it automatically eliminates contributions from ambient light, represented by vector \mathbf{a} (Section 1.1.2). This is because the same amount of ambient light is recorded in both photos, namely $\mathbf{T}(\mathbf{p})_+ + \mathbf{a}$ and $\mathbf{T}(\mathbf{p})_- + \mathbf{a}$.

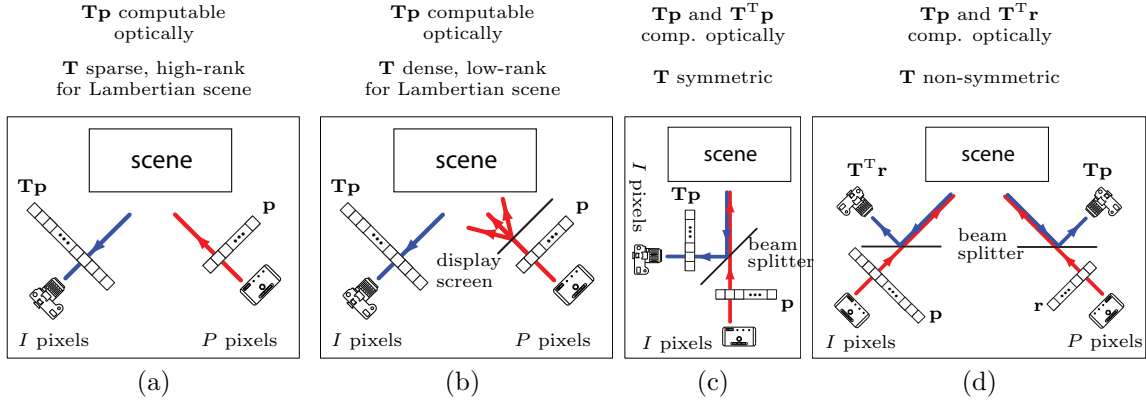


Figure 2.3: **Camera-projector arrangements.** We use three degrees of freedom in our arrangements: number of distinct viewpoints (one or two); ability to compute products with \mathbf{T}^T (yes or no); and type of illumination, *i.e.*, projecting patterns directly onto the scene as in (a), or displaying them on a screen, as in (b). Four of the eight possible arrangements are shown above.

Symmetric vs. non-symmetric transport matrices The convergence behavior of Krylov subspace methods depends quite significantly on whether or not \mathbf{T} is a symmetric matrix [78]. We restrict ourselves to the symmetric case, where convergence is well understood, by choosing appropriate projector-camera arrangements.

There are two general ways to enforce symmetry when implementing Krylov subspace methods in optics. The first is to make sure that \mathbf{T} itself is symmetric. This can be done with the coaxial arrangement of Figure 2.3(c) and is quite common [112, 142]. We also assume that the projector and camera resolutions match. Interchanging the position of the projector and camera in a coaxial system by definition produces the same light transport matrix, because it neither changes the point of view of the projector or the camera; therefore, according to Helmholtz reciprocity, the value of elements $\mathbf{T}[m, n]$ and $\mathbf{T}[n, m]$ must be the same (*i.e.* $\mathbf{T}^T = \mathbf{T}$). This arrangement is also quite limited because it does not allow any viewpoint variations between the projector and camera.

A second way to enforce symmetry is to apply the methods to a different matrix whose symmetry is guaranteed:

$$\mathbf{T}^* = \mathbf{T}^T \mathbf{T} . \quad (2.5)$$

Optically multiplying \mathbf{T}^* with a vector, however, involves matrix-vector products with both \mathbf{T} and its transpose:

$$\mathbf{T}^* \mathbf{p} = \mathbf{T}^T (\mathbf{T} \mathbf{p}) . \quad (2.6)$$

A single camera-projector pair is not enough to compute both products optically. For this, we use the arrangement of Garg et al. [30] shown in Figure 2.3(d). This arrangement uses two camera-projector pairs and enables two distinct project-and-capture operations. The first camera-projector pair computes the result of the light transport equation, $\mathbf{T} \mathbf{p}$ (Eq. (1.1)); in Figure 2.3(d), this operation “illuminates from the left, captures from the right”. By swapping the positions of the camera and projector, the second camera-projector pair optically evaluates the dual equation, $\mathbf{T}^T \mathbf{r}$ (Eq. (1.12)); in Figure 2.3(d), this operation “illuminates from the right, captures from the left”.

Algorithm	Step 1	Step 4	Step 5
Power iteration (Section 2.1.1)	$\mathbf{p}_1 = \text{positive vector}$	$\mathbf{p}_{k+1} = \mathbf{i}_k / \ \mathbf{i}_k\ _2$	return \mathbf{p}_{k+1}
Arnoldi (Section 2.2)	$\mathbf{p}_1 = \text{non-zero vector}$	$\mathbf{p}_{k+1} = \text{ortho}(\mathbf{p}_1, \dots, \mathbf{p}_k, \mathbf{i}_k)$ $\mathbf{p}_{k+1} = \mathbf{p}_{k+1} / \ \mathbf{p}_{k+1}\ _2$	return $[\mathbf{i}_1 \cdots \mathbf{i}_K][\mathbf{p}_1 \cdots \mathbf{p}_K]^\top$
GMRES (Section 2.3)	$\mathbf{p}_1 = \text{target photo } \mathbf{i}$	$\mathbf{p}_{k+1} = \text{ortho}(\mathbf{p}_1, \dots, \mathbf{p}_k, \mathbf{i}_k)$ $\mathbf{p}_{k+1} = \mathbf{p}_{k+1} / \ \mathbf{p}_{k+1}\ _2$	return $[\mathbf{p}_1 \cdots \mathbf{p}_K][\mathbf{i}_1 \cdots \mathbf{i}_K]^\dagger \mathbf{i}$

Table 2.1: The Krylov subspace methods we implement, along with their differences from power iteration. In the numerical steps listed above, the function *ortho()* projects the last vector \mathbf{i}_k in its parameter list onto the subspace orthogonal to the columns of $\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_k]$; the orthogonal projection of \mathbf{i}_k is given by the formula $\mathbf{p}_{k+1} = \mathbf{i}_k - \mathbf{P}(\mathbf{P}^\top \mathbf{P})^{-1} \mathbf{P}^\top \mathbf{i}_k$. See the relevant sections for a discussion of the specific algorithms and their optical implementation.

Rank and sparsity of transport matrices The performance of Krylov subspace methods also depends on the rank and sparsity of the light transport matrix, properties that depend on the projector-camera arrangement, the scene, and—most importantly—the light paths they induce.

Light paths that scatter only once (*e.g.* direct paths, specular indirect paths) produce *sparse high-rank transport matrices*. These light paths interact with only one diffuse point within the scene. Ordinarily, only one specular path connects this diffuse scene point to the camera (*i.e.*, only one camera pixel observes this point), and a single specular path connects the same diffuse point to the projector (*i.e.*, only one projector pixel illuminates this point). This one-to-one correspondence between projector pixels and camera pixels creates a sparse high-rank transport matrix.

Light paths scattered by two or more scene points (*e.g.* diffuse indirect paths) produce *dense low-rank transport matrices*. A diffuse point observed by a camera pixel receives light from many other diffuse scene points lit by different projector pixels; this creates a many-to-many mapping between projector pixels and camera pixels. Moreover, two nearby diffuse points observed by two adjacent camera pixels receive similar contributions of light; as a result, many rows of the transport matrix are linearly dependent on each other.

Consider imaging a convex, diffuse object with the arrangement illustrated in Figure 2.3(a). Since the object is convex, this arrangement produces only direct light paths. As a result, the corresponding transport matrix is high-rank and sparse. Note that displaying a projector pattern directly onto the object will transfer it to the photo in the form of a distorted texture, *i.e.*, it transfers a projector’s high-frequency content to the camera.

Placing a display screen (a diffuser) between the projector and the scene, as illustrated in Figure 2.3(b), produces a very different response for the same diffuse object. Because this arrangement now produces diffuse indirect paths, the transport matrix is low-rank and dense. Displaying a pattern onto a screen transfers only low frequencies to the camera in the form of shading.

General scenes make acquisition much more challenging by bridging the gap between these two extremes, *i.e.*, by reducing rank and sparsity in Figure 2.3(a) and increasing them in Figure 2.3(b). Note that \mathbf{T} would be sparse and high-rank for (b) if the object was a perfect mirror, because light transport would consist of only specular indirect paths.

Arnoldi and GMRES Krylov subspace methods come in many flavors depending on the numerical objective (eigenvalue estimation, solution of linear systems, *etc.*); type of matrix (symmetric, non-symmetric, positive definite, *etc.*); and error tolerance. We explore two of these methods here: Arnoldi iteration and GMRES.

When implemented in optics, Arnoldi and GMRES follow the same basic loop as optical power iteration. They capture a photo, process it, project the result back onto the scene, and repeat for a fixed number of iterations. Both methods differ from power iteration in just three steps. These differences are summarized in Table 2.1.

2.2 Optical Arnoldi for Transport Acquisition

We now show how to efficiently capture low-rank approximations of full-resolution transport matrices with the Arnoldi method.

The Arnoldi method can be thought of as a much more efficient version of power iteration. At the k^{th} iteration, the algorithm has enough information to estimate the top k eigenvectors (or singular vectors) of a matrix rather than just one of them. Unlike power iteration, Arnoldi does not estimate these eigenvectors directly. It generates a sequence of orthogonal vectors, $\mathbf{p}_1, \dots, \mathbf{p}_k$, whose span approximates the span of the top k eigenvectors. This approximation, which tightens as k increases, makes it possible to compute a rank- k approximation of the matrix without direct access to its elements. We refer the reader to Saad [109] for the basic theory behind the Arnoldi method and to [36, 117] for a convergence analysis and detailed error bounds on its rank- k approximation.¹

Algorithm 2 *Optical Arnoldi for rank- K transport acquisition.*

For symmetric \mathbf{T} :

For non-symmetric \mathbf{T} :

In: iterations K

In: iterations K

Out: rank- K approximation of \mathbf{T}

Out: rank- K approximation of \mathbf{T}

1: $\mathbf{p}_1 =$ non-zero vector

$\mathbf{p}_1 =$ non-zero vector

2: **for** $k = 1$ to K

for $k = 1$ to K

3: illuminate with $(\mathbf{p}_k)_+$ and $(\mathbf{p}_k)_-$
capture photos $(\mathbf{i}_k)_+$ and $(\mathbf{i}_k)_-$
 $\mathbf{i}_k = (\mathbf{i}_k)_+ - (\mathbf{i}_k)_-$

left-illuminate with $(\mathbf{p}_k)_+$ and $(\mathbf{p}_k)_-$
right-capture photos $(\mathbf{d}_k)_+$ and $(\mathbf{d}_k)_-$
 $\mathbf{d}_k = (\mathbf{d}_k)_+ - (\mathbf{d}_k)_-$

$\mathbf{r}_k = \mathbf{d}_k / \|\mathbf{d}_k\|_2$

right-illuminate with $(\mathbf{r}_k)_+$ and $(\mathbf{r}_k)_-$
left-capture photos $(\mathbf{s}_k)_+$ and $(\mathbf{s}_k)_-$
 $\mathbf{s}_k = (\mathbf{s}_k)_+ - (\mathbf{s}_k)_-$

4: $\mathbf{p}_{k+1} = \text{ortho}(\mathbf{p}_1, \dots, \mathbf{p}_k, \mathbf{i}_k)$

$\mathbf{p}_{k+1} = \text{ortho}(\mathbf{p}_1, \dots, \mathbf{p}_k, \mathbf{s}_k)$

$\mathbf{p}_{k+1} = \mathbf{p}_{k+1} / \|\mathbf{p}_{k+1}\|_2$

$\mathbf{p}_{k+1} = \mathbf{p}_{k+1} / \|\mathbf{p}_{k+1}\|_2$

5: **return** $[\mathbf{i}_1 \cdots \mathbf{i}_K][\mathbf{p}_1 \cdots \mathbf{p}_K]^T$

return $[\mathbf{d}_1 \cdots \mathbf{d}_K][\mathbf{p}_1 \cdots \mathbf{p}_K]^T$

Two optical implementations of Arnoldi are shown in Algorithm 2, one for symmetric and one for non-symmetric \mathbf{T} . For a symmetric \mathbf{T} , the implementation amounts to substituting three steps in

¹Note that another often-used method, *Lanczos*, is the Arnoldi method optimized for symmetric matrices.

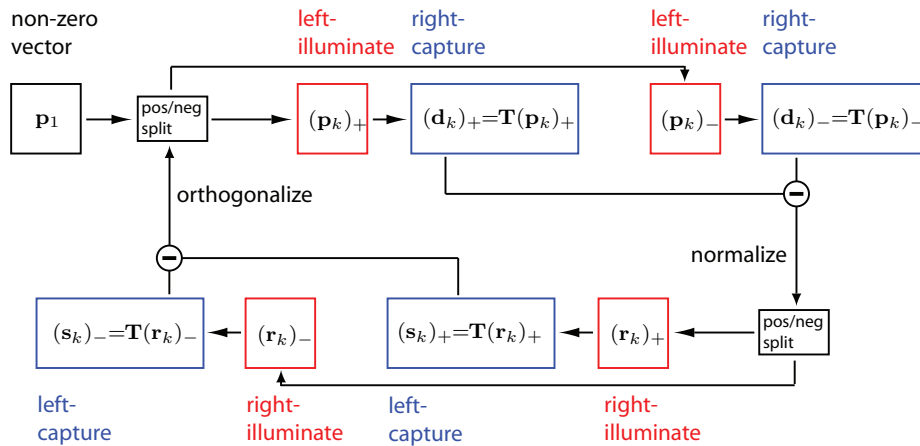


Figure 2.4: Arnoldi iteration with two projectors and two cameras.

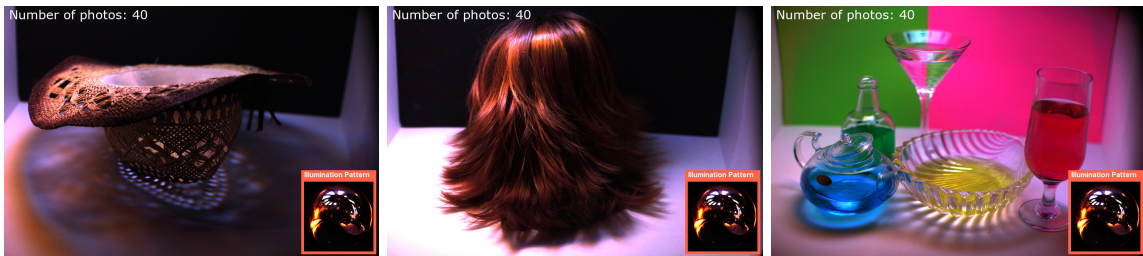


Figure 2.5: **Image-based relighting.** To create the relighting results shown above, just forty photos were used to acquire $700K \times 100K$ transport matrices with the optical Arnoldi algorithm. Note the complex shadows cast by the hat (both sharp and soft), the complex highlights on the hair and the shadows it casts, and the many shadows, caustics and indirect lighting effects in the glass scene. The entire process (photo capture, matrix reconstruction, relighting) took four minutes on a standard PC for each scene.

power iteration’s optical implementation with the corresponding steps from Table 2.1. Because the vectors generated by Arnoldi contain negative values, we use two illuminate-and-capture operations per matrix-vector product, one for their positive and one for their negative component. The optical implementation for non-symmetric transport matrices uses two sets of illuminate-and-capture operations, applied to T and T^T respectively, in order to implement Eq. (2.6). Therefore, we capture four photos per Arnoldi iteration when T is non-symmetric. Figure 2.4 shows the schematic diagram for this case, Figure 2.5 shows relighting results after reconstructing T , and Figure 2.6 shows photos captured during the run.

The optical Arnoldi algorithm can be thought of as building two bases simultaneously: an orthogonal basis for the *subspace of illumination vectors* and a basis for the *subspace of photos*. For a symmetric T , the illumination basis contains all vectors the algorithm used to illuminate the scene; the photo basis contains all photos it captured. In the non-symmetric case, the illumination basis contains the left-illumination vectors (first row of Figure 2.6) and the photo basis the right-captured photos (second row of Figure 2.6). The matrix itself, returned in the algorithm’s last step, just multiplies these two bases together.

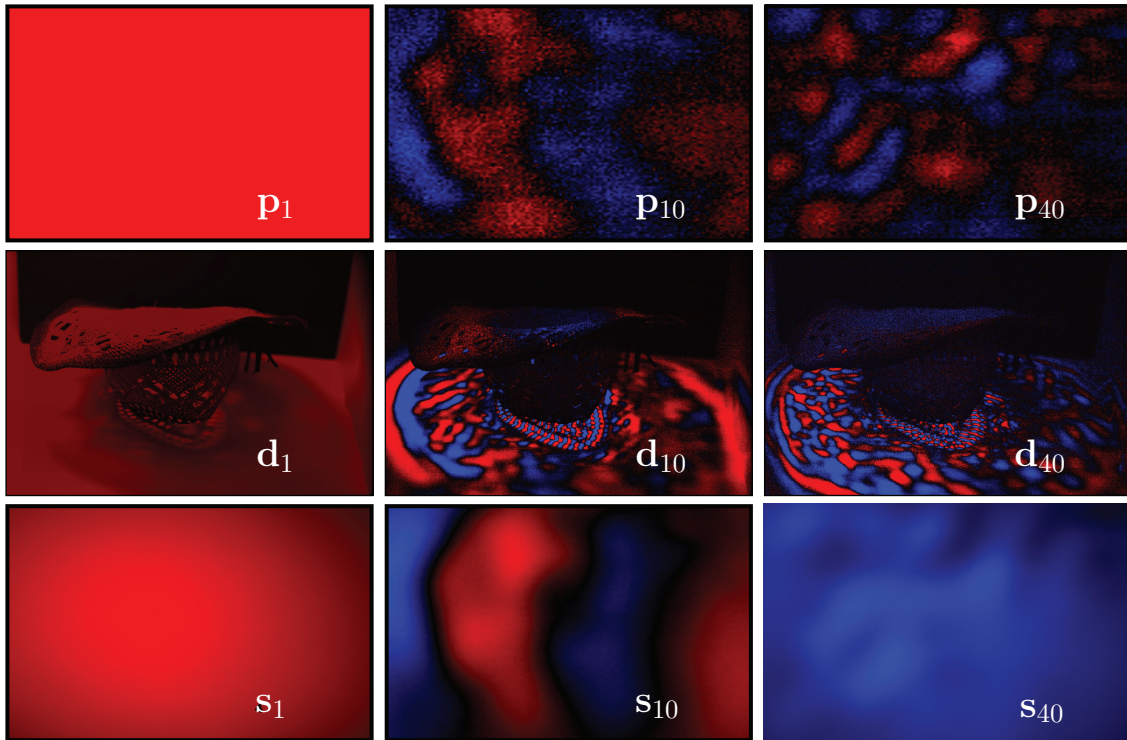


Figure 2.6: **Optical Arnoldi in action** for the hat scene rendered in Figure 2.5. We used the non-symmetric configuration in Figures 2.3(d) and 2.12(a), with the left camera-projector pair behind a display screen. The photo and illumination vectors contain both positive (red) and negative (blue) values. We started with a constant positive left illumination vector (\mathbf{p}_1) so the first photo captured with the right camera (\mathbf{d}_1) was positive as well. This photo was then projected onto the scene with the right projector. Note that since the left camera views the scene from behind the display screen, which is essentially a diffuser, its photos contain mainly low frequencies (\mathbf{s}_1). We produced the relighting result in Figure 2.5 by estimating a low-rank \mathbf{T} from ten Arnoldi iterations (*i.e.*, 40 photos).

Scene relighting To render a scene under a novel illumination vector \mathbf{p} , we rewrite Eq. (1.1) in terms of the captured illumination and photo bases. The equation for the non-symmetric case becomes

$$\mathbf{i} = [\mathbf{d}_1 \cdots \mathbf{d}_K][\mathbf{p}_1 \cdots \mathbf{p}_K]^T \mathbf{p} . \quad (2.7)$$

This equation can be thought of as a two-step relighting procedure: first we compute \mathbf{p} 's coordinates in the left-illumination basis by projecting \mathbf{p} onto it; then we linearly combine the right-captured photos to obtain the relighting result, \mathbf{i} .

2.2.1 Relation to Prior Work on Transport Acquisition

We discuss related work from a numerical perspective in terms of four properties— \mathbf{T} 's rank, sparsity, row space, and symmetry.

An important distinction between methods is the rank and sparsity of matrices they acquire, and

this distinction is implicit in the choice of a light source and scene. Techniques geared toward sparse high-rank matrices [30, 98, 114] rely on \mathbf{T} 's ability to transfer both high and low frequencies from the illumination domain to the camera domain; techniques acquiring dense low-rank matrices [22, 27, 130] assume that high-frequency illumination does not propagate to the camera domain. Optical Arnoldi is primarily applicable to dense low-rank matrices. These are often representative of natural settings, where illumination comes from point or area sources and where mirror reflection and sharp shadows do not dominate light transport.

The choice of illumination ensemble used to acquire the transport matrix is critical because it controls the basis for \mathbf{T} 's row space. To maximize efficiency, this ensemble should allow accurate reconstruction of \mathbf{T} 's rows from as few illumination vectors as possible. Many ensembles have been used for this purpose, including Haar wavelets [97], Hadamard patterns [111] and single-source illuminations [27]. For instance, Wang et al. [130] use the low-rank configuration in Figure 2.3(b) and single-source illumination vectors to reconstruct \mathbf{T} 's rows with the kernel Nyström method. These illumination ensembles have been *scene independent* in all previous work on transport acquisition.² For low-rank matrices, no scene-independent ensemble is optimal. The optimal ensemble under the Frobenius norm is *scene dependent* and consists of \mathbf{T} 's singular vectors [123]. This is precisely the ensemble optical Arnoldi approximates.

Garg et al. [30] and Wang et al. [130] used coaxial camera-projector arrangements to exploit the fact that knowing a subset of both rows and columns of \mathbf{T} makes it easier to reconstruct the rest. Numerically, however, symmetry has much more fundamental effect on a matrix, as it affects its eigenstructure. While we use similar camera-projector arrangements, our choices are guided primarily by numerical convergence considerations.

Sen et al. [115] and Peers et al. [99] recently used compressed sensing techniques to reconstruct individual rows of \mathbf{T} . These methods are complementary to our own, as they apply to a different matrix class—sparse, high-rank matrices—for which a low-rank approximation might lead to rendering artifacts. The scene-independent ensembles of these methods, however, are inefficient for capturing dense low-rank matrices. They are also very expensive computationally and depend on the size of \mathbf{T} . Here, by seeking to maximize the “information content” of each captured photo, optical Arnoldi makes the number of iterations required for convergence dependent on \mathbf{T} 's singular value distribution, not its size.

Computing transport eigenvectors Eigenvectors and singular vectors of real-world transport matrices have been used for compression [81] and to accelerate rendering [79]. In all cases, they were computed after acquiring \mathbf{T} . With optical Arnoldi, we analyze light transport in reverse: we first construct a basis that approximates the span of the top K transport eigenvectors, and then use that basis to reconstruct the matrix.

²Although techniques have been proposed for sampling vectors from *within* an ensemble in a scene-dependent way (e.g., [27]), the ensemble itself is fixed and independent of the scene.

2.3 Optical GMRES for Inverse Transport

We now consider an optical solution to the following problem. We are given a target photo \mathbf{i} and seek an illumination vector \mathbf{p} that produces it. Mathematically, this can be expressed as a solution to Eq. (1.1) where the unknown is \mathbf{p} , not \mathbf{i} ; that is, we want to compute the following:

$$\mathbf{p} = \mathbf{T}^\dagger \mathbf{i} \quad (2.8)$$

where the matrix \mathbf{T}^\dagger is the Moore-Penrose pseudoinverse of a potentially singular matrix \mathbf{T} .

Generalized minimal residual (GMRES) is a Krylov subspace method that iteratively solves this problem for unobservable matrices without inverting them, using just matrix-vector products [109]. As shown in Table 2.1, the method is almost identical to Arnoldi: the only difference is its initial vector (it is always \mathbf{i}) and its return value, which is a solution to the least-squares problem:

$$\mathbf{p} = \arg \min_{\mathbf{x}} \left\| [\mathbf{i}_1 \cdots \mathbf{i}_K] [\mathbf{p}_1 \cdots \mathbf{p}_K]^T \mathbf{x} - \mathbf{i} \right\|_2, \quad (2.9)$$

where \mathbf{i}_k and \mathbf{p}_k are computed in the k^{th} Arnoldi iteration. In essence, GMRES builds a rank- K approximation of \mathbf{T} and then inverts it to compute \mathbf{p} .

Despite its apparent simplicity, GMRES is an extremely powerful algorithm. It applies to any matrix (low-rank, high-rank, dense, sparse, *etc.*) and converges rapidly for arbitrary non-singular symmetric matrices [78]. Intuitively, GMRES does this by “exploring” only a portion of \mathbf{T} ’s row space, *i.e.*, the subspace that is precisely suitable for inverting the initial vector \mathbf{i} .

The optical implementation of GMRES is identical to Arnoldi’s. We simply run optical Arnoldi with a photo \mathbf{i} as the initial illumination vector and, after the algorithm terminates, we solve Eq. (2.9) computationally (Step 5 of GMRES in Table 2.1).³

In principle, it should be possible to use optical GMRES to invert light transport efficiently for any full-rank transport matrix, regardless of sparsity and size. This, for instance, would allow us to infer the illumination that produced a given photo of a scene, even when both the scene and the illumination are very complex. Figures 2.7, 2.8 and Section 2.5.2 show initial demonstrations of such a capability on both high-rank and low-rank matrices, at high resolution.

For singular transport matrices two possibilities exist: there may be many different illuminations that can produce a given photo (Eq. (1.1) has multiple solutions) or none at all (Eq. (1.1) is infeasible). When many solutions exist, optical GMRES can efficiently return one of them, although not necessarily the one used to produce the original photo. When no solutions exist, it will return the best-possible approximation lying within the rank- K subspace it built.

2.3.1 Relation to Prior Work on Inverse Transport

Implicit in all prior work is the assumption that to invert light transport we must first acquire \mathbf{T} . Grossberg et al. [37] consider this problem under the assumption of no indirect transport. This

³ The solution to Eq. (2.9) may contain negative elements. We just clamp them to zero, at the cost of increasing residual errors. Another possible approach is to add a non-negativity constraint and solve the equation using MATLAB’s `lsqnonneg()` function.

makes \mathbf{T} maximally sparse and full rank, and therefore easy to capture and invert. To deal with non-negligible (but still small) indirect illumination, Wetzstein and Bimber [132] capture the full-resolution \mathbf{T} , enforce sparsity by thresholding, and invert it with a sparse-system solver. For more significant indirect lighting effects, Ng et al. [92] and Bai et al. [7] rely either on low-resolution or on sparse transport matrices to make brute-force acquisition tractable. In other applications, Zhang and Nayar [142] compute \mathbf{T}^\dagger to account for projector defocus and Seitz et al. [112] use it to decompose an image of a diffuse scene into the contribution of individual light bounces. Both approaches rely on low-resolution transport matrices and brute-force acquisition.

Here we show that acquiring \mathbf{T} is not necessary as long as we pursue a more modest goal—rather than solve inverse light transport for *every* possible input photo, which is equivalent to computing \mathbf{T}^\dagger and hard, we use optical GMRES to solve it for *specific* ones, which is much easier. Moreover, GMRES can be generalized to handle many input vectors/photos at a time [45].

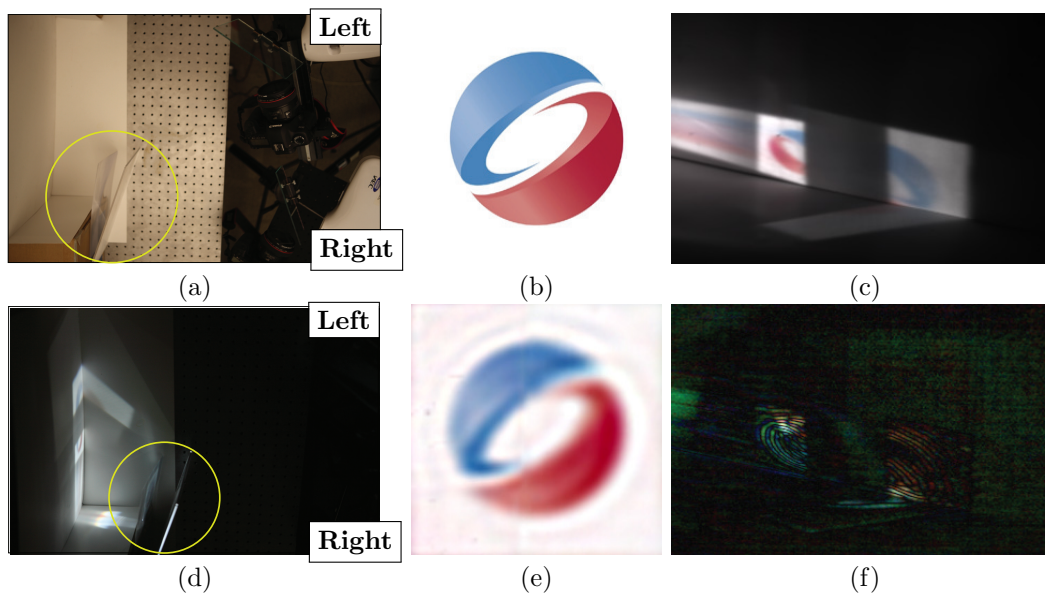


Figure 2.7: **Inverse light transport for a high-rank \mathbf{T} .** A top view of our setup is shown in (a). We used two camera-projector pairs to create the arrangement of Figure 2.3(d), and placed a fresnel lens and an LCD backlight diffuser (both circled) in front of the right pair. We then displayed the SIGGRAPH logo (b) onto the scene with the right projector and captured a photo with the left camera. A photo of the scene under this illumination is shown in (c) and the top view of the scene in (d). Note the evidence of complex low-rank and high-rank transport in this photo (diffused regions and distorted/replicated/broken-up logo, respectively). We then provided the photo in (c)—with no additional information—to optical GMRES to test its ability to reassemble the logo we actually displayed. The result of optical GMRES is shown in (e): the logo is reconstructed successfully although some ringing exists. It took 20 iterations of optical GMRES, *i.e.*, 80 photos, to get this result. Displaying the image in (e) onto the scene with the right projector produces a left-camera photo that is indistinguishable from the one in (c). The difference between those two photos, amplified $50\times$, is shown in (f). This suggests \mathbf{T} is mildly singular, confirming the ringing.

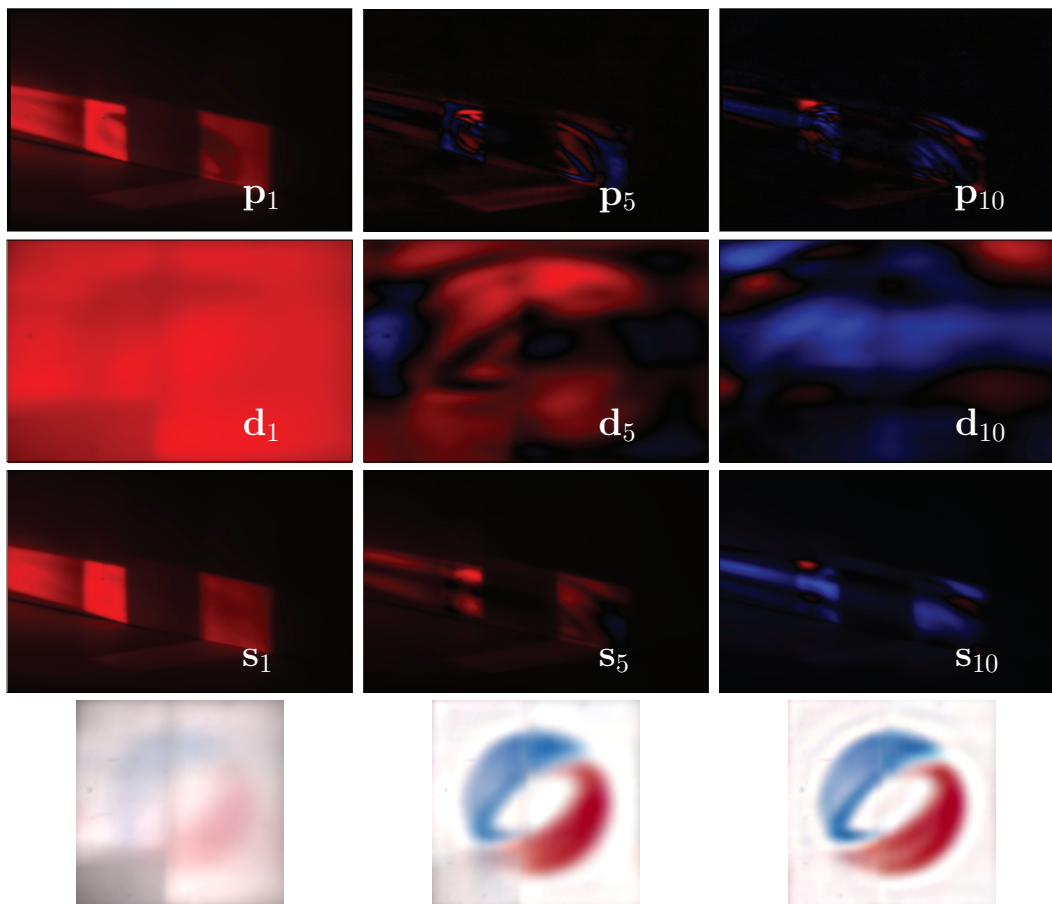


Figure 2.8: **Optical GMRES in action** for the inverse transport task of Figure 2.7. Although GMRES operates in full color, we show only the red channel with its positive (red) and negative (blue) components. The algorithm begins by displaying the input photo in Figure 2.7(d) with the left projector. The last row shows the GMRES solution after the corresponding iteration.

2.4 Implementation

Cameras and projectors We use either one or two coaxial camera-projector pairs, each composed of an NEC LT30 DLP projector, a beamsplitter, and a Canon EOS 5D Mark II camera with a 50 mm $f/1.2$ lens, chosen for its light-gathering abilities. The projectors are calibrated geometrically and radiometrically relative to the cameras. See Figures 2.7(a) and 2.12(a) for typical arrangements.

We operate each projector at 85 Hz and capture photos with a $1/13$ s exposure time at ISO 100. The only exception is when taking photos of the display screen, which is typically quite dim, where we use ISO 200 instead (*e.g.*, for the left camera in Figure 2.12(a)).⁴ We capture only low dynamic range photos and use a fixed exposure level for each run of our algorithms. To control it, we adjust the lens aperture and the maximum intensity of projector pixels so that photos taken under maximum illumination have an intensity at most 80% of the maximum possible. Beyond arranging

⁴To increase contrast even further for photos of the screen, one can block the aperture of the “inactive” projector (*i.e.*, left projector in Figure 2.12(a)).

the cameras and projectors; adjusting exposure; choosing a photo resolution; and choosing the number K of Arnoldi iterations, there are no user-tunable parameters in our implementation. To minimize defocus in all experiments, we kept the scene within the camera’s depth of field as much as possible.

Optical Arnoldi Numerical packages for large matrices typically allow the matrix-vector calculation function to be passed as a parameter. To acquire \mathbf{T} , we simply pass a function that implements matrix-vector products with two illuminate-and-capture operations. For example, acquisition of a symmetric \mathbf{T} with the *PROPACK* package [74] for *MATLAB* amounts to one line of code:

```
[V, D] = LANEIG(opt-prod, I, K, "LM")
```

where *opt-prod()* is a function that computes optical vector products with \mathbf{T} and *LANEIG()* solves the eigenvalue problem⁵ $\mathbf{T}\mathbf{v} = \lambda\mathbf{v}$. The implementation for a non-symmetric \mathbf{T} is similar:

```
[U, S, V] = LANSVD(opt-prod, opt-prod-t, P, I, K, "L")
```

where *opt-prod-t()* computes optical vector products with \mathbf{T}^T and *LANSVD()* performs singular value decomposition.

Optical matrix-vector products Our implementation essentially amounts to implementing the *opt-prod()* and *opt-prod-t()* functions for use in *MATLAB*. Given an illumination vector \mathbf{p} supplied by *MATLAB*, we reshape it into a 2D image and divide it by the maximum absolute pixel value to bring it to the range $[-1, 1]$. We then implement Eq. (2.4) directly, by splitting the vector into its positive and negative components; multiplying each component by the maximum intensity of projector pixels; capturing one photo per displayed component; and taking the difference of the two photos. The differencing operation removes any contributions from ambient light. We multiply the result with the maximum absolute pixel value of \mathbf{p} .

Since demosaicing large RAW photos is computationally expensive, we capture and store them on disk but do not return them as vectors to *MATLAB*. Instead, we return the 362×234 thumbnail images contained in the RAW files, which are linear, demosaiced, and small enough to fit in main memory. Steps 3 and 4 of optical Arnoldi and GMRES thus operate on a $(362 \times 234)^2$ version of \mathbf{T} .

High-resolution acquisition and relighting Even though the illumination vectors in Steps 3 and 4 of optical Arnoldi are computed from low-resolution thumbnails, the RAW photos we capture do enable high-resolution acquisition and relighting. For relighting in the non-symmetric case, we simply replace in Eq. (2.7) the low-resolution basis for \mathbf{T} ’s photo subspace with a high-resolution one:

$$\mathbf{i} = [\mathbf{d}_1^h \cdots \mathbf{d}_K^h][\mathbf{p}_1 \cdots \mathbf{p}_K]^T \mathbf{p} \quad (2.10)$$

where \mathbf{d}_k^h is the RAW photo captured in the k^{th} Arnoldi iteration. To reconstruct a high-resolution \mathbf{T} , we replace the matrix returned in Step 5 of the algorithm with the matrix in Eq. (2.10).⁶

⁵Strictly speaking, *LANEIG()* and *LANSVD()* use the Lanczos method.

⁶While thumbnails make each Arnoldi iteration more efficient, computing illumination vectors directly from demosaiced RAW photos may lead to lower reconstruction error for a given number of iterations.

Color handling Mismatches in the spectral response of projectors and cameras can lead to cross-talk between color channels. To avoid such issues in the Arnoldi method, the scene is lit with grayscale illumination patterns by converting each illumination vector to grayscale prior to projection. For GMRES, we capture each color channel with a separate illuminate-and-capture operation.

Timings Our entire relighting pipeline consists of running optical Arnoldi for K iterations, demosaicing the high-resolution photos it captured, and computing the relit images with Eq. (2.10). We use an Apple iMac with a 2.8 GHz Intel Core i7 processor and 16 GB of memory for all capture, processing and display. Optical Arnoldi for a non-symmetric \mathbf{T} runs at a rate of 12 seconds per iteration, including all processing and photo acquisition. Demosaicing also takes 12 seconds to process the positive and negative components of \mathbf{d}_k captured in each iteration. We use an image resolution of 1080×720 for all relighting results, with each image taking $0.3 \times K$ seconds to compute.

2.5 Results

2.5.1 Acquiring Transport Matrices with Optical Arnoldi

Ground-truth comparisons Although optical Arnoldi computes rank- K matrix approximations at the resolution of cameras and projectors, it is not possible to capture ground-truth matrices of that size for evaluation. To validate our algorithm and compare against other low-rank approximation methods, we acquired low-resolution transport matrices for the four scenes shown in Figure 2.9. We deliberately chose these scenes to exhibit a variety of challenging lighting effects, including shadows, inter-reflections, caustics, *etc.*

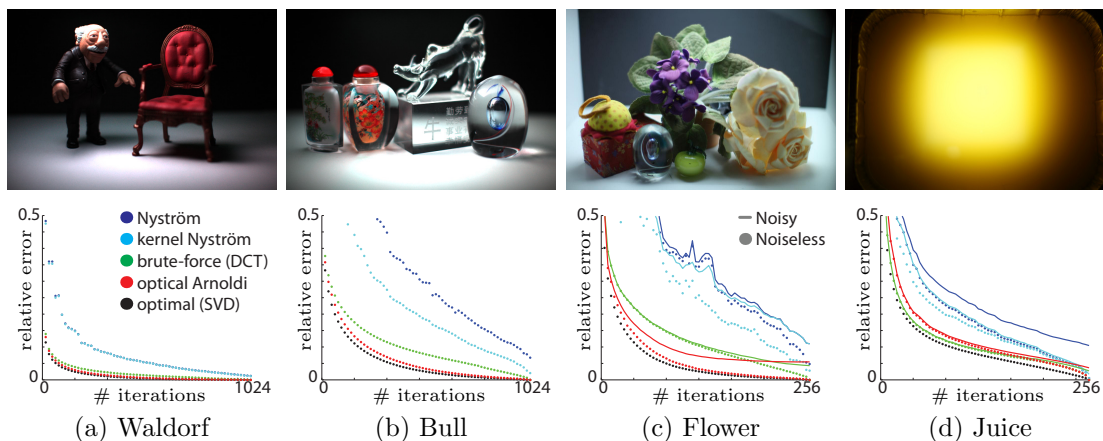


Figure 2.9: Error comparisons for five methods applied to scenes shown above. We used the non-symmetric arrangement in Figure 2.12(a) for the Waldorf, Bull and Flower scenes. Since the left projector was behind a diffuser, the transport matrix for these scenes was dense; the SVD error profiles suggest the rank is not very low for Bull and Flower because of mirror reflections from the glass objects in those scenes. For Juice, we used the symmetric arrangement in Figure 2.3(c), with light projected directly onto the scene. The transport matrix in this case was 256×256 and, as the SVD error profile indicates, has a high rank.

To capture the ground-truth matrix, we treated the left projector as a low-resolution display with 32×32 or 16×16 superpixels. We used the full set of Hadamard illumination patterns because of their good noise properties [111] and captured photos at thumbnail resolution (362×234).

The accuracy and efficiency of a particular acquisition technique can be measured by how well the ground-truth matrix can be reconstructed after a certain number of steps. We compute the relative reconstruction error under the Frobenius norm after k iterations:

$$\epsilon_k = \frac{\|\mathbf{T}_k - \hat{\mathbf{T}}\|_F}{\|\hat{\mathbf{T}}\|_F}, \quad (2.11)$$

where $\hat{\mathbf{T}}$ is the ground-truth matrix and \mathbf{T}_k is the transport matrix computable after k iterations. We report this error for \mathbf{T}_k computed five different ways:

- running k iterations of optical Arnoldi;
- acquiring k random rows and columns of the transport matrix and then using the Nyström method of Wang et al. [130];
- applying the kernel Nyström method [130] to the same input, with its exponent parameter chosen to minimize ϵ_k (this represents the best-case scenario for that method);
- acquiring k photos under the discrete cosine transform (DCT) illumination basis (*i.e.*, a brute-force technique);
- computing the best rank- k approximation of $\hat{\mathbf{T}}$ numerically using singular value decomposition (this is optimal under the Frobenius norm).

To assess the relative performance of these methods in the absence of measurement noise, we simulated photo acquisition for the four scenes shown in Figure 2.9. Specifically, instead of capturing more photos as needed by each method, we used the ground-truth transport matrix to compute them.

The plots in Figure 2.9 show that optical Arnoldi’s convergence rate under noiseless conditions is close to optimal, and significantly faster than the Nyström methods. Note that it should be possible, in principle, to reconstruct $\hat{\mathbf{T}}$ perfectly once the number of iterations becomes equal to the number of rows in $\hat{\mathbf{T}}$ (*i.e.*, 1024). We found that this was not the case with Nyström because it depends on a pseudoinverse tolerance parameter that does not guarantee perfect reconstruction when $\hat{\mathbf{T}}$ is singular. This is the case in the *Waldorf* and *Bull* scenes, as can be seen by SVD’s error profile. Finally, note that brute-force acquisition can outperform other methods in some cases (*e.g.*, *Juice*).

Figure 2.9 also shows comparisons under noisy conditions for the *Flower* and *Juice* scenes. In this case, all photos were captured rather than simulated. Since the single-source illuminations required by the Nyström methods (*i.e.*, only one superpixel is on) produce very dim images, we doubled the exposure time of the photos provided to Nyström. Thus, the total exposure time in Figure 2.9 is the same after k iterations for the Nyström methods and for optical Arnoldi (the latter captures twice as many photos per iteration).⁷ Despite the increase in the per-photo exposure time

⁷We consider constant total exposure time a more accurate baseline, given the different illumination conditions required by the Nyström methods. See Figure 2.10 for a comparison with a constant number of photos.

for the Nyström methods, their input images are still very noisy. Arnoldi, on the other hand, uses illumination vectors with much broader spatial support (*e.g.*, see Figure 2.6) and therefore has some of the advantages of the Hadamard patterns.

Since noise is present in all photos of *Flower* and *Juice*—including those we used for ground-truth capture—perfect reconstruction of $\hat{\mathbf{T}}$ cannot be guaranteed. Indeed, no method other than SVD, which depends only on $\hat{\mathbf{T}}$, achieves perfect reconstruction. As in the noiseless case, Arnoldi performs substantially better than the Nyström methods. Its convergence rate resembles that of SVD, although it converges to a non-zero reconstruction error. This is expected because it computes matrix-vector products with noisy 8-bit camera pixels rather than noiseless floating point arithmetic.

Relighting results and discussion Figure 2.10 shows relighting results for a point light source, and Figures 2.5, 2.11 show results for several more scenes. Unlike the ground-truth experiments, these scenes are relit with high-resolution, $(362 \times 234) \times (1080 \times 720)$ transport matrices.

These results suggest that it is possible to create good-quality relit images from a small number of captured photos, even in the presence of complex light transport phenomena. Clearly, since we acquire a rank- K approximation of the transport matrix from $4K$ captured photos, high-rank transport components cannot be acquired if K is very small. This is evident in the second row of the *Hat* scene in Figure 2.11, where a rank-10 approximation cannot reproduce the light passing through holes in the hat’s brim. These are reproduced much better with a rank-50 matrix. The *Crystal* scene in the same figure is more challenging, with numerous mirror reflections and light transported to the backdrop in complex ways. The result is a significant difference between the rank-10 and ground-truth photos.

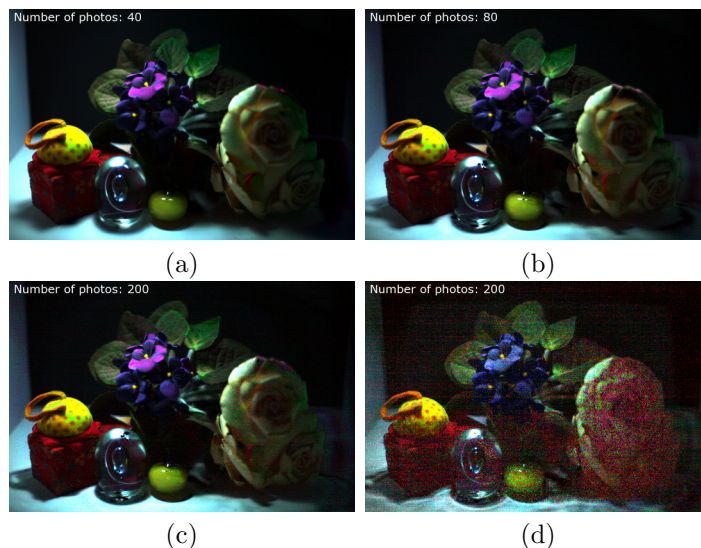


Figure 2.10: Relighting results for the Flower scene in Figure 2.9 under point-source illumination. We computed \mathbf{T} after (a) 10, (b) 20, and (c) 50 iterations of optical Arnoldi (*i.e.*, 200 photos, 200/13 s total exposure time). In (d) we show an image relit with the matrix \mathbf{T} computed by kernel Nyström from 100 row and 100 column samples (200 photos, 400/13 s total exposure time).

