# Inverse Problems in Time-of-Flight Imaging: Theory, Algorithms and Applications

by

Ayush Bhandari

B. Tech., Jaypee Institute of Information Technology (2007)

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

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

Signature redacted

Author ..................................................
Program in Media Arts and Sciences
August 08, 2014

Signature redacted

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

Signature redacted

Accepted by ..................................................
Pattie Maes
Interim Academic Head,
Program in Media Arts and Sciences

# Inverse Problems in Time-of-Flight Imaging:
# Theory, Algorithms and Applications

by

Ayush Bhandari

Submitted to the the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 08, 2014, in partial fulfillment of the
requirements for the degree of
Master of Science

## Abstract

Time–of–Fight (ToF) cameras utilize a combination of phase and amplitude information to return real-time, three dimensional information of a scene in form of depth images. Such cameras have a number of scientific and consumer oriented applications.

In this work, we formalize a mathematical framework that leads to unifying perspective on tackling inverse problems that arise in the ToF imaging context. Starting from first principles, we discuss the implications of time and frequency domain sensing of a scene. From a linear systems perspective, this amounts to an operator sampling problem where the operator depends on the physical parameters of a scene or the bio-sample being investigated. Having presented some examples of inverse problems, we discuss detailed solutions that benefit from scene based priors such sparsity and rank constraints.

Our theory is corroborated by experiments performed using ToF/Kinect cameras. Applications of this work include multi-bounce light decomposition, ultrafast imaging and fluorophore lifetime estimation.

Thesis Supervisor: Ramesh Raskar
Title: Associate Professor

# Inverse Problems in Time-of-Flight Imaging: Theory, Algorithms and Applications

by

Ayush Bhandari

Signature redacted

Thesis Advisor ..............................................................

**Ramesh Raskar**
Associate Professor
Program in Media Arts and Sciences

Signature redacted

Thesis Reader ..............................................................

**Laurent Demanet**
Assistant Professor
Department of Mathematics
Massachusetts Institute of Technology

Signature redacted

Thesis Reader ...................................... ......

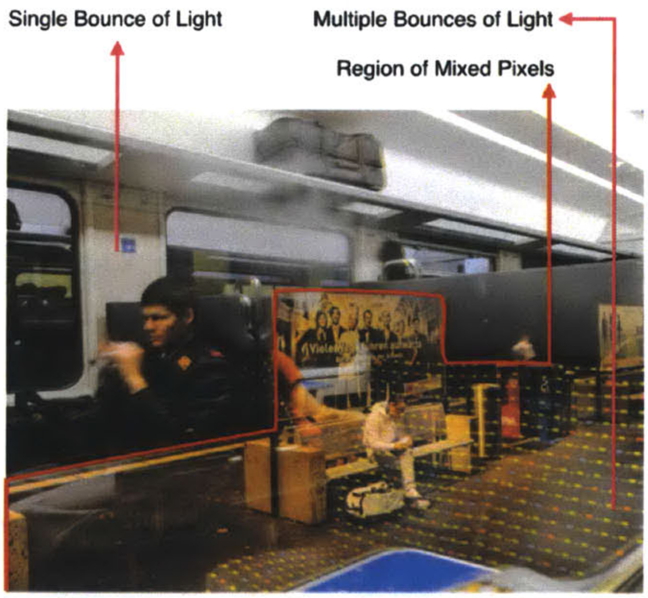
**Thierry Blu**
Professor
Department of Electronic Engineering
The Chinese University of Hong Kong

# Preface

While time–resolved imaging was already being explored in the group, first formal steps towards continuous wave time–of–flight (ToF) imaging were a consequence of the *Time–of–Flight Imaging: Algorithms, Sensors and Applications Seminar* which was held from October 21 to October 26 in 2012 at Dagstuhl.

Dr. Raskar was at the workshop. When he got back to Cambridge, he described the so called mixed–pixel problem in ToF imaging. For a given scene of interest, the ToF cameras provide an amplitude image (the usual photograph) and a depth image, that is, two images for the same exposure. The depth image is unusual. It encodes the distance of the object from the camera. Chapters 1 and 2 are dedicated to the ToF principle and how the ToF cameras compute depth.

The ability to compute depth is a speciality of the ToF cameras. That said, the ToF cameras are unable to discern the right depth/distance when multiple bounces of light combine at the sensor. To give an example, consider the setting when you photograph an object through a window. This is shown in the figure on the right. Indeed, one part of the photograph is due to the actual scene (or what you would have captured without the window) and the other part of the photograph is due to the reflections from the window pane. Ideally what one wants is a photograph of the scene without the window being there but in many cases of interest, for example, photographs taken through a aircraft, this situation is inevitable. Even the usual consumer camera will not be able to undo the effect of the window pane (unless you buy a polarizing filter). This example motivates the mixed–pixel problem, that is, multiple light bounces adding on the image as if two or more images were superimposed.



At the time when the problem was described to me, it was clear that this was a non–linear problem. Let me present an intuitive explanation. The ToF cameras measure depth by computing the the amount of time it takes for light to reflect from an object, just like the traditional SONAR[1]. That is to say, if the camera sends a signal $p$ to the object, what comes back is a delayed version of $p$. Mathematically, let the camera system send flashes of light which we model by $p(t)$ where the variable $t$ denotes time. An object at some distance $d_0$ meters from the camera will reflect the light back after $t_0 = d_0/c$ where $c = 3 \times 10^8 \mathrm{m/s}$ is the speed of light. Since this phenomenon happens only after the light has travelled a round trip, the camera records the reflected signal $r(t) = p(t - 2t_0)$. Here we have assumed a perfect reflection. Ideally, one should observe $\Gamma_0 p(t - 2t_0)$ where $\Gamma_0$ is the reflection coefficient of the scene. In summary, this interaction is modeled as,

$$p(t) \rightarrow \boxed{\text{Object}} \rightarrow r(t) = \Gamma_0 p(t - 2t_0), \quad \text{with} \quad t_0 = \frac{d}{c}, \Gamma_0 = 1.$$

This method of estimating depth or range is used across several disciplines of science and engineering (such as SONAR, RADAR and OCT[2]). Some choices of the function $p$ include Dirac's Delta function (femtosecond photography), sinusoids, Gaussian function and linear frequency modulated function (or chirps). Clearly, estimating $t_0$ and hence $d_0$ from $p$ is a non–linear problem. Consumer grade ToF cameras (such as the PMD and Microsoft Kinect) use an astonishingly simple method to decode $t_0$ from $r(t)$ with the choice of $p(t) = \cos(\omega_0 t)$.

Now if we consider the case when the camera records two reflections, one due to the window and another due to the object, the reflected signal takes form of[3],

$$r(t) = \Gamma_1 p(t - 2t_1) + \Gamma_2 p(t - 2t_2).$$

In this setting, given some form of $p$, if one can magically recover the quantities $\{\Gamma_1, \Gamma_2\}$ and

---

[1] Sound Navigation and Ranging.

[2] Optical Coherence Tomography.

[3] Also, following the same arguments as above, it is some what reasonable to believe that for the general case of $K$–bounces of light, one will end up observing, $r_K(t) = \sum_{k=0}^{K-1} \Gamma_k p(t - 2t_k)$. In this case, the problem is that of recovering the $2K$ quantities $\{\Gamma_k, t_k\}_{k=0}^{K-1}$.

$\{t_1, t_2\}$, then it is possible to completely characterize the scene! As will be shown later, this model also mimics the physics of the fluorescence phenomenon. Using similar principles, we will show that a variant of the same model can be used for fluorescence lifetime estimation.

This brings us back to the first ideas that were discussed in November 2012. The main theme of the problem was the following: Let $p(t) = e^{\jmath\omega t} \equiv \cos(\omega t) + \jmath \sin(\omega t)$. Given that for two bounces of light, the camera measures,

$$\begin{aligned} r(t) &= \Gamma_1 e^{\jmath\omega(t-2t_1)} + \Gamma_2 e^{\jmath\omega(t-2t_2)} \\ &= e^{-\jmath\omega t} \underbrace{\left(\Gamma_1 e^{-\jmath\phi_1\omega} + \Gamma_2 e^{-\jmath\phi_2\omega}\right)}_{m(\omega)}, \quad \phi_k = 2t_k, \end{aligned}$$

it is clear that for some fixed $\omega = \omega_0$, by no means one can estimate $\{\Gamma_1, \Gamma_2\}$ and $\{t_1, t_2\}$. This is simply because the quantities of interest, that is $\{\Gamma_1, \Gamma_2\}$ and $\{t_1, t_2\}$, are a function of $\omega$.

Hence if we at all seek to solve for $\{\Gamma_1, \Gamma_2\}$ and $\{t_1, t_2\}$, we need to measure $m(\omega)$,

$$m(\omega) = \Gamma_1 e^{-\jmath\phi_1\omega} + \Gamma_2 e^{-\jmath\phi_2\omega}.$$

The physical significance of this fact is that the ToF camera needs to probe the scene at multiple frequencies $\omega$.

