Acquisition of Transparent Refractive Media

by

Bradley Atcheson

M.Sc., The University of British Columbia, 2007

B.Sc., University of the Witwatersrand, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Computer Science)

The University Of British Columbia

(Vancouver)

November 2012

Abstract

Transparent refractive media are invisible but for the distortions they impart upon a background scene. Computerised acquisition of such media can therefore often not be performed via traditional scanning methods. By capturing refracted backgrounds rather than reflections off the target media itself, we develop techniques for reconstructing the intervening refractive index distribution for both static and time-varying media. The approach is based on tracking optical distortions and then performing tomographic reconstruction. For multi-view tomography we first require a suitably calibrated camera array. To this end we show how to temporally synchronise and geometrically calibrate an array of consumer-grade video cameras that can scale to larger sizes, and at lower cost, than a comparative array of machine vision cameras.

For media of low dynamic refractive index range, such as mixing gases, we show how to acquire data and formulate a linear least-squares problem to solve for the refractive index distribution. Unlike traditional methods of fluid flow measurement, ours is non-invasive and fully volumetric. For materials of higher dynamic refractive index range, we develop an alternative acquisition method based on temporally-encoded structured light patterns. Media causing significant distortion of light rays give rise to a large, nonlinear inverse problem. Results indicate that grid resolution relative to the minimum refractive feature size is a key factor limiting the accuracy of reconstructions.

Preface

The published papers on which Chapters 3 through 6 are based are listed below. In addition to the co-authors indicated, some of the ideas in this thesis are based on conversations I had with Lukas Ahrenberg, James Gregson, Matthias Hullin and Gordon Wetzstein.

Chapter 3 B. Atcheson, F. Heide, and W. Heidrich. "CALTag: High Precision Fiducial Markers for Camera Calibration." In *International Workshop on Vision, Modeling and Visualization,* pages 1-8, Siegen, Germany, 2010. © Eurographics 2010.

The concept for this project was my own. I conducted most of the experiments, wrote the primary software implementation and wrote the majority of the published paper. Felix Heide wrote an alternate software implementation and portions of the paper, contributed algorithmic ideas and improvements, and delivered the oral conference presentation. Dr Heidrich supervised the project and wrote and edited portions of the paper.

Chapter 4 D. Bradley, B. Atcheson, I. Ihrke and W. Heidrich. "Synchronization and Rolling Shutter Compensation for Consumer Video Camera Arrays." In *IEEE Conference on Computer Vision and Pattern Recognition Workshops,* pages 1–8. IEEE, 2009. © IEEE 2009.

The original ideas on which this project is based came from Dr Heidrich in conversation with Dr Jack Tumblin. Derek Bradley was responsible for the stroboscopic method and I was responsible for the image-warping method. We worked closely together on analysing the camera model, removing distortions and writing the paper. Dr Bradley conducted additional experiments

in response to reviewer comments, and delivered the oral conference presentation. Dr Ihrke conducted background and related work research and helped to write the paper. Dr Heidrich supervised the project and wrote parts of the paper.

Chapter 5 B. Atcheson and W. Heidrich. "Non-Parametric Acquisition of Near-Dirac Pixel Correspondences," In *International Conference on Computer Vision Theory and Applications,* pages 247–254, Rome, Italy, 2012.

The concept for this project was my own. I conducted all of the experiments and wrote all of the code as well as the majority of the paper, and delivered the oral conference presentation. Dr Heidrich supervised the project and edited the paper.

Chapter 6 B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H.-P. Seidel. "Time-resolved 3D Capture of Non-stationary Gas Flows." *ACM Transactions on Graphics,* 27(5):132, 2008.

The original idea for this project was my own, but inspired by Dr Heidrich's initial investigations into Schlieren imaging. I was responsible primarily for the 2D data capture and processing. Dr Ihrke conducted most of the calibration and 3D tomography work. We wrote the paper and delivered the oral conference presentation together. Dr Heidrich supervised the project and wrote significant portions of the paper. Art Tevs produced the animated renderings used in the video. Dr Bradley assisted with experiments and editing. Dr Magnor helped to edit the paper.

Table of Contents

List of Tables

List of Figures

Glossary

API Application Programming Interface

ART Algebraic Reconstruction Technique

AR Augmented Reality

AD Automatic Differentiation

BOS Background Oriented Schlieren

BRDF Bidirectional Reflectance Distribution Function

BVP Boundary Value Problem

CCD Charge-Coupled Device

CG Conjugate Gradients

CGLS Conjugate Gradients Least Squares

CMOS Complementary Metal Oxide Semiconductor

CRB Cramér-Rao Bound

CRC Cyclic Redundancy Check

CRT Cathode Ray Tube

DC Direct Current

DFT Discrete Fourier Transform

DLP Digital Light Projector

DSLR Digital Single Lens Reflex

DSP Digital Signal Processing

ESPRIT Estimation of Signal Parameters by Rotational Invariance Techniques

FBSS Forward Backward Spatial Smoothing

FWI Full Waveform Inversion

IR Infrared

ISI Inter Symbol Interference

IVP Initial Value Problem

LCD Liquid Crystal Display

LED Light Emitting Diode

MDL Minimum Description Length

MUSIC Multiple Signal Classification

NTSC National Television System Committee

ODE Ordinary Differential Equation

PIV Particle Image Velocimetry

POCS Projection Onto Convex Sets

PSF Point Spread Function

PSNR Peak Signal to Noise Ratio

RANSAC Random Sample and Concensus

RMS Root Mean Square

SART Simultaneous Algebraic Reconstruction Technique

SIRT Simultaneous Iterative Reconstruction Technique

SNR Signal to Noise Ratio

STP Standard Temperate and Pressure

SVD Singular Value Decomposition

UV Ultraviolet

Acknowledgments

I wish to thank my supervisor, Dr Wolfgang Heidrich, for all his guidance and unwavering support over these years. And for giving me the freedom to pursue ideas, both good and bad, and to learn from them.

Thank you also to my committee members, Dr Robert Bridson and Dr Michiel van de Panne, whose comments and questions were always insightful and encouraging.

I was fortunate to have had Dr Ivo Ihrke as a mentor and a friend.

This dissertation was supported in part by a UBC University Graduate Fellowship and a UBC Four Year Fellowship.

To my colleagues in the Imager and PSM labs, thank you for all your help, encouragement and friendship. In particular, I would like to thank: Derek, Cheryl, Abhi, Allan, Anika, Ben, Felix, Gordon, James, Lukas, Matthias, Mert, Mike, Nasa, Stelian, Steven, and Trent.

The city of Vancouver and it's people have opened a whole new world unto me, and for that I am eternally grateful. Two people in particular have shared in my journey here, through the highs, lows and in-betweens. Maeve and Anupam, thank you for everything.

Finally, thank you to my parents. You gave me everything I needed and more to make it through this.

Chapter 1

Introduction

"Do not Bodies and Light act mutually upon one another; that is to say, Bodies upon Light in emitting, reflecting, refracting and inflecting it, and Light upon Bodies for heating them, and putting their parts into a vibrating motion wherein heat consists?"

— Sir Isaac Newton (1704)

When light interacts with a material surface it will reflect and refract in varying proportions. Reflections are more easily noticeable in everyday life, and we are accustomed to using them to make measurements – think of any photographic camera or laser rangefinder. Refractions are less often exploited, but in niche applications do provide useful data for taking measurements.

Computerised acquisition of physical objects and phenomena allows us to represent the real-world digitally. This is useful for both aesthetic purposes, such as movie and game art assets, as well as functional ones, where we wish to study an object's behaviour under varying conditions. Thanks to a large body of research, huge quantities of data about many classes of media can be captured today fairly easily. The macroscale geometry of objects both small and large can be captured using structured light [e.g., Rusinkiewicz et al., 2002] or laser range scanners [e.g., Levoy, 1999]. More complicated surfaces like human skin [e.g., Bradley et al., 2010] and cloth [e.g., White et al., 2007] can be acquired using stereo vision. Entire cities are currently being acquired through large collections of photographs [e.g., Snavely et al., 2006], while at the same time we have access to interferometric

devices for measuring surface profiles at microscopic scales [e.g., Kumar et al., 2009]. More ethereal phenomena such as smoke [e.g., Hawkins et al., 2005] and fire [e.g., Ihrke and Magnor, 2004] can also be acquired.

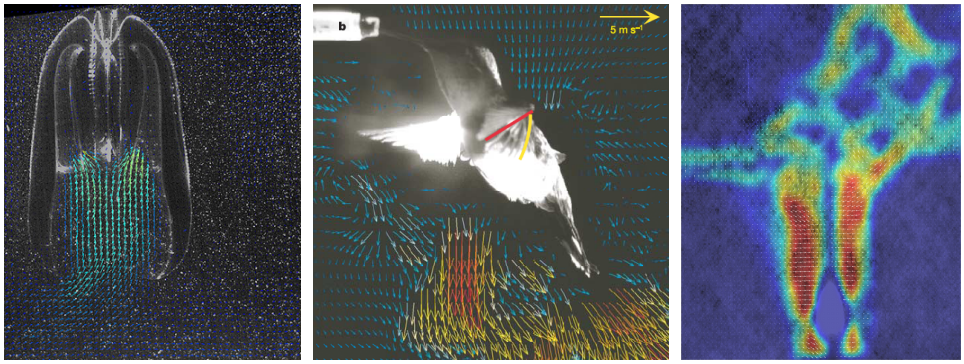
Certain classes however remain especially elusive. Transparent glass figurines do not reflect much light and so must first be coated in reflective powder before being placed into a laser scanner. There have been attempts at capturing such transparent objects but it remains a difficult problem [Ihrke et al., 2008]. Some media interact so weakly with light that they are practically invisible. Parcels of gas, heated or having a different chemical composition from their surroundings, are all around us but go by practically unnoticed. When their flow is too complex to simulate numerically, we must obtain the flow data via computerised acquisition. Such acquisition demands appropriate data collection tools and analysis algorithms. Dedicated hardware tools for certain very specific applications (i.e., Particle Image Velocimetry (PIV) [Grant, 1997]) are available, but their high cost and specificity render them inaccessible to most. A theme of this thesis is to move complexity away from the hardware side towards the computational side, allowing us to drastically reduce the cost of and increase access to refractive acquistion tools.

This thesis describes a set of techniques for imaging refractive media. In particular, we develop a suitable toolset using only low-cost conventional consumer equipment, and algorithms for processing and inverting the measurements. We make contributions to camera calibration and synchronisation, as well as 2D refractive imaging (environment matting) and 3D reconstruction (refractive tomography). To motivate this work, the following section describes a few of the potential applications to which it could be directed. Specific contributions are then summarised in Section 1.2 along with the structure for the remainder of the thesis.

1.1 Motivation and Applications

1.1.1 Fluid Imaging

Imaging fluid flowfields can provide important insights in biological studies. For example, *Mnemiopsis leidyi* (a comb jellyfish) has recently expanded its habitat



(a) (b) (c)

Figure 1.1: (a) 2D PIV flowfield around a comb jellyfish illustrates the stealth current it generates to capture prey. Reprinted with permission from [Colin et al., 2010]. (b) 2D PIV airflow velocity around a hovering hummingbird. Reprinted with permission from [Warrick et al., 2005]. (c) BOS photograph of a candle plume.

and caused significant changes in the plankton stocks upon which it feeds. To understand its remarkable effectiveness as a predator, scientists have employed PIV to examine the flow field around its tentillae during feeding. The 2D velocity field in Figure 1.1(a) illustrates the reason for its success – a highly laminar flow created by cilia between the oral lobes. This "stealth" flow is virtually undetectable to prey swimming nearby, until they enter its gentle current directing them towards their inevitable demise [Colin et al., 2010].

Hummingbirds posess a unique mode of flight that grants them remarkable agility. With high speed cameras one can image the vortices created under the wings and learn, for example, the relative contributions of the fore- and backstroke of the wings. Figure 1.1(b) illustrates the transport of air in a plane aligned with the bird's medio-lateral axis [Warrick et al., 2005].

These images were produced by seeding the flows with reflective particles (e.g., 10 µm glass beads) and illuminating them with a laser. However, such specialised hardware need not be necessary. In previous work, we have demonstrated the use of BOS imaging to acquire similar flowfields in heated air. Figure 1.1(c) illustrates how one can qualitatively examine the flow of heated gases – here, a candle plume

disturbed by a high speed horizontal injection. Aside from data processing, this image required nothing more than a consumer-grade camera and a printed page of random noise to produce. The method's drawback is that one obtains not an airflow velocity distribution, but rather a projection of the gradients of the 3D refractive index field. Its benefit lies in being able to do so noninvasively, by imaging refraction rather than reflection. The refractive index of dry air decreases approximately linearly with rising temperature in the regime of Standard Temperate and Pressure (STP) [Ciddor, 1996]. The imaging techniques we use are able to resolve these changes and generate a distortion map from the point of view of a camera looking through a scene of varying refractive index.

Multiple cameras directed through the same scene allow for 3D reconstruction via tomography. In Chapter 6 we demonstrate time-varying reconstructions of heated gas plumes using this technique. One of the simplifying assumptions made in that work is that light rays travel along linear paths, unperturbed by the changing index. This corresponds to the often used paraxial approximation of ray optics. However, for more accurate results, one should take into account the curved nature of the ray paths. In this thesis we explicitly consider ray bending and develop an analysis-by-synthesis framework for reconstructing the refractive index distribution, given only the entry and exit positions and angles of each ray traversing the scan volume.

Ideally we would like to convert the reconstructed refractive index into other physical properties of the medium. In gases close to STP, with knowledge of the refractive index n and wavelength λ one can approximate density ρ and temperature T through the Gladstone-Dale relation [Gladstone and Dale, 1863]

$$n(\lambda) - 1 = k\rho \tag{1.1}$$

with k a constant, or Minneart's approximation [Stam96]

$$\frac{n(\lambda) - 1}{n_0(\lambda) - 1} = \frac{T_0}{T}, \tag{1.2}$$

where $n_0 = 1.00023$ and $T_0 = 273\,\mathrm{K}$. Such formulae provide a rough estimate of gas parameters, but for accurate analysis via refractive index measurement it

is necessary to account for a host of other properties (e.g., humidity, CO_2 content etc.).

1.1.2 Manufacturing Inspection

Many industries manufacture glass products to various degrees of optical purity. For example, microscope objective lenses require diffraction-limited performance, whereas the design imperfections in cheap camera lenses will generally dwarf minor inhomogeneities in the raw glass from which they are ground. The luminaire design industry must meet more relaxed quality constraints while maintaining low cost. Vehicle headlamps are typically produced by slicing and grinding "pucks" from a single cylindrical rod. Optical impurities in the rod may force the entire piece to be discarded, or else a metric of the degree to which the rod is impure may allow for a judgement call to be made on its use. Quality control of transparent solids is currently performed manually [Wild, 2008] but could be automated using the techniques described in this thesis.

Bottling plants, and the burgeoning photovoltaic industry, also inspect glass for inhomogeneities that could cause explosions upon exposure to thermal shock. Due to the high speed requirements, specialised probes are currently in use for obtaining sparse glass wall thickness measurements [Michelt and Schulze, 2006]. While our more extensive acquisition and processing requirements are currently unsuitable for production lines, it may prove useful to scan occassional parts for more extensive verification checks.

We employ geometric optics exclusively throughout this thesis, since our focus is on refractive features down to a few millimetres in size, well beyond the wavelength of light at which it becomes necessary to employ wave-like models to handle scattering, wavefront healing and other diffraction effects. For precision optics, one is better served by interferometric techniques that can detect ray path length deviations (and hence measure lens surface profiles) down to submicron resolutions [Steinmetz, 1990]. In contrast, our imaging techniques operate at the (ever-increasing) resolutions of spatial light modulators and camera sensors.

1.1.3 Other Applications

Settles [2001] mentions many of the areas in which Schlieren imaging has had an impact. A partial listing includes: acoustic design, aerodynamics, ballistics, boundary layer interaction, convective heat transfer, glass purity, instability and concentration of mixing liquids, jet engine noise control, leak detection, medical diagnostics, thermal cutting and even artistic applications. In many cases a single 2D Schlieren photograph provides all the necessary information, but complete 3D reconstructions of fine refractive index variations can help us learn more about processes such as botanical transpiration [Gates and Benedict, 1963], or to optimise room ventilation [Heinsohn et al., 1986].

Refraction is not restricted to visible wavelengths, and although that is the focus of this thesis, similar principles apply to any imaging process in which refraction occurs and geometrical rays are an adequate model. Terahertz tomography in medical diagnosis is one potential new application [Abraham et al., 2010].

1.2 Contributions and Outline of Dissertation

In Chapter 2 we review some of the related work on imaging refractive media. In computer graphics the primary focus in this area has been image-based rendering, where the goal is to capture images representative of the real world and use them to render novel scenes. Our 2D refraction data is similar to that obtained in another graphics application – environment matting. We also examine the theory behind Schlieren photography, a scientific imaging technique underlying our approach to reconstructing index distributions from refraction measurements. Our 3D reconstructions resemble experiments from the literature to reconstruct profiles of symmetric, or stationary gas flows, but we aim to support more general, and higher ranges of, refractive index distributions.

1.2.1 Camera Calibration

In Chapters 3 through 5 we describe the camera-based tools developed for acquiring refraction data. One of the issues that arises in working with camera arrays for tomography is that the amount of manual labour involved quickly becomes unmanageable. Work must be repeated for each camera, and ensuring that all cameras

are recording a sufficient portion of both the calibration target and capture scene before beginning the experiment and data download process is difficult. In particular, the popular approach to geometrically calibrating cameras involves using a planar chequerboard target and mapping pixel coordinates of detected points to their known real-world positions on the grid [Tsai, 1986]. In practice, locating and identifying chequerboard feature points in an image is tedious to do manually, and difficult to do automatically, owing to the aperture problem. To address this, we developed a self-identifying pattern and detection algorithm, described in Chapter 3, to completely automate this process.

1.2.2 Camera Synchronisation

Two other troublesome properties of arrays of consumer camcorders are their lack of temporal synchronisation and their use of rolling shutters. In Chapter 4 we describe how to exploit the rolling shutter to obtain very precise relative temporal offsets amongst the cameras and how to use these to align the data in time. This enables us to construct a low-cost camera array that is both flexible and capable of high quality recordings.

1.2.3 Pixel Correspondences

Chapter 5 describes the structured-light system we developed that allows for mapping pixels from a spatial light modulator (LCD monitor, DLP projector etc.) to a camera. Whereas previous BOS work employed single-image optical flow based acquisition methods [Atcheson, 2007], for more extreme refractions we must find more reliable methods. Our particular interest lies in mapping from single pixels on the source to multiple, but still high (spatial) frequency groups of pixels on the detector. Existing light transport acquisition methods can efficiently map from single pixels to single pixels, or to a single parametrically-described group of pixels, but we have found complex refractive media to exhibit much more varied point spread behaviour that is better modelled via non-parametric approaches. Our solution addresses the middle ground between point-to-point correspondences and lower frequency light transport approximation methods.

Mapping pixels on planar display surfaces allows us to parametrise 3D rays that

pass through a refracting medium. Comparing these to undeflected ray paths gives us a per-ray deflection measurement that can be used for environment matting or tomographic reconstruction purposes.

1.2.4 Refractive Tomography

Finally, in Chapters 6 and 7 we examine the problem of tomographic reconstruction of refractive index distributions given the acquired deflection data. We are able to demonstrate successful 3D reconstructions of low-index media (gas flows) as well as partially successful reconstructions for higher-index media where significant ray curvature occurs.

We develop two different reconstruction algorithms. In Chapter 6 we tomographically reconstruct the gradient of the index field and then integrate that solution to obtain the final output. This method is suitable for media having a small refractive index range. In Chapter 7 we diverge from that approach and cast the problem instead as one of nonlinear optimisation. This enables us to solve directly for the index itself, allowing for more natural regularisation of the objective function and more efficient representations of the medium. As with most inverse problems, the objective is highly non-convex and so 3D reconstructions of complex high-index media are not currently possible. In the event that these algorithmic problems are one day solved, we show how to construct and calibrate a suitable acquisition setup that can very accurately measure rays deflected through high-index media, while simultaneously ignoring those that scatter.

Using 2D simulations we then develop a guided gradient-descent based optimisation method for finding the refractive index on a discretised domain. We solve the nonlinear inverse problem employing geometric optics for tracing rays in the forward model. From these simulations we identify and solve some of the key difficulties of minimising data misfit terms in such problems.

Previous work on refractive tomography has considered explicit linearisation of the ray trajectories and then iteration towards a solution. However, they generally use methods that require processing the entire captured dataset each iteration. In contrast, we describe how to move towards the solution using only one ray at a time. To make the problem more feasible on current hardware (3 GHz quad-core

CPU, 8 GB RAM), we employ stochastic gradient descent and use automatic differentiation to obtain local gradient information. We also show how this problem relates to seismic imaging, where the travel-time of seismic waves is used to recover the "slowness" distribution in the shallow earth (analogous to our refractive index), and discuss the relationship between our approach and algebraic reconstruction techniques.

Chapter 2

Background and Related Work

"... some of the best schlieren images are obtained in time-honored fashion by patience and by seizing the right moment."

— Gary Settles (2001)

In this chapter we begin by describing the physics underlying light's propagation through the media we aim to acquire, and then review some of the techniques available for imaging transparent refractive media. The focus here is on the broader picture of refractive acquisition; more project-specific related work will be discussed in the relevant chapters that follow.

Currently, the industry-standard method for acquiring fluid flowfields is PIV. It involves seeding a flow with reflective particles and illuminating them with a plane of laser light. This naturally allows 2D slices to be recorded, although extension to 3D via steroscopic [Arroyo and Greated, 1991] and tomographic [Elsinga et al., 2006] methods are possible. Fluids can also be acquired via an old optical technique called Schlieren* photography [Settles, 2001]. Also 2D, it differs in two key respects that make it complementary to, rather than a replacement for PIV:

- rather than extracting 2D slices, it records 2D projections, and
- rather than visualising transport inside the fluid, it captures the gradient of the refractive index.

*German for "streak" – referring to optical distortions caused by inhomogeneous refractive index.

Despite these differences, there have been many attempts to scan fluids in motion via Schlieren. Section 2.3 covers some of these, as well as a primer on the underlying optical theory.

Schlieren methods work well for nearly homogeneous materials that closely match their surrounding media (e.g., heated air). However, in scanning solid objects we must usually accommodate much greater refractive index ranges. Due to their extreme sensitivity, this is not a domain in which classical Schlieren methods shine. Fortunately however, whereas PIV and Schlieren are typically concerned with fast-moving dynamic scenes, solids can be photographed from multiple angles over an extended time. Exploiting this, we have developed a method to acquire Schlieren-like data despite the high index gradients (Chapter 5). Others approaches for scanning transparent media, ranging from purely image-based 2D methods to full simulation have been proposed in the computer graphics literature, and are discussed in Section 2.2.

Once the Schlieren-like data is available, it can be inverted to obtain a 3D refractive index distribution. The local index affects light velocity, analogous to the acoustic refractive index in sound waves, and the way in which the bulk modulus (resistance to uniform compression) of rock affects seismic wave propagation. Lacking a high-frequency pulsed laser, the data used in optical acquisition differs significantly from the time-of-flight information available in seismic methods. Nevertheless, the underlying inversion algorithms are similar and are reviewed in Section 2.4.1.

2.1 Ray Propagation in Inhomogeneous Media

Refraction at a planar interface is described by Snell's law. As shown in Figure 2.1(a), a ray entering a medium of higher optical density will bend towards the interface normal, with angles varying according to the ratio

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1} \tag{2.1}$$

of refractive indices n_i. Snell's law is the limit case behaviour of a more general continuous model. For graduated media the same rule applies in a differential

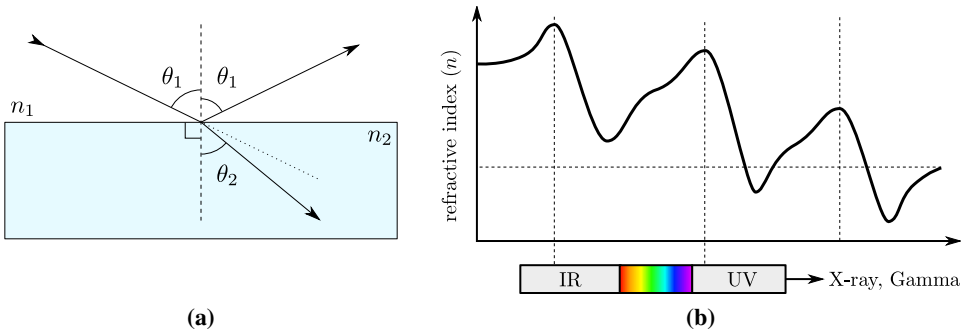


Figure 2.1: (a) Snell's law for $n_2 > n_1$ showing incident, reflected and refracted rays. (b) Schematic plot of refractive index dependence on wavelength in a fictional transparent material. Three resonance frequency modes are shown.

sense, resulting in rays that curve smoothly during traversal, obeying the ray equation derived below.

The refractive index is defined as the ratio of the speed of light in a vacuum to that of light in the medium[†]. It should be noted that this is both a material and wavelength dependent property. In glass, blue light refracts more strongly than red, as experimentally verifiable by observing the dispersion through a prism. This suggests that scanning methods that aim to reduce refraction as much as possible may consider the use of infrared illumination in addition to index-matching fluids. For the reader curious as to how it can be that red light refracts less strongly than blue, yet more so than X-rays, which pass unperturbed through practically most materials, the answer lies in resonance. The general trend along the electromagnetic spectrum is for higher frequencies to refract *less*. In water for example, radio waves have an index of about 9.73 at λ =24 cm, about 8.36 at λ =9 cm and since they hardly refract, 1.0 for X-rays [Goldsmith, 1937]. The refractive index itself is a function of the relative permittivity and permeability of a medium (specifically, the square root of their product). Electromagnetic fields in the medium oscillate naturally, and interact with passing waves. The degree of interaction is proportional to

[†]negative refractive index metamaterials are the subject of much recent interest. Indices below 1.0 are possible since they refer to the *phase velocity* of light, which carries no information and can therefore travel faster than 3×10^8 m s^{-1}.

the relative permittivity. Higher frequency waves have *less* influence on the vibration of charged particles because their force is more uniform when integrated over the time scale of the particle's vibration. However at certain resonance frequencies (electrons, atoms and molecules all contribute their own resonance freqencies) the effect is greatly amplified. Most common transparent materials have such a resonance in the near ultraviolet [Bach and Neuroth, 1995]. Hence, despite the overall trend towards lower refraction with higher frequency, as one approaches Ultraviolet (UV) (i.e., going from red to blue) the index *increases* and then decreases sharply once past the resonance mode. The refractive index is in general a complex value related to the dielectric function of a material [Cai and Shalaev, 2009]. However the imaginary component relates to the absorption coefficient and is negligible for everyday transparent materials at optical frequencies. We are concerned primarily with the real part, which can be modelled by the Sellmeier dispersion formula

$$n^2(\lambda) = 1 + \sum_j \left(\frac{S_j \lambda^2}{\lambda^2 - \lambda_j^2} \right) \tag{2.2}$$

where S_j and λ_j are the strength and wavelength of the j^{th} resonance mode. Figure 2.1(b) illustrates such an index profile with resonances in the far-IR and near-UV. Theoretical formula and empirically-determined constants can be found by consulting online databases [RID, 2012].

To combat dispersion, scanning methods employ laser light or else place a narrowband filter over a broad spectrum source (or the camera lens) [Trifonov et al., 2006]. Unfortunately this is at odds with the desire to maximise transmission to reduce sensor noise.

Throughout this thesis we employ geometrical optics (also known as the *infinite frequency approximation* in geophysical literature). Its use is predicated on the wavelength being significantly smaller than the scale of refractive index variations. When the scale of the media's inhomogeneities is on the order of a wavelength, then diffraction and interference effects will dominate. This dramatic simplification nevertheless provides good results when analysing everyday scenes and camera lenses. When much longer wavelengths are used, for example in seismic tomography, finite frequency models must be employed. These take diffraction

and travel-time into account, leading to the wonderfully-named *banana-doughnut* theory, after the longitudinal and cross-sectional shapes of the Fresnel zone surrounding the ray and influencing its path [Tromp et al., 2004].

In order to derive a model for the propagation of light rays through an inhomogeneous medium, one can begin with the Maxwell equations and apply the infinite frequency limit to arrive at the Eikonal[‡] equation [Kriezis et al., 1992]

$$|\nabla S|^2 = n^2. \tag{2.3}$$

Its scalar field solution $S(\mathbf{r})$ describes the phase evolution of a wave. Essentially this tells us the minimum travel time for a wave to propagate to point $\mathbf{r}$ from the source. A solution can only be obtained if the wavefronts do not cross (i.e., caustics do not form). The level sets of S are geometrical wavefronts and a ray is thus defined precisely as *a curve tangent to some wavefront unit normal $\hat{s}$ at each point* $\mathbf{r}$ *on the curve.* We therefore have by definition that

$$\widehat{\mathbf{s}} = \alpha \nabla S \tag{2.4}$$

From its unit magnitude we get that

$$\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}} = \alpha^2 |\nabla S|^2 \tag{2.5}$$

$$= \alpha^2 n^2 \tag{2.6}$$

$$= 1 \tag{2.7}$$

$$\implies \quad \widehat{\mathbf{s}} = \tfrac{1}{n} \nabla S. \tag{2.8}$$

If we let s be the differential arc length along a ray (parametrised by $\mathbf{r}$) between two adjacent points, then we get the unit tangent vector

$$\widehat{\mathbf{s}} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} \tag{2.9}$$

[‡]Greek for "image".

and from (2.8) and (2.9) we get an equation relating rays to the Eikonal Equation:

$$\nabla S = n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}. \tag{2.10}$$

A more useful form is obtained by eliminating S. To do so we make use of the following identity for a vector field V (illustrated here in 2D Cartesian coordinates)

$$\begin{aligned}
\frac{\mathrm{d}V}{\mathrm{d}s} &= \frac{\mathrm{d}V_x}{\mathrm{d}s}\widehat{\mathbf{x}} + \frac{\mathrm{d}V_y}{\mathrm{d}s}\widehat{\mathbf{y}} && (2.11)\\
&= \widehat{\mathbf{x}}\left(\frac{\partial V_x}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\partial V_x}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}s}\right) + \widehat{\mathbf{y}}\left(\frac{\partial V_y}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\partial V_y}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}s}\right) && (2.12)\\
&= \widehat{\mathbf{x}}\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\cdot\nabla\right) + \widehat{\mathbf{y}}\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\cdot\nabla\right) && (2.13)\\
&= \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\cdot\nabla\right)V && (2.14)
\end{aligned}$$

and differentiate the right hand side of Equation 2.10 with respect to s to obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right) &= \frac{\mathrm{d}}{\mathrm{d}s}\Big(\nabla S\Big) && (2.15)\\
&= \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\cdot\nabla\right)\nabla S && (2.16)\\
&= \left(\frac{1}{n}\big(\nabla S\big)\cdot\nabla\right)\nabla S && (2.17)\\
&= \frac{1}{2n}\nabla\Big(\nabla S\cdot\nabla S\Big) && (2.18)\\
&= \frac{1}{2n}\nabla\Big(\nabla S\Big)^2 && (2.19)\\
&= \frac{1}{2n}\nabla n^2 && (2.20)\\
&= \nabla n && (2.21)
\end{aligned}$$

leaving us with what is called the *Ray Equation of Geometric Optics* relating the ray's trajectory to the refractive index gradient [Kriezis et al., 1992]

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right) = \nabla n \tag{2.22}$$

The ray equation has been used in computer graphics to render atmospheric effects [Gutierrez et al., 2006; Stam and Langue, 1996] and complex refractive objects [Ihrke et al., 2007]. After a simple substitution it can be reformulated as a system of first order Ordinary Differential Equations (ODEs)

$$n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \mathbf{d} \tag{2.23}$$

$$\frac{\mathrm{d}\mathbf{d}}{\mathrm{d}s} = \nabla n \tag{2.24}$$

Note that we have here a parametrisation in terms of constant step size because $|\mathbf{s}| = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right| = 1$. When implementing a numerical ray tracer, using interpolated refractive indices in a discretised model, one must take care to enforce this constraint. Ihrke et al. [2007] derive the equivalent equations for constant *temporal* step size:

$$n\frac{\mathrm{d}}{\mathrm{d}t}\left(n^2\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right) = \nabla n \tag{2.25}$$

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathbf{w}}{n^2} \tag{2.26}$$

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}t} = \frac{1}{n}\nabla n \tag{2.27}$$

Note that the ray equation applies only to smoothly varying isotropic media and not sharp discontinuities, since S is not differentiable there. This, along with our use of finite differences to obtain gradients in a discretised synthetic model, makes it impossible to recover sharp edges.

In addition to refraction and absorption, light will also reflect when passing into a new medium. The reflection coefficient for unpolarised light is described by the Fresnel equation (see Figure 2.1(a))

$$2R_s = \left|\frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2}\right|^2 + \left|\frac{n_1\cos\theta_2 - n_2\cos\theta_1}{n_1\cos\theta_2 + n_2\cos\theta_1}\right|^2. \tag{2.28}$$

This approaches a maximum of 1.0 as $\theta_1 \to \frac{\pi}{2}$, indicating that all materials become progressively more reflective as one views them at shallower angles. When moving into a less dense medium ($n_2 < n_1$) there is a certain *critical angle* (approximately 41° for glass in air), at which the maximum is reached and all light is reflected

(total internal reflection). The sudden disappearence of the refracted component causes discontinuities in functions computed over rays traced according to these formulae, which adversely affects optimisation algorithms. In the opposite case ($n_1 < n_2$) there is a similar special value, called Brewster's angle, at which the reflected ray disappears completeley, but only for p-polarised light. In this thesis we ignore the effects of polarisation.

A ray trajectory through the scan volume can therefore be traced using Equations 2.23 and 2.24 given an initial position and direction vector. In reconstructing the refractive index field of gases based only on ray measurements (Chapter 6), the field $n(\mathbf{r})$ is unknown, and so a synthetic proxy $\tilde{n}(\mathbf{r})$ must be used instead. The particular acquisition setup we use also provides the exit position and direction for each ray. One therefore has two options, as shown in Figure 2.2:

- Solve a Boundary Value Problem (BVP) over $\tilde{n}$ for each ray in order for the trajectory to match known entry and exit conditions. This would involve numerous shooting method [Stoer and Bulirsch, 2002] iterations in order to arrive at one possible solution for the voxels through which the ray currently travels, and which will likely be inconsistent with the solutions of other rays.

- Solve a deterministic Initial Value Problem (IVP) for each ray to obtain its exit position and direction. The difference (*misfit*) between these and the measured exit parameters can then be iteratively minimised by varying $\tilde{n}(\mathbf{r})$.

As with iterative solvers in general, it is often better to take smaller step in the right direction than a large step in the wrong one. In solving the BVP the former approach is likely to expend too much effort in obtaining too good of a solution for just one ray and so the latter is to be preferred. Minimising misfits of measured data to forward simulations is also the dominant approach used in seismic tomography (see Section 2.4.1).

2.2 Imaging Transparent Media

Transparent objects are inherently more difficult to capture digitally than those that reflect light. In the remainder of this chapter we mention some of the approaches that have been used in the past. These range from the very simple (environment

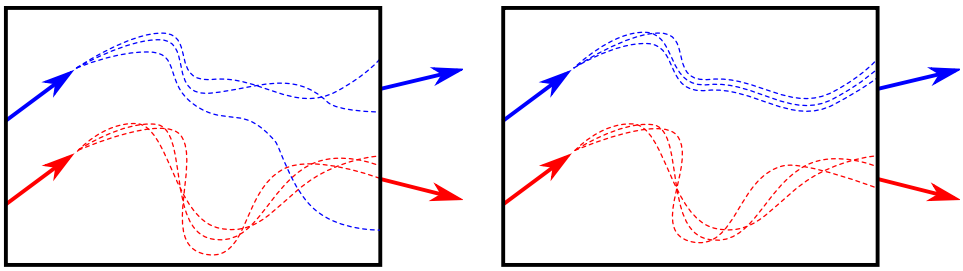


Figure 2.2: (a) Solving for $\tilde{n}$ as a BVP with the red ray having constrained entry and exit parameters. The blue ray may not simultaneously satisfy its constraints. (b) Solving as an IVP minimises error across all rays in each iteration.

matting) to those that make explicit use of the physics described above (Schlieren tomography) and that exploit refraction as their primary acquisition modality.

2.2.1 Environment Matting

First described by Zongker et al. [1999], environment matting is a relatively simple technique for capturing complex, occlusion-free lighting reflections and refractions of an object in a simple 2D data structure. The object can then be re-rendered into novel scenes while maintaining visually correct interaction with its new surroundings. A complete environment matte could potentially describe multiple properties (opacity, colour etc.) but we are primarily concerned here with recording the outgoing direction of refracted camera rays.

The technique differs from similar *environment mapping* [Blinn and Newell, 1976] methods (cube maps, sphere maps etc.) in that it relies on purely image-based capture and rendering, rather than a geometric model of the object, and in its focus on refraction and translucency rather than reflection.

An environment matte is usually captured by projecting seqeuences of structured background patterns behind the static object and filming it. Various patterns, from sweeping lines [Zongker et al., 1999] to natural images [Wexler et al., 2002] to wavelets [Peers and Dutré, 2003] have been used to obtain the mapping between sensor and background pixels. Subjects of interest for rendering have typically

been glass or glossy surfaces, with many sharp edges and rough surfaces causing beams to spread over large areas. Environment mattes therefore consist of mappings of rays to *regions* of the background (e.g., parametrically described by oriented Gaussian blobs [Chuang et al., 2000]). Applications include the generation of novel images by compositing mattes with new backgrounds (Figure 2.3(a)). Chapter 5 describes a method for capturing similar data, but in a non-parametric form to better represent the complex shapes a beam footprint can take when passing through a strongly refracting medium.

Peers and Dutré [2003] proposed the use of wavelets as illumination patterns for environment matting. Their initial algorithm was adaptive, i.e., it required processing the results of captured images to decide which patterns to project next, increasing overall acquisition time. This was remedied in later work [Peers and Dutré, 2005], in which the authors use sparsity priors to project results obtained with a fixed set of illumination patterns into a new wavelet representation. While this method produces excellent results for wide Point Spread Functions (PSFs), it is less applicable to sharp PSFs that occur on highly glossy surfaces. More images must be captured for these scenes since sharper PSFs require more basis functions to represent them. The method we use is both non-adaptive and tailored towards high frequency PSFs.

Environment matting relates to this thesis in two ways. First, it represents a simplified form of the data we wish to capture. As a purely single-view image-based representation it captures only the *effect* of a medium's refraction, and only from one point of view. Our goal is more general in that we seek a model of the refractive media so as to be able to render it into arbitrary scenes more accurately (without the assumption of the surrounding environment being infinitely far away). Second, our method uses as input a restricted class of environment matte – we capture only the refracted direction of rays that pass through in specular fashion i.e., do not spread out over large areas like in Figure 2.3(e). Environment matting can represent more general scenes than this, but we use this simplified data input in order to construct a far more general model.