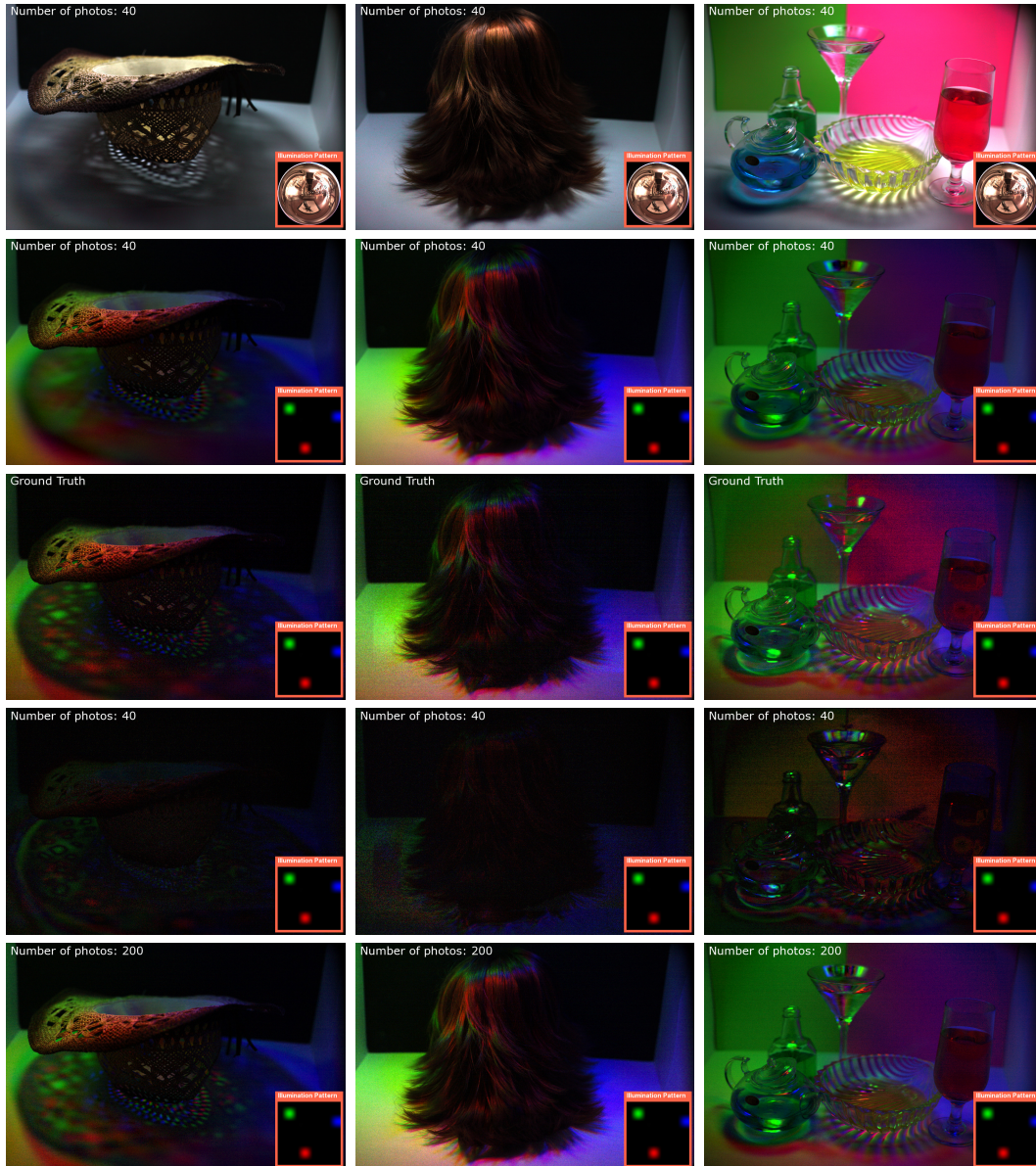


Figure 2.11: **First row:** Relighting results under complex illumination with a rank-10 transport matrix. **Second row:** We use spatially-localized lighting and a rank-10 matrix to identify inaccuracies in high-rank light transport. **Third row:** Actual photo of the scene captured under the same illumination with identical exposure settings. **Fourth row:** Difference between actual photo and the relit image above it. **Fifth row:** Relighting with a rank-50 matrix.

An interesting property of optical Arnoldi is that even when the relit images are not entirely correct, as in the *Crystal* scene, they are still visually plausible. This is mainly due to the scene-dependent basis. All relit images are linear combinations of photos of the scene itself.

We deliberately did not use high dynamic range photos for experiments. This makes noise apparent in some of our photos (both real and rendered). The purpose was to illustrate the performance of the optical-domain approach in a challenging setting. Cameras and projectors have limited dynamic range, are subject to noise, and represent numbers with limited precision. As such, optical matrix-vector products cannot be considered as accurate as matrix-vector products done with floating-point arithmetic. Nevertheless, we observe that despite these unavoidable issues, the presence of noise does not cause additional rendering artifacts. For example, the ground-truth photos in Figure 2.11, which were captured with the same exposure settings as Arnoldi’s, appear noisier than the rendered ones. The good behavior of optical Arnoldi comes from the overall stability of Krylov subspace methods in the presence of matrix-vector product errors [118].

2.5.2 Inverting Light Transport with Optical GMRES

The experiment in Figures 2.7 and 2.8 showed the ability of GMRES to invert light transport at high resolution for a high-rank \mathbf{T} . To test GMRES for a low-rank \mathbf{T} , we consider the problem of creating an “illumination impostor”, *i.e.*, using one light source (a projector) to simulate the complex lighting effects produced by another (a flashlight). This experiment is described in Figure 2.12.

The experiment shows that GMRES is able to invert light transport at high resolution in a very challenging setting, with complex occlusions, shadows and inter-reflections. This would not have been possible with existing methods without capturing the full \mathbf{T} first. Note that even though the illumination that GMRES estimates is similar to the one we used, it is not exactly the same. The photo of the scene it produces, however, matches that of the flashlight-illuminated scene very well. This means that \mathbf{T} has a whole subspace of feasible illuminations that produce the same photo—and GMRES just found one of them.

2.6 Relation to Work in Optical Computing

Numerical linear algebra with optics In discrete analog optical processing, matrices and vectors are encoded by transparency masks or spatial light modulators [35]. Athale and Collins [5] showed how a sequential arrangement of such elements (along with light sources, lenses and detectors) can be used to implement matrix multiplication optically.

For more complex operations, the idea of using optical feedback loops for processing was well established [10]. Kumar and Casasent [72] describe several optical feedback loops for computing eigenvectors of a matrix, including power iteration. Rajbenbach et al. [105] computed matrix inverses optically by implementing the Neumann series in a feedback loop. Our approach can be thought of as a re-interpretation of this general approach, where the matrix is encoded implicitly by the scene and where all other optical elements are replaced by cameras, projectors and a PC. In this respect, it is a hybrid system [9].

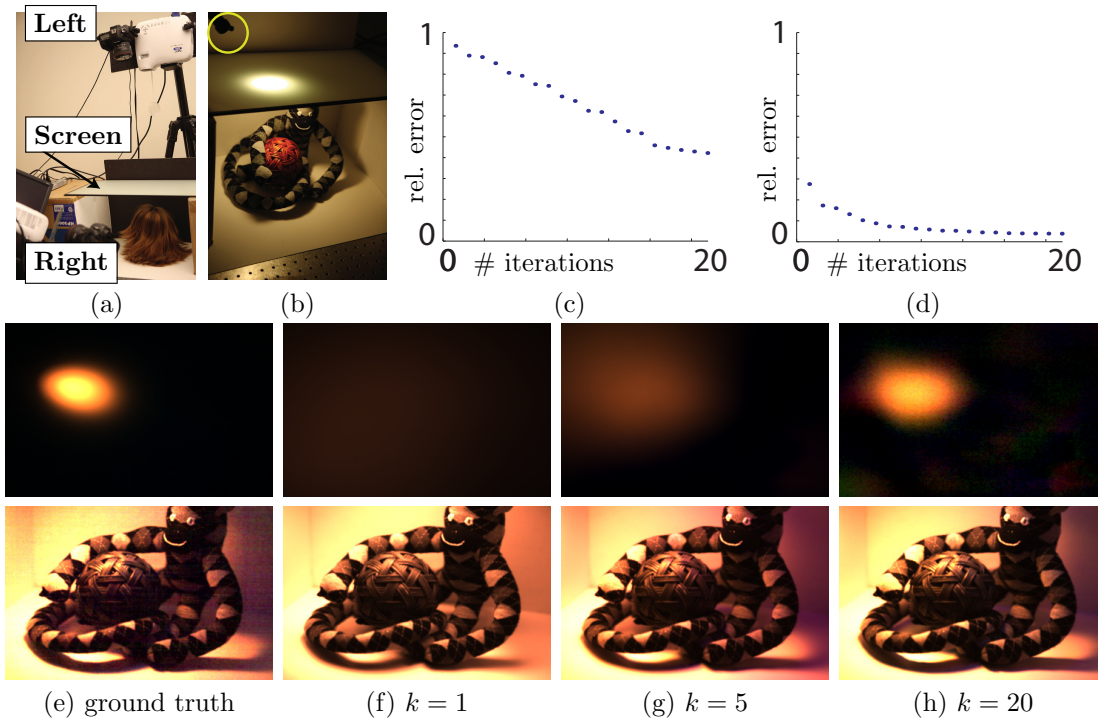


Figure 2.12: **Simulating one light source with another.** We used a non-symmetric arrangement with a display screen between the left camera and the scene (a). We then placed a toy on the other side (b) and used a flashlight (circled) to illuminate it. We captured a photo with the right camera (e, bottom) and gave it as input to optical GMRES. The goal was to get the left projector to illuminate the scene exactly like the flashlight. A left-camera photo of the display screen under flashlight illumination is shown in (e, top). This photo was not given to GMRES. It took GMRES 20 iterations to compute it (h, top). Note how the illumination computed by GMRES correctly replicates the shadows in the input photo, both hard and soft. The plot (c) shows the relative ℓ_2 reconstruction errors of the flashlight illumination (top of (e)-(h)) and plot (d) shows the errors for the input photo (bottom of (e)-(h)).

We refer readers interested in exploring this area to the review articles of Leith [75] and Ambs [4] as well as the critical analysis of Psaltis and Athale [104] on the feasibility of using linear optics for numerical calculations.

Context-aware illumination Wang et al. [131] recently used the coaxial arrangement in Figure 2.3(c) to enhance the appearance of a real-world scene in real time. Their idea is to illuminate the scene with a processed version of the live video in which certain features, like edges, have been enhanced. Their approach cannot be considered a form of optical computation because it explicitly prevents feedback loops. No photo is captured when the scene is illuminated by their context-aware light source.

Perhaps closer in spirit to our work is the dynamic contrast enhancement technique of Amano and Kato [3]. To enhance the contrast of a real-world planar scene under room illumination, they capture a photo, contrast-enhance it, and project it back onto the scene in a simple feedback loop. Contrary to our work, their analysis is done per pixel and assumes no indirect light transport.

2.7 Inexact Krylov Subspace Methods

As demonstrated empirically in Figure 2.9, sensor noise negatively affects the convergence properties of Krylov subspace methods. This is because the optical implementation of the matrix-vector product cannot be computed exactly.

Although the effect of noise on Krylov subspace methods is beyond the scope of this work, we note that Simoncini and Szyld [119] developed a general framework for understanding the properties of Krylov subspace methods under inexact arithmetic. Instead of exact matrix-vector products as in Eq. (1.1), the theory assumes each step of a Krylov subspace method computes the following product:

$$\mathbf{i} = (\mathbf{T} + \mathbf{E})\mathbf{p} \quad (2.12)$$

where the error matrix \mathbf{E} changes every time a matrix-vector product is evaluated. Krylov subspace methods are remarkably robust with respect to inexact matrix-vector products; the theory explains that the norm of the error, $\|\mathbf{E}\|$, can even grow as the Krylov iteration progresses without degrading the solution, and provides computable criteria to bound $\|\mathbf{E}\|$ at each step so that the residual norm of the solution falls below some tolerance.

2.8 Summary and Contributions

We introduced the idea of performing numerical computations on the full-resolution transport matrix in the optical domain. We have shown the following:

- computations often assumed to require complete acquisition of the matrix \mathbf{T} —such as computing eigenvectors or inverting light transport—can actually be done efficiently without it;
- numerical methods are widely available for use in the optical domain; and
- these methods also bring with them a wealth of numerical research that can potentially be transferred to the optical domain as well.

Our efforts were specifically aimed at low-rank approximation and inverse light transport. Although we believe the low-rank constraint can be a powerful one, clearly many transport phenomena cannot be represented by low-rank matrices. The question of how to bridge the gap between the low-rank constraint we exploit and the sparsity constraints employed recently [99] is wide open. One advantage of optical GMRES (at least for inverse transport) is that it can operate regardless of matrix rank, a feature that also applies to the optical computing technique presented in the next chapter.

Chapter 3

Optical Probing by Primal-Dual Coding

We observe that a much more general imaging regime exists for active illumination, one that operates *almost exclusively* in the optical domain. This regime, which we call *optical probing*, is an extension of *matrix probing*, a topic in numerical mathematics concerned with efficiently estimating the trace or diagonal of very large unobservable matrices of any rank, like \mathbf{T} [8, 121]. Optical probing enables direct acquisition of photos that seem to violate the scene’s light transport equation (*e.g.* the ability to capture a photo of the diagonal elements of matrix \mathbf{T}). These photos have a bilinear, rather than linear, relationship to conventional photos and appear as though the transport matrix of the scene itself has been manipulated. Unlike techniques that rely on exotic technologies [68], optical probing is realized with standard components (*i.e.*, a camera, a projector, and a controllable pixel mask).

More specifically, optical probing acquires photos governed by the following *transport probing equation* (Figure 3.1):

$$\mathbf{i} = (\mathbf{\Pi} \odot \mathbf{T}) \mathbf{1} \quad (3.1)$$

where the symbol \odot denotes the element-wise multiplication of two equal-sized matrices. According to this equation, a photo is formed by multiplying element-wise the scene’s transport matrix with a *probing matrix* $\mathbf{\Pi}$ (with unitless elements) and then multiplying the result with a constant vector of all ones.¹ Intuitively, the probing matrix can be thought of as a generalized illumination condition: it is under the complete control of the image acquisition process, it is known, and can be arbitrary. In contrast to conventional active illumination where there are P degrees of freedom for controlling the output photo (*i.e.*, the size of the illumination vector \mathbf{p}), the probing matrix has $I \times P$ degrees of freedom (the number of elements it contains). This gives tremendous flexibility over image acquisition, well beyond what is implied by the light transport equation.

To achieve this, optical probing exerts simultaneous control over two aspects of the image for-

¹The probing matrix as defined in Eq. (3.1) describes the transfer of radiant energy in relative terms, and does not describe the absolute amount of radiant energy transmitted to the photo. In Chapter 5, we introduce a scalar γ , measured in joules, that scales the unitless elements of $\mathbf{\Pi}$ to ones that represent absolute radiant energy, *i.e.*, $\gamma\mathbf{\Pi}$. Until then, we simply assume $\gamma = 1$ and omit this scalar entirely.

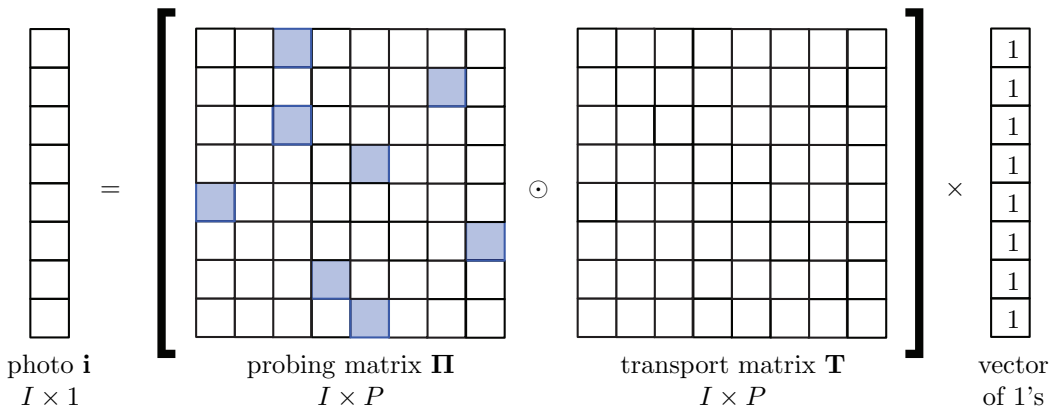


Figure 3.1: The transport probing equation.

mation process: it controls the scene’s illumination (the *primal domain*) and it modulates the light arriving at individual pixels on the camera’s sensor (the *dual domain*). To acquire a photo, we project a sequence of patterns onto the scene while the sensor is exposed to light. At the same time, a second sequence of patterns, derived from the first and applied in lockstep, modulates the light received at individual sensor pixels. Capturing a photo by optical probing therefore involves two basic choices—what primal-domain patterns to use for scene illumination and how to convert them into dual-domain patterns for modulating the sensor.

Optical probing is closely related to optical imaging techniques used in microscopy, namely confocal imaging [15], aperture correlation [134] and differential spinning disk microscopy [84]. These widely-employed techniques increase the optical resolution and contrast of microscopes by eliminating out-of-focus light. They are a special case of optical probing where the primal- and dual-domain patterns are binary, and either coincide or are complements of each other. To our knowledge, they have not been used for one-shot imaging in standard photography settings, although “synthetic” implementations of aperture correlation have been demonstrated by Levoy and colleagues [26, 76]. These implementations keep optical processing to a minimum and rely on substantial image acquisition and computational processing to synthesize individual photos. In contrast, we rely almost exclusively on optical processing; analyze the general case, where primal and dual codes may differ; and introduce the transport probing equation as a general image formation model for optical probing. This equation characterizes the space of possible photos and their relationship to a scene’s transport matrix.

Optical probing can also be thought of as a combination of illumination coding, which operates exclusively in the primal domain [111, 126] and coded-exposure photography, which operates in the dual [55, 89, 133]. These coding techniques have been applied very successfully in recent years but have not been combined for one-shot photo acquisition. Like many coded-exposure methods, we use relay optics and a display panel to modulate the light arriving at sensor pixels without having to access the camera’s interior.

As an initial feasibility test of the optical probing framework, we present results from a prototype system where a camera and a projector share the same viewpoint. We show how to capture photos

where the direct component of light transport is missing, and where components of direct or indirect illumination are selectively attenuated or enhanced. Compared to recent primal-domain methods with similar goals [90, 112, 142], a major advantage of our approach is that it is independent of the frequency content of light transport, the reflectance of the scene’s surfaces, and the rank of its transport matrix. It thus applies to fully-general scenes where a mixture of indirect transport phenomena may occur—including diffuse inter-reflections, specular reflections, caustics, sub-surface scattering, *etc.*

This chapter’s investigation into optical probing is limited in two respects, however, one conceptual and one practical. On the conceptual side, while optical probing and the transport probing equation apply to any camera-projector arrangement, all our experimental results rely on the camera and projector being coaxial. From a practical standpoint, our initial prototype system is built with off-the-shelf hardware, significantly restricting the rate by which primal- and dual-domain patterns change; we compensated by capturing photos with an exposure of several seconds, with all the disadvantages this entails (slightly-increased noise, static scenes, *etc.*).

3.1 Probing Light Transport

We begin by taking a closer look at the transport probing equation and its relation to light transport analysis. The question of how to implement it is considered in the following sections.

In a typical photography setting we may know very little—or nothing at all—about how the scene transports incident light. This black-box view of the scene and of its transport matrix leaves just one way to get information about them, *i.e.*, by capturing one or more photos. Unfortunately, there is a fundamental dimensionality gap between the transport matrix \mathbf{T} and the photos it can produce: the transport matrix has $I \times P$ elements whereas each photo can only provide I measurements.² We therefore have an extremely limited bandwidth for acquiring information about \mathbf{T} . The key advantage of the transport probing equation is the control it offers over how the elements of \mathbf{T} are mapped onto a photo’s limited pixels.

To gain further intuition about the space of allowable mappings, it is helpful to consider the probing equation in the special case when the probing matrix is binary and has only one non-zero element per row (dark squares in Figure 3.1). Multiplying element-wise such a matrix with \mathbf{T} leaves only one non-zero element on each row, equal to the corresponding element of the transport matrix. Multiplication with a constant vector then transfers it unaltered to the photo. Thus, by choosing different binary probing matrices we can directly acquire different I -dimensional slices of the $I \times P$ transport matrix. Table 3.1 shows several such examples, along with other probing operations made possible with non-binary probing matrices.

²This gap is even larger when considering physical properties beyond just radiant energy, such as polarization state [31], phase [102], and time-of-flight [68].

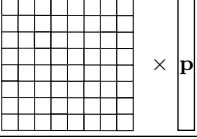
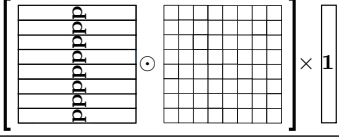
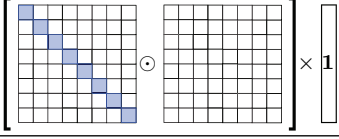
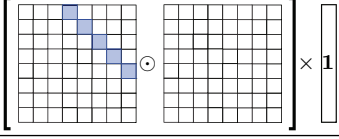
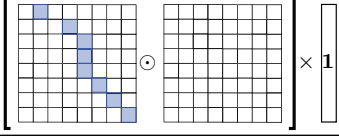
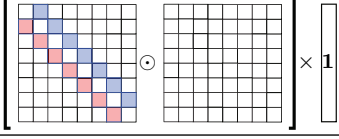
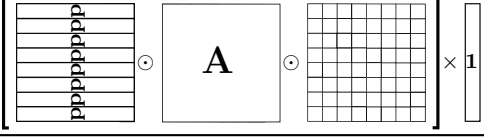
Transport Probing Equation	Expression for $\Pi[m, n]$	Expression for $\mathbf{i}[m]$
		$\sum_n \mathbf{T}[m, n] \mathbf{p}[n]$
	$\mathbf{p}[n]$	$\sum_n \mathbf{\Pi}[m, n] \mathbf{T}[m, n]$ $= \sum_n \mathbf{p}[n] \mathbf{T}[m, n]$
	$\delta(m - n)$	$\sum_n \mathbf{\Pi}[m, n] \mathbf{T}[m, n]$ $= \mathbf{T}[m, m]$
	$\delta(m - n + w)$	$\mathbf{T}[m, (m + w)]$
	$\delta(\mathbf{a}[m] - n)$	$\mathbf{T}[m, \mathbf{a}[m]]$
	$a \delta(\mathbf{a}[m] - n)$ $+ b \delta(\mathbf{b}[m] - n)$	$a \mathbf{T}[m, \mathbf{a}[m]]$ $+ b \mathbf{T}[m, \mathbf{b}[m]]$
	$\mathbf{A}[m, n] \mathbf{p}[n]$	$\sum_n \mathbf{A}[m, n] \mathbf{T}[m, n] \mathbf{p}[n]$

Table 3.1: Probing the transport matrix. Probing makes it possible to compute general linear mappings from transport matrix elements to camera pixels without having to capture (or even approximate) the full transport matrix first. **Row 1:** Capture a conventional photo under illumination \mathbf{p} (Eq. (1.1)). **Row 2:** Capture a conventional photo under illumination \mathbf{p} (Eq. (3.1)). Rows 1 and 2 coincide when $\mathbf{\Pi}[m, n] = \mathbf{p}[n]$. **Rows 3:** Capture the diagonal of \mathbf{T} . Note that the function $\delta(\cdot)$ is the discretized version of the Dirac delta function, *i.e.*, $\delta(x) = 1$ for $x = 0$, and $\delta(x) = 0$ otherwise. **Rows 4:** Capture the w^{th} off-diagonal of \mathbf{T} . **Rows 5:** Sample one element from each row of \mathbf{T} according to a given sampling vector \mathbf{a} . **Row 6:** Capture a linear combination of two elements in each row, according to sampling vectors \mathbf{a} and \mathbf{b} . **Rows 7:** Capture a photo that simulates conventional photography under illumination \mathbf{p} of a scene whose transport matrix is $\hat{\mathbf{T}} = \mathbf{A} \odot \mathbf{T}$, where \mathbf{A} is another user-defined probing matrix (*e.g.* one of the probing matrices from Rows 1-6).

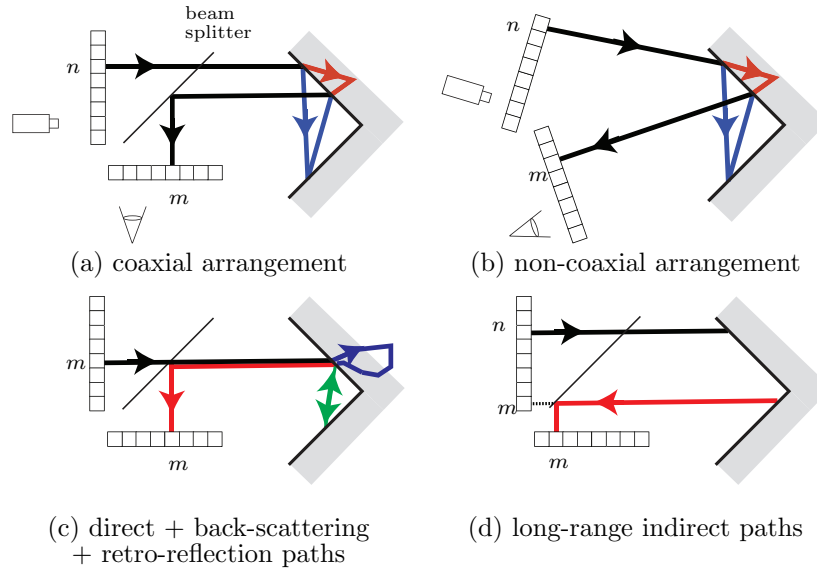


Figure 3.2: **Basic light paths.** (a)-(b) Element $\mathbf{T}[m, n]$ of the transport matrix takes into account all light paths that begin at projector pixel n and terminate at sensor pixel m . Two such paths are shown: inter-reflections (blue) and sub-surface scattering (red). (c) For coaxial arrangements, the direct transport path (red) always contributes to the diagonal element, $\mathbf{T}[m, m]$. This element may also include contributions from back-scattering (blue) and retro-reflection paths (green) which also begin and end at the same pixel. (d) In coaxial arrangements it is also possible to distinguish between long- and short-range transport paths by the distance $|m - n|$ between their endpoints.

Probing matrices for light transport calculations Observe that we can re-write the transport probing equation as

$$\mathbf{i} = \hat{\mathbf{T}} \mathbf{1} \quad (3.2)$$

by setting $\hat{\mathbf{T}} = \mathbf{\Pi} \odot \mathbf{T}$. This brings it into the familiar form of Eq. (1.1) and suggests a physical interpretation for the acquired photo: \mathbf{i} is the photo we would have captured by conventional photography if the scene had transport matrix $\mathbf{\Pi} \odot \mathbf{T}$ and was illuminated uniformly. Viewed this way, $\mathbf{\Pi}$ re-scales the elements of the transport matrix individually for the purpose of image acquisition.

Each element of the transport matrix describes the transport of radiant energy along a distinct collection of light paths from the illumination source to the sensor (Figures 3.2). By re-scaling these elements, the probing matrix gives the ability to directly acquire photos in which light transport along specific paths has been enhanced, attenuated, or blocked entirely. See Table 3.2 for several examples of imaging tasks, along with their associated probing matrices.

3.2 Photography by Optical Probing

Let us now look at how to capture photos governed by the transport probing equation. The key ingredient is an ability to simultaneously modulate light at its source (the projector) and its destination (the sensor). We first consider a naive algorithm that serves as an existence proof and illustrates the basic properties of the approach.

Acquisition Goal	Expression for Element $\mathbf{\Pi}[m, n]$
Capture direct plus retro-reflection and back-scattering (Figure 3.2(c))	$(m = n)$
Capture all other indirect (Figure 3.2(a))	$(m \neq n)$
Capture long-range indirect (Figure 3.2(d))	$(m - n > w)$
Capture short-range indirect (Figure 3.2(d))	$(m - n \leq w) \wedge (m \neq n)$
<i>Transport masking:</i> block all light paths from object \mathcal{B} to object \mathcal{A} but leave other paths unaffected	$(m \notin \mathcal{A}) \vee (n \notin \mathcal{B})$
<i>Local de-scattering:</i> capture direct plus high-frequency retro-reflection and back-scattering	$(m = n) - (m = n + 1)$
<i>Transport-domain filtering:</i> use a filter kernel \mathbf{f} to modulate transport paths according to their distance from the direct component	$(m = n + w)\mathbf{f}[w]$

Table 3.2: Probing matrices useful for light transport calculations in **coaxial arrangements**. We slightly abuse notation in the expressions for $\mathbf{\Pi}[m, n]$: logical operations (equality, inequality, set membership) are treated as functions returning 1 if true and 0 if false. For transport masking, we assume that the objects \mathcal{A} and \mathcal{B} are specified as sets of pixels on the image plane. The first five operations can be thought of as applying a binary mask to the transport matrix (*e.g.*, the matrix shown in Figure 3.3(b)). The last two operations involve cross-correlations along rows of the matrix. Note that unlike methods that rely on assumptions about the frequency content of light transport [90], none of the above operations do so, with the exception of local de-scattering. The de-scattering expression was proposed by Fuchs et al. [26] to remove back-scattering contributions that are constant in the neighborhood of the direct component. To our knowledge, transport masking and filtering have not been considered previously. For **non-coaxial arrangements**, the same relations apply with one modification: n must be equal to $\mathbf{a}[m]$, where vector \mathbf{a} encodes stereo correspondences. These relations are less useful when stereo correspondences are unknown.

3.2.1 Naive Approach: Path Isolation

Since we have complete control over which projector pixels emit light and which sensor pixels receive it, we can control in a very precise way the light transport paths that contribute to a photo. In particular, by turning on a single projector pixel and unmasking a single sensor pixel we guarantee that only the transport paths between those two pixels will contribute to the photo. The naive algorithm, shown in Algorithm 3, applies this basic idea in a time-multiplexed fashion: it allocates a time slice τ within the exposure period to every possible combination of projector pixel n and sensor pixel m , and uses $\mathbf{\Pi}[m, n]$ for the projector pixel’s intensity during that time slice. This ensures that transport paths that are “active” during a particular time slice will contribute radiant energy to the final photo precisely as described in Eq. (3.1).

Path isolation is an extremely inefficient approach for two reasons. Firstly, every element of the probing matrix gets equal time even though its value may be zero. Since many probing matrices of interest are sparse (*e.g.*, see Tables 3.1 and 3.2), the sensor would not be integrating any light at all for much of the exposure time. An obvious improvement is to allocate time proportionally to $\mathbf{\Pi}[m, n]$. This solution, for the special case where the $\mathbf{\Pi}$ is the identity matrix, describes the

basic confocal microscopy technique³ [15]. A second—much more serious—issue is that only one pixel integrates light at any given point in time. Given that a typical image may contain millions of pixels, the short time slices mean that light received at a pixel would fall well below the additive noise floor under typical photography conditions.

³The coaxial arrangements used in confocal imaging guarantee that the diagonal of the transport matrix contains all direct (*a.k.a.* in-focus) paths.

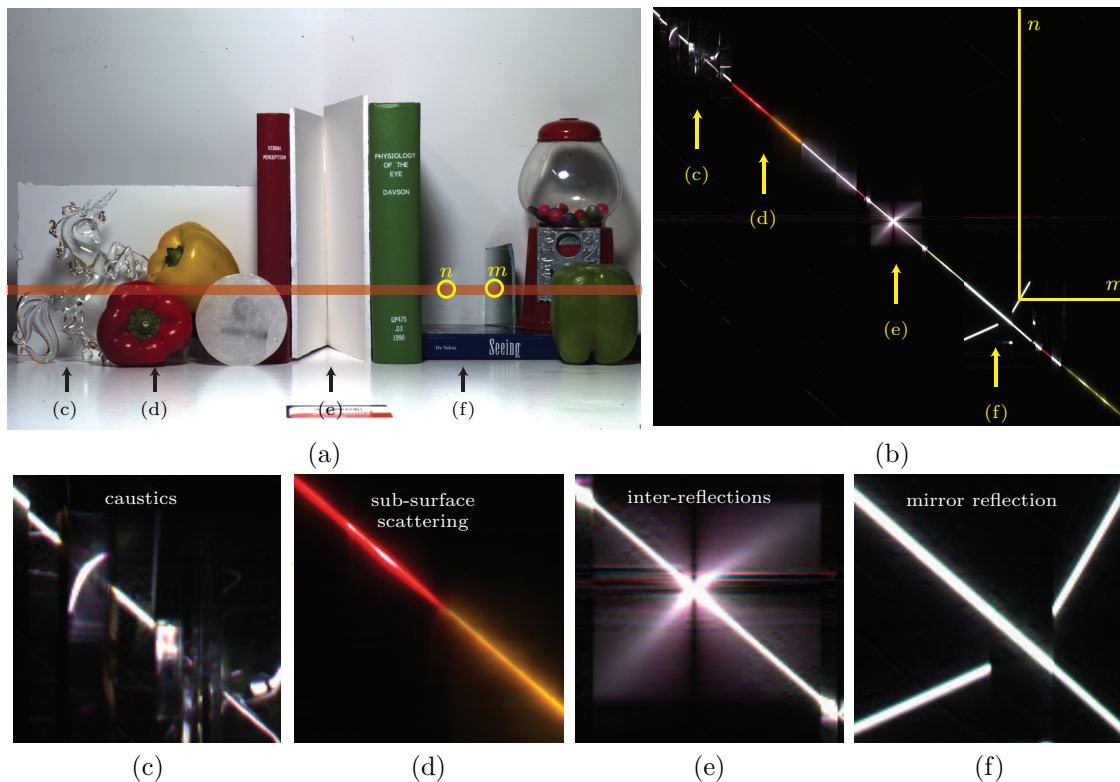


Figure 3.3: Transport matrix of a complex scene lit by a projector (coaxially-aligned with the camera). **(a)** A photo of the scene. From left-to-right, the scene contains a glass unicorn, a pair of bell peppers, a translucent wax disk, a V-wedge between a pair of books, a mirror reflecting light onto the wall, a gumball dispenser and a green bell pepper. **(b)** Since the complete transport matrix is too large to capture and visualize, we show a 2D slice of the matrix for the row highlighted in (a). The slice was acquired with a coaxial arrangement using the method of Schechner et al. [111]. Element $\mathbf{T}[m, n]$ in this slice represents light paths that originate from projector pixel n and reach camera pixel m in the highlighted row, shown as yellow circles in (a). Note that non-zero elements in this slice mostly concentrate around the diagonal. This indicates that most light is transported between nearby camera/projector pixels. Nevertheless, the matrix does include several interesting structures, magnified in (c)-(f). **(c)** Caustics from the glass unicorn appear as “streaks” in the transport matrix. **(d)** Sub-surface scattering from the yellow and red bell peppers produces a thick band along the diagonal, with isolated white specularities on the diagonal itself. **(e)** Inter-reflections from the V-wedge appear as fans of light. Dark regions in these fans are due to occlusion, which blocks inter-reflection paths in certain directions. **(f)** Mirror reflections produce high-frequency, off-diagonal structures.

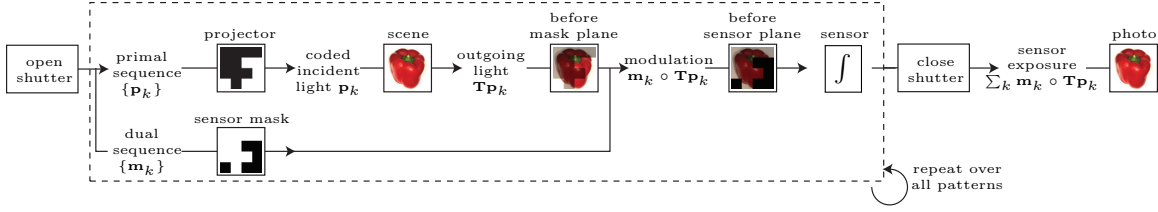


Figure 3.4: Image formation pipeline for a single photo captured by optical probing. The sequence of mask and projector patterns, $\{\mathbf{p}_k\}$ and $\{\mathbf{m}_k\}$, are computed in a preprocessing step. Patterns in the primal and dual domains are changed synchronously.

3.2.2 Optical Matrix Probing

Acquisition schemes that make efficient use of the sensor are highly desirable. For this, we turn to work in numerical linear algebra [8]. Let $\{\mathbf{p}_k\}$ and $\{\mathbf{m}_k\}$ be two sequences of column vectors corresponding to a decomposition of the probing matrix into a sum of rank-1 matrices:

$$\mathbf{\Pi} = \sum_{k=1}^K \mathbf{m}_k (\mathbf{p}_k)^T \quad (3.3)$$

This allows us to re-write the probing equation as a sum identical to the pipeline in Figure 3.4:

$$(\mathbf{\Pi} \odot \mathbf{T}) \mathbf{1} = \sum_{n=1}^P \mathbf{\Pi}[n] \odot \mathbf{T}[n] \quad (3.4)$$

$$= \sum_{n=1}^P \left(\sum_{k=1}^K \mathbf{m}_k \mathbf{p}_k[n] \right) \odot \mathbf{T}[n] \quad (3.5)$$

$$= \sum_{k=1}^K \left(\mathbf{m}_k \odot \sum_{n=1}^P \mathbf{T}[n] \mathbf{p}_k[n] \right) \quad (3.6)$$

$$= \sum_{k=1}^K \mathbf{m}_k \odot \mathbf{T} \mathbf{p}_k \quad (3.7)$$

Here we used $\mathbf{T}[n]$ to denote the n^{th} column of \mathbf{T} , and $\mathbf{p}_k[n]$ to denote the n^{th} element of \mathbf{p}_k . It follows that every decomposition of the probing matrix into a sum of rank-1 matrices provides a candidate sequence of illumination patterns $\{\mathbf{p}_k\}$ and masks $\{\mathbf{m}_k\}$ for optical probing.

Stochastic diagonal estimators Not all rank-1 decompositions are equally efficient. Bekas et al. [8] study this problem for the case where the probing matrix is the identity matrix \mathbf{I} . This matrix allows us to acquire the diagonal of \mathbf{T} (third row of Table 3.1 and first row of Table 3.2). The main idea is to consider rank-1 decompositions of the identity matrix created by an infinite sequence of random vectors $\{\mathbf{p}_k\}$ drawn from a distribution with mean zero and unit variance:

$$(\mathbf{I} \odot \mathbf{T}) \mathbf{1} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \mathbf{p}_k \odot \mathbf{T} \mathbf{p}_k \quad (3.8)$$

They show that the provably-optimal convergence rate is achieved when these randomized vectors are drawn from the Rademacher distribution, whose elements have a 50% chance of being either

+1 or -1 .⁴ Moreover, as shown in Appendix B.1.2, this distribution minimizes the variance of the K -term estimate of diagonal element $\mathbf{T}[m, m]$:

$$\frac{1}{K} \sum_{n=1, n \neq m}^P \mathbf{T}[m, n]^2 \quad (3.9)$$

Therefore, to double the accuracy of our estimate of $\mathbf{T}[m, m]$ we must use twice as many vectors. Importantly, Eq. (3.9) tells us that this accuracy depends on the transport matrix itself and thus on the type of indirect transport within a scene: it is low when the total intensity of non-zero elements along a row of \mathbf{T} is distributed over many elements (*e.g.*, diffuse inter-reflections) and high when it is concentrated at a few isolated ones (*e.g.*, caustics).

Stochastic estimators for general probing Generalizing the above estimator to the case of a general probing matrix $\mathbf{\Pi}$ is trivial in some respects and non-trivial in others. On the trivial side, extracting the w^{th} off-diagonal of \mathbf{T} , as shown in the fourth row of Table 3.1, simply requires a mask that is a w -shifted version of the illumination pattern:

$$(\mathbf{\Pi} \odot \mathbf{T}) \mathbf{1} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \text{shift}(\mathbf{p}_k, w) \circ \mathbf{T} \mathbf{p}_k \quad (3.10)$$

More generally, left-multiplying both sides of Eq. (3.8) by $\mathbf{\Pi}$ gives the following:

$$(\mathbf{\Pi} \odot \mathbf{T}) \mathbf{1} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \mathbf{\Pi} \mathbf{p}_k \circ \mathbf{T} \mathbf{p}_k \quad (3.11)$$

This expression is straightforward to implement optically (see Algorithm 3). Unfortunately, the optimality guarantees of the Rademacher distribution do not extend to general probing matrices.

⁴Although not described in these terms, aperture correlation [134] uses this sequence for extracting the diagonal of \mathbf{T} . This was proposed well before its optimality was established.

Algorithm 3 Basic algorithms for optical probing photography.

Path isolation:

In: exposure time T , probing matrix $\mathbf{\Pi}$

Out: photo equal to $(\mathbf{\Pi} \odot \mathbf{T}) \mathbf{1}$

- 1: $\tau = T/IP$
 - 2: open camera shutter
 - 3: **for** $m = 1$ to I
 - 4: unmask pixel m
 - 5: **for** $n = 1$ to P
 - 6: turn on projector pixel n for
 - 7: time τ with intensity $\mathbf{\Pi}[m, n]$
 - 7: mask all pixels
 - 8: close shutter
 - 9: **return** captured photo
-

Optical matrix probing:

In: exposure time T , probing matrix $\mathbf{\Pi}$,
 K illumination vectors $\{\mathbf{p}_k\}$

Out: photo equal to $(\mathbf{\Pi} \odot \mathbf{T}) \mathbf{1}$

- 1: $\tau = T/K$
 - 2: open camera shutter
 - 3: **for** $k = 1$ to K
 - 4: apply mask $\mathbf{m}_k = \mathbf{\Pi} \mathbf{p}_k$
 - 5: project vector \mathbf{p}_k for time τ
 - 6: close shutter
 - 7: **return** captured photo
-

Negative values Up to now we have ignored the fact that the illuminations and masks may contain negative values which are not physically realizable. Negative values may occur because of the random vector sequence being used or because $\mathbf{\Pi}$ itself contains negative values. Correct treatment of such cases requires capturing at most two photos followed by a (computational) subtraction. In particular, by expressing \mathbf{p}_k and \mathbf{m}_k as a difference of two non-negative vectors, it is possible to break each term in Eq. (3.3) into four terms, two of which are positive and two of which are negative. The first two terms are used during the exposure period of the positive photo and the last two of the negative one:

$$\underbrace{\left[\underbrace{(\mathbf{m}_k)_+(\mathbf{p}_k)_+^\top}_{\#1} + \underbrace{(\mathbf{m}_k)_-(\mathbf{p}_k)_-^\top}_{\#2} \right]}_{\text{positive terms}} - \underbrace{\left[\underbrace{(\mathbf{m}_k)_-(\mathbf{p}_k)_+^\top}_{\#3} + \underbrace{(\mathbf{m}_k)_+(\mathbf{p}_k)_-^\top}_{\#4} \right]}_{\text{negative terms}} \quad (3.12)$$

where $\mathbf{p}_k = (\mathbf{p}_k)_+ - (\mathbf{p}_k)_-$ and $\mathbf{m}_k = (\mathbf{m}_k)_+ - (\mathbf{m}_k)_-$. It follows that implementing a K -vector optical probing sequence with both positive and negative values requires capturing two photos using a $2K$ -long sequence during each photo's exposure.

Fortunately, useful probing matrices exist that produce strictly non-negative illuminations and masks. This makes it possible to probe efficiently with just one photo. Two important examples are capturing an indirect-only photo of the scene (second row in Table 3.2) and capturing the direct component plus one-half the indirect (*i.e.*, contrast-enhancing the direct component). See Appendix B.1.1 for a brief derivation of how Rademacher sequences can be replaced by non-negative Bernoulli sequences in such cases.⁵

3.3 Implementation

Hardware Figure 3.5 illustrates our hardware setup. We use an Epson PowerLite G5000 3-LCD projector, a disassembled Barco E-2320 PA monochrome LCD panel, a 50R/50T Edmund Optics beam splitter, two spherical plano-convex relay lenses, a Canon EOS Rebel XSi camera, a 2.67 GHz Intel Core i7 workstation with 3 GB of RAM and two NVIDIA GeForce 9500GT graphics cards. The polarizers on each side of the Barco LCD display had been removed for another project; we reassemble the LCD display by positioning it between two linear polarizing sheets. We also place a linear polarizer in front of the 3-LCD projector so that the projector illumination has a consistent polarization state. We operate the projector at resolution 1024×768 and use two VSYNC-enabled borderless OpenGL windows to control the projector and LCD at a 60 Hz refresh rate. Pattern generation is done through MATLAB.

The rightmost relay lens in Figure 3.5 focuses light from the scene onto the LCD display. The display attenuates this incident light using a 1600×1200 resolution pattern. The leftmost relay lens, along with the camera's own 50 mm focal length lens, is then used to capture a focused image of the LCD display. The focal length of the leftmost relay lens was 330 mm in all experiments; the other relay lens was 330 mm in all experiments except one (Figure 3.8), where a 220 mm focal length lens

⁵According to Eq. (3.3), if $\mathbf{\Pi}$'s rank is low applying non-negative factorization also yields efficient non-negative illuminations and masks.

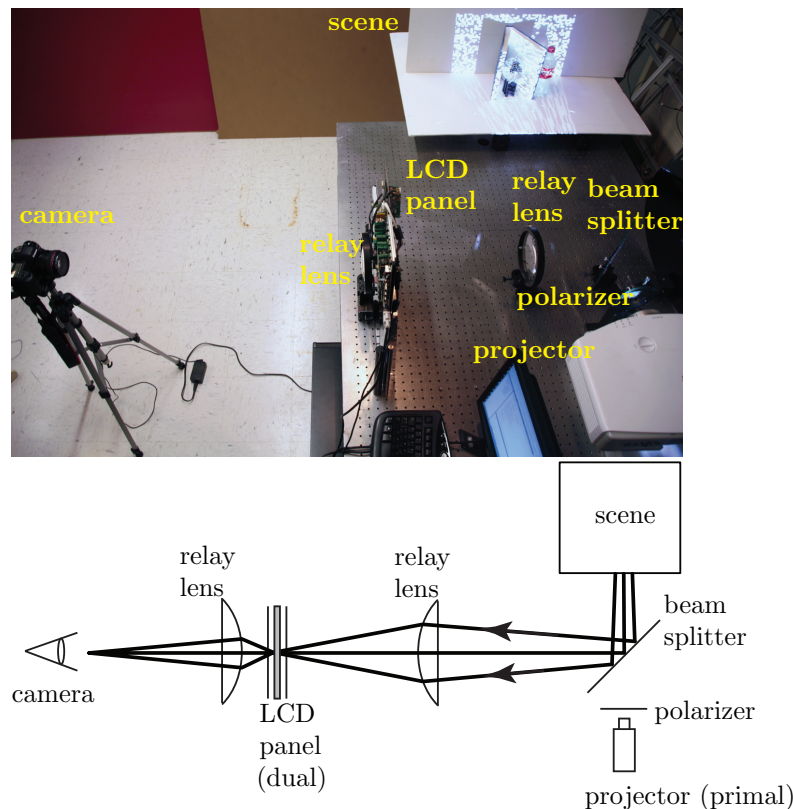


Figure 3.5: Photo and diagram of our optical probing prototype. We use a pair of relay lenses and an LCD panel with its backlight removed to modulate the light reaching the camera’s sensor. To acquire an optically probed photo, we project a sequence of illumination patterns onto the scene and simultaneously display a sequence of modulation patterns on the LCD panel. The camera’s shutter remains open throughout this process.

was used. These relay lenses were slightly larger than the active area of the LCD display (15 cm versus 10 cm) and positioned asymmetrically, with the leftmost one much closer to the display. We align the projector and camera using a beamsplitter to ensure a coaxial arrangement.