Adrian Dorrington[4] and co–workers from University of Waikato made one of the first attempts to solve this problem by using two frequency measurements. Their idea was simple. Let $\widetilde{m}(\omega_0)$ be the measurement at $\omega = \omega_0$. Then, the solution for the unknowns may be obtained by solving for,

$$\underset{\{\Gamma_1,\Gamma_2,\phi_1,\phi_2\}}{\arg\min} \left(|m(1) - \widetilde{m}(1)|^2 + |m(2) - \widetilde{m}(2)|^2\right).$$

Unfortunately, with only two measurements, the problem is under–determined and the cost function is non–convex. Therefore it is safe to assume that this method may not always

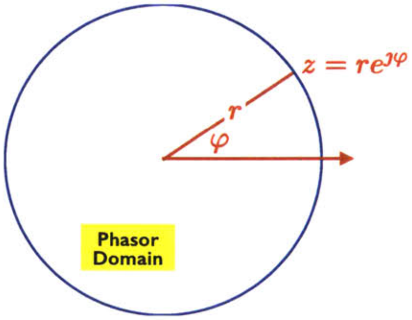
---

[4] Dorrington, Adrian A., et al. "Separating true range measurements from multi-path and scattering interference in commercial range cameras." IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2011.

guarantee a robust, unique solution.

For any one who has spent time with Thierry Blu and Martin Vetterli, it is clear that given $2K$ samples of $\widetilde{m}(\omega)$, at least in the noiseless setting, it is possible to compute all the unknowns $\{\Gamma_k, t_k\}_{k=0}^{K-1}$.

This is the classic spectral estimation problem[5]. In fact, a more general version of the problem that is linked with Fractional Fourier Transforms was dealt with in one my previous work on sparse sampling[6]. At first it might look impossible to solve for the unknowns in the non–linear mixture linked with $m$, but thanks to the age old result of de Prony[7].



Assuming a modest background in complex analysis, let me denote a phasor by the complex number $z = re^{\jmath\varphi}$ which is nothing but a vector in polar co–ordinates. This vector has length $r$ and is oriented at an angle $\varphi$. For the two bounce problem—the simplest example of mixed–pixel problem—the measurements $m(\omega)$ amount to addition to two phasors at some given frequency $\omega = \omega_0$. If $m(\omega)$ was known for all values of $\omega$, the question we should be asking is: How can we recover phasors,

$$\Gamma_1 e^{-\jmath\phi_1\omega} \quad \text{and} \quad \Gamma_2 e^{-\jmath\phi_2\omega}.$$

Concretely, in Fig. 0-1 we describe the geometry of the problem. Note that when $\omega = k\omega_0$ for integer values of $k$ and some fixed $\omega_0$, that is to say,

$$m(k\omega_0) = \Gamma_1 e^{-\jmath\phi_1 k\omega_0} + \Gamma_2 e^{-\jmath\phi_2 k\omega_0},$$

the two phasors $\Gamma_1 e^{-\jmath\phi_1\omega}$ and $\Gamma_2 e^{-\jmath\phi_2\omega}$ rotate at proportional angles. This subtle fact can be exploited as follows. Without loss of generality let $\omega_0 = 1$ and let,

$$m_k = \Gamma_1 e^{-\jmath\phi_1 k} + \Gamma_2 e^{-\jmath\phi_2 k}.$$

---

[5] Petre Stoica and Randolph L. Moses. "Introduction to Spectral Analysis." 1997.

[6] Ayush Bhandari and Pina Marziliano. "Sampling and reconstruction of sparse signals in fractional Fourier domain." 2010.

[7] Baron Gaspard Riche de Prony, "Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l?alkool,a différentes températures." 1795.

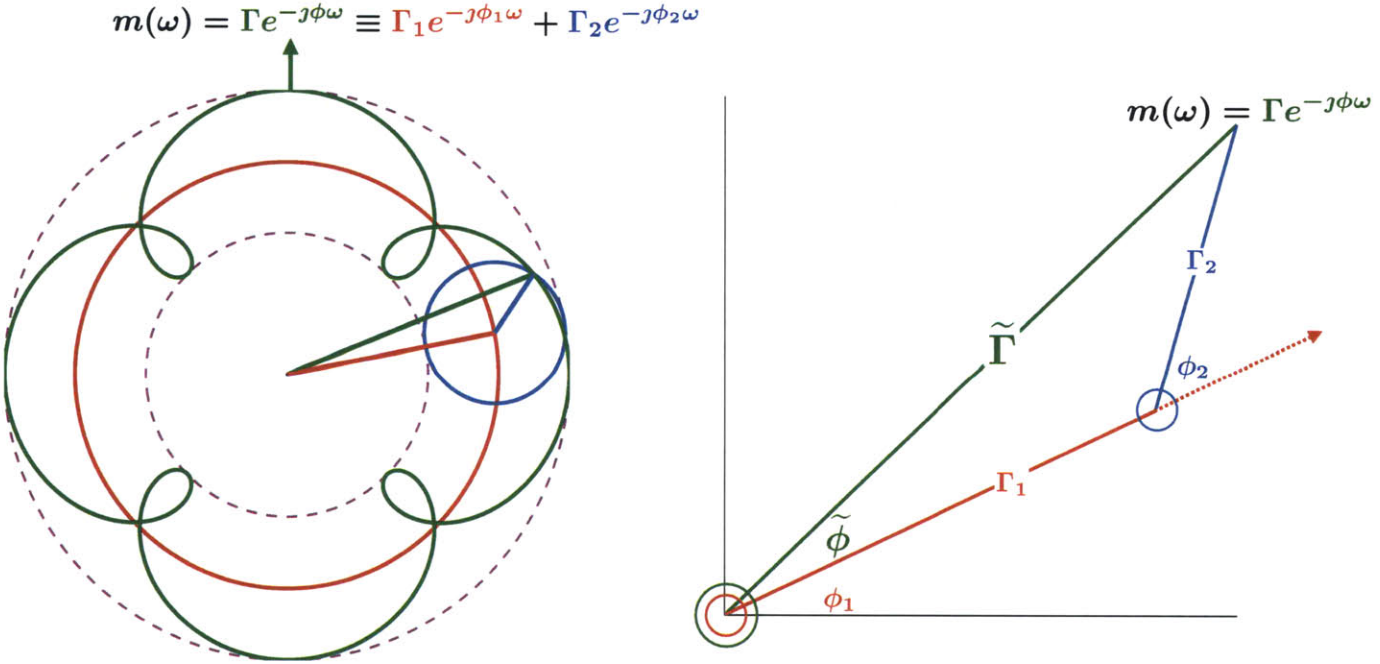


Figure 0-1: Geometry of the mixed–pixel measurements $m(\omega) = \Gamma_1 e^{-\jmath\phi_1\omega} + \Gamma_2 e^{-\jmath\phi_2\omega}$. For the two bounce case, each depth $d_k$ resulting in the time delay $t_k$ parametrizes the angle of the phasor with $\phi_k = 2t_k$. The intensity of the pixel $\Gamma_k$ parametrizes the length of the phasor. The two phasors add to produce the measurements $m(\omega) = \Gamma e^{-\jmath\phi\omega}$. The inverse problem is to estimate $\{\Gamma_1, \Gamma_2\}$ and $\{t_1, t_2\}$ given $m$.

Since $K = 2$ for the case of two bounces of light, we will show that algebraically, it is possible to find the unknowns or the phasors by using $2K = 4$ measurements.

Let us start with 4 measurements stacked in a vector–matrix form,

$$\begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ e^{\jmath\phi_1} & e^{\jmath\phi_2} \\ e^{\jmath 2\phi_1} & e^{\jmath 2\phi_1} \\ e^{\jmath 3\phi_1} & e^{\jmath 3\phi_2} \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \Leftrightarrow \begin{aligned} m_0 &= \Gamma_1 + \Gamma_2 \\ m_1 &= \Gamma_1 e^{\jmath\phi_1} + \Gamma_2 e^{\jmath\phi_2} \\ m_2 &= \Gamma_1 e^{\jmath 2\phi_1} + \Gamma_2 e^{\jmath 2\phi_2} \\ m_3 &= \Gamma_1 e^{\jmath 3\phi_1} + \Gamma_2 e^{\jmath 3\phi_2}. \end{aligned}$$

Setting $u_1 = e^{\jmath\phi_1}$ and $u_2 = e^{\jmath\phi_2}$ allows one to write,

$$\begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ u_1 & u_2 \\ u_1^2 & u_2^2 \\ u_1^3 & u_2^3 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \Leftrightarrow \begin{aligned} m_0 &= \Gamma_1 + \Gamma_2 \\ m_1 &= \Gamma_1 u_1 + \Gamma_2 u_2 \\ m_2 &= \Gamma_1 u_1^2 + \Gamma_2 u_2^2 \\ m_3 &= \Gamma_1 u_1^3 + \Gamma_2 u_2^3. \end{aligned}$$