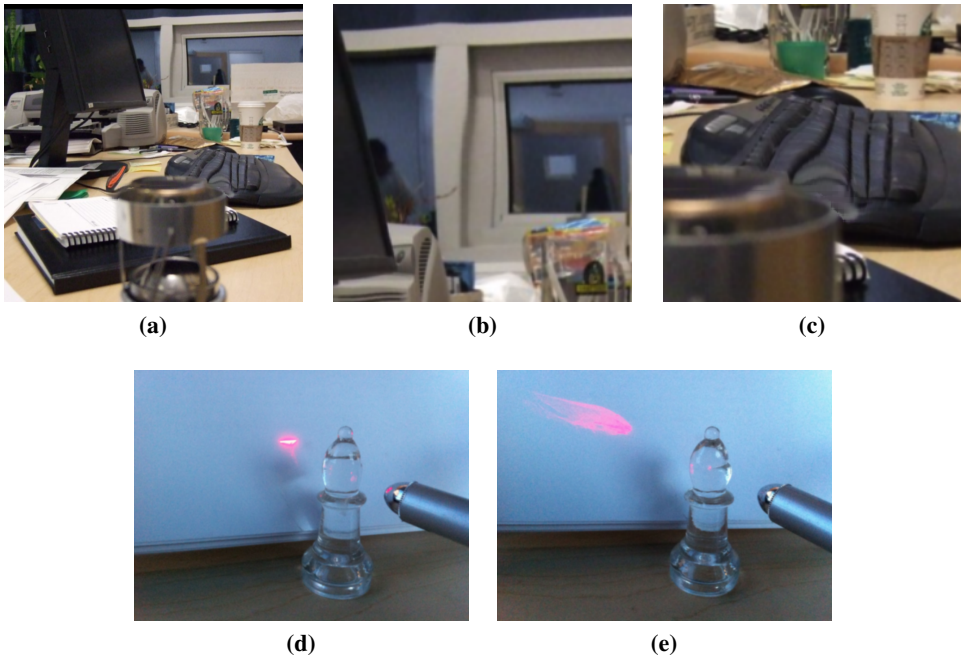


Figure 2.3: (a) Novel scene synthesised using captured gas plume environment matte (exaggerated for clarity – notice distortion in the keyboard and window frame) (b) Zoom-in on window. (c) Zoom-in on keyboard. (d) Specular beam footprint (pseudo-PSF) of laser pointer through glass object. (e) Wider non-specular footprint due to stronger, discontinuous refraction.

2.2.2 3D Acquisition

Moving away from image-based models towards full 3D representations, there has been some exploratory work over the past decade using various approaches. Some modify the medium itself (or the surrounding medium) in order to make acquisition easier, and some acquire just the first surface geometry whereas others obtain volumetric data.

Building on the much larger body of work in scanning opaque diffuse objects, Goesele et al. [2004] covered their transparent models in diffuse dust. This allowed for acquisition of surface geometry via standard laser triangulation. Going beyond just the surface geometry, the authors also captured the PSF at each sur-

face position, allowing them to record and render the subsurface scattering using an image-based model.

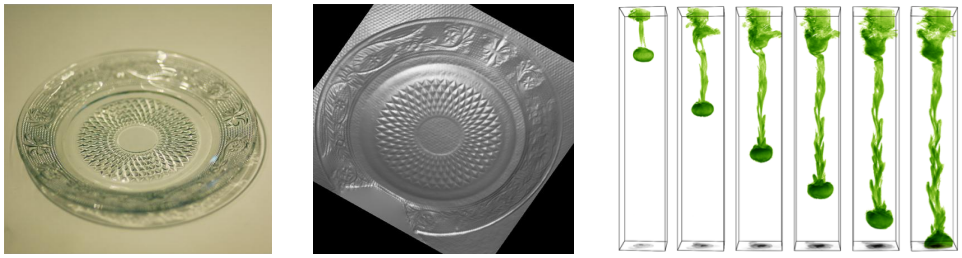
Applying a similar idea to fluids in motion, Wang et al. [2009a] added a scattering agent (white paint) to water squirted upwards in a fountain. This allowed for reflection of a projected noise pattern so that stereo reconstruction could be performed. The novel aspect of this work was in the direct inclusion of the Navier Stokes equations into the reconstruction algorithm to help guide the stereo reconstruction.

Gas and liquid flows have also traditionally been acquired by ignoring any refractions and instead placing reflective particles into the flow. This approach is termed PIV and is the mainstay of fluid acquisition today. The basic approach involves shining a sheet of laser light into the dust-filled medium and recording the motion with a high temporal resolution. Cross-correlation or other means are then used to find the in-plane motion [Grant, 1997; Westerweel, 1997].

Chen et al. [2007] propose a method for acquiring first surface geometry. It exploits the fact that polarization of light is modified by multiple scattering to filter out those effects and record just the direct illumination, from which reconstruction can be conducted using structured light. Nayar et al. [2006] propse an alternative method for separation of the direct illumination.

An interesting class of dynamic transparent media for which surface acquisition has been investigated is shallow water surfaces. Murase [1990] employed a single-camera setup directed at a noise pattern placed in the bottom of a shallow pool. The apparent position of points in the pattern are tracked over time via optical flow [Lucas and Kanade, 1981]. Their approach solves simultaneously for the pattern itself and the water surface normals, which are then integrated up to form a surface. Morris and Kutulakos [2005] improved the method by using a stereo camera pair along with a known background, and were able to acquire surface normals (hence geometry) as well as depth. Wetzstein et al. [2011] used a single-image light field probe to acquire surface normals for dynamic liquids as well as simple near-flat glass solids. An example object and reconstruction is shown in Figures 2.4(a) and 2.4(b).

High resolution tomographic acquisition of dynamic fluids has recently been presented by Gregson et al. [2012]. Using fluorescent dye it is an emissive straight-



(a) (b) (c)

Figure 2.4: (a) and (b) Thin glass dish and its heightfield reconstruction. Image used with permission of Gordon Wetzstein. (c) Fluorescent dye acquired tomographically. Image used with permission of James Gregson.

ray method that works with relatively few camera views and has low memory requirements (see Figure 2.4(c)).

Methods that aim to acquire volumes have generally attempted to mitigate refraction through immersion in an index-matching fluid. Trifonov et al. [2006] used the hazardous KSCN (potassium thiocyanate) in water to obtain a solution of index 1.5 at 80% concentration [Budwig, 1994]. If the target glass object is homogenous then this will remove refraction and allow for straight-ray absorption tomography. The authors dyed the solution to obtain the required measurement contrast. A representative result is shown in Figure 2.5(b). Other liquids (e.g., water: 1.333, ethanol: 1.361, glycerol: 1.473) may be used to match materials of lower index, or to provide an approximate match if the reconstruction method is robust to ray refraction. Index-matching silicone gels are also available, but are expensive to produce [Stone and Connor, 2000].

Rather than trying to observe the target object itself, one could instead acquire the surrounding medium. Hullin et al. [2008] also immerse their glass objects in fluid, but use a single-scattering fluorescent dye (Eosin Y) solution instead. The dye is excited by a sheet of laser light, which does not scatter inside the transparent object (Figure 2.5(d)). Sweeping the plane through the volume, or rotating the object, allows them to track the position of the fluid/object interface and reconstruct surface geometry. Alternatively, by also matching the fluid's index to the glass,

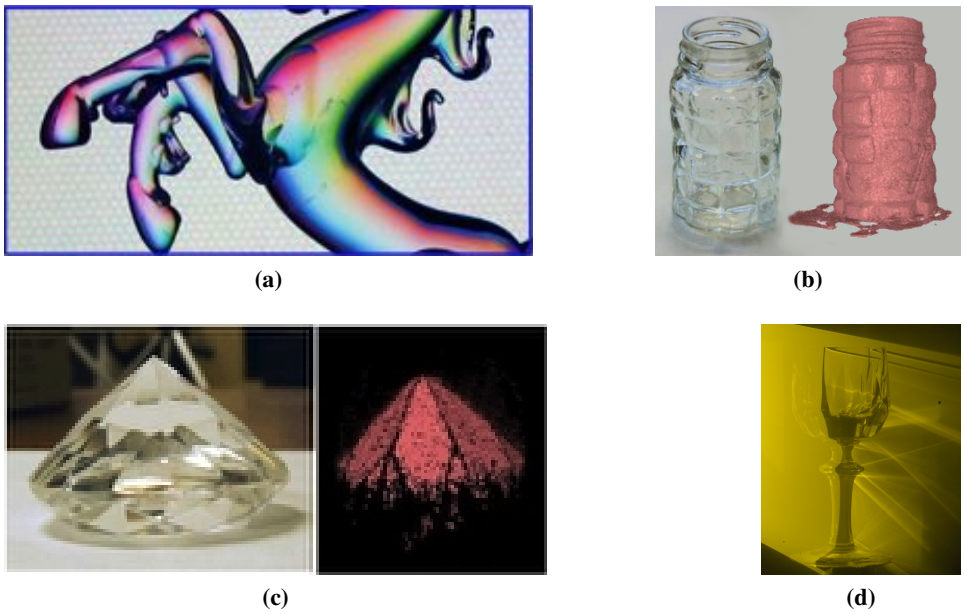


Figure 2.5: (a) Lightfield probe behind glass figurine convert angular deflection directly into colour. Image used with permission of Gordon Wetzstein. (b) Input and reconstructed glass model from absorption tomography in index-matched fluid. Reprinted with permission from [Trifonov et al., 2006]. (c) Input and reconstructed glass model from light field triangulation. Image used with permission of Kyros Kutulakos. (d) Wineglass immersed in fluorsecent dye solution, illuminated by laser light. Image used with permission of Matthias Hullin.

they can obtain volumetric slices directly.

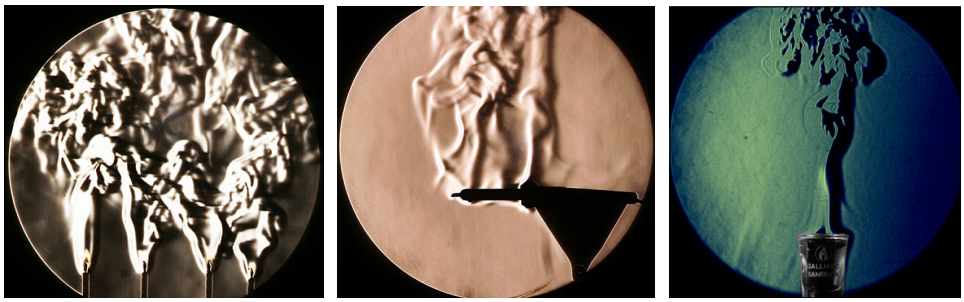
A thorough analysis of specular surface reconstruction was performed by Kutulakos and Steger [2008]. Similar to our approach, they measure exit rays after reflection or refraction at a finite number of surfaces. The authors categorise the problem based on the number of camera views, surface boundaries and measurement points on exit rays. In addition to reconstructing glass objects (Figure 2.5(c)), they prove that such reconstructions cannot be performed when rays cross more than two specular interfaces. However, this argument applies only to methods based purely on ray-triangulation – tomography relaxes this constraint and is discussed further in Section 2.4.

2.3 Schlieren Imaging

The three primary optical tools used in the study of optical density are, in order of increasing sensitivity: shadowgraphs, Schlieren, and interferometry [Weinstein, 1993]. Our focus here is on Schlieren, although the particular implementation bears strong resemblance to shadowgraphy, as described in Section 2.3.1. Interferometry employs superposition of electromagnetic waves to measure displacements at the wavelength scale – small enough to render the infinite frequency approximation invalid and therefore not discussed any further here.

Despite their status as a niche domain for acquisition in computer graphics, refractive media have in fact been studied for a long time. A dedicated scientific instrument, the Schlieren imaging system, is centuries old with some of the first uses being recorded by Robert Hooke (1635-1703). Significant further developments were thanks to the work of August Toepler (1836-1912) and our modern theoretical understanding is largely due to Hubert Schardin (1902-1965). A remarkably thorough history of the Schlieren technique is chronicled in Gary Settles' book [Settles, 2001] which remains the most complete reference on the topic today.

Images are essential in conveying exactly what Schlieren photography is. Figure 2.6 shows some examples of easily recognisable scenes, but rendered in a way that many people have never seen. What the method does is to translate phase information, to which the human eye is insensitive, into intensity information. The images use greyscale intensity to represent ray deflection. To understand how this can be done, the prototypical Schlieren system is diagrammed in Figure 2.7. A small light source is collimated via lens or off-axis parabolic mirror, directed through the scan volume, and then refocused back down to a point (actually, a finite image of the source itself) at which an opaque cutoff filter is placed. A camera then records the light that passes the filter. The near-parallel rays between the first two lenses are mapped one-to-one to points on the image plane, which is what gives rise to the sharply focused nature of Schlieren images. In the nominal state, these rays are undeflected, and pass by the filter. However, the introduction of a refracting intrusion into the scan volume causes a deflection in the rays passing through it. Specifically, an angular deflection of θ after passage through the scan volume translates into a spatial deflection of $f \cdot \tan\theta$ at the fiter plane, where f is the focal length



(a) (b) (c)

Figure 2.6: (a) Interaction of thermal plumes from four adjacent candles and (b) a hot clothing iron. Image used with permission of Andrew Davidhazy. (c) Alcohol vapour rising from a shot glass. Image used with permission of Kasi Metcalfe.

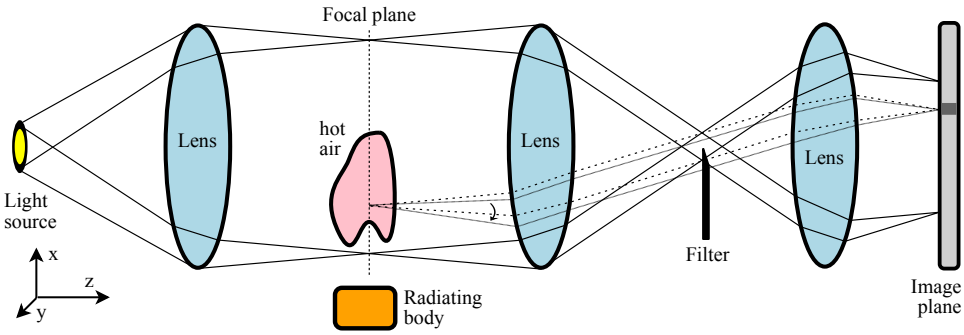


Figure 2.7: Classical lens-based Schlieren system using a point light source. The camera records an in-focus image of the refracting object placed in the scan volume. Angular variations in light rays are converted optically to intensity variations in the image. Redrawn from [Settles, 2001].

of the decollimating lens. Due to the shift, these rays are then partially occluded by the carefully-positioned filter, resulting in less light reaching the camera and a correspondingly darker spot on the image.

The basic system described here has been modified in various ways by many people. A standard horizontal razor blade filter will affect only the vertical component of ray deflection, so circular apertures have been used to used to render gradient magnitude images. In place of the binary aperture, some have used colour

graduated filters to produce *Rainbow Schlieren* [Howes, 1984], allowing one to capture the directional component of ray deflection in the hue. Graduated filters are to be preferred in general, since they reduce diffraction effects. Grid-based systems break free of the size restriction imposed by the high cost of Schlieren-grade optical components [Weinstein, 1993]. Irrespective of the particular implementation, Schlieren systems are capable of producing beautiful images, providing useful qualitative insights, and with sufficiently careful calibration, even quantitative estimates of refractive index for media defined by analytic distributions [Howes, 1984].

Digital processing of Schlieren data was prompted by the rapid rise in computational power over the past decade. Today, the BOS method [Meier, 2002] removes the need for delicate optical components, replacing them instead with a simple background pattern and a computer (see Figure 2.8). A reference image of the undisturbed background is compared with each image captured while the refracting object is present. Optical flow [Horn and Schunck, 1981; Lucas and Kanade, 1981] then provides displacement vectors in the background plane representing the deflection of each camera ray. Ambient illumination is used, rather than requiring a controlled darkroom as with classical Schlieren. The price paid for these gains in practicality is that images are now no longer always in focus. Instead, the user must employ a very bright light so as to reduce the camera aperture as much as possible, and position the components carefully to maximise depth of field in the scan region while keeping the background in focus.

Despite the practical difficulties, the method is in fact quite workable and we have demonstrated its use in capturing 3D reconstructions of turbulent gas flows [Atcheson et al., 2008]. This differs significantly from previous work in refractive gas flow tomography in that we captured time-varying fully general flows. When limited by having only one camera, previous authors made simplification in that only stationary [Schwarz, 1996] or else rotationally symmetric [Agrawal et al., 1999] flows were considered. These allow for emulation of multi-view acquisition with only a single camera, taking multiple images over time, or assuming that all views will be the same. Restricted inputs (symmetry) also allow for analytic expression of ray trajectories and a reduction in the number of model parameters, simplifying the reconstruction [Agrawal et al., 1999].

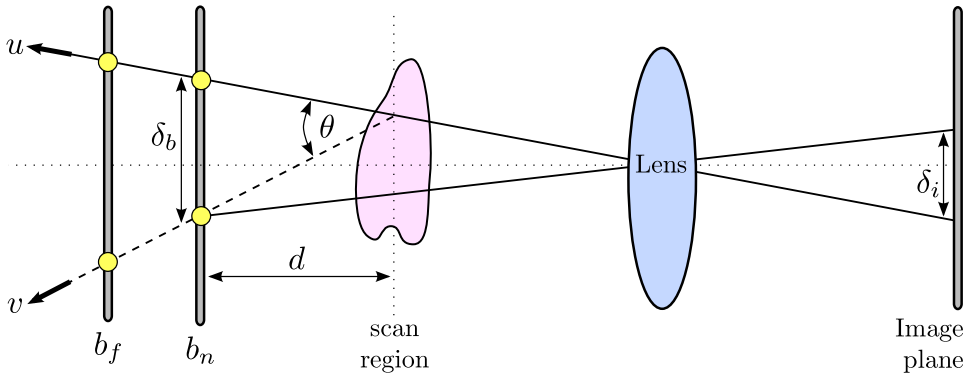


Figure 2.8: Background Oriented Schlieren ray diagram.

Wetzstein et al. [2011] further improved the BOS technique by manufacturing a lightfield [Levoy and Hanrahan, 1996] probe that could be placed behind the target object. Once the microlens array is produced and bonded to the pattern, this makes data capture significantly easier. Similar to Rainbow Schlieren, it requires precise radiometric calibration before quantitative results can be obtained. Nevertheless, for poured water scenes, it significantly outperforms optical flow-based approaches.

Optical flow remains an active research topic in the computer vision community [Baker et al., 2007]. Despite the existence of some Schlieren-specific optical flow methods [Agarwal et al., 2004], for strongly refracting glass objects with sharp edges optical flow performs poorly. Translucency poses addition problems, due to violation of brightness constancy. Use of gradient constancy instead can improve matters [Brox et al., 2004], but does not necessarily produce better results [Atcheson, 2007].

2.3.1 Shadowgraphy

A shadowgraph is quite literally the shadow of a refracting medium. Rays emanating from a bright point light source can be redirected by refraction. This leads to patterns of light (to which a net surplus of rays have been directed) and dark (from where the rays came) in the shadow. While not particularly useful (at present) as a scientific imaging technique, we mention it specifically for its similarity to BOS so

as to avoid confusion. On the distinction between true Schlieren and shadowgraph methods, Settles [2001] draws the line in terms of both the method's simplicity and the measured quantity. In the author's opinion, only the latter is useful as a defining characteristic, since simple methods have become available that also capture first derivative information [Atcheson, 2007].

In Schlieren images, the measured quantity is proportional to the gradient of the refractive index through which the ray passes. In direct shadowgraphs, a uniform gradient throughout the medium would result in all rays being uniformly shifted laterally, and hence no observable distinction between that object and one of zero gradients. Hence it is the *change* in gradients that produces shadows and so the measured quantity responds to the second derivative. The optical configuration of the BOS method is more similar to shadowgraphy than Schlieren, however in the case of a uniform lateral ray shift it is actually able to record nontrivial data, and thanks to calibration data can produce deflection angles proportional to first derivatives.

2.4 Tomography

Tomography is the process of reconstructing an N-dimensional image from a set of its (N−1)-dimensional projections. In X-ray tomography one rotates a linear emitter array around the target in opposition to a linear detector array. For each ray, the intensity I_0 at the source is known, and a measurement of the intensity I at the detector after passing through the scan volume is taken. Under an exponential absorption model the following equation describes the process [Iyer and Hirahara, 1993]:

$$I = I_0 \cdot \exp\left(-\int_\Gamma g(s)\,\mathrm{d}s\right). \tag{2.29}$$

The scan volume is thus parametrised by its absorption coefficient. Taking logarithms leads to the form shown in Table 2.1. X-rays are useful because they do not refract (simplifying the reconstruction) and because they have direct medical applications. The general technique applies to many different domains and wavetypes. The differences between radiological and seismic tomography are listed in Table 2.1 [Iyer and Hirahara, 1993], to which the optical refraction tomography of

	Radiology	Seismology	Optical Refraction
Equation	$\ln \frac{I_0}{I} = \int_\Gamma g(s)\,\mathrm{d}s$	$T - T_0 = \int_\Gamma \frac{1}{v(s)}\,\mathrm{d}s$	$\mathbf{d} - \mathbf{d}_0 = \int_\Gamma \frac{\nabla n(s)}{n(s)}\,\mathrm{d}s$
Unknown	g = absorption	v = speed	n = index
Measure	I = intensity	T = time	$\mathbf{d}$ = direction
Ray path	straight line	multiple curves	single curve
Type	controlled	un/controlled	controlled
Sources	many	single/few	many
Detectors	single	many	single

Table 2.1: Comparison between different types of tomography. Seismic sources are natural or artificial earthquakes, while detectors are geophones. X-rays and camera rays are emitted from dense 1D/2D arrays and detected in a single plane.

concern in this thesis has been added for comparison.

Various reconstruction algorithms are available. The most efficient are based on the Fourier Slice Theorem, which states that the Fourier Transform of a projection of an image (volume) is equal to a linear (planar) slice of the Fourier transform of the original image (volume) [Mersereau and Oppenheim, 1974]. In this setting the orthogonal projection operator is onto an (N−1)-dimensional hyperplane through the origin, parallel to the slice. A practical implementation of this theorem leads to the filtered backprojection methods, which account for the difference in sampling between the volume's Cartesian grid and the rotated planes [Kak and Slaney, 1988]. Unfortunately, limitation to orthographic, or restricted fan-beam ray configurations often renders this approach unusable, and so the alternative Algebraic Reconstruction Technique (ART) was developed.

ART is merely a rediscovery of the Kaczmarz method for iteratively solving a system of linear equations [Kaczmarz, 1937] (also known as Projection Onto Convex Sets (POCS) in signal processing [Strohmer and Vershynin, 2007]. Given such a system $A\mathbf{x} = \mathbf{b}$ one can converge to a solution using the update equation

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \left(\frac{\mathbf{b}^{(i)} - \left(\mathbf{a}^{(i)} \cdot \mathbf{x}_k\right)}{\left\|\mathbf{a}^{(i)}\right\|^2} \right) \mathbf{a}^{(i)} \tag{2.30}$$

for $\mathbf{a}^{(i)}$ the i^{th} row of A and $\mathbf{b}^{(i)}$ the i^{th} element of $\mathbf{b}$. The original method iterates over the rows of A in sequence (i.e., $i \equiv k \,(\mathrm{mod}\, m) + 1$) but better performance can be obtained by randomising the order of rows to focus more on linear independence amongst them [Strohmer and Vershynin, 2007]. Kak and Slaney [1988] illustrate it with the visualisation in Figure 2.9, using the undamped case of $\lambda = 1$. The algorithm is known to converge as long as $0 < \liminf_{k\to\infty} \lambda_k \leq \limsup_{k\to\infty} \lambda_k < 2$ [Herman et al., 1978]. At each iteration, the current solution is projected onto the next linear equation. In ART each row of the matrix A corresponds to the trajectory of a ray path through the medium, with values describing the contribution each voxel makes towards the final absorption value. This flexibility allows for more general ray paths than with filtered backprojection, although tracing rays and storing the matrix can be expensive. Matrix-free solvers can be used if one is able to implement the operations $A\mathbf{x}$ and $A^T\mathbf{b}$ efficiently (tracing rays and smearing residuals back into voxels along the ray path, respectively). Variations on this theme include Simultaneous Iterative Reconstruction Technique (SIRT) and Simultaneous Algebraic Reconstruction Technique (SART) [Kak and Slaney, 1988]. In SIRT, instead of updating $\mathbf{x}$ at each iteration, one instead computes projections onto all of the equations and uses their average as the next iterate. This converges more slowly but produces better-looking results. SART is the variant most often implemented today, and is essentially equivalent to SIRT with some additional weighting applied (i.e., interpolation kernels and Hamming windows to downplay data in the periphery where fewer rays contribute) [Andersen and Kak, 1984].

2.4.1 Seismic Tomography

In studying the shallow earth (to depths of a few kilometres), geologists often employ travel-time tomography. The propagation of certain seismic waves is affected predominantly by the bulk modulus of the rock. These waves can be generated mechanically (or explosively) and detected at distances of up to 100 km away (see Figure 2.10(a)). Similar acquisition methods are also used in ocean acoustic tomography [Munk et al., 2009]. As per Fermat's Principle, the ray normal to the wavefront will travel along the shortest path underground. This path is strongly curved, meaning that inversion algorithms must explicitly take into account the ray

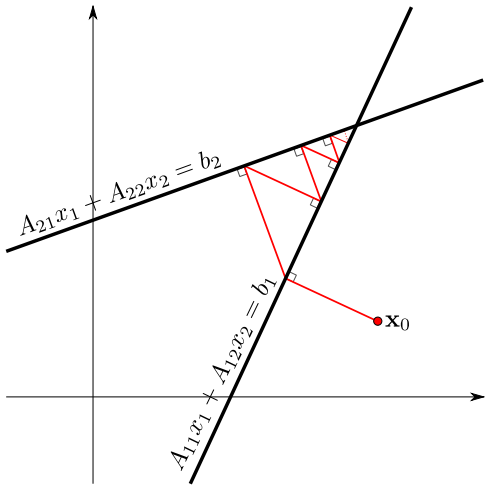
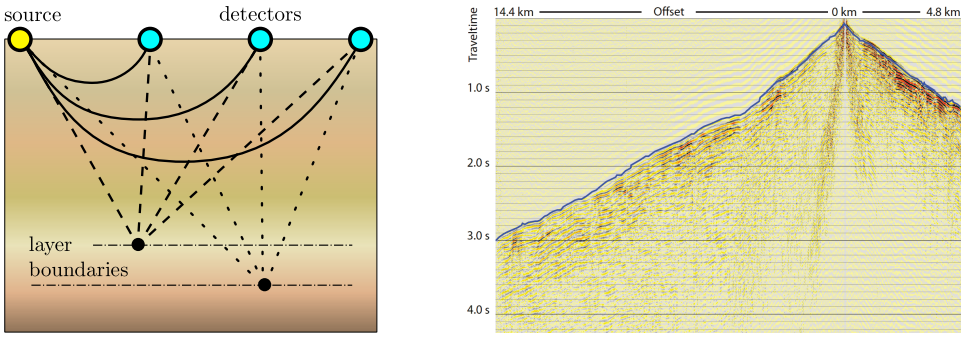


Figure 2.9: Schematic illustration of Kaczmarz (ART) method in 2D. Redrawn from [Kak and Slaney, 1988].

trajectories. Although the acquired data (travel-time) is very different to that used throughout this thesis (ray deflection), the reconstruction algorithms are similar and we stand to gain by cross-pollination of ideas. Specifically, the data resolutions we obtain via relatively cheap cameras are far higher than what is practical with very expensive seismic detectors, and consequently we propose the use of stochastic optimisation which may also potentially accelerate seismic tomography.

Seismic tomography follows an analysis-by-synthesis framework that iterates a version of our "gradient field tomography" (Chapter 6) in order to minimise a misfit function between synthesised and measured data. Within the framework, nonlinear optimisation of this function, and iteratively applying ART are essentially two equivalent methods in terms of result quality [Iyer and Hirahara, 1993]. Whereas the iterative approach is conceptually simpler (iterate ART, each time using the index distribution from the previous iteration to construct the coefficient matrix), the nonlinear optimisation approach is more useful in terms of problem analysis. One requires two main components: a forward model, and an optimisation routine. The forward model produces virtual measurements under a synthetic refractive index field. Methods for tracing optical [Andersen, 1982] and seismic [Pereyra, 1992] rays are well known. Wave-based forward models are more appropriate in some



(a) **(b)**

Figure 2.10: (a) Seismic tomography schematic using layered Earth model. Solid curved lines show first arrival refracted rays. Dashed lines indicate rays arriving later, after reflecting off boundaries. (b) Time vs distance cross-correlation of seismic dataset showing multiple arrivals of the signal at each distance. Image used with permission of Brendan Smithyman.

cases, and methods like Fast Marching [Sethian and Popovici, 1999] and Eikonal Rendering [Ihrke et al., 2007] can be used. Optimisation can conceivably be performed using any number of methods, although nonconvexity presents a significant practical problem. Many algorithms require derivative information, and it is this aspect that plays a dominant role in obtaining solutions in reasonable time.

A key development in seismic tomography was the development of the Adjoint State Method [Talagrand and Courtier, 1987] in which a time-reversed wave is propagated *backwards* from the receiver to the source in order to obtain gradient information. This replaces an extremely expensive finite difference approximation with only two passes of the forward model in order to obtain both function values and gradients. The approach is similar to time-reversal imaging in which one transmits a time-reversed version of an acoustic signal in order to locate its source [Tromp et al., 2004]. Chapter 8 discusses derivatives in more detail.

Current research in seismic (and ocean acoustic) tomography is moving towards Full Waveform Inversion (FWI) in order to extract higher resolution from the large datasets [Bozda et al., 2011]. In travel time tomography, only the first arriving wavefronts are considered (excluding surface waves). This corresponds to only

the *envelope* of the collected dataset. Figure 2.10(b) shows the intermediate measurement data. The horizontal axis represents distance to a receiver, placed in a 1D line leading away from the source near the top left. Time increases downwards on the vertical axis. Each datapoint represents the cross-correlation coefficient of the source seismic event to geophone station readings. Examining a vertical slice of the data reveals that multiple copies of the signal arrive at each station at different times. The blue line indicates the envelope of the data, which is the only part of the data used in travel-time tomography. It is this first-arrival data that we measure. FWI is a promising area, but beyond the scope of this thesis.

Chapter 3

Automatic Camera Calibration

"It is said that the camera cannot lie, but rarely do we allow it to do anything else, since the camera sees what you point it at: the camera sees what you want it to see."

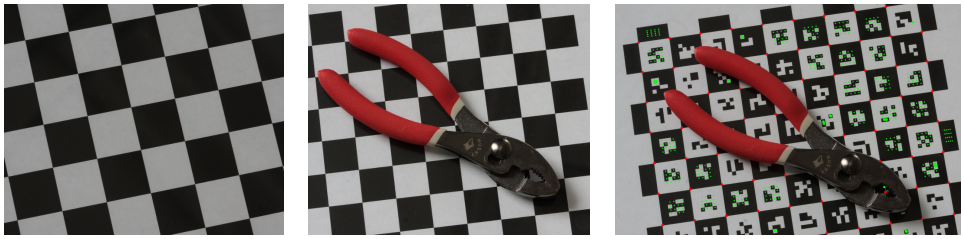
— James Baldwin (1924–1987)

In this chapter we present a self-identifying marker pattern for camera calibration, together with its associated detection algorithm. The pattern is designed to support high-precision, fully automatic localisation of calibration points, as well as identification of individual markers in the presence of significant occlusions, uneven illumination, and observation under extremely acute angles. The detection algorithm is computationally efficient and in the majority of cases requires no parameter tuning. After calibration we obtain reprojection errors up to 50% lower than with ARTag, another state-of-the-art self-identifying pattern.

The source code to a reference implementation of our algorithm has been made available at the author's web site [Atcheson et al., 2010b]. Since publication, multiple users have reported success in using CALTag for their needs.

3.1 Overview

Geometric calibration is a necessary first step in most image-based reconstruction methods. It provides an answer to the question: "given these photographs of a scene, where exactly is the camera, relative to some arbitrary world coordinate



(a) (b) (c)

Figure 3.1: Partial visibility due to (a) clipping or (b) occlusion are common failure modes of calibration methods involving chequerboards. By comparison, a system using fiducial markers such as ours (c) can easily deal with partial visibility.

system?" The typical process for calibrating cameras involves photographing a specially-designed target from multiple viewpoints and then identifying calibration points in the image that correspond to known points on the target. One of the most frequently used targets is a black and white planar chequerboard, where the calibration points are the corner points between squares. This pattern is simple to produce and allows for high accuracy because the corner points can be detected to subpixel precision [Lucchese and Mitra, 2002]. With the correspondences established across multiple images, intrinsic and extrinsic camera parameters can be obtained through nonlinear minimisation of the reprojection error [Zhang, 2000].

The difficulty with using chequerboards for camera calibration applications lies in how each internal corner point is detected and identified. Figures 3.1(a) and (b) show common failure cases for automatic chequer detection: partial visibility due to clipping against the image boundary, or to occlusion. In addition, it would be desirable to be able to simply place a scan target directly on top of a calibration pattern for stereo acquisition with a handheld camera. This is difficult with a plain chequerboard due to occlusion and shadows. For those applications the calibration points would have to be well separated from the scan object, thus reducing both the calibration accuracy and the useful image resolution for the actual target object. Manual intervention and labeling can overcome this limitation to some degree, but is cumbersome for videos or large image sequences.

An alternative to the common chequerboard are fiducial (individually identifi-

able) markers that allow for detection (and thus calibration) even if only a small percentage of markers are visible. Unfortunately for our purpose, most fiducial markers are designed with Augmented Reality (AR) applications in mind, where it is important to create isolated markers at a low spatial density. As we shall see in Section 3.5, this design compromises the precision of the marker localisation. In this work, we focus on the development of a fiducial marker system, which we dub CALTag (CALibration Tags), that provides:

- accurate localisation of calibration points using subpixel saddle point finders,
- high area density of both calibration points and markers,
- robustness under occlusion, uneven illumination, radial distortion and observation under acute angles,
- and automatic processing without parameter tweaking for convenient handling of videos and large image sequences.

As a result, our method also supports fully automatic calibration of complex multi-camera configurations where it is difficult or impossible to obtain "nice" views (i.e., ones in which each camera sees the entire calibration pattern). Although our discussion here focuses on dense, planar calibration grids, the method extends naturally to non-planar configurations. The use of individual markers in AR-style settings is possible through a separation of the marker identification and the point localisation method [Atcheson et al., 2010a].

3.2 Background and Related Work

3.2.1 Chequerboards

As mentioned, chequerboards are among the most commonly used calibration patterns, and the popular OpenCV library contains rudimentary functionality to automatically locate plain chequerboards. Since the corners of adjacent squares in a chequerboard touch each other, a saddle point finder can find the subpixel location of the calibration points with high accuracy and robustness. The disadvantage of

chequerboards is that it is almost impossible to automatically identify calibration points unless the full pattern or some other fiducial is visible.

The basic chequer pattern can be augmented with additional markers to aid in identification. For example, Yu and Peng [2006] add five double-triangles to the corners and center of a chequerboard and locate those markers using template matching. This works only when the entire board is visible in the field of view, and with their particular layout the orientation cannot be uniquely determined.

Calibration grids have also been augmented with physical devices to aid detection. House and Nickels [2006] place LEDs on the grid to allow for automatic identification of the four extremal corners, and Mohan et al. [2009] use a clever arrangement of microlenses and defocused cameras to read relative angular information from a distant target.

De La Escalera and Armingol [2010] proposed a method for automatic chequerboard detection via the Hough Transform. They first find edges in the image and seek two sets of lines with different vanishing points. This approach works well for images where the entire unoccluded grid lies in the field of view, but achieving this in practice can be a challenge.

Wang et al. [2007] also seek to automatically identify chequerboard patterns in images. They identify chequerboard corners as being those points that respond to a corner detector (Harris), as well as being simultaneously the intersection of two black squares and two white squares, and the intersection of two grid lines going through separate vanishing points. This approach works well if the grid is the dominant scene object, as is often the case in camera calibration. For our purposes however, we desire a method that can accommodate significant occlusions as well as cropping.

Mallon and Whelan [2007] analyse the impact of perspective bias and lens distortion on point detection accuracy for planar calibration grids. They find that saddle point finders are effectively bias-free, whereas methods based on circle centroids (another common calibration target) are significantly affected by both perspective and distortion.

3.2.2 Fiducial Markers

Fiducial (individually identifiable) markers have become increasingly popular in recent years. Such markers can be used in a variety of settings. Individual large-area markers are used as matrix codes (the 2D equivalent of a barcode) to encode data beyond a simple identifier [ISO/IEC 16022:2006, 2006; ISO/IEC 18004:2006, 2006]. More interesting for camera calibration are smaller fiducial markers that only encode a unique value for identification purposes. Even in this category, there are a large number of markers documented in the literature.

Some of the most common fiducial marker designs include concentric rings, where the center is the calibration point and the pattern of surrounding rings identifies the marker [Cho and Neumann, 1998; Gortler et al., 1996; Sattar et al., 2007], central dots demarcating the calibration point combined with radially arranged code patterns [López de Ipiña et al., 2002; Naimark and Foxlin, 2002], and finally rectangular patterns with identification codes in the interior [Zhang et al., 2002] such as ARTags [Fiala, 2005; Fiala and Shu, 2007]. An interesting property of the rectangular design is that every marker encodes four calibration points rather just than one. These points are usually localised by fitting lines to the edges of the rectangle, and computing the intersection points. While this approach provides better accuracy than the center-of-mass-style calculations used in many circular designs (which is subject to error from lens distortion [Mallon and Whelan, 2007], we show that it falls short of the precision provided by saddle point finders employed in chequerboard patterns and in our design.

Another shortcoming of many existing fiducial markers is that they require large areas of whitespace between them, and thus cannot be packed tightly on a calibration pattern. This is particularly true for the circular designs. However, a high density of calibration points is very desirable for camera calibration for two reasons – first, a large number of point correspondences improves the fitting results for homographies and other camera models, and second, many small markers make detection more robust under occlusion and high frequency illumination than with fewer, large markers.

Our design is based on rectangular encodings, but can be packed tightly so as to allow for both a high marker density and the use of high precision saddle point

finders. Like some other recent designs (e.g., [Fiala and Shu, 2007]), our marker IDs allow for error detection. They do not, however, provide error correction, since we anticipate CALTags will be used in larger groups, where errors may be corrected through a group majority.

3.3 Marker Design

For robustness under different lighting conditions and easy printing, we choose a binary marker design. Each CALTag marker consists of an $M \times N$ matrix of black and white squares (pixels), surrounded by a K pixel boundary that is either solid white or solid black. While we have conducted experiments with other configurations, we restrict ourselves to configurations with $M = N = 4$ and $K = 2$ for this discussion (see Figure 3.2(a)). The choice of code resolution is a tradeoff between the size of the codebook and the physical size of the pattern. Not every possible code can be used, so a small pattern limits the number of available markers and hence the number of corner points in a calibration grid. On the other hand, for the same physical marker area, smaller code patterns afford a larger printed pixel (payload dot) size δ.

Of the total $MN = 16$ code bits, we use the first $p = 10$ bits to represent the identifier, and the remaining $MN - p = 6$ bits for a checksum (CRC-6-ITU). The binary string is then simply rearranged columnwise into a 2D matrix to form the code. This allows for 2^p potential codes. However, not all of these codes can be used simultaneously, for two reasons. The first is that, in order to avoid inter-marker confusion under a bit flip, we must ensure that all rotated versions of marker codes have a minimum Hamming distance of 2 from all other used marker codes. The second reason is that patterns that are mostly white or mostly black are more likely to occur as textures or random patterns in normal images. For this reason, we choose only those codes with between 25% and 75% of their total pixels "on". This second criterion eliminates a relatively small percentage of codes in which both the data portion and the Cyclic Redundancy Check (CRC) portion together have a very one-sided intensity distribution.

The net effect of the two constraints is that, out of 1024 codes for a 4×4 grid layout, 18 codes are rejected due to the bit count constraint, and 300 codes

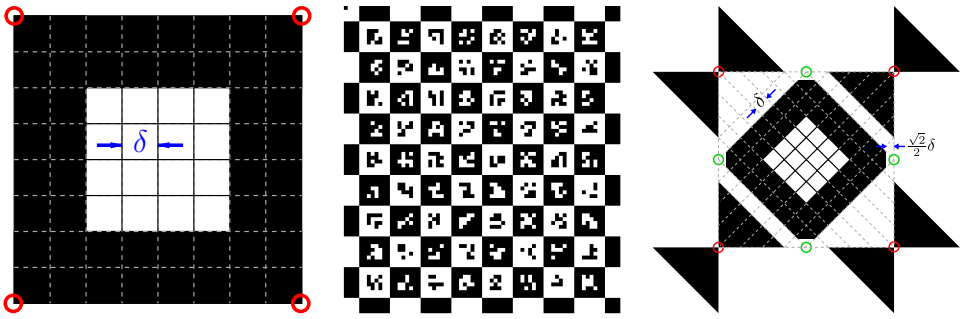


Figure 3.2: (a) Individual markers are binary 8×8 grids with a central 4×4 portion dedicated to the payload. (b) A calibration pattern is a grid of such markes, where the background colour alternates between black and white. The upper-left dot indicates the origin of the grid's coordinate frame. (c) Rotated markers are surrounded by whitespace making detection easier, but must be enlarged by a factor of $\sqrt{2}$ to maintain payload size.

due to the rotational incongruency constraint. In total, 712 codes remain to be used as valid calibration patterns. Enforcing a minimum Hamming distance of 3 under all rotations would reduce the number of codes to 247. When assembling a calibration pattern, we use all valid codes in numerical order, without further attempts to maximise Hamming distance. Due to the minimum distance, these codes allow for the detection of *any single* bit flip under *any rotation*. However, the CRC codes are more powerful. In 1D, they can also detect any burst errors (flips of subsequent bits) with burst lengths of up to 6 bits. In practice bit errors do occur in bursts because occluders, specular highlights, sensor bloom and blur are usually larger than individual dots. Although our 2D layout reduces the usefulness of this property, there are situations where this feature of CRC codes is helpful. For example, if a whole row of code pixels is occluded, the resulting pattern change can be detected. Empirically, in thousands of test images, we do not observe false-positive marker identification arising from incorrect interpretation of background scene structure.