Calibration All experiments require a coaxial camera and projector and aligned projector and LCD patterns. We do this as follows. We temporarily place a diffuser on the LCD panel; focus the image of the scene onto the diffuser by adjusting the position of the rightmost relay lens; align the projector to produce a coaxial arrangement; and move and resize the OpenGL windows so that the pixels of the projector and LCD panel overlap from the camera’s perspective. Since we relied on binary patterns for our experiments, no radiometric calibration was required. For more general probing experiments, where illuminations and/or masks are not binary, radiometric calibration of both the LCD panel and projector becomes necessary.

Codes robust to misalignment Even after calibration, the illumination and mask patterns in our prototype are not aligned with pixel accuracy. Moreover, despite driving the projector and the

LCD panel with the same vertical synchronization signal, the two devices do not refresh in perfect synchrony. These misalignments forced us to operate at a reduced spatio-temporal resolution. To account for temporal misalignment, we interleave black images on the LCD panel before and after displaying each mask pattern, reducing the effective optical probing display rate to 20 Hz. To deal with pixel misalignment, we rely on reduced-resolution, 64×48 optical probing sequences and use four times as many codes (included in our code counts for experiments). In particular, we replace every pair $\mathbf{p}_k, \mathbf{m}_k$ in the optical probing sequence with a four-code sequence $\mathbf{p}_{kl}, \mathbf{m}_{kl}$, that eliminates artifacts due to misaligned pixel boundaries (Figure 3.6). This sequence exposes the center $\frac{1}{4}$ th of reduced-resolution pixels and then shifts both the illumination pattern and the mask three times to expose the complete coarse-resolution pixels. The sum of the four mask patterns, \mathbf{m}_{kl} , produces the original dual code \mathbf{m}_k . Together with the black masks, we get the following twelve-code sequence:

Code #:	1	2	3	4	5	6	7	8	9	10	11	12
Primal code:	\mathbf{p}_{k_1}	\mathbf{p}_{k_1}	\mathbf{p}_{k_1}	\mathbf{p}_{k_2}	\mathbf{p}_{k_2}	\mathbf{p}_{k_2}	\mathbf{p}_{k_3}	\mathbf{p}_{k_3}	\mathbf{p}_{k_3}	\mathbf{p}_{k_4}	\mathbf{p}_{k_4}	\mathbf{p}_{k_4}
Dual code:	$\mathbf{0}$	\mathbf{m}_{k_1}	$\mathbf{0}$	$\mathbf{0}$	\mathbf{m}_{k_2}	$\mathbf{0}$	$\mathbf{0}$	\mathbf{m}_{k_3}	$\mathbf{0}$	$\mathbf{0}$	\mathbf{m}_{k_4}	$\mathbf{0}$

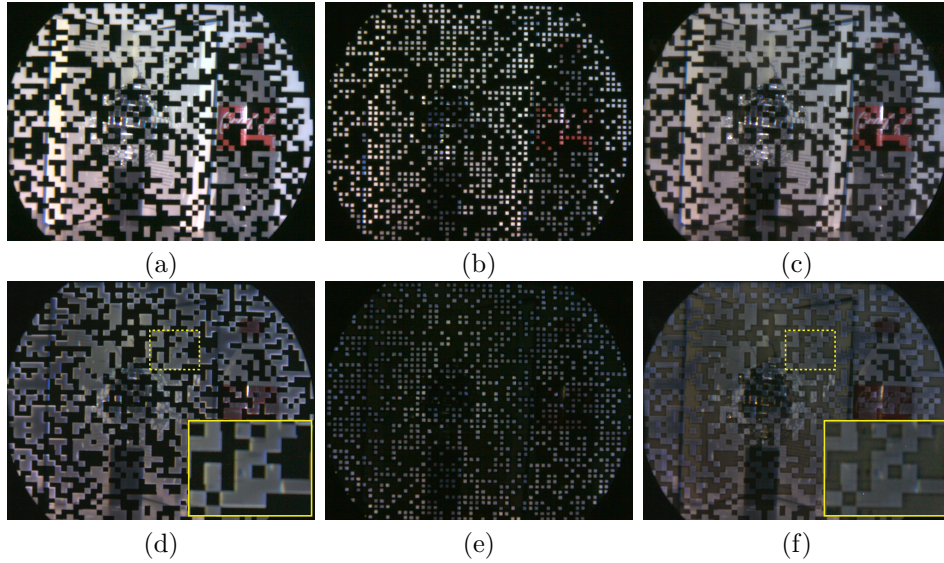


Figure 3.6: Codes robust to pixel misalignment. **Column 1:** Images of a scene taken with a mask/projection pattern pair defined on a coarse pixel grid. The projector and mask patterns are identical in (a), *i.e.*, $\mathbf{p}_k = \mathbf{m}_k$, whereas in (d) they are complements of each other. Misalignment artifacts appear as dark or bright pixels near coarse pixel boundaries. For example, the inset in (d) contains bright ghosting artifacts located near these boundaries, a result of the mask not blocking all direct light from the scene. **Column 2:** Photo of the scene corresponding to one member, $(\mathbf{p}_{kl}, \mathbf{m}_{kl})$, of the 12-code sequence we use to implement the codes in Column 1. Each code in the sequence samples the center $\frac{1}{4}$ th of each coarse pixel to avoid exposing the region near its boundary. The mask/projection pattern pair shown in Column 2 is shifted four times during the sequence. These shifts expose the area blocked by the previous codes in the sequence. The other eight codes in the sequence are blank frames captured with a mask consisting of all zeros. **Column 3:** Summing the photos corresponding to the 12-code sequence results in the photos shown in (c),(f). Note the lack of bright ghosting artifacts near the boundary of coarse pixels.

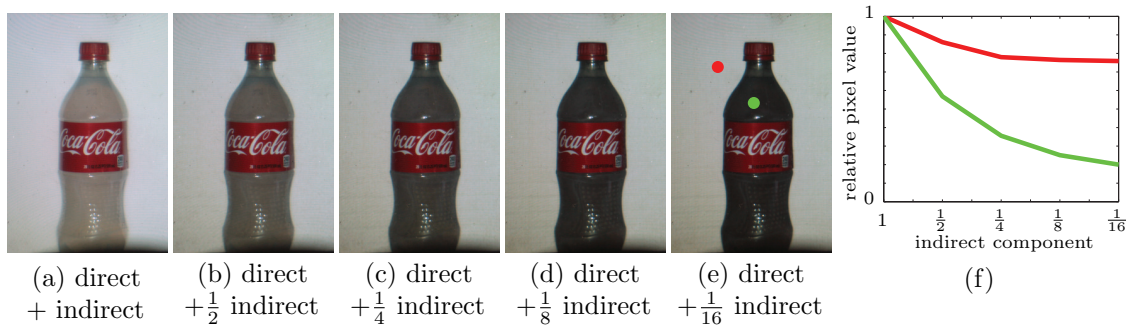


Figure 3.7: Contrast-enhancing the direct component of a coke bottle filled with milky water, which strongly scatters light. **Left to right:** A photo of the scene lit by white illumination contains full direct and indirect; the scene with 50% indirect illumination uses 64 mask and illumination codes; 25% indirect illumination uses 128 codes, 12.5% indirect uses 256 codes, and 6.25% indirect uses 512 codes. The plot on the right examines the intensity of two points in the scene, a point lit mostly by indirect (green), and a point lit mostly by direct (red). As we reduce the percentage of indirect light, the intensity of the point on the background decreases slightly, indicating a small indirect component. In contrast, the intensity of the point on the bottle diminishes, an indication of the strong indirect transport effects within the bottle.

Figure 3.6 demonstrates the effect of using this 12-code sequence for indirect-only imaging. For high code resolutions, the center $\frac{1}{4}^{\text{th}}$ of each pixel may no longer mask the misaligned pixels. This imposes a maximum working resolution to any optical probing setup (ours is 64×48). This means that all light paths within a coarse pixel are clumped together in the photos we capture.

Image capture and adjustment We capture 4272×2848 resolution RAW photos and decode them using DCRAW. We then rotate, flip, tone map and white balance the photos. We use gamma correction with $\gamma = 2.2$ for tone mapping and manually choose a white object in each photo to set the white point through MATLAB.

3.4 Results

Single-photo enhancement of direct component We first consider the problem of capturing single photos in which the direct component has been contrast-enhanced. Figure 3.7 shows the captured RAW images for a scene containing a coke bottle filled with milky water. The milky medium scatters light entering within the coke bottle, producing a strong indirect lighting effect.

To do this, we generate a sequence of projector and mask patterns using the Bernoulli distribution. These codes satisfy $\mathbf{p}_k = \mathbf{m}_k$, have elements that are 1 with probability p and 0 with probability $1-p$, and converge to a probing matrix with value p along the diagonal and p^2 everywhere else. Since this matrix attenuates the direct component by a factor of p , the mean brightness of direct-enhanced photos changes with p too. To keep it approximately constant in Figure 3.7, we compensate for reductions in p by increasing exposure time, *i.e.*, using more codes.

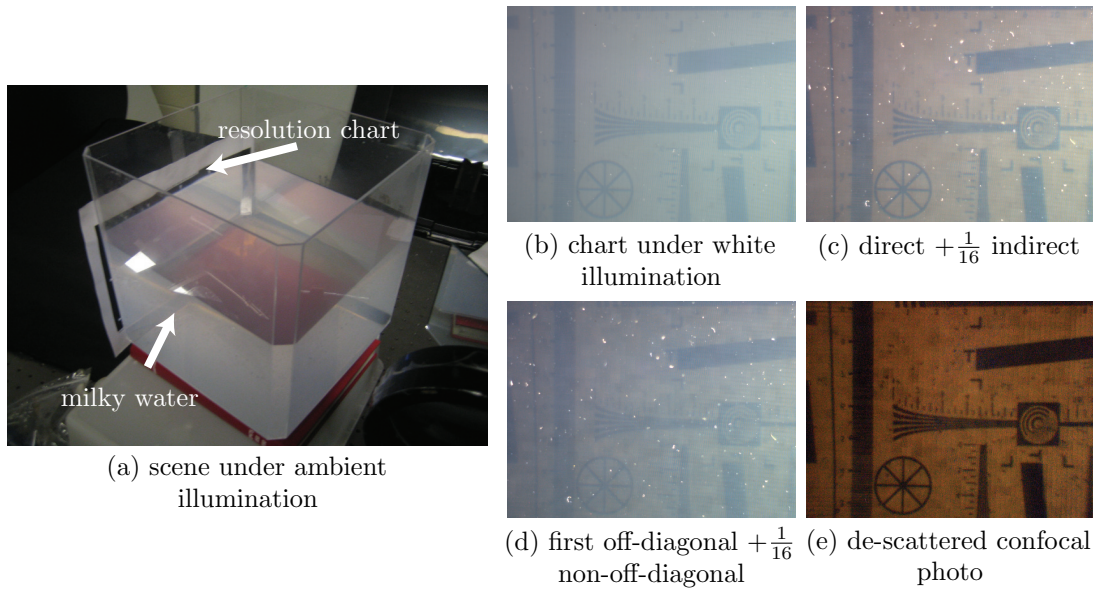


Figure 3.8: Imaging a resolution chart through a tank of milky water with optical probing. **(a)** A tank filled with water and a small amount of milk scatters light. The objective is to image a resolution chart taped to the back side of the tank using no more than two photos. **(b)** Illuminating the scene with a white illumination pattern produces an image with poor contrast because of back-scattering. **(c)** A contrast-enhanced direct photo, obtained using the same procedure as in Figure 3.7, uses 1024 mask and illumination patterns. This produces a photo where the indirect component is reduced. **(d)** Enhancing the first off-diagonal of the transport matrix involves using a shifted version of the codes used in (c), according to Eq. (3.10). **(e)** By subtracting the off-diagonal-enhanced photo in (d) from the direct-enhanced photo in (c), we obtain a de-scattered confocal photo. This significantly enhances contrast relative to the original floodlit image in (b).

De-scattering from two shots Fuchs et al. [26] describe an approach for imaging through scattering media that combines confocal imaging and de-scattering. It requires scanning a line across the scene and capturing, storing and analyzing many photos. We do this in just two shots by implementing their procedure optically.

For this experiment, we replace the rightmost 330 mm focal length lens with a 220 mm focal length lens and focus the camera and projector onto a resolution chart through a tank of milky water (Figure 3.8). We then take two photos in order to implement the probing matrix for local de-scattering in Table 3.2. The first photo captures the direct-enhanced component, corresponding to the direct plus $\frac{1}{16}$ th indirect, as in Figure 3.7. The second captures the first off-diagonal of the transport matrix plus $\frac{1}{16}$ th of contributions not on that off-diagonal (Eq. (3.10)). The difference of the two photos corresponds to a matrix with 1 on the diagonal and -1 on the first off-diagonal. Note that this difference greatly enhances the contrast relative to the floodlit image.

Just like in our direct enhancement experiment, we control the exposure level of our photos by the number of mask/projection pattern pairs used to capture them. Since we expect the diagonal entries of the transport matrix to be smaller than those of Figure 3.7 because of scattering, we double the total exposure time using 1024 codes.



Figure 3.9: **Left to right:** Photo of scene under white light, indirect component of scene captured by a single RAW photo, direct component computed by subtracting the indirect image from the scene under white light. No processing was done on the photos, other than a resize, flip, white balance, gamma correction and crop. All scenes use 512 mask and illumination codes, with the exception of Row 4 which uses 1024 codes. **Row 1:** A disco ball produces specular indirect transport across the scene, with diffuse inter-reflections occurring near the edge of the floor. Optical probing cleanly separates the photo into its direct component and its indirect component, which contains inter-reflections and specularities. **Row 2:** A scene composed of two styrofoam balls inside a glass container generates specular reflections, sub-surface scattering and caustics. We used small aperture to photograph this scene in order to avoid pixel saturation due to caustics. The photo is slightly noisier as a result. **Row 3:** A coke bottle filled with milky water strongly scatters light that enters the bottle. **Row 4:** Observe that the direct component contains none of the light that is reflected by the mirror (left) onto the vase (right) and the wall. However, the mirror does not appear black in the direct photo even though it has little to no direct component. This is because of back-scattering and retro-reflection (Figure 3.2(c)): light leaves a projector pixel, hits the mirror at a point, diffusely reflects off the wall, then hits the mirror at the *same* point, and returns to the same pixel. This light path will always contribute to the diagonal of the transport matrix and cannot be distinguished from direct illumination using methods that rely purely on intensity, and without regard to arrival time. Note that according to Eq. (3.9), the variance of stochastic diagonal estimation may be higher for indirect specular transport than for diffuse transport. To compensate for this, we used 1024 codes for this scene.

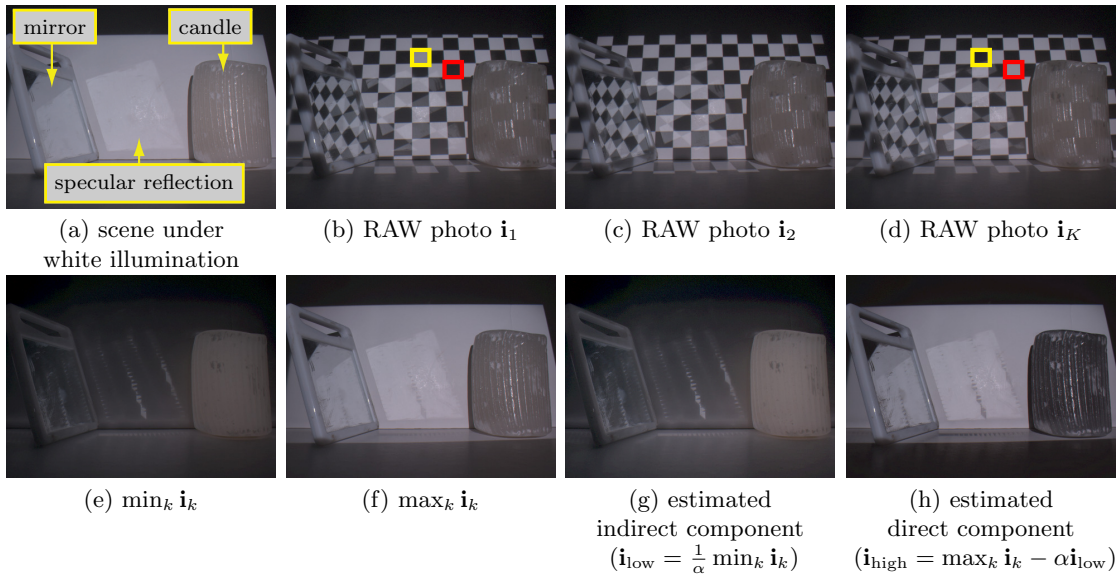


Figure 3.10: Demonstration of Nayar et al.’s direct and indirect separation technique [90]. (a) The scene contains a mirror (left), a diffuse backdrop (middle), and a translucent candle (right). (b) The first step of the procedure involves capturing a photo \mathbf{i}_1 of the scene lit by a high-frequency binary projection pattern (e.g. a checkerboard pattern). The separation procedure assumes every pixel $\mathbf{i}_1[m]$ either receives all of the direct light (e.g. region highlighted in yellow) or none of the direct light (e.g. region highlighted in red). At the same time, the procedure also assumes that all pixels $\mathbf{i}_1[m]$ always receive a constant fraction α of the indirect light; for example, since only $\alpha = \frac{1}{2}$ of the projector pixels are “on” in a checkerboard pattern, all pixels are assumed to receive exactly half of the indirect light present in (a). (c) The procedure captures another $K - 1$ photos of the scene under shifted versions of the same projection pattern. (d) For comparison purposes, this RAW photo is lit by the complement version of the illumination pattern used in (b). Note that the region highlighted in yellow now receives no direct light, and the region highlighted in red receives all of the direct light. (e)-(f) The next step is to compute the minimum and maximum brightness observed at each camera pixel across all K photos. The minimum pixel value is assumed to be half of the indirect component, and the maximum pixel value is assumed to be the direct component plus half of the indirect component. (g) Amplifying the values in (e) by a factor of $\frac{1}{\alpha}$ produces the indirect component. However, the separation method fails for regions containing specular indirect transport (e.g. the pixels containing specular reflections); these regions violate the assumption that pixels always receive exactly α of the indirect light. (h) Computing the difference between (f) and (e) produces an estimate of the direct component.

Separating indirect and direct components in the presence of specular indirect transport

Figure 3.9 shows several examples of capturing photos that contain only indirect or direct light. Each scene contains a variety of effects, including sub-surface scattering, caustics, specular reflections, and diffuse inter-reflections. Direct images are computed by subtracting the indirect image from an image of the scene under white illumination.

The mask and projector patterns are similar to those used for direct enhancement: \mathbf{p}_k is drawn from the Bernoulli distribution and $\mathbf{m}_k = \mathbf{1} - \mathbf{p}_k$, resulting in a probing matrix that has 0 along the diagonal and $p(p - 1)$ everywhere else. We use $p = 0.5$ in our experiments because it maximizes the magnitude of the probing matrix.

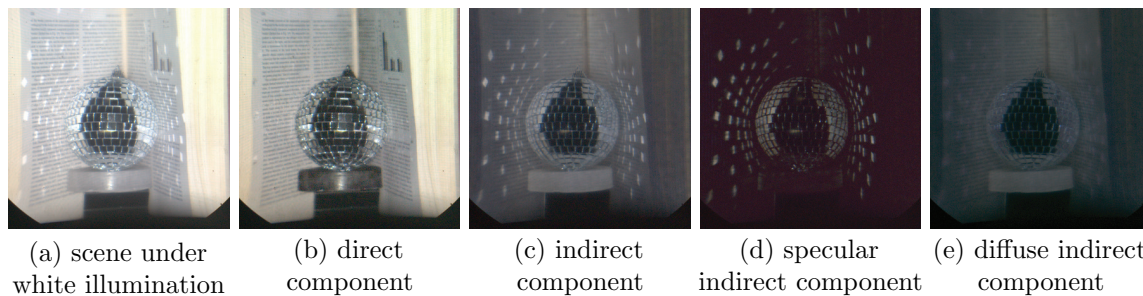


Figure 3.11: **(a)** A disco ball ornament placed on top of a wax disk and in front of an open book. The scene is lit by projecting a white pattern. Observe that the open book causes diffuse inter-reflections, the wax disk exhibits strong sub-surface scattering, whereas the disco ball produces specular reflections across the scene. **(b)** Optical probing produces a direct component from two photos. This removes light due to inter-reflections, sub-surface scattering and specular inter-reflections even though significant specular indirect transport effects are present. **(c)** We also obtain the indirect component in a single RAW photo. The LCD panel always transmits a small amount of light, even when all pixels are set to 0; to correct for this, we compute the difference between the indirect photo and a second photo where all the dual codes are set to 0 (not shown here). By capturing the indirect illumination of the scene under shifted high-frequency patterns, it is possible to separate **(d)** specular indirect transport (due to mirror reflections) from **(e)** diffuse indirect transport (due to inter-reflections and sub-surface scattering).

A similar, multi-photo approach to computing the indirect and direct components of a scene was first proposed by Nayar et al. [90]. They show that the direct and indirect components can be efficiently measured using high-frequency illumination patterns (Figure 3.10). This, however, assumes the presence of low-frequency indirect transport (*i.e.* the indirect component of every row of the transport matrix \mathbf{T} is a low-frequency signal), meaning that shifting a high-frequency illumination pattern across a scene produces no change in the indirect component. Nayar et al.’s method works well for diffuse indirect transport effects such as diffuse inter-reflections and sub-surface scattering, but fails in the presence of specular indirect transport effects like caustics and specular reflections.

Separating diffuse and specular indirect transport The indirect-only probing matrix (Table 3.2) removes any influence of the diagonal on images from the scene. To further analyze the indirect component, we apply the approach of Nayar et al. [90] *only to the indirect component*. To do this, we use probing to optically “simulate” conventional photography on the indirect-only transport matrix, as described in the last row of Table 3.1; here, we set the matrix \mathbf{A} to be the indirect-only probing matrix, and the pattern \mathbf{p} to be a high-frequency illumination pattern (*e.g.* a checkerboard). The mask and illumination codes of this probing matrix are a slightly modified version of the indirect-only probing matrix, where we multiply pattern \mathbf{p} with each illumination code.

Following the approach illustrated in Figure 3.10, we capture four indirect-only photos illuminated by a shifting high-frequency pattern \mathbf{p} to separate specular indirect light (specularities and caustics) from diffuse indirect light (inter-reflections and sub-surface scattering). Four “black” photos were used to counteract the transmissivity of the LCD panel.

Figure 3.11 shows that our algorithm successfully separates the specular indirect reflections of the disco ball from the inter-reflections and sub-surface scattering of the book and the scattering wax disk, respectively. The number of the codes for this experiment was 128 per photo.

Separating diffuse and specular direct transport The polarization state of light provides a visual cue that enables the separation of diffuse and specular reflections [135]. Specular reflections preserve the polarization state of light, whereas diffuse reflections do not. As a result, two crossed linear polarizers can be used to block all specular direct light from contributing to a photo: one to polarize the illumination at its source, and the other to filter out any linearly-polarized light emanating from the scene. Computing the specular direct component requires subtracting this photo from a regular photo of the scene. We refer to Wolff and Boult [135] for a comprehensive discussion on polarization and reflectance.

This same approach can be used to also recover the diffuse direct component, as demonstrated in Figure 1.3. The approach is simple: probe for direct light paths while using polarizers to block all specular direct paths. (Note that the separation results achieved in Figure 1.3 uses the optical probing prototype proposed in Chapter 4.)

3.5 Discussion

Limitations of our prototype Our prototype system is fixed to an optical bench, uses a restrictive coaxial camera-projector arrangement, is limited to displaying mask and illumination patterns at 20 Hz, has spherical aberrations, and uses a relatively low-contrast projector and LCD. There is also a small amount of light that transmits through the LCD panel which becomes visible over long exposures; compensating for this light requires a second photo with the mask set to black. Our single-photo experiments assume no ambient light, though ambient light can be eliminated using two photos.

Comparison to mask-less multi-image acquisition An alternative way of implementing optical probing is to capture a separate photo for each illumination pattern without using any mask, and then perform masking and integration computationally rather than optically. Such an approach might seem advantageous because it does not require hardware for optical masking and because individually-captured photos might be useful for other purposes. In Appendix A.1, we use an example to show that the utility of such photos is far less clear when acquisition time is taken into account. This is because high-speed multi-image acquisition is limited by *additive sensor noise* whereas optical probing is photon limited, and mainly constrained by the *display rate* of projectors and masks. Looking forward, we expect display technologies to improve much faster than additive sensor noise, conferring further advantage to our all-optical approach. Indeed, using a DMD for masking enables far higher display rates without moving outside our photon-limited regime.

3.6 Summary and Contributions

We introduced optical probing as a general imaging technique that offers fine-grain control over how light paths contribute to a photo. Probing is another example of an optical algorithm, requiring little or no post-processing after acquisition. We showed that optical probing

- greatly increases the available degrees of freedom in photography and allows direct acquisition of photos that cannot be captured in one shot by conventional methods;
- provides the ability to attenuate or enhance different light transport components (*e.g.* Figures 3.7, 3.8, and 3.9); and
- and can be combined with other computational methods to analyze light transport (*e.g.* Figure 3.11).

Although these probing results used a coaxially-aligned projector-camera system, the probing procedure itself is not restricted to any configuration of lights and cameras. In particular, the next chapter considers non-coaxial systems by bringing epipolar constraints into the formulation.

Chapter 4

Structured Light Transport

The assumption that light travels along direct paths is actually quite common in computer vision. While this assumption works well in many cases, we’ve seen that light propagation through natural scenes is actually a much more complex phenomenon: light reflects and refracts, it undergoes specular and diffuse inter-reflections, it scatters volumetrically and creates caustics, and may do all of the above in the same scene. Analyzing all these phenomena with lights and cameras *in general positions* (e.g. non-coaxial arrangements) is a hard, open problem—and is even harder when the scene is dynamic and light transport changes unpredictably. It is also a major factor preventing broader use of active illumination, including *structured-light techniques* used for triangulation-based 3D scanning (Section 1.4.2) which largely assume direct or diffuse indirect light transport [18, 34, 42, 110].

Working from first principles, we show that two families of light paths dominate image formation in a non-coaxial projector-camera system: *epipolar paths*, which satisfy the familiar epipolar constraint and contribute to a scene’s direct image, and *non-epipolar paths* which contribute to its indirect. Crucially, while the contributions of these paths are hard to separate computationally once an image has been captured, the paths themselves can—once again—be blocked optically *before* acquisition takes place.

Using this idea as a starting point, we show how to realize this type of control over light transport with the help of optical probing. Here we apply a generalized technique, which we call *Structured Light Transport (SLT)*, to four imaging problems and derive their associated mask and projection patterns:

- **non-epipolar (indirect-only) imaging:** capture an image that records only indirect light;
- **two-shot epipolar (direct-only) imaging:** capture two images whose difference contains only the direct light;
- **indirect-invariant structured light:** given any conventional structured-light pattern used for 3D shape acquisition, capture a view of the scene under that pattern that is guaranteed to be invariant to indirect light; and
- **one-shot multi-pattern imaging:** given any $S \geq 2$ conventional structured-light patterns

used for 3D shape acquisition, capture one image that contains S separate views of the scene “packed” into it, each corresponding to a different structured-light pattern.

Little is currently known about how to solve these problems in the general setting we consider. Our solutions, while firmly rooted in computer vision, operate exclusively in the optical domain and require no computational post-processing: our implementation is a physical device that just outputs live video. This video is optionally processed after acquisition by standard 3D reconstruction algorithms [110] which can be oblivious to the complexity of light transport occurring in a scene. The device itself is a novel combination of existing off-the-shelf components—a conventional video camera operating at 28 Hz, a pair of synchronized DMDs operating at 2.7 kHz to 24 kHz for sensor masking and pattern projection, and optics for coupling them.

From a practical point of view, this work offers four main contributions over the state of the art. First, it is the first demonstration of an “indirect-only video camera,” *i.e.*, a camera that outputs a live stream of indirect-only video for general scenes—exhibiting arbitrary motion, caustics, inter-reflections and numerous other transport effects. Prior work on indirect imaging was either constrained to static scenes [129], or assumed diffuse indirect transport [2, 90] and accurate 2D motion estimation [2]. Second, we show how to capture—with just one SLT shot—views of a scene that are invariant to indirect light. This is particularly useful for imaging dynamic scenes and represents an advance over direct-only imaging [90], which requires at least two images. Third, we show that *any* ensemble of structured-light patterns can be made robust to indirect light, regardless of the patterns’ frequency content. This involves simply switching from conventional to SLT imaging—without changing the patterns or the algorithm that processes them. As such, our work stands in contrast to prior work on transport-robust structured light, which places the onus on the design of the patterns themselves [11, 17, 41, 42]. Fourth, we show that SLT imaging can turn any multi-pattern 3D structured-light method into a one-shot technique for dynamic shape capture. Thus an entire family of previously-inapplicable techniques can be brought to bear on this much-studied problem [29, 54, 65, 110, 141] in order to improve depth map resolution and robustness to indirect light. As a proof of concept, we demonstrate in the reconstruction of dense depth and albedo from individual frames of monochrome video, acquired by combining indirect-invariant SLT imaging and conventional six-pattern phase-shifting.

While SLT imaging builds on the optical probing work of Chapter 3, its premise, theory, applications, and physical implementation are different.

4.1 The Stereo Transport Matrix

We begin by relating scene geometry to the light transported from a projector to a camera in general position.

Anatomy of the stereo transport matrix Since a projector and a camera in general position define a stereo pair, their transport matrix is best understood by taking two-view geometry into account. More specifically, we classify the elements of \mathbf{T} into three categories based on the geometry of their transport paths (Figure 4.1):

- **Epipolar elements**, whose projector and camera pixels are on corresponding epipolar lines. These are the only elements of \mathbf{T} whose transport paths begin and end on rays that can intersect in 3D. By performing stereo calibration [49] and vectorizing patterns and images according to Figure 4.2, these elements can be made to occupy a known, time-invariant, block-diagonal subset of the transport matrix.
- **Non-epipolar elements**, whose projector pixel and camera pixel are not on corresponding epipolar lines. Non-epipolar elements are significant because they vastly outnumber the other elements of \mathbf{T} and *never* account for direct transport. This is because their transport paths begin and end with rays that do not intersect, so light must bounce at least twice to follow them.
- **Direct elements**, whose camera and projector pixels are in stereo correspondence, *i.e.*, they are the perspective projections of a visible surface point. Direct elements are where direct surface reflection actually occurs in the scene; although they always lie within \mathbf{T} 's epipolar blocks, their precise location is scene dependent and thus unknown. Indeed, locating the direct elements is equivalent to computing the scene's instantaneous stereo disparity map (Figure 4.3).

We can therefore express every image of the scene as a sum of three components that arise from distinct “slices” of the transport matrix:

$$\mathbf{i} = \underbrace{\mathbf{T}^{\text{D}} \mathbf{p}}_{\text{direct image}} + \underbrace{\mathbf{T}^{\text{EI}} \mathbf{p}}_{\text{epipolar indirect image}} + \underbrace{\mathbf{T}^{\text{NE}} \mathbf{p}}_{\text{non-epipolar indirect image}} \quad (4.1)$$

where the $I \times P$ matrices \mathbf{T}^{D} , \mathbf{T}^{EI} and \mathbf{T}^{NE} hold the direct, epipolar indirect, and non-epipolar elements, respectively, and are zero everywhere else.

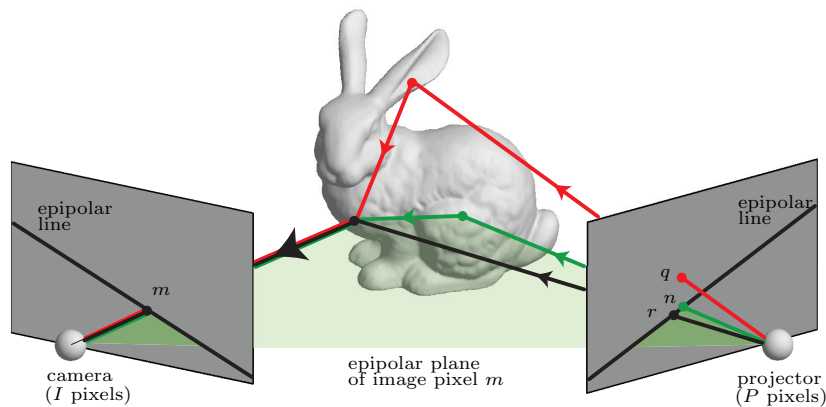


Figure 4.1: Light transport in a stereo projector-camera system. Light can reach pixel m on the image in one of three general ways: by indirect transport from an arbitrary pixel n on the corresponding epipolar line (green path); by indirect transport from a pixel q that is *not* on that line (red path); or by direct surface reflection, starting from projector pixel r on the epipolar line (black path).

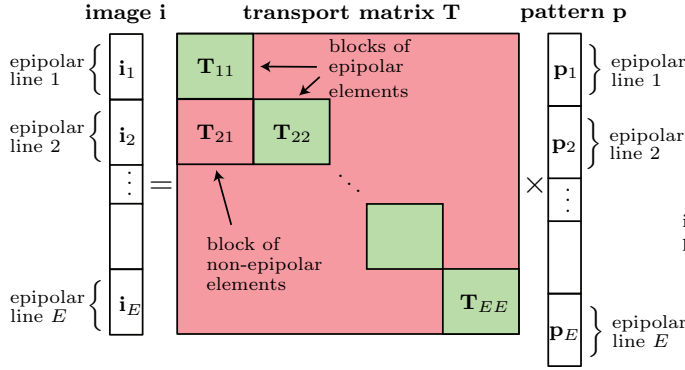


Figure 4.2: The light transport equation when patterns and images are vectorized so that consecutive pixels on corresponding epipolar lines form subvectors \mathbf{p}_e and \mathbf{i}_e , respectively. The matrix \mathbf{T} is also partitioned into $E \times E$ submatrices, where each submatrix \mathbf{T}_{ef} is known as a block and E is the number of epipolar lines. Under this vectorization scheme, block \mathbf{T}_{ef} describes transport from epipolar line f on the pattern to epipolar line e on the image. Blocks \mathbf{T}_{ee} , shown in green, contain the epipolar elements.

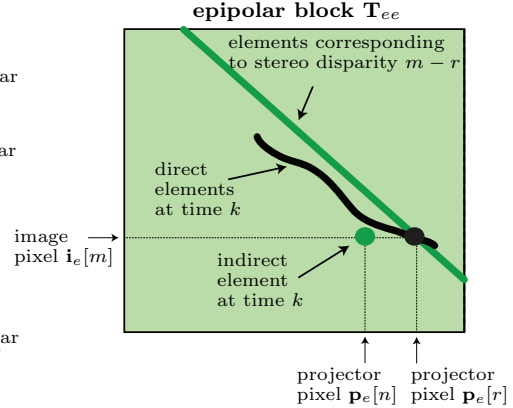


Figure 4.3: Structure of an epipolar block \mathbf{T}_{ee} . Element $\mathbf{T}_{ee}[m, r]$ describes transport from projector pixel $\mathbf{p}_e[r]$ to image pixel $\mathbf{i}_e[m]$. This element is direct if and only if the scene point projecting to both pixels is the same, *i.e.*, the point’s stereo disparity is $m - r$. The set of direct elements therefore represents the scene’s instantaneous disparity map. Conventional stereo algorithms attempt to localize this set while assuming that the transport matrix is zero everywhere else—both inside and outside its epipolar blocks.

4.2 Dominance of Non-Epipolar Transport

Although in theory all three image components in Eq. (4.1) may contribute to scene appearance, in practice their contributions are not equal. The key observation underlying our work is that the non-epipolar component is very large relative to the epipolar indirect for a broad range of scenes:

$$\mathbf{i} \approx \underbrace{\mathbf{T}^D \mathbf{p}}_{\text{direct image}} + \underbrace{\mathbf{T}^{\text{NE}} \mathbf{p}}_{\text{non-epipolar indirect image}}. \quad (4.2)$$

We call this the *non-epipolar dominance assumption*. The transport matrix is much simpler when this assumption holds because we can treat it as having a time-invariant structure with two easily-identifiable parts: the epipolar blocks, which contribute only to the direct image, and the non-epipolar blocks, which contribute only to the indirect.

To motivate this assumption on theoretical grounds, we prove that it holds for two very general scene classes: (1) scenes whose transport function for indirect elements is square-integrable and (2) generic scenes containing pure specular reflectors and transmitters. These two cases can be thought of as representing opposite extremes, with the former covering diffuse indirect transport phenomena such as diffuse inter-reflection and diffuse isotropic subsurface scattering [60] and the latter covering specular indirect transport such as caustics. In particular, we prove the following:

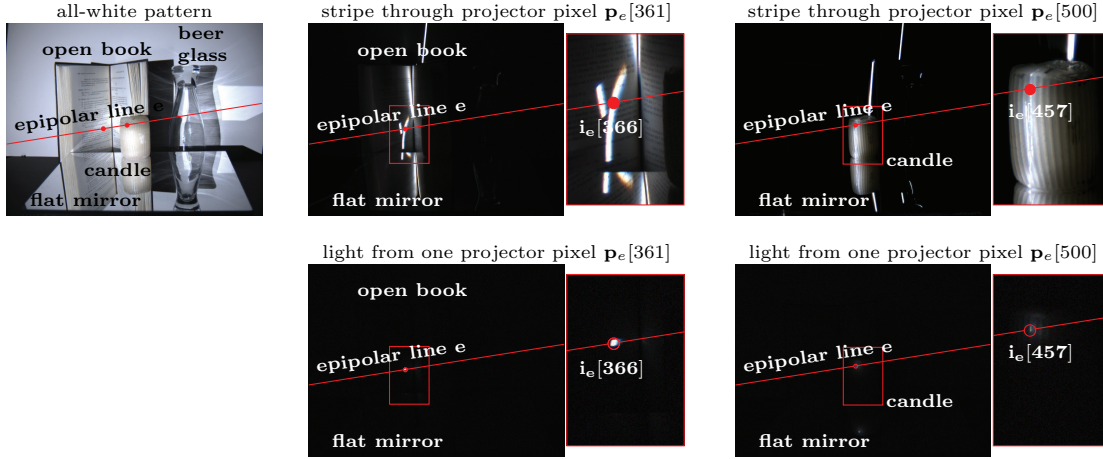


Figure 4.4: Experimental validation of non-epipolar dominance for a scene containing diffuse, translucent, refractive and mirror-like objects. From left to right: (1) View under an all-white projection pattern. (2) View when just one white vertical stripe is projected onto the scene. The many bright regions in this image occur because the stripe illuminates the book’s pages in three different ways: directly from the projector, by diffuse inter-reflection from the opposite page, and by specular reflection via the mirror. Their existence makes the scene hard to reconstruct with conventional techniques such as laser-stripe 3D scanning [18]. A magnified view of these regions is shown in the inset on the right. (3) View for another vertical stripe, part of which falls on the candle. The stripe appears very broad and poorly localized there, because of strong sub-surface scattering. (4) View when just one projector pixel illuminates the scene. Camera pixels along the epipolar line receive light travelling along both direct and epipolar indirect paths; note that, unlike (2), these camera pixels receive no light travelling along non-epipolar indirect paths. (5) View when a single pixel illuminates a point on the candle.

Proposition 1. *If \mathbf{T}^{NE} and \mathbf{T}^{EI} are discretized forms of transport functions that are square-integrable and positive over the rectified projector and image planes, then*

$$\lim_{\epsilon \rightarrow 0} \frac{\mathbf{T}^{\text{EI}} \mathbf{p}}{\mathbf{T}^{\text{NE}} \mathbf{p}} = \mathbf{0} \quad (4.3)$$

where division is entrywise, $\mathbf{0}$ is a vector of zeros, and ϵ is the pixel size for discretization.

Proposition 2. *In a generic scene, a specular transport path does not intersect any of the k -bounce specular transport paths that originate from the corresponding epipolar line for $k \geq 1$.*

See Appendix B.2 for proofs. Intuitively, both propositions are consequences of a “dimensionality gap”: the set of transport paths contributing to the epipolar indirect image has lower dimension than the set of paths contributing to the non-epipolar image (Figure 4.1). Thus contributions accumulated in one image are negligible relative to the other in generic settings.

On the practical side, we have found non-epipolar dominance to be applicable quite broadly; see Figures 4.4 and 4.5 for a detailed analysis of non-epipolar dominance in a complex scene, Figures 4.15 to 4.16 for more examples, and [94] for videos confirming the assumption’s validity.

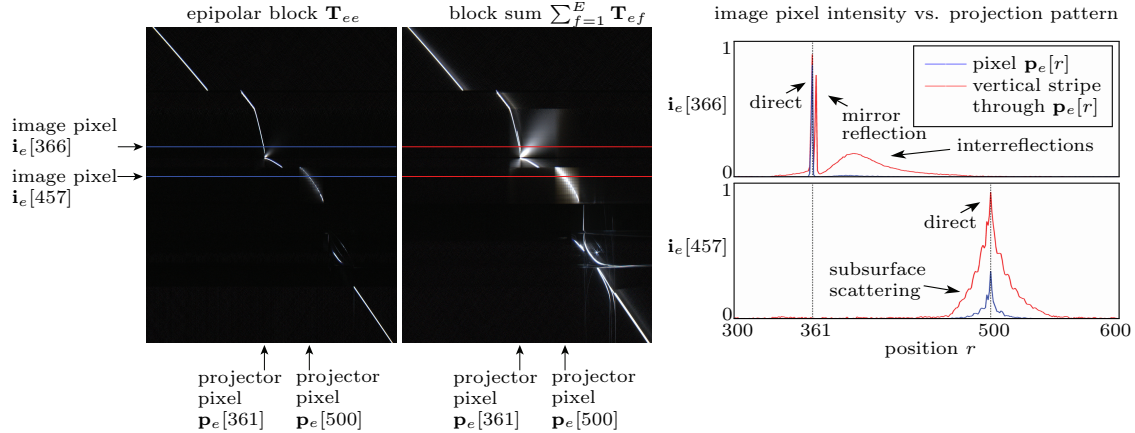


Figure 4.5: **Left:** The epipolar block \mathbf{T}_{ee} for epipolar line e of Figure 4.4. We show \mathbf{T}_{ee} using the conventions of Figure 4.3, *i.e.*, its r^{th} column comes from an image of the scene acquired with only projector pixel $\mathbf{p}_e[r]$ turned on. **Middle:** To assess the image contribution of non-epipolar transport, we acquire the block sum $\sum_{f=1}^E \mathbf{T}_{ef}$ and compare it to block \mathbf{T}_{ee} —observe that non-epipolar contributions indeed far surpass the epipolar indirect ones. To acquire the block sum, we capture images of the scene while sweeping a vertical stripe on the projector plane (see [94] for a video of the captured image sequence). The r^{th} column of the block sum is given by the pixels on epipolar line e when the stripe is at $\mathbf{p}_e[r]$. **Right:** Horizontal cross-section of \mathbf{T}_{ee} and $\sum_{f=1}^E \mathbf{T}_{ef}$ for two image pixels. Observe that \mathbf{T}_{ee} 's cross-section (blue) is sharp and unimodal whereas the block sum's (red) is trimodal for one pixel and very broad for the other.

4.3 Imaging by Structured Light Transport

The rich structure of the stereo transport matrix cannot be exploited by simply projecting a pattern onto the scene. This is because projection gives no control over how light flows through the scene: all elements of \mathbf{T} —regardless of position—will participate in image formation. To make full use of \mathbf{T} 's structure, *we structure the flow of light itself*.

Our starting point is the optical probing procedure outlined in Chapter 3. Two basic questions arise when considering the probing equation (Eq. (3.1)) for image acquisition and shape recovery: (1) what should $\mathbf{\Pi}$ be, and (2) how to design an imaging system that implements the equation? We previously restricted optical probing to static scenes and projector-camera arrangements that share a single viewpoint, none of which apply here. Below we focus on the first question—designing $\mathbf{\Pi}$ —and discuss live imaging of dynamic scenes in Section 4.4.

Conventional structured-light imaging To gain some insight, let us re-cast as a probing operation the act of projecting a pattern \mathbf{p} and capturing an image \mathbf{i} . Applying the vectorization scheme of Figure 4.2 to the light transport equation and re-arranging terms we get for epipolar line e :

$$\mathbf{i}_e = \sum_{f=1}^E \mathbf{T}_{ef} \mathbf{p}_f = \left[\sum_{f=1}^E \underbrace{(\mathbf{1} \mathbf{p}_f^T)}_{\text{block of probing matrix}} \odot \underbrace{\mathbf{T}_{ef}}_{\text{block of } \mathbf{T}} \right] \mathbf{1} \quad (4.4)$$

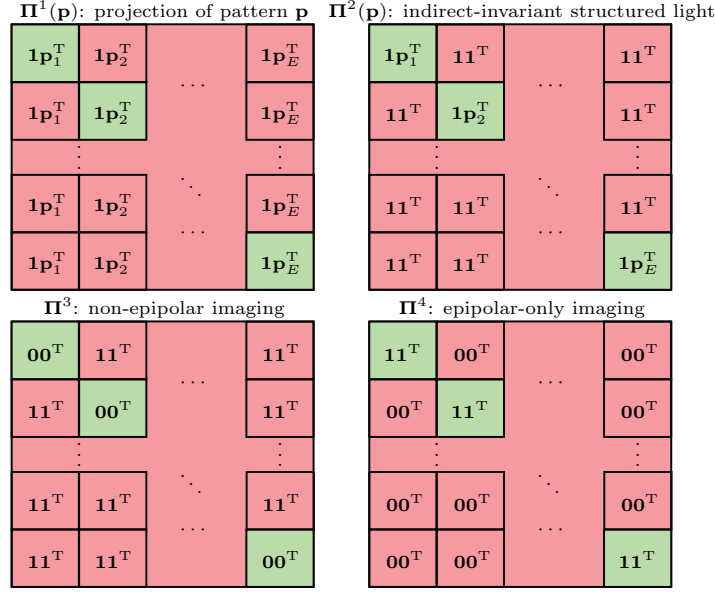


Figure 4.6: The four basic probing matrices used in Chapter 4. Their block structure mirrors the structure of \mathbf{T} in Figure 4.2.

where E is the number of epipolar lines. Eq. (4.4) implies that projecting \mathbf{p} is equivalent to probing with the matrix $\Pi^1(\mathbf{p})$ shown in Figure 4.6. Observe that if we capture images for a whole sequence of projection patterns—as is often the case in structured-light systems—the non-epipolar blocks of the probing matrix will be different for each pattern. Indirect transport will therefore contribute to each captured image differently in a way that strongly depends on the particular pattern. This makes structured-light 3D scanning difficult when indirect transport is present because its contributions cannot be easily identified and removed.