Now if we define a polynomial $\mathrm{P}(z) = (z - u_1)(z - u_2) = z^2 + \rho_1 z + \rho_0$, it is clear that,

$$\Gamma_1 \underbrace{\overbrace{\mathrm{P}(u_1)}^{u_1^2+\rho_1 u_1+\rho_0}}_{=0} + \Gamma_2 \underbrace{\overbrace{\mathrm{P}(u_2)}^{u_2^2+\rho_1 u_2+\rho_0}}_{=0} = 0$$

and,

$$\Gamma_1 u_1 \underbrace{\overbrace{\mathrm{P}(u_1)}^{u_1^2+\rho_1 u_1+\rho_0}}_{=0} + \Gamma_2 u_2 \underbrace{\overbrace{\mathrm{P}(u_2)}^{u_2^2+\rho_1 u_2+\rho_0}}_{=0} = 0.$$

Combining the two equations, we have,

$$\begin{aligned} \rho_0 (\Gamma_1 + \Gamma_2) + \rho_1 (\Gamma_1 u_1 + \Gamma_2 u_2) + (\Gamma_1 u_1^2 + \Gamma_2 u_2^2) &= 0 \\ \rho_0 (\Gamma_1 u_1 + \Gamma_2 u_2) + \rho_1 (\Gamma_1 u_1^2 + \Gamma_2 u_2^2) + (\Gamma_1 u_1^3 + \Gamma_2 u_2^3) &= 0 \end{aligned}$$

or simply,

$$\underbrace{\begin{bmatrix} m_0 & m_1 \\ m_1 & m_2 \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} \rho_0 \\ \rho_1 \end{bmatrix}}_{\mathbf{r}} = \underbrace{\begin{bmatrix} -m_2 \\ -m_3 \end{bmatrix}}_{\mathbf{m}}.$$

This simple computation shows that it is possible to calculate $\rho_0, \rho_1$ from the measurements,

$$\left\{ \begin{matrix} m_0 & m_1 & m_2 & m_3 \end{matrix} \right\}.$$

Provided that $m_2 m_0 \neq m_1^2$, the matrix $\mathbf{M}$ is invertible and we have,

$$\rho_0 = \frac{m_2^2 - m_1 m_3}{m_1^2 - m_0 m_2} \quad \text{and} \quad \rho_1 = \frac{m_0 m_3 - m_1 m_2}{m_1^2 - m_0 m_2}.$$

Having computed $\rho_1$ and $\rho_2$, solving for, $z^2 + \rho_1 z + \rho_0 = 0$ gives $u_1 = e^{\jmath\phi_1}$ and $u_2 = e^{\jmath\phi_2}$. With $\phi$ and $m$ known, $\Gamma$ can be computed using least–squares problem,

$$\begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ e^{\jmath\phi_1} & e^{\jmath\phi_2} \\ e^{\jmath 2\phi_1} & e^{\jmath 2\phi_1} \\ e^{\jmath 3\phi_1} & e^{\jmath 3\phi_2} \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix}.$$

As has been established, a simple series of algebraic manipulations decouples the non–linear estimation problem into a root–finding problem followed by a linear inverse problem. This connection was known to me for a while and hence the extension to the case of $K$–bounces was straightforward.

The idea that the delays $\phi_k$ can be decoupled from the non–linear mixture $m$ has an interesting link with systems identification theory. An overview of the idea is shown in Fig. 0-2. For the case of $K$–bounces of light, we have,

$$m(\omega) = \sum_{k=0}^{K-1} \Gamma_k u_k^{\omega}, u_k = e^{\jmath\phi_k}.$$

Given discrete measurements of form,

$$m_\ell = m(\ell\omega_0), \quad \ell \geq 0 \text{ and } \ell \in \text{Integers},$$

its Z–transform is defined as,

$$\mathrm{M}(z) = \underbrace{\sum\nolimits_{\ell} m_{\ell} z^{-\ell}}_{\text{Z−transform}} = \sum\nolimits_{k=0}^{K-1} \frac{\Gamma_k}{1 - u_k z^{-1}} = \frac{\mathrm{N}(z)}{\mathrm{D}(z)},$$

provided that the series converges. As can be seen, $\mathrm{M}(z)$ blows–up whenever $z = u_k$. Also, since $\mathrm{M}(z)$ can be factored into a polynomial ratio, the roots of denominator, that is, $\mathrm{D}(z)$ encode the locations of $u_k$.

In this way, by the end of the year 2012, we were convinced that there are some interesting directions linked with mixed–pixel problem. That led to investigation of this topic. While a lot of the work is still in progress, one of the common themes in this thesis is linked with the idea of converting inverse problems into a rational polynomial estimation problem. In this context, there are three major directions.

The first direction solves the mixed–pixel problem in closed–form and for the general setting of $K$–bounces. Our theory is corroborated by experiments on the customized PMD sensor as well as the Kinect Xbox One. The second direction deals with super–resolution ToF imaging where the sparse deconvolution problem is converted to a frequency estimation problem. Finally, the last theme of this work is linked with fluorescence lifetime estimation. In this case, the lifetime parameter as well as the distance of the bio–sample from the wall are encoded in the depth image in form of a rational polynomial.

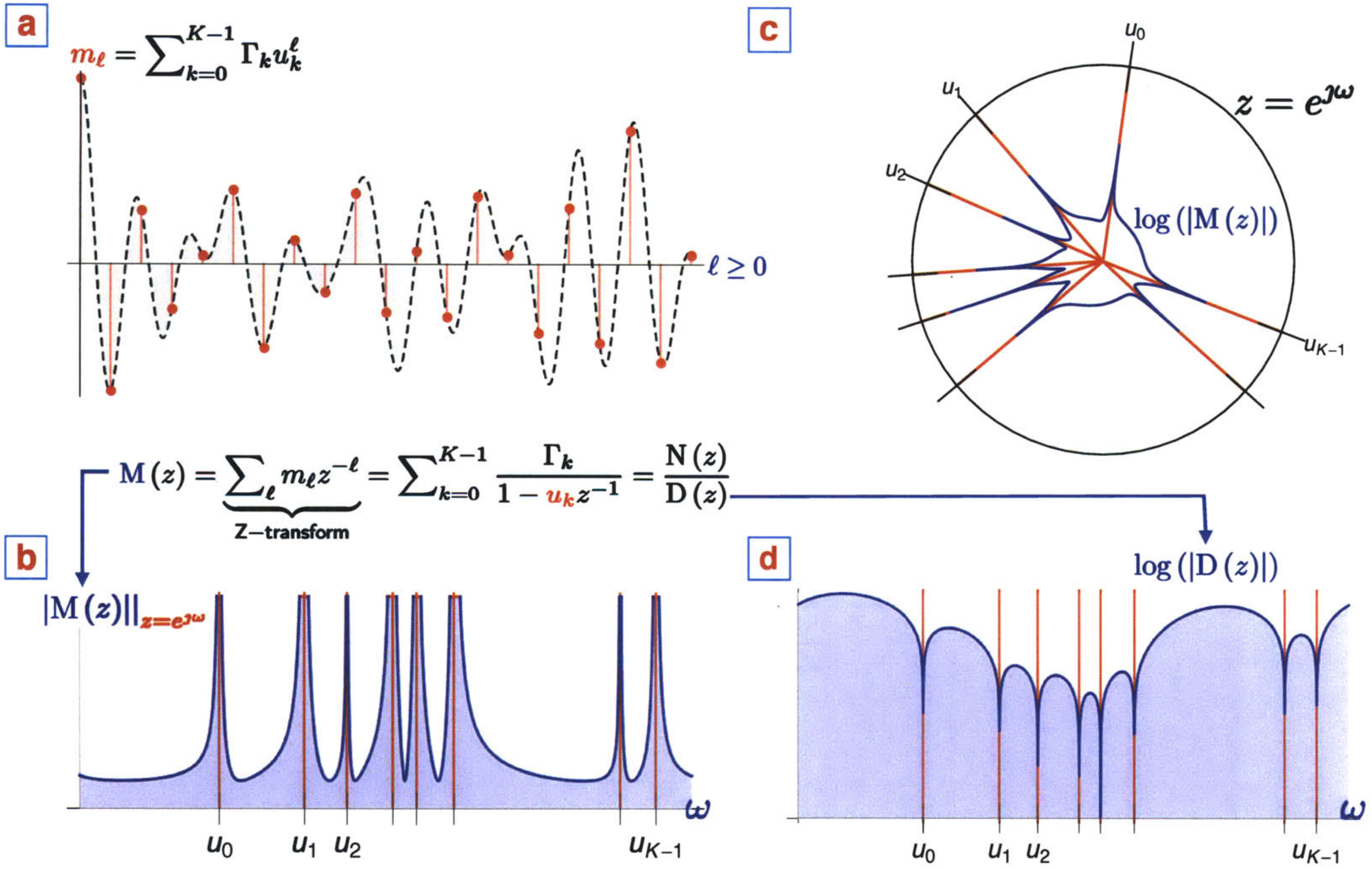


Figure 0-2: Mixed–pixel measurements for $K$–bounce case. The non–linear problem of estimating $\{\Gamma_k u_k\}_{k=0}^{K-1}$ from $m$ can be decoupled to first provide $u_k$. (a) The sequence $m_\ell = m(\ell\omega_0)$ with $m(\omega) = \sum_{k=0}^{K-1} \Gamma_k u_k^\omega$ is an additive composition of $K$–phasors, $\{\Gamma_k u_k\}_{k=0}^{K-1}$ where $u_k = e^{\jmath\phi_k}$. (b) The Z–transform of the discrete sequence $m_\ell$ is given by $\mathrm{M}(z) = \sum_{k=0}^{K-1} \frac{\Gamma_k}{1-u_k z^{-1}} = \frac{\mathrm{N}(z)}{\mathrm{D}(z)}$. The poles of M encode the locations $\{u_k\}_k$. (c) Since $m$ is a discrete sequence, $|\mathrm{M}(z)||_{z=e^{\jmath\omega}}$ is defined on a unit disk. $|\mathrm{M}(z)| \to \infty$ at polar angles $u_k = e^{\jmath\phi_k}$. (d) Alternatively, on factorizing $\mathrm{M} = \mathrm{N}/\mathrm{D}$, the roots of D encode the locations $\{u_k\}_k$.