Once the markers have been defined, they need to be arranged into a calibration pattern. We desire a dense packing of the markers to maximise both the number of

markers and the calibration points per unit area. Also, we would like to derive a layout in which the calibration points are given by local "bowtie" image topologies, in which black and white image portions touch like the corners in a chequerboard. With this kind of layout, calibration points can be localised with very high accuracy using a saddle point finder.

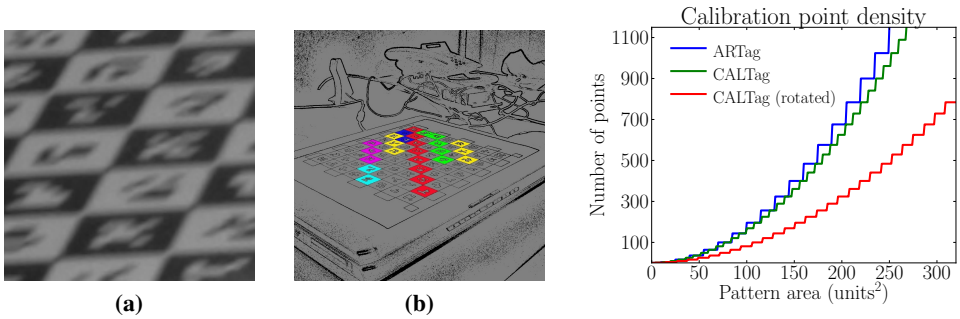
A straightforward layout that achieves these goals is to pack markers with black and white borders like the squares in an $S \times T$ chequerboard (Figure 3.2(b)). This layout optimises marker density, although detectability may suffer given the merging of different marker regions under difficult photometric conditions. Motion and defocus blur in particular can hamper marker detection. Figure 3.3(a) shows how diagonally-adjacent squares in a chequerboard can become joined. Since our detection algorithm is based on finding quadrilaterals, these chains must be broken. However, doing so reliably using only primitive image processing operations is difficult. To address this problem, we also tested an alternate design, shown in Figure 3.2(c), where each marker is rotated by 45° (see Figure 3.6(c) for the pattern layout). Because each marker is its own connected component, detection is simpler and more robust. However, we aim to maximise both the physical size of the payload as well as the total number of calibration (saddle) points. Figure 3.3(c) plots the total calibration points contained in three different patterns when cropped to the same physical size, all having payload dot sizes of one square unit.

3.4 Detection Algorithm

The detection algorithm is depicted in Figure 3.4 and described below. Beginning with the recorded image, we first find the potential markers using simple image processing techniques and some carefully chosen filtering criteria. The true markers are then confirmed by reading their binary codes. Finally, any missed calibration points are located using prior knowledge of the chequerboard layout. The output is a set of ordered 2D image coordinates corresponding to the calibration points.

3.4.1 Connected Components

The input image I is first converted from colour to greyscale, and then a Sobel filter is applied to extract edges. We then threshold the edge image to find pixels



(a) **(b)** **(c)**

Figure 3.3: (a) Motion blur causes diagonally adjacent squares to become connected. (b) Connected chains of makers form and are difficult to separate reliably. (c) Rotated markers do not suffer from this problem, but equivalent grid patterns contain fewer total saddle points.

on strong edges. The lighting may be uneven, so we use an adaptive threshold for this [Kovesi, 2000]. This mask is thinned to produce a binary edge image E which we then invert before extracting the connected components.

3.4.2 Potential Marker Identification

The previous stage outputs many more connected components than markers in the image; random background objects, as well as small segments of highly textured regions all result in components. The following two criteria are used to reject components that cannot possibly be markers:

- **Area:** We assume that each code pixel must cover an area of at least 2×2 image pixels in order to be reliably resolvable. Our markers are 8×8 units, so each one must cover at least 16^2 pixels. This lower bound often helps to remove the thousands of tiny regions that often occur in highly textured regions, such as grass or carpet. For an upper bound we use the fraction $1/\sqrt{ST}$ of the input image size, since having fewer than $\sqrt{ST}$ calibration points would would be below our minimum desired number of points (four non-collinear points are the absolute minimum necessary in order to be able to fit a planar homography).

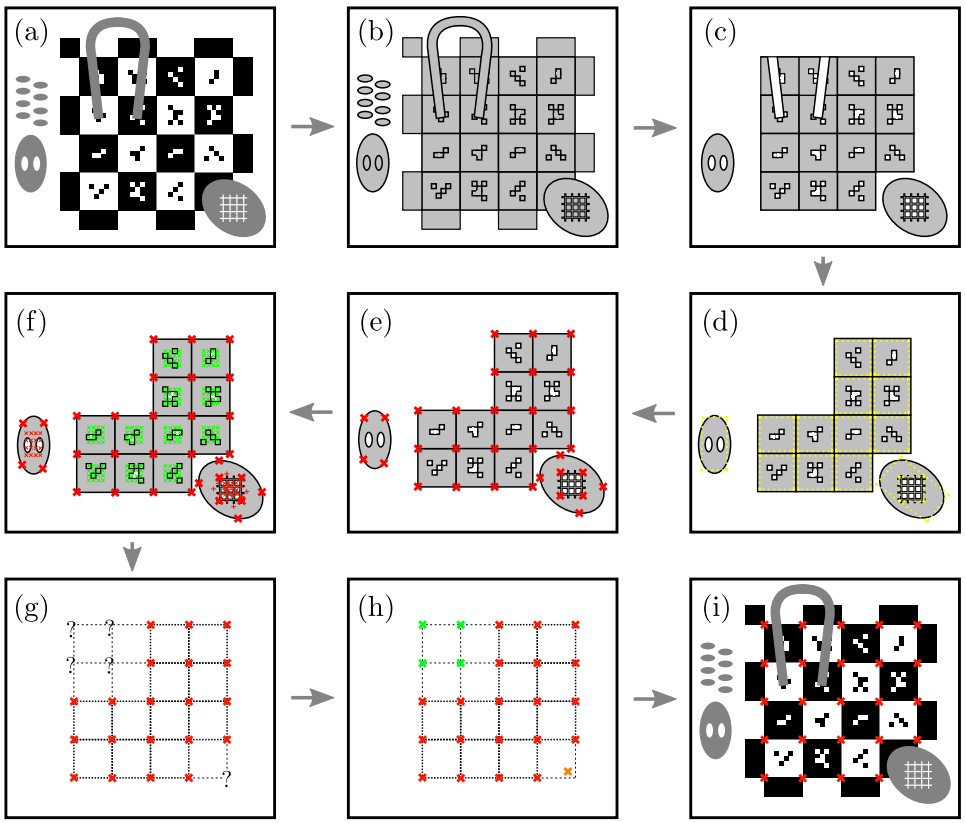


Figure 3.4: Flowchart of CALTag detection process. Given input image (a), we find connected components (b) and filter them based on area and Euler number (c). After detecting quadrilaterals (d) we find their mutual corners (e), refine them to exact saddle locations (f) and sample the payload (green crosses). After fitting the entire grid to the detected points (g) we estimate missed corner locations and search for them using the saddle finder (h). After validating the corners we obtain the final output (i).

- **Euler number**: The Euler number of an image is defined as the total number of objects in the image, minus the number of holes in those objects. Computing the Euler number for an individual connected component gives us a measure of how many interior holes there are. This calculation can be performed very efficiently [Gray, 1971]. The maximum possible number of holes would arise in the case of a marker with alternating black and white code dots, so we use a threshold of $-(MN/2)$, although in practice most markers have between 1 and 3 holes. Nested holes do not pose a problem - the entire internal code region would be considered as a separate marker, fully enclosed by the surrounding chequerboard square, and then rejected due to it having either too small an area, or an invalid binary code. The advantage of filtering based on Euler numbers is that they are resolution independent and require no parameter tweaking.

Approximate convexity was also investigated as a filtering criterion (markers are often not truly convex, due to image noise, edge detection errors and aliasing), but we found it to be expensive to compute and unnecessary given the success of the aforementioned criteria.

3.4.3 Quadrilateral Fitting

We next attempt to fit quadrilaterals to the remaining components. While the chequerboard as a whole may be distorted, the individual squares should be small enough that their boundaries can be well approximated by four linear segments. As Figure 3.6(d) shows, images with high radial distortion can be easily handled.

Quadrilateral fitting is a surprisingly poorly studied problem in the computer vision literature. The Hough Transform does not accommodate the severe perspective distortion we typically encounter in calibration images. In our approach, the first step is to trace the outline of the region, in either direction, to obtain image coordinates for the region's edge pixels. For each sample point on this boundary we compute the approximate gradient direction using central differences. These gradients are fed into Lloyd's K-Means clustering algorithm [Lloyd, 1982], with $K = 4$, to obtain the four dominant edge orientations. A least-squares line fit through each of these clusters is then used as the initial guess in finding the four boundary lines,

again via Lloyd's algorithm. At this point we have the four best fitting boundary lines (regardless of what shape the region is and how many edges it actually has) without any ordering. To extract a quad we therefore find the two most parallel lines, taking these to be opposite edges. This is sufficient in most cases to obtain a cyclic ordering of the corner points, which are themselves obtained via intersections with the other pair of lines. It is possible however, under strong perspective distortion, for opposing edges to be less parallel then two adjacent edges.

3.4.4 Saddle Points

We can now find subpixel-accurate saddle points in the greyscale image I using the same iterative algorithm as that used by Bouguet's camera calibration toolbox [Bouguet, 2004] and OpenCV [Willow Garage, 2010]. It considers all points p within a small window $\mathcal{N}_x$ around an approximate saddle point x. Nonzero image gradients only occur along edges, where they are orthogonal to the edge itself. Hence, if x is a saddle point

$$\nabla I(p) \cdot (px) = 0 \qquad \forall p \in \mathcal{N}_x. \tag{3.1}$$

This leads to a system of linear equations that can be iteratively solved for successively more accurate saddle point positions [Henrichsen, 2000]. Initial guesses are provided by the intersections of the four fitted edge lines.

There are two difficulties with applying the saddle finder to every corner point: first, it can have an impact on performance if there are many points, and second, the guesses arising from line intersections can be so poor that the corner cannot be found. But due to the layout of the markers, we know that each corner point should have up to four guesses corresponding to it, from each of the detected adjacent markers. We therefore cluster together nearby guessed corner points and consider only their centroid. Doing so provides us with an improved initial guess, eliminating the redundancy of searching for saddle points multiple times in the same image region. We use half of the average side length of the associated marker as a Euclidean distance threshold for clustering nearby points.

3.4.5 Marker Validation

At this point we have a collection of regions, most likely (although not guaranteed to be) quadrilaterals, along with four corner points for each region. Our task is to read the binary code depicted in the middle of the marker. Given a uniform square, the positions $\mathbf{c}_i$ of the code dots inside this square are known by construction of the markers. We must therefore map a unit square to the region's corners and then sample the image at the points dictated by applying the same mapping to the $\mathbf{c}_i$.

The corner points are ordered cyclically, clockwise around their centroid, but we do not yet know which point corresponds (arbitrarily) to the top left corner of the marker. All four possible orientations must therefore be considered in searching for a valid code. A 2D homography from the unit square to corner points is generated, giving us the sampling points for the code pixels in the adaptively thresholded image I. In this case, the radius of the Gaussian smoothing kernel is chosen to be three times the width of the marker. The filtering neighborhood therefore will contain enough black and white parts of the pattern that a local average can be reasonably estimated. The thresholded image is now sampled at the supposed code dot points, and converted in columnwise order to a string of binary characters.

The binary code is validated by computing the checksum of the first p bits and comparing it to the sampled checksum under all four possible rotations. Had an error-correcting coding scheme been employed, we could also correct for small errors in sampling the pattern, or for partially occluded patterns, but we found the 16 bit combination to work well enough in practice that most of the markers are detected correctly. False negatives do not pose a problem, since the chequerboard can be detected even when only a few (or potentially even just one) of the markers are correctly detected.

The markers with valid checksums are now filtered to remove any that have significantly different orientations to the others, where the orientation is taken to be the angle that the vector from the top left to the top right corner makes with the horizontal axis.

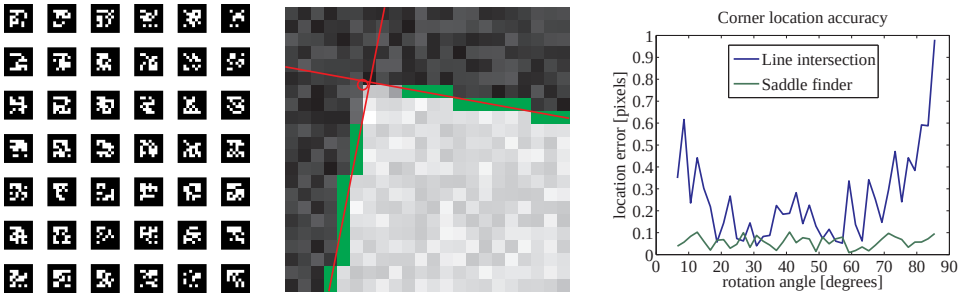
3.4.6 Locating Missed Points

In the final stage, we attempt to find any calibration points that are visible in the input image, but that were missed during detection, for example because the surrounding markers could not be identified. If at least one marker is correctly identified, then because we know where it lies in the chequerboard pattern from its ID we can guess where the remaining saddle points should lie in the image. We fit a homography to the detected points, using RANSAC this time to account for potential erroneous points, and from that obtain the approximate image coordinates of the missing points. At these points we run the saddle finder, and if it converges we add that point to the collection of calibration points.

3.5 Results

As a primary point of reference for our approach we use the ARTag markers [Fiala, 2005; Fiala and Shu, 2007], since they represent a state-of-the-art fiducial marker system and can been applied to camera calibration [Bradley et al., 2008a,b]. Like our approach, ARTags consist of a binary 2D matrix (Figure 3.5(a)). Various image processing techniques are used to locate potential markers, and then sample the interior code points to obtain a binary sequence. The sequence comprises 36 bits, 10 of which encode the marker ID, while the remainder include a CRC checksum and a Reed-Solomon error correction code.

Although ARTags can be detected and identified reliably, they are not ideal for camera calibration, primarily because the corner localisation is comparatively poor. Each ARTag marker is reported along with the positions of the quadrilateral corners. These are found by detecting edges in the image, linking them to make up quadrilaterals, fitting lines through adjacent edges and computing their intersections. Localisation of the corner is thus dependent on a line fit through pixels far away from the actual point. Since this takes place before calibration, the image edge may not be straight due to lens distortion [Mallon and Whelan, 2007]. In addition, edges cannot be both perfectly detected and localised, and so the choice of filter kernel used in edge detection could compromise the accuracy. Our method detects saddle points instead. The error plot in Figure 3.5(c) illustrates their superiority over fitting lines through quad edges.



(a) **(b)** **(c)**

Figure 3.5: (a) The ARTag calibration grid consists of separated square markers. (b) Simulated corner finding using ARTag's approach. The red circle is the true location. (c) The error becomes very high at certain angles.

Calibration point density is an important characteristic of a pattern. ARTag uses 36 bit Reed-Solomon error-corrected codes, whereas CALTag uses 16 bit error detection codes. While the larger code size and error correction ability are useful in AR applications, they do not provide additional advantages for camera calibration and instead consume space that could be used for more markers. The requirement to separate the ARTag markers by whitespace further reduces their density. However, each ARTag marker provides four calibration points, while the calibration points are shared between different markers in the CALTag system, yielding a 1:1 ratio between markers and calibration points. The net effect is that for our layout the point density is always lower, but not by much, than that of ARTag.

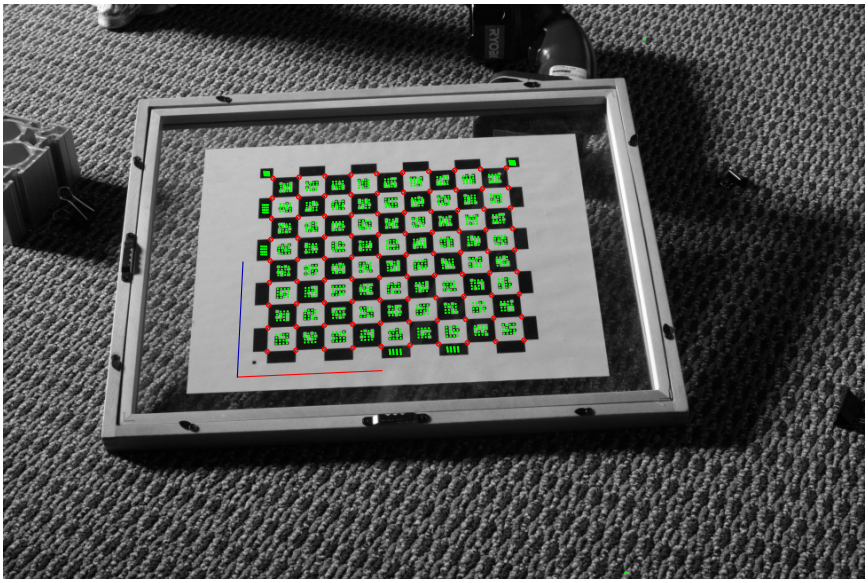
To quantify the effect of this tradeoff for increased accuracy at the cost of lower point density, we performed calibrations using both the ARTag and CALTag patterns, both having the same printed area and code pixel size. For ARTag this meant we could fit 5×6 markers (120 calibration points) whereas CALTag had 8×9 (90 calibration points). Table 3.1 shows the calibration results for ARTag vs. CALTag under a variety of different conditions. Under perfect conditions (uniform illumination, no occlusion, low radial distortion), the ARTag calibration using 4 corners per marker produced an average reprojection error of 0.918 pixels. We then tested if the reprojection error can be improved by first averaging the four

Pattern	Clear			Shadow			Occluded		
	M	P	RE	M	P	RE	M	P	RE
ARTag	180	720	0.918	180	720	0.908	138	552	0.878
ARTag (avg.)	180	180	0.394	180	180	0.403	138	138	0.368
CALTag	415	539	0.274	412	540	0.288	265	473	0.349

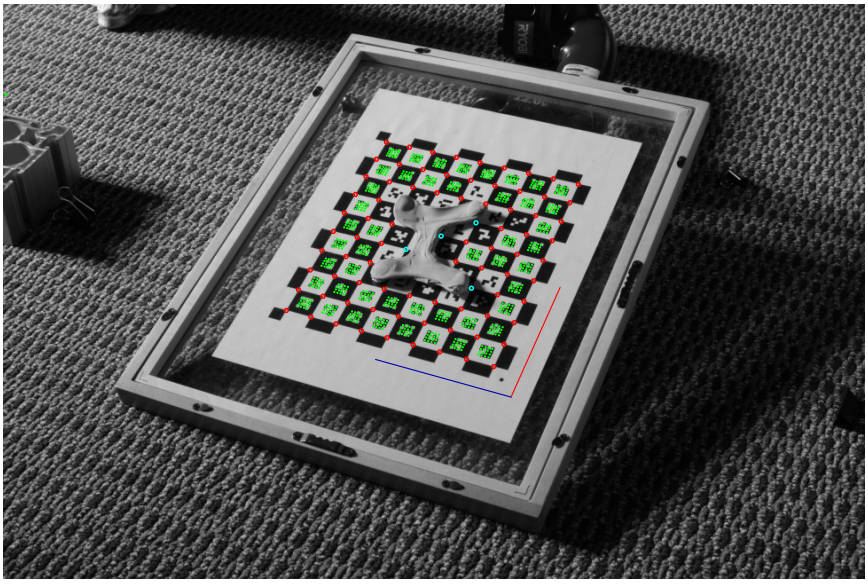
Table 3.1: Calibration results table using 6 images per setup. M = number of markers found (total across all images). P = number of points found (total). RE = mean reprojection error. ARTag pattern had 5×6 markers (180 total markers, 720 total points), CALTag had 8×9 (432 total markers, 540 total points). Normalised to same code pixel size (one unit square). ARTag area = 70×85 units. CALTag area = 72×80 units.

corner points for each ARTag marker to obtain fewer calibration points of higher precision. Doing so produced a single point per marker (30 per image) and reduced the error to 0.394 pixels. Fewer but more accurate points therefore produced a better result. However, the two variants of CALTag produce an average reprojection error of 0.274 and 0.248, respectively. Similar results were obtained in less than ideal conditions (see table). We also note that, for ARTag, in some images we had to manually select a bounding region for the marker pattern, since the system could not cope with large amounts of clutter. The CALTag results were obtained fully automatically.

Figure 3.6 shows several more results of the CALTag detection. The algorithm is successful and robust even under extremely difficult conditions. Since originally developed, CALTag has been used to calibrate camera arrays for emissive fluid tomography [Gregson et al., 2012] as well as for multiple other currently unpublished projects.

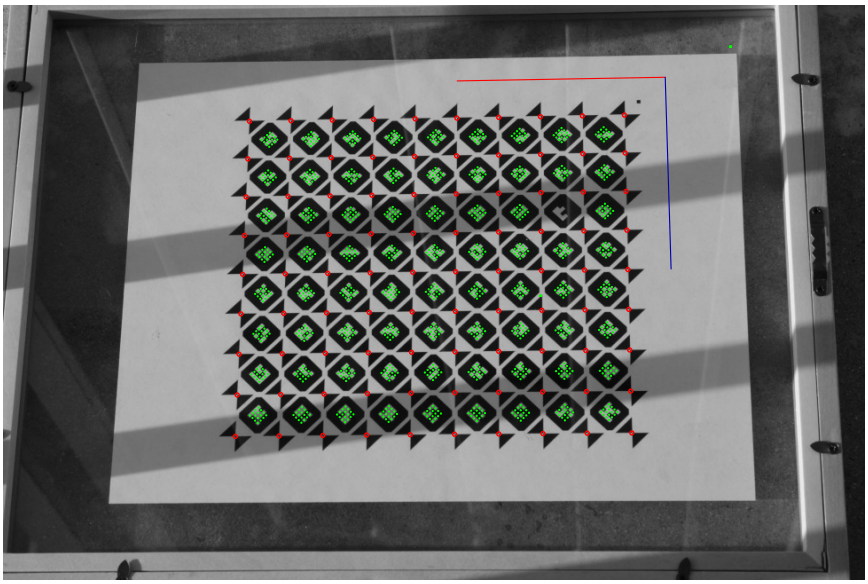


(a)

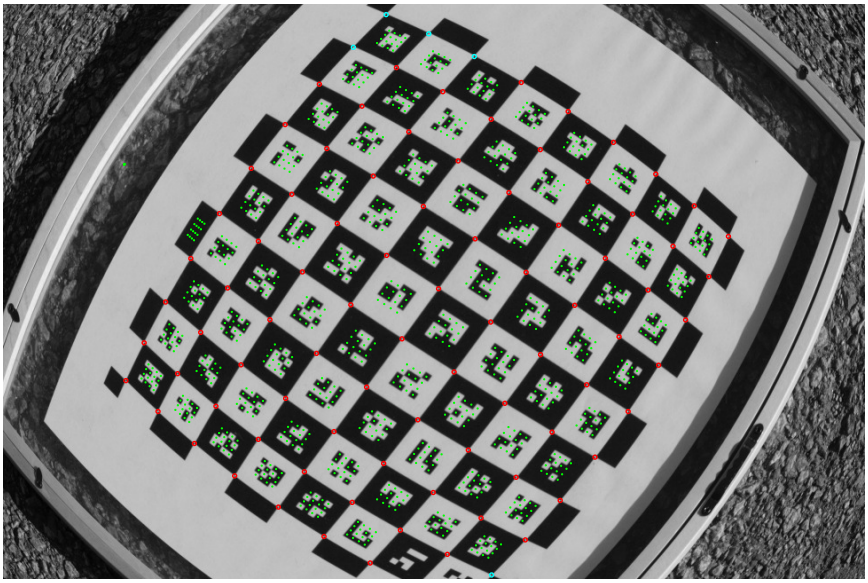


(b)

Figure 3.6: Results of CALTag detection under various conditions. Points were missed during initial marker detection but found later through the RANSAC homography are indicated in cyan.



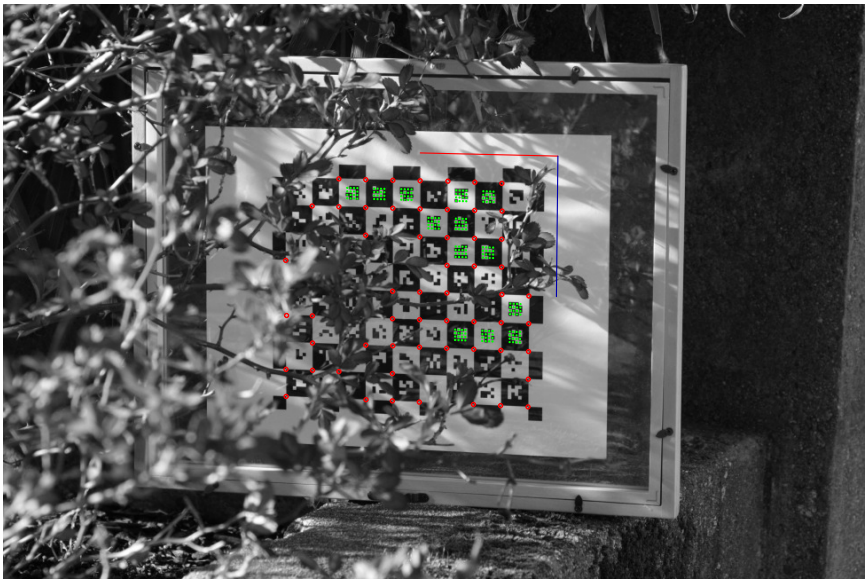
(c)



(d)

Figure 3.6: Further results. (c) Rotated markers as shown in Figure 3.2(c).

(e)



(f)

Figure 3.6: Further results.

Chapter 4

Camera Synchronisation

"From then on, Edgerton would point his flash at the world around him, using photography to discover what the unaided eye couldn't see."

— J. Kim Vandiver (1945–)

Camera arrays are very useful scientific imaging instruments with many applications [Atcheson, 2007; Bradley et al., 2008a]. In tomography and multi-view stereo settings they allow for acquisition and reconstruction of dynamic media, since multiple views can be captured "almost" simultaneously. The accuracy to which the different exposures must be synchronised depends on the time-scale of the target. In many instances, simply triggering all cameras with a common infra-red remote may be sufficient, but some applications, like the gas capture in Chapter 6 are far more demanding. Even when accurate to within a single frame (i.e., $1/30$ s at National Television System Committee (NTSC) rates) artefacts can be visible, as we shall see in Section 4.6.

One potential solution to this problem is to use better hardware. Perfectly synchronised arrays of high quality machine vision cameras can be constructed, but these are very expensive, and come with the associated nuisance of prodigious computer infrastructure. We opted instead to construct a large (16 camera) array from ordinary consumer camcorders (Sony HDR-SR7). Two major obstacles to the use of consumer camcorders in computer vision applications are the lack of

synchronisation hardware, and the use of a "rolling" shutter, which introduces a temporal shear in the video volume.

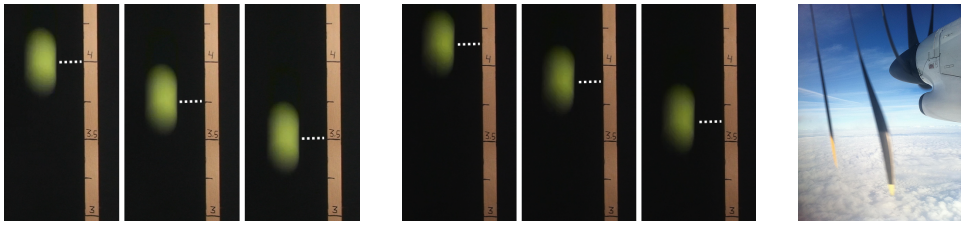
This chapter describes two simple approaches for solving both the rolling shutter shear and the synchronisation problem at the same time. The first is based on stroboscopic illumination, while the second employs a subframe warp along optical flow vectors. The resulting camera array has been successfully used in multiple configurations to reconstruct gases [Atcheson, 2007], fluids [Gregson et al., 2012], garments [Bradley et al., 2008b] and faces [Bradley et al., 2010].

4.1 Overview

Consumer camcorders are evolving as promising alternatives to scientific cameras in many computer vision applications. They offer high spatial resolution and guaranteed high frame rates at significantly reduced cost. Also, integrated hard drives or other storage media eliminate the need to transfer high-bandwidth video sequences in real-time to a computer.

These benefits must be weighed against the costs – some of which can be compensated for. The most significant loss is in control. Machine vision cameras provide Application Programming Interfaces (APIs) for adjusting nearly every parameter from shutterspeed to gamma curve to trigger offset, as well as the ability to load, save and remotely set these values. Camcorders on the other hand, tend towards automatic image adjustments and provide coarse-grained at-best controls for a scant few parameters. Nevertheless, through careful experimental procedure and controlled environments, we can coax useful data out of them.

The other significant costs are those for which this chapter provides algorithmic solutions. First, consumer camcorders typically do not have support for hardware synchronisation. Second, these cameras (as well as high-end Digital Single Lens Reflexs (DSLRs) capable of video capture, and mobile phone cameras) employ a "rolling" shutter, in which the individual scanlines use a slightly different temporal offset for the exposure interval (see, e.g. [Wilburn et al., 2004]). The resulting frames represent a sheared slice of the spatio-temporal video volume that cannot be used directly for many computer vision applications. Visually, one can see the problems in Figure 4.1 where an event is actually captured at different points in



(a) (b) (c)

Figure 4.1: Small errors in synchronisation become evident with fast-moving subjects. (a) and (b) show two closely-spaced cameras observing a ball falling next to a static ruler. The images are consecutive frames, left to right. (c) Rolling shutter artefacts can produce "impossible" images when the scene contains temporal frequencies near to, or higher than, the camera framerate. Image used with permission of Jon Page.

time, either by different cameras, or by different scanlines in the same camera.

In this chapter we discuss two different approaches for solving both problems at the same time. The first method performs optical synchronisation via stroboscopic illumination. Strobes ensure that all cameras are exposed instantaneously. Where applicable, this method solves the synchronisation problem directly. When it cannot be used directly, it nevertheless provides a useful signal against which we can synchronise using other methods. The simultaneous strobe flash also addresses rolling shutter problems, although the scanlines for a single flash are usually distributed across pairs of consecutive frames (or fields, with interlacing).

As alluded to above, the strobes cannot be used everywhere. In particular, outdoor or brightly lit scenes drown out the strobe light. In these cases we first acquire accurate per-camera temporal offsets using a strobe or other mechanism. We then use more expensive image processing (optical flow and morphing) to warp the frames into alignment, and to remove the rolling shutter shear.

In the following, we review relevant work on camera synchronisation (Section 4.2), before elaborating on the rolling shutter camera model on which we base our experiments (Section 4.3). We then discuss the details of our two synchronisation methods in Sections Section 4.4 and 4.5. Finally, experimental results are presented in Section 4.6.

4.2 Related Work

Due to their idiosyncrasies, rolling shutter cameras are not commonly used in computer vision. However, the rapid growth in the consumer camera market in recent years, as well as ever-improving image quality, has prompted some analysis and applications to appear in the literature. We demonstrate the novel use of these techniques for the realisation of low-cost camera arrays with good synchronisation characteristics.

Stroboscopic illumination has been used to capture multi-exposure images. Classic examples include early photographic work by Harold E. Edgerton and Gjon Mili to capture high-speed events on film. Lately, computer vision techniques have used this principle to recover trajectories of high speed motions, e.g., Theobalt et al. [2004] track the hand motion and ball trajectory of a baseball player. Linz et al. [2008] recover flow fields from multi-exposure images to generate intermediate single exposure views and synthetic motion blur.

4.2.1 Rolling Shutters

Wilburn et al. [2004] use an array of rolling shutter cameras to record high-speed video. The camera array is closely spaced and groups of cameras are hardware triggered at staggered time intervals to record high-speed video footage. Geometric distortions due to different view points of the cameras are removed by warping the acquired images. To compensate for rolling shutter distortions, the authors sort scanlines from different cameras into a virtual view that is distortion free. Ait-Aider et al. [2007] recover object kinematics from a single rolling shutter image using a-priori knowledge of straight lines that are imaged as curves.

Although there are hardware solutions for the Complementary Metal Oxide Semiconductor (CMOS) rolling shutter problem [Wäny and Israel, 2003], these are often not desirable since the transistor count on the chip increases significantly, which reduces the pixel fill-factor of the chip. Lately, camera models for rolling shutter cameras have been proposed, taking camera motion and scene geometry into account. Meingast et al. [2005] develop an analytic rolling shutter projection model and analyse the behaviour of rolling shutter cameras under specific camera or object motions. Alternatively, rolling shutter images can be undistorted in soft-

ware. Liang et al. [2005, 2008] describe motion estimation based on coarse block matching. They then smooth the results by fitting Bézier curves to the motion data. The motion vector field is used for image compensation, similar to our approach described in Section 4.5, however we perform dense optical flow and extend the technique to a multi-camera setup to solve the synchronisation problem as well. Nicklin et al. [2007] describe rolling shutter compensation in a robotic application. They simplify the problem by assuming that no motion parallax is present.

Since publication of our method, Baker et al. [2010] have addressed rolling shutter wobble by assuming that the camera is undergoing high frequency jitter and using temporal superresolution techniques to render the corrected image.

Some authors have considered the use of externally-acquired information in addition to image analysis. In particular, Karpenko et al. [2011] use the gyroscopes available on commodity mobile phones to estimate the camera motion and compensate for distortion.

Grundmann et al. [2012] have also sought to simultaneously remove rolling shutter distortions alongside a related defect i.e., camera shake. They robustly track feature points across frames and assume that each scanline undergoes motion described by a homography. Their approach allows for near real-time correction of video without any prior calibration. Our approach in contrast is computationally very intensive and operates on a per-pixel level rather than exploiting similarities amongst pixels within a common scanline (continuity can be enforced by smoothing).

4.2.2 Multi-View Synchronisation

Wang and Yang [2005] consider dynamic light field rendering from unsynchronised camera footage. They assume that images are tagged with time stamps and use the known time offsets to first compute a virtual common time frame for all cameras and afterwards perform spatial warping to generate novel views. Camera images are assumed to be taken with a global shutter.

Computer vision research has been concerned with the use of unsynchronised camera arrays for purposes such as geometry reconstruction. For this it is necessary to virtually synchronise the camera footage of two or more independent cam-

eras. Most work in this area has so far assumed the use of global shutter cameras. The problem of synchronising two video sequences was first introduced by Stein [1999]. Since Stein's seminal work, several authors have investigated this problem. Most algorithms are based on some form of feature tracking [Caspi et al., 2006]. Often, feature point trajectories are used in conjuction with geometric constraints relating the cameras like homographies [Dai et al., 2006; Stein, 1999], the fundamental matrix [Carceroni et al., 2004; Sinha and Pollefeys, 2004] or the tri-focal tensor [Lei and Yang, 2006]. The algorithms differ in how the feature information is matched and whether frame or sub-frame accuracy can be achieved. Most authors consider the two-sequence problem, but N-sequence synchronisation has also been considered [Carceroni et al., 2004; Lei and Yang, 2006].

A different approach to N-sequence synchronisation has been proposed by Shrestha et al. [2006]. The authors investigate the problem when given video sequences from different consumer camcorders recording a common indoor event. By assuming that in addition to the video cameras, the event is being captured by visitors using still cameras with flashes, they propose to analyse flash patterns in the different video streams. By matching binary flash patterns throughout the video sequences, frame-level synchronisation can be achieved.

4.3 Camera Model

Both of our synchronisation methods target inexpensive consumer-grade video cameras and camcorders. In this market segment, there has been a recent push to replace Charge-Coupled Device (CCD) chips with CMOS sensors. There are pros and cons to both technologies, with CMOS generally employing rolling shutters. Unlike global shutters, in which all pixels begin and end their exposures simultaneously, rolling shutters trigger the exposure windows of consecutive scanlines in a staggered fashion. We aim to model this process explicitly in order to eliminate the introduced distortion, as well as to exploit it in order to obtain very accurate synchronisation events.

One reason for staggering the readout is to reduce the expensive buffer memory required to hold the data coming off the sensor and awaiting processing and compression. The exposure window for each scanline may begin at any point up

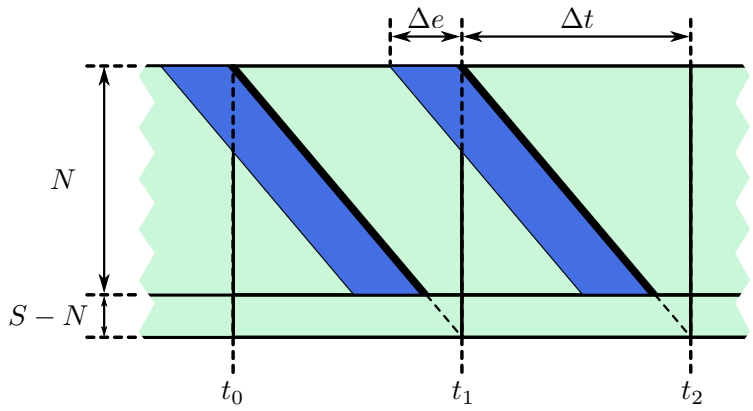


Figure 4.2: Rolling shutter camera model. Just-in-time exposure and readout of the individual scanlines (vertical axis) creates a shear of the exposure intervals along the horizontal time axis. The slope of this shear is a function of the frame rate and the period is determined by the number of scanlines in the video format.

to $1/\Delta t$ sec before the readout. Maximising the exposure duration ensures that a scene event will be captured by every scanline. This is important when setting off a brief flash to serve as a synchronisation trigger. Such a flash would result in a dark-to-light transition line appearing somewhere in the image, where the higher scanlines finished exposure before the flash and so appear darker than those below that recorded the flash. Each scanline's window is of the same duration, and they begin one after the other from top to bottom, starting again with the topmost scanline after the bottom one begins. Effectively this means that the lower portions of the frame record events that occurred after those in the upper portions of the frame. A time-vs-scanline diagram of the processes is shown in Figure 4.2. Specifically, we can model the readout time $r_j^{(y)}$ for scanline y in frame j as follows:

$$r_j^{(y)} = t_j + \frac{y}{S}\Delta t \tag{4.1}$$

$$= t_0 + \left(j + \frac{y}{S}\right)\Delta t, \tag{4.2}$$

where Δt is the frame duration (one over the frame rate), S the total number of scanlines per frame, and t_j the readout time of the topmost (visible) scanline in

frame j. Readout duration is effectively instantaneous for our purposes. The exposure interval for scanline y in frame j is then given as

$$E_j^{(y)} = \left[r_j^{(y)} - \Delta e,\, r_j^{(y)} \right], \tag{4.3}$$

where Δe is the duration of exposure (exposure time).

Note that the total number S of scanlines may be larger than the number N of *visible* scanlines. For example, the specification for high definition video [ITU, 2002] calls for $S = 1125$ total lines for video standards with $N = 1080$ visible lines. The extra 45 invisible lines correspond to the vertical synchronisation signal. Standard definition video uses 39 ($S = 525$) invisible scanlines in NTSC [ITU, 2007].