Indirect-invariant structured light The contribution of indirect transport becomes much easier to handle if we ensure it is *the same for every pattern*. Since this contribution is dominated by the non-epipolar blocks of the transport matrix, we can achieve (almost) complete invariance to indirect transport by probing with a matrix whose non-epipolar blocks are independent of \mathbf{p} . In particular, probing with the matrix $\Pi^2(\mathbf{p})$ in Figure 4.6 yields

$$\mathbf{i}_e = \underbrace{\left[(\mathbf{1p}_e^T) \odot \mathbf{T}_{ee} \right] \mathbf{1}}_{\text{direct image (depends on } \mathbf{p})} + \underbrace{\left[\sum_{f=1, f \neq e}^E \mathbf{T}_{ef} \right] \mathbf{1}}_{\text{indirect image (ambient)}}. \quad (4.5)$$

The image in Eq. (4.5) has two properties: (1) its direct component is identical to the direct component we would get by projecting \mathbf{p} conventionally onto the scene, and (2) its non-epipolar component is independent of \mathbf{p} . This independence essentially turns indirect contributions into an “ambient

light” term that does not originate from the projection pattern.¹ To see the practical significance of this independence, Figure 4.13 compares views of a scene under conventional and one-shot indirect-invariant structured light, for the same projection pattern.

An important corollary of Eq. (4.5) is that indirect-invariant structured light images can be acquired for *any* sequence of patterns—regardless of frequency content or other properties—using the corresponding sequence of probing matrices.

Non-epipolar imaging A notable special case of indirect-invariant structured light is to set \mathbf{p} to zero (matrix $\mathbf{\Pi}^3$ in Figure 4.6). This yields an image that has no contributions from direct transport. Moreover, almost all indirect light will be recorded when non-epipolar dominance holds.

Epipolar-only imaging The exact opposite effect can be achieved with a probing matrix that is zero everywhere except along the epipolar blocks (matrix $\mathbf{\Pi}^4$ in Figure 4.6). When non-epipolar dominance holds, images captured this way can be treated as (almost) purely direct.

One-shot, multi-pattern, indirect-invariant structured light All four probing matrices in Figure 4.6 produce views of the scene under a fixed illumination pattern \mathbf{p} . With probing, however, it is possible to capture—in just one shot—spatially-multiplexed views of the scene for a whole sequence of structured-light patterns, $\mathbf{p}_1, \dots, \mathbf{p}_S$. The probing matrix to achieve this can be thought of as defining a “projection pattern mosaic,” much like the RGB filter mosaic does for color (Figure 4.7). Moreover, we can confer invariance to indirect light by defining the mosaic in terms of probing matrices rather than conventional patterns.

Specifically, suppose we partition the I image pixels into S sets and let $\mathbf{b}_1, \dots, \mathbf{b}_S$ be binary vectors of size I indicating the pixel membership of each set. The matrix

$$\mathbf{\Pi}^5(\mathbf{p}_1, \dots, \mathbf{p}_S) = \sum_{s=1}^S [\mathbf{b}_s \mathbf{1}^T] \odot \mathbf{\Pi}^2(\mathbf{p}_s) \quad (4.6)$$

interleaves the rows of S indirect-invariant probing matrices. Thus, probing with this matrix yields an image containing S sub-images, each of which is a view of the scene under a specific structured-light pattern in the sequence.

¹ Other examples of ambient terms with identical behavior include image contributions from the projector’s black level and contributions from light sources other than the projector. Because such terms are often unavoidable yet easy to handle, many structured-light algorithms are designed to either recover them explicitly or be robust to their existence [110]. Non-zero ambient terms do, however, reduce contrast and may affect SNR.

color filter mosaic			6-pattern mosaic			6-pattern indirect-invariant mosaic		
R	G	R	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	$\mathbf{\Pi}^2(\mathbf{p}_1)$	$\mathbf{\Pi}^2(\mathbf{p}_2)$	$\mathbf{\Pi}^2(\mathbf{p}_3)$
G	B	G	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	$\mathbf{\Pi}^2(\mathbf{p}_4)$	$\mathbf{\Pi}^2(\mathbf{p}_5)$	$\mathbf{\Pi}^2(\mathbf{p}_6)$

Figure 4.7: Example layouts for color RGB, monochrome 6-pattern, and monochrome 6-pattern indirect-invariant structured light imaging.

4.4 Live Structured-Light-Transport Imaging

Optical probing is possible by (1) opening the camera’s shutter, (2) projecting pattern \mathbf{q}_k onto the scene, (3) using a pixel mask \mathbf{m}_k to modulate the light arriving at individual camera pixels, (4) changing the pattern and mask synchronously K times, and (5) closing the shutter. We implemented this procedure in Chapter 3 for single-image acquisition and low-resolution probing matrices using an LCD panel for pixel masking, an SLR camera for image acquisition, and $K \in [100, 1000]$.

Although the results were promising, LCDs are not suitable for video-rate (30 Hz) probing: they refresh at 30-200 Hz, limiting K to an unusable 1-6 masks/projections per frame; and they have low transmittance, requiring long exposure times.

The approach here, on the other hand, is to use a pair of off-the-shelf DMDs for projection and masking (Figure 4.8 and Figure 4.9). These devices are compact, incur no light loss and can operate synchronously at 2.7 – 24 kHz. To implement Eq. (3.1), we couple them with a conventional video camera operating at 28 FPS. This allows 96 – 800 masks/projections within the 36 ms exposure of each frame. To our knowledge, such a coupling has not been proposed before.

A major difference between LCDs and DMDs is that DMDs are *binary*. This turns the derivation of masks and projection patterns into a combinatorial optimization problem. Formally, given an *integer*² probing matrix $\mathbf{\Pi}$ and an upper bound on K , we seek a length- K rank-1 decomposition into binary vectors such that the decomposition approximates $\mathbf{\Pi}$ as closely as possible. Estimating the length of the shortest *exact* decomposition is itself NP-hard [143].

The approach below is to derive randomized decompositions of $\mathbf{\Pi}$ that approximate Eq. (3.1) in expectation. Although our experience is that this approach works well in practice, it should not be treated as optimal.

²Since any grayscale structured-light pattern \mathbf{p} must be quantized before projection, probing matrices are always integer, including $\mathbf{\Pi}^2(\mathbf{p})$.

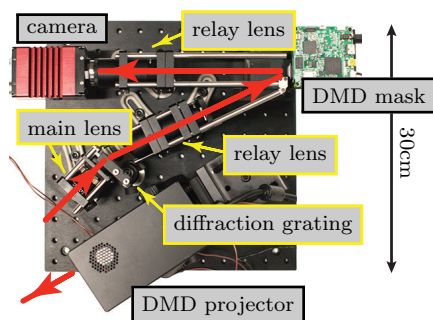


Figure 4.8: Photo of our low-speed, low-cost prototype. The projector can be detached to change the stereo baseline. The optical path is shown in red.

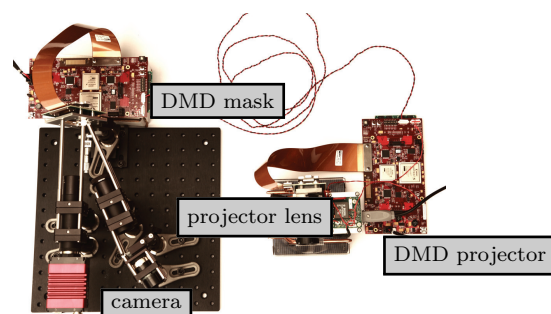


Figure 4.9: Our high-speed system. The key differences between this system and that shown in Figure 4.8 are a monochrome camera, the DMD mask, and the DMD projector.

Non-epipolar imaging Matrix $\mathbf{\Pi}^3$ is a special case where short decompositions are easy. Let \mathbf{q}_e be a pattern whose pixels are 1 along epipolar line e and 0 everywhere else and let \mathbf{m}_e be a mask that is 1 everywhere except at epipolar line e . Then it is easy to show that $\mathbf{\Pi}^3 = \sum_{e=1}^E \mathbf{m}_e (\mathbf{q}_e)^\top$. This corresponds to a sequence of mask/projection pairs where only one epipolar line is “off” in the mask and only the corresponding epipolar line is “on” in the pattern. Even though this decomposition is exact—and feasible for near-megapixel images—it has poor light efficiency because only one epipolar line is “on” at any time. To improve light efficiency we use random patterns instead, which yield good approximations that are much shorter.

Specifically, consider the random pattern

$$\mathbf{q} = \{\text{each epipolar line is 1 with probability 0.5}\} , \quad (4.7)$$

let the projection pattern \mathbf{q}_k be a sample of \mathbf{q} , and let the mask \mathbf{m}_k be equal to $\overline{\mathbf{q}_k} = 1 - \mathbf{q}_k$. See Figure 4.10, Row 1 for an example of \mathbf{q}_k , $\overline{\mathbf{q}_k}$, and \mathbf{m}_k . Taking expectations in Eq. (3.4), the epipolar line e of the expected image is given by

$$\mathbb{E}[\mathbf{i}_e] = \mathbb{E}[\overline{\mathbf{q}_e}] \circ \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \mathbb{E}[\mathbf{q}_f] = 0.25 \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \mathbf{1} \quad (4.8)$$

where $\mathbb{E}[\]$ denotes expectation. The expanded derivation is available in Appendix B.1. This is the result of probing with matrix $\mathbf{\Pi}^3$, albeit at $\frac{1}{4}$ th of the “ideal” image intensity.³ Note that corresponding epipolar lines are never on at the same time in the pattern and mask; thus no epipolar transport path ever contributes to the captured image.

Epipolar-only imaging Matrix $\mathbf{\Pi}^4$ is a special case at the other extreme, where *no* short rank-1 decompositions exist. Since $\mathbf{\Pi}^4 = \mathbf{\Pi}^1(\mathbf{1}) - \mathbf{\Pi}^3$, we compute the result of probing with $\mathbf{\Pi}^4$ by subtracting two adjacent video frames—one captured by projecting an all-white pattern and one captured by non-epipolar imaging. Naturally, two-frame motion estimation may be necessary to handle fast-moving scenes (but we do not estimate motion in our experiments).

Indirect-invariant structured light A perhaps counterintuitive result is that even though epipolar-only imaging requires two frames, indirect-invariant structured light requires just one. This is important because probing with matrix $\mathbf{\Pi}^2()$ is all we need for reconstruction with structured light. Let \mathbf{p} be an arbitrary structured-light pattern scaled to $[0, 1]$. Define mask \mathbf{m}_k to be a sample of \mathbf{q} from Eq. (4.7) and the pattern to be

$$\mathbf{q}_k = \mathbf{m}_k \text{ XNOR } \mathbf{r}_k \quad (4.9)$$

$$\stackrel{\text{def}}{=} \mathbf{m}_k \circ \mathbf{r}_k + \overline{\mathbf{m}_k} \circ \overline{\mathbf{r}_k} \quad (4.10)$$

³Intuitively, since half the epipolar lines are “off” in the pattern and the mask, only a quarter of the total light is transported from projector to camera.

where XNOR is the exclusive nor operator of binary vectors \mathbf{m}_k and \mathbf{r}_k , and \mathbf{r}_k is a sample of yet another random pattern:

$$\boldsymbol{\tau} = \{\text{pixel } n \text{ on epipolar line } e \text{ is } 1 \text{ with probability } \mathbf{p}_e[n]\} . \quad (4.11)$$

A pictorial illustration of Eq. (4.10) can be found in Figure 4.10 on Row 4, with example random binary patterns \mathbf{r}_k sampled from $\boldsymbol{\tau}$ (Eq. (4.11)) shown on Row 3. From calculations similar to Eq. (4.8), the expected image is

$$\mathbb{E}[\mathbf{i}_e] = 0.5\mathbf{T}_{ee}\mathbf{p}_e + 0.25 \sum_{f=1, f \neq e}^E [\mathbf{T}_{ef}\mathbf{p}_f + \mathbf{T}_{ef}(\mathbf{1} - \mathbf{p}_f)] \quad (4.12)$$

$$= \underbrace{0.5\mathbf{T}_{ee}\mathbf{p}_e}_{\substack{\text{direct image} \\ \text{(depends on } \mathbf{p})}} + \underbrace{0.25 \sum_{f=1, f \neq e}^E \mathbf{T}_{ef}\mathbf{1}}_{\text{indirect image (ambient)}} , \quad (4.13)$$

which is equivalent to the result of probing with $\mathbf{\Pi}^2()$; see Appendix B.1 for derivation.

One-shot, multi-pattern, indirect-invariant imaging Here we use the mask for indirect-invariant imaging and temporally multiplex S random projection patterns—each defined by Eq. (4.10) and corresponding to a different structured-light pattern—across our “budget” of K total projections per video frame. Row 5 of Figure 4.10 illustrates the construction of a multi-pattern indirect-invariant mask and corresponding projection pattern for $S = 6$. After the video is recorded, we “demosaic” each frame \mathbf{i} independently to infer S full-resolution images, one for each structured-light pattern. Following work on compressed sensing [55, 107] we do this by solving for S images that reproduce frame \mathbf{i} and are sparse under a chosen basis \mathbf{W} :

$$\min_{\mathbf{i}_1, \dots, \mathbf{i}_S} \left\| \mathbf{W}^T \begin{bmatrix} \mathbf{i}_1 & \dots & \mathbf{i}_S \end{bmatrix} \right\|_p \quad (4.14)$$

$$\text{subject to} \quad \left\| \sum_{s=1}^S \mathbf{b}_s \circ \mathbf{i}_s - \mathbf{i} \right\|_2 \leq \epsilon \quad (4.15)$$

where $\|\cdot\|_p$ is a sparsity-inducing norm, the vector \mathbf{b}_s holds pixel memberships for pattern s , and the scalar ϵ is a user-specified error tolerance to account for the noise in image \mathbf{i} .

We use the $\ell_{2,1}$ matrix norm here (the sum of the ℓ_2 norms of the rows) because it promotes group sparsity and thus concentrate non-zero terms along rows of matrix $\mathbf{W}^T [\mathbf{i}_1 \dots \mathbf{i}_S]$. This is a sensible norm for one-shot imaging because the visual features of the S images $\{\mathbf{i}_s\}$ are largely consistent, and we therefore expect the non-zero elements of sparse vectors $\{\mathbf{W}^T \mathbf{i}_s\}$ to have the same structure across views.

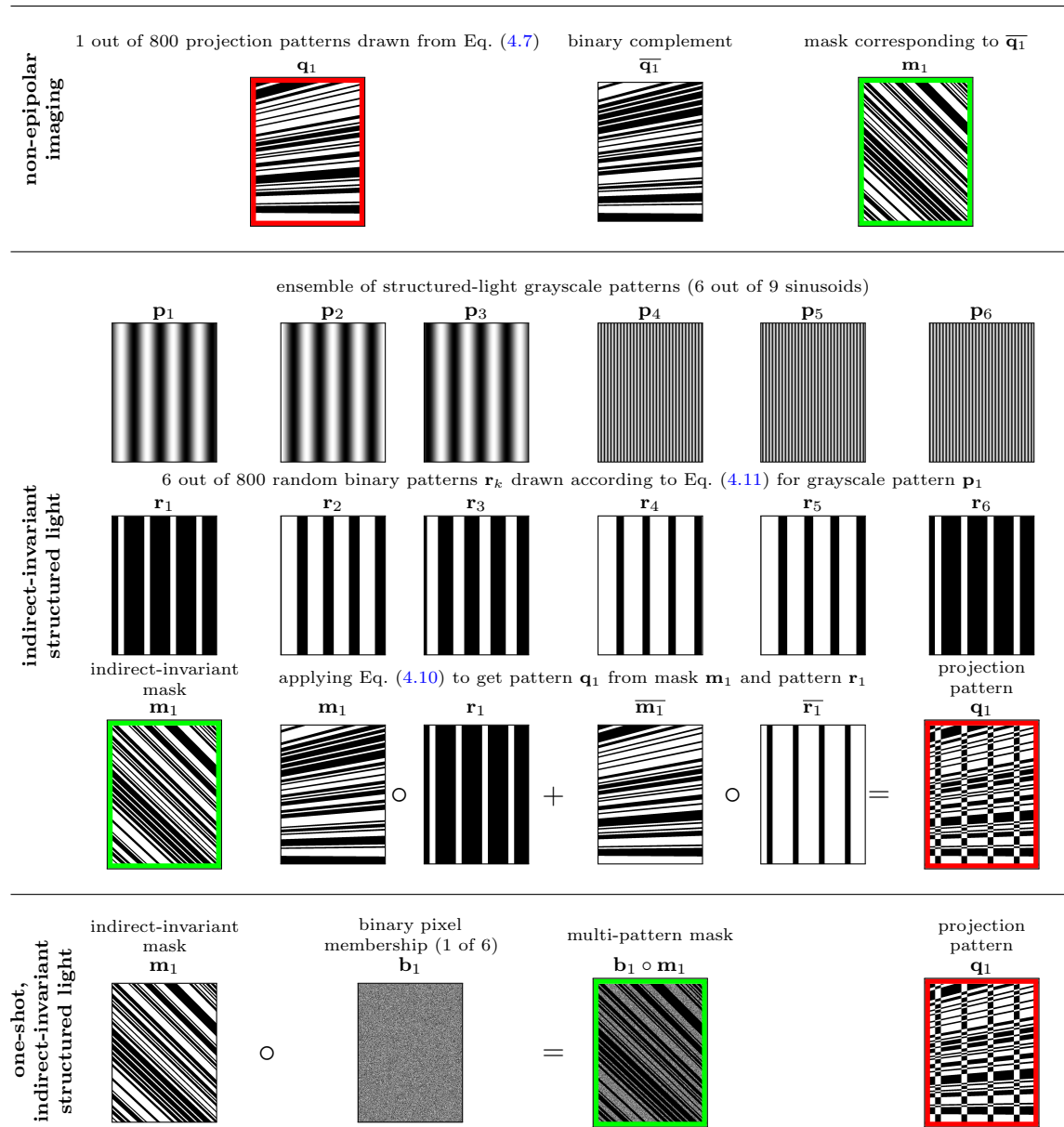


Figure 4.10: Deriving random pattern/mask pairs for three cases of SLT imaging. The derived patterns and masks are indicated with red and green borders, respectively. **Row 1:** For non-epipolar imaging, the patterns and masks are constant along epipolar lines, with approximately half of them “on.” **Row 2:** Six of the nine structured-light patterns we used. **Rows 3-4:** The masks for indirect-invariant structured light are identical to those for non-epipolar imaging but the projection patterns differ. To generate them for a given grayscale structured-light pattern, we first generate a random sequence of binary patterns (Row 3) and then use that sequence, along with the sequence of masks, to compute the projection patterns. Row 4 shows one such example. **Row 5:** We generate pattern/mask pairs for 6-shot imaging as follows: (1) create 6 random binary images representing pixel membership for each pattern; (2) generate a sequence of 132 indirect-invariant binary pattern/mask pairs for each of 6 grayscale structured-light patterns, as outlined in Rows 2-4; (3) use the 792 projection patterns as is, and (4) multiply the masks element-wise with the associated pixel memberships. Row 5 shows one such calculation, for grayscale pattern p_1 .

4.5 Implementation

Experimental prototypes We created two additional experimental systems for performing optical probing:

- a low-speed, low-cost system for video-rate non-epipolar and epipolar imaging (Figure 4.8) whose components are listed in Table C.1; and
- a high-speed system for indirect-invariant structured light shape acquisition and one-shot multi-pattern structured light (Figure 4.9), whose components are also listed in Table C.1.

Our low-speed, low-cost system included a color AVT GT1920C camera for acquisition, a Texas Instruments LightCrafter for pixel masking and a 100 lm Keynote Photonics LightCrafter kit for projection. The DMDs were synchronized at 2.7 kHz, permitting $K = 96$ patterns and masks per video frame. The camera and DMD resolutions were quite different— 1936×1456 versus 608×684 —with each DMD pixel mapping to a 2×2 block of camera pixels. System calibration consists of computing the epipolar geometry between the two DMDs. We did this by first computing correspondences between the camera and each DMD separately. Patterns are uploaded to both DMDs once, at the beginning of an imaging session.

For our high-speed system, we used a monochrome AVT GT1920 camera and a pair of high-end DMDs from Texas Instruments (DLi 4130) with a 2000 lm light source. These operate at 22.2 kHz, permitting $K = 800$ patterns per video frame. Although the DMD resolution was fairly high at 1024×768 , its effective resolution was much lower, 484×364 , because of the different physical dimensions and orientation of the camera sensor and DMD.

In one-shot multi-pattern imaging, the effective DMD resolution was even lower, 256×256 , because of the scene’s limited extent within the camera’s field of view.

Non-epipolar mask & projection patterns We use random mask/projection pairs like those shown in Row 1 of Figure 4.10. To reduce the sensation of flicker by users who are physically present during video acquisition, we generate a random sequence of $K/2$ mask/projection pairs and then generate a second mask/projection sequence whose projection patterns are the binary complement of the first $K/2$ projection patterns. This ensures a stable perception because the image integrated by the eye (or by a mask-less camera) over the period of one video frame corresponds to a view of the scene under an all-white projection pattern.⁴

For non-epipolar imaging, it is also important to ensure that no direct light “leaks” accidentally through the DMD mask. Such leaks can occur because of pixel misalignments between the DMD mask and the camera’s sensor; because of the binary rasterization of epipolar lines; and because of projector/camera defocus. To make non-epipolar acquisition robust to such effects, we slightly dilate the “off” regions on the generated masks, similar to the approach taken in Figure 3.6 for LCD masks. This reduces the occurrence of such leaks at the expense of a slight reduction in light efficiency. We found this approach to be very effective in practice.

⁴We emphasize that flicker is a purely subjective sensation that may be experienced by users who view the scene directly, without the benefit of the DMD mask. In particular, flicker *does not* occur in the videos captured by our prototypes.

Epipolar-only patterns We generate epipolar-only video by operating the camera at 56 FPS and configuring the DMD of our low-speed prototype as follows:

- *odd video frames*: display $K/2$ all-on mask/projection pairs
- *even video frames*: display a sequence of $K/2$ non-epipolar mask/projection pattern pairs.

Epipolar-only video at 28 FPS is generated by scaling the odd frames by 0.25 to account for the reduced intensity of non-epipolar imaging (Eq. (4.8)) and subtracting in real time the even frames from the scaled odd ones.

Indirect-invariant patterns We generate a sequence of K mask/projection pairs for each of S grayscale structured-light patterns, as illustrated in Rows 2-4 of Figure 4.10. We then capture one RAW image of the scene for each of the S generated mask/projection sequences. These S images are supplied, unaltered, to the 3D reconstruction algorithm.

One-shot, multi-pattern, indirect-invariant patterns We generate a sequence of K mask and pattern pairs, as outlined in Row 5 of Figure 4.10, and upload them to the DMDs. We then apply the algorithm outlined in Section 4.4 independently to each frame of the RAW live video stream.

4.6 Results

Non-epipolar and epipolar-only imaging We used our low-speed, low-cost prototype with a total of $K = 96$ mask/projection patterns per frame for non-epipolar and epipolar-only imaging. For calibration, we computed the epipolar geometry between the two DMDs by first relating them to the image plane. Overall resolution was equal to the resolution of our DMDs, *i.e.*, 608×684 . See Figures 4.11, 4.12, 4.15, and 4.16 for examples of non-epipolar and epipolar-only images.

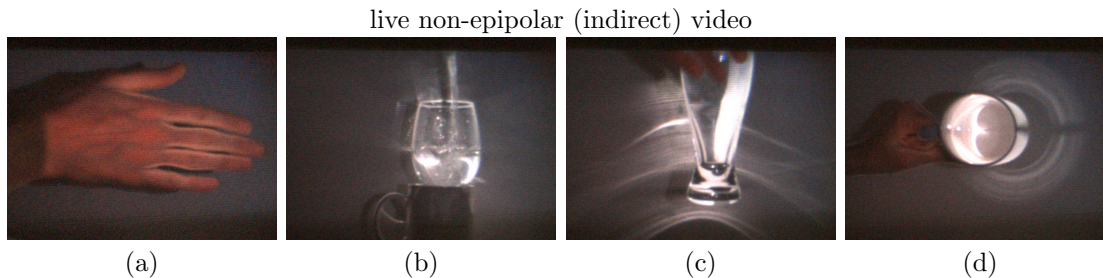


Figure 4.11: Snapshots from RAW live non-epipolar (indirect) video. (a) A hand; note the vein pattern and the inter-reflections between fingers. (b) Pouring water into a glass. (c) Refractions and caustics from a beer glass. (d) Caustics formed inside a mug from specular inter-reflections; note the secondary reflections to the board behind the mug and from the board onto the mug’s exterior surface. See Figure 4.15 for more non-epipolar images and [94] for videos.

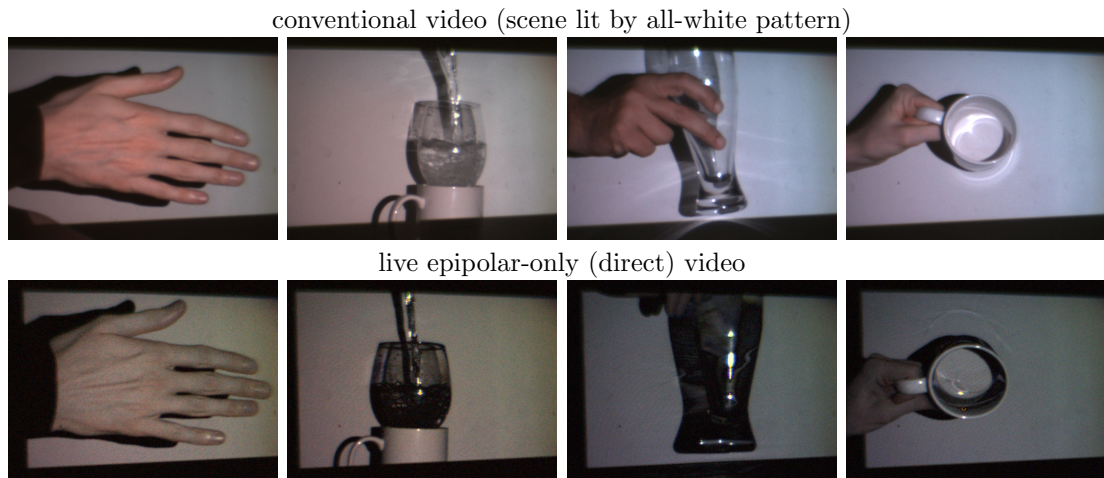


Figure 4.12: Frames from conventional and epipolar-only (direct) video, corresponding to the live non-epipolar video frames of Figure 4.11. From left-to-right: (1) A hand; note the significant difference in apparent color of the hand in the non-epipolar and epipolar-only components, due to sub-surface absorption and direct surface reflection, respectively. (2) Pouring water into a glass demonstrates our ability to successfully image highly-complex, time-varying phenomena. The water appears dark in the epipolar-only image because the light refracted by the water does not satisfy epipolar constraints. (3) This beer glass appears essentially opaque in the epipolar-only component. Again, this is because the light transmitted through the glass undergoes refraction, yielding non-linear paths that almost never lie on a single epipolar plane. (4) A mug. Note that artifacts appear on the white background; these occur because the mug moved very quickly during acquisition, and the frame-differencing we do for epipolar-only imaging caused ghosting.

Indirect-invariant structured light We used high-end DMDs and a monochrome camera to capture $S = 9$ indirect-invariant structured light frames (Figure 4.13) with $K = 800$ patterns/masks per frame. We then supply these frames as input to a reconstruction algorithm to compute 3D shape (Figure 4.14), and compare our approach to conventional approaches. The effective DMD resolution was approximately 484×364 . The scenes occupied a 40^3 cm^3 volume about 70 cm away from the camera. To show the effectiveness of SLT imaging, we chose the most basic pattern and technique to efficiently solve the correspondence problem outlined in Section 1.4.2—phase-shifting [110] with 9 sinusoids total, at frequencies 1, 8 and 64. We also demonstrate an example of indirect-invariant structured light on a refractive object in Figure 4.17.

Dense depth and albedo from one shot We used $S = 6$ sinusoids at frequencies 4 and 32 for the experiment in Figure 4.18, and a random assignment of pixels to sinusoids, rather than the regular assignment illustrated in Figure 4.7. We recorded multi-pattern, indirect-invariant video at 28 FPS and reconstructed each frame independently by (1) solving for the 6 demosaiced patterns using SPGL1 [124] for optimization and the JPEG2000 wavelet basis, and (2) using them to get per-pixel depth and albedo. Our reconstruction procedure took, as input, a RAW frame cropped to a 1024×1024 region of interest (with an effective DMD resolution of 256×256), and recovered the depth and albedo in under 8 minutes per frame on an Apple iMac with a 2.8 GHz Intel Core i7 processor and 16 GB of memory.

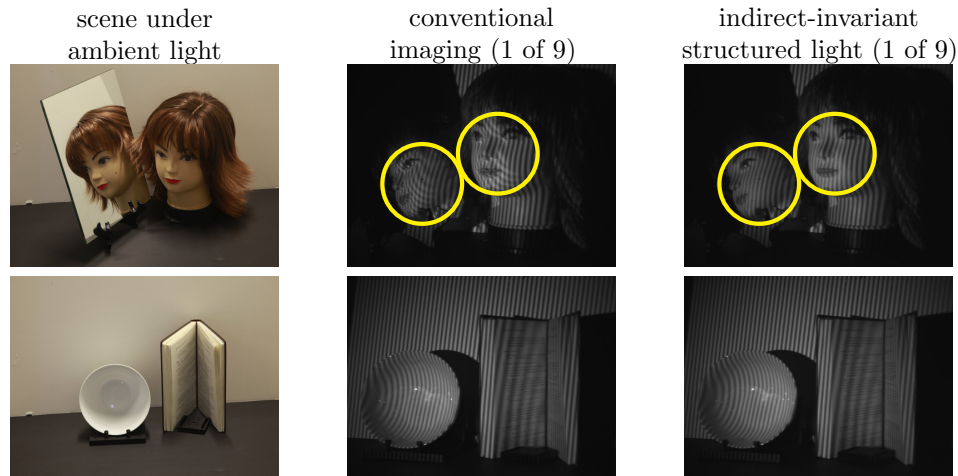


Figure 4.13: We imaged the scene on the left in two ways: (1) projecting 9 phase-shifted patterns directly onto it and (2) capturing indirect-invariant structured light images for the same patterns. For the top row, note that the conventional image contains “double fringes” (circled) from secondary reflections whereas the indirect-invariant one does not; this “double fringe” effect occurs because of the interference between the phase-shifted pattern transmitted through the direct channel and the same pattern specularly reflected by the mirror. Although the interference is less prominent in the bottom row, the conventional images are equally problematic for 3D reconstruction.

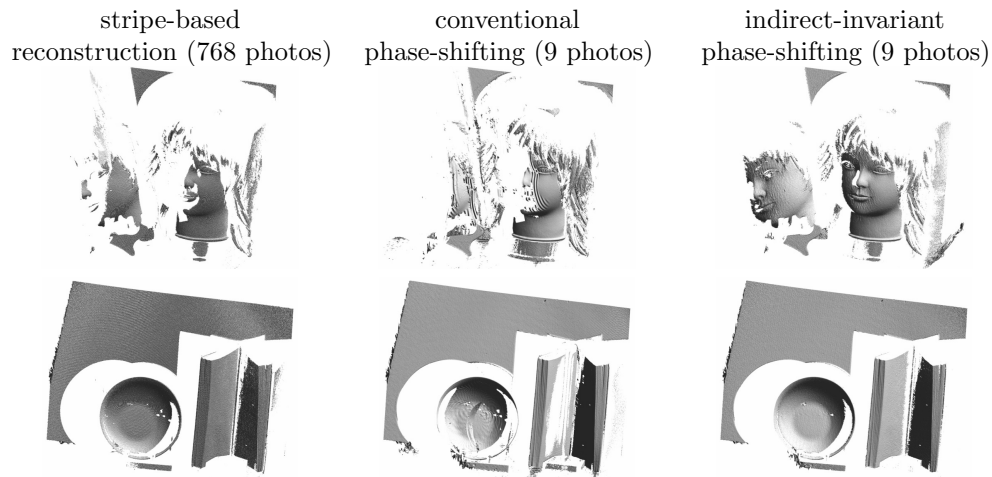


Figure 4.14: This figure contains 3D results for the two scenes shown in Figure 4.13, using a conventional stripe-based reconstruction, conventional phase-shift reconstruction, and our indirect-invariant phase-shift reconstruction. **Column 1:** We first show reconstruction results obtained by sweeping a vertical stripe across the scene, as done in conventional triangulation-based 3D laser scanning (768 images total). **Column 2:** The conventional phase-shift reconstruction procedure takes 9 input images acquired by conventional projection of phase-shifted patterns to compute RAW 3D points. **Column 3:** Our indirect-invariant phase-shift results use the same 9 patterns and reconstruction algorithm as in conventional phase-shifting, combined with indirect-invariant structured light imaging. The phase-shift reconstruction algorithm fails catastrophically for the conventionally-acquired images, whereas with SLT imaging it is able to reconstruct even the hidden side of the face, from the mirror’s indirect view. Also note that, despite the fact that stripe scanning relies on a much larger input dataset, our indirect-invariant approach produces comparable results for the bowl scene and a far more complete model for the face scene.

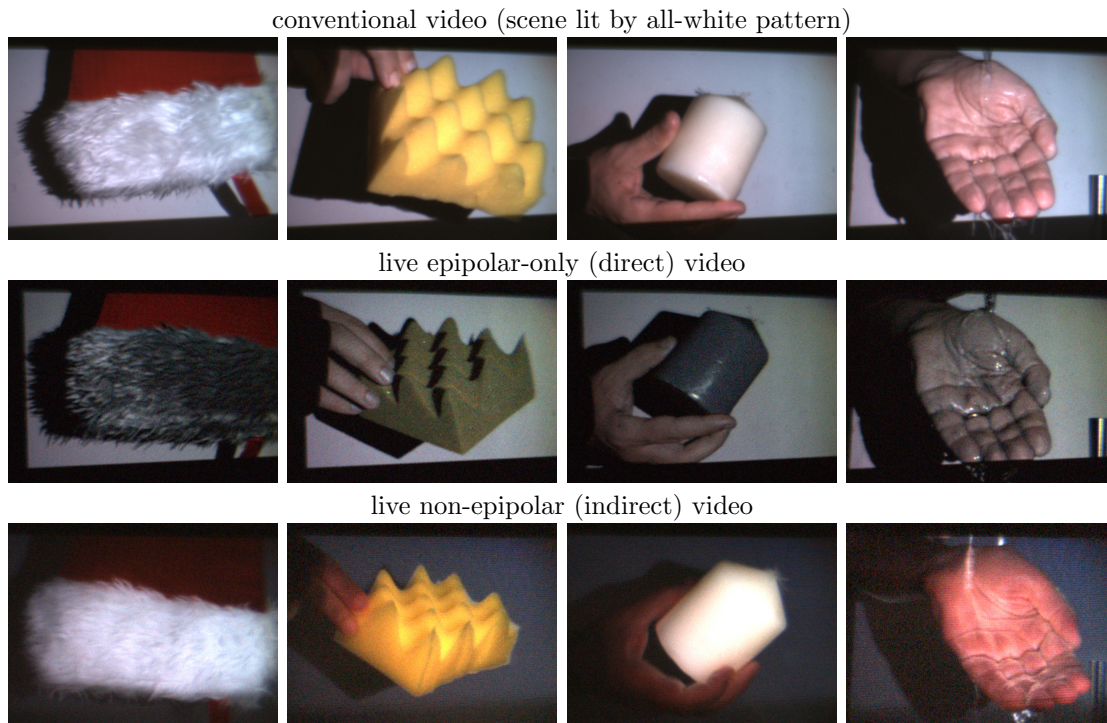


Figure 4.15: Frames from conventional, epipolar-only (direct), and non-epipolar (indirect) video. From left to right: (1) Faux-fur; note the marked difference between the epipolar-only component, which appears very shiny due to direct near-specular reflection, versus the diffuse appearance of the non-epipolar component, caused by sub-surface scattering. (2) A piece of packing foam. (3) A translucent candle. The non-epipolar frames of both the packing foam and candle demonstrate that the color of volumetric or translucent materials is often attributed to light traveling through the sub-surface. (4) Water flowing over a hand; the non-epipolar frame makes apparent the very dramatic change in a hand's reflectance properties when water flows over it. We hypothesize that these changes are caused by scattering in the thin film of water flowing over the hand.



Figure 4.16: Frames from a conventional video (left) and RAW non-epipolar video (right) of a tight-fitting latex glove. Note the seemingly transparent nature of the glove in non-epipolar mode. The hand is dark in regions where the glove makes contact with the skin; otherwise, light scatters within cavities between the glove and skin.

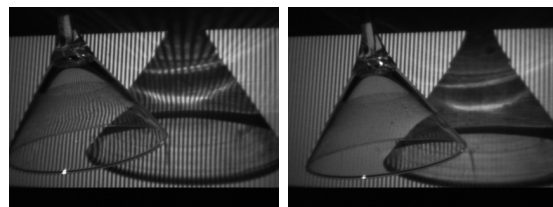


Figure 4.17: Conventional (left) and indirect-invariant (right) structured light video frames of a martini glass lit by a high-frequency pattern. Note that the pattern appears to have disappeared from regions where light undergoes refraction, with the exception of degenerate regions where light paths are doubly refracted.

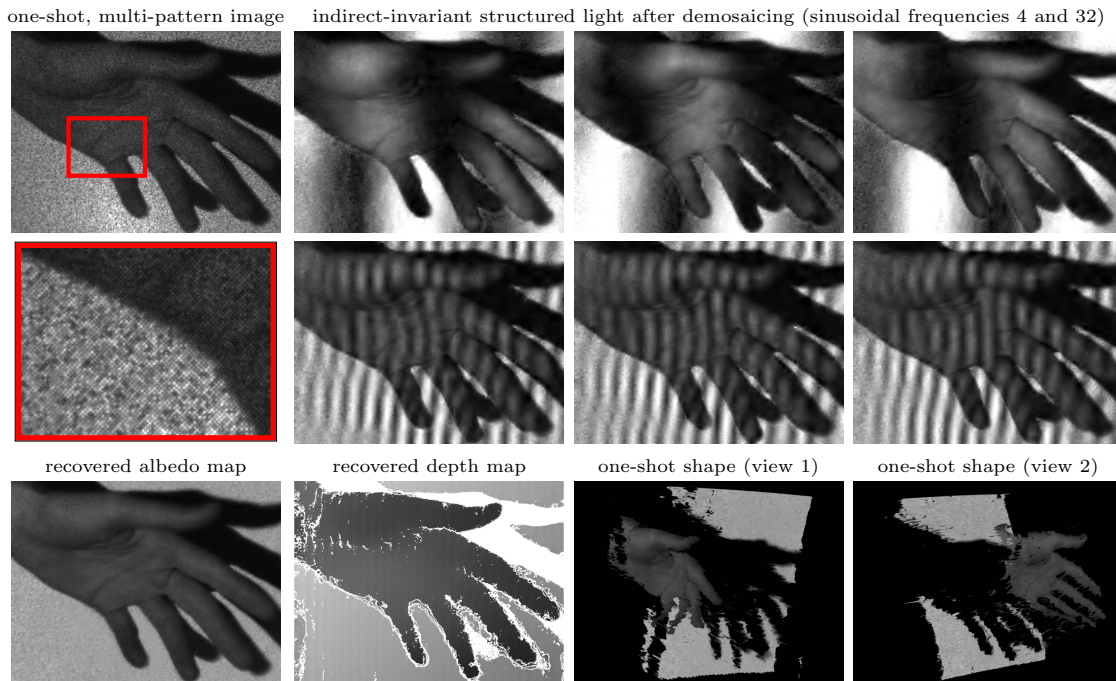


Figure 4.18: Reconstructing dense depth and albedo from a video (frame 131 of 169) of a moving hand, recorded live using one-shot, indirect-invariant, multi-pattern imaging. From this frame, our demosaicing algorithm recovers 6 full-resolution indirect-invariant structured light images of the hand, for 6 sinusoidal patterns. These images yield albedo and depth maps on the bottom, and texture-mapped geometry shown from two viewpoints.

4.7 Summary and Contributions

We believe that optical-domain processing—and SLT imaging in particular—offers a powerful new way to analyze the appearance of complex scenes, and to boost the abilities of existing reconstruction algorithms. We have demonstrated the following:

- there exists a crucial link between stereo geometry and light transport, where direct light always obeys epipolar geometry and indirect light overwhelmingly does not;
- a camera and projector in general positions can exploit this geometry to manipulate the direct and indirect light flowing through a fully-general, time-varying scene; and
- SLT is well-suited for making existing structured-light shape recovery algorithms robust to indirect transport, and can even turn them into one-shot methods for dynamic 3D capture.

Although our focus was mainly on monochromatic light and conventional cameras, SLT imaging depends on neither; in particular, we explore its connection to time-of-flight imaging in Chapter 6. Last but not least, although these prototypes rely on DMD masks/projectors and several optical components, we demonstrate a new *energy-efficient* prototype in the next chapter where per-pixel processing is implemented directly on the sensor.

Chapter 5

Energy-Efficient Probing by Homogeneous Matrix Factorization

When we capture an image under active illumination, the power of the light source matters a lot: all things being equal, brighter sources will send more radiant energy to a sensor during an exposure, producing a brighter image and a higher SNR. The brightness of the source, however, is just one way to control how much light reaches the sensor. For example, we know that there exist both efficient codes (Sections 3.2.2) and not-so-efficient codes (Section 3.2.1) for optical probing, although we have said very little so far about their optimality. This brings up a natural question: how should we compute codes that maximize *energy efficiency*, *i.e.*, the radiant energy that can be transmitted from sources to sensors for a given imaging task, power, and exposure time?

Studies of this problem began in the 1960s for the special case of arrangements with just three active components: a light source that is always turned on, a light-blocking mask that is controlled by a binary code, and a sensor [59, 62]. The optimal sequence of codes for this *single-layer* coding arrangement is derived from the Hadamard matrix [50] and enjoys widespread use [111], mainly because most conventional computational imaging systems are arranged this way.

Unfortunately, neither the original Hadamard multiplexing theory nor its recent extensions [16, 86, 87] apply to *multi-layer* coding arrangements, such as the mask/projection pattern pairs used in optical probing. As a result, the problem of computing energy-efficient codes for such arrangements is poorly understood—and even less is known about how to design them optimally. These problems are especially relevant for live imaging, where short exposures and low-power restrictions leave no room for wasting light.

We present a mathematical framework to derive energy-efficient codes and present a novel experimental testbed for live energy-efficient optical probing. We make the following contributions:

Energy-efficient codes for probing light transport We give a principled way to compute code sequences for optical probing. The code sequences in the previous chapters were constructed somewhat heuristically. In contrast, the energy-efficient codes we compute in this chapter transmit more radiant energy to the camera for a given light source power (*i.e.*, brighter images, all else

being equal) and produce superior images for a given amount of radiant energy transmitted (*i.e.*, fewer artifacts, all else being equal). We confirm these gains experimentally with our high-speed DMD-based prototype (Figure 4.9).

Live energy-efficient probing with lasers and rolling shutters Although energy-efficient codes do improve image quality, far bigger gains are possible by optimizing the system itself. To this end, we demonstrate highly efficient transport probing with a novel combination of a low-power laser projector and a rolling shutter camera. This laser-based prototype (1) captures epipolar-only and non-epipolar live video without wasting any light; (2) captures live epipolar structured-light video of very bright scenes—even a light bulb that has been turned on; (3) reconstructs 3D objects in challenging conditions such as strong indirect and ambient light; and (4) records live video from a projector’s—rather than the camera’s—point of view (*i.e.* dual *videography*).

Unified mathematical model for redistributive projection We show that despite the significant differences between our DMD- and laser-based prototypes, it is possible to describe both of them with the same underlying model. This makes it easy to compare their energy efficiency, and to define an efficiency criterion for code sequences that applies to a whole spectrum of projection devices—from conventional DMD projectors to rapidly-steerable laser beams.

Energy-efficient codes by homogeneous factorization We show that finding energy-efficient codes under our model is equivalent to solving a *homogeneous* constrained matrix factorization problem, which captures the physical constraints on the mask and projection patterns. We solve this factorization problem by minimizing an objective function that includes the projective tensor norm; we use the fast local optimizer of Haeffele et al. [47] in our implementation, to compute energy-efficient codes for our DMD-based prototype.

Impulse illumination is globally optimal We show that the global minimum of our objective function can be derived in closed-form and is actually very simple: it is a sequence of impulse illuminations produced by a projector having the smallest possible redistribution ratio for the exposure time, *i.e.*, it emits all available radiant energy into a tight beam that can be redirected very quickly. A key corollary of this result is that our laser-based prototype, whose illumination sequence is fixed by the hardware and not under our control, is globally optimal for the tasks of epipolar-only and non-epipolar imaging. Moreover, global optimality holds even in the presence of ambient light. For the other tasks, however, a significant gap from global optimality still exists. This can only be closed with yet-unavailable hardware for more flexible laser projection and electronic masking.

Energy-efficient codes for redistributive projectors Of course, between the two extremes of ideal impulse projection (minimum redistribution ratio) and conventional mask-based projection (maximum redistribution ratio) lies a whole spectrum of arrangements that redistribute light partially [44, 57, 83]. We explore this spectrum briefly through simulations, noting that a rapid “phase transition” seems to occur in our energy-efficient illumination codes, with pure impulses on one end switching to dense codes on the other.