# Acknowledgments

Doing what I do has been an enjoyable experience over last five years and five labs[8]. This particular work is an outcome of last eighteen months of research performed at Ramesh Raskar's lab at the Massachusetts Institute of Technology. This is also the longest amount of time I have spent in one lab.

I thank Ramesh for introducing me to the world of inverse problems that are linked with the trending area of Time–of–Flight (ToF) imaging. Ramesh's outlook and insight for problems, that have practical significance and industrial applications, have given an interesting direction to my otherwise theoretically inclined skill set. I have absorbed some of Ramesh's enthusiasm for transforming theory in to practice and this has provided me with a new class of problems to think about.

My parents[9] have provided me with tremendous support and freedom to explore and experiment with my career as a researcher. I am indebted to them for being perennially receptive and patient.

Before arriving in Cambridge, I have had the pleasure of working with various people who have deeply influenced the way I think today. I would like to thank my previous advisors, mentors and hosts: Thierry Blu, Michaël Unser, Martin Vetterli, Pina Marziliano, Ahmed Zayed and Frederic Bimbot/Rémi Gribonval. I hope to live by the standards they have set. Among others, Pedro Pinto and Christophe Vignat's encouragement has been instrumental in my move from EPFL to MIT.

Laurent Demanet and Thierry Blu accepted to be jury members for this work. I thank them for taking out time for going through this work. Laurent's encouragement, on–and–off advises and suggestions (in MIT hallways) have been useful. Skype meetings and exchange of emails/ideas (for last five years) with Thierry have been very beneficial and enriching. Thierry's painstaking proof-reading of the draft has made sure that all the equations belong here[10]. Laurent Daudet's friendly advises about science and philosophy are duly acknowledged!

---

[8] After my nomadic stops at INRIA–Rennes (France), NTU (Singapore), CUHK (Hong Kong) and BIG/LCAV EPFL (Switzerland).

[9] Together with my mentors.

[10] and that, Fourier and Prony *rest in peace*.

Refael Whyte's expertise with ToF cameras has played a pivotal role in making this work practicable. Achuta Kadambi, Christopher Barsi and Anshuman Das have contributed to many discussions and experiments which has led to fruitful collaborations. Shahram Izadi, Christoph Rhemann and Micha Feigin have been patient with experimental designs.

I thank all the members of the lab for their help. Taya Leary, Maggie Church and Keira Horowitz—thank you for being generous with your time and assistance. I am very thankful to Linda Peterson for preventing sky fall time and again (specially at the eleventh hour). Her infinite patience and support has been very helpful.

Adrien Clidière, Simon Arberet and Magnus Andermo have been a source of support all these years. Thanks for all those discussions. Nicloas/Lucie, Sunny, Aurelién,Jean–Charles, Lada and Taya—I have always enjoyed your coffee company in Cambridge. Uncle Raj and his family have been home for me right from the day I landed in Boston. I look forward to spending more time with you!

I owe a *thank you* to many more people who can not all be listed here. Inspired by MGK, I have tabulated a "Table of Thanks". I apologise if I forgot anyone.

After advising a number of friends to not to cite their adjoints, dual–spaces, partners, girlfriends/boyfriends and all that, in any form of academic acknowledgement, I'll make an exception for my own cause—Bhavika, you are the best! You'll have to be patient with me for a few more years!

# Acknowledgements

| | | Fom Switzerland (EPFL Years) | From France (INRIA Years) | From Cambridge (MIT Years) | Coffee/Restaurants | Mentors–Advisors–Lab Hosts | Scientific Discussions | MIT Related | 4 AM Friends | Family | Teachers | Help |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Simon | Arberet | 👍 | 👍 | | | | 👍 | | 👍 | | 👍 | 👍 |
| Magnus | Andermo | 👍 | | | | | | | | 👍 | | 👍 |
| Aydin | Arpa | | | 👍 | | | | 👍 | | | | |
| Jean–C. | Baritaux | 👍 | | 👍 | 👍 | | 👍 | | | | 👍 | |
| Christopher | Barsi | | | 👍 | 👍 | | 👍 | 👍 | | | | |
| Hisham | Bedri | | | | | | | 👍 | | | | 👍 |
| Frederic | Bimbot | | 👍 | | | 👍 | 👍 | | | | | |
| Thierry | Blu | | | | 👍 | 👍 | 👍 | | | | | |
| Nicolas | Bonneel | | | 👍 | 👍 | | | | | | | |
| Aurélien | Bourquard | 👍 | | 👍 | 👍 | | 👍 | 👍 | 👍 | | | 👍 |
| Kunal | Chaudhury | 👍 | | | | | | | | | 👍 | |
| Maggie | Church | | | | | | | 👍 | | | | 👍 |
| Adrien | Clidière | | 👍 | | | | | | 👍 | | 👍 | 👍 |
| Anshuman | Das | | | 👍 | 👍 | | 👍 | 👍 | | | | |
| Laurent | Daudet | | 👍 | 👍 | | | | | | | 👍 | 👍 |
| Laurent | Demanet | | | | | | 👍 | 👍 | | | 👍 | 👍 |
| Xiaowen | Dong | 👍 | | 👍 | | | | | | | | |
| Julio | Estrada | | | | | | | 👍 | | | | |
| Davide | Frey | | 👍 | | 👍 | | | | | | | |
| Micha | Feigin | | | | | | 👍 | 👍 | | | | |
| Cristina | Gabriela | 👍 | | 👍 | 👍 | | | | | | | |
| Remi | Gribonval | | 👍 | | | 👍 | 👍 | | | | 👍 | |
| Paul | Hand | | | | | | | 👍 | | | 👍 | |
| Thomas | Heinis | 👍 | | 👍 | 👍 | | | | 👍 | | | |
| Barmak | Heshmat | | | | | | | 👍 | | | | |
| Keira | Horowitz | | | | | | | 👍 | | | | 👍 |
| Hayato | Ikoma | | | | | | 👍 | 👍 | | | | |
| Shahram | Izadi | | | | | | 👍 | 👍 | | | | 👍 |
| Sunny | Jolly | | | 👍 | 👍 | | | 👍 | | | | |
| Raj | Joshi | | | 👍 | 👍 | | | | | 👍 | | 👍 |
| Achuta | Kadambi | | | | | | 👍 | 👍 | | | | 👍 |
| Taya | Leary | | | 👍 | 👍 | | | | | | | 👍 |
| Michael | Leigsnering | | | | 👍 | | 👍 | | 👍 | 👍 | | 👍 |
| Maria | Lindqvist | 👍 | | 👍 | 👍 | | | | | | | |
| Kshitij | Marwah | | | | | | | 👍 | | | | 👍 |
| Pina | Marziliano | 👍 | | | 👍 | 👍 | 👍 | | | | 👍 | 👍 |
| Belen | Masia | | | 👍 | 👍 | | | 👍 | | | | |