Most consumer cameras trade spatial for temporal resolution by recording the even and the odd scanlines in separate fields (interlacing). The model above still holds in this case, after halving the parameters $\Delta t, N$ and S. One must bear in mind that the y^{th} row in the even and odd fields making up a frame correspond to distinct rows of pixels on the sensor, and so do not record exactly the same content.

For synchronisation of multiple cameras, we assume that all cameras follow the same video standard, i.e., that Δt, S, and N are identical for all cameras, and that either all or none of the cameras use interlacing. These assumptions are easily met if all cameras in the array are the same model. A possible concern is the potential for slight differences in the frame rate across individual cameras. However, even inexpensive cameras appear to have very good accuracy and stability with respect to frame rate. In our experiments with up to 16 cameras and several minutes of video, per-camera differences did not appear to have a visible impact.

In practice, the camera's operating system polls the trigger in a loop waiting for input. This means that there is an arbitrary delay between triggering of the recording, and when the first scanline begins exposure. It is therefore impossible to synchronise multiple cameras even when a common trigger is used. We experimented with both Infrared (IR) and wired remotes and were only able to achieve synchronisation to within three frames on average. The onboard processing may also apply radial distortion correction, which can result in curved transition lines between light and dark regions in the strobe flash images, rather than seeing the expected perfectly straight horizontal scanlines.

4.4 Stroboscopic Illumination

Stroboscopes have long been used for obtaining instantaneous exposures of moving objects using standard cameras with global shutters (e.g., [Theobalt et al., 2004]). An extension of this approach to multi-camera systems results in multiple video streams that are *optically* synchronised through illumination. Unfortunately, this straightforward approach requires additional processing to work with rolling shutter cameras, which we address here.

With our first approach, we can solve the rolling shutter problem for individual cameras, or simultaneously solve the synchronisation and rolling shutter problems for an array of cameras, as long as the lighting in the environment can be controlled. With no ambient illumination, stroboscopes create simultaneous exposures for all cameras. However, with rolling shutters, the exposed scanlines are usually divided between two adjacent frames (or fields). In our technique, we combine two partially exposed frames to form a single synchronised exposure image for each camera. Since all scanlines are exposed by the flash at the same time, this method exhibits no temporal shear regardless of the individual read-out times.

In the single camera setting, the camera starts recording a dark scene in a normal way. Stroboscopic illumination is then activated, creating the exposures. The flash is captured by all scanlines that are exposing at the time of the flash. The number of scanlines that record the event is determined by the exposure time of the camera, Δe. We therefore ensure that all scanlines record the flash by maximising Δe (see Figure 4.3), creating a continuous exposure with respect to the camera. Due to the overlapping exposure windows of the scanlines in rolling shutter cameras, the strobe flash is usually split between two consecutive frames. The two frames containing the instantaneous exposure can be combined by summing consecutive frames, or else taking a pixelwise maximum. Note the choice of whether to pair a frame with the preceeding or succeeding one. If the incorrect choice is made, a visible discontinuity will appear in the output, similar to the tearing produced by Cathode Ray Tube (CRT) displays without properly synchonised input signals. To avoid this, one should begin recording in a darkened environment and scan the frames in temporal order to find the first nonzero one. This one should be paired with the succeeding frame. As a practical matter, be aware that even in a

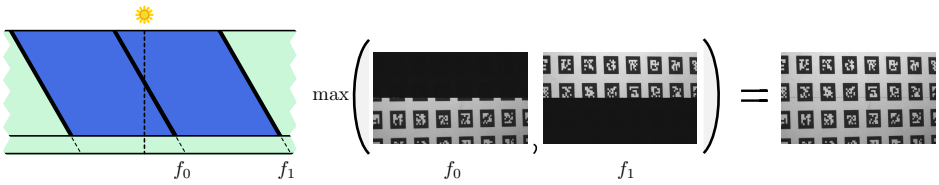


Figure 4.3: In rolling shutter cameras, consecutive frames that contain the instantly exposed scanlines are combined to make the final image.

dark room, noise and video standards will conspire to produce nonzero data in the unilluminated scanlines. Also, detecting the first "bright" scanline automatically can be difficult if the scene is only partially filled – a full-frame reflective target is recommended.

In a multi-camera setting, each camera independently captures the scene with the strobe illumination. The per-camera rolling shutter compensation as described above automatically synchronises the array.

Although the cameras record frames at a certain frame rate, the frequency of the strobe lighting can be set independently to create a virtual frame rate for the video sequence. This is because one output frame is generated for each flash of the stroboscope. The maximum frequency that avoids double-exposure of scanlines is $1/\Delta t$. However, flashing at this frequency tightly packs the instantaneous exposures with no gap of dark pixels between them. Leaving a gap between the exposures helps to separate them, especially in the case of minor drift if the stroboscope frequency cannot be set precisely to $1/\Delta t$. The simplest approach is to set the strobe frequency to half the camera frame rate, creating a full frame of unexposed scanlines between every exposure. Note that the unexposed scanlines are also split between two consecutive frames, exactly like the exposed scanlines. If this reduction in temporal resolution is acceptable, then every pair of adjacent frames can be combined in the straightforward manner described above. If a higher virtual frame rate is desired, the strobe rate can be increased. The frames can be combined automatically with a little more computational effort to explicitly search for the unexposed scanlines that separate the frames. This technique is robust to any minor drift that may occur over time, if the strobe frequency cannot be set with high precision.

An additional benefit of this method is the complete elimination of motion blur (see Figure 4.4). The price to pay is having a potentially insufficient amount of total light, resulting in underexposed images, or increased noise associated with high gain. To accommodate this, flashes of longer duration can be used. Details are ommitted from this thesis and can be found in [Bradley et al., 2009].

Our stroboscope illumination technique works by first starting the cameras in a darkened environment and only then activating the strobe lighting. The first exposure flash can then be located in each camera, identifying the first synchronised frame. After this, one can continue using the strobes to keep the cameras in sync, or else activate other lighting and simply record the first flash for synchronisation via our second method, described next.

4.5 Subframe Warping

Our second technique, while being less accurate than the previous, is applicable to more general illumination conditions. It is based on interpolating intermediate frames. Given two consecutive recorded frames, I_n and I_{n+1}, the temporal shear can be removed by interpolating or warping between the two frames using different offsets for each scanline.

Linear interpolation may work for some scenes but in general, especially with higher frequency content, better results can be obtained by morphing. We obtain optical flow vectors $\mathbf{u}(x,y) = (u,v)$ describing the displacement of a pixel (x,y) from I_n to I_{n+1}. We then warp along these optical flow vectors to create a morphed image M as follows

$$\begin{aligned} M(x,y) = (1-\alpha)\cdot I_n\Big(x+\alpha u,\ y+\alpha v\Big) \\ + \alpha\cdot I_{n+1}\Big(x-(1-\alpha)\,u,\ y-(1-\alpha)\,v\Big), \end{aligned} \tag{4.4}$$

where $\alpha = y/S$ is a blending weight that varies as a function of the scanline index. The drawbacks to using optical flow are that it is expensive to compute and can fail in textureless regions or at depth discontinuities. As long as scene motion is sufficiently slow we can often obtain good flow estimates. The result is a vertical slice through the spatio-temporal volume in Figure 4.2 at timestep $t_{(n+1)}$. In the

case of interlaced video we compute optical flow between successive fields, after shifting every second field vertically by half a scanline.

There is nothing to prevent us from shifting α by an arbitrary offset δ, which allows for multiple cameras to be synchronised if we know their relative offsets. Finding such an offset is easy if stroboscopes are available. Even in outdoor environments a strobe light can be aimed directly into the camera lens. As shown in Figure 4.3, the scanline y at which we observe a transition from a block of bright scanlines back to darker ones indicates the time at which the strobe was flashed. Assuming we have already naïvely synchronised the cameras to integer field precision, the subfield offset (in seconds) between cameras C_p and C_q is

$$\left(\frac{y_p - y_q}{S}\right)\Delta t \tag{4.5}$$

Dividing this by Δt gives the offset $\pm\delta$ (depending on the ordering of cameras). Note that if $\delta \neq 0$ then when computing M, $|\alpha + \delta|$ will exceed 1 for some scanlines. In this case we have stepped across into the next sheared slice of the volume and have to work with I_{n+1} and I_{n+2} (or I_{n-1} and I_n) instead.

If stroboscopes are not available, then temporal offsets can be obtained via other means, such as by filming continuous periodic motion and detecting trajectories [Carceroni et al., 2004].

4.6 Experiments

For our experiments, we use up to 16 Sony HDR-SR7 camcorders. These camcorders follow the 1080i/30 format [ITU, 2002]. That is, video is recorded at 29.97 frames per second (approximately 60 fields per second interlaced), and each frame has a final visible resolution of 1920×1080. Like many consumer devices, the video is recorded in *anamorphic* mode, where the horizontal direction is undersampled by a factor of $4/3$, meaning that each frame is represented as two fields with a resolution of 1440×540.

For the stroboscope illumination experiments with instantaneous exposure, we use three hardware-synchronised Monarch Instrument Nova-Strobe DAX stroboscopes. This model allows very precise flash rates (between 30 and 20,000 flashes

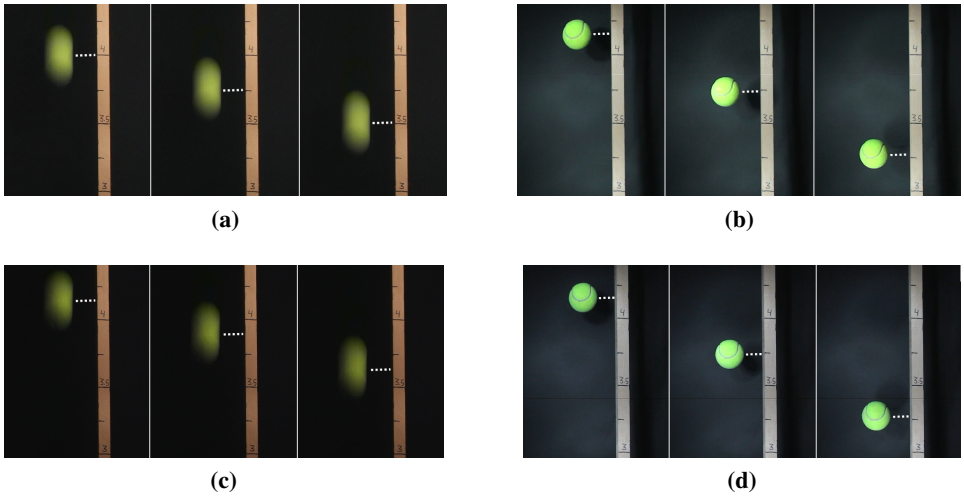


Figure 4.4: Synchronisation of two closely-spaced cameras via stroboscope. Three consecutive frames (left to right) in the first camera (a) capture the falling ball at an earlier time than the other camera (c) in the unsynchronised case. The same cameras become synchronised under stroboscopic illumination in Figures (b) and (d).

per minute) and short flash durations (20 µs). We use multiple spatially distributed strobes instead of just one in order to increase the intensity and uniformity of the illumination.

4.6.1 Synchronisation

Our first method for synchronisation is demonstrated with the example of a falling ball, in Figure 4.4. Two cameras observe a tennis ball falling beside a measuring stick. On the left side we show three consecutive frames for each camera, captured with regular, constant illumination. The ball is falling quickly, so the images contain motion blur. We measure the height of the ball at the center of the blur, as indicated by the dashed white line. It is clear that the cameras are not synchronised. On the right side of the figure, we show the same example using stroboscopic illumination. Measuring the height of the ball demonstrates the precise optical synchronisation. Note also that this method avoids motion blur, since the frames are captured at instantaneous moments in time. This benefit allows us to accurately

(a) (b) (c)

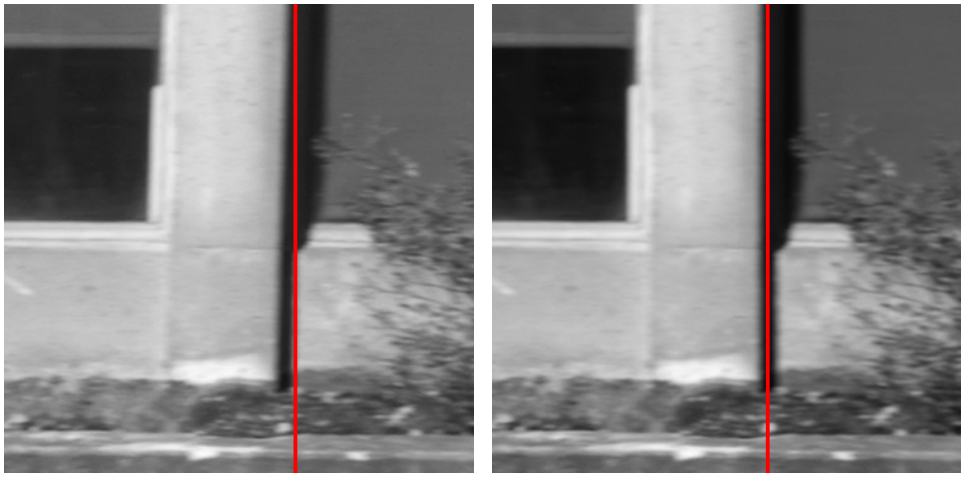
Figure 4.5: Subframe warping synchronisation. (a) Two consecutive fields from first camera, superimposed onto one image. (b) Closest integer frame aligned field from second camera. (c) Warped field from first camera, synchronised to match the second.

capture very fast motions.

Note that the amount of motion between consecutive frames in the synchronised example is roughly twice that of the unsynchronised case, since the strobe frequency was set to half the camera frame rate as discussed in Section 4.4.

In multi-camera setups where stroboscopic illumination cannot be used, we can still perform synchronisation via the subframe warping method. Figure 4.5 shows a rotating arrow filmed from two different cameras. A single stroboscope flash was used to obtain the relative time offset.

We tested framerate stability and consistency for all cameras in our array. The cameras were arranged in a semicircle and pointed at a diffuse target in the center. The ball was illuminated by stroboscopes set to the NTSC frame rate of 29.97 Hz. As we discussed in Section 4.3, flashing at this rate results in a split image where the scanline at which the split occurs should be stable over time. If either the camera frame rate or the strobe were exhibiting temporal drift, the split scanline would move up or down, indicating a mismatch in illumination and recording frequency. While observing video footage of the 16 cameras recorded for more than two minutes, we did not see temporal drift in any camera. Since all cameras were observing the same strobe signal this indicates that framerates are very stable across different cameras of the same model.



(a) (b)

Figure 4.6: (a) Original frame from a handheld panning sequence. The red line shows how the vertical wall is displaced by as much as 8 pixels in the lower portion of the image. (b) Corrected image after warping.

4.6.2 Rolling Shutter Compensation

Any motion orthogonal to the rolling shutter's direction results in a warping of straight lines. Similarly, vertical motion results in stretch. Static scenes are obviously unaffected by the rolling shutter, whereas too fast a motion causes blur that somewhat hides the distortion. However, at reasonably quick handheld panning speeds, the distortion can be quite severe, as shown in Figure 4.6. Since the wall edge covers many scanlines, there is a relatively long time difference between when the top and bottom of it are captured. Horizontal edges by contrast are not affected to such a large degree. The rotating chequerboard in Figure 4.7 shows how straight lines are rendered as curves under a rolling shutter.

Our stroboscopic illumination technique completely avoids distortion caused by rolling shutter cameras since all scanlines in a frame are exposed simultaneously by the strobe lighting. This is demonstrated in Figure 4.7(d). Despite the fast motion, straight lines are captured correctly.

Figure 4.7 also shows our rolling shutter correction results for continuous lighting. Here, we used Horn-Schunck optical flow to compute the warped image. Note

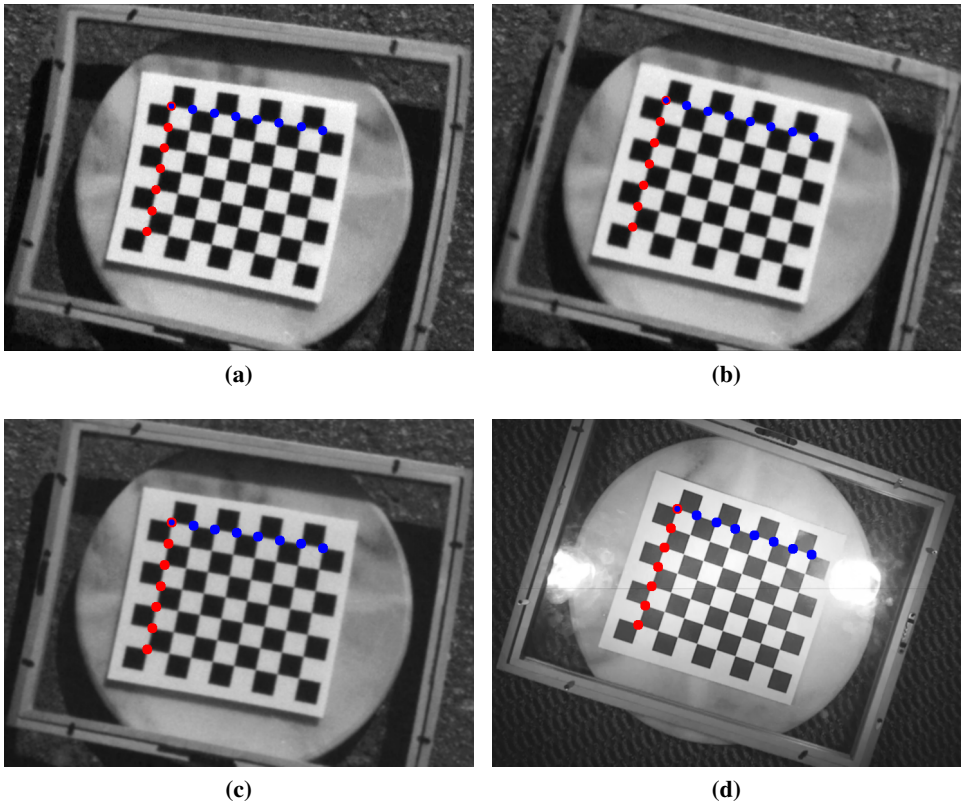


Figure 4.7: (a) A fast rotating chequerboard (1 revolution per second) in which straight lines are warped into curves. (b) After rolling shutter correction by subframe warping. (c) Slower motion (half the speed) still produces noticeable distortion. (d) The same scene captured with stroboscopic illumination.

that problems can therefore occur in scenes containing occlusions. We were able to correct the fast-moving chequerboard to match the undistorted lines captured in the *stationary* and *stroboscope* cases.

Table 4.1 contains the residual vector error (L_2) as well as the worst case perpendicular distance (L_∞) for the (approximately) vertical and horizontal rows of indicated corner points to straight line fits. As we can see, both approaches effectively remove the rolling shutter distortion.

	Vertical		Horizontal	
	L_∞	L_2	L_∞	L_2
Fast-moving	1.38	2.55	0.19	0.29
Slow-moving	0.57	0.98	0.20	0.37
Stationary	0.10	0.15	0.11	0.19
Subframe warp	0.08	0.14	0.14	0.28
Stroboscope	0.10	0.18	0.09	0.17

Table 4.1: Distortion metrics for lines warped by rolling shutter. Values show norms of perpendicular distance error vectors for the indicated corner points to straight line fits. The vertical line (red), which crosses many scanlines, is more distorted than the horizontal (blue) line.

4.7 Conclusion

We have explicitly modelled the cause of the spatio-temporal distortions in video sequences generated by rolling shutter cameras. Given this, we have presented two methods to compensate for these distortions as well as to synchronise multiple cameras. The first uses stroboscopic illumination. This method requires active illumination, limiting its applicability to relatively small-scale indoor environments. The approach also results in a loss of frame rate and potentially increased camera noise. On the other hand, the method is very easy to set up, and the post-processing can easily be performed in real-time. It is therefore ideally suited for on-the-fly processing of video streams. Since the stroboscope can eliminate motion blur, it naturally suits scenes with fast motion that require more accurate synchronisation. With a controllable virtual exposure time we allow a trade-off between motion blur and camera noise.

The second method on the other hand, requires computationally more expensive optical-flow-based morphing, and is thus best suited for offline processing. As is typical for optical flow methods, it can fail when the scene has little high frequency structure or excessive motion, and it can result in distortions and other artefacts near occlusion boundaries. The key advantage of this approach is that it does not require active lighting, and is thus ideal for outdoor and other large-scale environments. No additional hardware is required beyond the camera(s).

One thing to note is that in addition to synchronising and undistorting raw cap-

tured images, one may also work with computed results. For example, in our gas reconstruction experiments (Chapter 6), the raw data contained very high frequency noise patterns. Warping this data adversely affected the reconstruction quality. Instead, we opted to perform 2D processing using unsynchronised and distorted data, giving sheared models. However, these models had the benefit of being smooth and easier to warp without introducing objectionable artefacts. In this case, we recorded the temporal offsets using stroboscopes before capture, reconstructed as usual, and then applied the synchronisation and undistortion to our generated results as a post-process.

We have used both methods presented here in several capture projects, including time-resolved reconstruction of non-stationary gas flows [Atcheson et al., 2008], multi-view stereo reconstruction of video sequences [Bradley et al., 2008a], time-varying deformable object capture [Bradley et al., 2008b] and emissive fluid tomography [Gregson et al., 2012]. Together, they enable the use of rolling shutter cameras in a large range of computer vision applications. With these issues resolved, we believe that consumer camcorders are very suitable for camera arrays due to their low cost, guaranteed frame rate, and easy handling. We believe that software-based compensation for rolling shutters will become more prevalent as CMOS cameras are increasingly mounted on mobile devices (phones and unmanned aerial vehicles).

Chapter 5

Pixel Correspondences

"Measurement, we have seen, always has an element of error in it. The most exact description or prediction that a scientist can make is still only approximate."

— Abraham Kaplan (1998)

BOS-based tomography requires that one obtain measurements of 3D ray deflections. We infer these from 2D deflection measurements in a plane behind the scan volume. For small deflections, a noise pattern can be printed on such a plane and then apparent motion tracked via optical flow. For larger deflections and changes in focus, optical flow breaks down and we must instead obtain direct correspondences between camera pixels and points on this plane. With suitable hardware, the plane can be translated by a fixed distance and then correspondences recomputed. This gives us two points in 3D space through which a line can be fit, giving us the exit ray measurement. Unfortunately, obtaining the correspondences is rife with practical difficulties. In this chapter we address these difficulties with a structured light-based algorithm that addresses a particular class of the problem not well served by existing methods.

Many computer vision and graphics applications require the acquisition of these correspondences between the pixels of a 2D illumination pattern and those of captured 2D photographs. Trivial cases with only one-to-one correspondences require only a few measurements. In more general scenes containing complex

inter-reflections, capturing the full reflection field requires more extensive sampling and complex processing schemes. We present a method that addresses the middle ground: scenes where each pixel maps to a small, compact set of pixels that cannot easily be modelled parametrically. The coding method is based on optically-constructed Bloom filters and frequency coding. It is non-adaptive, allowing fast acquisition, robust to measurement noise, and can be decoded with only moderate computational power. It requires fewer measurements and scales up to higher resolutions more efficiently than previous methods.

5.1 Introduction

Many problems in computer vision require the establishment of correspondences between camera pixels and either a single or multiple points on scene objects or illuminants. For example, in 3D scanning it is common to project a sequence of light stripes or encoded patterns onto an object in order to reconstruct the geometry via the observed displacement of projector pixels. In these settings, each camera pixel receives only contributions from a single point on the illuminant, i.e., the PSF is a Dirac peak. Binary encodings such a Gray codes [Bitner et al., 1976] are an excellent theoretical solution to this pixel correspondence problem. In practice however, they suffer from errors since the PSFs are rarely perfectly Dirac, and such binary codes do not readily admit subpixel-accurate correspondences.

In this work, we focus on the intermediate problem of small, near-Dirac PSFs which must be captured with high subpixel precision [Atcheson and Heidrich, 2012]. This allows for mapping camera rays through transparent solids for 3D tomographic reconstruction, using an acquisition setup like that shown in Figure 7.1. Unlike with the relatively small refraction induced by gases, such rays are subject to significant defocus blur. For such applications it is not only necessary to estimate small PSFs, but we must do so robustly, and with high accuracy. Due to their high-frequency anisotropic nature, a non-parametric description of the PSFs is preferable to the axis-aligned box [Zongker et al., 1999] or oriented Gaussian models [Chuang et al., 2000] used in environment matting.

Bloom filters are extremely space-efficient data structures for probabilistic set membership testing [Bloom, 1970]. We show how such structures can be opti-

cally constructed in the context of the pixel correspondence problem, and then inverted using heuristics and compressive sensing algorithms. To this we add a frequency-based environment matting scheme [Zhu and Yang, 2004], but modified to increase efficiency. It naturally handles one-to-many pixel correspondences in a non-parametric fashion. The result is a combined binary/frequency-based coding scheme that requires a comparatively small number of input images while being robust under noise. Our method is non-adaptive in that the structured light patterns need not be modified at runtime. All images can therefore be acquired at the maximum frame rate of the camera and illuminant, reducing overall acquisition times to a few seconds. Processing time is on the order of minutes on a desktop machine, which is significantly faster than the general light transport acquisition methods based on compressed sensing.

5.2 Related Work

The estimation of pixel correspondences is a common problem in vision research, with applications ranging from structured light scanning to environment matting to novel uses of the light transport matrix.

5.2.1 Structured Light Scanning

Structured light applications typically employ efficient encodings such as Gray codes [Bitner et al., 1976] that require only a small number of images. For scanning moving objects, other codes have been developed which allow tracking over time [Hall-Holt and Rusinkiewicz, 2001; Rusinkiewicz et al., 2002]. These stripe encodings are efficient for the purpose of structured light scanning, but can only determine one-to-one pixel mappings. While acceptable for many 3D scanning purposes, the inability to deal with mixtures of pixels can result in artefacts.

Scharstein and Szeliski [2003] projected both Gray-coded stripes as well as sine waves of different spatial frequencies. They note that binary codes can be difficult to measure in the presence of low scene albedo or low signal-to-noise ratio and overcame this by projecting both the binary code and its inverse. In general, binary codes are very robust. Methods based on absolute amplitude measurements are highly dependent upon accurate radiometric calibration and consistent scene

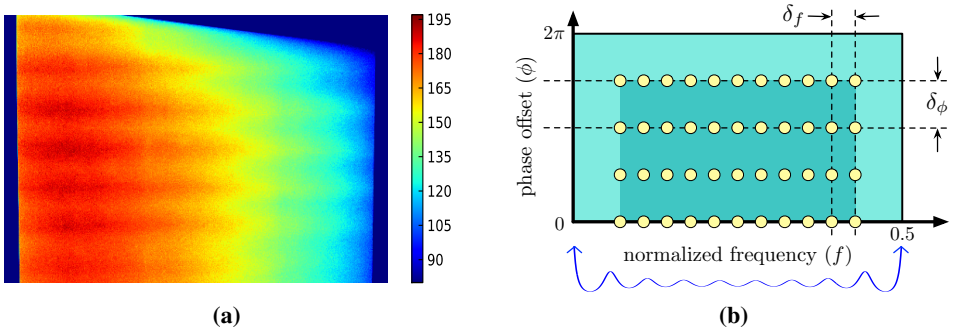


Figure 5.1: (a) Off-axis 8-bit photograph of an LCD monitor displaying a uniform gray image. Note the severe intensity falloff due to viewing angle and Moiré. (b) Sample points in frequency/phase space. δ_f and δ_ϕ may be arbitrarily small. The graph below represents the CRB for the variance on frequency estimates. Note that accuracy degrades significantly near the 0 and 0.5 cycles/sample limits.

albedo. Figure 5.1(a) depicts a common scenario, where the degree of variation that must be calibrated for is of almost the same magnitude as the projected image's intensity range. Scharstein and Szeliski [2003] used relative amplitude measurements of the sine waves to account for the varying albedo, but the calibration problem remains a challenge.

5.2.2 Environment Matting

A high-level description of environment matting can be found in Section 2.2.1. Such a matte represents a 4D slice of the full 8D reflectance field relating incident to outgoing illumination [Debevec et al., 2000]. Many methods use horizontal and vertical stripe structured light patterns to obtain correspondences in the form of singular regions on the background for each camera pixel. This incurs ambiguities in cases where two disjoint regions on the background map to a single camera pixel (e.g., combinations of reflective and a refractive rays). Chuang et al. [2000] resolve the ambiguity via the addition of extra (potentially redundant) diagonal line sweeps, whereas our method does so via more efficient encoding.

Zhu and Yang [2004] have proposed a temporal frequency-based coding scheme whereby the intensity of each pixel is modulated according to a 1D signal (sinu-

soid). Our intra-tile coding scheme is based on this method but employs a second carrier, ninety degrees out of phase of the primary sinusoid, in order to double the information density at no extra cost. The use of only integral frequencies satisfies the Nyquist Inter Symbol Interference (ISI) criterion and allows for very fast, easy and robust Discrete Fourier Transform (DFT)-based decoding. We choose to uniquely code individual pixels (within each tile) rather than coding whole rows and columns of the illuminant. This allows our method to scale up to higher illuminant resolutions, and to naturally handle PSFs of arbitrary (small) shape, rather than assuming a parametric form.

5.2.3 Light Transport Matrix

A number of recent papers have focused on the general problem of estimating the light transport matrix between illuminant and camera pixels. With the complete matrix, one can perform interesting operations such as synthetically interchanging the positions of the illuminant and the camera [Sen et al., 2005]. Most of these methods employ strategies similar to those used in environment matting. Sen et al. [2005] propose a hierarchical decomposition into non-interfering regions. The adaptive approach requires many images to resolve PSFs partially overlapping multiple regions. Our method naturally handles such overlap without requiring additional scans.

Garg et al. [2006] note that the light transport matrix is often data-sparse. They exploit this, along with its symmetry due to Helmholtz reciprocity, in their adaptive acquisition algorithm that divides the matrix into blocks and approximates each with a rank-1 factorisation. Wang et al. [2009b] similarly seek a low-rank approximation to the full matrix. However, they do so by densely sampling rows and columns of the matrix (which requires a complex acquisition setup) and then using a kernel Nyström method to reconstruct the full matrix. These methods assume the matrix to be data-sparse. That is to say, while it may not be predominantly zero, it is sparse in some transform domain (compressible).

Methods based on compressed sensing are beginning to appear. Sen and Darabi [2009] and Peers et al. [2009] both describe promising, non-adaptive, methods that transform the light transport into a wavelet domain in which it is more sparse.

While these methods allow for capturing very complex light transport, they still require on the order of hundreds to thousands of images at typical resolutions, and many hours of decoding time to obtain results.

Our method combines advantages of many of the aforementioned works in that it is both scalable and robust, while remaining conceptually simple and easy to implement. For typical configurations we require on the order of a few hundred images that can be acquired non-adaptively in seconds and then processed in minutes on a standard desktop computer. Unlike more advanced light transport acquisition methods, we cannot acquire large, diffuse PSFs (one-to-many correspondences). But for the case of small, finite PSFs, those methods require many images to resolve high frequency detail. In contrast, our method efficiently captures accurate data at much lower cost in terms of acquisition and processing time.

5.3 Algorithm

We propose a combined binary/frequency-coded structured light pattern for estimating pixel correspondences. Appropriate acquisition setups are simple and inexpensive. All that is required is a spatially-addressable background illuminant (projector or LCD monitor), a camera, and a reflective or refractive scene. Projected patterns are acquired by a synchronised camera and then decoded offline.

The detection algorithm is divided into two phases. First, the background is partitioned into small rectangular tiles (we use 8×8 pixels). Each tile is assigned a unique temporal binary code. A sequence of images is acquired where the tiles flash white or black according to their bit pattern. Since each camera pixel maps to a small area of the background, the measured signal consists of the superposition of these bit patterns. The task is then to determine which codes are present in the observed signal. We use sparsity and spatial coherence heuristics to solve it.

In the second phase we obtain per-pixel weights corresponding to the PSF. Each pixel within a tile is assigned a unique integral frequency and phase combination. We then acquire a sequence of patterns in which each pixel's intensity varies according to the amplitude of its corresponding sinusoidal waveform.

The first phase (inter-tile coding) may optionally use a frequency encoding similar to that of the second phase, but at higher resolution. We describe this

method first in Section 5.3.1 and note that it performs very well in simulation. However, with real data that may not be subject to our simulated assumptions of additive white noise, we turn instead to the Bloom filter-based method described in Section 5.3.2. The second phase (intra-tile) is then described in Section 5.3.3.

5.3.1 Inter-Tile Frequency Coding

As previously mentioned, we assign each tile a unique code. By enumerating tiles this way in 2D, we avoid the ambiguity suffered by methods that partition the background into rows and columns [Chuang et al., 2000; Zongker et al., 1999]. In those schemes, a pixel containing contributions from rows $x_1 \neq x_2$ and columns $y_1 \neq y_2$ has four possible intersection points. The actual beam may have struck two, three or four of these points, and the natural way to eliminate the phantom points is to perform an additional scan pass using a different orientation (e.g., diagonal lines). However, for the unambiguous beams this pass is redundant and reduces efficiency.

The disadvantage of using 2D enumeration is that there are usually far more tiles requiring unique identifiers than either rows or columns. For example, a 1600×1200 monitor could be partitioned into 30 000 tiles of size 8×8. Were we to directly employ frequency-based environment matting [Zhu and Yang, 2004] on these, we would have a maximum frequency of 30 kHz and thus require more than 60 000 captured images (no useful information is encoded in the DC component). Even the improvement we describe in Section 5.3.3 only halves this. But this does assume only integral frequencies and only two phases. We are in fact free to choose any appropriate frequency/phase sampling resolution. Figure 5.1(b) shows an example sampling lattice in frequency/phase parameter space. In the diagram a regular grid is used, with buffer regions in the very low and very high frequencies. Frequency estimation accuracy in these boundary regions is degraded, as predicted via the Cramér-Rao Bound (CRB), which places a lower bound on the variance of an unbiased estimator [Kay, 1993]. While the CRB suggests that an optimal lattice would be nonuniformly spaced with frequency sampling density varying according to f, in practice the oscillations are small and we prefer a regular grid for simplicity. However, frequencies near 0 Hz and the Nyquist limit should nevertheless be avoided.

This very dense sampling requires a signal parameter estimation algorithm that can very accurately detect the frequencies. Periodograms, as used in the intra-tile coding step, are most useful when only integral frequencies are present. Otherwise, spectral leakage interferes. In higher resolution scenarios, better accuracy can be obtained via subspace methods such as Multiple Signal Classification (MUSIC) [Schmidt, 1986] and Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT) [Roy and Kailath, 1989]. These are state of the art harmonic retrieval methods for extracting accurate frequency estimates from small quantities of noisy data. Based on eigendecomposition of the signal covariance matrix, the ESPRIT algorithm is particularly well suited to the case of sinusoidal parameter estimation in a signal corrupted by additive white Gaussian noise. Details on its implementation are given in Appendix A.

Despite their great accuracy, subspace methods can fail when signals contain multiple components of very similar frequency. This is likely to occur if we number the tiles in row- or column-wise order and map these directly to consecutive points in frequency/phase space, because many beams will strike near the tile boundaries and receive contributions from adjacent tiles. To ensure that spatial neighbours are not also frequency/phase-space neighbours it is necessary to label them according to a random, or low discrepancy sequence , such as the Van der Corput [1935].

Our simulations in Section 5.4 indicate that 225 000 unique codes can be represented in 64 images. Unfortunately, real-world experiments could not reproduce these synthetic results. One possible explanation is that Gaussian noise is a poor model of the actual measurement noise and response-curve linearization error in our acquisition setup. While we believe that high resolution spectral methods show promise for pixel coding, our experiments suggest that too many images need be captured in order to obtain accurate estimates. For this reason we also developed the better-performing binary coding scheme described next.

5.3.2 Inter-Tile Binary Coding

A set of N distinct tiles can easily by coded as consecutive natural numbers, whose binary representations require the acquisition of only $\log_2 N$ images. This scheme suffers from reliability problems, in that a single incorrectly-read bit can drastically

alter the number. Gray codes ameliorate this problem by ordering the binary codes such that adjacent codes differ in only a single bit [Bitner et al., 1976; Scharstein and Szeliski, 2003]. Camera beams that strike a boundary between two tiles will measure the superposition of two very similar codes, and the most likely error to occur (in the bit position that differs between the two tiles) will result in a localisation error of at most one tile. In general, the superposition of binary codes separated by large Hamming distances leads to measurements that are difficult to interpret and that lack a reliability metric. We would like to be able to measure such arbitrary binary superpositions, and in cases where only a few codes are present would like to be able to discern which they are. For superpositions of many codes, rather than approximating a broad PSF, we discard the signal as being of no use in our final application.

We choose to acquire codes in a Bloom filter. It is represented as a vector of m bits, all initialised to 0. To insert an object, one computes k independent hash values, all in the range $[1,m]$ and sets the corresponding bits to 1. To query whether an object is in the set, one computes its hash values and checks whether those bits are all on (an $\mathrm{O}(1)$ operation). False negatives are impossible (assuming no error in reading the bit values), although there is a probability of approximately

$$f = \left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^{k} \tag{5.1}$$

of returning a false positive, when the set contains n elements. This probability is minimised by choosing $k = \lfloor (m/n) \ln 2 \rfloor$ to arrive at a false positive rate of approximately $f = (0.6185)^{m/n}$ [Kirsch and Mitzenmacher, 2006].

In the context of our pixel (tile) correspondences, the Bloom filter is constructed optically. We decide beforehand on an acceptable error rate f or else a fixed image acquisition budget m, and compute the optimal k value. Each tile is then assigned a binary code based on those k uniformly-distributed hash values. Because the number of tiles is smaller than the universe of $\binom{m}{k}$ keys, it is feasible to explicitly enumerate them all as the columns of a "code matrix" C, as depicted in Figure 5.2.

The camera acquires a sequence of images, which are then thresholded to bi-

nary values. Each pixel therefore records a signal vector y that corresponds to a Bloom filter containing the hash codes of all the tiles struck by that camera beam. By our assumption of near-Dirac PSFs, there is an upper bound of 4 on the number of elements in the set (n). With 64 images, this gives a false probability rate of approximately 0.05%. Since they are sparsely distributed, these errors can be detected via a spatial median filter.

Decoding the measured signals y_i is a matter of inverting the Bloom filter. Since we have the matrix C, this can be done by solving the equation

$$y_i = (Cx > 0). \tag{5.2}$$

The underdetermined system can only be solved by assuming that x is sparse, which is the case for near-Dirac PSFs. This is similar to the standard basis pursuit problem

$$\min_x ||x||_1 \text{ subject to } y_i = Cx, \tag{5.3}$$

encountered in compressed sensing problems. Having chosen the columns of C independently to be sparse binary vectors, they are incoherent (mutually orthogonal), satisfying the restricted isometry property [Candes and Tao, 2005]. The primary difference between Equation 5.2 and basis pursuit is that we cannot measure Cx directly and must make do with only its sparsity pattern. In practice, solutions can be found with the aid of heuristics. To solve the equation we first compute

$$v_i = C^T y_i, \tag{5.4}$$

which corresponds to taking the dot product of the measured signal with each tile code. Since the matrices are sparse and binary, this can be done for each pixel y_i reasonably efficiently. The result is an integer-valued vector v_i. The indices of entries of v_i equal to k correspond to a superset of codes that make up the solution. Extracting only those columns of C to form C' allows us to instead solve the much smaller problem

$$y_i = C'x > 0. \tag{5.5}$$

Due to partial overlap it is possible for codes to be erroneously included in C'. For

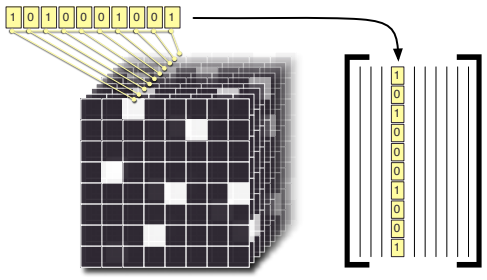


Figure 5.2: Binary temporal codes. Each tile is assigned a unique binary code across the projected image sequence. The codes form the columns of the code matrix.

example, given binary codes $U = (0,1,1)$, $V = (1,1,0)$ and $W = (1,0,1)$ then a ray striking tiles coded by U and V will produce the measurement $X = (1,1,1)$. W will then be included in C' since $W \cdot X = 2 = k$. Our objective therefore is to find a minimal subset of the active codes that adequately explain the measurement.