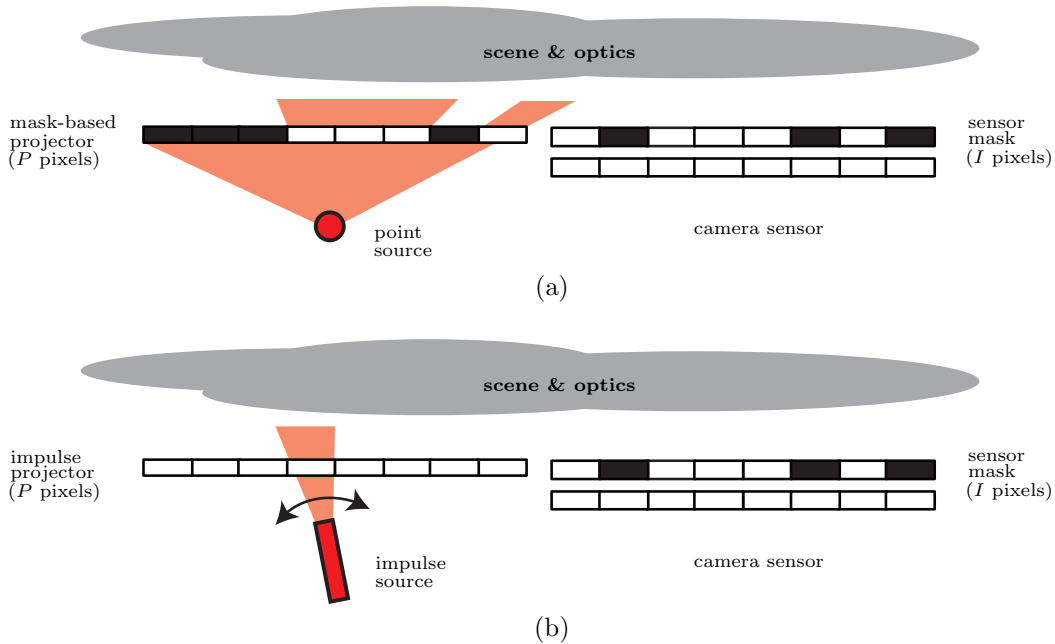


Figure 5.1: Two-layer arrangements considered in this work. **(a)** At the low end of the efficiency spectrum are projectors that always spread light to all pixels on a mask (*i.e.*, redistribution ratio is P). These projectors waste energy because they create patterns by blocking light—and that light is not used for imaging. **(b)** At the other extreme are “impulse” projectors, able to concentrate all their energy to just one pixel (*i.e.*, redistribution ratio is 1). In between are arrangements that can concentrate light partially, which we also analyze.

5.1 Lights, Masks and Energy Efficiency

We begin by revisiting the oft-used concepts of a projection pattern \mathbf{p} and a mask \mathbf{m} , and focus on their physical units and constraints. These lead to a definition of energy efficiency for the two-layer mask/projection sequences used in optical probing.

5.1.1 Redistributive Projection and Sensor Masking

We assume that the projector’s light source is always on and emits light at a constant rate Φ , measured in watts. Illuminating a scene for an exposure time T means that the total radiant energy generated by the source is ΦT and given in joules.

Each element of the illumination pattern \mathbf{p} gives the radiant energy emitted from a specific projector pixel during the exposure time. The sum of all elements of \mathbf{p} cannot exceed the radiant energy generated by the source:

$$0 \leq \mathbf{p}, \quad \|\mathbf{p}\|_1 \leq \Phi T \quad (5.1)$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm of a vector.

The redistribution ratio of a projector A key measure of projector flexibility is how well the projector can “channel” to individual pixels all the radiant energy generated by the source. We express this ability as an upper bound on the individual elements of \mathbf{p} :

$$\|\mathbf{p}\|_\infty \leq \Phi T / \varrho \quad (5.2)$$

where $\|\cdot\|_\infty$ is the ℓ_∞ norm, giving the largest element of a vector, and ϱ is a projector-specific parameter we call the redistribution ratio, that is equal to the ratio of the bounds in Eqs. (5.1) and (5.2). For a P -pixel projector, this ratio takes values between 1 and P and models energy redistribution: the larger its value is, the lower the energy we can send through any one pixel, and the more energy we waste when projecting a pattern with just one pixel turned on (or just a few).

The specific value of ϱ depends on the projection technology. At the far end of the range, with $\varrho = P$, are conventional projectors (Figure 5.1(a)). These projectors (*e.g.* those using DLP and LCD technology) use mirrors and lenses to distribute light evenly over a controllable mask, which may then block all or part of it. Any one of the P projector pixels emits at most $\frac{1}{P}$ th of the total available light ΦT . The near end of the range, with $\varrho = 1$, represents an idealized projector that is perfectly efficient (Figure 5.1(b)). This projector can send all its light through just one pixel for the *entire* exposure time T but can also distribute it—without any blocking—according to an arbitrary illumination pattern \mathbf{p} . Between these two extremes lies a whole spectrum of projection technologies that approach this ideal to a greater or lesser extent [19, 44, 57].

Norm bound on illumination vectors The ℓ_1 and ℓ_∞ bounds on \mathbf{p} can be written as

$$0 \leq \mathbf{p}, \quad \|\mathbf{p}\|_{\dagger\varrho} = \max\left(\frac{\varrho\|\mathbf{p}\|_\infty}{\Phi}, \frac{\|\mathbf{p}\|_1}{\Phi}\right) \leq T \quad (5.3)$$

where $\|\cdot\|_{\dagger\varrho}$ is the max of two norms and therefore also a norm. These bounds are useful in three ways. First, we can optimize arrangements with very different light redistribution properties by adjusting the redistribution ratio. Second, the dependence on exposure time makes a distinction between systems that conserve energy and those that merely conserve power (but require long exposures [44]). Third, they explicitly account for timescale-dependent behavior (*e.g.*, raster-scan projectors, like the MEMS projector used in our laser-based prototype, can act like a beam, light sheet, or point source depending on T).

Norm bound on mask vectors Each element of the mask vector \mathbf{m} is a unitless scalar between zero and one describing attenuation at a specific sensor pixel:

$$0 \leq \mathbf{m}, \quad \|\mathbf{m}\|_\infty \leq 1 \quad (5.4)$$

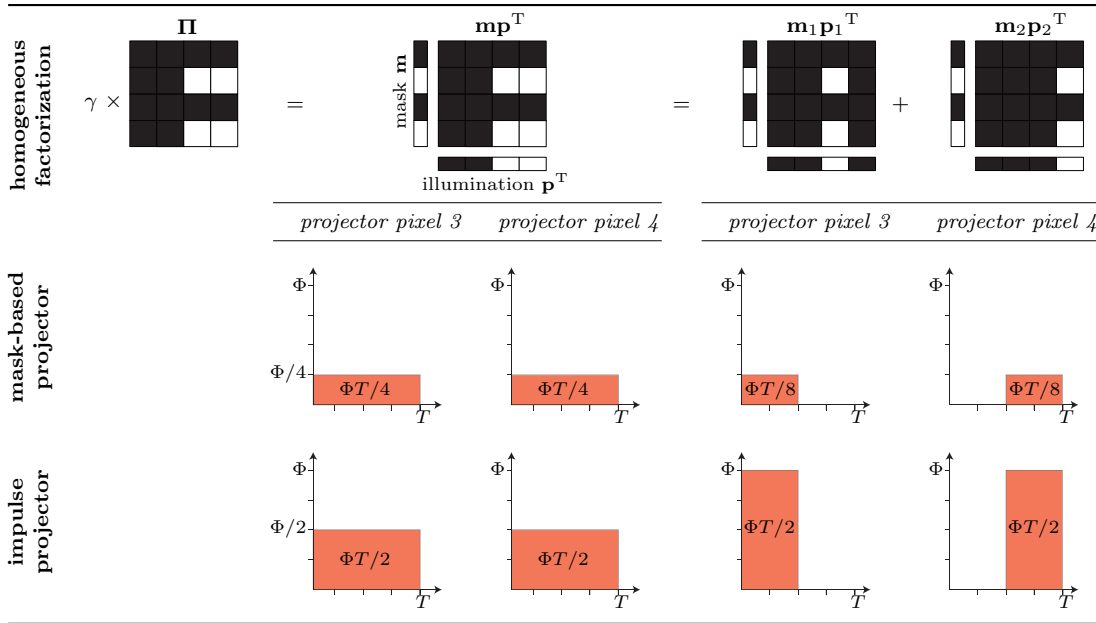


Figure 5.2: **Top row:** Realizing a rank-1 probing matrix of size 4×4 (left) in two ways—with a single mask/projection pair (middle) and with a sequence of two illumination patterns and masks (right). **Middle row:** Plots of power versus time for the “active” pixels in a mask-based projector. The projector’s redistribution ratio is $\varrho = 4$ for this 4-pixel projector because any given projector pixel can emit at most one quarter of the total available energy. For the single mask/projection pair case, projector pixels 3 and 4 are active throughout exposure period T , each generating $\Phi T/4$ J of radiant energy. For the two mask/projection pair case, projector pixels 3 and 4 are only active for half the exposure period, producing $\Phi T/8$ J of radiant energy each. Although the single mask/projection pair is more efficient than the length-two sequence ($\gamma = \Phi T/4$ versus $\gamma = \Phi T/8$) neither makes full use of the source’s energy. **Bottom row:** With an impulse projector on the other hand ($\varrho = 1$), both cases are optimal ($\gamma = \Phi T/2$); this is because the total radiant energy emitted by the impulse projector is exactly ΦT J, which is the total energy generated by its source over period T .

5.1.2 Optical Probing

The elements of the probing matrix Π specify energy attenuation from a specific projector pixel to a specific sensor pixel. Just like masks, probing matrices contain unitless scalars in the range $[0, 1]$.

Energy efficiency of mask/projection sequences Probing matrices describe attenuation in relative terms and do not tell us how much energy can actually be transmitted from the projector to the sensor. Indeed, the energy efficiency of optical probing depends on the precise sequence of projection patterns and masks used.

More specifically, we can realize any rank-1 probing matrix Π using just one projection pattern and one sensor mask (Figure 5.2):

$$\gamma \Pi = \mathbf{m} \mathbf{p}^T \quad (5.5)$$

where vectors \mathbf{p} and \mathbf{m} depend on matrix Π and satisfy all physical constraints, and γ is a scalar that is measured in joules. Intuitively, γ converts the unitless elements of Π into actual energies that

can be potentially transmitted during the exposure time. As mentioned in Footnote 1, we previously assumed that $\gamma = 1$. We call γ the *energy efficiency* of the mask/projection pair \mathbf{p} and \mathbf{m} .

Higher-rank probing matrices require changing projection patterns and masks $K > 1$ times during the exposure time. Mathematically this is equivalent to expressing matrix $\mathbf{\Pi}$ as a sum of outer products, subject to the physical constraints on illumination patterns and masks:

$$\gamma \mathbf{\Pi} = \sum_{k=1}^K \mathbf{m}_k (\mathbf{p}_k)^T \quad (5.6)$$

$$0 \leq t_k, \quad \sum_{k=1}^K t_k \leq T \quad (5.7)$$

$$0 \leq \mathbf{m}_k, \quad \|\mathbf{m}_k\|_\infty \leq 1, \quad 0 \leq \mathbf{p}_k, \quad \|\mathbf{p}_k\|_{\dagger\varrho} \leq t_k \quad (5.8)$$

where the mask \mathbf{m}_k and pattern \mathbf{p}_k are active for a time interval t_k , and scalar γ is the energy efficiency of the whole sequence. The sum of the time intervals t_k is the time necessarily to cycle through the entire mask/projection sequence, which cannot be longer than the total available exposure time T .

Note that the scalar γ is also inherently constrained from above by Eqs. (5.7) and (5.8):

$$|\gamma| \|\mathbf{\Pi}\|_{\max} \leq \sum_{k=1}^K \|\mathbf{m}_k (\mathbf{p}_k)^T\|_{\max} \quad (5.9)$$

$$\leq \sum_{k=1}^K \|\mathbf{m}_k\|_\infty \|\mathbf{p}_k\|_\infty \quad (5.10)$$

$$\leq \sum_{k=1}^K \|\mathbf{m}_k\|_\infty \frac{\Phi}{\varrho} \|\mathbf{p}_k\|_{\dagger\varrho} \quad (5.11)$$

$$\leq \frac{\Phi}{\varrho} \sum_{k=1}^K t_k \quad (5.12)$$

$$\leq \frac{\Phi T}{\varrho} \quad (5.13)$$

where the matrix norm $\|\cdot\|_{\max}$ is the maximum absolute element of a matrix. Note that this inequality is not a tight upper bound on the value of energy efficiency γ ; the inequality simply states that the radiant energy transmitted from any projector pixel n to camera pixel m , represented by $\gamma \mathbf{\Pi}[m, n]$, cannot exceed the total energy ΦT for impulse projectors and the distributed energy $\frac{\Phi T}{P}$ for mask-based projectors.

5.1.3 Homogeneous Factorization

The greater the energy efficiency of a sequence, the more total energy will flow from the projector to the camera in any given scene. We therefore seek mask/projection sequences that maximize γ .

Re-writing Eq. (5.6) in the form of a matrix factorization equation

$$\gamma \mathbf{\Pi} = \underbrace{[\mathbf{m}_1 \mathbf{m}_2 \cdots \mathbf{m}_K]}_{\text{masks } \mathbf{M}} \underbrace{[\mathbf{p}_1 \mathbf{p}_2 \cdots \mathbf{p}_K]}_{\text{illuminations } \mathbf{P}}^T \quad (5.14)$$

leads to a homogeneous factorization problem whose goal is to find the masks \mathbf{M} , patterns \mathbf{P} , and timeslices t_k that maximize energy efficiency:

$$\max_{\gamma, \mathbf{M}, \mathbf{P}, t_1, \dots, t_K} \gamma \quad (5.15)$$

$$\text{subject to} \quad \gamma \mathbf{\Pi} = \mathbf{M}\mathbf{P}^T \quad (5.16)$$

$$0 \leq \mathbf{m}_k, \quad \|\mathbf{m}_k\|_\infty \leq 1$$

$$0 \leq \mathbf{p}_k, \quad \|\mathbf{p}_k\|_{\dagger\ell} \leq t_k$$

$$0 \leq t_k, \quad \sum_{k=1}^K t_k \leq T$$

The energy efficiency of a particular factorization—and thus the solution to this optimization problem—depends to a large degree on a projector’s ability to redistribute light, *i.e.*, on its redistribution ratio. See Figure 5.2 for a detailed illustration.

5.2 Homogeneous Low-Rank Factorization

While the optimization in Eq. (5.15) is hard to solve directly, it can be relaxed into the following:

$$\begin{aligned} \min_{\mathbf{M}, \mathbf{P}} \quad & \|\mathbf{\Pi} - \mathbf{M}\mathbf{P}^T\|_F^2 + \lambda \sum_{k=1}^K \|\mathbf{m}_k\|_\infty \|\mathbf{p}_k\|_{\dagger\ell} \\ \text{subject to} \quad & 0 \leq \mathbf{m}_k, 0 \leq \mathbf{p}_k \end{aligned} \quad (5.17)$$

where λ is a regularization parameter that balances energy efficiency and the reproduction of $\mathbf{\Pi}$. Details of this relaxation, which absorbs timeslices and inverse energy efficiency γ^{-1} into the second term of Eq. (5.17), are in Appendix B.1.5. Its main intuition is to convert the homogeneous maximization of Eq. (5.15) into an inhomogeneous one by seeking the masks and projection patterns with the *minimum* sum of norm products. Once this decomposition is found, we can “brighten” matrix $\mathbf{\Pi}$, and thus find γ , by re-scaling \mathbf{M} and \mathbf{P} to saturate their upper-bound constraints; specifically:

$$\gamma = T \left(\sum_{k=1}^K \|\mathbf{m}_k\|_\infty \|\mathbf{p}_k\|_{\dagger\ell} \right)^{-1} \quad (5.18)$$

Note that, according to Eq. (5.18) and the definition of norm $\|\cdot\|_{\dagger\ell}$, the energy efficiency term γ grows linearly with both exposure time T and light source power Φ .

The relaxation in Eq. (5.17) has received some attention recently in computer vision and machine learning [6, 47]. Here we focus on aspects of that work that are specific to our optimization problem.

The projective tensor norm By leaving the sequence length K unconstrained and dropping non-negativity constraints, Eq. (5.17) becomes equivalent to a minimization of the following [47]:

$$\min_{\mathbf{X}} \quad \|\mathbf{\Pi} - \mathbf{X}\|_F^2 + \lambda h(\mathbf{X}) \quad (5.19)$$

where function $h(\mathbf{X})$ is the *projective tensor norm*, defined as

$$h(\mathbf{X}) = \min_{\mathbf{X}=\mathbf{M}\mathbf{P}^T} \left\{ \sum_{k=1}^K \|\mathbf{m}_k\|_p \|\mathbf{p}_k\|_q \right\} \quad (5.20)$$

with $p = \infty$ and $q = \dagger\varrho$ according to Eq. (5.17).

5.2.1 Implications of the Theory

Impulse illumination is globally optimal A closed-form solution to Eq. (5.17) exists for the special case of impulse projectors. This is because the norms $\|\mathbf{p}_k\|_1$ and $\|\mathbf{p}_k\|_{\dagger\varrho}$ in that equation always coincide when the projector’s redistribution ratio is equal to one. In particular, Bach et al. [6] showed that for $p = \infty$ and $q = 1$, the projective tensor norm—which is equal to the global minimum of Eq. (5.17)—is given by the sum of the ℓ_∞ norms of the mask vectors. The factorization corresponding to this minimum is

$$\mathbf{P} = \mathbf{I}, \quad \mathbf{M} = \mathbf{\Pi} \mathbf{P} \quad (5.21)$$

where \mathbf{I} is the $P \times P$ identity matrix, *i.e.* the optimal projection patterns \mathbf{P} are *impulses*.

Epipolar illumination is globally optimal for epipolar-only and non-epipolar imaging

The global optimality of impulse projectors has another, more practical implication: scanning-based laser projectors that can concentrate all their energy onto individual scanlines are also globally optimal for probing—but only when their scanlines are along the *epipolar lines*, and only for probing matrices whose elements do not vary within an epipolar block (*e.g.*, the first three matrices in Figure 5.3). In that case, each epipolar line can be treated as a “pixel” without loss of generality, and each epipolar block of $\mathbf{\Pi}$ can be treated as a single element. The globally-optimal solution is again given by Eq. (5.21), with “impulses” corresponding to a single epipolar line turned on.

Epipolar illumination and epipolar masking confer robustness to ambient light

A second practical implication of Eq. (5.21) concerns probing matrices $\mathbf{\Pi}$ that are permutations of the identity matrix. The globally-optimal solution is to unmask each camera pixel exactly once. This solution exposes camera pixels for the shortest possible time interval among all possible mask/projection sequences. In other words, *impulse illumination and impulse masking* simultaneously maximize the energy efficiency of probing and minimize the impact of ambient light. Similarly, *epipolar illumination and epipolar masking* are optimal for epipolar probing in the presence of ambient light—a principle we directly exploit with our laser-based prototype (Section 5.4).

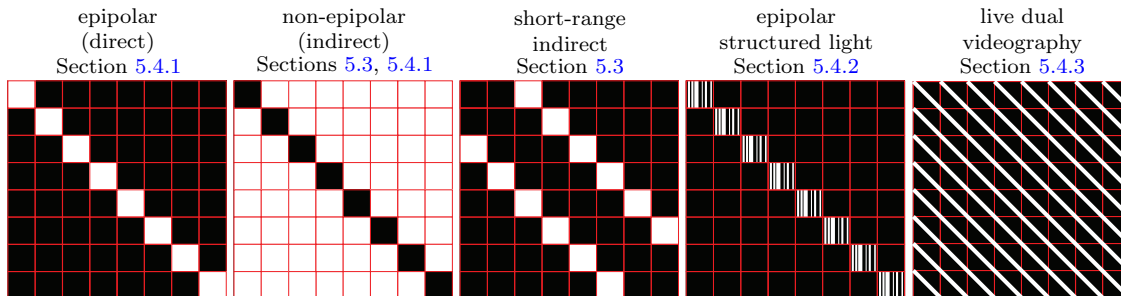


Figure 5.3: The five basic probing matrices used in Chapter 5. We follow the conventions of the stereo light transport matrix (Figure 4.2) for a projector and a sensor that have just eight epipolar lines worth of pixels (but many pixels along each epipolar line). The epipolar and non-epipolar matrices are identical to those in Figure 4.6. Individual elements range from zero (black) to one (white). Red lines indicate row boundaries in the raster-scan ordering of individual pixels. These lines partition $\mathbf{\Pi}$ into contiguous epipolar blocks (Figure 4.3), with each block controlling energy transmission from projector pixels on a specific epipolar line to sensor pixels on another epipolar line.

Code optimization algorithm for non-impulse projectors No closed-form solution is known for the projective tensor norm when ϱ is greater than one. We use the *structured low-rank matrix factorization* algorithm of Haeffele et al. [47] to locally optimize this norm with convergence guarantees. The base requirement of their algorithm is the ability to evaluate the *proximal operator*. For a norm $\|\cdot\|_p$ this operator is defined as

$$\mathbf{prox}_{\lambda\|\cdot\|_p}(\mathbf{v}) = \operatorname{argmin}_{\mathbf{x}} \left(\|\mathbf{x}\|_p + \frac{1}{2\lambda} \|\mathbf{x} - \mathbf{v}\|_2^2 \right) \quad (5.22)$$

The proximal operator for $\|\cdot\|_{\dagger\varrho}$ has a simple form when $\varrho = P$ and a very efficient algorithm exists for computing it [96]. We use this approach to compute optimized code sequences for our DMD-based prototype (Section 5.3). For projectors with redistribution ratios between 2 and $P - 1$, we found no simple solution. We evaluate Eq. (5.22) explicitly in such cases, by solving a constrained minimization problem. To impose non-negativity constraints on masks and projection patterns, we replace each negative component of the proximal operator’s output with zero [96, 140].

5.3 DMDs for Live Energy-Efficient Imaging

DMDs offer flexible masking but have two important limitations: they are very inefficient when used for light projection ($\varrho = P$) and they can only support short code sequences for live imaging (*e.g.*, $K \leq 96$ for the Texas Instruments LightCrafter DMD). This makes the design of energy-efficient codes all the more important, so that images are as bright and artifact-free as possible.

Optimized codes for epipolar imaging Probing with the epipolar-only matrix $\mathbf{\Pi}$ (Figure 5.3) yields an image that is almost exclusively due to direct surface reflections.

Figure 5.5 shows the projector patterns generated by Eq. (5.17). These codes are far from random. First, the matrix $\mathbf{M}\mathbf{P}^T$ converges to the probing matrix as K approaches $\mathbf{\Pi}$ ’s size. This

has a clear, positive impact on image quality (Figure 5.4). Second, when $K = P$ is divisible by four, the square matrix \mathbf{P} is exactly the Hadamard matrix [50], *i.e.*, its elements have value $+1$ or -1 and its inverse is $\frac{1}{P}\mathbf{P}^T$. Third, for other sizes, the codes we generate exhibit similar characteristics to the Hadamard matrix (Figure 5.5). Fourth, we observe that energy efficiency can be improved even further by computing code sequences that are *longer* than $\mathbf{\Pi}$'s size.

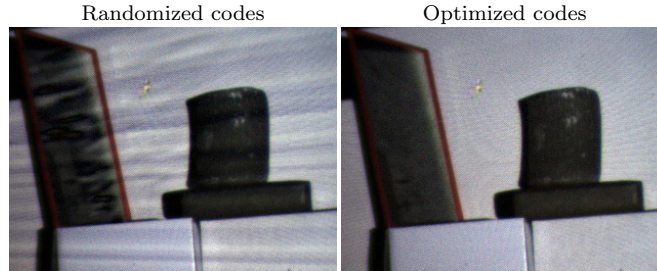


Figure 5.4: Epipolar imaging with the high-speed DMD-based prototype. The scene has a mirror on the left and a translucent candle on top of a box on the right, all in front of a white wall (see Figure 5.9(a) for a conventional photo). We generated randomized and optimized codes for $K = 512$. Both codes are equally energy efficient since they produce images of similar brightness. Our optimized codes, however, are not just efficient, they approximate $\mathbf{\Pi}$ well too. This yields images with far fewer artifacts (note the bright streaks on the left image).

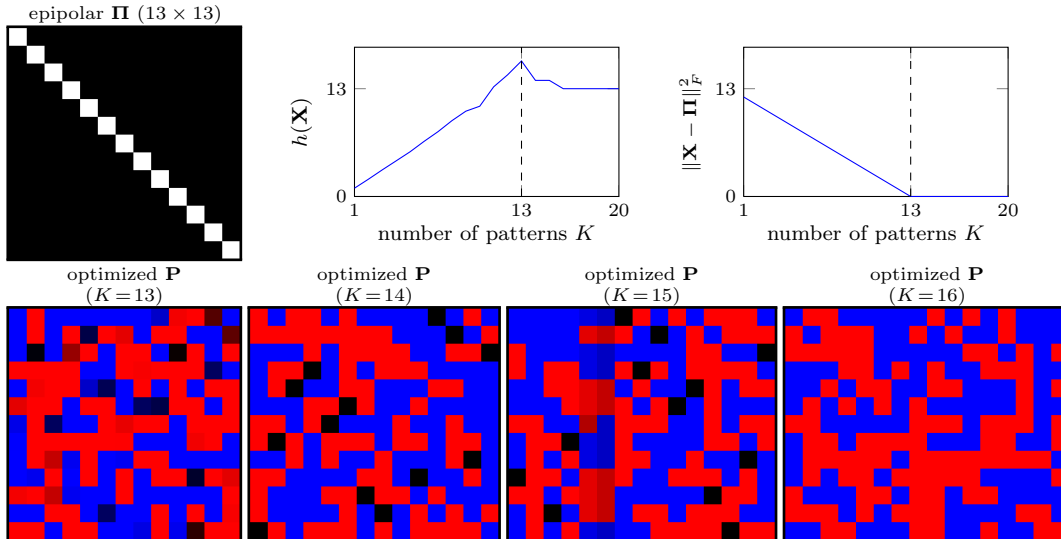


Figure 5.5: Codes optimized for different sequence lengths K , where $\varrho = 13$, $\Phi = 1$, and $T = 1$. Here we dropped the non-negativity constraint and computed codes \mathbf{M} and \mathbf{P} that require two photos to implement efficiently on a DMD-based projector (*i.e.*, one for the positive and one for the negative elements of \mathbf{P}). **Top-Left:** A probing matrix. **Top-Middle:** Sequence length versus energy efficiency as measured by the projective tensor norm (lower is better, as per Eq. (5.18)). **Top-Right:** Sequence length versus reconstruction error of matrix $\mathbf{\Pi}$. Note that energy efficiency starts improving once $\mathbf{\Pi}$ can be reconstructed exactly, *i.e.*, for $K \geq 13$. **Bottom Row:** Optimized projector patterns for various sequence lengths (mask matrices omitted for brevity). Positive and negative elements are color-coded in red and blue, respectively.

Optimized codes for non-epipolar imaging Figure 5.6 illustrates the visual quality improvement from our optimized codes in this task. The task requires probing with the non-epipolar matrix of Figure 5.3 which guarantees that no direct light reaches the sensor. Short mask/projection sequences from the randomized algorithm result in strong visual artifacts (bright diagonal streaks along epipolar lines). These occur because the energy that is transmitted from projector to sensor is attenuated non-uniformly over the sensor plane (*i.e.*, \mathbf{MP}^T is a poor approximation of $\mathbf{\Pi}$). Our optimized codes, on the other hand, yield bright and artifact-free images for a short sequence of the same length. These codes are shown in Figure 5.7.

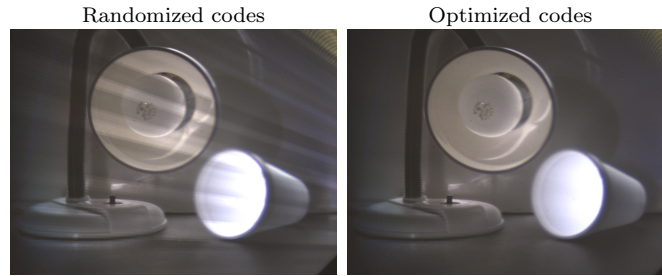


Figure 5.6: Non-epipolar imaging using the optimized codes versus the randomized codes. The scene consists of a styrofoam cup and a lamp (turned off). We generated a short sequence of codes ($K = 32$) using the two algorithms, and captured the images shown above. Again, both code sequences have similar energy efficiency but our optimized codes produce no discernible artifacts. The bright diagonal streaks in the left image is the result of poorly approximating matrix $\mathbf{\Pi}$.

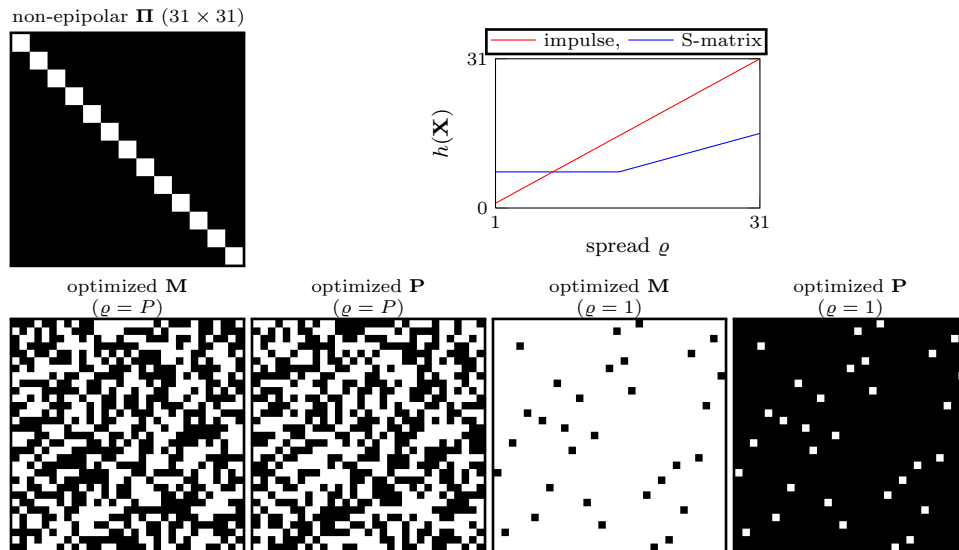


Figure 5.7: Optimized codes for different values of the redistribution ratio ρ , and where $\Phi = 1$ and $T = 1$. For a DMD-based projector ($\rho = 31$), our algorithm returns masks \mathbf{M} and illuminations \mathbf{P} that are complementary S-matrices [50, 111]. For the impulse projector ($\rho = 1$) we get impulses. The projective tensor norm plot of the S-matrix and impulse solutions is shown on the right for different values of ρ (lower is better).

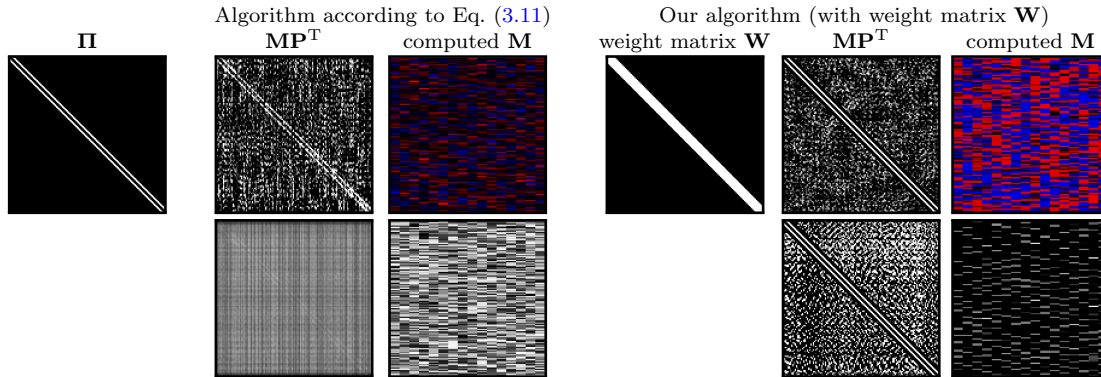


Figure 5.8: Computing a $K = 16$ code sequence for a rank-128 short-range indirect probing matrix. The top row shows the mask matrix computed without enforcing non-negativity whereas the bottom row shows the matrix with those constraints enforced. We can use this matrix to capture images with only “short-range indirect” contributions, *i.e.*, light that contains neither direct reflections nor contributions transported between distant epipolar planes. We are aware of no techniques to capture such images, either by factorizing $\mathbf{\Pi}$ or by other means. In particular, the randomized algorithm fails to produce usable codes because $\mathbf{\Pi}$ ’s off-diagonals are reconstructed poorly. By incorporating the weight matrix \mathbf{W} shown into the Frobenius norm of Eq. (5.17), our algorithm produces a low-rank approximation of $\mathbf{\Pi}$ whose near off-diagonals are reconstructed very well. Use of these weights was partly inspired by Lanman et al. [73]. Figures 5.9(c) shows images captured with our DMD-based prototype using this code.

Optimized codes for high-rank probing Many matrices useful for probing have rank much higher than the maximum attainable value of K . In such cases, the matrix $\mathbf{\Pi}$ can only be approximated by a rank- K approximation of $\mathbf{\Pi}$. Figure 5.8 compares our codes to those obtained by the randomized approach for the hitherto not studied task of *short-range indirect imaging*. This task involves probing with the corresponding matrix from Figure 5.3, which blocks all energy transmissions except those occurring between any two nearby epipolar planes.

For such high-rank cases, we modify the objective function in Eq. (5.17) by replacing the Frobenius norm, $\|\cdot\|_F$, with a weighted Frobenius norm, $\|\cdot\|_{\mathbf{W}}$. The role of the weight matrix, \mathbf{W} , is to effectively reduce the rank of the probing matrix $\mathbf{\Pi}$ by targeting a subset of elements that we want to reconstruct accurately. For example, in Figure 5.8, the weight matrix is chosen such that the elements near the diagonal are reconstructed perfectly. If we assume the unknown light transport matrix is also band-limited (*i.e.* its nonzero elements are located near the main diagonal), this low-rank probing operation $\mathbf{M}\mathbf{P}^T$ produces the same result as the high-rank probing matrix $\mathbf{\Pi}$.

Figure 5.9 shows captured images for several high-rank probing tasks. These tasks were impossible to perform previously because no efficient low-rank factorizations of $\mathbf{\Pi}$ were known (*i.e.*, with enough energy efficiency for a useful image signal).

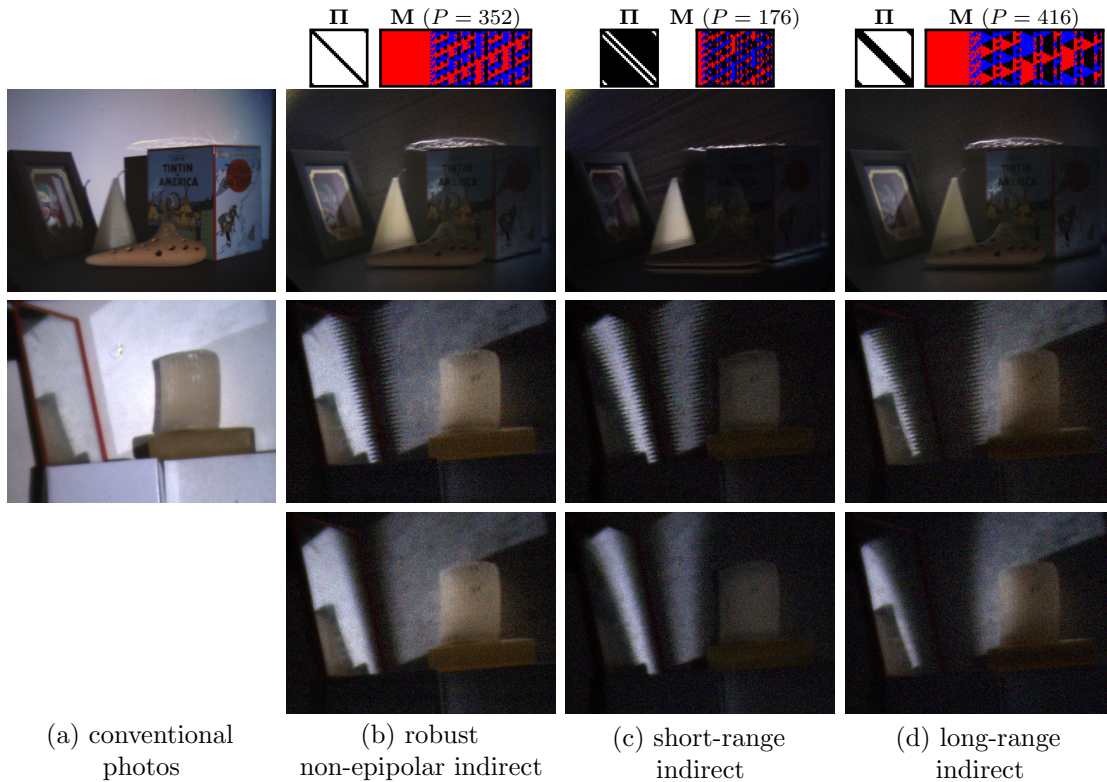


Figure 5.9: New transport-probing abilities enabled by homogeneous factorization of high-rank matrices. Probing matrix Π and mask matrix M (after dithering) are shown above each column. **(a)** We show results for two scenes: (Top) A scene containing, from left to right, a Chinese mask within a display case; a translucent candle; an ocarina; and a tin box. (Middle and bottom) A scene with a mirror and a translucent candle. **(b)** Making non-epipolar probing robust to geometric misalignments by building robustness into matrix Π itself. Here we set the elements of Π to zero within a small band of the diagonal to ensure no direct light “leaks” into the photo, and compute a code sequence of length $K = 352$. **(c)** Short-range indirect imaging captures inter-reflections, sub-surface scattering, and a bright vertical band of mirror reflections. We used $K = 176$. **(d)** Long-range indirect photos captured using a code sequence of length $K = 416$. Since sub-surface scattering is a short- to mid-range transport phenomenon, candles appear darker compared to (c). In the bottom row we used $4\times$ longer sequences to improve image quality. These imaging tasks are considerably “harder” than epipolar-only and non-epipolar probing, requiring less energy-efficient codes: exposure time was $T = 1/3$ s for all images, about $10\times$ longer than in Figures 5.4 and 5.6.

Implementation details The projective tensor norm is a prior that naturally favors binary codes. Nevertheless, the codes computed by our algorithm are not binary in general. We implement non-binary codes by temporally dithering each code with the DMD, *i.e.*, projecting a sequence of binary codes that achieve the desired intensity value. This increases the number of codes but does not affect energy efficiency.

Ignoring non-negativity constraints in Eq. (5.17) can significantly increase the energy efficiency of code sequences. Codes with both positive and negative values, however, are not physically realizable. We convert a length- K mask/projection sequence into two sequences of length $2K$ with Eq. (3.12); this also decreases energy efficiency by a factor of 4.

Energy-efficient codes for redistributive projectors Finding globally-optimal codes for general redistributive projectors remains an open problem. Nevertheless, we can draw interesting conclusions about the characteristics of (locally-optimized) energy-efficient codes for different redistribution ratios and different probing matrices. Take, for instance, the case of a non-epipolar probing matrix $\mathbf{\Pi}$ of size $P \times P$ where $P + 1$ is divisible by four. For $\varrho = P$, our algorithm automatically returns illuminations \mathbf{P} similar to an S-matrix [50, 111] (Figure 5.7). This matrix has several characteristics: (1) it is binary, where each element has value 0 or 1, (2) each row or column has exactly $\frac{P+1}{2}$ elements with value 1 and $\frac{P-1}{2}$ elements with value 0, and (3) its inverse is $\frac{2}{P+1}(2\mathbf{P}^T - 1)$. At the other extreme of $\varrho = 1$ it returns the impulse illuminations predicted by Eq. (5.21). Between these two extremes, we found no solution that improves upon either the S-matrix or the impulse illuminations. This suggests a rapid “phase transition” in the energy efficiency of these two code sequences; this rapid transition occurs at the intersection point of the two plots in Figure 5.7.¹

5.4 Lasers for Live Energy-Efficient Imaging

Scanning-based laser projectors operate on a different principle than mask-based projectors. These MEMS-based projectors steer a laser beam across the scene, sending all their energy to just one pixel at a time. We consider these projectors at the more granular timescale of a single *scanline*, where they act as impulse projectors whose “impulses” are scanlines (Section 5.2.1). By aligning the projector’s scanlines with the rows of a rolling-shutter camera—whose ability to mask individual rows electronically provides a sequence of “impulse” masks—we obtain a system with very high energy efficiency for many transport probing tasks.

Our prototype is portable and consists of nothing more than an off-the-shelf camera and off-the-shelf projector whose only modifications are synchronization electronics—no passive or active optical components are required beyond a simple color filter. This avoids the need for sophisticated hardware devices and extra optics (*e.g.* relay lenses, diffraction grating), all of which introduce aberrations, limit working volume, reduce light throughput, and adversely impact size, weight and cost.

Laser projectors for “impulse” illumination We use a Microvision ShowWX+ pico-projector (resolution: 848×480 , frame rate: 60 Hz, light power: 5 lm per color channel) for light projection. Fortunately, even though the projector’s raster-scan path is fixed in hardware, it implements a globally-optimal sequence of illumination codes for many of the probing tasks we consider (Section 5.2.1).

Rolling-shutter cameras for “impulse” masking For outdoor experiments we use an IDS Imaging UI-3250CP-M-GL (1600×1200) monochrome CMOS rolling-shutter camera, fitted with a 532 nm laser line cut filter (centered on the projector’s green laser wavelength) with a passband

¹One can draw an analogy to the work of Cossairt et al. [16] who observed similar behavior in their study of computational imaging in read-noise versus shot-noise limited settings.

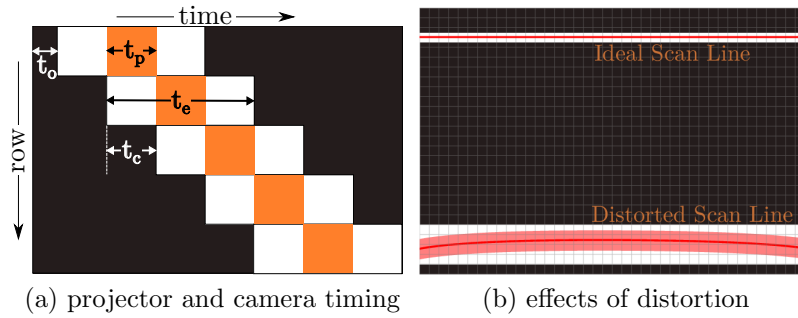


Figure 5.10: **(a)** How to combine masking with a rolling shutter camera and raster scanning with a laser projector. At each timestep (of duration t_p), the projector illuminates a single scanline (orange). The camera’s rolling shutter exposes one or more rows of pixels (white). The masks are defined by three controllable parameters: the exposure time t_e ; the time it takes the rolling shutter to read a row of pixels t_c ; and the offset between the projector sync output and the camera trigger input t_o . **(b)** Ideally, a single projector scanline corresponds to a single camera row (top). In practice, because of distortion and jitter, each scanline corresponds to a band of sensor pixels (bottom).

of 10 nm. For indoor experiments we use an IDS Imaging UI-3240CP-C-HQ (1280×1024) color CMOS camera. In both cases, we use the cameras’ rolling shutter to implement sensor masking.

At any point in time, the rolling shutter can expose either a single row or a band of rows. The rolling shutter is triggered by the VSYNC signal generated by the projector. Figure 5.10 shows our cameras’ timing diagram. We use t_p to denote the timeslice for which the projector dwells on a single scanline. The speed at which the rolling shutter progresses down the rows of the image (t_c) is determined by the pixel clock frequency; we choose this frequency and the focal length of camera lens so that the downward velocity of exposed camera rows matches the downward velocity of the projector’s scanline. Increasing the row exposure (t_e) increases the thickness of the band of camera rows exposed for each projector scanline. Changing the delay (t_o) between the VSYNC signal from the projector and the trigger signal passed to the camera changes the offset between the illuminated row on the projector and the imaged row(s) on the camera.

5.4.1 Epipolar and Non-Epipolar Imaging

To probe with an epipolar matrix $\mathbf{\Pi}$ (Figure 5.3), we position the camera and projector so that they emulate a rectified stereo configuration (Figure 5.11). This is key, because the epipolar lines for a rectified system align with the rows of the projector and camera sensor. Low-distortion camera lenses are also critical for this alignment to minimize distortion; we use the Lensagon CVM0411ND varifocal lens in all our experiments.

To capture live epipolar video, we project a plain white pattern continuously and choose sensor masks so that for each projector scanline, only the corresponding row of sensor pixels is exposed. This has the effect of transmitting energy from a projector row to a sensor row on the same epipolar plane, while blocking everything else. We do this by matching the camera’s row exposure time (t_e) and row readout time (t_c) to the time the projector dwells on a single scanline (t_p).

By changing the timing parameters on the camera so that the masks are inverted, the same setup

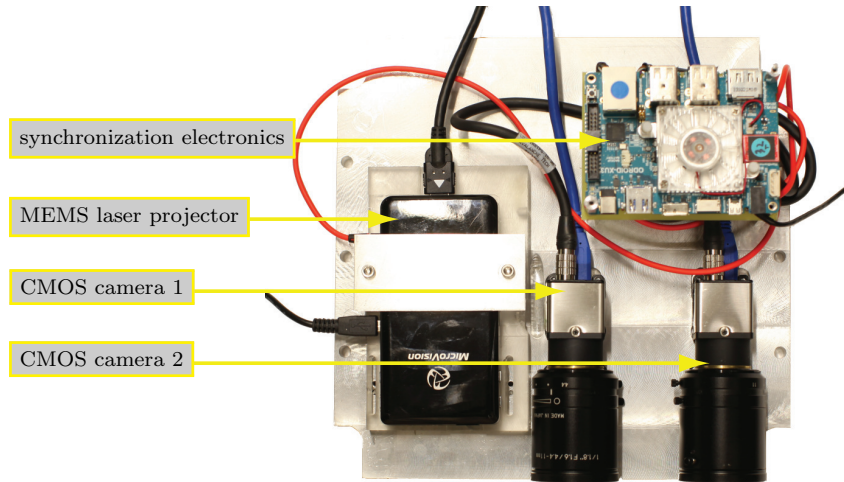


Figure 5.11: Photo of our laser-based prototype, consisting of an off-the-shelf laser projector and two rolling-shutter CMOS cameras, each fitted with an ordinary lens and a red band-pass filter. A microcontroller synchronizes the rolling-shutter of the cameras with the MEMS of the projector. We only use one camera to generate the energy-efficient imaging results in this work; the second (optional) camera can be used to implement other active illumination techniques that require two cameras (*e.g.* spacetime stereo [20]).

can be used for probing with non-epipolar $\mathbf{\Pi}$ (Figure 5.3). This is done by setting t_e to be t_p less than the projector cycle time and adding t_p to camera trigger offset t_o . Thus, at any point in time, every sensor row is exposed except the one corresponding to the current projector scanline. This blocks all energy transmission between projector rows and sensor rows on the same epipolar plane while leaving all other light unblocked.

Ideally, we would be able to configure the rolling shutter so that only the rows of sensor pixels illuminated by the projector at any timestep would be exposed (Figure 5.10). In practice, the projector we use generates distorted scanlines that are not absolutely straight. Additionally, we observe synchronization jitter and small perturbations in the trajectory of the projector’s laser during each exposure cycle. This means that the region in the camera image corresponding to a projector scanline is constrained to lie inside a narrow band in the image, not along a single row of pixels. To accommodate these bands, during epipolar imaging we thicken the region of unblocked pixels in each mask by increasing the pixel exposure time t_e and adjusting the trigger offset t_o . As a result, some short-range indirect light may leak into the epipolar image.

Figure 5.12 shows frames from live epipolar and non-epipolar video captured with our prototype, for scenes with significant indirect light transport. This prototype performs non-epipolar imaging at least four times more efficiently than with a conventional projector because DMD-based codes must block 75% of the energy from the projector’s light source. Moreover, it captures epipolar video at full video rates and in a single readout, whereas energy-efficient codes for DMD-based epipolar imaging require two readouts and a computational subtraction step (Section 5.3).