## Acknowledgements

| | | Fom Switzerland (EPFL Years) | From France (INRIA Years) | From Cambridge (MIT Years) | Coffee/Restaurants | Mentors–Advisors–Lab Hosts | Scientific Discussions | MIT Related | 4 AM Friends | Family | Teachers | Help |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Shay | Maymon | 👍 | | 👍 | 👍 | | 👍 | 👍 | | | | |
| Anthea | Monod | 👍 | | 👍 | 👍 | | | | | | | 👍 |
| Annette | Mossel | | | 👍 | 👍 | | | 👍 | | | | |
| Youssef | Mroueh | | | 👍 | 👍 | | 👍 | 👍 | | | | |
| Nikhil | Naik | | | 👍 | | | | 👍 | | | | 👍 |
| Christine | Neuberg | 👍 | | | 👍 | | | | | | | |
| Daan | Noordermeer | 👍 | | 👍 | 👍 | | | | | | | |
| Lubos | Omelina | | | 👍 | 👍 | | | 👍 | 👍 | | | |
| Linda | Peterson | | | | | | | 👍 | | | | 👍 |
| Francisco | Pinto | 👍 | | | 👍 | | | | | | 👍 | 👍 |
| Pedro | Pinto | 👍 | | 👍 | 👍 | | | | | | 👍 | 👍 |
| Adrian | Popescu | 👍 | | | 👍 | | | | | | | |
| Vasant | R. | | | 👍 | 👍 | | | 👍 | | | | |
| Ramesh | Raskar | | | | | 👍 | 👍 | 👍 | | | 👍 | |
| Dan | Raviv | | | | | | 👍 | 👍 | | | | |
| Christoph | Rhemann | | | | | | 👍 | | | | | |
| Heverson | Ribeiro | | 👍 | | 👍 | | | | | | | |
| Karin | Roesch | | | | | | | 👍 | | | | |
| Mune | Sato | | | | | | | 👍 | | | | |
| Guy | Satat | | | 👍 | 👍 | | 👍 | 👍 | | | | |
| Nick | Savidis | | | 👍 | 👍 | | | 👍 | | | | |
| Jon | Seaton | | | 👍 | 👍 | | | 👍 | | | | |
| Sagar | Sen | | 👍 | 👍 | 👍 | | | | | | | |
| Orit | Shamir | 👍 | | 👍 | 👍 | | | 👍 | | | | 👍 |
| Boxin | Shi | | | | | | | 👍 | | | | |
| Ashish | Sinha | | | | 👍 | | | | 👍 | | | 👍 |
| Lada | Sycheva | 👍 | | 👍 | 👍 | | | | | | | 👍 |
| Immanuel | Trummer | 👍 | | | 👍 | | | | | | | |
| Jay | Unnikrishnan | 👍 | | | 👍 | | | | | | | |
| Michael | Unser | 👍 | | | 👍 | 👍 | 👍 | | | | 👍 | 👍 |
| Martin | Vetterli | 👍 | | | | 👍 | 👍 | | | | | |
| Cristoph | Vignat | 👍 | | 👍 | 👍 | 👍 | 👍 | | 👍 | | 👍 | 👍 |
| Aaron | Wetzler | | | 👍 | 👍 | | | 👍 | | | | |
| Gordon | Wetzstein | | | 👍 | | | 👍 | 👍 | | | | |
| Refael | Whyte | | | 👍 | 👍 | | 👍 | 👍 | | | | |
| Ahmed | Zayed | | | | | 👍 | 👍 | | | | | |

# Contents

**Mathematical Introduction to Time–of–Flight Imaging** **27**

**1 A Unifying Framework for Solving Inverse Problems in Time–of–Flight Imaging** **28**

1.1 Generalized Image Formation Model for ToF Sensors . . . . . . . . . . . . 29
1.2 Shift–Invariant Kernels . . . . . . . . . . . . . . . . . . . . . . . . . . 30
  1.2.1 Fourier Domain Description . . . . . . . . . . . . . . . . . . . . 31
  1.2.2 Connection with Sampling Theory for Shift–Invariant Spaces . . . 31
1.3 Specification of Inverse Problems . . . . . . . . . . . . . . . . . . . . 32
  1.3.1 Single Depth Imaging . . . . . . . . . . . . . . . . . . . . . . . 33
  1.3.2 Multiple Depth Imaging . . . . . . . . . . . . . . . . . . . . . . 34
  1.3.3 Fluorescence Lifetime Estimation . . . . . . . . . . . . . . . . . 35
  1.3.4 Transient Imaging . . . . . . . . . . . . . . . . . . . . . . . . . 36
1.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

**Inverse Problems in Time–of–Flight Imaging** **37**

**2 Time–of–Flight Imaging** **38**

2.1 The Principle of Time–of–Flight . . . . . . . . . . . . . . . . . . . . . 38
  2.1.1 A Brief History of Optical ToF Principle . . . . . . . . . . . . . . 39

2.1.2 ToF Sensing Meets 3D Imaging . . . . . . . . . . . . . . . . . . . 40
2.2 Depth Sensing with Time–of–Flight Principle . . . . . . . . . . . . . . . . 43
2.2.1 Amplitude modulated continuous-wave (AMCW) ToF Imaging . . 43
2.2.2 Impulse Based ToF Imaging . . . . . . . . . . . . . . . . . . . . . 45
2.3 Industrial and Consumer ToF Cameras . . . . . . . . . . . . . . . . . . . 46

**3 Resolving Multi-path Interference in Time-of-Flight Imaging 49**
3.1 Overview of Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
3.2 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.3 Problem Formulation . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
3.4 Experimental Verification . . . . . . . . . . . . . . . . . . . . . . . . . 56
3.5 Concluding Remarks . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

**4 Sparse Linear Operator Identification for Time–of–Flight Imaging 60**
4.1 Overview of Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
4.2 Introduction and Motivation . . . . . . . . . . . . . . . . . . . . . . . . 61
4.2.1 From Sampling Functions to Sensing Operators . . . . . . . . . . . 61
4.2.2 Motivation: Mixed–Pixel Problem in ToF Imaging . . . . . . . . . 62
4.2.3 Contributions and Organization of this chapter . . . . . . . . . . . 64
4.3 The Sparse Operator Identification Problem . . . . . . . . . . . . . . . . 64
4.3.1 Link with the Mixed–Pixel Problem in ToF Imaging . . . . . . . . 65
4.3.2 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
4.3.3 Sparse Linear Operator Identification without Sparse Regularization: Forward Model and Solution to Inverse Problem . . . . . . . 67
4.3.4 Note on Enforcing Sparsity and Computational Complexity . . . . 70
4.4 From Theory to Practice . . . . . . . . . . . . . . . . . . . . . . . . . . 70
4.4.1 Design of Practical Experiments with ToF camera . . . . . . . . . . 70
4.4.2 Cramér–Rao Lower Bounds (CRB): Overview of Results . . . . . . 71
4.5 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

**5 Multi–bounce Decomposition of Light with Microsoft Kinect Xbox One 79**

5.1 Problem Statement . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79

5.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80

5.2.1 Multipath Interference in ToF Cameras . . . . . . . . . . . . . . . . 81

5.2.2 Contribution and Organization of this Chapter . . . . . . . . . . . 81

5.3 ToF Image Formation Model . . . . . . . . . . . . . . . . . . . . . . . 82

5.3.1 Modelling Single Depth . . . . . . . . . . . . . . . . . . . . . . . . 82

5.3.2 Multipath Interference Problem . . . . . . . . . . . . . . . . . . . . 84

5.3.3 Closed Form Solution for Retrieving $\{\Gamma_k, d_k\}_{k=0}^{K-1}$ . . . . . . . . . . 85

5.3.4 Remarks and Discussion . . . . . . . . . . . . . . . . . . . . . . . . 86

5.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87

5.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

**6 Fluorescence Lifetime Estimation via Coded Time–of–Flight Cameras 90**

6.1 Overview of Problem Statement and Contributions . . . . . . . . . . . . . 90

6.2 Inverse Problem Framework for FLI–Time-of-Flight Imaging . . . . . . . . 91

6.3 Forward Model for FLI using ToF Sensors . . . . . . . . . . . . . . . . . 92

6.4 Coded Illumination for FLI . . . . . . . . . . . . . . . . . . . . . . . . 95

6.5 Experimental Verification . . . . . . . . . . . . . . . . . . . . . . . . . 96

**7 Conclusion 98**

# List of Figures

0-1 Geometry of the mixed–pixel measurements $m(\omega) = \Gamma_1 e^{-\jmath\phi_1\omega} + \Gamma_2 e^{-\jmath\phi_2\omega}$. For the two bounce case, each depth $d_k$ resulting in the time delay $t_k$ parametrizes the angle of the phasor with $\phi_k = 2t_k$. The intensity of the pixel $\Gamma_k$ parametrizes the length of the phasor. The two phasors add to produce the measurements $m(\omega) = \Gamma e^{-\jmath\phi\omega}$. The inverse problem is to estimate $\{\Gamma_1, \Gamma_2\}$ and $\{t_1, t_2\}$ given $m$. . . . . . . . . . . . . . . . . . . . . . . . . 8

0-2 Mixed–pixel measurements for $K$–bounce case. The non–linear problem of estimating $\{\Gamma_k u_k\}_{k=0}^{K-1}$ from $m$ can be decoupled to first provide $u_k$. (a) The sequence $m_\ell = m(\ell\omega_0)$ with $m(\omega) = \sum_{k=0}^{K-1} \Gamma_k u_k^\omega$ is an additive composition of $K$–phasors, $\{\Gamma_k u_k\}_{k=0}^{K-1}$ where $u_k = e^{\jmath\phi_k}$. (b) The Z–transform of the discrete sequence $m_\ell$ is given by $\mathrm{M}(z) = \sum_{k=0}^{K-1} \frac{\Gamma_k}{1-u_k z^{-1}} = \frac{\mathrm{N}(z)}{\mathrm{D}(z)}$. The poles of M encode the locations $\{u_k\}_k$. (c) Since $m$ is a discrete sequence, $|\mathrm{M}(z)||_{z=e^{\jmath\omega}}$ is defined on a unit disk. $|\mathrm{M}(z)| \to \infty$ at polar angles $u_k = e^{\jmath\phi_k}$. (d) Alternatively, on factorizing $\mathrm{M} = \mathrm{N}/\mathrm{D}$, the roots of D encode the locations $\{u_k\}_k$. . . . . . . . . . . . . . . . . . . . . . . . . 12