Any algorithm for solving the basis pursuit problem will give us an estimate of the solution. We additionally impose the constraint that $0 \leq x \leq 1$. Unfortunately, since overlapping nonzeros in the codes produce values that exceed the range of y_i, an exact solution is unlikely and we must instead threshold the resultant x at an empirically-determined value (0.1 in our experiments).

Another heuristic is to enforce spatial coherence, which will be satisfied by all near-Dirac beams. The tile coordinates corresponding to the codes in C' are clustered according to a Chebychev distance threshold of 1. This gives us separate islands of tiles, each of which is checked to see if its constituent tile codes can account for all the observed "on" pixels. If so, then that one island is a solution to Equation 5.5.

During processing, any pixels that cannot be decoded are recorded for further examination during the postprocessing phase. At that time, the neighbours have been determined, so any tile islands that lie suffiently close to any of the neighbours are considered to be valid solutions, even if a few code bits do not match (the result of thresholding errors during acquisition).

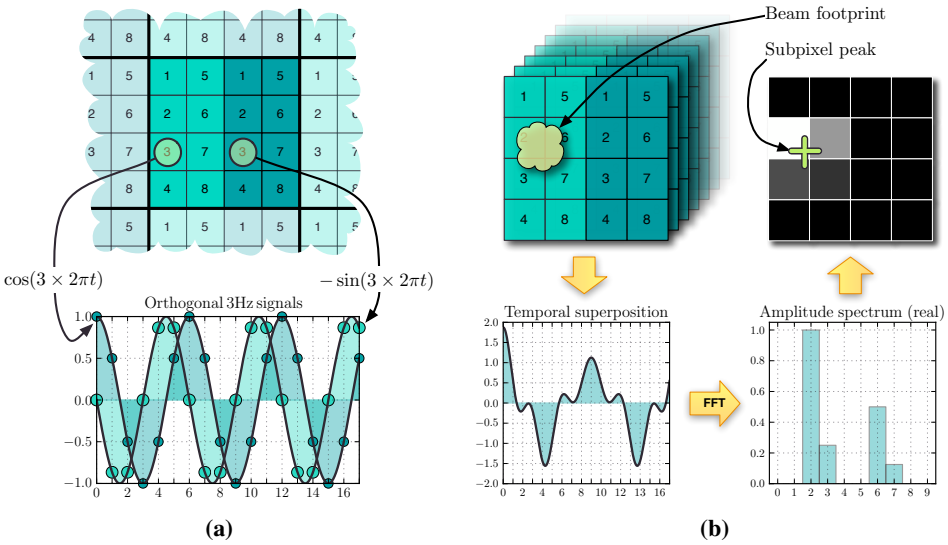


Figure 5.3: (a) Sample 4×4 tile. In this case, $f_{\max} = 8$ Hz. The phase is 0 in the left half of the tile and $\pi/2$ in the right. (b) Temporal superposition of signals under the beam footprint.

5.3.3 Intra-Tile Coding

When a camera beam neither splits into multiple paths, nor spreads out over a large area, we expect a small PSF lying either entirely within one tile, or across the boundaries of two, three or four neighboring tiles. Because the pixels struck by a beam within a tile are somewhat analogous to the tiles struck on the background, we could use the same strategy for detecting them: uniquely coding each pixel within a tile. There are however, two key differences here that call for a different method. First, a greater proportion of the pixels within a tile will be struck than the proportion of tiles struck within the background. Both inter-tile coding methods break down when too many codes are superimposed. Second, there are relatively few pixels in total within each tile, making a more direct, non-parametric method feasible.

In particular, we use the frequency coding method described by Zhu and Yang [2004], but modify it to require only half as many images. Figure 5.3 shows an example of a tile surrounded by segments of its eight neighbours. The $N \times N$ tile

is split vertically in half, and each side is enumerated as indicated. The label of the k'th pixel corresponds directly to its temporal frequency f_k. The spatial location is encoded by setting the frequency and phase of a complex exponential

$$s(t) = Ae^{i(2\pi ft+\phi)}, \quad t \in [0,1) \tag{5.6}$$

and modulating the intensity s of each background pixel over time as

$$s'(t) = \lfloor 0.8D/2 \rfloor s(t) + D/2 \tag{5.7}$$

in order to take the effective dynamic range D of the display into account ($D = 256$ for an 8-bit LCD). The factor 0.8 is chosen empirically to avoid the extremes of the display's intensity range, where clipping can occur. The pixel's location is hence transmitted as a sampled waveform. The maximum frequency $f_{\max}$ is $N^2/2$ Hz and so we set $F_s = 2(f_{\max}+1)$ to satisfy the Nyquist rate, and the sampling rate to $T_s = 1/F_s$. The projected frames then correspond to discrete times $n \in \{0, T_s, 2T_s, \ldots, 1-T_s\}$. If the phase were unrestricted we could generate a discrete sequence of pixel intensities for the k'th pixel as

$$s'_k[n] = s'(s_k[n]) = s'\left(Ae^{i(2\pi f_k t+\phi_k)}\right). \tag{5.8}$$

But we can ease the spectral estimation by allowing only two phases spaced exactly one quarter-period apart, chosen for convenience to be 0 and $\pi/2$. Hence we assign the following signals within a tile:

$$s_k[n] = \begin{cases} \cos(2\pi f_k n), & \text{left half,} \\ -\sin(2\pi f_k n), & \text{right half.} \end{cases} \tag{5.9}$$

The encoder assigns to all signals a unit magnitude ($A = 1$) so that we can easily recover relative contributions from multiple frequencies when camera rays strike multiple pixels. The receiver measures a superposition of signals from the p pixels struck by the beam, corrupted by what we model as additive white Gaussian noise w:

$$x[n] = \sum_{l=1}^{p} A_l e^{i(2\pi f_l n+\phi_l)} + w[n] \tag{5.10}$$

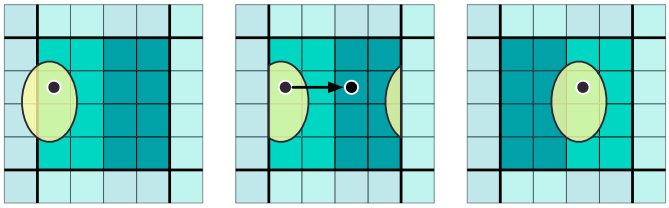


Figure 5.4: Wraparound effect caused by tiling. When a beam strikes the boundary between two tiles (left), we observe a magnitude spectrum with peaks on opposing edges at the perimeter (middle). We circularly shift the maximum pixel (black dot) towards the centre (right) allowing for subpixel peak interpolation.

Our goal is to estimate the parameters f_l and ϕ_l, which together encode the position of each component pixel, and A_l which will represent the relative amount of light arriving at the sensor from it. To estimate these spectral parameters we use the periodogram, which represents the magnitude-squared Fourier transform of the signal, divided by the number of time samples [Kay, 1993]. After performing a per-pixel DFT, we scale by T_s and discard the redundant copy of the spectrum. The real component (in-phase channel) then directly corresponds to the relative contribution A_k towards the PSF from pixels in the left half and the imaginary component (quadrature channel) likewise corresponds to contributions from the right. The PSF can be directly visualised by plotting these results as an $N \times N$ intensity plot, as in Figure 5.3 (right, top). It is thus described non-parametrically, and a subpixel-accurate location of the peak may be interpolated and added to the tile's global coordinates. An approximate interpolant may be obtained via the amplitude spectrum's centroid, or a local 3×3 Gaussian fit [Thomas et al., 2005] around the maximum using the equation

$$x = \frac{\ln p_{x-1,y} - \ln p_{x+1,y}}{2\left(\ln p_{x-1,y} - 2\ln p_{x,y} + \ln p_{x+1,y}\right)} \tag{5.11}$$

for the x component and a similar one for the y component.

A complication arises if the beam crosses a tile boundary. Previous methods for handling boundary overlaps in tile-based schemes have involved scanning additional passes with translationally offset tile grids [Sen et al., 2005] and considering

only one of these passes: that which finds the PSF fully enclosed by a single tile. Our method requires only a single pass, as long as the PSF is smaller than a single tile. We locate the maximum value in the magnitude spectrum and circularly shift this to the centre of the tile, recording the shift vector so that we can subtract it and still obtain an absolute position in global coordinates. Figure 5.4 illustrates the process.

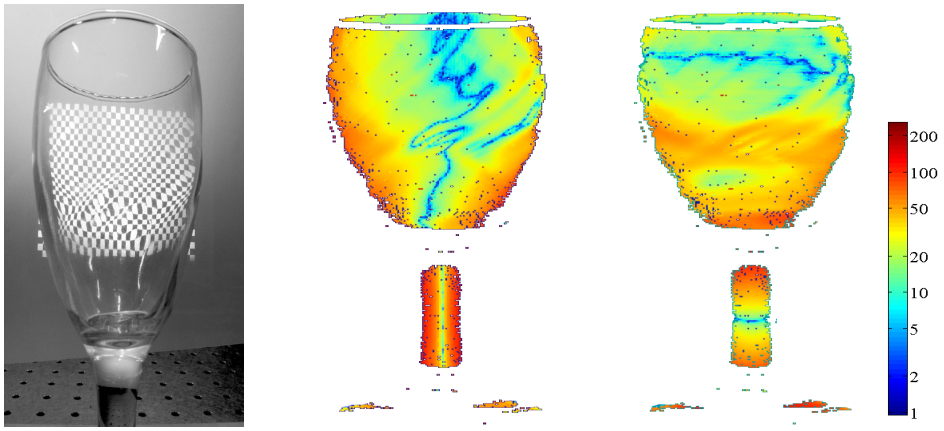
When a PSF overlaps the boundary between two tiles, we will have obtained the addresses of both neighbouring tiles from the inter-tile coding, as well as an $N \times N$ intensity image from the intra-tile coding which, by its wraparound design, corresponds to *both* tiles. When two copies of the magnitude spectrum are placed side by side, the PSF is formed along the shared edge, and a repeated copy is split across the outer perimeter. Circularly shifting the maximum intensity pixel into one of the central pixels will have the same effect as placing copies of the tile side by side, and allows us to interpolate a subpixel peak.

5.4 Results

We first demonstrate successful capture of simple environment mattes using the binary/frequency coding method, then present simulations indicating the expected performance of a method based on high resolution spectral estimation. We include them since they suggest a way to increase the information throughput for a given image aquisition budget, but note that a more accurate measurement setup would be necessary to achieve such results in practice.

To test the algorithm we computed optical flow by comparing the correspondences before and after placing a refracting object in front of the camera. Figure 5.5 shows the displacement vectors and a sample photograph of the scene (from a different viewpoint). The Bloom filter parameters for this dataset were $m = 60$ and $k = 4$. Aside from missing data in regions of total internal reflection, the errors are few and easily filtered out.

In some cases we require a single corresponding point on the background for each camera pixel, in others we require the whole PSF. Figure 5.6(a) shows how our method can provide both an accurate non-parametric PSF, as well as a reasonably accurate point correspondence. Since the precise location of non-Dirac PSFs is



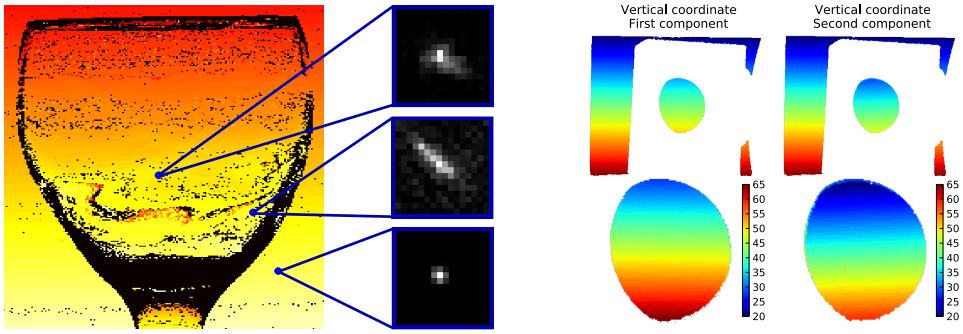
(a) (b) (c)

Figure 5.5: Background distortion through a poorly-manufactured wineglass. The rightmost images show log magnitude of vertical and horizontal apparent displacement of background pixels when viewed through the glass. The scale is 0.01 to 250 pixels.

undefined, we choose it to be the centroid of the neighborhood around the brightest pixel. In a moving scene one may compute the optical flow between PSFs from one time step to the next, without needing to know their precise location. Figure 5.6(a) also shows a near-failure case where too few binary code images were captured ($m = 40, k = 4$), resulting in many undetectable pixels.

Unlike Gray codes, our method is capable of detecting PSFs composed of multiple near-Dirac components. Figure 5.6(b) shows an example where a beam splitter (mounted inside the occluding housing) and mirror combination are used to direct camera rays to two distinct points on the illuminant. This example shows only the inter-tile binary coding result, since frequency-based intra-tile coding would require larger tiles when acquiring larger, or multi-component, PSFs. In this case, we eliminate the tiles and apply binary coding to each pixel. The result is that fewer images need be captured, at the cost of losing subpixel precision.

Capture paramaters were $m = 112$ and $k = 10$. We assumed that at most 8 tile codes would be present in any one Bloom filter to accommodate the worst case of both beams striking at the intersection of four neighboring tiles. The false positive



(a) (b)

Figure 5.6: (a) Examples of spread-out, bimodal and point-like point spread functions. The colour gradient indicates the vertical component of the detected background pixels' coordinates. (b) Multipath correspondences. A beam splitter inside the occluder reflects some light onto a mirror that directs it to another point on the illuminant. The bottom row of images show a closeup of the central region (the beam splitter) on a different colour scale.

probability in this case is 0.098%.

To verify the accuracy of our frequency estimations and to determine appropriate parameter values, we conducted simulations under expected conditions. Figure 5.7 shows the results. In the leftmost graph, we analysed the impact of measurement noise for the case where only a single frequency is embedded in the signal. The graphs show median absolute error, relative to the Nyquist frequency, so an upper error value of 0.5×10^{-3} indicates that we could choose a sampling lattice spacing of double this, i.e., $\delta_f = 0.001 \times N/2$ Hz. Error values asymptotically approach a lower bound as the number of captured images increases, but going beyond 100 images leads to diminishing returns. Too few images however, lead to very high error, indicating that ESPRIT would not be suitable for detecting frequencies within a tile.

Next, we investigated how superposition of signals degrades performance. The second graph shows cases with up to four simultaneous frequencies, chosen randomly, but spaced far enough apart so as not to be strongly correlated. The amplitudes were all set to 1.0 and the simulation was run at a Signal to Noise Ratio (SNR)

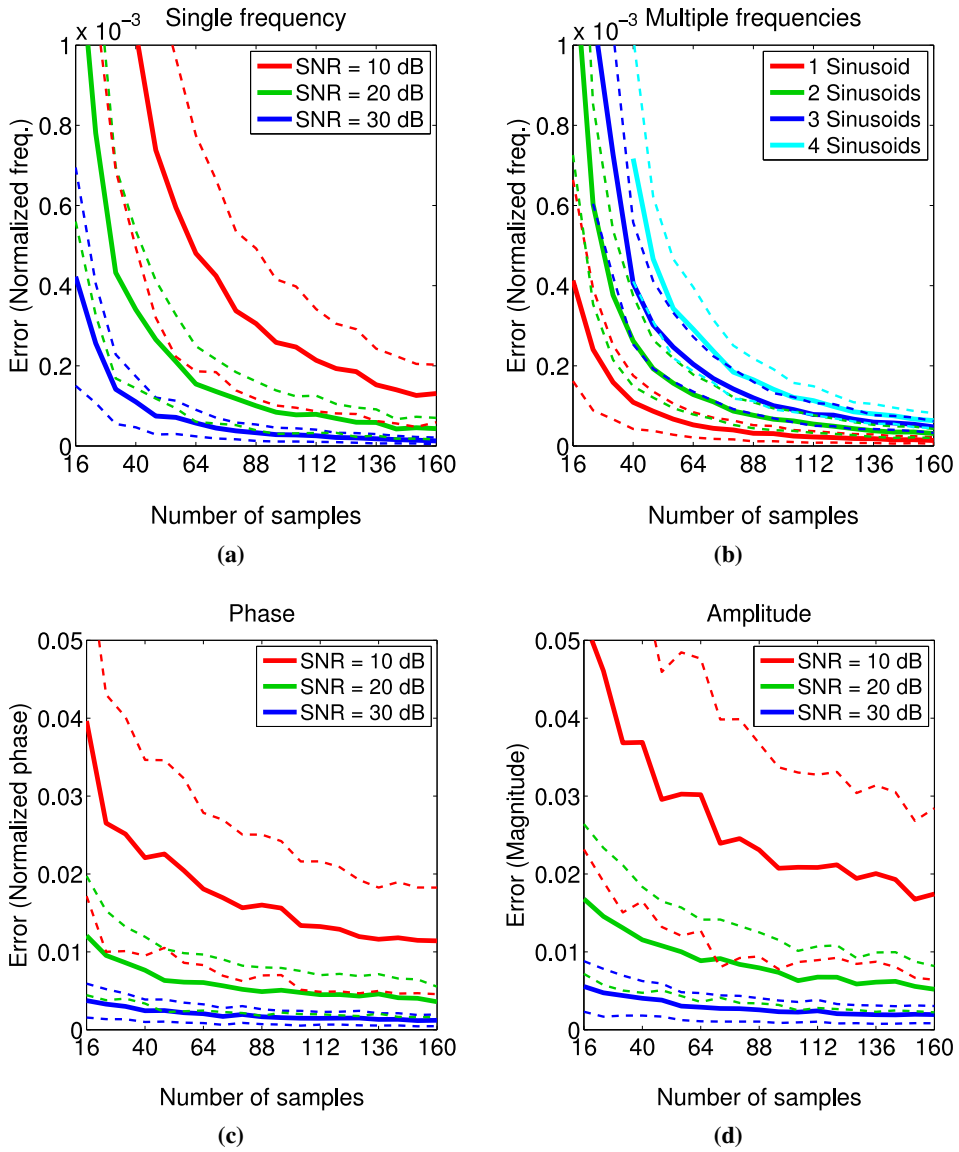


Figure 5.7: Synthetic experiment results. Solid lines show the median absolute error, while dashed lines indicate the median absolute deviation. 500 trials were performed for each tested sample size.

of 30 dB. Accuracy does degrade as more signals are added, but the effect becomes relatively small as N increases.

The third graph shows that we are unable to detect phase as accurately as frequency. For this reason, the sampling interval δ_ϕ depicted in Figure 5.1(b) must be much larger than δ_f. The vertical axis in this graph is relative to π rad/sample.

The final graph shows amplitude accuracy, at which we obtain similar performance to phase (as is to be expected, since both values result from the solution of the same linear system). The vertical axis is relative to the unit input signal amplitude.

Given these results, we can determine the number of tiles than can adequately be coded given a fixed image acquisition budget. For a typical case of $N = 64$ images taken at an SNR of 30 dB, when four sinusoids are present, frequency can reliably be detected to within $0.0004 \times N/2 = 0.0128$ Hz, and the phase is accurate to within $0.005 \times \pi$ rad/sample. Avoiding the lower 5% and upper 5% of frequencies, and covering this space with a lattice of points spaced $\delta_f = 2 \times 0.0128$ Hz and $\delta_\phi = 2 \times 0.005\pi$ rad/sample apart gives us 225k sample points, i.e. 64 images is enough to support 225k tiles, so long as no more than four of them are superimposed at one pixel.

Chapter 6

Gas Flow Acquisition

"Think of the transition air/matter in a CT scan: as far as classical physics is concerned, this transition is abrupt and cannot be expressed as a bandlimited function. Further, there exist obviously no way at all to perform any kind of antialiasing filter on physical matter (before sampling). Most patients would certainly object to any attempt of the sort."

— Thévenaz et al. [2000]

Fluid simulation is widely used in computer graphics applications. However, it remains difficult to obtain measurements of the corresponding real fluid flows for validation purposes, or in cases where such simulations do not adequately model reality. In this chapter, we take a step in the direction of capturing gas flow data for such applications. Specifically, we present the first time-resolved Schlieren tomography system for capturing full 3D, non-stationary gas flows on a dense volumetric grid. Schlieren tomography uses 2D ray deflection measurements to reconstruct a time-varying grid of 3D refractive index values, which directly correspond to physical properties of the flow. We build upon the tools described in previous chapters to capture data with a relatively low-cost consumer camera array. The reconstruction algorithm is a variant of ART that produces high quality results from even a small number of cameras. The method is suitable for use in cases where the range of refractive indices is small i.e., rays can be reasonably approximated by straight lines.

6.1 Overview and Related Work

Computer graphics research has for a long time been interested in capturing properties and behaviours of real-world objects, both for direct use of the captured data in rendering, and, perhaps more importantly, for deepening the understanding of the principles underlying specific phenomena. For example, in studying material reflection, a significant body of work has been developed for measuring aspects such as Bidirectional Reflectance Distribution Functions (BRDFs) [Marschner et al., 1999; Ward, 1992], and subsurface and volumetric scattering [Goesele et al., 2004; Hawkins et al., 2005; Jensen et al., 2001; Narasimhan et al., 2006].

However, there has been comparatively little work done on fluid capture within computer graphics. Some notable examples are the capture of shallow surface waves [Morris and Kutulakos, 2005; Murase, 1990], the surface geometry of simple fountains [Wang et al., 2009a] and the volumetric emissivity of flames [Hasinoff and Kutulakos, 2003; Ihrke and Magnor, 2004]. For real measurements, we turn instead to the fluid imaging community which has a long history in this topic. However, these measurements are typically either sparse, or only capture 2D slices or projections of the flow. A simple way to acquire sparse measurements is to insert fixed probes (e.g., thermocouples) directly into the flow and to record temporal data. For some cases (e.g., rocket engines) this is the only viable acquisition mode, although any increase in measurement density will necessarily influence the flow itself. A more useful technique is PIV. Although predominantly 2D, it can be extended to volumetric imaging either by hardware-based solutions (sweeping the laser plane through the volume [Van Vliet et al., 2004]) or else algorithmically via stereography and tomography [Grant, 1997]. For reasons of hardware complexity, an alternative to PIV would be desirable.

A promising candidate has recently emerged: Schlieren tomography (see Section 2.3 for an overview). Rather than tracking transport within a fluid, this method measures dense, volumetric refractive index distributions. For a turblent fluid, tracking the distribution over time can provide fluid transport information. Tracking density and particles has been used in the capture of dynamic participating media such as smoke [Fuchs et al., 2006; Hawkins et al., 2005]. Transparent gases however, such as heated air, tell-tale mixtures of compounds from gas pipe leaks,

or helicopter rotor blade tip vortices [Kindler et al., 2007], must instead rely on refraction-based techniques like Schlieren tomography. This has the added benefit of imaging the fluid directly, rather than any carried particles whose inertia may cause them to behave slightly differently.

Until now, experiments based on this technique have been restricted to two simplified cases:

- *Stationary flows* [Agrawal et al., 1999; Goldhahn and Seume, 2007; Schwarz, 1996]. These exhibit complex structure, and although fluid transport occurs, their spatially-varying refractive index at any particular point remains constant over time. They can therefore be recorded by a single moving camera. An example of such a flow would be the multiple laminar plumes depicted in Figure 6.6.

- *Axisymmetric* [Faris and Byer, 1988; Venkatakrishnan and Meier, 2004]. These appear the same when viewed from any angle around a particular axis, and can therefore also be captured with just one camera. Examples of such flows include laminar candle plumes and, more interestingly, nozzle jets of various shapes.

These limitations are primarily due to the inability to acquire sufficient synchronised, time-varying Schlieren data. The machine vision hardware required to capture such data remains expensive and unwieldly today. However, our acquisition setup demonstrates that much cheaper consumer camera arrays can instead be used for this purpose, enabling us to scale up more easily and acquire the additional views relatively easily. Aside from hardware, most previous work on Schlieren tomography has involved flows (or approximations thereof) with simple structure that can be described using invertible analytical models (e.g., axisymmetric plumes [Agrawal et al., 1999]). We show here how one can instead cast the problem as a standard tomography problem in the ART framework. The still limited number of views poses a problem in terms of reconstruction quality, and to that end we also show how to employ visual hull constraints to allow for the use of only a modest 16 cameras.

6.2 Algorithm

In this section we first describe the theory relating our measurements to ray propagation and the ensuing inversion process. We then describe our data acquisition process, the physical measurement setup, and the tomographic reconstruction process.

6.2.1 Theory

The reconstruction algorithm is based upon the observation that the measured change in a ray's direction corresponds to a line integral over all the differential changes along its trajectory. Recall the image formation model described in Chapter 2, in particular the ray equation as a system of ODEs:

$$n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \mathbf{d} \tag{6.1}$$

$$\frac{\mathrm{d}\mathbf{d}}{\mathrm{d}s} = \nabla n. \tag{6.2}$$

Integrating it along the ray path Γ produces the result

$$\int_{\Gamma} \frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right)\mathrm{d}s \;=\; \int_{\Gamma} (\nabla n)\,\mathrm{d}s. \tag{6.3}$$

Dividing through by the index at each point and recalling that $\frac{d\mathbf{r}}{ds}$ is a unit vector representing the ray's direction, we get that

$$\widehat{\mathbf{d}}\,\Big|_{\Gamma_a}^{\Gamma_b} = \int_{\Gamma}\left(\frac{\nabla n}{n}\right)\mathrm{d}s. \tag{6.4}$$

In order to solve this problem efficiently we linearise it by assuming that the refractive index range is small enough to ignore. This has two main effects: first, that we can neglect the factor n in the denominator, and second that we can represent the ray paths as data-independent straight lines that are determined solely by the camera calibration. This is similar to the paraxial approximation made in optics, in that we assume $n \approx 1$ and update $\mathbf{r}$ based on $\mathbf{d}$, while leaving $\mathbf{d}$ constant. Representing the ray's direction as it enters and leaves the reconstruction volume at Γ_a and Γ_b

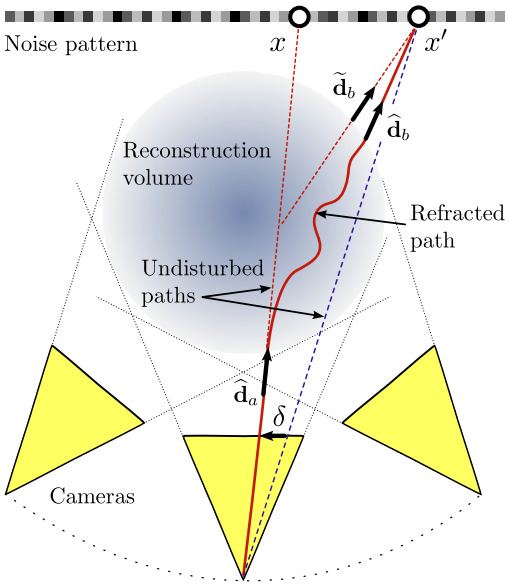


Figure 6.1: Principle of the BOS deflection sensor. A plane with a high frequency noise pattern is placed behind the scene of interest and an image is recorded without the object. Then the inhomogeneous refractive index field is inserted between the camera and background plane. Another image is taken and the deflection of the light rays in the image plane is computed using optical flow.

as, respectively $\widehat{\mathbf{d}}_a$ and $\widehat{\mathbf{d}}_b$, we get the final equation in the form listed in Table 2.1:

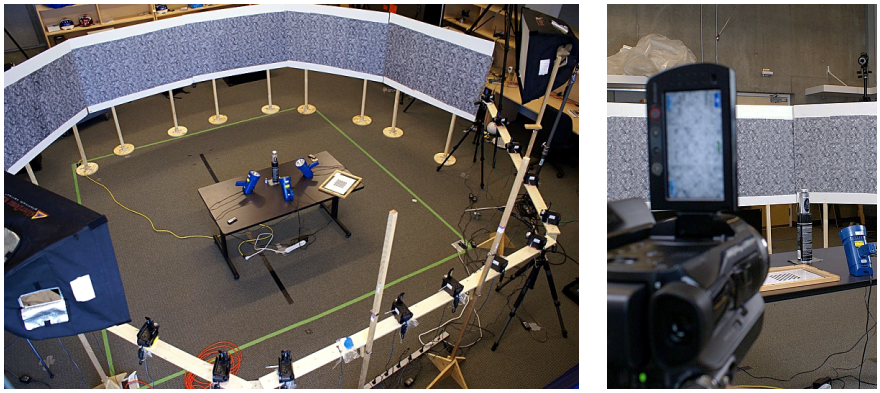
$$\widehat{\mathbf{d}}_b - \widehat{\mathbf{d}}_a = \int_\Gamma \nabla n \, \mathrm{d}s. \tag{6.5}$$

Incoming ray directions are obtained from the camera calibration in the world coordinate frame. Outgoing ray directions are provided by measurements. We use BOS to obtain 2D deflection vectors in a background plane, and from these estimate the 3D deflections. Figure 6.1 shows a ray from one of the cameras being deflected as it passes through the scan volume. Helmholtz reciprocity allows us to think in terms of rays travelling from the camera to the plane/light source. In the figure, the cylindrical scan volume has lower index than the surrounding air, and so the ray is bent away from the centre. A reference image with no refracting medium will have the ray striking point x on the background; after perturbation we instead

observe point x' at the same camera pixel. The apparent motion *from* x' *to* x is represented by the 2D optical flow vector δ. Note that in computing optical flow, simply interchanging the order of the operands (images) and negating the vector field does not produce an identical result. Rather, it produces a flowfield similar to that of the reverse ordering, but warped by the flowfield of the forward ordering. One should therefore compute the flow from the distorted to the undistorted image and not vice versa. From the calibration and measured vector δ we would like to obtain $\widehat{\mathbf{d}}_b$ but unfortunately cannot, since we do not yet know the exact point at which the (curved) ray exits the scan volume. Although our reconstruction algorithm employs a simplified straight ray model, the actual ray exit point is measureably different from the straight ray exit point, thanks to the large scale of our acquisition setup. Knowing exactly where the scan volume lies, we instead assume that all refraction happens at a single point, half way along the ray's passage through the volume. We thus obtain and use the estimated direction vector $\widetilde{\mathbf{d}}_b$ instead. Results in Table 6.1 show that the approximation error is minor and has negligible impact on reconstruction quality. The accuracy of $\widetilde{\mathbf{d}}_b$ is improved as the distance between background and scan volume is reduced relative to the distance between scan volume and camera.

6.2.2 Data Acquisition

We acquire raw data using the consumer camcorder array shown in Figure 6.2. It consists of 16 high definition (1440×1080, interlaced) Sony HDR-SR7 camcorders (with rolling shutters). Temporal synchronisation is done according to the method described in Chapter 4. The cameras surround a measurement volume of roughly $30\,\text{cm} \times 15\,\text{cm} \times 15\,\text{cm}$ in an almost 180° arc. At long focal lengths it becomes difficult to aim the cameras accurately at a distant target such that their view frusta intersect in such a small region of space. It is also very difficult to position a calibration grid in such a way as to be entirely visible to multiple cameras. It is situations such as these where fiducial calibration grids like CALTag become extremely useful. Geometric calibration of the cameras must be performed with respect to a global coordinate frame, but in a semicircular configuration the cameras cannot all see the same planar grid simultaneously. We therefore calibrate groups of nearby



(a) **(b)**

Figure 6.2: Photographs of our acquisition setup.

cameras independently from each other and then merge them together. Using a 3D calibration target, such as a precision-manufactured cube with unique CALTag grids on each face, would be a superior approach. Behind the scan volume we place high frequency noise patterns that are illuminated with both sunlight and 800 W halogen stage lights. Strong lighting is required to keep exposure times as short as possible so as to minimise motion blur. Overhead fluorescent lights flicker at the mains power frequency, out of sync with the cameras, with the resulting spatio-temporally varying illumination changes (due to rolling shutter) in the images causing optical flow artefacts.

In order to maximise the detectable light ray deflection, the background should be positioned as far as possible behind the measurement volume. The cameras should use a long focal length, and be focused on the background plane for the optical flow to work reliably. This means that, in common with most BOS systems, the flow volume itself will be out of focus. To account for this the camera aperture can be reduced and the cameras moved further away from the volume. Since we require a large aperture for reasons of light sensitivity, we compromised by positioning the measurement volume in the centre of a 7 m diameter ring formed by the cameras and backgrounds.

In previous work, we showed that multiscale noise patterns make for an ideal high-frequency background for BOS imaging [Atcheson et al., 2009]. A camera

recording such a pattern will observe dense, locally distinct features everywhere in the image, independent of the magnification factor. This decoupling of camera and pattern resolution reduces the effort for setting up a BOS system. We also determined that simple gradient-based optical flow algorithms perform adequately on this data, whereas more complex variational methods [Brox et al., 2004] require more parameter tweaking and provide little benefit at high computational cost [Atcheson et al., 2009]. New optical flow algorithms continue to be developed [Baker et al., 2007] but often seek to improve results on the difficult stereo vision occlusions that do not occur in BOS. Their benefits should be weighed against the cost of parameter tuning, a significant endeavour when processing many thousands of frames. In our experiments we used the Lucas-Kanade algorithm [Lucas and Kanade, 1981] which proved to be less sensitive to parameter choice than the slightly better-performing Horn-Schunck [Horn and Schunck, 1981]. Alternatively, dynamic environment matting techniques [Chuang et al., 2000; Wetzstein et al., 2011] could be employed to measure the deflection vectors. Use of these methods to acquire quantitative measurements would however involve significant work to handle the limited colour fidelity, compression artefacts and radiometric calibration of the cameras.

During capture, we record a reference frame from each camera, and use this to compute optical flow for each input video frame. Flow fields are smoothed and filtered to remove outliers. After filtering, we then downsample the deflection fields to 480×270. This reduced resolution is sufficient for tomographic reconstruction at our target resolution, while easing the memory requirements that arise from such a large system. We found that it is important to perform the downsampling on the optical flow fields, rather than on the raw images before computing flow, which resulted in significantly poorer flow estimates in our experiments. Similarly, temporal synchronisation and rolling shutter removal should be performed on the flow results rather than on the raw images. The reason for this is that flow fields are significantly smoother than the high frequency noise patterns, and so are less prone to error upon warping (which involves a second level of optical flow). Figures 6.3(a) through (c) show a representative input image and optical flow result.

The tomographic algorithm we use requires as input the visual hull of the 3D flow. To generate this hull, we need a conservative binary mask of the 2D optical

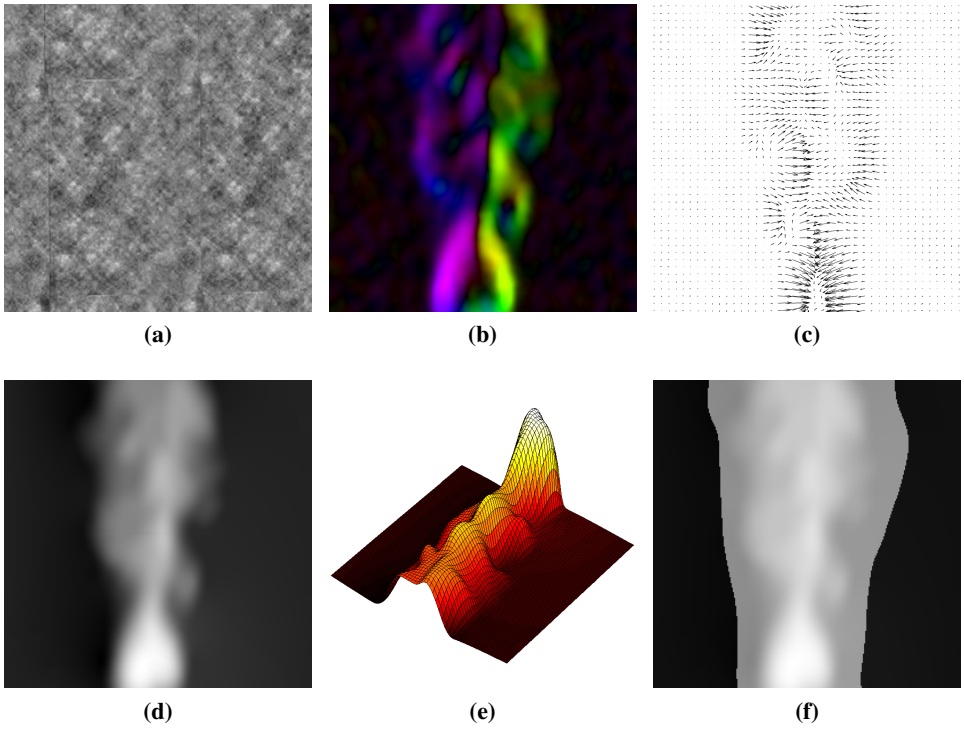


Figure 6.3: 2D data processing. (a) Raw camera image. (b) Optical flow result with direction encoded as hue. (c) Quiver plot of optical flow result. (d) Poisson integrated deflection vectors. (e) 3D rendering of integrated virtual heightfield. (f) Binary mask used for conservative visual hull estimate, after filtering (overlaid on heightfield).

flow for each camera and frame. Note that we cannot simply segment the gas flow from the background by thresholding on the magnitude of the 2D vectors, since these can be zero even if the 3D gradient is not. Equation 6.5 shows that this happens when the index gradient is parallel to the ray's direction. Figure 6.3(b) shows that this occurs in practice for rays passing through the centre of cylindrical plumes.

Fortunately, the 2D optical flow vectors follow a specific pattern. Consider a cylindrical plume of hot air. Light will be refracted *away* from the central axis, and hence apparent motion will be inwards (Figure 6.3(c)) from either side. This is similar to the gradient of a virtual heightfield, and so we perform a 2D Poisson

integration of the vectors (Figure 6.3(d)) and threshold the virtual height instead to produce the binary mask. The threshold value is chosen automatically to include a given percentage of pixels across the entire frame sequence, ensuring approximate consistency across all cameras. Finally, a spatio-temporal dilation is applied to the masks to remove any remaining temporal artefacts and make the mask a conservative estimate of the true visual hull.

Numerically integrating the 2D Schlieren gradients is approximately equivalent to projecting the results of a 3D refractive index gradient integration. Venkatakrishnan and Meier [2004] use this to obtain accurate tomographic reconstructions of shock waves against wedges for which analytic solutions are known. In practice the 2D vectors fields are smooth, and consistent enough with gradient fields so as to be integrable. To avoid problems arising from unknown boundary constraints, occluders should not be present, and the field of view should be large enough so as to encompass both the flow and a sufficiently large empty boundary region.

6.2.3 3D Tomography

Given the per-ray 3D exit direction estimates, we can set up a linear system based on Equation 6.5 using an ART framework. We propose a two-phase reconstruction algorithm to recover first, the index gradient field, and then from that the refractive indices themselves.