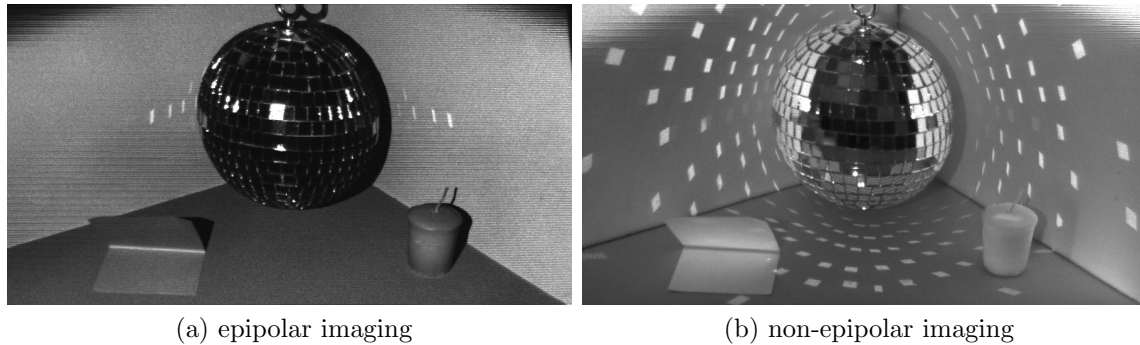


Figure 5.12: Separating components of illumination: we can capture live epipolar images in one shot with our laser-based prototype. Note that most inter-reflections appear in the non-epipolar component: reflections from the ball appear in the epipolar image only when they agree with the epipolar geometry. These images were captured by exposing epipolar stripes wider than a pixel; as a result, some short-range sub-surface scattering appears in the epipolar image instead of the non-epipolar image. Vignetting artifacts at the corners are due to projector distortion.

5.4.2 Epipolar Structured Light

Combining epipolar imaging with the projection of structured-light patterns is equivalent to probing with the epipolar structured light matrix $\mathbf{\Pi}$ in Figure 5.3. By probing with $\mathbf{\Pi}$ instead of projecting structured light patterns conventionally, we gain two benefits: robustness to indirect light transport and robustness to ambient illumination.

Robustness to indirect light transport As demonstrated throughout Chapter 4, indirect transport effects like inter-reflections and scattering can cause severe, systematic errors when using structured light to reconstruct the 3D shape of scenes with concavities and objects made of optically challenging materials. Unlike indirect invariant structured light imaging where photos are merely invariant to indirect light, epipolar structured light eliminates indirect light altogether. The reconstructions obtained in Figure 5.13 show how effective epipolar structured light is in the face of inter-reflections and sub-surface scattering.

Structured light in bright sunlight Active light sources are typically orders of magnitude weaker than daylight. Since conventional cameras have limited dynamic range, the signal from the source is overwhelmed by ambient illumination when regular imaging is used. This makes active illumination methods, such as structured light, hard to use outdoors. Even though in theory our prototype minimizes pixel exposure only for the task of epipolar probing (Section 5.2.1), in practice the exposure time of individual scanlines is sufficiently small that the energy received from ambient light is small compared to the energy transmitted from the projector—even under challenging illumination conditions and even when the projector light source is weak. Refer to Figures 5.13, 5.14, and 5.15 for some example reconstructions performed under challenging ambient illumination conditions. All reconstructions were computed using ten high-frequency gray code patterns.

It should be noted that our method is complementary to other methods for suppressing ambient

light like placing a narrow-band spectral filter on the camera that matches the output wavelength of the active light source. Also note that, unlike the ambient light suppression done in time-of-flight cameras (which is electronic and subject to shot noise), epipolar structured light blocks ambient illumination before it reaches the sensor.

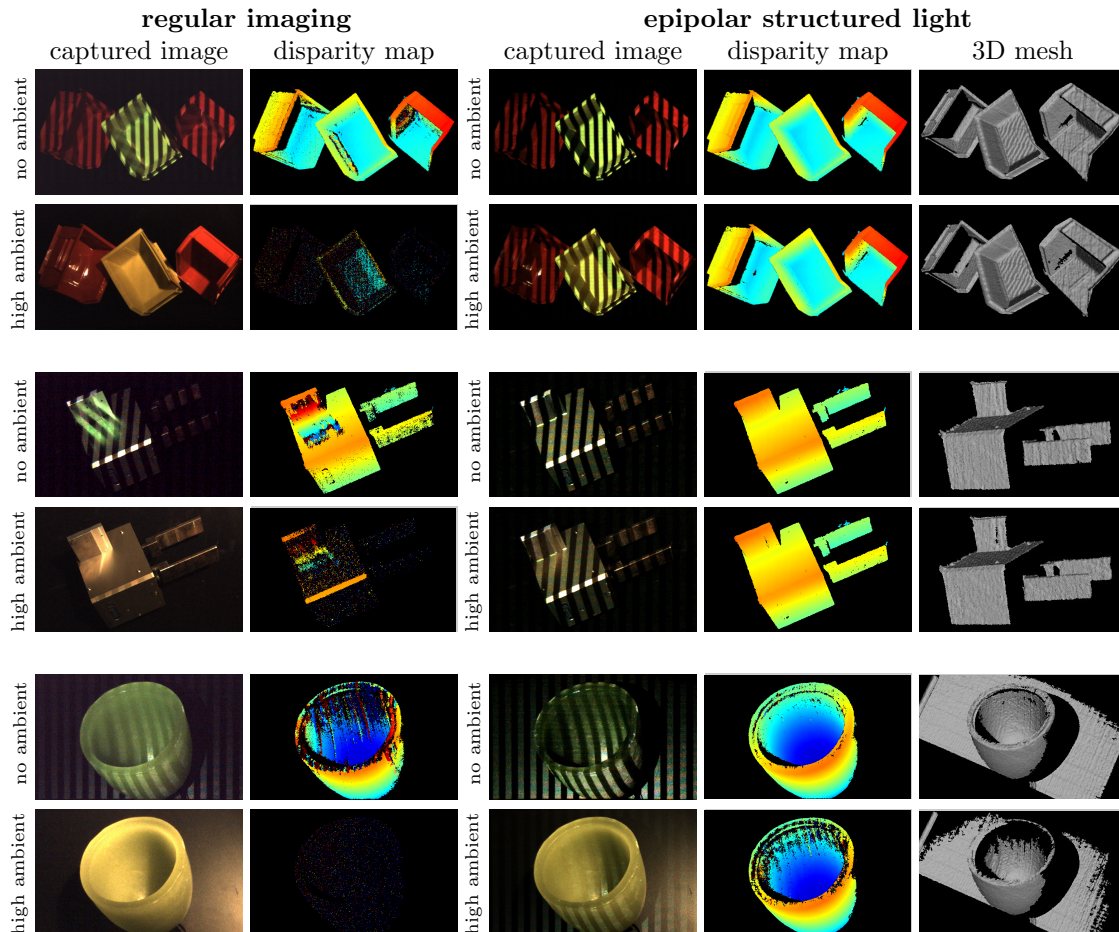


Figure 5.13: Imaging and reconstructing challenging objects with strong indirect illumination effects under ambient lighting. Structured light with regular imaging is prone to errors due to indirect light transport effects (notice the effects of inter-reflections on the plastic bins and industrial gripper and sub-surface scattering in the green wax bowl). Additionally, since active illumination sources are often orders of magnitude weaker than ambient light, structured light patterns are easily overwhelmed by ambient light. Epipolar structured light blocks a significant fraction of both. In these examples an ambient light level of $10,000 \text{ lm/m}^2$ causes structured light with regular imaging to break down. In contrast, epipolar structured light still works well. We used the same acquisition time for both methods in each scene.

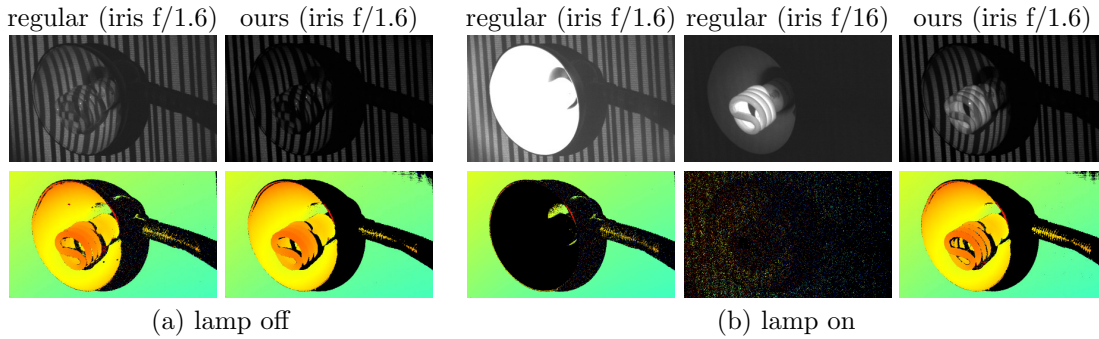


Figure 5.14: Imaging and scanning a 1600 lm lamp with a 5 lm projector: When the lamp is off (a), the pattern projected by the projector is visible with both regular imaging and our method. By projecting a series of structured light patterns the lamp can be reconstructed. When the bulb is turned on (b), regular imaging breaks down. With a large aperture the image is saturated, and with a small aperture the projected pattern is not visible to the camera on the bulb or the shade. Our rolling shutter based implementation exposes sensor pixels only while they can receive light from the projector and as a result, most of the light from the bulb is blocked, the pattern is clearly visible even on the bulb and the fixture can be reconstructed even though it is on.

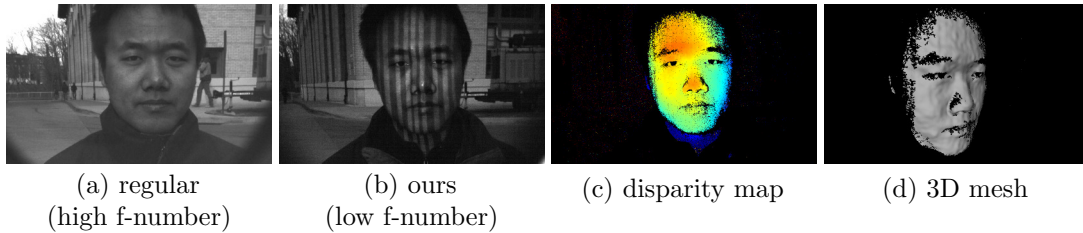


Figure 5.15: Active illumination with a 5 lm laser projector in bright sunlight ($80,000 \text{ lm/m}^2$): With regular imaging (a), the active illumination patterns are overwhelmed by sunlight and are not visible despite using a wavelength filter. Our rolling shutter based implementation blocks a large fraction of the ambient light. This allows the projected pattern to be seen by the camera (b) and makes 3D structured light reconstruction possible (c),(d).

5.4.3 Live Dual Videography

Dual photography [71, 114] (Section 1.4.3) makes it possible for a projector and a camera in general position to “exchange” their viewpoints; however, the technique by Sen et al. [114] first requires capturing the full transport matrix.

We demonstrate that while capturing the full transport matrix of a scene is still very hard, capturing an approximation of the *epipolar* image from the projector’s viewpoint is not only easy, it can be done in real time with no computational processing whatsoever (Figure 5.16).

Because the maximum exposure period of the rolling-shutter camera is $1/60$ s, we instead use our high-speed system from Figure 4.9 combined with the laser-based projector to extend the exposure period to $1/5$ s and improve SNR. The scanlines of the camera and projector are perpendicular to the epipolar planes, and we adjust timings so that the projector’s scanline and the camera’s “active” row are identical. This corresponds to the dual videography probing matrix in Figure 5.3. We then

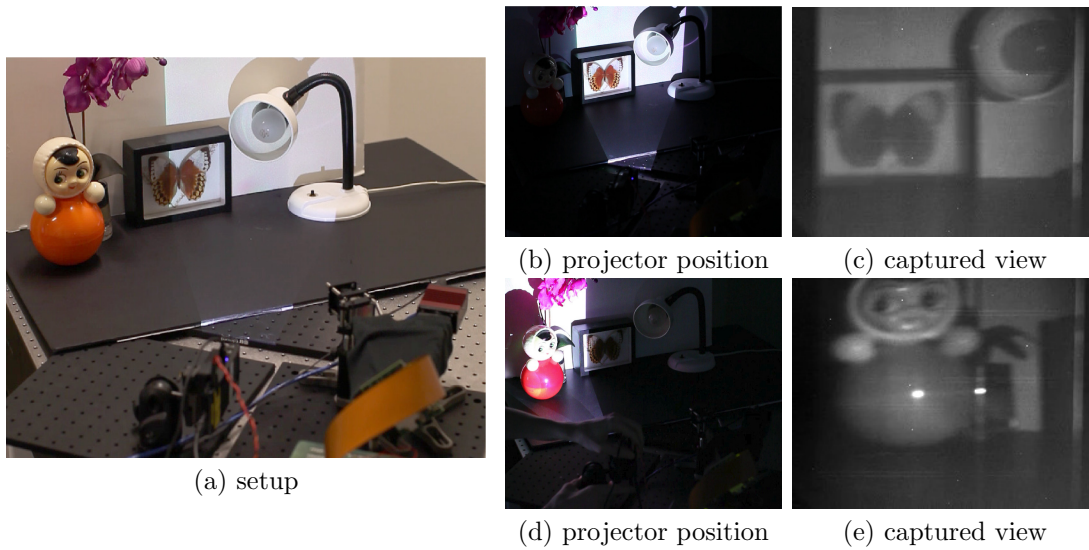


Figure 5.16: Live dual videography. We demonstrate how combining an impulse projector with a masked camera can be used to capture live video of a scene from the point of view of the projector. As the projector pans from right (b) to left (d), the apparent viewpoint of the video frame captured by the static camera changes from (c) to (e).

place a line diffuser in front of the camera, with its axis of diffusion oriented horizontally, *i.e.*, along the epipolar planes. This smears the image of every scene point along the epipolar planes. As the projector scanline and active row sweep across the scene, an image is formed on the sensor from the projector’s viewpoint. Key to this method’s success is again our prototype’s energy efficiency: even though the diffuser scatters light a great deal, enough of it reaches the camera to create a clear image that is relatively free of noise.

5.5 Summary and Contributions

We presented a novel framework for modeling and maximizing the energy efficiency of optical probing. We introduced the notion of *redistribution ratio* that models the degree to which a light source’s energy can be used for imaging rather than be wasted due to blocking/attenuation. In particular,

- we prove that optimal energy efficiency requires impulse illumination, where all the energy of the source can be concentrated in just one projector pixel that can be scanned rapidly;
- for projectors that cannot do this, we provide an algorithm that computes energy-efficient codes by solving a homogeneous matrix factorization problem through local optimization;
- this energy-efficient formulation enables new transport-probing abilities (*e.g.* short-range and long-range indirect imaging) with our DMD-based prototype;
- we combine a laser projector and a rolling-shutter camera to demonstrate a new energy-efficient probing prototype that makes near-optimal use of its light source energy; and

- this energy-efficient prototype enables several new imaging capabilities, which include epipolar and non-epipolar imaging with no wasted light (Figure 5.12), robust 3D scanning of objects made of shiny and translucent materials (Figure 5.13), structured-light imaging under strong ambient lighting (Figures 5.14 and 5.15), and performing dual photography at live video rates (Figure 5.16).

That said, much remains to be understood on the problem of maximizing energy efficiency—especially for projectors that cannot use the energy of their source very efficiently (no closed-form solution exists in that case). Last but not least, as dynamic vision sensors become more practical [80], and as redistributive laser systems and electronic masks become more flexible, it should be possible to build systems that push the energy efficiency frontier even further.

Chapter 6

Probing Transient Light Transport

Transient imaging has recently emerged as an alternative paradigm for light transport analysis that exploits light’s finite speed. Instead of illuminating a scene with spatially-varying patterns, they rely on temporally-varying ones using MHz to THz lasers and sensors sensitive to these rates (*e.g.*, streak cameras [129] and photonic mixer devices [53, 63]). These techniques have opened new frontiers in imaging—looking around corners [68], time-of-flight depth imaging [69, 128], lensless imaging [138], and capturing propagating optical wavefronts [129]—but are fundamentally limited in their ability to analyze complex transport.

In this chapter, we combine both paradigms by considering, for the first time, the generation and acquisition of transient space-time patterns for scene analysis. In this new analysis regime, the projector emits a 3D signal (2D space \times 1D time) and the camera receives a transformed 3D version of it. This brings two sets of constraints to bear on the same problem: constraints on the *spatial layout* of light paths arriving at a given pixel and constraints on their *travel time*. These constraints are complementary but *not* orthogonal; thus, by considering them jointly we can draw far stronger conclusions about light transport than when spatial or temporal light patterns are used in isolation, or sequentially.

To demonstrate the practical advantages of this approach, we apply it to the problem of probing specific light transport paths to analyze a scene’s transient appearance. We explored this thoroughly for conventional imaging, but it is poorly understood in transient settings. Here, we give a solution to this problem for imaging three transport components demonstrated in Figures 1.3 and 3.11: (1) the direct component, (2) the specular indirect component (*e.g.* caustics and specular reflections) and (3) the diffuse indirect component (*e.g.* sub-surface scattering and diffuse inter-reflections). We then implement this basic imaging ability in a functional prototype to

- improve the robustness of time-of-flight sensors against indirect transport by acquiring and processing only the direct time-of-flight component;
- capture sharp, evolving wavefronts of “light-in-flight” that so far have been directly observed only with very expensive streak camera technology; and
- conduct space-time light path analysis to separate “true” scene points from their mirror reflections.

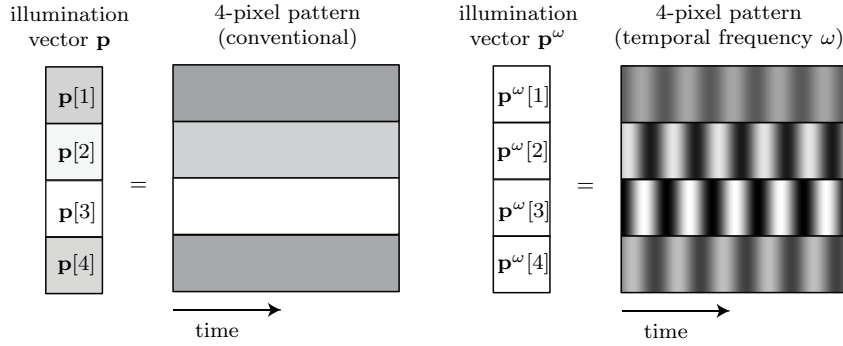


Figure 6.1: Spatial vs. spatio-temporal patterns. On the left, pixel intensities remain fixed over the exposure time of the illumination pattern. On the right, pixel intensities vary sinusoidally with a common temporal frequency ω but their amplitudes and phases differ. The amplitude a and phase offset ϕ of each sinusoid is represented by a complex value $\mathbf{p}^\omega[n] = ae^{-i\phi}$. Note that we assume these sinusoids oscillate between intensities a and $-a$.

Toward these goals, our key contribution is a new form of the light transport equation that makes many classical transport analysis problems easy to formulate and solve in the transient domain. This new *transient frequency light transport equation* assumes that we image the scene by projecting a pattern onto it—a spatio-temporal signal whose pixel intensities vary sinusoidally in time with a common frequency ω (Figure 6.1):

$$\mathbf{i}^\omega = \mathbf{T}^\omega \mathbf{p}^\omega, \quad (6.1)$$

where the pattern \mathbf{p}^ω is a column vector of P complex numbers, each representing the sinusoid’s amplitude and phase for a specific projector pixel; the image \mathbf{i}^ω represents the per-pixel sinusoids received at the camera; and \mathbf{T}^ω is the scene’s $I \times P$ *transient frequency transport matrix* for emission frequency ω . Intuitively, this matrix tells us how the temporal sinusoid received at a specific camera pixel is affected by temporal sinusoids emitted by different projector pixels, after accounting for all light transport paths and the delays they induce (Figure 6.2). In this sense, the transient frequency transport matrix describes transport in the exact same way the conventional matrix \mathbf{T} does, except that it deals with per-pixel temporal sinusoids instead of per-pixel intensities. This matrix is different for different emission frequencies and reduces to the conventional transport matrix for the DC frequency ($\omega = 0$), where the projection pattern does not vary over time. Taken together, this continuous family of discrete transport matrices is a five-dimensional structure that fully describes light transport at transient timescales—from a projector plane to a camera plane via a 1D set of emission frequencies.

The fact that Eq. (6.1) exists has two implications for light transport analysis. First and foremost, we can directly apply conventional light transport techniques to the transient case—without compromising their effectiveness or adding extra assumptions. This involves simply replacing conventional projectors and cameras with transient ones, and imaging the scene by emitting from every projector pixel a temporal sinusoid of a common frequency ω . We use this idea extensively in a new prototype system to acquire specific components of light transport through optical probing. These components cannot be captured robustly with existing transient imaging techniques because temporal signals may propagate through the scene along many different paths and combine upon arrival at

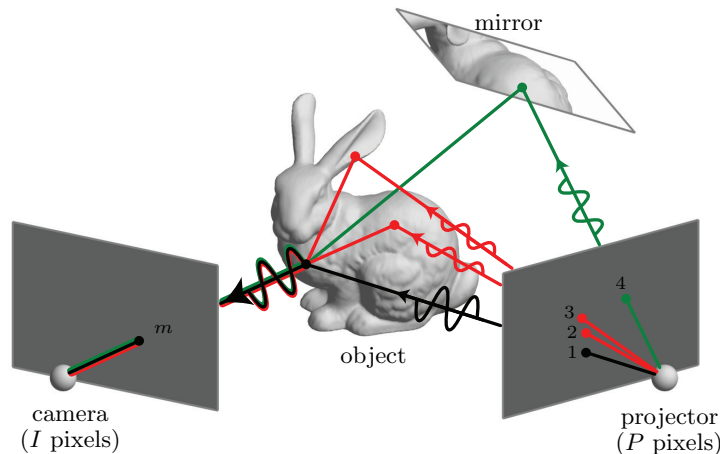


Figure 6.2: Visualizing the transient frequency transport equation. The projector emits sinusoids of frequency ω from pixels 1, 2, 3 and 4 and these signals propagate to camera pixel m along four paths: a direct path (black), where light bounces only once; two diffuse indirect paths (red) where light undergoes at least two diffuse bounces; and a specular indirect path (green) where light undergoes at most one diffuse bounce. Since all sinusoids have the same frequency, their superposition, recorded by camera pixel m , will have the same frequency as well. The amplitude and phase of the sinusoid received by this camera pixel depends on path-specific attenuations and delays, specified in the elements of \mathbf{T}^ω (e.g., for the green path it is $\mathbf{T}^\omega[m, 4]$). For an example of a retro-reflective path, which we also consider when working with coaxial projector-camera arrangements, see Figure 6.8(a).

a pixel, making it impossible to separate them without strong assumptions about the actual signal received (e.g., diffuse one-[67] or three-bounce transport [68], dominant peak [137], parametric [53], and/or sparse [24, 63]).

Second, probing the scene with a specific temporal frequency is relatively easy to implement with photonic mixer devices (PMDs). These devices offer affordable spatio-temporal imaging and can be configured to operate at a single emission frequency. Our prototype is built around such a camera, with spatio-temporal projection made possible by replacing the light source of a conventional projector with the PMD’s laser diode. In this respect, our prototype can be thought of as generalizing the point-source/transient-camera system of Heide et al. [53] and the transient-projector/single-pixel-camera system of Kirmani et al. [67].

6.1 Related Work on Transient Imaging

Transient imaging technology. The measurement of optical impulse responses with picosecond resolution was first demonstrated by Duguay and Mattick [25] using a Kerr shutter. Abramson [1] developed a holographic approach using an ultrashort pulse as the reference wave to image light in flight, an idea later transferred to digital sensors [101]. The high-framerate operation of opto-electronic high-speed imagers, imposes major technological challenges on shutter and readout. Direct sampling of the time dimension could be demonstrated for single-pixel opto-electronic photodetectors [32, 66, 68]. The highest-resolution videos of light in flight to this day were obtained by Velten et al. [129] using a streak tube which they synchronized to a pulsed laser, an extremely sensitive setup

with long capture times. To make transient imaging affordable and practical, Heide et al. [53] introduced the use of demodulation (PMD) sensors, which they showed produce correlation data that encodes the transient image. Kadambi et al. [63] later combined the same type of sensor with temporally coded illumination and a sparse deconvolution approach, achieving reconstructions of similar quality.

Spatio-temporal light transport analysis/inversion. The advent of transient imaging enabled the analysis of light transport on the time axis, giving rise to interesting application directions. For example, Naik et al. [88] demonstrated the use of transient data to recover surface reflectance distributions; Kirmani et al. [66] combined an omnidirectional time-resolved sensor with modulation masks for compressive depth imaging, and Velten et al. [128] reconstructed object geometry “around corners”, *i.e.*, from light diffusely reflected off a wall. Although in this work, we also propose a novel technique for enhanced range scanning, we consider our main contribution to be on the theoretical end. Here, Smith et al. [120], who first formulated the problem of transient rendering, laid the foundations for understanding time-resolved light transport in macroscopic scenes. In addition to the aforementioned application-centered works, Wu et al. provided a phenomenological characterization of temporal profiles encountered in transient images [137] as well as a more comprehensive spatio-temporal treatment [139]. In this paper, we suggest to treat the time dimension in the Fourier domain, which is particularly well suited for demodulation sensors used in the most recent works on transient imaging. To our knowledge, we are the first to provide a unifying framework for now-classic works in spatial light transport, and extend it towards transient imaging.

6.2 Light Transport in Space and Time

We begin by deriving the transient frequency transport equation from first principles by considering propagation in space and time.

The space-time impulse response The conventional light transport equation ignores time completely. One way to interpret that equation is to think of the projector as emitting a time-invariant pattern; the camera then captures a photo only after light propagation through the scene has reached a steady state.

To take transient phenomena into account, we must consider the case where light has not yet reached a steady state. The most straightforward way to do this is to add a time dimension to the basic light transport equation:

$$\tilde{\mathbf{i}}(\tau) = \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(0) \quad (6.2)$$

Here $\tilde{\mathbf{p}}(0)$ is an illumination pattern that is emitted for an infinitely-brief interval at time zero, and $\tilde{\mathbf{i}}(\tau)$ records only light that arrived at the camera exactly at time τ . The time-varying matrix $\tilde{\mathbf{T}}(\tau)$ therefore describes the part of the light transport in the scene that has a travel time of exactly τ . From a signal processing perspective, this non-negative matrix-valued function can be thought of as the scene’s space-time impulse response [68].

The relationship between the space-time impulse response and the conventional light transport matrix is a simple time integral:

$$\mathbf{T} = \int_0^\infty \tilde{\mathbf{T}}(\tau) d\tau \quad (6.3)$$

Transport of general spatio-temporal patterns Now consider a general pattern $\tilde{\mathbf{p}}(t)$ that varies in both space and time. If the projector emits $\tilde{\mathbf{p}}(t_0)$ for an infinitely-brief interval at time instant $t = t_0$, the time sequence of camera pixel intensities that results from that emission will be

$$\tilde{\mathbf{i}}(t_0 + \tau) = \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t_0) \quad (6.4)$$

Emitting the full spatio-temporal pattern $\tilde{\mathbf{p}}(t)$ will produce a time-varying image that takes into account all possible travel times:¹

$$\tilde{\mathbf{i}}(t) = \int_{-\infty}^\infty \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t - \tau) d\tau \quad (6.5)$$

$$\stackrel{\text{def}}{=} (\tilde{\mathbf{T}} * \tilde{\mathbf{p}})(t) \quad (6.6)$$

where the operator $*$ convolves a matrix function and a vector function over time.² More information on this form of the convolution operator and a brief overview of algebra and notation on matrix and vector functions can be found in the Appendix B.3.

When the scene's space-time impulse response is known, we can use the convolution integral of Eq. (6.6) to render the scene under any space-time illumination pattern (Figure 6.3). Applying this equation in practice, however, is difficult for several reasons. First, rendering even a single transient snapshot $\tilde{\mathbf{i}}(t_0)$ requires the full 5D space-time impulse response because of the convolution integral involved. Second, this function can be extremely large because of the extra time dimension, compared to the conventional light transport matrix, making it even more challenging to measure, store, and analyze directly. Third, representing this function as a discrete 5D array makes it difficult to infer properties of transient light transport because light travels along continuous and unbounded path lengths, and its dynamic range can be extremely high (*e.g.*, direct paths consist of Dirac peaks). Fourth, using impulse-like illumination patterns to analyze transient transport typically requires expensive equipment, long capture times, and exotic techniques to overcome SNR issues.

The transient frequency transport equation Observe that the convolution in Equation 6.6 is only in the temporal domain. To derive the transient frequency transport equation we apply the convolution theorem to the time axis only, independently for each element of matrix $\tilde{\mathbf{T}}(\tau)$:

$$\underbrace{\mathcal{F}\{\tilde{\mathbf{i}}\}(\omega)}_{\text{image } \mathbf{i}^\omega} = \mathcal{F}\{\tilde{\mathbf{T}} * \tilde{\mathbf{p}}\}(\omega) \quad (6.7)$$

$$= \underbrace{\mathcal{F}\{\tilde{\mathbf{T}}\}(\omega)}_{\text{matrix } \mathbf{T}^\omega} \underbrace{\mathcal{F}\{\tilde{\mathbf{p}}\}(\omega)}_{\text{pattern } \mathbf{p}^\omega} \quad (6.8)$$

¹Since light cannot travel back in time, $\tilde{\mathbf{T}}(\tau)$ is zero for $\tau < 0$.

²Our definition of convolution here is consistent with convolution operators on tensor fields [56].

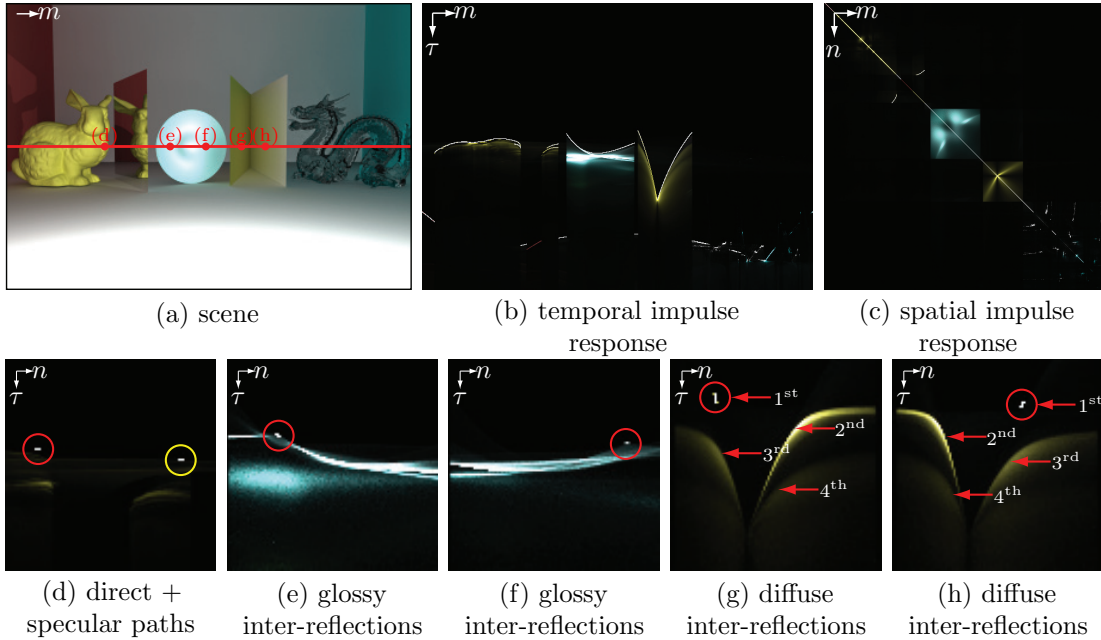


Figure 6.3: Simulated transient light transport, rendered using a modified path tracing integrator for diffuse transport and a photon mapping integrator for specular indirect transport. For each camera pixel m , we simulate the travel time τ of all light paths from the scene lit by a projector pixel n . **(a)** The scene contains, from left to right, a diffuse bunny, a mirror casting specular indirect light on the bunny, a glossy bowl, a diffuse V-shaped wedge, and a glass dragon. **(b)** A transient image of pixels m across times τ . **(c)** A conventional light transport matrix, tabulating transport for each camera pixel m in response to projector pixel n . **(d)-(h)** Space-time impulse response functions for distinct camera pixels, as labelled in (a). Both the direct light paths, highlighted by the red circles, and specular indirect light paths, highlighted in yellow, are impulses in space-time. The responses in (g) and (h) show the first four diffuse bounces of light that occur between the two faces of the V-shaped wedge; the light bounces back and forth between the left and right sides of this wedge as travel time τ increases. Note that each column in (b) is given by summing the columns of the corresponding space-time impulse response, and each column in (c) is given by summing the rows; as a result, while the individual bounces of light produce distinct space-time responses in (g) and (h), these signals become mixed in (b) and (c).

where $\mathcal{F}\{\}$ denotes the element-wise Fourier transform along the time axis and ω denotes temporal frequency. For a fixed frequency ω , Eq. (6.8) is a matrix-vector product whose factors we denote as \mathbf{T}^ω and \mathbf{p}^ω for notational convenience. This brings it into the form shown in Equation 6.1.

The transient frequency transport equation can be interpreted as an image formation model for patterns like those shown in Figure 6.1, which contain just one temporal frequency. The most important advantage of this model is that it is separable in three domains simultaneously: the temporal frequency domain and the domains of the projector and the camera pixels, respectively. This leads to a mathematically simpler representation for light transport analysis. In particular, the contribution of all light paths from a specific projector pixel to a specific camera pixel—including all attenuations and time delays—is captured in just one complex number per temporal frequency.

Analyzing light transport one frequency at a time offers computational advantages as well: in contrast to the scene’s space-time impulse response which is 5D, the transient frequency transport

matrix is a 4D object that has exactly the same size as the conventional transport matrix. Moreover, inexpensive and light-efficient PMD cameras can be configured to follow this single-frequency image formation model exactly, making it possible to perform transient light transport analysis *directly* in the temporal frequency domain.

6.3 Analysis by Temporal Frequency Probing

Much of the theory established for conventional light transport analysis applies to the transient case, as long as we restrict illumination to patterns of a single temporal frequency. Below we consider transient versions of two techniques we have implemented, as well as of the rendering equation.

Probing the temporal frequency transport matrix We can readily use the probing model of Eq. (3.1) for transient imaging because matrix elements $\mathbf{T}[m, n]$ and $\mathbf{T}^\omega[m, n]$ represent the exact same set of light paths for any m, n . As a result, matrix $\mathbf{\Pi}$ has the same effect in both imaging regimes:

$$\mathbf{i}^\omega = (\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1} \quad (6.9)$$

The transient rendering equation This equation represents all light transport as a function of time t [68]:

$$l_{ij}(t) = q_{ij}(t) + \sum_{k=1}^K f_{kij} l_{ki}(t - \tau_{ki}) \quad (6.10)$$

where $l_{ij}(t)$ captures radiant energy along a ray, $q_{ij}(t)$ is the emitted radiant energy, f_{kij} describes the proportion of radiant energy from point k to point i that will be reflected towards point j , and τ_{ki} corresponds to the flight time between two points k and i . Note that this spatially-discrete expression is based on a continuous-valued (i.e. non-discretized) set of travel times τ_{ki} between pairwise points.

The transient rendering equation is itself separable in the temporal frequency domain:

$$\begin{aligned} l_{ij}^\omega &= q_{ij}^\omega + \int_{-\infty}^{\infty} \left(\sum_{k=1}^K f_{kij} l_{ki}(t - \tau_{ki}) \right) e^{-2\pi i t \omega} dt \\ &= q_{ij}^\omega + \sum_{k=1}^K f_{kij} e^{-2\pi i \tau_{ki} \omega} \int_{-\infty}^{\infty} l_{ki}(t) e^{-2\pi i t \omega} dt \\ &= q_{ij}^\omega + \sum_{k=1}^K f_{kij} e^{-2\pi i \tau_{ki} \omega} l_{ki}^\omega \end{aligned} \quad (6.11)$$

where l_{ij}^ω denotes the Fourier coefficient of temporal function $l_{ij}(t)$ for frequency ω . The key difference between Eq. (6.11) and the conventional rendering equation [64] are the complex exponential factors that represent phase delay between pairs of scene points. Most of the equations in Table 6.1 follow

as direct consequences of this equation. Note that the equation remains valid for discrete temporal frequencies as well.

Separating specular and diffuse indirect components Nayar et al.’s separation technique discussed in Chapter 3 can be applied to the transient domain too. Following the approach outlined in Section 3.4, we capture the direct and indirect components by probing the transient transport matrix at a single frequency; then we apply Nayar et al.’s method to decompose the indirect component further, into its specular and diffuse indirect components.

Relation between matrices \mathbf{T}^ω for different frequencies ω Per-frequency analysis is most effective when transport matrices at different temporal frequencies are related, so that analyzing one tells us something about the others. Fortunately strong correlations do exist, and we use three of them here. First and foremost, element $\mathbf{T}^\omega[m, n]$ represents the same physical 3D transport path(s) from projector pixel n to camera pixel m regardless of the frequency ω . Thus, if it represents direct or specular indirect transport at one frequency, it will do so at all others. Second, a direct path between these pixels contributes a Dirac peak to pixel m ’s temporal profile, which has a flat spectrum. We can therefore use element $\mathbf{T}^\omega[m, n]$ to predict amplitude at all frequencies, and to predict phase up to a discrete (phase-unwrapping) ambiguity. Third, a similarly strong correlation occurs for elements representing specular indirect paths, which also produce Dirac temporal peaks in typical settings.

In contrast, correlations are much weaker when the transport between pixels n and m is due to diffuse indirect paths. Such transport often involves a broad distribution of path lengths and contributes temporal profiles with complex shape and small spectral support. Although this makes it hard to predict the phase and amplitude in general, the contribution of diffuse indirect paths is known to vanish for large enough frequencies ω [43].

#	Description	Reference(s)	Conventional Light Transport	Single-Frequency Transient Light Transport
1	light transport equation	Eq. (1.1)	$\mathbf{i} = \mathbf{T} \mathbf{p}$	$\mathbf{i}^\omega = \mathbf{T}^\omega \mathbf{p}^\omega$
2	correspondence equation	Eq. (1.11)	$n = \arg \max_k \mathbf{T}[m, k]$	$n = \arg \max_k \mathbf{T}^\omega[m, k]$
3	dual equation	Eq. (1.12)	$\mathbf{i} = \mathbf{T}^\top \mathbf{p}$	$\mathbf{i}^\omega = (\mathbf{T}^\omega)^\top \mathbf{p}^\omega$
4	transport eigenvectors	Eq. (2.1)	$\lambda \mathbf{v} = \mathbf{T} \mathbf{v}$	$\lambda \mathbf{v} = \mathbf{T}^\omega \mathbf{v}$
5	inverse equation	Eq. (2.8)	$\mathbf{i} = \mathbf{T}^\dagger \mathbf{p}$	$\mathbf{i}^\omega = (\mathbf{T}^\omega)^\dagger \mathbf{p}^\omega$
6	probing equation	Eq. (3.1)	$\mathbf{i} = (\mathbf{T} \odot \mathbf{\Pi}) \mathbf{1}$	$\mathbf{i}^\omega = (\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1}$
7	low/high-frequency transport separation	Figure 3.10	$\mathbf{i}_{\text{low}} = \frac{1}{\alpha} \min_k \mathbf{T} \mathbf{p}_k$ $\mathbf{i}_{\text{high}} = \max_k \mathbf{T} \mathbf{p}_k - \alpha \mathbf{i}_{\text{low}}$	$\mathbf{i}_{\text{low}}^\omega = \frac{1}{\alpha} \min_k \mathbf{T}^\omega \mathbf{p}_k^\omega$ $\mathbf{i}_{\text{high}}^\omega = \max_k \mathbf{T}^\omega \mathbf{p}_k^\omega - \alpha \mathbf{i}_{\text{low}}^\omega$

Table 6.1: Light transport analysis techniques that have simple extensions to the single-frequency transient domain. In each instance, the transient formulation becomes the conventional (steady state) formulation at $\omega = 0$. We once again combine optical probing (row 6) with Nayar’s frequency separation technique (row 7) to capture three transport components: a direct component, a specular indirect component, and a diffuse indirect component.

Algorithm 4 *Acquire a PMD photo for illumination pattern \mathbf{p} .*

In: frequency ω and real-valued spatial illumination pattern \mathbf{p}

Out: photo equal to $\mathbf{i}^\omega = \mathbf{T}^\omega \mathbf{p}^\omega$

- 1: given frequency ω , set hardware-defined modulation functions $f(t)$, $g(t)$ such that $h(\tau) = (f * g)(\tau) = \cos(\omega\tau) + b$ for some arbitrary constant offset b
 - 2: define phase delay vector $\boldsymbol{\phi} = [0, \frac{\pi}{2\omega}, \frac{\pi}{\omega}, \frac{3\pi}{2\omega}]$
 - 3: **for** $d = 1$ to 4 **do**
 - 4: display pattern \mathbf{p}
 - 5: modulate sensor and source with $f(t)$, $g(t - \phi_d)$ so that $h(\tau) = \cos(\omega(\tau + \phi_d)) + b$
 - 6: capture image \mathbf{i}_d^ω satisfying Eq. (6.12)
 - 7: **end for**
 - 8: set $\mathbf{i}^\omega = (\mathbf{i}_1^\omega - \mathbf{i}_3^\omega) + i(\mathbf{i}_2^\omega - \mathbf{i}_4^\omega)$
 - 9: **return** captured PMD photo \mathbf{i}^ω
-

6.4 Implementation

Basic imaging procedure The working principle behind photonic mixer devices (PMDs) is the synchronous modulation of both the incident and outgoing light at high frequencies. Modulating the incident light by function $f(t)$ and projecting pattern $\tilde{\mathbf{p}}(t) = g(t)\mathbf{p}$, with temporal modulation function $g(t)$ and 2D spatial pattern \mathbf{p} , yields the following image formation model:

$$\begin{aligned}
 \mathbf{i} &= \int_0^{NT} f(t) \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t - \tau) d\tau dt \\
 &= \left[\int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \left(N \int_0^T f(t)g(t - \tau) dt \right) d\tau \right] \mathbf{p} \\
 &= \left[N \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) h(\tau) d\tau \right] \mathbf{p}
 \end{aligned} \tag{6.12}$$

where the function $h(\tau)$ is the convolution between the two modulation functions $f(t)$ and $g(t)$, $T = \frac{1}{\omega}$ is the modulation period, and the integer N is the number of periods captured during a single exposure. This becomes exactly Eq. (6.1) when the convolution function is chosen to be the complex exponential $h(\tau) = e^{-2\pi i\tau\omega}$.

In practice, the modulation functions and their convolution have non-negative, real values. The imaging procedure of a PMD camera synthesizes images for mean-zero, complex-valued $h(\tau)$ by capturing and linearly combining four images: two for the real term and two for the imaginary term. See Algorithm 4 for the basic imaging procedure when the illumination pattern \mathbf{p} is real valued.³

In real PMD cameras, the specific modulation functions $g(t)$ and $f(t)$ are usually determined by hardware constraints, and cannot be chosen arbitrarily. In particular, $h(\tau)$ is not always an exact sinusoid. We note, however, that, if $g(t)$ and $f(t)$ are periodic with frequency ω , then so is $h(\tau)$. Therefore $h(\tau)$ is a superposition of the base frequency ω and its harmonics. In this case, the transport matrices are simply a weighted sum of the transport matrices for the base frequency and

³Illuminating the scene with a complex pattern \mathbf{p}^ω is not physically realizable. It can be simulated, however, by capturing four PMD photos in a way similar to Algorithm 4. We do not use such patterns in our experiments.

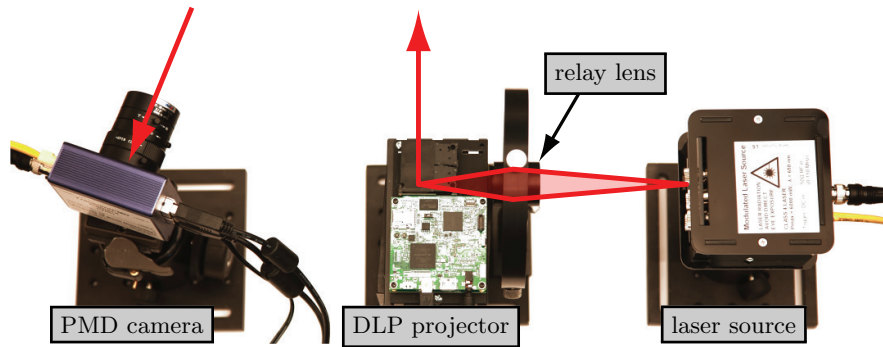


Figure 6.4: Overhead view of our prototype. The modulated laser source (right) emits light that passes through a lens and the existing optics of a DLP LightCrafter projector (middle) on its way to the scene. The PMD camera (left) captures the light returning from the scene.

its harmonics. In practice, we conduct the bulk of our analysis at frequency $\omega_1 = 100$ MHz (*i.e.*, depth acquisition and phase estimation for direct and specular indirect paths), where our prototype’s deviation from a perfect sinusoid is negligible.

Hardware Figure 6.4 shows an overhead photo of our prototype in its non-coaxial configuration, in which the projector and camera can be thought of as forming a stereo pair. We modified the 160×120 -resolution sensor of a PMD PhotonICs 19k-S3 by replacing the internal modulation signal with our own external modulation signals, outputting frequencies that range from 12 MHz to 140 MHz. We illuminate the scene with a custom laser projector, built by replacing the RGB light source of an off-the-shelf DLP LightCrafter projector with the light emitted from six 650 nm laser diodes. A single 40 mm lens (Thorlabs TRS254-040-A-ML) directs the modulated laser illumination through the existing optics of the projector, from which the RGB LEDs and dichroic mirrors were removed. For coaxial camera and projector arrangements, we add a 50/50 beamsplitter (Edmund Optics #46-583) to optically align the projector and the camera.

The exposure time of our PMD camera was strictly limited to the range of 1 to 8 ms. We used a 1 ms exposure time (Step 5 of Algorithm 4) when operating the camera in a stereo arrangement as shown in Figure 6.4; in coaxial arrangements we increased it to 8 ms to compensate for the system’s 25% ideal light efficiency and for beamsplitter imperfections. Thus, capturing one PMD photo with Algorithm 4 takes 4 or 32 ms, depending on the arrangement.

Calibration The sensing behavior of individual pixels is not perfectly uniform over the entire PMD sensor. We model these deviations as an element-wise product between two complex images—the PMD photo of the scene and a 2D “noise” pattern \mathbf{w}^ω . The pattern’s amplitude, $|\mathbf{w}^\omega|$, models the non-uniformity of pixel sensitivities and represents *fixed pattern noise* (FPN). Its phase, $\arg \mathbf{w}^\omega$, models pixel-specific offsets in the phase of the sensor’s modulation function, and thus represents *fixed phase-pattern noise* (FPPN). Removing this deterministic noise pattern involves two steps: (1) pre-compute \mathbf{w}^ω by capturing a photo of a scene that is constant in both amplitude and phase, and (2) multiply its reciprocal $(\mathbf{w}^\omega)^{-1}$ with every captured PMD photo.