1-1 Mixed pixel problem: Example of case where multiple bounce of light corrupt the image and hence distort the depth information. Information below the —line is corrupted with to multiple bounces of light. . . . . . . . . . . . 33

2-1 Early pioneers of optical time–of–flight principle. Image source: Wikimedia Commons/CC-PD. (a) Galileo Galilei, portrait by Giusto Sustermans (b) Ole Rømer, portrait by Jacob Coning (1700) (c) Isaac Newton, portrait by Godfrey Kneller (1689) (d) Hippolyte Fizeau. . . . . . . . . . . . . . . . . 39
2-2 Example of 3D image. (a) 3D image of a mannequin. (b)–(d) 3D image of a mannequin behind a transparent sheet with different viewpoints. . . . . . 41
2-3 Overview of 3D imaging techniques. . . . . . . . . . . . . . . . . . . . . . 42
2-4 Continuous wave and impulse based ToF imaging for single depth estimation. 43
2-5 Consumer and industrial grade ToF cameras. (a) PMD CamCube (photonics mixer device) (b) Mesa (also known as the Swiss Ranger) (c) Microsoft Kinect (d) PMD Nano. . . . . . . . . . . . . . . . . . . . . . . . . . 46
2-6 Comparison of various industrial and consumer ToF cameras. Comparison parameters (a) modulation frequency (b) wavelength (c) resolution (d) frame rate (e) measurement accuracy (f) maximum range. . . . . . . . . . . . . . 48

3-1 (a) ToF principle: the phase delay of an emitted AMCW wave proportionally encodes the distance of the reflecting object. (b) Mirror-like and (c) semi-transparent reflections produce MPI at a given camera pixel and yields an incorrect phase. (c) A complicated scene with severe MPI. . . . . . . . . . . 51
3-2 Left: experimental setup. Two transparencies block the left side of the camera (for a three-component measurement), and one transparency blocks the right (two-component measurement). Right: measured amplitude and depth at $\omega = 3\omega_0$. Dashed line indicates edge of second transparency. . . . 55
3-3 Reconstructed amplitudes and depths via sparse regularization. Dashed lines indicate edge of second transparency. . . . . . . . . . . . . . . . . . . . 57
3-4 Phase histogram for reconstructed and measured depth maps. Reconstructed phases cluster around the correct depths, whereas the measured depth map has a wide variance across the entire range. . . . . . . . . . . . . . . . . . 58

4-1 (a) The ToF camera emits a reference signal. Time delay of arrival from direct reflection encodes the depth $d_1$. (b) Demonstration of mixed–pixel problem (MPP): two or more paths contribute to the pixel. The inter–reflections between sheets are assumed to be negligible. . . . . . . . . . . . 65

4-2 (a) Performance analysis of methods used for solving $b_m = \sum_{k=0}^{K-1} \beta_k e^{-\jmath\omega_0 m t_k}$ (4.13) with different oversampling factors $N = 2\eta K + 1$. We compare Matrix Pencil Method [76] with Cadzow's algorithm [75] for the case of $K = 3$. The parameters are chosen from a uniform distribution. For each SNR value, we average the result over 5000 realizations. (b) Starting with M–sequence in (4.15) we obtain the calibrated function $\mathsf{C}_{p,\overline{p}}$ in —- and its $\Omega = 29$–term Fourier Series approximation, $\phi \in \mathsf{B}_\Omega$ in – –. The inset shows the Fourier Series coefficients (4.9). (c) Mixed–Pixel Problem for $K = 3$ mixing paths. The measurements $\mathbf{y}$ are marked in —-. Ground truth and estimates using OMP in (4.7) and our method are also marked. (d) We report the computation time involved with using LASSO, OMP and our non–iterative method. OMP requires $\sim$ 0.03 seconds per pixel while our method (using Matrix Pencils) requires $\sim$ 0.002 seconds. This result is consistent over a set of 961 pixels. LASSO is computationally intensive. For one pixel, the requirement is 689.40 seconds. . . . . . . . . . . . . . . . . 72

4-3 (a) Measurements (—) of a scene where a translucent/diffusing sheet $\{\beta_0, t_0\}$ hides a placard $\{\beta_1, t_1\}$. The mixed–pixels correspond to the case of $K = 2$. Using our method, the mixed–pixels in (a) are decomposed into the diffusing sheet in (b) and the placard that reads, "Time of Flight" in (c). (d) We show measurements $\mathbf{y}$ for one of the pixel of $120 \times 120$ image. These measurements are decomposed into shifted and amplitude scaled versions of $\mathsf{C}_{p,\overline{p}} = \phi$ marked in - - - - . Depths $\propto \{t_0, t_1\}$ are marked with $\rightarrow$. . . . . . 73

5-1 (a) The ToF camera emits a reference signal. Time delay of arrival from direct reflection encodes the depth *d*. (b) Specular or mirror like and, (c) semi–transparent reflections cause multiple light paths to mix at the sensor which leads to multipath interference or the mixed–pixel problem (MPP). (d) Case of continuous multipath reflections. (e) Time–of–flight imaging system pipeline for solving inverse problems. . . . . . . . . . . . . . . . . . . . . . . . 80

5-2 Images (a) and (b) show one of the input amplitude and phase images while (g) and (h) show the embedding of the intensity image onto the depth data. The scene, depicted in (j) is constructed from a mannequin on a table with a transparency between the camera and the scene. As can be seen, there is a strong flashback from the transparency as well as corruption of the depth image due to multi bounce. (c) and (d) show the intensity and depth reconstruction of the transparency while (e) and (f) show that of the scene. (i) Shows the embedding of the intensity image onto the depth reconstruction. Not only have the intensity and depth information been recovered correctly for both the scene and transparency, we have also recovered both amplitude and depth information that was completely occluded by the flashback (highlighted). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88

6-1 (a) Setup for fluorescence lifetime estimation using ToF cameras. Pixels at $(x_b, y_b)$ measure depth by using time–of–flight principle in accordance with (6.2). Pixels $(x_f, y_f)$ correspond to fluorescent sample. Impinging light signal $p$ undergoes two transformations: first, amplitude scaling and delay $((\mathcal{T}_d \circ \rho\mathcal{I})\,[p])$ due to the depth $d$ and second, due to its interaction with the fluorophore sample which is in effect $y = \mathcal{T}_d \circ \mu\mathcal{L}_{\tau}^{-1}\,[p]$. (b) Mathematical notation for the physical set up described in part (a). The system response in context of (6.1) is given by $\mathcal{H} = \mathcal{T}_d \circ (\mu\mathcal{L}_{\tau}^{-1} + \rho\mathcal{I})$ and the measurements read, $m(t) = \mathcal{H}[\phi], \phi = \mathsf{C}_{p,\bar{p}}$. (c) Time–domain FLIM is modeled by $m(t) = \mathcal{H}[\phi]$ with $\phi = p = \delta$ and $d = 0$. (d) Frequency–domain FLIM is modeled by $m(t) = \mathcal{H}[\phi]$ with $p = \delta$ and some calibrated $d$. (e) ToF–FLIM uses $m(t) = \mathcal{H}[\phi]$ where $\phi$ is some band–limited function with utmost $M_0$ Fourier components. We show $\mu\mathcal{L}_{\tau}^{-1} \to \mathcal{T}_d \circ \mu\mathcal{L}_{\tau}^{-1} \to \mathcal{H}$—the sequence of transformations on $\phi$. . . . . . . . . . . . . . . . . . . . . . . . . . . . 93

6-2 Overview of forward model for coded FLIM imaging. . . . . . . . . . . . . 94

6-3 (a) Measurements for a quantum dot with $\tau = 32$ ns acquired by a ToF camera that uses a time code $\phi$ (in the inset). (b) Fourier Transform of $\phi$ and its band-limited approximation via Fourier Series consisting of $M_0 = 51$ components.(c) We estimate lifetime and depth from measurements using $\angle\widehat{\mathbf{h}} = \angle\mathbf{D}_{\widehat{\phi}}^{-1}\mathbf{V}^{+}\mathbf{m} \approx -(\varphi_{\tau}(m\omega_0) + \varphi_d(m\omega_0))$. Curves in —— show measurements and the expected phase function is marked with - - -. . . . . 96

6-4 Histogram of estimated lifetimes. . . . . . . . . . . . . . . . . . . . . . . . 97

# List of Tables

2.1 Comparison Between Impulse and AMCW Based ToF . . . . . . . . . . . . 45

*In loving memory of my grandfather.*

# Mathematical Introduction to Time–of–Flight Imaging

# 1

# A Unifying Framework for Solving Inverse Problems in Time–of–Flight Imaging

The main goal of this chapter is to introduce the reader to the most general form of image formation model for Time–of–Flight (ToF) sensors. What do we mean by "the most general form of image formation model" and what are the advantages of this unifying perspective will be clear once we outline how the problem is dealt with in literature. While consumer ToF sensors are studied in detail in the preceding chapters, here, we will only consider the abstract, mathematical model for such sensors.