For the gradient field tomography we discretise the vector-valued function ∇n using a set of normalised basis functions $\hat{\phi}_i$ with coefficient vectors $\mathbf{n}_i$

$$\nabla n = \begin{pmatrix} \sum_i n_i^{(x)} \hat{\phi}_i \\ \sum_i n_i^{(y)} \hat{\phi}_i \\ \sum_i n_i^{(z)} \hat{\phi}_i \end{pmatrix} = \sum_i \mathbf{n}_i \hat{\phi}_i \tag{6.6}$$

and thus obtain a discrete version of Equation 6.5

$$\widetilde{\mathbf{d}}_b - \widehat{\mathbf{d}}_a = \int_\Gamma \nabla n \, \mathrm{d}s \tag{6.7}$$

$$= \int_\Gamma \left(\sum_i \mathbf{n}_i \hat{\phi}_i \right) \mathrm{d}s \tag{6.8}$$

$$= \sum_i \left(\mathbf{n}_i \int_\Gamma \hat{\phi}_i \, \mathrm{d}s \right). \tag{6.9}$$

Here $\mathbf{n}_i = \left(n_i^{(x)}, n_i^{(y)}, n_i^{(z)} \right)^T$ is a 3D vector independently parametrising the three refractive index gradient components. The discretisation results in a separate system of linear equations for each of the components

$$A\mathbf{n}^{(x,y,z)} = \mathbf{d}_b^{(x,y,z)} - \mathbf{d}_a^{(x,y,z)}. \tag{6.10}$$

Note that the (sparse) coefficient matrix A is the same for each of the gradient components; only the right hand sides differ. The entries of matrix A consist of line integrals over the basis functions:

$$A = \begin{pmatrix} \int_{\Gamma_1} \hat{\phi}_1 \, \mathrm{d}s & \cdots & \int_{\Gamma_1} \hat{\phi}_N \, \mathrm{d}s \\ \int_{\Gamma_2} \hat{\phi}_1 \, \mathrm{d}s & \cdots & \int_{\Gamma_2} \hat{\phi}_N \, \mathrm{d}s \\ \vdots & \ddots & \vdots \\ \int_{\Gamma_M} \hat{\phi}_1 \, \mathrm{d}s & \cdots & \int_{\Gamma_M} \hat{\phi}_N \, \mathrm{d}s \end{pmatrix} \tag{6.11}$$

where N is the number of basis functions and M is the total number of deflection measurements from all cameras. We approximate the line integrals by ray casting and sampling the basis functions. In general the curved rays Γ_j are defined by a solution of the ODE from Equation 2.22, which necessarily involves interpolation of the refractive index gradients. The fact that we do not yet have a solution for the index gradient field is what motivates the simplification made in Equation 6.5 (dropping the factor n in the denominator). We therefore treat ray trajectories as straight lines. This is consistent with the paraxial approximation typically used in Schlieren photography [Settles, 2001]. Simulations of our setup showed a mean deviation from straight line path of less than 0.1 voxels in a 128^3 regular grid discretisation of typical flow data (i.e., about 0.2 mm across the scan volume).

The choice of basis function is important for the tractability of the problem. We use radially symmetric functions because the integrals are dependent only upon the Euclidean distance of the ray to the basis function's centre, and can therefore be precomputed and accessed via lookup table. Kaiser-Bessel functions are often used in tomography since they minimise spectral energy above a certain frequency,

and provide an easily tunable parameter to trade resolution for ripple [Costa et al., 1983]. The results in this chapter were generated using simpler, linear basis functions

$$\phi(s) = \beta(r) = \max(0, 1-r), \tag{6.12}$$

with one voxel overlap in each dimension. Here r is the radial distance from the centre of the basis function. They are arranged on a regular lattice, although we exclude those with support lying entirely outside the visual hull [Laurentini, 1994]. Visual hull restricted tomography was introduced in the context of flame reconstruction [Ihrke and Magnor, 2004] and is useful in obtaining high quality tomographic reconstructions with a sparse set of input views. The visual hull serves as an effective regulariser on the shape of the reconstructed volume and suppresses projection artefacts. Use of these functions preserves the sparseness of the linear system while still allowing for interpolation in the 3D solution space. The finite support of a single basis function is illustrated in Figure 6.4. Due to symmetry we need only know the perpendicular distance of a ray to the voxel centre in order to calculate the line integral

$$\int_\Gamma \phi(s)\,\mathrm{d}s \;\approx\; 2\int_0^{\sqrt{r^2-d^2}} \beta(\sqrt{d^2+y^2})\,\mathrm{d}y. \tag{6.13}$$

Curved ray integrals can be precomputed in a similar fashion by averaging over tangents fitted to a sequence of sampled points (see Figure 6.4(b)). When tracing rays we look up an integral $\int_\Gamma \phi_i(s)\,\mathrm{d}s$ for each basis intersected by the ray. Upon exiting the volume, the entries of the corresponding row in coefficient matrix A are output after normalisation

$$\int_\Gamma \hat{\phi}_i\,\mathrm{d}s = \frac{\int_\Gamma \phi_i\,\mathrm{d}s}{\sum_j \int_\Gamma \phi_j\,\mathrm{d}s} \tag{6.14}$$

After solving for $\mathbf{n}$ via CGLS, integration of the gradient field is analogous to computing a surface from (potentially noisy) normals. We use a discretised version of the definition of the Laplacian operator

$$\Delta n = \nabla \cdot \nabla n \tag{6.15}$$

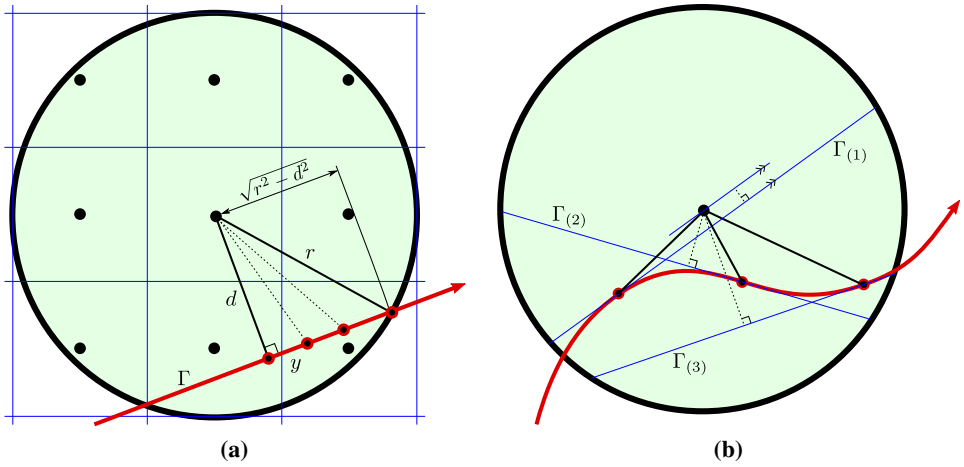


Figure 6.4: (a) 2D representation of a basis function superimposed on a 3×3 voxel grid. A line integral along straight ray Γ (sampled at the red circles) is a function of radial distance d. (b) Integrals along curved rays can be approximated by averaging over tangents fitted to points sampled at sub-voxel resolution.

to compute n. The left hand side of Equation 6.15 is discretised, while the right hand side is computed using the recovered ∇n and the resulting Poisson equation solved for n. The basic Poisson integration scheme assumes a consistent set of curl-free gradient vectors, i.e., $\nabla \times \nabla n = 0$. However, due to measurement errors, the reconstructed vector field does not, in general, meet this condition. As a result, the standard Poisson formulation often results in overshoots by attempting to fit inconsistent gradient vectors in a least-squares sense. Agrawal et al. [2006] present a technique for integrating inconsistent gradient fields in two dimensions. Their method is based on anisotropic diffusion and can be formulated as

$$\nabla \cdot (\mathbf{D} \nabla n) = \nabla \cdot \left(\mathbf{D} \widehat{\nabla n} \right). \tag{6.16}$$

Here $\mathbf{D}$ is a diffusion tensor that weighs gradient information from different directions. For standard Poisson integration $\mathbf{D} = \mathbf{1}$. In our work we use an edge-preserving, anisotropic diffusion tensor similar to that used by Weickert [1998] and Agrawal et al. [2006], but extended to three dimensions by Ihrke [Atcheson

et al., 2008]. This involves more complex analysis of face, edge and corner situations in 3D as compared to the 2D case where only corners and straight edges must be dealt with. Intuitively, **D** prefers gradient information taken from similar isosurfaces of the integrated function and weighs down gradient information orthogonal to it. The definition of **D** and its computation can be found in Appendix B.

We discretise Equation 6.16 using a combination of first order forward and backward differences, which results in a numerical approximation similar to central differences. The anisotropic Poisson equation is again discretised within the visual hull only. This measure saves computation time and avoids blurring of the result into the surrounding empty volume. We use Dirichlet boundary conditions outside the visual hull.

The resulting linear system is large, sparse, and positive definite. It can be solved most efficiently with multi-grid solvers. However, since we have to perform the integration only once per frame we use a less efficient but easier to implement Jacobi-preconditioned Conjugate Gradients (CG) method [Barrett et al., 1994].

6.3 Results

We evaluated our Schlieren imaging and tomographic reconstruction system both quantitatively with synthetic data, and qualitatively through measurements. Synthetic flow data is available and provides us with ground truth against which we can analyse each stage of the reconstruction algorithm (i.e., 2D optical flow, 3D tomography, and gradient integration). In addition, we wish to know the sensitivity to parameters such as the number of cameras. We conducted tests using data generated by a fluid simulator, as well as with data captured by our setup that was subsequently used as virtual ground truth in further simulations. The fluid flow results reported here are from one particular, but representative, dataset: a fuel injection simulation [SFB 382, 2005] shown in Figure 6.5(a). We report errors as both relative Root Mean Square (RMS) errors and Peak Signal to Noise Ratio (PSNR) defined as $-20 \cdot \log_{10} \mathrm{RMS}$.

First, we evaluated the impact of the Poisson solver and its interaction with the discretisation of the normal field. In many gradient-based algorithms, the Poisson solver operates on a gradient field that has been numerically computed. In such

a setting, it is possible to carefully select the discretisation of the Poisson solver to match that of the normal estimation, such that the result is exact up to floating point precision. However, for the measured gradient fields in our setting, the discretisation of the normal field is implicit in the measurement setup and tomographic reconstruction, and thus the one subsequently imposed by the Poisson solver will introduce an additional numerical error. To estimate this error, we started from the ground truth volume data, computed its gradient field with an "unknown" discretisation, and used the anisotropic Poisson solver to recover the original volume. We obtain a PSNR of 42.15 dB (RMS error of 0.78%) on the fuel injection data set, and similar numbers on other data. These numbers provide a baseline for the quality that can be achieved with perfect optical flow estimation, an unlimited number of views, and perfect tomographic reconstruction. A comparison between the ground truth data and the Poisson reconstruction is shown in Figure 6.5.

Next, we analysed the impact of the number of cameras on the tomographic reconstruction step. We ray-traced light paths from virtual cameras through the ground truth volume, and recorded the direction of the rays as they exited the volume. These normalised direction values $\widehat{\mathbf{d}}_b$ were then used as input to the tomographic reconstruction algorithm, whose solution was then integrated using the anisotropic Poisson solver. The resulting errors are shown in the first row of Table 6.1. Total reconstruction error for 16 cameras (PSNR: 41.29 dB, RMS: 0.86%) is already very close to the error bound obtained from the Poisson integration alone. Additional cameras do not result in significant further reductions of error. While the numbers depend somewhat on the volume resolution and the complexity of the flow, we found that 16 cameras generally provide the best tradeoff between hardware requirements and precision.

In the previous simulation we assumed that the exact refracted light direction $\widehat{\mathbf{d}}_b$ was known. However, the deflection measurements obtained by BOS correspond to an approximation $\left(\widetilde{\mathbf{d}}_b\right)$ made by assuming that the refraction occurs at a single point (see Figure 6.1). We expect the approximation to be a good one, given that the scan volume is small relative to the distance between camera and background, and that the total ray deflections are so small. Indeed this is the case, as confirmed by our experiments comparing the two outgoing vector directions (second row of Table 6.1).

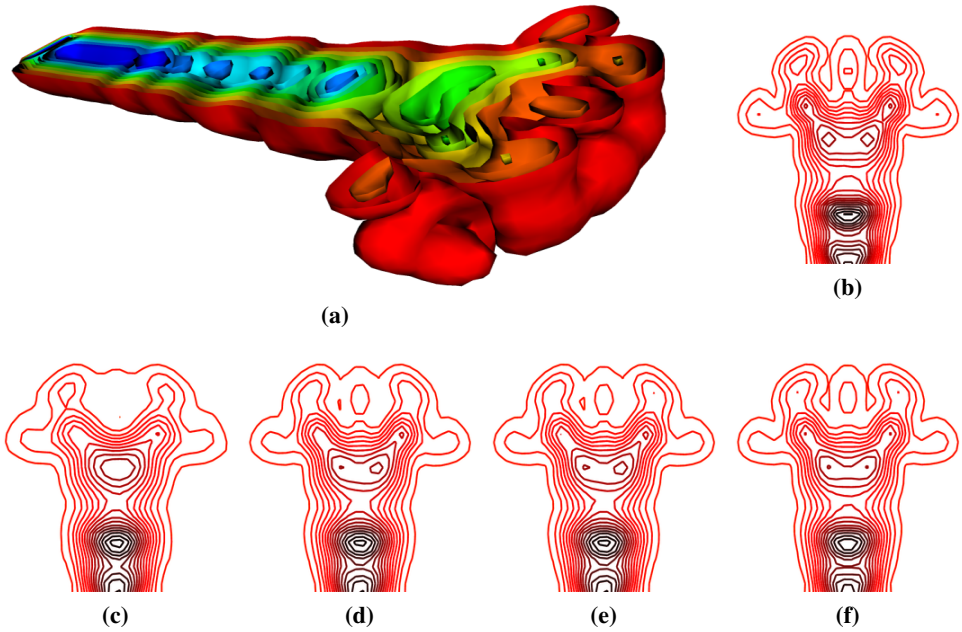


Figure 6.5: (a) Cut-plane view of 3D isosurface rendering of ground truth fuel injection dataset. (b) Ground truth cross-sectional contour map. (c) Contour map for tomographic reconstruction from optical flow data using 8 cameras. (d) 16 cameras. (e) 32 cameras. (f) Poisson integration from ground truth 3D gradients.

Direction estimate	Half ring setup					
	8 Cameras		16 Cameras		32 Cameras	
	PSNR	RMS	PSNR	RMS	PSNR	RMS
Ground truth $\widehat{\mathbf{d}}_b$	40.55	0.94%	41.29	0.86%	41.39	0.85%
Approximate $\widetilde{\mathbf{d}}_b$	40.43	0.97%	40.73	0.91%	40.76	0.91%
Optical flow	39.29	1.09%	39.84	1.02%	39.88	1.01%
	Full ring setup					
	7 Cameras		15 Cameras		31 Cameras	
Approximate $\widetilde{\mathbf{d}}_b$	40.03	1.00%	40.74	0.92%	40.83	0.91%

Table 6.1: Errors statistics for experiments with synthetic data.

In order to estimate the full system error we traced rays through the volume and intersected the refracted rays with a virtual noise background. The resulting images were then processed by the complete pipeline of optical flow, tomography, and gradient integration. Row three in the table shows that the optical flow algorithm introduces additional error, but that overall error remains low, especially when considering the lower bound imposed by the Poisson solver (42.15 dB). Figure 6.5 shows visualisations of the original ground truth fluid flow along with reconstructions from different numbers of cameras.

We also studied the impact of the anisotropic Poisson solver, and found that it improves the PSNR of tomographically reconstructed datasets by about 1 dB when compared to an isotropic solver. We found that a regularisation value of $\alpha = 0.8$ produced the best results (see [Atcheson et al., 2008] for definition of α). All results in this chapter were computed using this value.

Finally, we tested whether it would be better to arrange the cameras in a full ring rather than the half ring used thus far. The last row in Table 6.1 shows the results obtained with 7, 15, and 31 virtual cameras and approximated deflection vectors. We chose an odd number of cameras for the full ring in order to avoid almost complete redundancy of information when two nearly-orthographic cameras are positioned directly opposite each other. There is almost no measureable difference between the half and full ring setup for the same approximate number of cameras. This justifies the use of the half ring setup, which is easier to construct physically.

Real measurements were performed in the setup described in Section 6.2.2. Figure 6.6 shows time sequences of volume renderings for four different gas flows. They demonstrate the ability of our system to capture both turbulent and laminar flows. The turbulent hot air above a camping stove in the top row clearly shows the advection of small scale detail. The laminar flows, including the hot air plumes above three tea lights in the bottom row, show the ability of our approach to clearly separate distinct features, as well as its temporal continuity and low noise. For the burner sequence, our most complex data set, the visual hull was filled by 150 000 basis functions and we aquired 700 000 pixel measurements per time frame. For the other sequences the linear system from Equation 6.10 is usually over-determined by a factor of 8–25. Finally, we show some additional results in Figure 6.7.

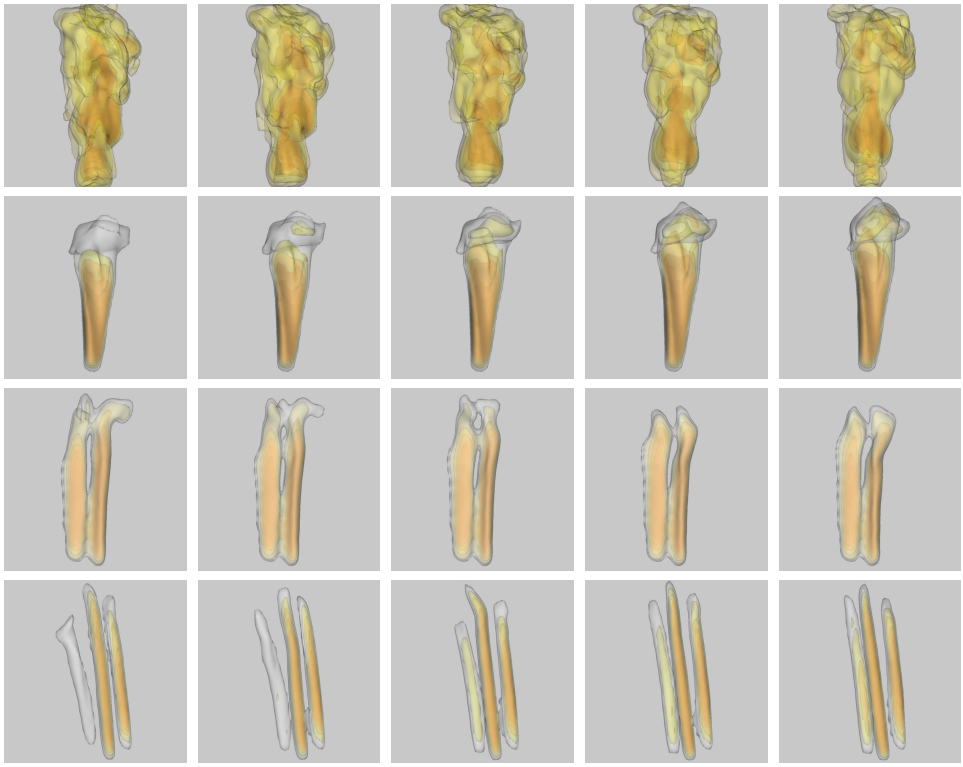


Figure 6.6: 3D reconstructions of data measured in our BOS tomography system. The images within each row are one frame ($1/60$ s) apart. Top row: Turbulent flow of hot air above a gas burner. The advection of features is clearly visible, as the hot air rises due to buoyancy. Second row: Hot air rising from a candle. The flow starts out almost laminar, but eventually breaks up into more turbulent behaviour. Third row: Hot air plume for two tea lights. The almost laminar flow is occasionally disrupted by ambient air movement. Bottom row: Very laminar flow above three tea lights. The individual plumes are clearly separated.

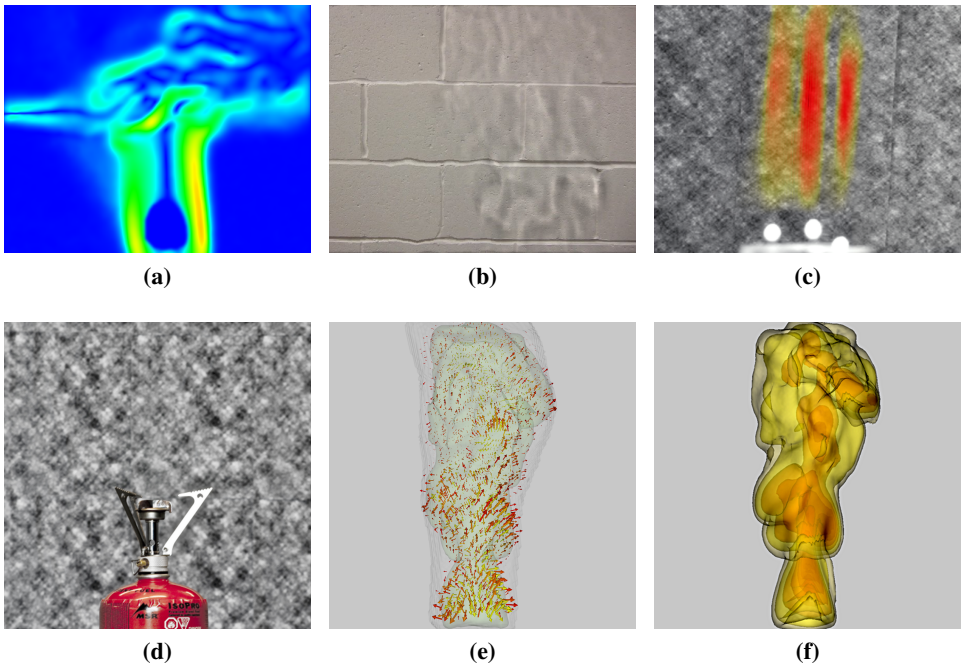


Figure 6.7: (a) 2D optical flow deflection magnitude image of a candle's hot air plume disturbed by a jet of compressed air. Simulating such a flow would be difficult for most fluid simulators, since it violates the incompressibility assumption. Information from the flame is lost due to sensor saturation. (b) A potential application for gas flow data is to render shadowgraphs. Rays are traced through the captured gas burner flow to produce caustics on the wall. (c) Maximum intensity projection of reconstructed laminar plumes of three tea light candles overlaid onto one of the original camera views. (d) Raw camera image for camping stove dataset. (e) Visualisation of the tomographic reconstruction of refractive index gradients from the turbulent gas burner flow. The visual hull is also represented. (f) 3D isosurface rendering of camping stove.

Chapter 7

High-Index Tomography

"Crude measurement usually yields misleading, even erroneous conclusions no matter how sophisticated a technique is used."

— H.T. Reynolds (1984)

In this chapter we explore the extension of gradient field tomography to refractive index fields of higher dynamic range. These are inherently more difficult to reconstruct because the ray paths can bend substantially and depend upon model parameters. Our gradient field tomography solution for the linear problem of gas reconstruction is not appropriate in this case and so we develop an alternative approach. This other method is cast as a nonlinear optimisation, making use of both positional and directional information of light rays. These measurements are provided at high resolution by an almost entirely automated acquisition setup. Our experiments indicate that while reconstruction is possible, its quality is currently too poor to be of use in real applications. We explore the various underlying reasons for this and contrast the method with another more successful refraction-based tomographic method.

7.1 Acquisition Setup

Schlieren tomography requires deflection measurements for each ray, from multiple viewing angles. In Section 6.2.2 we described a setup for capturing this data

for gases via optical flow. Unfortunately, optical flow algorithms are too unreliable and sensor resolutions too low for this to be a feasible approach on high index media (e.g., glass). One must therefore find alternative ways to capture data, such as:

- Immerse the solid media in an approximately index-matched fluid. This is similar to the approach of Trifonov et al. [2006] but with less stringent requirements for an exact match. Water or vegetable oils can more than halve the dynamic index range of glass in air, bringing the problem closer to the point at which optical flow becomes useful. Note that there are benefits to Schlieren tomography even in the event that an exactly-matching fluid is available – unlike with absorption tomography, there is no requirement that the scan object be perfectly homogeneous in its index. Indeed, a slight mismatch is *necessary* for Schlieren tomography to work at all and dilution can always be used with soluble fluids to adjust the discrepancy.

- Forgo the dense measurements of optical flow and instead track a sparse grid of rays. One way to obtain this is to pass a laser beam through a diffraction grating, mimicking a much larger array of separate lasers. Figure 7.1(b) shows such a device. With a sufficiently high temporal acquisition resolution, one can track each individual beam as it moves across a large area. This method would be useful for dynamic, high-index media like liquids. For static media, it may be difficult to identify the beams given only a reference and a highly distorted frame.

- Extend the idea of tracking individual beams by adding uniquely identifiable information to each one. We have a solution for this in the pixel correspondences method from Chapter 5. It also provides the same measurement density as optical flow. By temporally coding each point on a structured light background pattern, we can uniquely identify the rays passing through the scan volume in both refracted and unrefracted cases, thus providing the necessary deflection information.

The latter approach is most applicable to high index static media and is our method of choice. Figure 7.1(a) shows the physical setup. An LCD monitor is mounted on

a linear translation stage that can position the background plane at two discrete distances from the stationary camera. The camera is a Prosilica EC1350 monochrome machine vision camera. A rotational stage positioned just in front of the monitor is used to simulate a circular ring of cameras. Recall that for the straight ray case we observe no benefit in going from a semi-circular to a full circular setup, but because of the complex ray trajectories here we do use a 360° arrangement. The stage must be positioned as close to the monitor as feasible because of depth of field limitations. To address these, we focus the camera at a plane between the subject and background plane, and note that some blur can be beneficial in reducing Moiré effects arising when photographing LCDs screens (see Figure 7.2). Moiré can also be reduced via careful equipment positioning. This illustrates a significant benefit of the frequency-based coding scheme we use for intra-tile correspondences. Such variation in the background, let alone the scan object itself, renders accurate radiometric calibration virtually impossible for any method based on absolute measurements.

Geometric calibration of this setup is a delicate affair. We must calibrate for both the position of the background plane and the rotational axis of the scan volume relative to the camera. To do this we use a rapid prototyping machine to build a support structure that can rotate with the stage. Onto it we place a glass pane with an adhesive CALTag grid attached. Other thin materials, despite their appearance to the naked eye, are often slightly warped and lead to inaccurate calibrations. It is convenient to have the physical size of the CALTag markers on the rig be identical to those displayed on the LCD monitor. This allows for extrinsic parameters of both planes to be obtained in a common coordinate frame.

To calibrate the LCD monitor we manually adjust its pose while capturing approximately twenty images. A second set of images is captured by translating the monitor by fixed increments along a linear path. These known distances between planes provide essential additional constraint information. When calibrating cameras it is not uncommon for displacement along the optical axis to be confounded with the focal length parameter. We used our own implementation of Zhang's [2000] algorithm which enables us to insert these additional constraints arising from the physical setup.

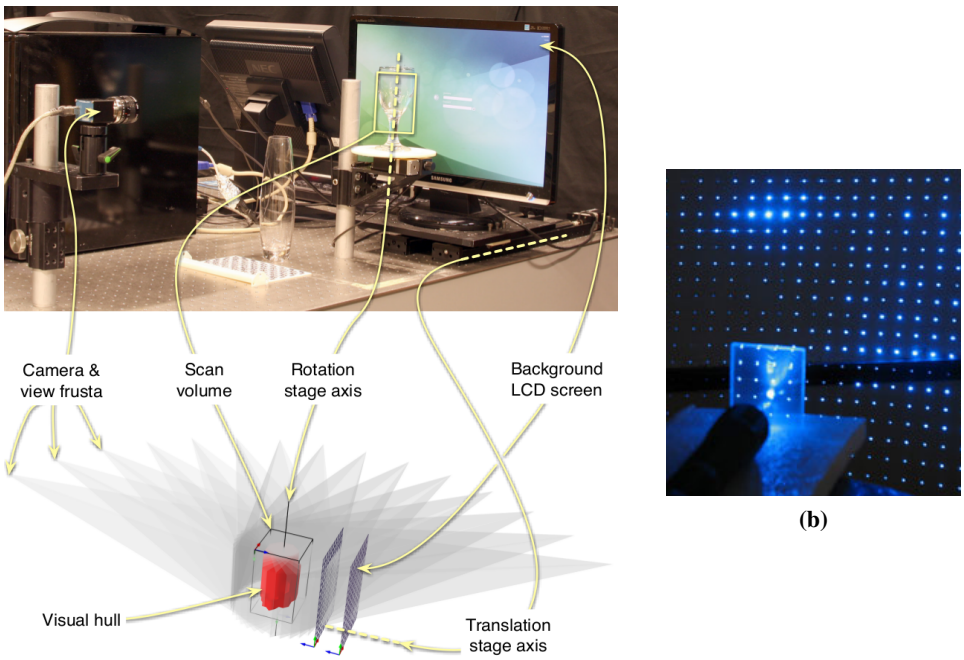


Figure 7.1: (a) Semi-automated acquisition setup. The background LCD monitor is mounted on a linear translation stage and the wineglass and calibration pattern are mounted in turn on a rotational stage. Display, capture and mechanical components are all controlled from a single computer. (b) Potential acquisition setup using laser beam and diffraction grating.

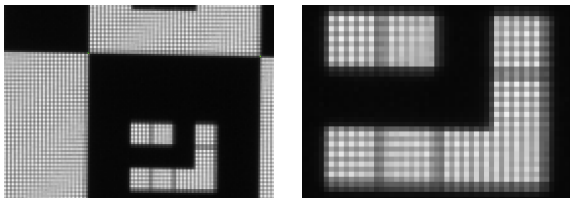


Figure 7.2: (a) Moiré from CALTag pattern displayed on LCD monitor. These are most pronounced when the resolution of the camera is very close to that of the portion of the monitor filling the frame. (b) Close-up view.

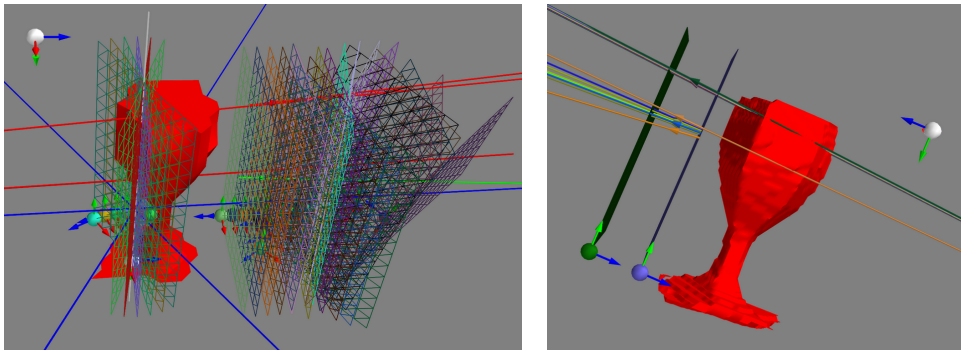
The scan volume is likewise calibrated by rotating the rig by fixed angular increments. We need not ensure that the calibration plane is physically coincident with the rotation axis, nor even that they be parallel (properties which are very difficult to achieve in practice). Once the grid positions are known relative to the camera, we solve for a best-fitting rotation axis. Figure 7.3(a) shows a rendering of some of these calibration planes in a common world space. The visual hull of a wineglass in the scan volume is also shown for reference.

With this setup we choose two positions for the background monitor and N viewing angles. Given a desired pixel correspondence resolution we require M frames per view. We then capture $2(M+NM)$ frames across all angle and background combinations, as well as a reference set with no scan object present. Depending on the opacity of the scan object and the brightness of the monitor, total capture time can range from a few minutes to a few hours. LCD brightness should be carefully adjusted to avoid highlight clipping.

Data is output for all pixels for which we can obtain correspondences in both the front and rear positions, for both reference and refracted views. Coordinates are mapped into world units by scaling according to the monitor's resolution. We then fit lines through both front and rear plane intersection points to obtain refracted and unrefracted rays. Figure 7.3(b) shows some of these rays. The refracted ray can be projected backwards onto the visual hull; however beyond that we know nothing of the ray path between that intersection and the point at which the unrefracted ray first enters the visual hull.

In reconstructing gas flows we made use of the difference in outgoing ray directions. However, for nonlinear ray problems, these measurements do not uniquely identify a ray. Rays may undergo multiple large refractions, be displaced by significant distances and yet still exit at the same angle as they entered. We therefore track changes in both position and angle between refracted and unrefracted rays. This raises the question of units – positional displacements are measured in arbitrary world units, whereas angles are measured in e.g., degrees. To ensure that one property does not overly skew the solution, we first normalise the data by determining scaling constants that result in evenly distributed residuals (see Equation 7.1).

The setup described here is suitable for capturing deflection data of glass objects in air. This is a particularly difficult subject, not just because of the high



(a) (b)

Figure 7.3: (a) Rendering of calibration planes used to orient the background planes and rotation axis to the camera. (b) Rendering of acquired rays. The upper horizontal fan of rays (viewed side-on) shows only unrefracted paths, while the lower fan shows refracted paths as well.

dynamic index range (approximately 0.5) but also because of the high frequency of index changes (step edges). In order to simplify the problem and and for ease of illustration, we revert to 2D simulation results for the remainder of this chapter.

7.2 From Gradient Field to Index Tomography

One of the challenges in using gradient field tomography is that each component of the index gradient is solved for independently. This contributes to the lack of integrability of the vector field. For any ray travelling through the medium, it is only the gradient components orthogonal to its direction of motion that contribute to its change of direction. The ray's final exitant direction is therefore independent of part of the information in the index gradient field, but it is precisely this ignored information that determines the exitant direction of rays from other cameras passing through points along the first ray's path. Given measurement noise and the fact that acquired information is not shared equally by reconstructed components, it is not surprising to see inconsistent gradient fields. In Chapter 6 we accounted for small nonzero curl via anisotropic Poisson integration, but stronger refractions result in much noisier gradient components.

Another drawback to gradient field tomography is the difficulty inherent in ap-

plying regularisation. It is more natural to express desired properties (high smoothness, low total variation, etc.) of the refractive index directly as a function of that output, as opposed to functions of intermediate fields (gradient components) which are nonlinearly combined to form the output. Therefore we would prefer to solve directly for the index field from the data.

This can be cast as a nonlinear optimisation problem where we seek to minimise a misfit function between acquired measurements and a synthesised output from our raytraced forward model using an iterative estimate of the solution. This formulation can be generically expressed as

$$\min_n \frac{1}{N} \sum_{i=1}^{N} w^{(i)} \rho \left(r^{(i)}(n) \right) + \lambda J(n) \tag{7.1}$$

which combines a weighted per-ray residual $r^{(i)}$ with a regularisation term $J(n)$. Because sharp edges are incompatible with the infinite frequency approximation (ray optics), in our experiments we imposed a smoothness prior on the index field using

$$J(n) = \int_\Omega \|\nabla n\|_2^2 \, \mathrm{d}x \, \mathrm{d}y. \tag{7.2}$$

When the solution is known to be piecewise constant it may be preferable to regularise based on total variation instead. For the i^{th} ray we obtain a residual

$$r^{(i)}(n) = \begin{pmatrix} \alpha_\theta & \\ & \alpha_\delta \end{pmatrix} \begin{pmatrix} f_\theta^{(i)}(n) - \mathbf{d}_\theta^{(i)} \\ f_\delta^{(i)}(n) - \mathbf{d}_\delta^{(i)} \end{pmatrix} \tag{7.3}$$

as the difference between measured data $\mathbf{d}$ and our forward model f, which consists of tracing through the current index field n and obtaining angular and positional differences between original and refracted rays (denoted u and v respectively). The angular difference is defined as

$$f_\theta^{(i)}(n) = \arccos\left(d_u \cdot d_v\right) \tag{7.4}$$

where d_u represents the normalised direction vector of ray u, while the positional

difference is defined as

$$f_{\delta}^{(i)}(n) = \|p_u - p_v\|_2 \tag{7.5}$$

where p_u represents the point of intersection of ray u with the background plane. Measurements $\mathbf{d}$ are computed similarly.

The scaling constants α_θ and α_δ are chosen so as to normalise the error components, accounting for the difference in scale between angular and positional units. We found these to be typically distributed approximately uniformly between zero and a maximum value, aside from a large peak near zero. Rather than dividing by the maximum (which would be sensitive to outliers) we fit normal distributions and used their inverse standard deviations as the α. These constants are determined during initialisation via a single pass through the data, or sampling thereof.

The penalty function $\rho(\mathbf{x}) = \sum_j \Psi(x_j)$ can be adjusted to compensate for outliers. In the case of $\Psi(x) = x^2$ we obtain a least squares solution. Better results are obtained by substituting a more slowly growing function for larger residuals, such as the Huber penalty function

$$\Psi_\varepsilon(x) = \begin{cases} x^2 & |x| \leq \varepsilon \\ \varepsilon(2|x| - \varepsilon) & |x| > \varepsilon. \end{cases} \tag{7.6}$$

Aravkin et al. [2011] argue that the nonconvex function

$$\Psi_\varepsilon(x) = \log\left(1 + \frac{x^2}{\varepsilon}\right) \tag{7.7}$$

is superior to other choices in these inversion problems because the influence of outliers diminishes as they become larger. We experimented with both of these penalty functions and found them to be superior to simple least squares penalties. However, no robust penalty emerged as the best overall choice.

The weights $w^{(i)}$ are also determined during initialisation. Careful weight selection can significantly speed convergence. SART introduced the use of weights to favour updates to voxels in the central portion of the scan region [Andersen and Kak, 1984]. The justification for this is that the correction terms computed for central voxels are influenced by many rays. The assumption is that this makes them

more reliable than updates to perimeter voxels. In SART this works well because multiple rays are used to compute each update, and because all reconstruction voxels are considered *equal*. This equality does not extend to the nonlinear case. When rays can bend, the very first refractive interface they encouter has a far greater influence on their eventual exit position and direction that the latter refractive events. Indeed, we would be far better off knowing the true solution for perimeter voxels than for central ones. For this reason we choose to down-weight the influence of ray residuals on central voxels, in direct constrast to SARTs Hamming window.

To set the weights we construct a synthetic distance field centered on the rotation axis at c, with the equation

$$W(x) = \|x - c\|_2^a, \tag{7.8}$$

where we adjust the amount of falloff towards the centre via a. In our experiments we typically used $a = 2$, although the optimal value is data-dependent. Higher values are more suitable for index fields with stronger gradients. We then trace each camera ray through the volume, integrating W along the ray path. For this task, we assume a uniform index distribution i.e., rays follow straight lines.

An alternate interpretation of the SART weighting scheme is given by Strohmer and Vershynin [2007] in the context of generalised Kaczmarz methods (ART). As others have recommended, the authors randomise the order of iteration through equations used to update the solution. Crucially, they select rows (rays) not with uniform probability, but rather according to their L_2 norm. In ART problems where the coefficient matrix row entries give the distance traversed through each voxel by a ray (or some other measure of the voxel's influence on the ray) the L_2 norm corresponds to a ray length metric. It was proven that favouring "longer" rays, i.e., those that pass through the central region, leads to faster convergence. Again though, we note that this is true only for linear problems. Nonlinear problems require problem-specific weighting schemes.

An unfortunate side effect of increasing residual weights on perimeter rays is that for some of the voxels they traverse, we sacrifice orthogonal view information. This is because orthogonal rays intersecting those voxels are central rays. Their weights are accordingly lowered and we must therefore rely on limited-angle in-

formation in order to solve for the outer voxels.

To solve Equation 7.1 we use an interior point constrained nonlinear optimisation algorithm [Wächter and Biegler, 2006]. Derivatives and Hessian matrices are computed using the ADOL-C automatic differentiation library [Walther and Griewank, 2012].

7.3 Synthetic Evaluation

For evaluation purposes we use two different 2D synthetic lenses, shown in Figure 7.4. The first is a Luneburg Lens [Luneburg, 1944]. It is a radially-symmetric function of distance from the origin, giving the refractive index

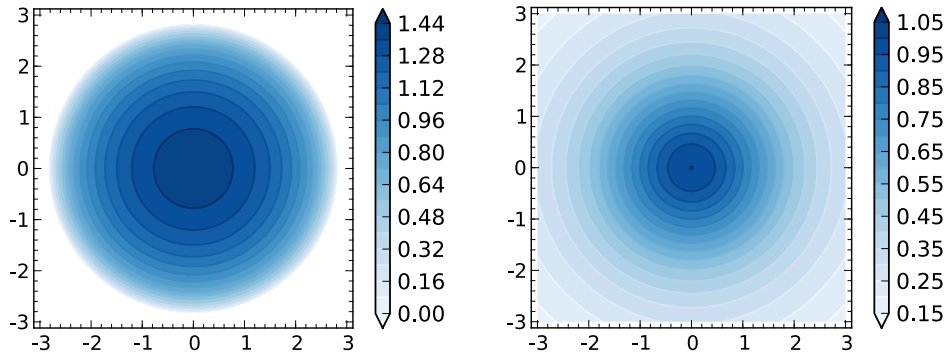
$$n(r) = \sqrt{2 - \left(\frac{r}{R}\right)^2}, \tag{7.9}$$

where R is the radius of the lens. This function has an analytic solution whereby parallel rays entering from one direction will all focus to a common point lying on the circumference of the lens on the opposite side [Andersen and Kak, 1984]. Use of this lens allows us to test the performance of our forward model (ray tracer) across internal parameters such as Runge-Kutta step sizes, and external parameters such as discretisation resolution. We found step sizes to have remarkably little influence on the results, provided they remain on the order of one pixel or less. Resolution however, has a significant influence. Figure 7.5 demonstrates this with ray trajectories from three orthographic cameras through a discretised Luneburg Lens, at both high and low resolutions. At sufficiently high resolutions the solutions behave as expected, converging to foci with little aberration. At low resolutions however we see significant artefacts – in particular an increase in outlier rays.