Algorithm 5 *Combine PMD imaging and matrix probing.*

In: phase delay ϕ_d ; spatial binary patterns $\mathbf{p}_1, \dots, \mathbf{p}_K$ and masks $\mathbf{m}_1, \dots, \mathbf{m}_K$ such that $\mathbf{\Pi} \approx \sum_1^K \mathbf{m}_k(\mathbf{p}_k)^\top$

Out: component \mathbf{i}_d^ω of PMD photo $(\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1}$

- 1: modulate sensor and source according to $f(t), g(t - \phi_d)$
 - 2: open camera shutter
 - 3: **for** $k = 1$ **to** K **do**
 - 4: apply pixel mask \mathbf{m}_k
 - 5: display pattern \mathbf{p}_k for $\frac{1}{K}$ th the exposure time
 - 6: **end for**
 - 7: close camera shutter
 - 8: **return** image \mathbf{i}_d^ω
-

Algorithm 6 *Acquire the direct component of a PMD photo.*

In: frequency ω and sequence length K

Out: direct PMD photo $\mathbf{i}_{\text{direct}}^\omega$

- 1: acquire conventional PMD photo \mathbf{i}^ω using Algorithm 4 and an all-white pattern
 - 2: construct indirect-only binary sequences $\mathbf{p}_1, \dots, \mathbf{p}_K$ and $\mathbf{m}_1, \dots, \mathbf{m}_K$
 - 3: acquire indirect-only PMD photo $\mathbf{i}_{\text{indirect}}^\omega$ using Algorithms 4 and 5
 - 4: **return** image $\mathbf{i}^\omega - \mathbf{i}_{\text{indirect}}^\omega$
-

Transport matrix probing with a PMD camera Probing in the transient domain comes from approximating Eq. (6.9) with the same sum of bilinear matrix-vector products as in Eq. (3.4):

$$(\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1} \approx \sum_{k=1}^K \mathbf{m}_k \odot \mathbf{T}^\omega \mathbf{p}_k \quad (6.13)$$

where the sequence of vectors \mathbf{p}_k and \mathbf{m}_k defines a rank- K approximation of the probing matrix: $\mathbf{\Pi} \approx \sum_1^K \mathbf{m}_k(\mathbf{p}_k)^\top$. When ω is the DC frequency, Eqs. (6.13) and (3.4) are identical.

Generalizing this procedure to arbitrary temporal frequencies ω and to complex-valued PMD photos is straightforward: we simply replace Steps 4 and 5 of Algorithm 4 (which acquire images without a mask under a fixed illumination pattern) with Steps 1-8 of Algorithm 5 (which change masks and illumination patterns K times). Note that this modification of Algorithm 4 does not change the total number of images captured, which remains equal to four. Unfortunately, hardware constraints prevented us from implementing Algorithm 5 exactly as shown, impacting the number of images we capture in experiments. We return to this at the end of Section 6.4.

Mask/projection sequences for indirect-only imaging Following the approach of Chapter 4, we compute the epipolar geometry between the projector and the PMD camera and construct each mask \mathbf{m}_k by randomly turning each of its epipolar lines “on” or “off” with probability 0.5. Given mask \mathbf{m}_k , we construct the corresponding pattern \mathbf{p}_k by turning “on” all epipolar lines that were turned “off” in the mask, and vice-versa.

Acquiring direct PMD photos The basic procedure is shown in Algorithm 6. It amounts to capturing a conventional PMD photo and then subtracting its indirect-only component.

Algorithm 7 *Acquire specular and diffuse indirect PMD photos.*

In: frequency ω and sequence lengths J and K

Out: photos $\mathbf{i}_{\text{low}}^\omega$ and $\mathbf{i}_{\text{high}}^\omega$ containing diffuse and specular indirect transport components, respectively

- 1: construct indirect-only binary sequences $\mathbf{p}_1, \dots, \mathbf{p}_K$ and $\mathbf{m}_1, \dots, \mathbf{m}_K$
- 2: construct high-frequency pattern sequence, $\mathbf{q}_1, \dots, \mathbf{q}_J$
- 3: set $\mathbf{i}_{\text{low}}^\omega = \mathbf{i}_{\text{high}}^\omega = 0$
- 4: **for** $j = 1$ **to** J **do**
- 5: acquire a PMD photo \mathbf{i}^ω using Algorithms 4 and 5, the mask sequence $\mathbf{m}_1, \dots, \mathbf{m}_K$ and the pattern sequence $\mathbf{p}_1 \odot \mathbf{q}_j, \dots, \mathbf{p}_K \odot \mathbf{q}_j$
- 6: set $\mathbf{i}_{\text{low}}^\omega = \min(\mathbf{i}_{\text{low}}^\omega, \mathbf{i}^\omega)$
- 7: set $\mathbf{i}_{\text{high}}^\omega = \max(\mathbf{i}_{\text{high}}^\omega, \mathbf{i}^\omega)$
- 8: **end for**
- 9: set $\mathbf{i}_{\text{high}}^\omega = \mathbf{i}_{\text{high}}^\omega - \mathbf{i}_{\text{low}}^\omega$
- 10: set $\mathbf{i}_{\text{low}}^\omega = \frac{1}{\alpha} \mathbf{i}_{\text{low}}^\omega$
- 11: **return** separated components $\mathbf{i}_{\text{low}}^\omega, \mathbf{i}_{\text{high}}^\omega$

Separating specular and diffuse indirect transport components We further decompose the indirect component of a PMD photo into its specular and diffuse indirect components. We do this with the approach described in Section 3.4 and Figure 3.11. Algorithm 7 shows the basic steps, adapted to the case of PMD imaging.

Depth acquisition from direct PMD photos The phase component of a PMD pixel encodes the depth of each scene point as a value that ranges from 0 to 2π . Specifically, a coaxial system produces pixel values of the following form:

$$\mathbf{i}^\omega = ae^{-2\pi id \frac{2\omega}{c}} \quad (6.14)$$

where a is the albedo of a scene point, d is its depth, $2d$ is the round-trip distance travelled by light to the camera, and c is the speed of light. This produces ambiguities in the relation between phase and depth. For example, frequency $\omega_1 = 100$ MHz only encodes depth for a maximum unambiguous range of $\frac{c}{2\omega_1} \approx 1.5$ m. For a greater depth range, we acquire direct PMD photos for two frequencies ω_1 and $\omega_2 = \omega_1/2$ and use phase-unwrapping [82] to calculate depth. Specifically, given photos \mathbf{i}^{ω_1} and \mathbf{i}^{ω_2} , the phase-unwrapped depth is

$$d = \frac{c}{2\omega_1} \left\{ 2 - \frac{\arg \mathbf{i}^{\omega_1}}{2\pi} - \left\lfloor \frac{\pi + 2 \arg \mathbf{i}^{\omega_2} - \arg \mathbf{i}^{\omega_1}}{2\pi} \right\rfloor \right\} \quad (6.15)$$

Since the scenes in our experiments were well within 1.5 m, imaging at a single frequency ω_1 with no unwrapping was sufficient.

Hardware limitations and computational masking The 8 ms maximum exposure time of our PMD camera prevented us from implementing Algorithm 5 in one shot. This is because our DLP LightCrafter kits can perform at most $K = 21$ projection/masking operations in that interval, leading to much poorer approximations of Eq. (6.13) compared to the 96 patterns that fit in a 36 ms video frame. To overcome this limitation we mask images computationally, by pushing Steps 2

PMD imaging task	Optical masking	Computational masking	Experiments
illumination pattern \mathbf{p}	4	4	4
indirect-only	4	$4K$	512
direct	8	$4 + 4K$	516
specular/diffuse indirect	$4J$	$4KJ$	3072
depth acquisition	8	$4 + 4K$	516
transport decomposition	$8 + 4J$	$4 + 4K + 4KJ$	3588
light-in-flight imaging	$8 + 4J + 4F$	$4 + 4K + 4KJ + 4F$	3718

Table 6.2: Images required for transient transport analysis. We use $K = 128$, $J = 6$, $F = 65$ in our experiments, with $\omega_1 = 100$ MHz and no phase unwrapping. See Sections 6.4 and 6.5 for explanations.

and 7 of Algorithm 5 inside the loop. In particular, we capture $K = 128$ images individually, each with a 1 ms exposure; we multiply element-wise the image captured in the k^{th} iteration with the associated mask \mathbf{m}_k ; and accumulate the result. This increased significantly the number of images we had to capture for the experiments in Section 6.5. Table 6.2 gives full details.

6.5 Results

Transport decomposition for time-of-flight imaging We start with separating a scene’s transient appearance into its three transport components—direct, specular indirect, and diffuse indirect. Since both the direct and the specular indirect components are due to distinct temporally-isolated reflection events, they correspond to Dirac peaks in the time domain. We run Algorithm 6 for frequency ω_1 to localize the former and Algorithm 7 for the same frequency to identify and localize the latter. This also gives us the diffuse indirect contributions for frequency ω_1 .

Figure 6.5 shows this decomposition for a scene containing a mirror and a miniature statue of Michelangelo’s David positioned near the corner of a room. We used a coaxial projector-camera arrangement for this example.

Time-of-flight depth images robust to indirect transport PMD cameras compute depth by acquiring PMD photos for one or more frequencies with a co-located light source, and then using Eq. (6.15) to turn phases into depth values. An unfortunate consequence of this approach is that indirect light has a pronounced influence on the measurements. Though methods exist for removing the influence of diffuse indirect light from a PMD image computationally by relying on predictive models [28, 33, 61] or optically by operating at higher modulation frequencies ω [43], these methods do not generalize for handling all forms of indirect light.

We demonstrate the ability to recover accurate depth images that are robust to indirect light transport using PMD cameras. Specifically, we use Algorithm 6 to capture the direct component of a scene for frequency ω_1 . This photo is by definition invariant to indirect transport so its phase yields transport-robust depth maps.

Figure 6.7 shows depth acquisition results for three scenes with significant specular and diffuse indirect transport. The scene in row 1 is a bowl pressed flush against a planar wall (Figure 6.6).

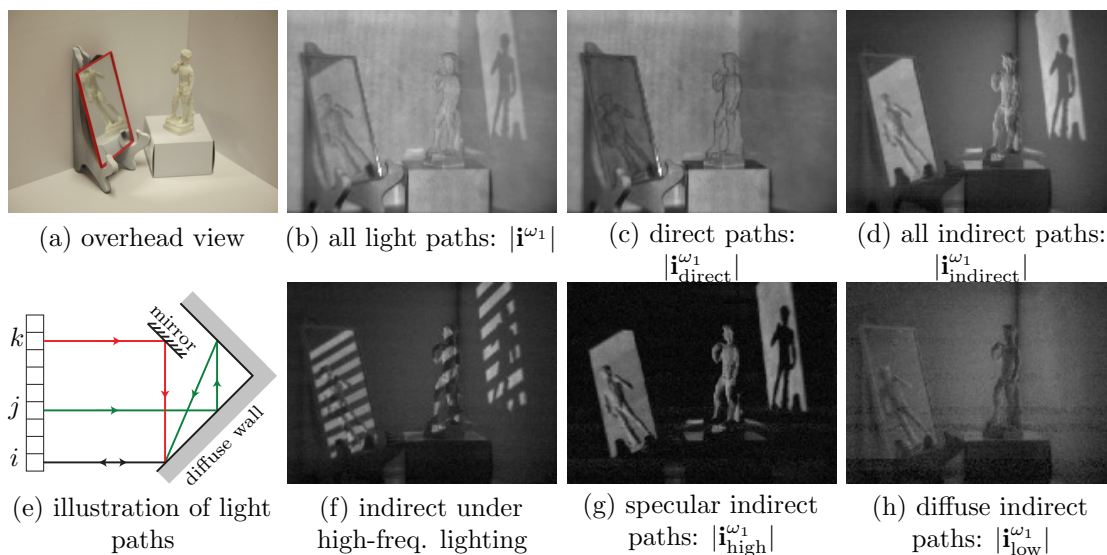


Figure 6.5: Capturing PMD photos corresponding to individual transport components. Since these photos are complex, we only show their amplitudes in this figure. (a) A scene containing a mirror and a miniature statue of David. (b) PMD photo returned by Algorithm 4 for an all-white illumination pattern. (c) PMD photo returned by Algorithm 6. (d) PMD photo returned by Algorithm 5. (e) Basic path geometries for a camera pixel i : direct light path $i \rightarrow i$ (black), specular indirect path $k \rightarrow i$ (red), and diffuse indirect path $j \rightarrow i$ (green). (f) One of six PMD photos acquired in Step 5 of Algorithm 7. This photo is an indirect-only transient view of the scene under a projection pattern consisting of binary stripes. (g)-(h) PMD photos returned by Algorithm 7.

The deep concavity of the bowl produces a significant amount of indirect light through diffuse inter-reflections. As a result, the RAW time-of-flight measurements place the base of the bowl 4 to 5 cm behind the wall itself (columns (a) and (c)). Our approach reconstructs a physically-valid solution, with the bowl’s base coinciding with the expected position of the wall (columns (b), (d) and (e)). Note that the reconstructions of the convex handle of the bowl are the same in both cases (lower-left region of the slice in column (e)). The concave lip, on the other hand, in the conventionally-acquired time-of-flight image is offset from its correct position (lower-right region of the slice in column (e)).

Row 2 of Figure 6.7 shows reconstruction results for an open book whose pages were kept as flat as possible. The significant diffuse inter-reflections between these pages distort their conventionally-acquired shape (columns (a) and (c)) but leave direct shape measurements unaffected: their flat shape is evident in columns (b), (d) and (e). Row 3 shows results for the David scene. The specular indirect light paths, reflected by the mirror onto the right wall, produce an embossed silhouette on the wall, indicated by the yellow arrow in column (c). In contrast, the direct component is not influenced by specular indirect paths—and also recovers the depth of objects viewed within the mirror itself.

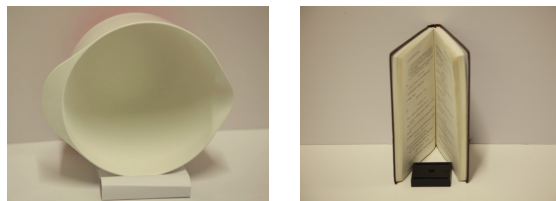


Figure 6.6: Conventional photos of scenes reconstructed in Figure 6.7. We used a coaxial projector-camera arrangement for these scenes.

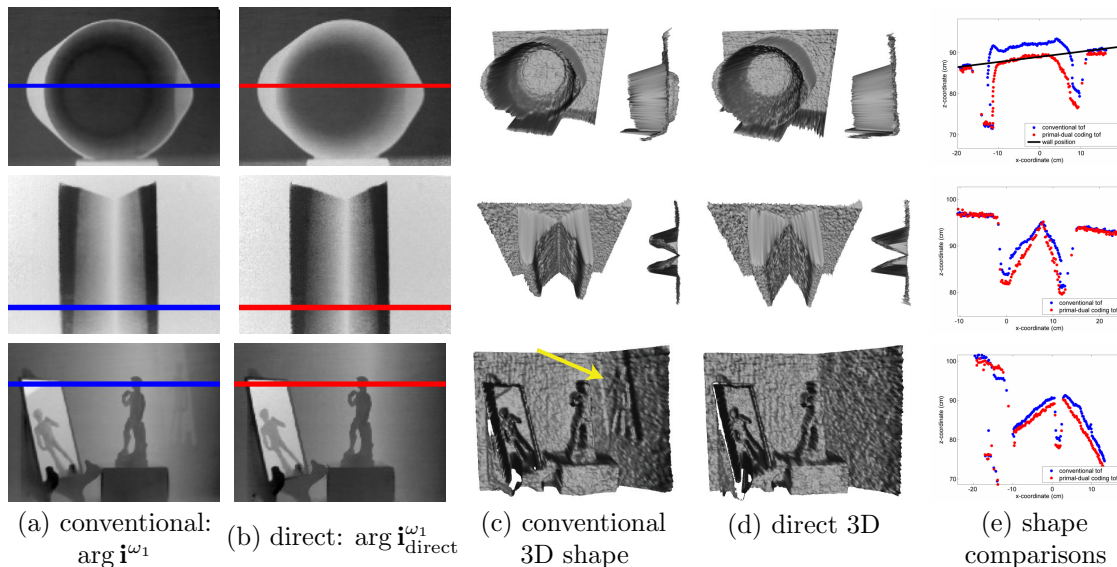


Figure 6.7: Geometry results for the three scenes in Figure 6.5(a) and 6.6. (a) The phase of conventionally-acquired PMD photos. (b) The phase of the direct component returned by Algorithm 6. (c)-(d) Views of the 3D meshes computed from (a) and (b), respectively. (e) Plots of the x - and z -coordinates for a slice of each scene, computed from the conventional (blue) and the direct (red) phases. Observe that the base of the conventionally-acquired bowl protrudes through the back wall by about 5 cm; the pages of the conventionally-acquired book appear curved; the corner of the room in the conventionally-acquired David scene is rounded, and the specular indirect paths illuminating the room’s right wall produce a 2 to 3 cm offset in depth values. None of these artifacts appear in (b) or (d).

Distinguishing between direct views and mirror reflections The direct component of a photo represents direct light paths, but it also contains retro-reflected light paths. We can use the phase of PMD photos to classify pixels according to whether they receive light directly or via a retro-reflective path (*e.g.*, as would occur if we viewed a diffuse point through a mirror). In particular, when the projector and camera are coaxial, direct light paths are always the shortest paths received by each camera pixel. Retro-reflective paths, on the other hand, do not have this property: if a pixel receives light from both retro-reflective and specular indirect light paths, the specular indirect paths always take the shortest route to the camera. This turns pixel classification into a simple pixel-wise comparison of phase values in direct and specular indirect components. See Figure 6.8 for a detailed demonstration.

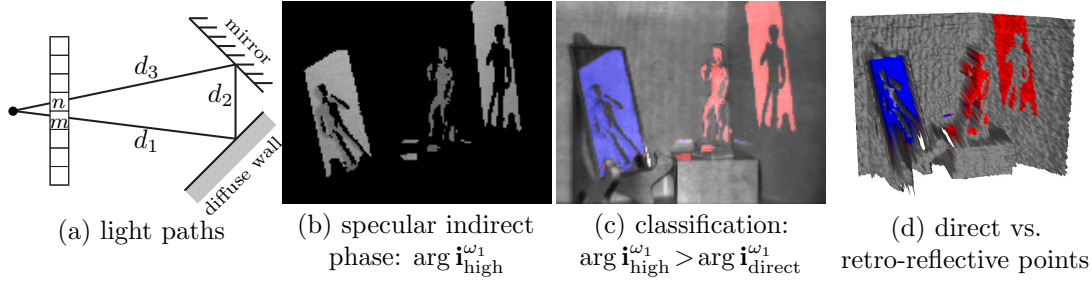


Figure 6.8: Distinguishing between direct views and mirror reflections. **(a)** For a co-located camera and projector, the direct light path $m \rightarrow m$ has length $2d_1$, the specular indirect paths $m \rightarrow n$ and $n \rightarrow m$ have lengths $d_1 + d_2 + d_3$, and the retro-reflective path $n \rightarrow n$ has length $2(d_2 + d_3)$. It follows that specular indirect paths are always longer than direct paths and shorter than retro-reflective paths. **(b)** The phase of the specular indirect component of the David scene, with phases zeroed-out at pixels where the amplitude $|\mathbf{i}_{\text{high}}^{\omega_1}|$ is low, and thus phase is uninformative. **(c)** The result of a pixel-wise “greater than” operator between the specular indirect component shown in (b) and the direct component in row 3 of Figure 6.7(b). Retro-reflective light paths are shown in blue and direct light paths in red. Note that we cannot distinguish between direct and retro-reflective pixels in the absence of a specular indirect component. **(d)** Color-coded mesh using the conventions in (c).

Capturing evolving wavefronts of “light-in-flight” PMD cameras provide a cheap and light-efficient alternative to capturing transient images [53]. Unfortunately, the maximum modulation frequency is limited by PMD hardware, making reconstructing transient images a highly ill-conditioned problem, and requiring strong regularizers to perform the reconstruction.

To overcome these limitations, we propose a transport-specific reconstruction for each of the direct, specular indirect, and diffuse indirect transport components of a scene. In particular, we reconstruct the temporal intensity profiles for direct and specular indirect light paths by running Algorithms 6 and 7 at frequency ω_1 , and fitting a Dirac peak to each pixel in $\mathbf{i}_{\text{direct}}^{\omega_1}$ and $\mathbf{i}_{\text{high}}^{\omega_1}$. To reconstruct the diffuse indirect wavefront of a scene, we (1) capture conventional PMD photos at $F = 65$ frequencies from 12 to 140 MHz in 2 MHz increments, (2) subtract the predicted contribution of the direct and specular indirect Dirac peaks from these photos, and (3) apply the reconstruction method of Heide et al. [53] to recover the temporal intensity profile due to diffuse indirect light.

Figure 6.9 compares our approach to that of Heide et al. [53], which does not perform transport decomposition. The first scene (rows 1 and 2) contains transparent objects that refract the wavefront travelling through them. The second scene (rows 3 and 4) produces strong diffuse inter-reflections from two large bowls that are positioned near a corner. The third scene (rows 5 and 6) includes a mirror and a jug filled with milky water; these produce specular indirect light paths via the mirror, as well as volumetric scattering through the water. Note that the direct and specular indirect wavefronts propagating through the scene are well-resolved in both space and time in our reconstructions, whereas they appear broad and poorly-localized in the absence of transport decomposition.



(a) PMD image (b) transient frame ($\tau = 1.9$ ns) (c) transient frame ($\tau = 3.2$ ns) (d) transient frame ($\tau = 5.0$ ns)

Figure 6.9: Transient image comparisons between Heide et al. [53] and our approach. These scenes were all imaged with a stereo projector-camera arrangement. (a) Steady state images of the scene captured with a normal camera under ambient illumination (rows 1, 3, and 5) and our PMD camera with Algorithm 4 under white projector illumination (rows 2, 4, and 6). (b)-(d) Frames from the temporal intensity profile reconstructed using the conventional approach (rows 1, 3, and 5) and our approach (rows 2, 4, and 6). Note the sharp Dirac impulses travelling along the walls in our reconstructions, which meet in the corner of the scene. These correspond to direct transport, although sharp specular wavefronts also occur in some cases (rows 2 and 6). Moreover, even though reconstructing the diffuse indirect time profile remains highly ill-conditioned, reconstructing the direct and specular indirect contributions separately simplifies the reconstruction process for the diffuse indirect component, and improves its accuracy. This is most evident in column (b), where contributions from diffuse scattering and inter-reflections appear to occur throughout the three scenes in the conventional reconstructions. This is physically impossible, however, since the elapsed time is too brief for light to have actually reached those regions. In contrast, diffuse indirect components are dark in our reconstructions and appear to trail the direct wavefronts.

6.6 Summary and Contributions

We uncovered a key mathematical link between time-of-flight and conventional photography, making it possible to readily transfer computational illumination techniques from one domain to the other. The approach hinges on an ability to probe scenes by illuminating them with coded space-time patterns that vary sinusoidally in time, with the same frequency at every pixel. Technology with this capability built-in is already entering the consumer market in the form of off-the-shelf devices such as PMD cameras and the Kinect 2.

On the practical side, we have demonstrated that

- optical probing can be combined with time-of-flight depth cameras to capture shape robust to all forms of indirect light;
- we can also use optical probing to significantly increase the temporal resolution of PMD-captured transient images; and
- we can distinguish between direct views and mirror reflections of an object—which is impossible to do from just stereo or time-of-flight constraints alone, and suggests that perhaps a great deal of untapped scene information is lurking around the corner.

The design of new 3D scene analysis techniques that integrates both geometric and time-of-flight constraints on light paths is an exciting avenue for future work.

Chapter 7

Conclusions and Future Work

Throughout this thesis, we establish a mathematical framework for active illumination, referred to as *optical linear algebra* for computational light transport. This framework describes active illumination techniques as matrix operators that we apply directly onto a scene’s light transport matrix \mathbf{T} . These matrix operations infer scene properties (*e.g.* shape, indirect appearance) from the transport matrix, and provide a rigorous mathematical basis for describing and analyzing active illumination.

We also identify the largest practical issue plaguing computational light transport today: the prohibitively-high cost of evaluating these operators on large, unknown transport matrices. Although many techniques have been proposed to capture the matrix \mathbf{T} and analyze it in the numerical domain [22, 27, 30, 97, 99, 111, 114, 115, 130], they nevertheless require hundreds or thousands of photos and rely on vast computational resources to store and process the matrix.

Our solution is simple: avoid measuring the matrix \mathbf{T} in the first place. Instead, we use optical computing principles to analyze the matrix “online” (*i.e.*, without direct access to the elements of \mathbf{T}). Specifically, we

- equate numerical calculations (*e.g.* matrix-vector products) to the optical processes used in light transport analysis (*e.g.* project-and-capture operations);
- associate the type of matrix \mathbf{T} (*e.g.* its symmetry, rank, sparsity) to optical properties (*e.g.* the arrangement of lights and cameras); and
- identify classes of numerical algorithms (*e.g.* Krylov subspace methods, matrix probing) that can be partially or fully implemented in the optical domain.

From these principles, we derived very efficient optical solutions to a variety of problems in computer vision and graphics.

In Chapter 2, we provided an overview of Krylov subspaces, and focused on optical implementations of two well-known Krylov subspace methods: Arnoldi and GMRES. Arnoldi iteration is particularly efficient at finding the top few eigenvectors of a matrix from only a few dozen matrix-vector operations, and the GMRES algorithm provides an efficient way to solve linear systems. Our optical Arnoldi algorithm captured the eigenvectors of \mathbf{T} to compute a low-rank approximation of it;

we then used this matrix approximation to relight photos of the scene under synthetic illumination conditions. Our optical GMRES algorithm inverted the light transport equation by taking a photo of the scene as input and computing the illumination conditions that produced it.

Chapter 3 introduced our optical matrix probing procedure. Optical probing is a new imaging technique that offers unprecedented control over the amount of radiant energy transmitted along specific light paths. This procedure uses programmable optics (*e.g.* DMDs, LCDs) to simultaneously code the radiant energy emitted into a scene and modulate the radiant energy incident on the sensor. For coaxial projector-camera arrangements, optical probing can be used to attenuate or enhance the direct and indirect components of a photo. A key feature of optical probing is that it operates almost exclusively in the optical domain: results consist of directly-acquired, unprocessed RAW photos or differences between them.

Chapter 4 extended optical probing to general projector and camera configurations. The approach rests on a crucial link between stereo geometry and light transport: while direct light always obeys the epipolar geometry of a projector-camera pair, indirect light overwhelming does not. We probed for these epipolar (direct) and non-epipolar (indirect) light paths with an experimental optical probing camera that operates at live video rates. We also used our prototype camera to make existing structured-light 3D shape algorithms robust to indirect transport, and also turn them into one-shot methods for dynamic 3D shape capture.

Chapter 5 explored how to make best use of the radiant power available for optical probing. We derived an optimization procedure to compute energy-efficient codes for our probing procedure, while respecting the physical constraints imposed on the coding mechanisms; these codes yield brighter images with fewer artifacts when probing for the direct and indirect components, and offer solutions to other probing tasks where codes were not known previously (*e.g.* short-range indirect imaging). Furthermore, we developed another prototype camera that maximizes energy efficiency for epipolar and non-epipolar imaging through the novel combination of a low-power laser projector and a rolling shutter camera; we used this prototype to demonstrate never-seen-before capabilities such as reconstructing the 3D shape of objects in the presence of strong indirect light and strong ambient light, and performing dual photography at live video rates.

Finally, Chapter 6 introduced a new paradigm for light transport analysis that exploits the finite speed of light. We jointly considered the spatial constraints on light paths and the temporal constraints on the transient time delays they induce. We showed that the regular light transport equation also models transient transport, provided that light sources emit a sinusoidally modulated signal at a given frequency. This made it possible to analyze transient light transport by trivially adapting standard analysis techniques to the transient domain. Specifically, we implemented a transient version of optical probing to capture the direct, diffuse indirect, and specular indirect components of a time-of-flight image. We used these images to reconstruct time-of-flight depth images robust to indirect transport, distinguish between direct views of objects and their mirror reflections, and also capture sharp evolving wavefronts of light propagating through a scene at sub-nanosecond time resolutions.

7.1 Limitations

While this thesis demonstrates several state of the art solutions to well-established problems in computer vision and graphics, optical linear algebra has both theoretical and practical drawbacks.

Our mathematical framework is limited in three respects. First, optical calculations are significantly less accurate than their numerical counterparts. Following from Eq. (1.1) and Eq. (1.3), computing a matrix-vector product optically is impacted by several physical factors: dynamic range, ambient illumination, and sensor noise. Second, because of the restricted access to the elements of \mathbf{T} , few numerical algorithms are found to be well-suited for the optical domain. Third, the impact of noise on our optical algorithms remains unclear. Specifically, there are open questions regarding the effect of sensor noise on the convergence rate of our optical Krylov subspace algorithms.

On the practical side, moving the onus of computation to the optical domain significantly increases hardware complexity. Our prototype systems comprise of optics (beamsplitters, relay lenses, diffraction gratings), programmable optical elements (DMDs, LCDs, MEMS), opto-mechanical components, and custom electronics. These components can be expensive, fragile, bulky, inflexible, and challenging to configure. Moreover, additional optical components and hardware imperfections can reduce image quality (*e.g.* lens aberrations lead to blurring of the image) and light efficiency (*e.g.* LCDs block at least 50% of the light). However, despite all these supposed limitations, we have addressed several longstanding problems using off-the-shelf parts; new hardware, such as rolling shutters that permit electronic masking, promises to make hardware designs for optical computing a lot less complex.

7.2 Future Research

This work represents only the first step in establishing a complete mathematical framework for active illumination, and there are many exciting directions left unexplored for optical linear algebra.

Although the light paths we consider are restricted to projector-camera systems, the theory itself is not restricted to any specific arrangement of lights and sensors, or even the size of the scene. The configuration space for light sources can range from a single point light source (an LED diode) to an array of projectors (a light-field projector); the space for sensors can range from a single sensor element to an array of cameras (a light-field camera).

Beyond the spatial and temporal constraints on light paths, there are several other imaging paradigms available for light transport analysis, including those based on the wavelength (*i.e.* color) and polarization. The wavelength and polarization state of light provide important radiometric cues for analyzing scenes. Fluorescent objects absorb light at one wavelength, and re-emit light at a different (larger) wavelength; this transport of radiant energy from one wavelength to another is represented by yet another extended version of the transport matrix [58]. The polarization state of radiant energy is characterized by a Stokes vector of length 4, and the transport of radiant energy from one polarization state to another is represented by a 4×4 matrix known as the Mueller matrix [13]. These additional radiometric constraints on light paths drastically increase the dimensionality of light transport, emphasizing once again the need for efficient optical computations.

There is, however, one physical characteristic incompatible with the current definition of the real-valued light transport matrix: the wavelike nature of light. In classical physics, radiant energy is the energy of electromagnetic waves (rather than photons), and two electromagnetic waves may interfere either constructively or destructively. In optics, the complex-valued *transmission matrix* maps monochromatic electromagnetic waves from source to sensor [95, 102], similar to our transient frequency light transport matrix. These transmission matrices have already found many exciting optical applications previously considered physically impossible, including the ability to focus light or transmit images through scattering materials [103, 127]. Using optical computing principles to analyze these transmission matrices is another exciting direction for this work.

Appendices

Appendix A

Extended Discussion

A.1 One-shot optical probing versus mask-less multi-image acquisition

We consider the following hypothetical procedures:

- *PDC*: Capture one optimally-exposed photo with 512 primal and dual codes and a 30-second exposure time at ISO 50 (*i.e.*, slightly more time than needed for a 20 Hz display rate).
- *MI*: Capture 512 photos without a mask, each exposed for 30/512 seconds at ISO 50 or ISO 1600 (*i.e.*, same total exposure time as PDC and therefore the same total photoelectrons).
- *PDC*×10 and *MI*×10: Reduce exposure time by a factor of 10 and use ISO 1600.

These procedures represent the idealized cases of a conventional high-speed system that captures and transfers photos with no overhead and of a spatiotemporally-aligned optical probing system with 100% mask transmissivity (representative of a DMD mask, and about twice the transmissivity of our LCD panel). We compare their SNR for a specific camera—the Canon EOS 1D Mark II—whose noise characteristics have been studied in detail [12].

The SNR of digital sensors (Eq. (1.8)) can be expressed as

$$\text{SNR}(\hat{\mathbf{i}}) = 20 \log_{10} \left[\frac{\Psi \mathbf{i} / g}{\sqrt{\sigma^2 + \Psi \mathbf{i} / g^2}} \right]$$

where we ignore the (negligible) effect of quantization for simplicity. The value $\Psi \mathbf{i}[m]$ is the number of photoelectrons received by pixel m ; gain g is measured in photoelectrons per digital number and depends on the camera’s ISO setting; and the terms in the denominator represent the variance of read noise and photon noise, respectively. According to Clark [12], the gain at ISO 50 is $g = 26.03$ photoelectrons per digital number, and the standard deviation of the read-noise component is $\sigma = \frac{30.62}{g} = 1.18$ digital numbers.

To calculate the SNR for the PDC procedure, we assume that the mean number of photoelectrons of the optimally-exposed optically-probed photo is 10,387 photoelectrons. This corresponds to 13% of the full-well capacity $i_{\max} = 79,900$ photoelectrons (for ISO 50) and is in accordance with general auto-exposure criteria [14]. This yields an SNR equal to 39.79 dB (against an additive floor of 1.18).

Since a sensor pixel is masked by approximately half the primal and dual codes on average, the photoelectrons in a PDC photo are due to $\frac{512}{2} = 256$ of those codes. This means that these photoelectrons are distributed across 256 photos in the MI procedure, or 40.57 photoelectrons per photo. Thus, the MI procedure is dominated by additive noise and has an extremely low SNR, just 2.26 dB (against the additive floor of 1.18). This makes the captured photos practically unusable on their own. Even at ISO 1600, where $g = 0.81$ photoelectrons per digital number and read noise is $\sigma = \frac{3.9}{g} = 4.81$, the SNR is just 14.70 dB.

We now examine the SNR of their masked sum according to Eq. (3.4). Since only 256 photos contribute to this sum on average, we must count the contribution of additive noise only from those photos. The standard deviation of read noise will therefore be $\sqrt{256} = 16$ times larger, or $\sigma = 18.88$ digital numbers at ISO 50 and $\sigma = 76.96$ digital numbers at ISO 1600. This yields an SNR of 26.34 dB at ISO 50 and 38.78 dB at ISO 1600. Thus, despite the significant computational burden of acquiring, transferring and storing 512 individual photos, these photos are not very useful on their own and do not offer an SNR gain over a well-engineered optical probing system.

Finally, applying PDC \times 10 and MI \times 10 means that ten times fewer photoelectrons contribute to a pixel. While the resulting under-exposed photo has an SNR of 30.10 dB under PDC \times 10, MI \times 10's reduction to 23.40 dB for the masked sum is much more severe.

Appendix B

Proofs and Derivations

B.1 Expanded derivations

B.1.1 Derivation of Direct-Enhanced and Indirect-Only Probing Matrices

Suppose $(\mathbf{p}_k)_+$ and $(\mathbf{p}_k)_-$ refer to the positive and negative components of a Rademacher vector \mathbf{p}_k , where $\mathbf{p}_k = (\mathbf{p}_k)_+ - (\mathbf{p}_k)_-$. Note that the codes $(\mathbf{p}_k)_+$ and $(\mathbf{p}_k)_-$ have a Bernoulli distribution with success probability $p = 0.5$, where the elements have value 1 with probability p and value 0 with probability $1 - p$. The following derives from Eq. (3.12):

$$\mathbf{p}_k(\mathbf{p}_k)^\top = \underbrace{[(\mathbf{p}_k)_+((\mathbf{p}_k)_+)^\top + (\mathbf{p}_k)_-((\mathbf{p}_k)_-)^\top]}_{\text{direct-enhancing term}} - \underbrace{[(\mathbf{p}_k)_-((\mathbf{p}_k)_+)^\top + (\mathbf{p}_k)_+((\mathbf{p}_k)_-)^\top]}_{\text{indirect-only term}} \quad (\text{B.1})$$

The *direct-enhancing term* converges to a probing matrix containing the value 1 for diagonal elements, and the value 0.5 for off-diagonal elements. The *indirect-only term* converges to a matrix with 0 for diagonal elements and 0.5 for off-diagonal elements. As expected, subtracting the indirect-only probing matrix from the direct-enhanced matrix produces the identity matrix.

Random vectors \mathbf{x}_k sampled from the Bernoulli distribution form the direct-enhanced matrix in the limit:

$$\mathbf{\Pi} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^\top \quad (\text{B.2})$$

The diagonal entries $\mathbf{\Pi}[m, m]$ of the probing matrix have the following expected value:

$$\begin{aligned} \mathbf{\Pi}[m, m] &= \text{E} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k[m] \mathbf{x}_k[m] \right] \\ &= \frac{1}{K} \sum_{k=1}^K \text{E} [(\mathbf{x}_k[m])^2] = p \end{aligned} \quad (\text{B.3})$$

A similar derivation produces the expected value of the off-diagonal entries $\mathbf{\Pi}[m, n]$ where $m \neq n$:

$$\begin{aligned} \mathbf{\Pi}[m, n] &= \mathbb{E} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k[m] \mathbf{x}_k[n] \right] \\ &= \frac{1}{K} \sum_{k=1}^K \mathbb{E} [\mathbf{x}_k[m] \mathbf{x}_k[n]] \\ &= \frac{1}{K} \sum_{k=1}^K \mathbb{E} [\mathbf{x}_k[m]] \mathbb{E} [\mathbf{x}_k[n]] = p^2 \end{aligned} \quad (\text{B.4})$$

In the limit, the probing matrix has value p on the diagonal, and p^2 for off-diagonal elements. As the value p becomes smaller, the diagonal terms become larger relative to the off-diagonal terms.

The same set of vectors \mathbf{x}^k , with a minor modification, forms the indirect-only probing matrix:

$$\mathbf{\Pi} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k (\mathbf{1} - \mathbf{x}_k)^T \quad (\text{B.5})$$

The diagonal entries $\mathbf{\Pi}[m, m]$ of the probing matrix converge to the following:

$$\begin{aligned} \mathbf{\Pi}[m, m] &= \mathbb{E} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k[m] (1 - \mathbf{x}_k[m]) \right] \\ &= \frac{1}{K} \sum_{k=1}^K \mathbb{E} [\mathbf{x}_k[m] (1 - \mathbf{x}_k[m])] \\ &= \frac{1}{K} \sum_{k=1}^K (\mathbb{E} [\mathbf{x}_k[m]] - \mathbb{E} [(\mathbf{x}_k[m])^2]) = 0 \end{aligned} \quad (\text{B.6})$$

Once again, we derive the off-diagonal entries $\mathbf{\Pi}[m, n]$ where $m \neq n$ for the indirect-only probing matrix:

$$\begin{aligned} \mathbf{\Pi}[m, n] &= \mathbb{E} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k[m] (1 - \mathbf{x}_k[n]) \right] \\ &= \frac{1}{K} \sum_{k=1}^K \mathbb{E} [\mathbf{x}_k[m] (1 - \mathbf{x}_k[n])] \\ &= \frac{1}{K} \sum_{k=1}^K (\mathbb{E} [\mathbf{x}_k[m]] - \mathbb{E} [\mathbf{x}_k[m]] \mathbb{E} [\mathbf{x}_k[n]]) = p(1 - p) \end{aligned}$$

The off-diagonal term is maximum when $p = 0.5$. Note that, unlike the direct-enhanced case, there is no benefit in setting p to any other value.

When computing the direct-enhanced and indirect-only probing matrices, we generate the illumination and mask codes such that their sum produces a uniform image, as in the case of the direct-enhancing and indirect-only terms of Eq. (B.1). This results in a set of mask codes that uniformly exposes each sensor pixel.

B.1.2 Derivation of Eq. (3.9)

The stochastic diagonal estimator converges to the diagonal matrix when using independent and identically distributed random vectors sampled from a distribution with mean 0 and variance 1. We derive the variance of pixel $\mathbf{i}[m]$ for a K -term estimate of the diagonal element $\mathbf{T}[m, m]$ as follows:

$$\begin{aligned}
\text{Var}(\mathbf{i}[m]) &= \text{Var}\left(\frac{1}{K} \sum_{k=1}^K \sum_{n=1}^P \mathbf{p}_k[m] \mathbf{p}_k[n] \mathbf{T}[m, n]\right) \\
&= \frac{1}{K^2} \sum_{k=1}^K \text{Var}\left(\sum_{n=1}^P \mathbf{p}_k[m] \mathbf{p}_k[n] \mathbf{T}[m, n]\right) \\
&= \frac{1}{K} \text{Var}\left(\sum_{n=1}^P \mathbf{p}[m] \mathbf{p}[n] \mathbf{T}[m, n]\right) \\
&= \frac{1}{K} \left(\mathbb{E} \left[\left(\sum_{n=1}^P \mathbf{p}[m] \mathbf{p}[n] \mathbf{T}[m, n] \right)^2 \right] - \mathbb{E} \left[\sum_{n=1}^P \mathbf{p}[m] \mathbf{p}[n] \mathbf{T}[m, n] \right]^2 \right) \\
&= \frac{1}{K} \left(\mathbb{E} \left[\sum_{n=1}^P \mathbf{p}[m]^2 \mathbf{p}[n]^2 \mathbf{T}[m, n]^2 \right] - \mathbb{E} \left[\mathbf{p}[m]^2 \mathbf{T}[m, m] \right]^2 \right) \\
&= \frac{1}{K} \left(\mathbb{E} \left[\mathbf{p}[m]^4 \mathbf{T}[m, m]^2 \right] + \mathbb{E} \left[\sum_{\substack{n=1 \\ n \neq m}}^P \mathbf{p}[m]^2 \mathbf{p}[n]^2 \mathbf{T}[m, n]^2 \right] - \mathbb{E} \left[\mathbf{p}[m]^2 \mathbf{T}[m, m] \right]^2 \right) \\
&= \frac{1}{K} \left(\mathbb{E} \left[\mathbf{p}[m]^4 \right] \mathbf{T}[m, m]^2 + \sum_{\substack{n=1 \\ n \neq m}}^P \mathbf{T}[m, n]^2 - \mathbf{T}[m, m]^2 \right) \tag{B.7}
\end{aligned}$$

The Rademacher sequence (where elements have a 50% chance of being either 1 or -1) is the optimal distribution for the stochastic diagonal estimator, in the sense that the sequence minimizes the variance Eq. (B.7). This is the case because $\mathbb{E}[\mathbf{p}[m]^4] \geq \mathbb{E}[\mathbf{p}[m]^2]^2 = 1$ for all distributions with variance 1, and $\mathbb{E}[\mathbf{p}[m]^4] = 1$ for the Rademacher distribution. The variance of the stochastic diagonal estimator using the Rademacher sequence is the following:

$$\begin{aligned}
\text{Var}(\mathbf{i}[m]) &= \frac{1}{K} \left(\mathbb{E} \left[\mathbf{p}[m]^4 \right] \mathbf{T}[m, m]^2 + \sum_{\substack{n=1 \\ n \neq m}}^P \mathbf{T}[m, n]^2 - \mathbf{T}[m, m]^2 \right) \\
&= \frac{1}{K} \sum_{\substack{n=1 \\ n \neq m}}^P \mathbf{T}[m, n]^2 \tag{B.8}
\end{aligned}$$

According to Eq. (B.8), the number of codes K must increase by a factor of 2 to decrease the variance by a factor of 2. Note that the variance term only depends on the off-diagonal matrix entries; when the off-diagonal elements are zero, the stochastic diagonal estimator returns the exact value of element $\mathbf{T}[m, m]$ for any $K \geq 1$.

B.1.3 Derivation of Eq. (4.8)

Combining Eqs. (3.4) and (4.4) we have

$$\mathbf{i}_e = \frac{1}{K} \sum_{k=1}^K \sum_{f=1}^E \overline{\mathbf{q}_{ek}} \circ [\mathbf{T}_{ef} \mathbf{q}_{fk}] \quad (\text{B.9})$$

where the $1/K$ factor captures the fact that each term in the sum is allocated $1/K$ of the total exposure time. We now split the sum into its epipolar and non-epipolar terms

$$\mathbf{i}_e = \frac{1}{K} \sum_{k=1}^K \left\{ \overline{\mathbf{q}_{ek}} \circ [\mathbf{T}_{ee} \mathbf{q}_{ek}] + \sum_{\substack{f=1 \\ f \neq e}}^E \overline{\mathbf{q}_{ek}} \circ [\mathbf{T}_{ef} \mathbf{q}_{fk}] \right\} \quad (\text{B.10})$$

and observe that the first term is always a vector of zeros. So,

$$\mathbf{i}_e = \frac{1}{K} \sum_{k=1}^K \sum_{\substack{f=1 \\ f \neq e}}^E \overline{\mathbf{q}_{ek}} \circ [\mathbf{T}_{ef} \mathbf{q}_{fk}] . \quad (\text{B.11})$$

Letting $K \rightarrow \infty$ and applying the Central Limit Theorem to Eq. (B.11) we get the expected image $\mathbb{E}[\mathbf{i}_e]$ for epipolar line e :

$$\begin{aligned} \mathbb{E}[\mathbf{i}_e] &= \mathbb{E} \left[\sum_{\substack{f=1 \\ f \neq e}}^E \overline{\mathbf{q}_e} \circ [\mathbf{T}_{ef} \mathbf{q}_f] \right] \\ &= \mathbb{E}[\overline{\mathbf{q}_e}] \circ \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \mathbb{E}[\mathbf{q}_f] \\ &= 0.25 \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \mathbf{1} , \end{aligned} \quad (\text{B.12})$$

where Eq. (B.12) follows from the fact that epipolar lines e and f are distinct and thus their corresponding random vectors \mathbf{q}_e and \mathbf{q}_f are independent.