## 1.1 Generalized Image Formation Model for ToF Sensors

Almost all ToF sensors are active devices. The ToF assembly has an emitter (LASER or LED source) which emits a signal that probes the scene. Let us call it the *probing function* . The probing function can be a time–localised pulse or a continuous wave (such as a sinusoid). The probing function $p$ interacts with the scene prescribed by a response kernel $h$. This interaction results in the reflected signal $r$. We model this process with a Fredholm integral operator,

$$p \rightarrow \boxed{h} \rightarrow r = \int_{\Omega} h\left(t, z\right) p\left(z\right) dz \tag{1.1}$$

where $h$ is the continuously defined kernel and $\Omega$ is some measurable set which defines the domain of integration. We are mainly interested in $L_2$–kernels that are measurable on $\Omega \times \Omega$, that is to say,

$$\int_{\Omega}\int_{\Omega} \left|h\left(t, z\right)\right|^2 dt dz < +\infty \qquad (L_2\text{–kernel}).$$

An equivalent operator formulation that is representative of this setup takes form of,

$$r = \mathcal{H}\left[p\right].$$

For example, the operator that models our fluorescence lifetime estimation problem is given in (6.3) and the corresponding kernel is given by (6.4),

$$\begin{aligned} h\left(t, z\right) &= h_0\left(t - z\right) \text{ where,} \\ h_0\left(t\right) &= \delta\left(t - \frac{2d}{c}\right) * \left(\rho\delta\left(t\right) + \mu e^{-(t)/\tau}\mathbf{1}_{t \geqslant 0}\left(t\right)\right) \end{aligned} \tag{1.2}$$

The reflected signal $r$ is then observed by the camera with response kernel $\psi$ and measurements at the sensor read,

$$\begin{aligned} m &= \langle r, \psi^* \rangle_z \\ &= \underbrace{\int_{\Omega_\psi} \psi(t,z)\, r(z)\, dz}_{\mathcal{G}[r]} \Leftrightarrow \mathcal{G}[r] \end{aligned} \tag{1.3}$$

which is another Fredholm integral with kernel $\psi$. The kernel $\psi$ can be used to model the point spread function (PSF) of the camera or could also be a waveform which facilitates demodulation. Concrete examples that follow later will make the context clear.

Finally, like all digital instruments, the camera samples the measurement to yield discrete measurements,

$$m[n] = m(t)|_{t=nT}, \quad n \in \mathbb{Z}, \quad T > 0.$$

The imaging pipeline then associates itself with the following chain that maps the probing function $p$ to measurements $m$,

$$p \to \boxed{h} \to \underbrace{r = \int_{\Omega} h(t,z)\, p(z)\, dz}_{\text{Scene response kernel}} \to \boxed{\psi} \to \underbrace{\int_{\Omega_\psi} \psi(t,z)\, r(z)\, dz}_{\text{Camera response kernel}} \to m.$$

## 1.2 Shift–Invariant Kernels

An important class of Fredholm kernels $h$ and $\psi$ that model the scene response and camera response, respectively, is the class of shift–invariant kernels which assume form of,

$$k_{\mathsf{SI}}(t,z) = k(t-z). \tag{1.4}$$

For the class of shift–invariant kernels, the interaction,

$$p \to \boxed{h_{\mathsf{SI}}} \to r = \int_{\Omega} h(t-z)\, p(z)\, dz \Leftrightarrow \underbrace{(h * p)(t)}_{\text{Convolution}}$$

boils down to filtering or convolution in context of Linear Time Invariant systems. As a result, the image formation model assumes a simple structure that takes form of,

$$m\left[n\right] = \left(\psi * h * p\right)\left(t\right)\big|_{t=nT}.$$

For simplicity of exposition, we drop the subscript SI and assume kernels to be shift–invariant unless pointed out explicitly.

### 1.2.1 Fourier Domain Description

Let us define the Fourier transform of a function by,

$$\widehat{f}\left(\omega\right) = \left\langle f\left(t\right), e^{\jmath\omega t}\right\rangle_t.$$

Thanks to the shift–invariant formulation, due to convolution–multiplication property, the measurements have a simple Fourier domain description,

$$\widehat{m}\left(\omega\right) = \left(\widehat{\psi}\widehat{h}\widehat{p}\right)\left(\omega\right).$$

In context of Coded–ToF imaging we will consider probing functions that are assumed to be sufficiently smooth and time–localized. For such functions, we characterize the spectrum using,

$$\left|\widehat{p}\left(\omega\right)\right| \leqslant \frac{C}{1+\left|\omega\right|^{m+1+\varepsilon}}, \quad C>0, \varepsilon>0, m\in\mathbb{Z}_{+} \tag{1.5}$$

that is to say, the function $p$ is bounded and $m$–times differentiable.

### 1.2.2 Connection with Sampling Theory for Shift–Invariant Spaces

In view of the sampling theory for shift–invariant subspaces [1, 2], the discrete measurements of the ToF imaging pipeline can be modelled as,

$$m\left(nT\right) = \left\langle p\left(t\right), \overline{\left(\psi * h\right)}\left(t-nT\right)\right\rangle, \quad p,\psi\in\mathbb{R}, \tag{1.6}$$

where we use the time–reverse operator $\overline{f}(\cdot) = f(-\cdot)$ and $\overline{(\psi * h)}$ is some pre–filter which projects the probing function onto subspace of functions,

$$\left\{\overline{(\psi * h)}(t - nT)\right\}_{n\in\mathbb{Z}}.$$

This is because,

$$\begin{aligned}
\left\langle p(t), \overline{(\psi * h)}(t - nT)\right\rangle &= \int p(t)\, \overline{(\psi * h)}(t - nT)\, dt \\
&= \int p(t)\, (\psi * h)(\tau - t)\, dt \\
&= (p * \psi * h)(\tau)|_{\tau = nT} \\
&= m(nT).
\end{aligned}$$

In what follows, we give a brief over view of inverse problems and how the image formation model developed so far can be used to solve problems of practical interest.

## 1.3 Specification of Inverse Problems

ToF imaging can be accomplished via impulse modulation where $p = \delta$ or amplitude modulation (AMCW–ToF) where $p$ is some continuous function of time. We identify four examples of inverse problems. In the remainder of this work, we will focus on the details of these problems.

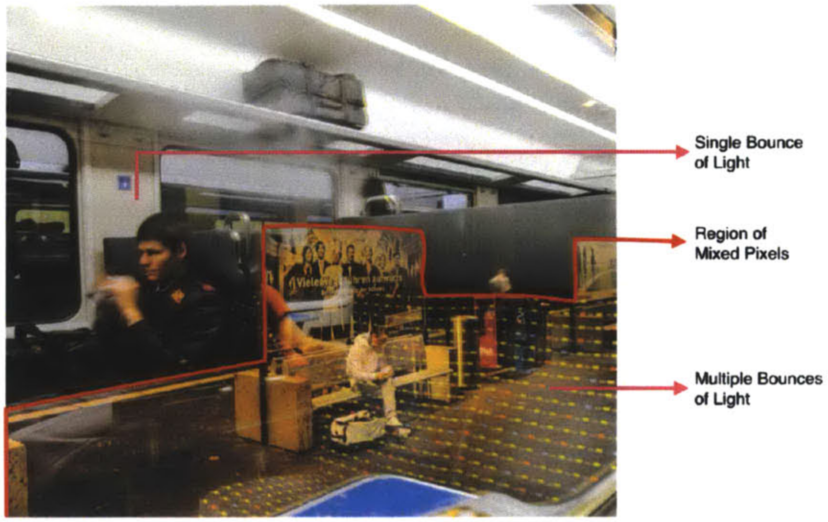


Figure 1-1: Mixed pixel problem: Example of case where multiple bounce of light corrupt the image and hence distort the depth information. Information below the —line is corrupted with to multiple bounces of light.

### 1.3.1 Single Depth Imaging

This is the default mode of operation of AMCW–ToF cameras. The camera modulates a cosine wave at some modulation frequency. The reflected signal then arrives at the sensor which demodulates the signal using the lock–in principle. The details of this configuration are below. This problem is related with the problem of "Tone Estimation" [3, 4].

► Example I: Depth Imaging via ToF Principles
AMCW–ToF

---

- Probing Function: $p(t) = \cos(\xi_0 t)$. — $\xi_0$: modulation frequency.
- Scene Response: $h_{\mathsf{SI}}(t, z) = \rho\delta\left(t - z - \frac{2d}{c}\right)$
- Reflected Signal: $r(t) = \rho\cos(\xi_0 t - \phi), \phi \triangleq \frac{2\xi_0 d}{c}$ — $\phi \in [0, 2\pi]$
- Measurements: $m(t) = \frac{\rho}{2}\cos(\xi_0 t + \phi)$. — $\psi_{\mathsf{SI}}(t, z) = \overline{p}(t - z)$

Inverse Problem: Estimate $\{d, \rho\}$.