The second synthetic test lens we used was Maxwell's Fisheye. It has the equation

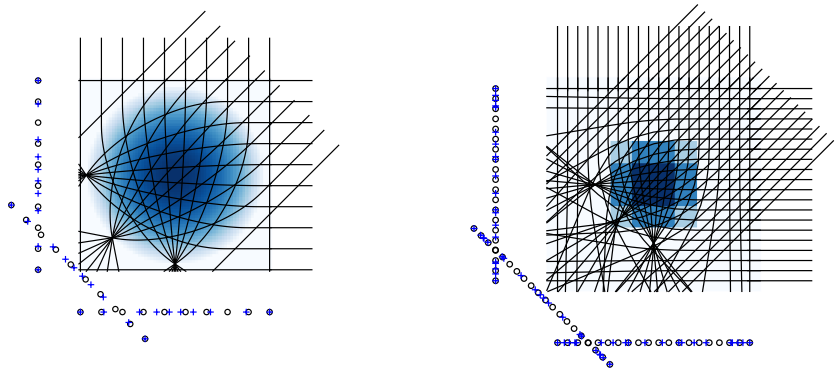
$$n(r) = \frac{n_0}{1 + \left(\frac{r}{a}\right)^2} \tag{7.10}$$

and the property that all rays emanating from a point p, regardless of their direction, will curve around to pass through a focal point p' lying on the line connecting p to the origin. Moreover, after passing p' the ray will continue around to again arrive



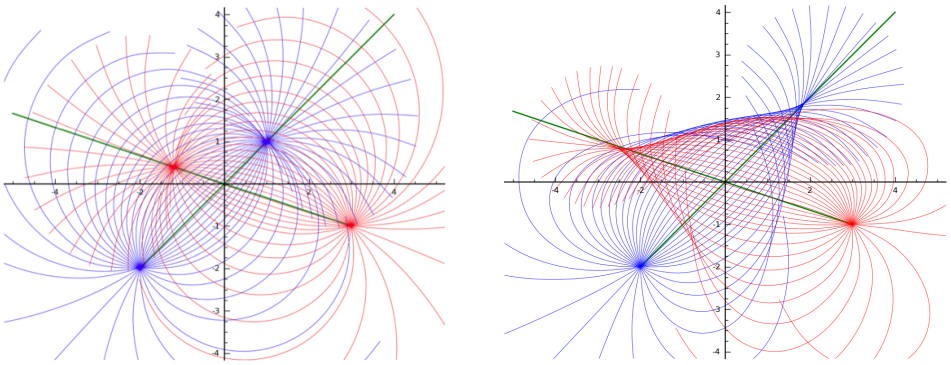
(a) **(b)**

Figure 7.4: Refractive index fields of synthetic (a) Luneburg and (b) Fisheye lenses. In both cases the values represent a refractive index delta above an ambient index n_0. Both functions are smooth but are represented here as contour plots for clarity.



(a) **(b)**

Figure 7.5: (a) Results of raytrace through discrete high resolution Luneburg Lens. Markers on the background planes denote intersections with unrefracted (o) and refracted (+) rays. (b) A much lower resolution discrete lens.



(a) **(b)**

Figure 7.6: (a) Results of raytrace through analytic fisheye lens. Two separate sources in the lower left and lower right are shown, from which rays emanate evenly across 360°. (b) Use of a low resolution discrete representation of the function and its gradients causes paths to deviate significantly.

at p in exactly the same initial direction [Andersen and Kak, 1984]. Figure 7.6(a) illustrates this behaviour for an analytic function and gradients. In comparison, the discrete version of the lens produces distorted ray paths. Figure 7.6(b) understates the severity of the problem because the underlying function has a structure that partially corrects for deviations from the correct path. In more general scenes, once a ray drifts from the true solution, it is likely to encounter other data that will cause it to deviate even further.

As with most inverse problems, the objective is highly non-convex, and so we require a good initial guess to avoid the trap of local minima. One way to obtain these is to construct a multiscale pyramid and use lower resolution solutions to initialise higher ones. Behaviour on low resolution grids therefore has a major influence on the final solution. Figure 7.7 provides further illustration of how traced ray paths can be very unreliable on coarse grids. It also shows how the choice of interpolation kernel can have a large impact.

At high resolutions there too can be challenges. Figure 7.8 shows the phantom head test scene at high resolution for an orthographic camera on the left and a perspective camera on the right. We see that rays become progressively more erratic

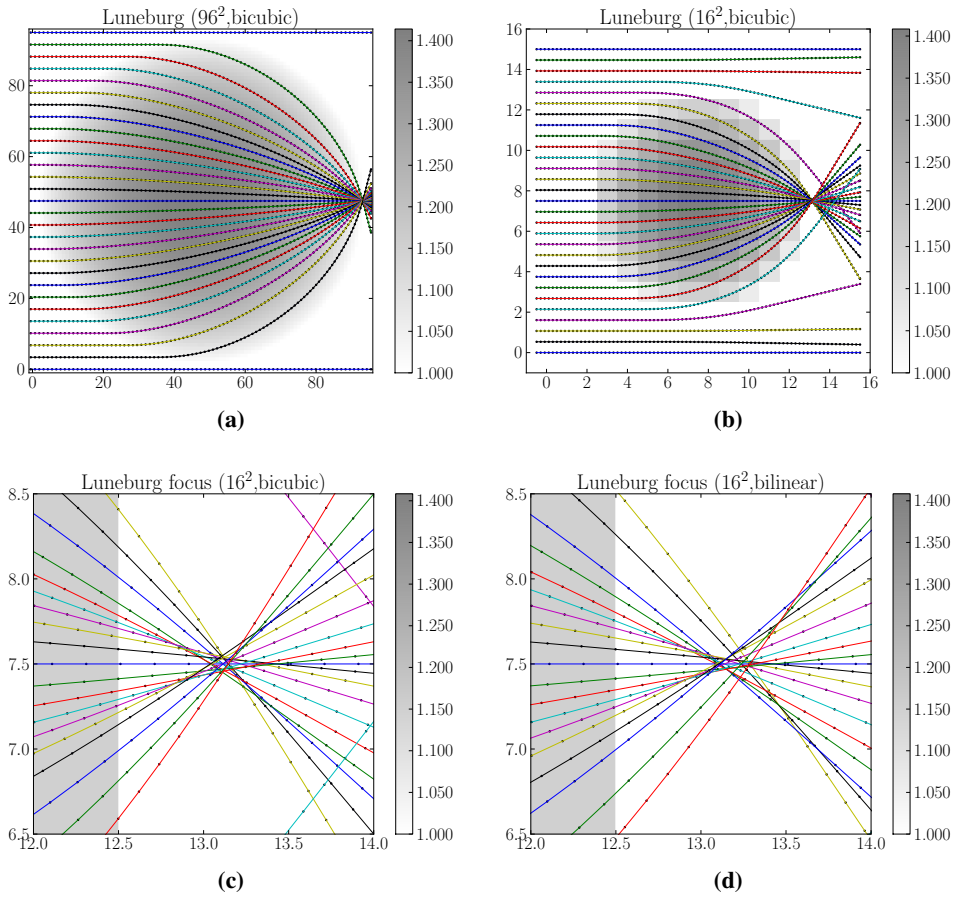


Figure 7.7: (a) Even rays that barely glance the surface of a high resolution Luneburg Lens behave as expected, regradless of the interpolation kernel used. (b) At lower resolutions errors become more predominant, as the precise location of edges becomes harder to define. (c) Close-up view of the focal point for low resolution lens, using bicubic interpolation kernel. Notice the spherical aberration. (d) Switching to a bilinear kernel results in slightly increased aberration for most rays, whereas some are missing completely. The grey region represents the underlying index field sampled with a nearest-neighbour kernel for illustrative purposes. The interpolated field used when tracing rays is smooth so we do not expect sharp bends at the interface.

as the index delta increases up to 0.5. Even at the lower end of the scale, rays can deviate from their unrefracted paths by distances many times that of the interpolation kernel support. Total internal reflection is also evident inside the thin skull wall. At high indices, high frequency features cause extremely erratic behaviour. Such scenes do not pose a problem for traditional ray tracers when the materials are opaque and reflective, and the surrounding medium is homogeneous. When the object geometry is represented explicitly as triangles or other such primities, ray intersections can be accurately computed and normals are easily obtained. However, when tracing through a continuous medium edge information is not available. We attempted to infer when rays were undergoing total reflection, or glancing off a nearly parallel edge, by computing the Fresnel term. This did not solve the problems because of the difficulty selecting an appropriate cutoff threshold for discarding rays.

In addition to analytic lenses, we also investigated discrete ones. Figure 7.9 shows ground truth and reconstructions for a two small synthetic test scenes, and Figure 7.10 shows the phantom head. In these tests we see that the rough shape can be reconstructed for low index deltas, although the results are certainly not of high quality. This illustrates the primary difficulty with descent-based optimisation schemes on this problem. The nonconvexity ensures that the algorithm often becomes trapped in local minima, and is very sensitive to the initial guess.

Aside from multiple local minima, one could ask whether multiple global minima i.e., isomorphic solutions, exist? In certain contrived examples they do. The N-Queens problem is well known to have multiple solutions where lines along the cardinal and diagonal axes pass through exactly one queen. With orthographic cameras positioned at 45° around such a board, an absorption tomography algorithm would be unable to distinguish between the various solutions. Refractive tomography would produce different measurements for each however, because the total deflection is a function of where exactly the ray strikes the inhomogeneity. For any fixed camera configuration, a pathological refractive index field of sufficiently tiny (below grid resolution) beads could be constructed such that each ray is deflected a finite number of times to exit in eactly the same position and direction as an undisturned ray would. In general though, any isomorphic index fields with respect to rays from one camera are likely to not be isomorphic with respect

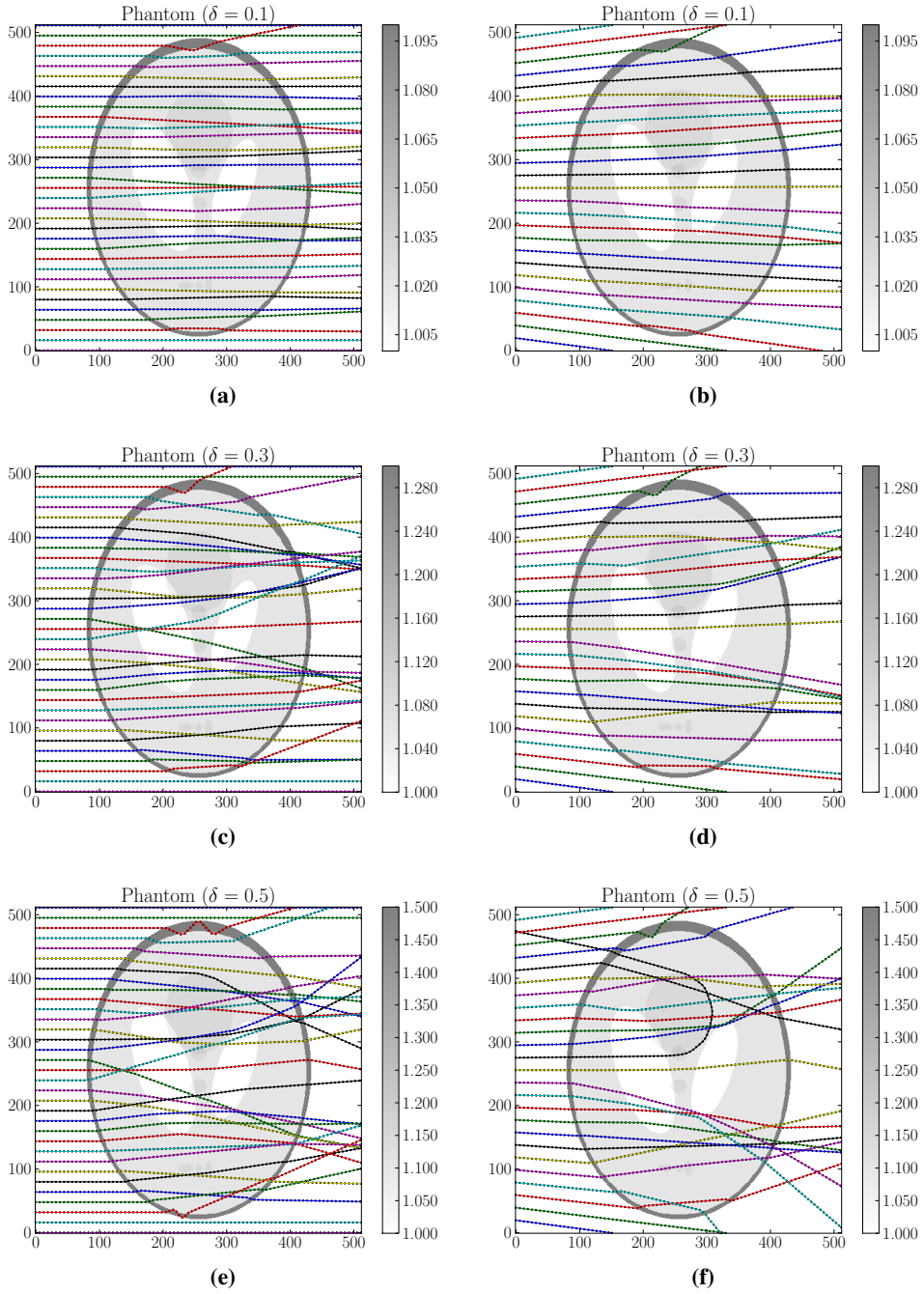


Figure 7.8: Ray trajectories for high resolution discrete phantom head for orthographic (left) and perspective (right) camera configurations.

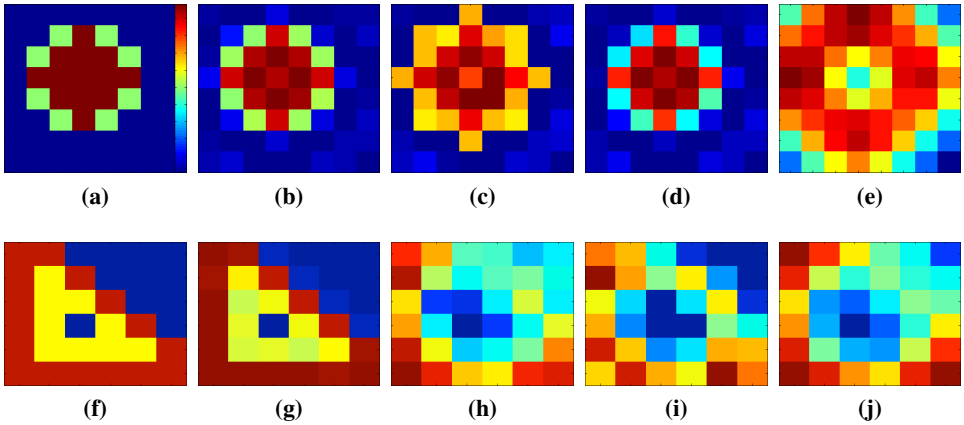
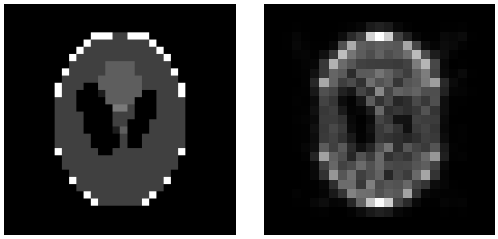


Figure 7.9: (a) "Blob" synthetic test scene ground truth. Resolution is 8×8 pixels. Refractive index delta is 0.2. (b) through (e) show representative results of local minima obtained while exploring the space of camera resolutions and algorithm parameters. (f) "Wedge" synthetic test scene ground truth. Resolution is 8×8 pixels (additional zero-padding around the scene is added during reconstruction). Refractive index delta is 0.1. (g) Best reconstruction, obtained with 64 cameras of 100 rays each. (h) through (j) show various local minima obtained while varying algorithm parameters.

to other cameras. A more realistic example of an unresolvable solution would be where caustics are cast upon the measurement planes. In such cases, immersion in water could remove the caustics and potentially lead to a solution.

7.4 Relation to Seismic Tomography

One domain in which inversion of large scale refractive problems is successful is seismic tomography. The primary difference between it and our method is that we make use geometric optics exclusively and acquire direction and displacement rather than travel time. The practical constraints of acquisition force these choices upon us, and at first glance they appear reasonable. Optical rays have much higher frequency than the features we wish to recover. However, it is on the algorithmic side where difficulties creep in to thwart reconstruction. To illustrate this, Tromp et al. [2004] consider a wave-based model and show precisely how its frequency



(a) (b)

Figure 7.10: (a) Ground truth 32×32 phantom head. Refractive index delta is 0.01. (b) Reconstruction after convergence (800 iterations). Results for larger index deltas do not contain any recognisable structure.

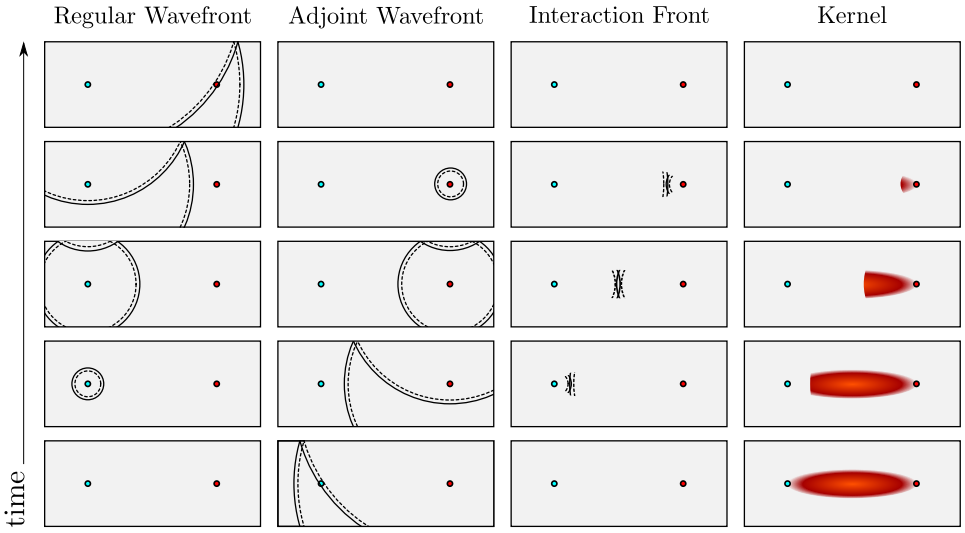


Figure 7.11: Evolution of adjoint wave interaction. Redrawn from [Tromp et al., 2004].

relates to our resolvable resolution.

Consider a single emitter buried slightly underground and a single receiver some distance away at the same depth. The intervening meterial is homogeneous. The first column of Figure 7.11 shows the propagation of two wavefronts from the source, beginning with t_0 at the bottom. The wave and its reflection off the surface expand outwards and eventually pass the receiver. The arrival time of this simulated wavefront is compared to measurements to obtain a misfit value, which

is then minimised in a gradient-descent based optimisation. In common with our problem, evaluating the gradient of this misfit function is both the key and the most computationally expensive part of solving the optimisation. Finite differences could be used but are too slow in practice. A method commonly used in seismic studies to obtain the gradient in only a single (additional) evaluation of the forward model (as opposed to one for every parameter of the model) is the Adjoint State Method [Talagrand and Courtier, 1987]. In this approach, the forward model is evaluated in time-reversed fashion from the receiver to the source. This strongly resembles what occurs in automatic differentiation and will be discussed further in the following chapter. The second column of the figure shows this *adoint* wave proceeding downwards from t_N at the top. The third column shows the interaction of the two wavefronts while the last shows the superposition of all such interactions as the adjoint wave is evaluated. It represents a kernel, and is typically toroidal in cross-section while curving downward from the source and then up again towards the receiver. It is therefore often called a *banana-doughnut* kernel. Elements of the model lying within the kernel's support affect the wavefront velocity. Notice also that it has a finite width. This width is dependent upon the wavelength – higher frequencies correspond to narrower kernels. In this we see a graphical interpretation of the infinite frequency approximation. In the limit (geometric optics) the kernel is a 1D curve, influenced only by the points through which it passes.

Now we are forced to consider the difference between the conceptual model and the implementation. Discretising the indices onto a grid restricts the minimum width of a kernel. It also necessitates indices being obtained by interpolation, which itself requires a kernel covering multiple voxels (cubic interpolation will sample 64 neighbouring voxels for each point interpolation in 3D). In addition, gradients of the index field must also be computed via finite differences, involving another small kernel, and these gradients too must be interpolated. In all, we see that in a discrete model, a ray at any point is influenced by many surrounding voxels, effectively preventing the banana doughnut kernel from collapsing to a line.

This effective minimum width of the kernel imposes a constraint on the minimum resolvable feature size. When using cubic interpolation and central divided differences for gradients, the kernel support is six pixels wide in each dimension. At very high resolutions this is less of an issue, but with low resolution grids this ef-

fectively means that the geometric ray model is only valid if the minimum feature size is a large fraction of the entire grid. We were thus unable to use a multi-scale pyramid strategy for solving the optimisation problem. Surprisingly, it is not the high resolution levels of a pyramid that are more difficult, but rather the low resolution levels, where the geometric ray model is no longer valid. As Dahlen et al. [2000] note, "geometrical ray theory provides an adequate basis for seismic traveltime tomography only if the cross-path scale length of the wave-speed heterogeneity DC is much greater than the width of the banana-doughnut kernel." Use of a geometric model requires us therefore to jump directly to a high resolution grid. The initial guess on such a grid would likely be very far away from the true solution and therefore it becomes very difficult to solve from a theoretical standpoint given the nonconvexity, and a practical standpoint given the size of the system. In the following chapter we discuss a potential approach that future research may take towards address the practical difficulties.

Chapter 8

Discussion and Conclusion

"Progress in science depends on new techniques, new discoveries and new ideas, probably in that order."

— Sydney Brenner (1980)

In the relatively underexplored area of transparent media acquisition we have investigated tools and algorithms for capturing gases, solids and even some special materials such as birefringent crystals. In this chapter we summarise our contributions to camera synchronisation and calibration, light transport acquisition, and refractive tomography.

8.1 Consumer Camcorder Arrays

Camera technology has seen steady progress since the advent of digital photography. Sensors have exceeded film in terms of resolution, dynamic range and sensitivity and will continue to improve. Cameras themselves have also gained sufficient bandwidth, storage and processing power as to enable new and interesting forms of image-based acquisition. One aspect of camera technology that has been relatively underserved by recent advances is multi-view acquisition. Light-field sensors are becoming available [Ng, 2006] and provide increased angular resolution. Integrated stereo cameras are not a new development but are now entering the consumer space and inspiring creative exploration. Wide baseline photograph collections are being combined to build models of objects at the scale of entire

buildings and cities [Snavely et al., 2006] but these operate by projecting down the temporal dimension onto a static model.

In all cases these technologies were previously inaccessible to most due to their prohibitive cost or technical limitations. Their democratisation into consumer space has sparked interest in hitherto unexplored applications and driven further improvements.

Temporally synchronised wide-baseline multi-view image acquisition currently requires specialised machine vision hardware. A complete system requires multiple components connected via high bandwidth wired links. While not a concern with only a few cameras, there are significant practical difficulties in accommodating the bandwidth and storage requirements of dozens of high resolution uncompressed video streams. It may seem counterintuitive to replace such hardware with comparatively poorer quality consumer-grade cameras. On-board compression and storage are great boons, but do significantly degrade image quality. However, the examples in this thesis, along with other related projects [Bradley et al., 2010; Gregson et al., 2012] have succesfully demonstrated that such an array can in fact be used for scientific applications. The image quality is good enough to perform visually-correct reconstructions of scenes at everyday human scales (e.g., fluids, cloth and human faces). Temporal synchronisation and accurate geometric calibration are the essential components necessary to make such arrays practical and inspire new applications, and in this thesis we have made contributions to both of these aspects.

We must mention one cause for concern: relying on consumer-grade hardware brings with it the benefits of steady improvement and vast economies of scale, but also the lack of emphasis on specialised needs. While scientific cameras provide highly configurable interfaces, consumer cameras unfortunately often provide only "fully automatic" modes and do not expose control over even such basic parameters as aperture size and exposure time, let alone full access to the image processing pipeline. In addition, capture parameters can change over time in response to scene content, and these workings often remain proprietary secrets. Future research in this area will likely account algorithmically for this variation, although we hope to see new and accessible tools made available by the open-source hardware community, spurred on perhaps by the Frankencamera [Adams et al., 2010].

8.1.1 Camera Synchronisation

Temporal synchronisation of a camera array is a key difference separating the reconstruction of static and dynamic models. Its level of accuracy determines the class of media that can be captured. Slow-moving objects can be captured with only a rough synchronisation, whereas faster media like turbulent fluids require more accurate (millisecond or better) synchronisation. In this thesis we have addressed both static and dynamic scenes, and for the dynamic scenes, developed a means for achieving very accurate synchronisation using a combination of physical and algorithmic controls.

While we did experiment with wide-angle IR triggers and custom wired control boxes to trigger multiple cameras, their polling mechanisms make it impossible to perfectly synchronise without external intervention. Our solution exploits an otherwise undesirable feature of the camera (the rolling shutter) to provide accurate synchronisation via an external piece of equipment over which we have more control (stroboscopes). In controlled darkroom environments use of strobes almost completely solves the problem, while in lit environments they give us initial temporal offsets that can be used to synchronise offline.

A key benefit of postponing the synchronisation is that we avoid the need to perform additional image processing on the raw data. Such operations are generally lossy, and we wish to delay them until after pipeline stages that require higher quality data. In our gas reconstruction pipeline, the 2D flow fields are of lower resolution than the input images, and so we were able to synchronise flow fields with little loss in image quality.

In the time since our work on camera synchronisation was published, others too have recognised the benefits to be had in analysing rolling-shutter video. In particular, Grundmann and Essa [2011] have used it to great effect in image stabilisation for amateur video. We see this as an area of significant innovation in the near future as large numbers of CMOS-based wearable and airborne sensors come online.

8.1.2 Calibration Tags

Another critical aspect of working with multi-view image data is accurate geometric calibration. This has been a well studied problem in computer vision and is considered a solved problem today [Hartley and Zisserman, 2004]. There are two fundamental approaches: target-based and structure from motion. The latter seeks to infer pose from feature points in natural scenes, and is clearly more desirable in real-world applications. In many cases it is the only feasible method, and active research in this area is making it more and more accurate. However, for controlled environments, target-based calibration is certainly feasible, and provides high accuracy. It is therefore a natural choice for our multi-camera array.

While the algorithmic aspect of determining 3D pose from 2D image correspondences is a solved problem, the practical aspect of obtaining those 2D correspondences is not. Chequerboards are frequently used as targets because they have large-scale structure that can be easily and quickly located in an image, as well as small scale features (corners) that can be accurately located. This localisation is easily automated with subpixel corner finders as long as an initial guess is provided. It is producing the initial guess that we sought to address with CALTag.

This problem has received little attention thus far because camera arrays such as ours are currently rare. Manual intervention is feasible when dealing with only a few cameras, but does not scale. Given their simple appearance, chequerboards ought to be straightforwardly detected automatically. The recent commercial success of human face and vehicle detection demonstrates that far more complex objects can routinely be detected with high accuracy. Somewhat surprisingly then, one does not today easily find reliable chequerboard detectors in popular vision software. The small body of recent work in this area has typically relied on parallel line detection and grouping via common vanishing points [Wang et al., 2007]. This does not however solve the important problem of identification. Only locating corner points in images does not provide the necessary information in order to perform calibration. Our solution takes its cue from the world of AR and augments a target with specially designed codes so as to robustly provide unambiguous identifying information. In addition to the code design, we also develop an image processing pipeline to locate these targets in images.

The resulting system both eliminates tedious and error-prone manual labour, and solves the aperture problem that arises when cameras are positioned with only a subset of the target in their field of view.

8.2 Pixel Correspondences

For nearly homogeneous refractive media, optical flow suffices to measure refractions. These can be used directly in environment matting, or else as input to a tomographic reconstruction. We have demonstrated that in cases where the paraxial approximation is appropriate, simple single-frame methods are all that is necessary for time-resolved reconstructions of interesting refracting media. However, the same cannot be said for high index-gradient media. Everyday glass objects cause so much distortion that traditional optical flow cannot be used. In this thesis we developed a different, multi-image approach, more akin to light transport acquisition. It provides the same data, at the cost of limiting us to static media.

The general light transport problem has many applications and can be solved in many ways. For our niche application of ray refraction measurements, the need is for a method that can acquire very small, high frequency PSFs that are characteristic of transport through transparent yet nontrivial media. Transport through trivial media (glass panes) is easily mapped via simpler techniques, whereas highly detailed media are beyond the capabilities of our method. However, some everyday objects are complex enough to significantly distort light passing through them, yet still simple enough that one could recognise a natural scene viewed through it. We believe that such a class of object, in addition to being something not easily scanned by existing technologies, is a reasonable proxy for interesting refractive media in other domains (geological, medical, etc.). In some cases, refraction can be mitigated by immersion in a suitable fluid. However we cannot rely solely on this approach since it is not always feasible, and since it does not at all address heterogeneous index materials. When it can be achieved, reducing refraction does make acquisition easier and potentially opens up the possibility for tomographic reconstruction along the lines of our low-index reconstructions in Chapter 6. Our results indicate that whenever geometric ray models are appropriate, and acquisition resolutions are high enough relative to the features in the media being scanned,

then we can accurately record exit ray deviations.

Most prior methods for establishing pixel correspondences are based on matching spatio-temporal intensity patterns. These produce qualitatively good visual results, but lack guarantees on correctness. We have proposed instead to assign unique temporal binary codes on the tile level and to demultiplex them after transmission through the optical projector-camera system. This opens up the possibility of using tools from Digital Signal Processing (DSP) to ensure that each code is accurately read. One possible direction for future work would be to insert error detection and correction codes into the signals.

Our current binary signal decoding scheme employs compressed sensing and spatial heuristics to demultiplex signals. We have introduced the Bloom filter as an optical computing tool for determining one-to-few pixel correspondences. Results show that it can recover them when those correspondence points are spatially distant. However, without more advanced DSP techniques, we cannot accommodate one-to-many correspondences. We therefore group pixels into tiles, and apply a separate frequency-based coding scheme to map the pixels within each tile. To this end, we have improved upon existing frequency coding methods by halving the required number of images, eliminating redundant sweep scans, and allowing for subpixel precision with nonparametric point-spread functions. Our method is also the first of which we are aware to capture high frequency multi-path light transport through birefringent materials.

8.3 Refractive Tomography

In this thesis we have demonstrated the ability to visualise time-varying gas flows based only on passive observations of their effects. The reconstructed quantity and imposed capture constraints are sufficiently different from PIV that we position our method as a complement to, rather than a replacement for PIV.

As input the method uses distortions caused by minor refractive index inhomogeneities, and for this reason it is inherently more capable than methods that seek to eliminate all refraction via immersion in a suitable fluid. For small variations in index, ray geometry is not significantly affected and we can precompute a coefficient matrix describing the interaction between each ray and voxel pair. This

matrix can be stored using a sparse representation and used to solve a set of linear least-squares problems, whose solutions are then integrated to produce our final solution. For stronger optical inhomogeneities, the problem is no longer linear, in that ray geometry is significantly dependent upon the unknown variables. Our approach in this thesis has been to address this new problem via an iterative extension of the linear version of the problem.

There are two primary challenges that arise when doing so. The first is related to the nonconvexity of the objective function. As is the case with most inverse problems, the objective has many local minima and the only feasible way to apply gradient-based optimisation methods is to have a sufficiently close initial guess. While we were able to make use of the visual hull and the expected binary distribution of indices in the glass-in-air case, this information is not sufficient. A commonly-used approach is to solve a sequence of smaller problems of increasing size, using a multiscale pyramid. We were unable to use this strategy because of the difficulties involved in tracing rays through low resolution grids. Synthetic experiments show that the interpolation kernel has an overly significant effect on the ray ODE solution and that rays cannot accurately be traced unless voxel size is smaller than the minimum feature size. This is a reflection of the difficulty inherent in choosing a model based purely on geometric optics. While we do gain in conceptual simplicity and ease of data acquisition, the geometric ray model is valid only for high frequency waves and macroscopic media. In the implementation, the use of a coarse voxel grid breaks these assumptions. One possible way to apply ray-based models would be to change the representation to something with no minimum feature size limit. Point-based [Gregson et al., 2012] or wavelet hierarchy [Peers and Dutré, 2005] representations can adapt to an arbitrary model, although one must be able to quickly sample the index value and its gradient at an arbitrary point, and to be able to express derivatives with respect to a reasonably small set of model parameters. Discretised voxel grids remain an attractive choice given these constraints.

A geometric ray model is more appropriate when tracing through a high resolution voxel grid. Unfortunately, this brings with it significant performance challenges. The fundamental operations in our framework are ray tracing, and derivative computation. Ray tracing is well suited to hardware implementation, but unfor-

tunately does not provide us with easy access to derivatives. At each iteration of the nonlinear optimisation process, we must obtain an objective value and a measure of the change in that objective with respect to each model parameter (voxel refractive index value). It is a trivial matter to retrace after slight perturbation to acquire derivatives via finite differences. Despite the low cost of hardware, this approach does not scale well with 2D and especially 3D reconstructions. Each evaluation of the forward model involves tracing rays from many pixels from many cameras, and is an expensive operation. Wrapping this process inside an $O(N^3)$ loop in order to take only one small step in an optimisation process is not a cost-effective solution.

Evaluation of the objective function takes time dependent upon grid and camera resolutions but each invocation for the perturbation of a single voxel is independent. This is therefore an embarrassingly parallel problem that can be solved with sufficient hardware. However, in a world where compute time relates directly to both money and carbon dioxide, one should consider seriously whether the effort expended in obtaining a simulation's result is worth the cost. Total cost includes development costs, and because the nature of our problem is dependent upon the relative scale of voxels to refractive features, we expect there to be significant additional development work at higher resolutions beyond what is necessary to reconstruct low resolution problems. Based on scaling experiments with small data we determined that the cost of running multiple experiments on a large GPU-equipped cluster would be prohibitively high and elected instead to investigate more efficient methods.

We chose to obtain gradients via a technique called Automatic Differentiation (AD) [Walther and Griewank, 2012]. It provides accurate derivatives up to machine precision, with the additional benefit of not requiring the user to select a finite difference step size parameter. Surprisingly, AD is a relatively under-exploited tool with a very common use-case. We anticipate that this will change as large-scale optimisation becomes more prevalent, and parallel hardware accelerators becomes more programmable. The appropriate form of AD for our problem is so-called "reverse accumulation" which records a trace of each operation performed during the objective evaluation [Rall, 1981]. It then sweeps backwards through the trace while applying the chain rule. There are strong parallels between reverse accumulation and Adjoint State time reversal [Talagrand and Courtier, 1987] (as well

as backpropagation used in neural networks [Russell and Norvig, 1995]). In our experiments on relatively small 2D synthetic problems, the temporary trace storage quickly grew to exceed main memory. Because the process is not easily parallelisable, we were constrained to small problems by the memory on a single machine.

Our approach can be described as a nonlinear variant of SART, where instead of backprojecting residual values, we iteratively adjust voxel values based on derivatives. This suggests that a less memory-intensive variant (i.e., ART) could be applied to our problem. Indeed, Aravkin et al. [2011] have shown that very large scale inversion problems can be solved via stochastic gradient descent. Their method describes an optimal strategy for approximating the gradient of a sum of many functions by a much smaller sampling of those functions. In our problem, the data misfit term is of precisely this form, and could therefore be split into sets of a few rays each, each of which could be evaluated on a separate compute node along with an automatically provided gradient. From a performance perspective this is likely the only feasible approach to solving the optimisation that allows for the very high resolution voxel grids necessary when using geometric ray optics. However, the need for an initial solution close to the global minimum remains, and we conclude that even very efficient gradient-based optimisation routines cannot succeed without having this initial guess.

Schlieren tomography is therefore a niche tool, best suited to refractive gas reconstruction. Since it requires optical transmission through the target, occlusions can pose a problem. However, the ability to acquire data using relatively cheap and easily obtainable equipment makes it a useful tool in the right circumstances.

8.4 Analysis

In the field of computer graphics this thesis relates most closely the problem of model acquisition. Ihrke et al. [2008] have divided transparent and specular scene reconstruction into a set of classes of increasing difficulty. The ideas described here advance the state of the art in their class of *volumetric, multiple scattering* materials. This is one level below the most general class, which includes occluders.

It is this inability to accommodate occlusions that is the most significant limitation of the research. Whereas PIV can be used to study airflow around bodies for

industrial applications, Schlieren tomography is fundamentally limited to purely transparent scenes. This is due to the core principle. The challenges we encountered in practice were related to implementation and computational difficulties.

The most significant strength of Schlieren tomography is its generality. Assuming a sufficiently fast and accurate measurement setup, and efficient solver, refraction-based tomography can accommodate many types of model. Unlike the methods that assume a simple heightfield model or a finite number of refractions, we could accommodate any arbitrary volume. Convexity of the boundary is a non-issue and in fact a well-defined boundary need not even exist. Dynamic scenes can be scanned if the index delta is sufficiently low. Spatial variation of opacity throughout the model has no influence on the result, provided that at least some light can pass through and the camera is sufficiently sensitive. We inherit the advantages of tomography (interior structure can be resolved) along with its impositions: multiple views are required, except in the rare cases where the model is axisymmetric.

Given the high cost of Schlieren tomography, it is unlikely to become a practical method in graphics applications. Cheaper alternatives, such as painting transparent surfaces to scan via traditional means, or computational fluid dynamics for simulating gas flows are preferable in most cases.

The relation to seismic and oceanographic tomography is more promising. This is an area currently growing to take advantage of recent advances in storing and processing massive amounts of data. While existing algorithms in these fields have been expressed in domain-specific terms, we have begun to explore the relationship between them and identify fundamental operations from other fields. Future work in refractive tomography is likely to employ a more appropriate basis than a regular grid, drawing on work in compressive sensing. The computer graphics literature describes many ways to accelerate raytracing and wavefront propagation, but there remains much to do in learning how to apply these methods to alternative bases. Automatic differentiation is currently difficult to implement in hardware, but given its myriad applications it is likely to soon become a standard tool. Stochastic gradient descent marks a return to simpler, but more scalable numerical optimisation techniques. The combination of all these emerging methods is likely to form the basis of algorithms for large scale reconstructions of the Earth, ocean, atmosphere

and potentially extra-terrestrial bodies.

While seismic tomography is known to work with wave-based models, the particular emphasis of this thesis has been on ray-based models. Our results indicate that reconstruction with such models is possible, but only if the grid resolution is high enough relative to the scale of variation of refractive index inhomogeneities. If so, there are two major requirements: an efficient solver, and a good initial guess. The former is attainable. The major challenge and most important avenue for future work therefore lies in obtaining an approximate solution close enough to the global optimum for local optimisation to work.