B.1.4 Derivation of Eq. (4.13)

Combining Eqs. (3.4) and (4.4) for the indirect-invariant mask and pattern we have:

$$\mathbf{i}_e = \frac{1}{K} \sum_{k=1}^K \sum_{f=1}^E \mathbf{m}_{ek} \circ \left\{ \mathbf{T}_{ef} [\mathbf{m}_{fk} \circ \mathbf{r}_{fk} + \overline{\mathbf{m}_{fk}} \circ \overline{\mathbf{r}_{fk}}] \right\} \quad (\text{B.13})$$

We split the sum into its epipolar and non-epipolar terms,

$$\mathbf{i}_e = \frac{1}{K} \sum_{k=1}^K \left\{ \mathbf{m}_{ek} \circ \mathbf{T}_{ee} [\mathbf{m}_{ek} \circ \mathbf{r}_{ek}] + \mathbf{m}_{ek} \circ \mathbf{T}_{ee} [\overline{\mathbf{m}_{ek}} \circ \overline{\mathbf{r}_{ek}}] + \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{m}_{ek} \circ \mathbf{T}_{ef} [\mathbf{m}_{fk} \circ \mathbf{r}_{fk}] + \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{m}_{ek} \circ \mathbf{T}_{ef} [\overline{\mathbf{m}_{fk}} \circ \overline{\mathbf{r}_{fk}}] \right\} \quad (\text{B.14})$$

and note that the second term of Eq. (B.14) is always a vector of zeros. Letting $K \rightarrow \infty$ and applying the Central Limit Theorem to Eq. (B.14) we get the expected image for epipolar line e :

$$\mathbb{E}[\mathbf{i}_e] = \mathbb{E} \left[\mathbf{q}_e \circ [\mathbf{T}_{ee} (\mathbf{q}_e \circ \mathbf{r}_e)] + \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{q}_e \circ [\mathbf{T}_{ef} (\mathbf{q}_f \circ \mathbf{r}_f + \overline{\mathbf{q}_f} \circ \overline{\mathbf{r}_f})] \right] . \quad (\text{B.15})$$

Now, \mathbf{q}_e is a random binary vector whose probability of being either $\mathbf{1}$ or $\mathbf{0}$ is 0.5. Using this fact as well as \mathbf{q}_e 's independence from all other random vectors, the expectation in Eq. (B.15) becomes

$$\mathbb{E}[\mathbf{i}_e] = 0.5 \mathbf{T}_{ee} \mathbb{E}[\mathbf{r}_e] + 0.5 \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \mathbb{E}[\mathbf{q}_f \circ \mathbf{r}_f + \overline{\mathbf{q}_f} \circ \overline{\mathbf{r}_f}] . \quad (\text{B.16})$$

Finally, using the definition of binary random vector \mathbf{r}_f in Eq. (4.11) the expectation becomes

$$\begin{aligned} \mathbb{E}[\mathbf{i}_e] &= 0.5 \mathbf{T}_{ee} \mathbf{p}_e + 0.5 \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \{ \Pr(\mathbf{q}_f = \mathbf{1}) \mathbf{p}_e + \Pr(\mathbf{q}_f = \mathbf{0}) (\mathbf{1} - \mathbf{p}_e) \} \\ &= 0.5 \mathbf{T}_{ee} \mathbf{p}_e + 0.25 \sum_{\substack{f=1 \\ f \neq e}}^E \mathbf{T}_{ef} \mathbf{1} . \end{aligned} \quad (\text{B.17})$$

B.1.5 Derivation of Eq. (5.15)

The homogeneous factorization equation maximizes energy efficiency subject to all imaging constraints, by maximizing the scalar γ or, equivalently, minimizing its reciprocal γ^{-1} :

$$\begin{aligned} &\min_{\gamma, t, \mathbf{M}, \mathbf{P}} \quad \gamma^{-1} && (\text{B.18}) \\ &\text{subject to} \quad 0 \leq \mathbf{m}_k, \quad \|\mathbf{m}_k\|_\infty \leq 1 \\ & \quad \quad \quad 0 \leq \mathbf{p}_k, \quad \|\mathbf{p}_k\|_{\dagger \ell} \leq t_k \\ & \quad \quad \quad 0 \leq t_k, \quad \sum_{k=1}^K t_k \leq T \\ & \quad \quad \quad \gamma \mathbf{\Pi} = \mathbf{M} \mathbf{P}^T \end{aligned}$$

This is a difficult problem to solve in general, given the nonlinear nature of the equality constraint. Instead, we reformulate and then relax this problem into a simpler form.

Reformulation There is a bilinear relation between the masks \mathbf{m}_k and illumination patterns \mathbf{p}_k . Substituting the mask and illumination variables with $a_k \mathbf{p}_k$ and $a_k^{-1} \mathbf{m}_k$ satisfies the equality constraint for all scalars $a_k > 0$. Eq. (B.18) can be rewritten as follows:

$$\begin{aligned}
& \min_{\gamma, t, \mathbf{M}, \mathbf{P}} \quad \gamma^{-1} & (B.19) \\
& \text{subject to} \quad 0 \leq \mathbf{m}_k, \quad a_k^{-1} \|\mathbf{m}_k\|_{\infty} \leq 1 \\
& \quad \quad \quad 0 \leq \mathbf{p}_k, \quad a_k \|\mathbf{p}_k\|_{\dagger \ell} \leq t_k \\
& \quad \quad \quad 0 \leq t_k, \quad \sum_{k=1}^K t_k \leq T \\
& \quad \quad \quad \gamma \mathbf{\Pi} = \mathbf{M} \mathbf{P}^T
\end{aligned}$$

The formulation can be relaxed by combining the inequality constraints on \mathbf{m}_k and \mathbf{p}_k :

$$\begin{aligned}
& \min_{\gamma, t, \mathbf{M}, \mathbf{P}} \quad \gamma^{-1} & (B.20) \\
& \text{subject to} \quad 0 \leq \mathbf{m}_k, 0 \leq \mathbf{p}_k \\
& \quad \quad \quad \|\mathbf{m}_k\|_{\infty} \|\mathbf{p}_k\|_{\dagger \ell} \leq t_k \\
& \quad \quad \quad 0 \leq t_k, \quad \sum_{k=1}^K t_k \leq T \\
& \quad \quad \quad \gamma \mathbf{\Pi} = \mathbf{M} \mathbf{P}^T
\end{aligned}$$

Similarly, we can eliminate the time variable t_k from the formulation by relaxing the inequality constraints, as follows:

$$\begin{aligned}
& \min_{\gamma, \mathbf{M}, \mathbf{P}} \quad \gamma^{-1} & (B.21) \\
& \text{subject to} \quad 0 \leq \mathbf{m}_k, 0 \leq \mathbf{p}_k \\
& \quad \quad \quad \frac{1}{T} \sum_{k=1}^K \|\mathbf{m}_k\|_{\infty} \|\mathbf{p}_k\|_{\dagger \ell} \leq 1 \\
& \quad \quad \quad \gamma \mathbf{\Pi} = \mathbf{M} \mathbf{P}^T
\end{aligned}$$

By replacing the mask with $\gamma^{-1} \mathbf{m}_k$, the formulation reduces to the following form:

$$\begin{aligned}
& \min_{\gamma, \mathbf{M}, \mathbf{P}} \quad \gamma^{-1} & (B.22) \\
& \text{subject to} \quad 0 \leq \mathbf{m}_k, 0 \leq \mathbf{p}_k \\
& \quad \quad \quad \frac{1}{T} \sum_{k=1}^K \|\mathbf{m}_k\|_{\infty} \|\mathbf{p}_k\|_{\dagger \ell} \leq \gamma^{-1} \\
& \quad \quad \quad \mathbf{\Pi} = \mathbf{M} \mathbf{P}^T
\end{aligned}$$

Note that the solution of Eq. (B.22) is closely related to that of Eq. (B.18). Suppose that γ' , \mathbf{m}'_k , and \mathbf{p}'_k minimize Eq. (B.22). The solution to Eq. (B.18) is then given by the following:

$$\gamma = \gamma' \quad (\text{B.23})$$

$$t_k = \|\mathbf{m}'_k\|_\infty \|\mathbf{p}'_k\|_{\dagger\ell} \quad (\text{B.24})$$

$$\mathbf{m}_k = \frac{\mathbf{m}'_k}{\|\mathbf{m}'_k\|_\infty} \quad (\text{B.25})$$

$$\mathbf{p}_k = t_k \frac{\mathbf{p}'_k}{\|\mathbf{p}'_k\|_{\dagger\ell}} \quad (\text{B.26})$$

Relaxation The final step is to relax the equality constraint through Lagrangian relaxation. First, we rewrite Eq. (B.22) equivalently in the following form:

$$\begin{aligned} \min_{\mathbf{M}, \mathbf{P}} \quad & \frac{1}{T} \sum_{k=1}^K \|\mathbf{m}_k\|_\infty \|\mathbf{p}_k\|_{\dagger\ell} \quad (\text{B.27}) \\ \text{subject to} \quad & 0 \leq \mathbf{m}_k, 0 \leq \mathbf{p}_k \\ & \|\mathbf{\Pi} - \mathbf{M}\mathbf{P}^T\|_F^2 \leq 0 \end{aligned}$$

The inequality with respect to the homogeneous factorization can be brought into the objective using a Lagrange multiplier, as follows:

$$\begin{aligned} \min_{\mathbf{M}, \mathbf{P}} \quad & \|\mathbf{\Pi} - \mathbf{M}\mathbf{P}^T\|_F^2 + \lambda \sum_{k=1}^K \|\mathbf{m}_k\|_\infty \|\mathbf{p}_k\|_{\dagger\ell} \quad (\text{B.28}) \\ \text{subject to} \quad & 0 \leq \mathbf{m}_k, 0 \leq \mathbf{p}_k \end{aligned}$$

for some scalar $\lambda > 0$. When λ is sufficiently small, the solution to Eq. (B.28) is a close approximation to the solution of the original problem.

B.2 Proofs of Propositions 1 and 2

B.2.1 Proof of Proposition 1

Proposition 1. *If \mathbf{T}^{NE} and \mathbf{T}^{EI} are discretized forms of transport functions that are square-integrable and positive over the rectified projector and image planes, then*

$$\lim_{\epsilon \rightarrow 0} \frac{\mathbf{T}^{\text{EI}} \mathbf{p}}{\mathbf{T}^{\text{NE}} \mathbf{p}} = \mathbf{0} \quad (\text{B.29})$$

where division is entrywise and ϵ is the pixel size for discretization.

Proof sketch. We begin by identifying the rectified projector and image planes with the continuous domain $\mathcal{D} = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$. Let $n = (n_x, n_y)$ be a point on the projector plane and let $\mathbf{I}_\epsilon(n)$ be an indicator function over \mathcal{D} that specifies the spatial extent of the discrete epipolar line through the origin:

$$\mathbf{I}_\epsilon(p) = \begin{cases} 1 & \text{if } |n_x| \leq \frac{\epsilon}{2} \text{ and } |n_y| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.30})$$

In the continuous setting, light transport from the projector plane to the image plane is described by the *rendering equation* [64]. Given an image point $m \in \mathcal{D}$ on the epipolar line through the origin, this equation describes the total radiant energy transported to m from points on the projector plane:

$$\mathcal{I}(m) = \underbrace{\mathcal{T}(\hat{n}, m) \mathcal{P}(\hat{n})}_{\text{direct}} + \underbrace{\int_{\mathcal{D}-\{\hat{n}\}} \mathcal{T}(n, m) \mathcal{P}(n) \, dn}_{\text{indirect}} \quad (\text{B.31})$$

where \hat{n} is the projector point in stereo correspondence with image point m ; $\mathcal{P}(n)$ is the radiant energy along the ray through projector point n ; and $\mathcal{T}(n, m)$ is the *transport function* describing the proportion of radiant energy from n that gets transported to m .

Without loss of generality, we prove the continuous form of the ratio in Eq. (B.29) for an image point i ; this point is taken to be inside a discrete image pixel of dimension $\epsilon \times \epsilon$ on the epipolar line through the origin.

More specifically, we consider the epipolar indirect, total indirect, and non-epipolar indirect contributions at m :

$$\mathcal{I}^{\text{EI}}(m) = \int_{\mathcal{D}-\{\hat{n}\}} \mathbf{I}_\epsilon(n) \mathcal{T}(n, m) \mathcal{P}(n) \, dn \quad (\text{B.32})$$

$$\mathcal{I}^{\text{I}}(m) = \int_{\mathcal{D}-\{\hat{n}\}} \mathcal{T}(n, m) \mathcal{P}(n) \, dn \quad (\text{B.33})$$

$$\mathcal{I}^{\text{NE}}(m) = \mathcal{I}^{\text{I}}(m) - \mathcal{I}^{\text{EI}}(m) \quad (\text{B.34})$$

We now show that for any $\delta > 0$, there is an $\epsilon > 0$ such that

$$\left| \frac{\mathcal{I}^{\text{EI}}(m)}{\mathcal{I}^{\text{NE}}(m)} \right| < \delta . \quad (\text{B.35})$$

We use the Cauchy-Schwarz inequality [108], which states that if two functions f and g are square-integrable (*i.e.* $\int |f(x)|^2 dx < \infty$), then

$$\left| \int f(x)g(x) dx \right| \leq \left\{ \int |f(x)|^2 dx \right\}^{\frac{1}{2}} \left\{ \int |g(x)|^2 dx \right\}^{\frac{1}{2}} \quad (\text{B.36})$$

Since $\mathcal{T}()$ is a square-integrable function, we can apply the Cauchy-Schwarz inequality to Eq. (B.32) to get an upper bound on the epipolar indirect contributions:

$$\begin{aligned} \mathcal{I}^{\text{EI}}(m) &\leq \left\{ \int_{\mathcal{D}-\{\hat{n}\}} \mathbf{I}_\epsilon(n) dn \right\}^{\frac{1}{2}} \left\{ \int_{\mathcal{D}-\{\hat{n}\}} [\mathcal{T}(n, m) \mathcal{P}(n)]^2 dn \right\}^{\frac{1}{2}} \\ &= (2\epsilon)^{\frac{1}{2}} \left\{ \int_{\mathcal{D}-\{\hat{n}\}} [\mathcal{T}(n, m) \mathcal{P}(n)]^2 dn \right\}^{\frac{1}{2}} . \end{aligned} \quad (\text{B.37})$$

By combining Eqs. (B.33), (B.34) and (B.37) we also get a lower bound on the non-epipolar contributions:

$$\mathcal{I}^{\text{NE}}(m) \geq \int_{\mathcal{D}-\{\hat{n}\}} \mathcal{T}(n, m) \mathcal{P}(n) dn - (2\epsilon)^{\frac{1}{2}} \left\{ \int_{\mathcal{D}-\{\hat{n}\}} [\mathcal{T}(n, m) \mathcal{P}(n)]^2 dn \right\}^{\frac{1}{2}} . \quad (\text{B.38})$$

Eq. (B.35) now follows by choosing ϵ to be

$$\epsilon = \frac{1}{2} \left(\frac{\delta}{2 + \delta} \right)^2 \frac{\left\{ \int_{\mathcal{D}-\{\hat{n}\}} \mathcal{T}(n, m) \mathcal{P}(n) dn \right\}^2}{\int_{\mathcal{D}-\{\hat{n}\}} [\mathcal{T}(n, m) \mathcal{P}(n)]^2 dn} . \quad (\text{B.39})$$

Substituting Eq. (B.39) into Eqs. (B.37) and (B.38) we get

$$\left| \frac{\mathcal{I}^{\text{EI}}(m)}{\mathcal{I}^{\text{NE}}(m)} \right| \leq \frac{\frac{\delta}{2+\delta}}{1 - \frac{\delta}{2+\delta}} = \frac{\delta}{2} < \delta . \quad (\text{B.40})$$

□

B.2.2 Proof Sketch of Proposition 2

We prove Proposition 2 for generic scenes consisting of a finite collection of objects, each of which is an open set in \mathbb{R}^3 bounded by a smooth generic surface [39, 70].

Proposition 2. *In a generic scene, a specular transport path does not intersect any of the k -bounce specular transport paths that originate from the corresponding epipolar line for $k \geq 1$.*

Proof. For simplicity, we reverse the direction of light travel through image pixels, treating the camera as a second projector that also sends light onto the scene. These image pixels generate j -bounce specular transport paths, which undergo j specular bounces before reaching a non-specular point within the scene. For example, a specular transport path with no specular bounces ($j = 0$) is a ray directly connecting an image pixel to a non-specular point in the scene.

Without loss of generality, let m be a point on the camera’s image plane that generates a specular transport path with $j \geq 0$ specular bounces. Also, let \mathcal{L} be the corresponding epipolar line on the (continuous) projector plane, and $n \in \mathcal{L}$ be a point on this epipolar line. We assume this point n generates a specular transport path with $k \geq 1$ bounces.

To prove the proposition, we show that the following cannot hold simultaneously:

1. the transport paths through m and n intersect at their $(j + 1)^{\text{th}}$ and $(k + 1)^{\text{th}}$ bounce respectively, *i.e.*, the paths meet at a common non-specular surface point within the scene; and
2. this intersection is generic, *i.e.*, it occurs for some point n in \mathcal{L} for (nearly) all arbitrarily small perturbations of the scene.

In particular, let $l_j(m)$ be the 3D ray that light follows after j specular bounces from point m . Similarly, let $l'_k(n)$ be the corresponding 3D ray for point n . Note that, for specular reflections, the 3D rays $l_j(m)$ and $l'_k(n)$ only coincide (*i.e.* they share the same origin and direction) if points m and n represent the same point in 3D space (*i.e.* the camera and projector are in a coaxial configuration).

Since these transport paths bounce off smooth open surfaces, the mapping $n \mapsto l'_k(n)$ is a smooth function for some open subset $\mathcal{Q} \subset \mathcal{L}$ where $n \in \mathcal{Q}$. This mapping defines a ruled surface in \mathbb{R}^3 ; intuitively, as point n ranges over \mathcal{Q} , the 3D ray $l'_k(n)$ twists and translates in space, tracing a ruled surface. Because the origins of 3D rays $l_j(m)$ and $l'_k(n)$ cannot coincide, we can freely perturb the ruled surface $l'_k(\mathcal{Q})$ by smoothly perturbing the surface normal at the origin of each ray $l'_k(n)$.

Now, for transport paths to meet at a non-specular point, two surfaces and a ray must intersect at a point: the ray $l_j(m)$, the ruled surface $l'_k(\mathcal{Q})$, and a surface in the scene. This, however, is *not* a generic condition because the intersection does not hold for almost all arbitrarily small perturbations of the ruled surface $l'_k(\mathcal{Q})$.

□

B.3 Fourier Transform and Convolution for Matrix-valued Functions

We start by defining a Fourier Transform for matrix-valued functions

$$\mathbf{T}^\omega = \mathcal{F}\{\tilde{\mathbf{T}}\}(\omega) = \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) e^{-2\pi i \omega \tau} d\tau \quad (\text{B.41})$$

and an analogous Fourier transform for vector valued functions

$$\mathbf{p}^\omega = \mathcal{F}\{\tilde{\mathbf{p}}\}(\omega) = \int_{-\infty}^{\infty} \tilde{\mathbf{p}}(t) e^{-2\pi i \omega t} dt \quad (\text{B.42})$$

Using the definition of convolution of a matrix-valued function and a vector valued function from the main paper

$$(\tilde{\mathbf{T}} * \tilde{\mathbf{p}})(t) = \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t - \tau) d\tau \quad (\text{B.43})$$

we can show that the convolution theorem holds:

$$\begin{aligned} \mathcal{F}\{\tilde{\mathbf{T}} * \tilde{\mathbf{p}}\}(\omega) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t - \tau) d\tau \right) e^{-2\pi i \omega t} dt \\ &= \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \left(\int_{-\infty}^{\infty} \tilde{\mathbf{p}}(t - \tau) e^{-2\pi i \omega t} dt \right) d\tau \\ &= \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \left(\int_{-\infty}^{\infty} \tilde{\mathbf{p}}(t') e^{-2\pi i \omega (t' + \tau)} dt' \right) d\tau \\ &= \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \left(\int_{-\infty}^{\infty} \tilde{\mathbf{p}}(t') e^{-2\pi i \omega t'} dt' \right) e^{-2\pi i \omega \tau} d\tau \\ &= \left(\int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) e^{-2\pi i \omega \tau} d\tau \right) \left(\int_{-\infty}^{\infty} \tilde{\mathbf{p}}(t') e^{-2\pi i \omega t'} dt' \right) \\ &= \mathcal{F}\{\tilde{\mathbf{T}}\}(\omega) \mathcal{F}\{\tilde{\mathbf{p}}\}(\omega) \end{aligned} \quad (\text{B.44})$$

Similar derivations can be made for other key properties known from the scalar Fourier transform (e.g. the shift and correlation theorems).

Appendix C

Hardware

To encourage reproducibility, we include the complete parts list for our two experimental systems from Chapter 4 in Table C.1.

#	Part Description	Quantity	Model Name	Company	Low-speed Part	High-speed Part
1	color camera	1	GT1920C	Allied Vision Technologies	✓	
2	monochrome camera	1	GT1920	Allied Vision Technologies		✓
3	connector housing	2	WM1722-ND	Digi-Key Corporation	✓	
4	connector housing	2	WM1728-ND	Digi-Key Corporation		✓
5	power supply	1	T1228-Z12P-ND	Digi-Key Corporation	✓	
6	power supply	1	LC3000-Pro Power Supply	Keynote Photonics	✓	
7	low-speed DMD (projector)	1	LC3000-Pro Pico Projector	Keynote Photonics	✓	
8	low-speed DMD (mask)	1	DLP LightCrafter	Texas Instruments	✓	
9	high-speed DMD	2	DL4130VIS-7XGA	Digital Light Innovations		✓
10	high-power LED light engine	1	High Power S2+ w/ LED	Digital Light Innovations	✓	✓
11	fixed filter holder 40 mm Sq.	1	#54-997	Edmund Optics		✓
12	45 degree mounting adapter	1	#59-001	Edmund Optics		✓
13	crimp	4	WM1142CT-ND	Digi-Key Corporation	✓	✓
14	Hirose contact plug	1	HR1623-ND	Digi-Key Corporation	✓	✓
15	12mm f/1.4 objective lens	1	Cinegon 1.4/12-0906	Schneider Optics	✓	✓
16	visible achromatic doublet pairs	2	MAP10100100-A	Thorlabs	✓	✓
17	300 grooves/mm transmission grating	1	GT25-03	Thorlabs	✓	✓
18	ring-activated threaded iris diaphragm	2	SM1D12D	Thorlabs	✓	✓
19	C-mount to SM1 adapter	1	SM1A9	Thorlabs	✓	✓
20	SM1 to C-mount adapter	1	SM1A10	Thorlabs	✓	✓
21	SM1 Coupler	1	SM1T10	Thorlabs	✓	✓
22	SM1 Lens Tube, 2 inch Thread Depth	1	SM1L20	Thorlabs	✓	✓
23	SM1 Lens Tube, 3 inch Thread Depth	1	SM1L30	Thorlabs	✓	✓
24	SM1-threaded cage plate	1	CP4S	Thorlabs	✓	✓
25	cage plate with 1.2 inch double bore	5	CP12	Thorlabs	✓	✓
26	cage plate with 35 mm aperture	4	CP03/M	Thorlabs	✓	✓
27	cylindrical lens mount	1	CH1A	Thorlabs	✓	✓
28	rod swivel coupler (set of four)	1	C2A	Thorlabs	✓	✓
29	rod end swivel connector (set of four)	1	C3A	Thorlabs	✓	✓
30	cage assembly rod, 2 inch long	2	ER2	Thorlabs	✓	✓
31	cage assembly rod, 3 inch long	4	ER3	Thorlabs	✓	✓
32	cage assembly rod, 4 inch long	4	ER4	Thorlabs	✓	✓
33	aluminum breadboard	1	MB3030/M	Thorlabs	✓	✓
34	12.7 mm x 40 mm optical post	1	TR40/M	Thorlabs	✓	✓
35	12.7 mm x 100 mm optical post	6	TR100/M	Thorlabs	✓	✓
36	post holder, 40 mm	1	PH40/M	Thorlabs	✓	✓
37	post holder, 100 mm	6	PH100/M	Thorlabs	✓	✓
38	studded pedestal base adapter	7	BE1/M	Thorlabs	✓	✓
39	small clamping fork	7	CF125	Thorlabs	✓	✓
40	30 mm single axis translation stage	1	#66-397	Edmund Optics	✓	✓
41	bottom adapter plate	1	#66-620	Edmund Optics	✓	✓
42	top adapter plate	1	#66-493	Edmund Optics	✓	✓
43	metric base plate	1	#54-975	Edmund Optics	✓	✓
44	thread-to-thread adapter	1	#56-323	Edmund Optics	✓	✓

Table C.1: List of parts for the low-speed system shown in Figure 4.8 and the high-speed system shown in Figure 4.9. Both systems use identical optics; the only differences between the two systems are (1) the DMD projector and mask, (2) their mounts, and (3) a color vs. monochrome camera. The camera in both systems outputs live video at a rate of 28 FPS. For the low-speed system, each video frame requires 96 binary masks/projection patterns, *i.e.*, each mask/projection pair is active for 375 μ s of the frame’s 36000 μ s total exposure time. Each video frame of the high-speed system consists of 800 binary mask/projection patterns, *i.e.*, each mask/pattern is active for 45 μ s.

Bibliography

- [1] Nils Abramson. Light-in-flight recording by holography. *Optics Letters*, 3(4):121–123, 1978.
- [2] Supreeth Achar, Stephen T. Nuske, and Srinivasa G. Narasimhan. Compensating for motion during direct-global separation. *Proc. ICCV*, pages 1481–1488, 2013.
- [3] Toshiyuki Amano and Hirokazu Kato. Real world dynamic appearance enhancement with procam feedback. *Proc. PROCAMS*, pages 5:1–5:2, 2008.
- [4] Pierre Ambs. A short history of optical computing: rise, decline, and evolution. *Proc. SPIE*, 7388(1):1–14, 2009.
- [5] Ravindra A. Athale and William C. Collins. Optical matrix-matrix multiplier based on outer product decomposition. *Applied Optics*, 21(12):2089–2090, 1982.
- [6] Francis Bach, Julien Mairal, and Jean Ponce. Convex sparse matrix factorizations. *arXiv:0812.1869*, pages 1–12, 2008.
- [7] Jiamin Bai, Manmohan Chandraker, Tian-Tsong Ng, and Ravi Ramamoorthi. A dual theory of inverse and forward light transport. *Proc. ECCV*, pages 294–307, 2010.
- [8] Constantine Bekas, Effrosyni Kokiopoulou, and Yousef Saad. An estimator for the diagonal of a matrix. *Applied Numerical Mathematics*, 57(11-12):1214–1229, 2007.
- [9] David Casasent. Optical information processing. *Topics in Applied Physics*, pages 181–233, 1981.
- [10] Jack Cederquist and Sing H. Lee. The use of feedback in optical information processing. *Applied Physics A: Materials Science & Processing*, 18(4):311–319, 1979.
- [11] Tongbo Chen, Hans-Peter Seidel, and Hendrik P. A. Lensch. Modulated phase-shifting for 3D scanning. *Proc. CVPR*, pages 1–8, 2008.
- [12] Roger N. Clark. Canon 1D Mark II analysis. <http://www.clarkvision.com/articles/evaluation-1d2/>, 2007.
- [13] Edward Collett. *Field guide to polarization*. SPIE Press, 2005.

- [14] Photography-Cameras-Automatic controls of exposure. *ISO 2721:1982*. International Organization for Standardization, 1982.
- [15] Timothy R. Corle and Gordon S. Kino. *Confocal scanning optical microscopy and related imaging systems*. Academic Press, 1996.
- [16] Oliver Cossairt, Mohit Gupta, and Shree K. Nayar. When does computational imaging improve performance? *IEEE Transactions on Image Processing*, 22(2):447–458, 2012.
- [17] Vincent Couture, Nicolas Martin, and Sébastien Roy. Unstructured light scanning to overcome interreflections. *Proc. ICCV*, pages 1895–1902, 2011.
- [18] Brian Curless and Marc Levoy. Better optical triangulation through spacetime analysis. *Proc. ICCV*, pages 987–994, 1995.
- [19] Gerwin Damberg, Anders Ballestad, Eric Kozak, Raveen Kumaran, and Johannes Minor. Efficient, high brightness, high dynamic range projection. *ACM SIGGRAPH: Emerging Technologies*, pages 18:1–18:1, 2014.
- [20] James Davis, Diego Nehab, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Spacetime stereo: a unifying framework for depth from triangulation. *IEEE TPAMI*, 27(2):296–302, 2005.
- [21] Paul Debevec. The light stages and their applications to photoreal digital actors. *ACM SIGGRAPH Asia: Technical Brief*, pages 1–4, 2012.
- [22] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. *ACM SIGGRAPH*, pages 145–156, 2000.
- [23] Paul Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. *ACM SIGGRAPH*, pages 31:1–31:10, 1997.
- [24] Adrian A. Dorrington, John P. Godbaz, Michael J. Cree, Andrew D. Payne, and Lee V. Streeter. Separating true range measurements from multi-path and scattering interference in commercial range cameras. *Proc. SPIE*, 7864(1):1–10, 2011.
- [25] Michel A. Duguay and Arthur T. Mattick. Ultrahigh speed photography of picosecond light pulses and echoes. *Applied Optics*, 10(9):2162–2170, 1971.
- [26] Christian Fuchs, Michael Heinz, Marc Levoy, Hans-Peter Seidel, and Hendrik P. A. Lensch. Combining confocal imaging and descattering. *Computer Graphics Forum*, 27(4):1245–1253, 2008.
- [27] Martin Fuchs, Volker Blanz, Hendrik Lensch, and Hans-Peter Seidel. Adaptive sampling of reflectance fields. *ACM Transactions on Graphics*, 26(2):1–18, 2007.
- [28] Stefan Fuchs. Multipath interference compensation in time-of-flight camera images. *Proc. ICPR*, pages 3583–3586, 2010.

- [29] Graham Fyffe, Xueming Yu, and Paul Debevec. Single-shot photometric stereo by spectral multiplexing. *Proc. ICCP*, pages 1–6, 2011.
- [30] Gaurav Garg, Eino-Ville Talvala, Marc Levoy, and Hendrik P. A. Lensch. Symmetric photography: Exploiting data-sparseness in reflectance fields. *Proc. EGSR*, pages 251–262, 2006.
- [31] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. Circularly polarized spherical illumination reflectometry. *ACM SIGGRAPH Asia*, 29(6):162:1–162:12, 2010.
- [32] Keisuke Goda, Kevin K. Tsia, and Bahram Jalali. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature*, 458(1):1145–1149, 2009.
- [33] John P. Godbaz, Michael J. Cree, and Adrian A. Dorrington. Closed-form inverses for the mixed pixel/multipath interference problem in AMCW lidar. *Proc. SPIE*, 8296(1):1–15, 2012.
- [34] Guy Godin, Marc Rioux, Jean-Angelo Beraldin, Marc Levoy, Luc Cournoyer, and François Blais. An assessment of laser range measurement on marble surfaces. *Proc. Conference on Optical 3D Measurement Techniques*, pages 1–8, 2001.
- [35] Joseph W. Goodman. Introduction to Fourier optics, 3rd edition. *Roberts & Company Publishers*, 2005.
- [36] Anne Greenbaum and Lloyd N. Trefethen. GMRES/CR and Arnoldi/Lanczos as matrix approximation problems. *SIAM Journal on Scientific Computing*, 15(2):359–368, 1994.
- [37] Michael D. Grossberg, Harish Peri, Shree K. Nayar, and Peter N. Belhumeur. Making one object look like another: controlling appearance using a projector-camera system. *Proc. CVPR*, pages 452–459, 2004.
- [38] Peter S. Guilfoyle and Richard V. Stone. Digital optical computer II. *Proc. SPIE*, 1563(1):214–222, 1991.
- [39] Victor Guillemin and Alan Pollack. *Differential Topology*. Prentice-Hall, 1974.
- [40] Mohit Gupta. *Scene Recovery and Rendering Techniques under Global Light Transport*. PhD thesis, Carnegie Mellon University, 2010.
- [41] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. Structured light 3D scanning in the presence of global illumination. *Proc. CVPR*, pages 713–720, 2011.
- [42] Mohit Gupta and Shree K. Nayar. Micro phase shifting. *Proc. CVPR*, pages 813–820, 2012.
- [43] Mohit Gupta, Shree K. Nayar, Matthias Hullin, and Jaime Martin. Phasor imaging: a generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics*, 34(5):1–18, 2015.
- [44] Mohit Gupta, Qi Yin, and Shree K. Nayar. Structured light in sunlight. *Proc. ICCV*, pages 545–552, 2013.

- [45] Martin H. Gutknecht. Block Krylov space methods for linear systems with multiple right-hand sides: an introduction. *Modern Mathematical Models, Methods and Algorithms for Real World Systems*, pages 420–447, 2007.
- [46] Paul Haeberli. Synthetic lighting for photography. <http://www.graficaobscura.com/synth/>, 1992.
- [47] Benjamin Haeffele, Eric Young, and René Vidal. Structured low-rank matrix factorization: Optimality, algorithm, and applications to image processing. *Proc. ICML*, pages 2007–2015, 2014.
- [48] Richard Hartley and Peter Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.
- [49] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [50] Martin Harwit and Neil J. A. Sloane. *Hadamard Transform Optics*. Academic Press, 1979.
- [51] Samuel W. Hasinoff. Noise-optimal capture for high dynamic range photography. *Proc. CVPR*, pages 553–560, 2010.
- [52] Tim Hawkins, Jonathan Cohen, Chris Tchou, and Paul Debevec. Light stage 2.0. *ACM SIGGRAPH: Technical Sketches*, page 217, 2001.
- [53] Felix Heide, Matthias B. Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM SIGGRAPH*, pages 45:1–45:10, 2013.
- [54] Carlos Hernandez, George Vogiatzis, Gabriel J. Brostow, Bjorn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. *Proc. ICCV*, pages 1–8, 2007.
- [55] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K. Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. *Proc. ICCV*, pages 1–8, 2011.
- [56] Mario Hlawitschka, Julia Ebling, and Gerik Scheuermann. Convolution and Fourier transform of second order tensor fields. *Proc. IASTED VIIP*, pages 78–83, 2004.
- [57] Reynald Hoskinson, Boris Stoeber, Wolfgang Heidrich, and Sidney Fels. Light reallocation for high contrast projection using an analog micromirror array. *ACM Transactions on Graphics*, 29(6):165:1–165:10, 2010.
- [58] Matthias B. Hullin, Johannes Hanika, Boris Ajdin, Hans-Peter Seidel, Jan Kautz, and Hendrik P. A. Lensch. Acquisition and analysis of bispectral bidirectional reflectance and reradiation distribution functions. *ACM Transactions on Graphics*, 29(4):97:1–97:7, 2010.
- [59] Roland N. Ibbett, David Aspinall, and John F. Grainger. Real-time multiplexing of dispersed spectra in any wavelength region. *Applied Optics*, 7(6):1089–1093, 1968.

- [60] Henrik Jensen, Stephen Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. *ACM SIGGRAPH*, pages 511–518, 2001.
- [61] David Jiménez, Daniel Pizarro, Manuel Mazo, and Sira Palazuelos. Modeling and correction of multipath interference in time of flight cameras. *Image and Vision Computing*, 32(1):1–13, 2014.
- [62] John A. Decker Jr. and Martin Harwit. Experimental operation of a Hadamard spectrometer. *Appl. Opt.*, 8(12):2552–2554, 1969.
- [63] 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 Transactions on Graphics*, 32(6):167:1–167:10, 2013.
- [64] James T. Kajiya. The rendering equation. *ACM SIGGRAPH*, pages 143–150, 1986.
- [65] Hiroshi Kawasaki, Ryo Furukawa, Ryusuke Sagawa, and Yasushi Yagi. Dynamic scene shape reconstruction using a single structured light pattern. *Proc. CVPR*, pages 1–8, 2008.
- [66] 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(22):21485–21507, 2011.
- [67] Ahmed Kirmani, Andrea Colaco, Franco N. C. Wong, and Vivek K. Goyal. CoDAC: a compressive depth acquisition camera framework. *Proc. ICASSP*, pages 5425–5428, 2012.
- [68] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using ultrafast transient imaging. *International Journal of Computer Vision*, 95(1):13–28, 2011.
- [69] Ahmed Kirmani, Dheera Venkatraman, Donggeek Shin, Andrea Colaco, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K. Goyal. First-photon imaging. *Science*, 343(6166):58–61, 2013.
- [70] Jan J. Koenderink. *Solid Shape*. Cambridge University Press, 1990.
- [71] Sanjeev J. Koppal and Srinivasa G. Narasimhan. Beyond perspective dual photography with illumination masks. *IEEE Transactions on Image Processing*, 24(7):2083–2097, 2015.
- [72] B. V. K. Vijaya Kumar and David Casasent. Eigenvector determination by iterative optical methods. *Applied Optics*, 20(21):3707–3710, 1981.
- [73] Douglas Lanman, Matthew Hirsch, Yunhee Kim, and Ramesh Raskar. Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. *ACM SIGGRAPH Asia*, pages 163:1–163:10, 2010.
- [74] Rasmus M. Larsen. PROPACK. <http://soi.stanford.edu/~rmunk/PROPACK>, 2010.

- [75] Emmett N. Leith. The evolution of information optics. *IEEE Journal of Selected Topics in Quantum Electronics*, 6(6):1297–1304, 2000.
- [76] Marc Levoy, Billy Chen, Vaibhav Vaish, Mark Horowitz, Ian Mcdowall, and Mark Bolas. Synthetic aperture confocal imaging. *ACM SIGGRAPH*, pages 825–834, 2004.
- [77] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital Michelangelo project: 3D scanning of large statues. *ACM SIGGRAPH*, pages 131–144, 2000.
- [78] Jörg Liesen and Petr Tichy. Convergence analysis of Krylov subspace methods. *GAMM-Mitteilungen*, 27(2):153–173, 2004.
- [79] Dhruv Mahajan, Ira Shlizerman, Ravi Ramamoorthi, and Peter N. Belhumeur. A theory of locally low dimensional light transport. *ACM SIGGRAPH*, pages 62:1–62:10, 2007.
- [80] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta. MC3D: Motion Contrast 3D Scanning. *Proc. ICCP*, pages 147–156, 2015.
- [81] Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, and Addy Ngan. Acquisition and rendering of transparent and refractive objects. *Proc. EGSR*, pages 1–12, 2002.
- [82] Jonathan Mei, Ahmed Kirmani, Andrea Colaço, and Vivek K. Goyal. Phase unwrapping and denoising for time-of-flight imaging using generalized approximate message passing. *Proc. ICIP*, pages 364–368, 2013.
- [83] Christoph Mertz, Sanjeev J. Koppal, Solomon Sia, and Srinivasa G. Narasimhan. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. *Proc. PROCAMS*, pages 15–22, 2012.
- [84] Jerome Mertz. Optical sectioning microscopy with planar or structured illumination. *Nature Methods*, 8(10):811–819, 2011.
- [85] Marvin Minsky. Microscopy apparatus, 1961. US Patent 3,013,467.
- [86] Kaushik Mitra, Oliver Cossairt, and Ashok Veeraraghavan. Can we beat Hadamard multiplexing? Data driven design and analysis for computational imaging systems. *Proc. ICCP*, pages 1–9, 2014.
- [87] Kaushik Mitra, Oliver Cossairt, and Ashok Veeraraghavan. A framework for analysis of computational imaging systems: Role of signal prior, sensor noise and multiplexing. *IEEE TPAMI*, 36(10):1909–1921, 2014.
- [88] Nikhil Naik, Shuang Zhao, Andreas Velten, Ramesh Raskar, and Kavita Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM SIGGRAPH Asia*, pages 171:1–171:10, 2011.

- [89] Shree K. Nayar, Vlad Branzoi, and Terry E. Boult. Programmable imaging using a digital micromirror array. *Proc. CVPR*, pages 436–443, 2004.
- [90] Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM SIGGRAPH*, pages 935–944, 2006.
- [91] Ren Ng, Ravi Ramamoorthi, and Pat Hanrahan. All-frequency shadows using non-linear wavelet lighting approximation. *ACM SIGGRAPH*, pages 376–381, 2003.
- [92] Tian-Tsong Ng, Ramanpreet S. Pahwa, Jiamin Bai, Tony Q. S. Quek, and Kar-han Tan. Radiometric compensation using stratified inverses. *Proc. ICCV*, pages 1889–1894, 2009.
- [93] Charles A. Northend. Lidar, a laser radar for meteorological studies. *Naturwissenschaften*, 54(4):77–80, 1967.
- [94] Matthew O’Toole, John Mather, and Kiriakos N. Kutulakos. Supplementary materials. <http://www.dgp.toronto.edu/~motoole/slt>, 2014.
- [95] Haldun M. Ozaktas, Serdar Yüksel, and M. Alper Kutay. Linear algebraic theory of partial coherence: discrete fields and measures of partial coherence. *Journal of the Optical Society of America*, 19(8):1563–1571, 2002.
- [96] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.
- [97] Pieter Peers and Philip Dutré. Wavelet environment matting. *Proc. Eurographics Symposium on Rendering*, pages 1–11, 2003.
- [98] Pieter Peers and Philip Dutré. Inferring reflectance functions from wavelet noise. *Proc. Eurographics Symposium on Rendering*, pages 1–11, 2005.
- [99] Pieter Peers, Dhruv K. Mahajan, Bruce Lamond, Abhijeet Ghosh, Wojciech Matusik, Ravi Ramamoorthi, and Paul Debevec. Compressive light transport sensing. *ACM Transactions on Graphics*, 28(1):3:1–3:18, 2009.
- [100] Photography—Electronic still-picture cameras—Determination of ISO speed. *ISO 12232:1998*. International Organization for Standardization, 1998.
- [101] Juan Pomarico, Ulf Schnars, Hans J. Hartmann, and Werner Jüptner. Digital recording and numerical reconstruction of holograms: a new method for displaying light in flight. *Applied Optics*, 34(35):8095–8099, 1995.
- [102] Sébastien M. Popoff, Geoffroy Loresey, Rémi Carminati, Mathias Fink, Albert Claude Boccara, and Sylvain Gigan. Measuring the transmission matrix in optics. *Physical Review Letters*, 104(10):1–4, 2010.

- [103] Sébastien M. Popoff, Geoffroy Lerosey, Mathias Fink, Albert Claude Boccara, and Sylvain Gigan. Image transmission through an opaque material. *Nature Communications*, 1(81):1–5, 2010.
- [104] Demetri Psaltis and Ravindra A. Athale. High accuracy computation with linear analog optical systems: a critical study. *Applied Optics*, 25(18):3071–3077, 1986.
- [105] Henri Rajbenbach, Yeshayahu Fainman, and Sing H. Lee. Optical implementation of an interactive algorithm for matrix-inversion. *Applied Optics*, 26(6):1024–1031, 1987.
- [106] Dikpal Reddy, Ravi Ramamoorthi, and Brian Curless. Frequency-space decomposition and acquisition of light transport under spatially varying illumination. *Proc. ECCV*, pages 1–14, 2012.
- [107] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. P2C2: Programmable pixel compressive camera for high speed imaging. *Proc. CVPR*, pages 329–336, 2011.
- [108] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.
- [109] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.
- [110] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, 2010.
- [111] Yoav Y. Schechner, Shree K. Nayar, and Peter N. Belhumeur. Multiplexing for optimal lighting. *IEEE TPAMI*, 29(8):1339–1354, 2007.
- [112] Steven M. Seitz, Y. Matsushita, and Kiriakos N. Kutulakos. A theory of inverse light transport. *Proc. ICCV*, pages 1440–1447, 2005.
- [113] Pradeep Sen. On the relationship between dual photography and classical ghost imaging. *arXiv:1309.3007*, pages 1–6, 2013.
- [114] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen Marschner, Mark Horowitz, Marc Levoy, and Hendrik P. A. Lensch. Dual photography. *ACM SIGGRAPH*, pages 745–755, 2005.
- [115] Pradeep Sen and Soheil Darabi. Compressive dual photography. *Computer Graphics Forum*, 28(2):609–618, 2009.
- [116] Yoshiaki Shirai and Motoi Suwa. Recognition of polyhedrons with a range finder. *Proc. International Joint Conference on Artificial intelligence*, pages 80–87, 1971.
- [117] Horst D. Simon and Hongyuan Zha. Low-rank matrix approximation using the Lanczos bidiagonalization process with applications. *SIAM Journal on Scientific Computing*, 21(6):2257–2274, 2000.
- [118] Valeria Simoncini and Daniel B. Szyld. Theory of inexact krylov subspace methods and applications to scientific computing. *SIAM Journal on Scientific Computing*, 25(2):454–477, 2003.

- [119] Valeria Simoncini and Daniel B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM Journal on Scientific Computing*, 25(2):454–477, 2003.
- [120] Adam Smith, James Skorupski, and James Davis. Transient rendering. Technical Report UCSC-SOE-08-26, School of Engineering, University of California, Santa Cruz, 2008.
- [121] Jok M. Tang and Yousef Saad. A probing method for computing the diagonal of a matrix inverse. *Numerical Linear Algebra with Applications*, 19(3):485–501, 2012.
- [122] Ambler Thompson and Barry N. Taylor. *Guide for the Use of the International System of Units (SI)*. National Institute of Standards and Technology / U.S. Department of Commerce, 2008.
- [123] Lloyd N. Trefethen and David Bau. Numerical linear algebra. *SIAM*, 1997.
- [124] Ewout van den Berg and Michael P. Friedlander. SPGL1: a solver for large-scale reconstruction. <https://www.math.ucdavis.edu/~mpf/spg11>, 2007.
- [125] Eric Veach. *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford University, 1998.
- [126] Ashok Veeraraghavan, Dikpal Reddy, and Ramesh Raskar. Coded strobing photography: compressive sensing of high speed periodic videos. *IEEE TPAMI*, 33(4):671–686, 2011.
- [127] Ivo M. Vellekoop and Allard P. Mosk. Focusing coherent light through opaque strongly scattering media. *Optics Letters*, 32(16):2309–2311, 2007.
- [128] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Mounqi G. Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, pages 1–8, 2012.
- [129] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Mounqi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics*, 32(4):44:1–44:8, 2013.
- [130] Jiaping Wang, Yue Dong, Xin Tong, Zhouchen Lin, and Baining Guo. Kernel Nyström method for light transport. *ACM SIGGRAPH*, pages 29:1–29:10, 2009.
- [131] Oliver Wang, Martin Fuchs, Christian Fuchs, James Davis, Hans-Peter Seidel, and Hendrik P. A. Lensch. A context-aware light source. *Proc. ICCP*, pages 1–8, 2010.
- [132] Gordon Wetzstein and Oliver Bimber. Radiometric compensation through inverse light transport. *Proc. Pacific Graphics*, pages 391–399, 2007.
- [133] Gordon Wetzstein, Wolfgang Heidrich, and David Luebke. Optical image processing using light modulation displays. *Computer Graphics Forum*, 29(6):1934–1944, 2010.

- [134] Tony Wilson, Rimas Juškaitis, Mark A. A. Neil, and Michal Kozubek. Confocal microscopy by aperture correlation. *Optics Letters*, 21(23):1879–1881, 1996.
- [135] Lawrence B. Wolff and Terrance E. Boult. Constraining object features using a polarization reflectance model. *IEEE TPAMI*, 13(7):635–657, 1991.
- [136] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [137] Di Wu, Matthew O’Toole, Andreas Velten, Amit Agrawal, and Ramesh Raskar. Decomposing global light transport using time of flight imaging. *Proc. CVPR*, pages 366–373, 2012.
- [138] 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 of Computer Vision*, pages 1–13, 2013.
- [139] Di Wu, Gordon Wetzstein, Christopher Barsi, Thomas Willwacher, Matthew O’Toole, Nikhil Naik, Qionghai Dai, Kyros Kutulakos, and Ramesh Raskar. Frequency analysis of transient light transport with applications in bare sensor imaging. *Proc. ECCV*, pages 542–555, 2012.
- [140] Yangyang Xu and Wotao Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Scientific Computing*, 6(3):1758–1789, 2013.
- [141] Li Zhang, Brian Curless, and Steven M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *Proc. 3DPVT*, pages 24–36, 2002.
- [142] Li Zhang and Shree K. Nayar. Projection defocus analysis for scene capture and image display. *ACM SIGGRAPH*, pages 907–915, 2006.
- [143] Jin Zhong. Binary ranks and binary factorizations of nonnegative integer matrices. *Electronic Journal of Linear Algebra*, 23(1):540–552, 2012.