Bibliography

Refractive Index Database, 2012. URL http://refractiveindex.info. Accessed on 20 May 2012. → page 13

E. Abraham, A. Younus, C. Aguerre, P. Desbarats, and P. Mounaix. Refraction Losses in Terahertz Computed Tomography. *Optics Communications*, 283(10): 2050–2055, 2010. → page 6

A. Adams, M. Horowitz, S. H. Park, N. Gelfand, J. Baek, W. Matusik, M. Levoy, D. E. Jacobs, J. Dolson, M. Tico, K. Pulli, E.-V. Talvala, B. Ajdin, D. Vaquero, and H. P. Lensch. The Frankencamera. *ACM Transactions on Graphics*, 29(4): 1, 2010. → page 129

S. Agarwal, S. P. Mallick, D. Kriegman, and S. Belongie. On Refractive Optical Flow. In *European Conference on Computer Vision*, pages 483–494. Springer, 2004. → page 27

A. Agrawal, B. Albers, and D. Griffin. Abel Inversion of Deflectometric Measurements in Dynamic Flows. *Applied Optics*, 38(15):3394–3398, 1999. → pages 26 and 92

A. Agrawal, R. Raskar, and R. Chellappa. What Is the Range of Surface Reconstructions from a Gradient Field? In A. Leonardis, H. Bischof, and A. Pinz, editors, *European Conference on Computer Vision*, volume 3951/2006 of *Lecture Notes in Computer Science*, pages 578–591, Berlin, 2006. Springer Berlin/Heidelberg. → page 102

O. Ait-Aider, A. Bartoli, and N. Andreff. Kinematics from Lines in a Single Rolling Shutter Image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, Minneapolis, 2007. IEEE Computer Society. → page 56

A. Andersen. Digital Ray Tracing in Two-Dimensional Refractive Fields. *The Journal of the Acoustical Society of America*, 72(5):1593, Nov. 1982. → page 31

A. Andersen and A. Kak. Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984. → pages 30, 116, 118, and 120

A. Aravkin, M. Friedlander, F. Herrmann, and T. Van Leeuwen. Robust Inversion, Dimensionality Reduction, and Randomized Sampling. *Inverse Problems*, 2011. → pages 116 and 136

M. Arroyo and C. Greated. Stereoscopic Particle Image Velocimetry. *Measurement Science and Technology*, 2(12):1181–1186, 1991. → page 10

B. Atcheson. *Schlieren-Based Flow Imaging*. MSc Thesis, The University of British Columbia, 2007. → pages 7, 27, 28, 53, and 54

B. Atcheson and W. Heidrich. Non-Parametric Acquisition of Near-Dirac Pixel Correspondences. In *International Conference on Computer Vision Theory and Applications*, pages 247–254, Rome, Italy, 2012. → page 72

B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H.-P. Seidel. Time-Resolved 3D Capture of Non-Stationary Gas Flows. *ACM Transactions on Graphics*, 27(5):132, 2008. → pages 26, 70, 102, and 106

B. Atcheson, W. Heidrich, and I. Ihrke. An Evaluation of Optical Flow Algorithms for Background Oriented Schlieren Imaging. *Experiments in Fluids*, 46(3):467–476, Oct. 2009. → pages 96 and 97

B. Atcheson, F. Heide, and W. Heidrich. CALTag: High Precision Fiducial Markers for Camera Calibration. In *International Workshop on Vision, Modeling and Visualization*, pages 1–8, Siegen, Germany, 2010a. → page 36

B. Atcheson, F. Heide, and W. Heidrich. CALTag: Automatic Marker Detection for Camera Calibration, 2010b. URL http://www.cs.ubc.ca/labs/imager/tr/2010/Atcheson_VMV2010_CALTag/. Accessed on 4 Aug 2012. → page 34

H. Bach and N. Neuroth, editors. *The Properties of Optical Glass*. Schott Series on Glass and Glass Ceramics. Springer, 1995. → page 13

S. Baker, S. Roth, D. Scharstein, M. J. Black, J. Lewis, and R. Szeliski. A Database and Evaluation Methodology for Optical Flow. In *IEEE International*

Conference on Computer Vision, pages 1–8, Rio de Janeiro, Oct. 2007. IEEE Computer Society. → pages 27 and 97

S. Baker, E. Bennett, S. B. Kang, and R. Szeliski. Removing Rolling Shutter Wobble. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2392–2399. Technical Report MSR-TR-2010-28, Microsoft Research, 2010, IEEE, Mar. 2010. → page 57

R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. V. D. Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, 2nd edition, 1994. → page 103

J. R. Bitner, G. Ehrlich, and E. M. Reingold. Efficient Generation of the Binary Reflected Gray Code and its Applications. *Communications of the ACM*, 19(9), 1976. → pages 72, 73, and 79

J. F. Blinn and M. E. Newell. Texture and Reflection in Computer Generated Images. *Communications of the ACM*, 19(10):542–547, 1976. → page 18

B. H. Bloom. Space/Time Trade-offs in Hash Coding with Allowable Errors. *Communications of the ACM*, 13(7):422–426, 1970. → page 72

J.-Y. Bouguet. Camera Calibration Toolbox for MATLAB, 2004. URL http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed on 1 August 2012. → page 45

E. Bozda, J. Trampert, and J. Tromp. Misfit Functions for Full Waveform Inversion based on Instantaneous Phase and Envelope Measurements. *Geophysical Journal International*, 185(2):845–870, 2011. → page 32

D. Bradley, T. Boubekeur, and W. Heidrich. Accurate Multi-View Reconstruction Using Robust Binocular Stereo and Surface Meshing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Anchorage, 2008a. → pages 47, 53, and 70

D. Bradley, T. Popa, and A. Sheffer. Markerless Garment Capture. *ACM Transactions on Graphics*, 27(3):99, 2008b. → pages 47, 54, and 70

D. Bradley, B. Atcheson, I. Ihrke, and W. Heidrich. Synchronization and Rolling Shutter Compensation for Consumer Video Camera Arrays. In *IEEE International Workshop on Projector-Camera Systems*, pages 1–8. IEEE Computer Society, 2009. → page 63

D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. High Resolution Passive Facial Performance Capture. *ACM Transactions on Graphics*, 29(4):41:1–41:10, 2010. → pages 1, 54, and 129

T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In *European Conference on Computer Vision*, pages 25–36. Springer, 2004. → pages 27 and 97

R. Budwig. Refractive Index Matching Methods for Liquid Flow Investigations. *Experiments in Fluids*, 17(5):350–355, 1994. → page 22

W. Cai and V. Shalaev. *Optical Metamaterials: Fundamentals and Applications*, volume 16. Springer, 2009. → page 13

E. Candes and T. Tao. Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51(12):22, 2005. → page 80

R. Carceroni, F. Padua, G. Santos, and K. N. Kutulakos. Linear Sequence-to-Sequence Alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages I–746–I–753, Washington, 2004. IEEE Computer Society. → pages 58 and 64

Y. Caspi, D. Simakov, and M. Irani. Feature-Based Sequence-to-Sequence Matching. *International Journal of Computer Vision*, 68(1):53–64, June 2006. → page 58

T. Chen, H. P. Lensch, C. Fuchs, and H.-P. Seidel. Polarization and Phase-Shifting for 3D Scanning of Translucent Objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. → page 21

Y. Cho and U. Neumann. Multi-Ring Color Fiducial Systems for Scalable Fiducial Tracking Augmented Reality. In *Virtual Reality Annual International Symposium*, page 212. IEEE, 1998. → page 38

Y.-Y. Chuang, D. E. Zongker, J. Hindorff, B. Curless, D. H. Salesin, and R. Szeliski. Environment Matting Extensions: Towards Higher Accuracy and Real-time Capture. In *ACM SIGGRAPH*, pages 121–130, 2000. → pages 19, 72, 74, 77, and 97

P. Ciddor. Refractive Index of Air: New Equations for the Visible and Near Infrared. *Applied Optics*, 35(9):1566–1573, 1996. → page 4

S. P. Colin, J. H. Costello, L. J. Hansson, J. Titelman, and J. O. Dabiri. Stealth Predation and the Predatory Success of the Invasive Ctenophore Mnemiopsis

Leidyi. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40):17223–7, Oct. 2010. → page 3

J. Costa, A. Venetsanopoulos, and M. Trefler. Design and Implementation of Digital Tomographic Filters. *IEEE Transactions on Medical Imaging*, 2(2): 89–100, 1983. → page 101

F. Dahlen, S. Hung, and G. Nolet. Fréchet Kernels for Finite-Frequency Traveltimes - I. Theory. *Geophysical Journal International*, 141(1):157–174, 2000. → page 127

C. Dai, Y. Zheng, and X. Li. Subframe Video Synchronization via 3D Phase Correlation. In *IEEE International Conference on Image Processing*, pages 501–504, 2006. → page 58

A. De La Escalera and J. M. Armingol. Automatic Chessboard Detection for Intrinsic and Extrinsic Camera Parameter Calibration. *Sensors*, 10(3): 2027–2044, 2010. → page 37

P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the Reflectance Field of a Human Face. In *ACM SIGGRAPH*, Computer Graphics Proceedings, Annual Conference Series, pages 145–156. ACM Press/Addison-Wesley Publishing Co., 2000. → page 74

G. Elsinga, F. Scarano, B. Wieneke, and B. Oudheusden. Tomographic Particle Image Velocimetry. *Experiments in Fluids*, 41(6):933–947, Oct. 2006. → page 10

G. W. Faris and R. L. Byer. Three-Dimensional Beam-Deflection Optical Tomography of a Supersonic Jet. *Applied Optics*, 27(24):5202–5212, Dec. 1988. → page 92

M. Fiala. ARTag, a Fiducial Marker System Using Digital Techniques. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 590–596, San Diego, 2005. IEEE. → pages 38 and 47

M. Fiala and C. Shu. Self-Identifying Patterns for Plane-Based Camera Calibration. *Machine Vision and Applications*, 19(4):209–216, July 2007. → pages 38, 39, and 47

C. Fuchs, T. Chen, M. Goesele, H. Theisel, and H.-P. Seidel. Volumetric Density Capture From a Single Image. In *International Workshop on Volume Graphics*, pages 17–22, 2006. → page 91

G. Garg, E.-V. Talvala, M. Levoy, and H. P. Lensch. Symmetric Photography: Exploiting Data-Sparseness in Reflectance Fields. In *Eurographics Symposium on Rendering*, pages 251–262. Eurographics Association, 2006. → page 75

D. Gates and C. Benedict. Convection Phenomena from Plants in Still Air. *American Journal of Botany*, 50(7):563–573, 1963. → page 6

A. Gershman. ECE 761: Advanced Topics in DSP, 2000. URL http://www.ece.mcmaster.ca/~gershman/alex_advanced.html. Accessed on 26 April 2012. → page 158

J. Gladstone and T. Dale. Researches on the Refraction, Dipsersion and Sensitiveness of Liquid. *Philosophical Transactions of the Royal Society of London*, 153:317–343, 1863. → page 4

M. Goesele, H. P. Lensch, J. Lang, C. Fuchs, and H.-P. Seidel. DISCO: Acquisition of Translucent Objects. *ACM Transactions on Graphics*, 23(3): 835–844, 2004. → pages 20 and 91

E. Goldhahn and J. Seume. The Background Oriented Schlieren Technique: Sensitivity, Accuracy, Resolution and Application to a Three-Dimensional Density Field. *Experiments in Fluids*, 43(2-3):241–249, July 2007. → page 92

T. T. Goldsmith. The Refractive Index of Water for Electromagnetic Waves Eight to Twenty-Four Centimeters in Length. *Physical Review*, 51(4):245–247, 1937. → page 12

S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The Lumigraph. In *ACM SIGGRAPH*, pages 43–54, New Orleans, 1996. ACM. → page 38

I. Grant. Particle Image Velocimetry: A Review. *Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science*, 211 (1):55–76, 1997. → pages 2, 21, and 91

S. Gray. Local Properties of Binary Images in Two Dimensions. *IEEE Transactions on Computers*, 20(5):551–561, 1971. → page 44

J. Gregson, M. Krimerman, M. B. Hullin, and W. Heidrich. Stochastic Tomography and its Applications in 3D Imaging of Mixing Fluids. *ACM Transactions on Graphics*, 31(4):52:1–52:10, 2012. → pages 21, 49, 54, 70, 129, and 134

M. Grundmann and I. Essa. Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 225–232. IEEE, 2011. → page 130

M. Grundmann, V. Kwatra, D. Castro, and I. Essa. Calibration-Free Rolling Shutter Removal. In *International Conference on Computational Photography*, 2012. → page 57

D. Gutierrez, F. J. Seron, A. Munoz, and O. Anson. Simulation of Atmospheric Phenomena. *Computers & Graphics*, 30(6):994–1010, Dec. 2006. → page 16

O. Hall-Holt and S. Rusinkiewicz. Stripe Boundary Codes for Real-time Structured-Light Range Scanning of Moving Objects. *IEEE International Conference on Computer Vision*, 2:359–366, 2001. → page 73

R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, volume 16. Cambridge University Press, 2004. → page 131

S. W. Hasinoff and K. N. Kutulakos. Photo-Consistent 3D Fire by Flame-Sheet Decomposition. In *IEEE International Conference on Computer Vision*, pages 1184–1191. IEEE Computer Society, 2003. → page 91

T. Hawkins, P. Einarsson, and P. Debevec. Acquisition of Time-Varying Participating Media. *ACM Transactions on Graphics*, 24(3):812–815, 2005. → pages 2 and 91

R. Heinsohn, S. Yu, C. Merkle, G. S. Settles, and B. Huitema. Viscous Turbulent Flow in Push-Pull Ventilation Systems. In *First International Symposium for Contamination Control*, pages 529–566, Amsterdam, 1986. Elsevier. → page 6

A. Henrichsen. *3D Reconstruction and Camera Calibration from 2D Images*. MSc Thesis, University of Cape Town, 2000. → page 45

G. T. Herman, A. Lent, and P. H. Lutz. Relaxation Methods for Image Reconstruction. *Communications of the ACM*, 21(2):282–288, 1978. → page 30

B. K. Horn and B. G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17(1-3):185–203, Aug. 1981. → pages 26 and 97

B. C. House and K. Nickels. Increased Automation in Stereo Camera Calibration Techniques. *Cybernetics*, 4(4):48–51, 2006. → page 37

W. L. Howes. Rainbow Schlieren and its Applications. *Applied Optics*, 23(14): 2449–2460, July 1984. → page 26

M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. Lensch. Fluorescent Immersion Range Scanning. *ACM Transactions on Graphics*, 27(3):1, 2008. → page 22

I. Ihrke and M. Magnor. Image-Based Tomographic Reconstruction of Flames. In *Symposium on Computer Animation*, pages 365–373, Grenoble, 2004. Eurographics Association. → pages 2, 91, and 101

I. Ihrke, G. Ziegler, A. Tevs, C. Theobalt, M. Magnor, and H.-P. Seidel. Eikonal Rendering: Efficient Light Transport in Refractive Objects. *ACM Transactions on Graphics*, 26(3):59, 2007. → pages 16 and 32

I. Ihrke, K. N. Kutulakos, H. P. Lensch, M. Magnor, and W. Heidrich. State of the Art in Transparent and Specular Object Reconstruction. In *STAR Proceedings of Eurographics*, pages 87–108, 2008. → pages 2 and 136

ISO/IEC 16022:2006. Information Technology – Automatic identification and Data Capture Techniques – Data Matrix Bar Code Symbology Specification, 2006. → page 38

ISO/IEC 18004:2006. Information Technology – Automatic Identification and Data Capture Techniques – QR Code 2005 Bar Code Symbology Specification, 2006. → page 38

ITU. ITU-R BT.709, Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange. Technical report, International Telecommunication Union, 2002. → pages 60 and 64

ITU. ITU-R BT.601, Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. Technical report, International Telecommunication Union, 2007. → page 60

H. Iyer and K. Hirahara. *Seismic Tomography: Theory and Practice*. Springer, 1st edition, 1993. → pages 28 and 31

M. Jansson. Forward-Only and Forward-Backward Sample Covariances A Comparative Study. *Signal Processing*, 77(3):235–245, 1999. → page 157

H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan. A Practical Model for Subsurface Light Transport. In *ACM SIGGRAPH*, pages 511–518, 2001. → page 91

S. Kaczmarz. Angenäherte Auflösung von Systemen Linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres*, A: 355–357, 1937. → page 29

A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*, volume 33. IEEE Press, 2nd edition, 1988. → pages 29, 30, and 31

A. Karpenko, D. E. Jacobs, and M. Levoy. Digital Video Stabilization and Rolling Shutter Correction using Gyroscopes. 2011. → page 57

S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall PTR, 1993. → pages 77 and 84

K. Kindler, E. Goldhahn, F. Leopold, and M. Raffel. Recent Developments in Background Oriented Schlieren Methods for Rotor Blade Tip Vortex Measurements. *Experiments in Fluids*, 43(2-3):233–240, June 2007. → page 92

A. Kirsch and M. Mitzenmacher. Less Hashing, Same Performance: Building a Better Bloom Filter. *Proceedings Algorithms ESA*, 4168:456–467, 2006. → page 79

P. D. Kovesi. MATLAB and Octave Functions for Computer Vision and Image Processing, 2000. URL http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/. Accessed on 1 March 2010. → page 42

E. Kriezis, D. Chrissoulidis, and A. Papagiannakis. *Electromagnetics and Optics*. World Scientific, 1992. → pages 14 and 15

U. Kumar, B. Bhaduri, M. Kothiyal, and N. Mohan. Two-Wavelength Micro-Interferometry for 3-D Surface Profiling. *Optics and Lasers in Engineering*, 47(2):223–229, Feb. 2009. → page 2

K. N. Kutulakos and E. Steger. A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation. *International Journal of Computer Vision*, 76(1): 13–29, July 2008. → page 23

A. Laurentini. The Visual Hull Concept for Silhouette-Based Image Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994. → page 101

C. Lei and Y.-H. Yang. Tri-Focal Tensor-Based Multiple Video Synchronization with Subframe Optimization. *IEEE Transactions on Image Processing*, 15(9): 2473–2480, 2006. → page 58

M. Levoy. The Digital Michelangelo Project. In *International Conference on 3-D Digital Imaging and Modeling*, pages 2–11, Ottawa, 1999. IEEE Computer Society. → page 1

M. Levoy and P. Hanrahan. Light Field Rendering. *ACM SIGGRAPH*, 30:31–42, 1996. → page 27

X. Li and K. Pahlavan. Super-Resolution TOA Estimation with Diversity for Indoor Geolocation. *IEEE Transactions on Wireless Communications*, 3(1): 224–234, 2004. → page 157

C.-K. Liang, Y.-C. Peng, and H. H. Chen. Rolling Shutter Distortion Correction. In S. Li, F. Pereira, H.-Y. Shum, and A. G. Tescher, editors, *Visual Communications and Image Processing*, volume 5960, pages 1315–1322. SPIE, July 2005. → page 57

C.-K. Liang, L.-W. Chang, and H. H. Chen. Analysis and Compensation of Rolling Shutter Effect. *IEEE Transactions on Image Processing*, 17(8): 1323–1330, 2008. → page 57

C. Linz, T. Stich, and M. Magnor. High-Speed Motion Analysis with Multi-Exposure Images. In *International Workshop on Vision, Modeling and Visualization*, pages 273–281, Constance, 2008. → page 56

S. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. → page 44

D. López de Ipiña, P. R. Mendonça, and A. Hopper. TRIP: A Low-Cost Vision-Based Location System for Ubiquitous Computing. *Personal and Ubiquitous Computing*, 6(3):206–219, May 2002. → page 38

B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981. → pages 21, 26, and 97

L. Lucchese and S. Mitra. Using Saddle Points for Subpixel Feature Detection in Camera Calibration Targets. In *Asia-Pacific Conference on Circuits and Systems*, volume 2, pages 191–195, Singapore, 2002. → page 35

R. Luneburg. Mathematical Theory of Optics. *Annual Review of Fluid Mechanics*, 14(1):131–151, 1944. → page 118

J. Mallon and P. Whelan. Which Pattern? Biasing Aspects of Planar Calibration Patterns and Detection Methods. *Pattern Recognition Letters*, 28(8):921–930, June 2007. → pages 37, 38, and 47

S. R. Marschner, S. H. Westin, E. P. Lafortune, K. E. Torrance, and D. P. Greenberg. Image-Based BRDF Measurement Including Human Skin. *Eurographics Workshop on Rendering*, 5(1):1–15, 1999. → page 91

G. Meier. Computerized Background-Oriented Schlieren. *Experiments in Fluids*, 33(1):181–187, July 2002. → page 26

M. Meingast, C. Geyer, and S. Sastry. Geometric Models of Rolling-Shutter Cameras. In C. Geyer, M. Pollefeys, and X. Ying, editors, *Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, page 8, San Diego, Mar. 2005. → page 56

R. Mersereau and A. Oppenheim. Digital Reconstruction of Multidimensional Signals From Their Projections. *Proceedings of the IEEE*, 62(10):1319–1338, 1974. → page 29

B. Michelt and J. Schulze. Contact-Free Thickness Measurement of Container Glass. *Glas Ingenieur*, pages 35–37, Mar. 2006. → page 5

A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar. Bokode: Imperceptible Visual Tags for Camera Based Interaction from a Distance. *ACM Transactions on Graphics*, 28(3):1, 2009. → page 37

N. J. Morris and K. N. Kutulakos. Dynamic Refraction Stereo. In *IEEE International Conference on Computer Vision*, volume 2, pages 1573–1580, Beijing, 2005. IEEE Computer Society. → pages 21 and 91

W. Munk, P. Worcester, and C. Wunsch. *Ocean Acoustic Tomography*. Cambridge University Press, 2009. → page 30

H. Murase. Surface Shape Reconstruction of an Undulating Transparent Object. In *IEEE International Conference on Computer Vision*, pages 313–317. IEEE Computer Society, 1990. → pages 21 and 91

L. Naimark and E. Foxlin. Circular Data Matrix Fiducial System and Robust Image Processing for a Wearable Vision-Inertial Self-Tracker. In *International Symposium on Mixed and Augmented Reality*, page 27. IEEE, 2002. → page 38

S. G. Narasimhan, M. Gupta, C. Donner, R. Ramamoorthi, S. K. Nayar, and H. W. Jensen. Acquiring Scattering Properties of Participating Media by Dilution. *ACM Transactions on Graphics*, 25(3):1003, 2006. → page 91

S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast Separation of Direct and Global Components of a Scene Using High Frequency Illumination. *ACM Transactions on Graphics*, 25(3):935, 2006. → page 21

R. Ng. *Digital Light Field Photography*. PhD thesis, Stanford, 2006. → page 128

S. Nicklin, R. Fisher, and R. Middleton. Rolling Shutter Image Compensation. In *RoboCup 2006: Robot Soccer World Cup X*, volume 4434 of *Lecture Notes in Computer Science*, pages 402–409. Springer Berlin / Heidelberg, Berlin, 2007. → page 57

P. Peers and P. Dutré. Wavelet Environment Matting. In *Eurographics Symposium on Rendering*, pages 157–166, 2003. → pages 18 and 19

P. Peers and P. Dutré. Inferring Reflectance Functions from Wavelet Noise. In K. Bala and P. Dutré, editors, *Eurographics Symposium on Rendering*, pages 173–182, Konstanz, 2005. → pages 19 and 134

P. Peers, D. K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi, and P. Debevec. Compressive Light Transport Sensing. *ACM Transactions on Graphics*, 28(1):3:1–3:18, 2009. → page 75

V. Pereyra. Two-Point Ray Tracing in General 3D Media. *Geophysical Prospecting*, 40(03):267–288, 1992. → page 31

L. B. Rall. *Automatic Differentiation: Techniques and Applications*, volume 120 of *Lecture Notes in Computer Science*. Springer, 1981. → page 135

R. Roy and T. Kailath. ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984–995, 1989. → pages 78 and 156

S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-Time 3D Model Acquisition. *ACM Transactions on Graphics*, 21(3):438–446, 2002. → pages 1 and 73

S. Russell and P. Norvig. *Artificial Intelligence A Modern Approach*, volume 9. Prentice-Hall, 1995. → page 136

J. Sattar, E. Bourque, P. Giguere, and G. Dudek. Fourier Tags: Smoothly Degradable Fiducial Markers for use in Human-Robot Interaction. In *Canadian Conference on Computer and Robot Vision*, pages 165–174. IEEE, 2007. → page 38

D. Scharstein and R. Szeliski. High-Accuracy Stereo Depth Maps using Structured Light. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages I–195–I–202, 2003. → pages 73, 74, and 79

R. Schmidt. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986. → page 78

A. Schwarz. Multi-Tomographic Flame Analysis with a Schlieren Apparatus. *Measurement Science and Technology*, 7(3):406–413, 1996. → pages 26 and 92

P. Sen and S. Darabi. Compressive Dual Photography. *Computer Graphics Forum*, 28(2):609–618, 2009. → page 75

P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. Lensch. Dual Photography. *ACM Transactions on Graphics*, 24(3):745–755, 2005. → pages 75 and 84

J. Sethian and A. Popovici. 3-D Traveltime Computation using the Fast Marching Method. *Geophysics*, 64(2):516–523, 1999. → page 32

G. S. Settles. *Schlieren and Shadowgraph Techniques*. Springer, 2001. → pages 6, 10, 24, 25, 28, and 100

SFB 382. Fuel Injection Volumetric Dataset, 2005. URL http://www.volvis.org. Accessed on 1 August 2012. → page 103

T. Shan, M. Wax, and T. Kailath. On Spatial Smoothing for Direction-of-Arrival Estimation of Coherent Signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(4):806–811, 1985. → page 157

P. Shrestha, H. Weda, M. Barbieri, and D. Sekulovski. Synchronization of Multiple Video Recordings based on Still Camera Flashes. In *International Multimedia Conference*, pages 137–140, Santa Barbara, 2006. ACM. → page 58

S. N. Sinha and M. Pollefeys. Synchronization and Calibration of Camera Networks from Silhouettes. In *IEEE International Conference on Pattern Recognition*, pages 116–119, Cambridge, 2004. IEEE Computer Society. → page 58

N. Snavely, S. M. Seitz, and R. Szeliski. Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics*, 25(3):835–846, 2006. → pages 1 and 129

J. Stam and E. Langue. Ray Tracing in Non-Constant Media. In *Eurographics Workshop on Rendering*, pages 225–234, Porto, 1996. Springer London. → page 16

G. Stein. Tracking from Multiple View Points: Self-Calibration of Space and Time. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 521–527, Fort Collins, 1999. IEEE Computer Society. → page 58

C. Steinmetz. Sub-Micron Position Measurement and Control on Precision Machine Tools with Laser Interferometry. *Precision Engineering*, 12(1):12–24, Jan. 1990. → page 5

J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Texts in Applied Mathematics. Springer, 3rd edition, 2002. → page 17

D. S. Stone and S. R. Connor. Engineering Properties of High Refractive Index Optical Gels for Photonic Device Applications. In *Proceedings of the International Society for Optics and Photonics*, volume 3937, pages 144–155, 2000. → page 22

T. Strohmer and R. Vershynin. A Randomized Kaczmarz Algorithm with Exponential Convergence. *Journal of Fourier Analysis and Applications*, 15(2): 262–278, 2007. → pages 29, 30, and 117

O. Talagrand and P. Courtier. Variational Assimilation of Meteorological Observations with the Adjoint Vorticity Equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478):1311–1328, 1987. → pages 32, 126, and 135

C. Theobalt, I. Albrecht, J. Haber, M. Magnor, and H.-P. Seidel. Pitching a Baseball: Tracking High-Speed Motion with Multi-Exposure Images. *ACM Transactions on Graphics*, 23(3):540 – 547, 2004. → pages 56 and 61

P. Thévenaz, T. Blu, and M. Unser. Interpolation revisited. *IEEE Transactions on Medical Imaging*, 19(7):739–758, 2000. → page 90

M. Thomas, S. Misra, C. Kambhamettu, and J. T. Kirby. A Robust Motion Estimation Algorithm for PIV. *Measurement Science and Technology*, 16(3): 865–877, 2005. → page 84

B. Trifonov, D. Bradley, and W. Heidrich. Tomographic Reconstruction of Transparent Objects. In *Eurographics Symposium on Rendering*, pages 51–60, 2006. → pages 13, 22, 23, and 110

J. Tromp, C. Tape, and Q. Liu. Seismic Tomography, Adjoint Methods, Time Reversal and Banana-Doughnut Kernels. *Geophysical Journal International*, 160(1):195–216, 2004. → pages 14, 32, 124, and 125

R. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374. IBM, IEEE, 1986. → page 7

J. Van der Corput. Verteilungsfunktionen. In *Nederlandse Akademie van Wetenschappen*, pages 813–821, 1935. → page 78

S. van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM, 1991. → page 158

E. Van Vliet, S. Van Bergen, J. Derksen, L. Portela, and H. Van Den Akker. Time-Resolved, 3D, Laser-Induced Fluorescence Measurements of Fine-Structure Passive Scalar Mixing in a Tubular Reactor. *Experiments in Fluids*, 37(1):1–21, Apr. 2004. → page 91

L. Venkatakrishnan and G. Meier. Density Measurements using the Background Oriented Schlieren Technique. *Experiments in Fluids*, 37(2):237–247, Apr. 2004. → pages 92 and 99

A. Wächter and L. Biegler. On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming. *Mathematical Programming*, 106(1):25–57, 2006. → page 118

A. Walther and A. Griewank. Getting started with ADOL-C. In U. Naumann and O. Schenk, editors, *Combinatorial Scientific Computing*, chapter 7, pages 181–202. Chapman-Hall CRC Computational Science, 2012. → pages 118 and 135

H. Wang and R. Yang. Towards Space-Time Light Field Rendering. In *Symposium on Interactive 3D Graphics*, pages 125 – 132, Washington, 2005. ACM. → page 57

H. Wang, M. Liao, Q. Zhang, R. Yang, and G. Turk. Physically Guided Liquid Surface Modeling From Videos. *ACM Transactions on Graphics*, 28(3):90, 2009a. → pages 21 and 91

J. Wang, Y. Dong, X. Tong, Z. Lin, and B. Guo. Kernel Nyström Method for Light Transport. *ACM Transactions on Graphics*, 28(3):29:1–29:10, 2009b. → page 75

Z. Wang, W. Wu, X. Xu, and D. Xue. Recognition and Location of the Internal Corners of Planar Checkerboard Calibration Pattern Image. *Applied Mathematics and Computation*, 185(2):894–906, 2007. → pages 37 and 131

M. Wäny and G. Israel. CMOS Image Sensor with NMOS-only Global Shutter and Enhanced Responsivity. *IEEE Transactions on Electron Devices*, 50(1): 57–62, 2003. → page 56

G. J. Ward. Measuring and Modeling Anisotropic Reflection. *ACM SIGGRAPH*, 26(2):265–272, 1992. → page 91

D. R. Warrick, B. W. Tobalske, and D. R. Powers. Aerodynamics of the Hovering Hummingbird. *Nature*, 435(7045):1094–1097, 2005. → page 3

M. Wax and T. Kailath. Detection of Signals by Information Theoretic Criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(2), 1985. → page 158

J. Weickert. *Anisotropic Diffusion in Image Processing*. B.G. Teubner, Stuttgart, 1998. → pages 102 and 160

L. M. Weinstein. Large-Field High-Brightness Focusing Schlieren System. *American Institute of Aeronautics and Astronautics Journal*, 31(7):1250–1255, 1993. → pages 24 and 26

J. Westerweel. Fundamentals of Digital Particle Image Velocimetry. *Measurement Science and Technology*, 8(12):1379–1392, 1997. → page 21

G. Wetzstein, R. Raskar, and W. Heidrich. Hand-Held Schlieren Photography with Light Field Probes. In *International Conference on Computational Photography*, 2011. → pages 21, 26, and 97

Y. Wexler, A. W. Fitzgibbon, and A. Zisserman. Image-Based Environment Matting. In *Eurographics Workshop on Rendering*, pages 279–290. Eurographics Association, 2002. → page 18

R. White, K. Crane, and D. Forsyth. Capturing and Animating Occluded Cloth. *ACM Transactions on Graphics*, 26(3):34, 2007. → page 1

B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and H. Mark. High-Speed Videography Using a Dense Camera Array. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 294–301, Washington, 2004. IEEE Computer Society. → pages 54 and 56

M. Wild. Ullrich GmbH. Personal Communication, 2008. → page 5

Willow Garage. OpenCV 2.0, 2010. URL http://www.willowgarage.com/pages/software/opencv. Accessed on 1 March 2010. → page 45

C. Yu and Q. Peng. Robust Recognition of Checkerboard Pattern for Camera Calibration. *Optical Engineering*, 45(9):093201–9, Sept. 2006. → page 37

X. Zhang, S. Fronz, and N. Navab. Visual Marker Detection and Decoding in AR Systems: A Comparative Study. In *International Symposium on Mixed and Augmented Reality*, page 97. IEEE, 2002. → page 38

Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000. → pages 35 and 111

J. Zhu and Y.-H. Yang. Frequency-Based Environment Matting. In *Pacific Graphics*, pages 402–410, 2004. → pages 73, 74, 77, and 82

D. E. Zongker, D. M. Werner, B. Curless, and D. H. Salesin. Environment Matting and Compositing. In *ACM SIGGRAPH*, pages 205–214, 1999. → pages 18, 72, and 77

Appendix A

Parameter Estimation

In harmonic retrieval problems, the widely-used ESPRIT [Roy and Kailath, 1989] subspace method has high accuracy and relatively low computational complexity. It was developed to work with sensor arrays consisting of at least two identical, but displaced, elements. This underlying translational invariance induces a rotational invariance in the signal subspace measured at each of the elements. Once identified, this rotational invariance can be combined with knowledge of the array configuration to extract signal parameters.

One commonly used sensor array configuration is the uniform linear array: identical detectors equally spaced along a line in space. This corresponds to uniform sampling of the time series in Equation 5.10. We can interpret the measured signal vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$ as being comprised of two measurements, $\mathbf{x}_a = (x_1, x_2, \ldots, x_{N-1})^T$ and $\mathbf{x}_b = (x_2, x_3, \ldots, x_N)^T$, the outputs of two overlapping subarrays. Ignoring noise, these subarrays measure identical signals, modulo a (known) phase delay. In general, more than two subarrays are formed. Given a measurement $\mathbf{x}$ then, the algorithm proceeds as follows:

1. **Preprocess data** to remove DC component by subtracting the mean from $\mathbf{x}$

$$\mathbf{x} \leftarrow \mathbf{x} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{A.1}$$

2. **Compute the sample covariance matrix** by averaging the outer products

of M overlapping windows of length $L = N - M + 1$.

$$R = \frac{1}{M} \sum_{k=1}^{M} \mathbf{x}_{(k)} \mathbf{x}_{(k)}^T, \tag{A.2}$$

where $\mathbf{x}_{(k)} = (x_k, x_{k+1}, \ldots, x_{k+L-1})^T$. This represents forward smoothing, a spatial smoothing technique commonly used to decorrelate signals. The presence of highly correlated signals can reduce the rank of R and make signal detection difficult. In our application, this may occur when, for example, a ray strikes two background pixels that happen to have very similar frequencies. Maximizing the spatial distance between nearby frequencies helps to avoid this, but cannot guarantee its prevention. Forward Backward Spatial Smoothing (FBSS) can help to overcome the effects of correlated signals and guarantee a covariance matrix of full rank [Shan et al., 1985]

$$R_{(FB)} = \frac{R + JR^T J}{2}, \tag{A.3}$$

where J is a square matrix with ones on the main antidiagonal and zeros elsewhere. Although the true covariance matrix would be Toeplitz, this modified $R_{(FB)}$ matrix is persymmetric [Jansson, 1999]. The parameter L trades off increased resolution (higher L) against the ability to detect multiple coherent signals (higher M) [Li and Pahlavan, 2004]. In our experiments we have used a value of $L = \lfloor 2N/3 \rfloor$.

3. **Perform an eigen decomposition** with eigenvalues in decreasing order of magnitude along the diagonal of Λ. In practice, the Singular Value Decomposition (SVD) may be used instead.

$$R_{(FB)} = Q \Lambda Q^{-1} \tag{A.4}$$

4. **Estimate the number of signals** p by partitioning the eigenvalues into two sets $\Lambda_s = \{\lambda_1, \lambda_2, \ldots, \lambda_p\}$ and $\Lambda_n = \{\lambda_{p+1}, \lambda_{p+2}, \ldots, \lambda_L\}$. The elements of Λ_n will cluster around the noise variance, while $\lambda \in \Lambda_s$ will be significantly larger. A suitable value for p may be chosen via the Minimum Description

Length (MDL) [Wax and Kailath, 1985] by searching for the minimum $p \in \mathbb{N}_0$ of

$$-M(L-p)\log\left\{\frac{\prod_{i=p+1}^{L}\lambda_i^{1/(L-p)}}{\frac{1}{L-p}\sum_{i=p+1}^{L}\lambda_i}\right\}+\frac{p(2L-p)\log M}{2}. \tag{A.5}$$

5. **Partition the eigenvectors** into $Q=[Q_s\,|\,Q_n]$ where $Q_s=[\mathbf{q}_1\,|\,\mathbf{q}_2\,|\,\ldots\,|\,\mathbf{q}_p]$ spans the signal subspace. This is the same subspace as that spanned by the Vandermonde matrix of steering vectors

$$A=\begin{pmatrix}\alpha_1 e^{-i\omega_1 0} & \cdots & \alpha_p e^{-i\omega_p 0}\\ \alpha_1 e^{-i\omega_1 1} & \cdots & \alpha_p e^{-i\omega_p 1}\\ \vdots & \ddots & \vdots\\ \alpha_1 e^{-i\omega_1 (N-1)} & \cdots & \alpha_p e^{-i\omega_p (N-1)}\end{pmatrix}, \tag{A.6}$$

where the $\omega_i = 2\pi f_i \tau$ represent the frequencies of the p sinusoids.

6. **Solve for the subspace rotation.** Denoting $\overline{X}$ as a matrix with the bottom row removed, and $\underline{X}$ as one with the top row removed, we note that $\overline{A}D=\underline{A}$ where $D=\mathrm{diag}\{e^{i\omega_1},e^{i\omega_2},\ldots,e^{i\omega_p}\}$ due to the phase shift between the two overlapping subarrays. Now since A and Q_s span the same subspace, there exists a C such that $Q_s=AC$. Hence, as Gershman [2000] shows,

$$\underline{Q_s}=\underline{A}C=\overline{A}DC \text{ and } \overline{Q_s}=\overline{A}C \tag{A.7}$$

$$\Rightarrow\quad \underline{Q_s}C^{-1}D^{-1}C=\overline{A}DCC^{-1}D^{-1}C=\overline{A}C=\overline{Q_s} \tag{A.8}$$

$$\Rightarrow\quad \underline{Q_s}=\overline{Q_s}C^{-1}DC=\overline{Q_s}\Phi \tag{A.9}$$

We seek the diagonal of D, and hence the eigenvalues $\lambda_i^{(\Phi)}$ of Φ. The matrix Φ is best obtained using Total Least Squares [van Huffel and Vandewalle, 1991].

7. **Estimate the signal parameters.** The frequency estimates are

$$\omega_i = -\text{angle}\left(\lambda_i^{(\Phi)}\right). \tag{A.10}$$

The amplitudes and phase estimates can now be obtained from the magnitudes and angles, respectively, of the complex-valued solution to $\mathbf{x} = A'\alpha$, where A' is formed from the A above by omitting the α_i. Before solving this system however, we round the frequencies to the nearest values that were used to generate signals at the transmitter.

Appendix B

Diffusion Tensor

The diffusion tensor is derived from the structure tensor [Weickert, 1998]

$$\mathbf{J}_\sigma = K_\sigma \left(\widehat{\nabla n} \widehat{\nabla n}^T \right) \tag{B.1}$$

of the refractive index field with its components smoothed independently. We use a Gaussian filter kernel K_σ with $\sigma = 0.5$. Using an eigendecomposition $\mathbf{J}_\sigma = \mathbf{V}\Lambda\mathbf{V}^{-1}$ with $\lambda_0 \geq \lambda_1 \geq \lambda_2$, we generate the diffusion tensor by changing the eigenvalues to

$$\widetilde{\Lambda} = \begin{pmatrix} \alpha & & \\ & \alpha + (1-\alpha) e^{-\frac{\max\left|\widehat{\nabla n}\right|}{k(\lambda_0 - \lambda_1)^2}} & \\ & & \alpha + (1-\alpha) e^{-\frac{\max\left|\widehat{\nabla n}\right|}{k(\lambda_0 - \lambda_2)^2}} \end{pmatrix}, \tag{B.2}$$

where $k = 0.5 \cdot 10^{-5}$ and α is a data fidelity parameter. The diffusion tensor $\mathbf{D}$ is obtained by computing $\mathbf{D} = \mathbf{V}\widetilde{\Lambda}\mathbf{V}^{-1}$. Choosing $\alpha = 1$ results in standard Poisson integration which can be used if $\widehat{\nabla n}$ is indeed a gradient field; lower α values result in better noise removal. By analysing tests on synthetic data we found $\alpha = 0.8$ to be a good choice for the noise levels introduced by optical flow and tomographic reconstruction.