### 1.3.2 Multiple Depth Imaging

When light undergoes multiple reflections, the depth data gets corrupted. This leads to an inverse problem linked with recovery of information linked with the bounces of light. The problem is illustrated in Fig. 1-1. Ideally, we hope to demix bounces of light so that we can separate the reflections from the direct bounce of light (linked with the shortest path). Exemplary solutions are presented in sections that follow. The experimental setup for AMCW–ToF imaging is shown in Fig. 3-2 and the demixed light paths that are used to recover the image are shown in Fig. 3-3. Similarly, 3D images linked with this problem are shown in Fig. 2-2 and Fig. 5-2.

Mathematical configuration [5, 6] of this problem shows that the associated inverse problem can be re–cast as a non–linear least squares problem. More specifically, a parameter estimation problem that is linked with spectral estimation [7]. The details of the probing function, the reflected signal and the camera measurements are below. The choice and characteristics of the probing function depend on the electronics and the scene response function models the physics of the problem.

▶ Example II: Multi–path Problem/Multiple Depth Estimation

Coded–ToF

---

- ■ Probing Function: $p\left(t\right) = \sum_{|m|\leqslant M_0} \widehat{p}_m \exp\left(\jmath\omega_0 m t\right)$ $\quad |\widehat{p}\left(\omega\right)| \leqslant \frac{C}{1+|\omega|^{m+1+\varepsilon}}$ (from (1.5))
- ■ Scene Response: $h_{\mathsf{SI}}^{(K)}\left(t,z\right) = \sum_{k=0}^{K-1} \rho_k \delta\left(t - z - \frac{2d_k}{c}\right)$
- ■ Reflected Signal: $r\left(t\right) = \sum_{k=0}^{K-1} \rho_k p\left(t - z - \frac{2d_k}{c}\right)$
- ■ Measurements: $m\left(t\right) = \sum_{k=0}^{K-1} \rho_k \left(p * \overline{p}\right)\left(t - z - \frac{2d_k}{c}\right)$ $\quad \psi_{\mathsf{SI}}\left(t,z\right) = \overline{p}\left(t-z\right)$

---

Inverse Problem: Estimate $\{d_k, \rho_k\}_{k=0}^{K-1}$.

---

### 1.3.3 Fluorescence Lifetime Estimation

Another interesting inverse problem that is linked with ToF sensors deals with fluorescence lifetime estimation [8]. When light interacts with a fluorescent object at some distance $d$ and with some lifetime $\tau$, the reflected light is composition of light bounces. This results in a non–linear, mixed–pixel problem. The specifics of the probing function and problem setup are described in **Example III** below.

▶ Example III: Fluorescence Lifetime Estimation
AMCW/Coded–ToF

---

- Probing Function: $p(t) = \sum_{|m|\leqslant M_0} \widehat{p}_m \exp(\jmath\omega_0 m t)$
- Scene Response: $h_{\mathsf{SI}}(t,z) = \delta\left(t - z - \frac{2d}{c}\right) * \left(\rho\delta(t-z) + \mu e^{-(t-z)/\tau}\mathbf{1}_{t-z\geqslant 0}(t-z)\right)$
- Reflected Signal: $r(t) = (h_{\mathsf{SI}} * p)(t)$
- Measurements: $m(t) = T^{-1}\sum_{|m|\leqslant M_0} |\widehat{p}_m|^2 \widehat{h}_m \exp(\jmath\omega_0 m t)$  $\psi_{\mathsf{SI}}(t,z) = \overline{p}(t-z)$.

  with $\widehat{h}_m = \widehat{h}(m\omega_0)$ and $\widehat{h}(\omega) = \left(\rho + \frac{1}{1/\tau + \jmath\omega}\right) e^{-\jmath\frac{2d}{c}\omega}$

---

Inverse Problem: Estimate $\{d, \tau\}$ (or the distance of fluorescent sample and its lifetime).

### 1.3.4 Transient Imaging

Next, we outline our final inverse problem: Transient Imaging. Given that the speed of light is finite $c = 3 \times 10^8$, can we image transient phenomenon? Indeed this is possible with ToF imaging. Only recently, progress has been made using both impulse–ToF [9, 10] and AMCW–ToF sensors [11]. While this work is concerned with AMCW–ToF imaging, our generic imaging pipeline very accurately models the transient imaging configuration.

▶ Example IV: Time–Resolved Imaging of Transient Phenomenon

Impulse Imaging/Ultrafast ToF

---

- Probing Function: $p(t) = \delta(t)$
- Scene Response: $h_{\mathsf{SI}}(t,z) = \underbrace{h_{\mathsf{D}}(t-z)}_{\text{Direct Reflection}} + \underbrace{h_{\mathsf{I}}(t-z)}_{\text{Inter–reflections}} + \underbrace{h_{\mathsf{S}}(t-z)}_{\text{Subsurface}}$

| | |
|---|---|
| —◇ Direct Reflection | $h_{\mathsf{D}}(t) = \alpha_{\mathsf{D}}\delta(t - 2d_{\mathsf{D}}/c)$ |
| —◇ Inter–reflections | $h_{\mathsf{I}}(t) = \sum_{k=0}^{K-1} \alpha_k \delta(t - 2d_k/c)$ |
| —◇ Subsurface Scattering | $h_{\mathsf{S}}(t) = \delta(t - 2d_{\mathsf{S}}/c) * \left(\alpha_{\mathsf{S}} e^{-\beta_{\mathsf{S}} t}\mathbf{1}_{t \geqslant 0}(t)\right)$ |

- Reflected Signal: $r(t) = h_{\mathsf{SI}}(t,0)$
- Measurements: $m(t) = \langle r(z), \psi_{\mathsf{SI}}(t,z)\rangle_z$ Camera PSF: $\psi_{\mathsf{SI}}(t,z) = C_0 \exp\left(-\frac{(t-z)^2}{2\sigma^2}\right)$

---

Inverse Problem: Estimate Transient Image Parameters $\left\{\alpha_{\mathsf{D}}, d_{\mathsf{D}}, \{\alpha_k, d_k\}_{k=0}^{K-1}, \alpha_{\mathsf{S}}, d_{\mathsf{S}}, \beta_{\mathsf{S}}\right\}$

---

## 1.4 Summary

Our ToF imaging pipeline unifies inverse problems linked with ToF imaging. Instead of dealing with inverse problems on case–by–cases basis, our pipeline encapsulates the essence of both impulse as well as AMCW imaging modalities. Even more so, whenever the scene response kernel has a shift–invariant structure, AMCW—ToF imaging amounts to computing Fourier–modes of the scene response function. This insight leads to interesting and efficient formulation of inverse problems where the scene response function is parametric in nature. For example, multi–depth estimation problem amounts to parameter estimation of $K$ complex exponentials where $K$ is the number of light bounces. In next chapters, we discuss these problems in details.

# Inverse Problems in Time–of–Flight Imaging

# 2

# Time–of–Flight Imaging

## 2.1 The Principle of Time–of–Flight

As the name suggests, in its most abstract form, the principle of Time–of–Flight deals with the following idea: Consider an emitter that emits a signal. This signal interacts with its physical environment and is received by the receiver. Having obtained time–resolved measurements at the receiver, can we say anything about the environment from the nature of measurements?

Nature is replete with evolved systems that rely on the Time–of–Flight or the ToF principle. For example, bats [12, 13, 14] and dolphins use the ToF principle for echolocation. This helps them navigate their environment and also serves the purpose of tracking and hunting. The ToF principle is also used by visually impaired and blind human beings to sense their environment [15]. This is known as acoustic wayfinding .

Figure 2-1: Early pioneers of optical time–of–flight principle. Image source: Wikimedia Commons/CC-PD. (a) Galileo Galilei, portrait by Giusto Sustermans (b) Ole Rømer, portrait by Jacob Coning (1700) (c) Isaac Newton, portrait by Godfrey Kneller (1689) (d) Hippolyte Fizeau.

Underlying both examples mentioned here, the common theme is that of using the ToF principle with a co–located emitter and receiver.

Within the ToF context, acoustic signals have also been used for measuring depth of wells by dropping stones and estimating the time–of–arrival of the echo. That said, humans have explored the ToF principle beyond the purpose of self–navigation via acoustic wayfinding. One of the main ideas here is to use ranging via ToF principle. While acoustic signals work well for large–scale scenes, optical signals are definitely the natural way out for high resolution sensing because of the wavelength of the electromagnetic waves.

This brings us to the topic of optical ranging—an idea that is at the heart of our discussion in this thesis.

### 2.1.1 A Brief History of Optical ToF Principle

Earliest work on the optical ToF principle dates back to the experiments of Galileo Galilei (cf. Fig. 2-1) who tried to measure the speed of light by ToF principle. Unfortunately, the choice of distance in his experiment did not give a conclusive result. To this end, the first break through came from the work of Danish astronomer Ole Rømer [16, 17] who was working at the Royal Observatory in Paris at the time. Rømer's experiment had over come the distance limitation—a bottleneck in Galileo's setup—as it was based on arguments that depended on distances at the scale of planetary separation. As is the case with most break throughs, the idea was rejected. However, the idea was supported by contemporary scientists such as Christiaan Huygens and Isaac Newton. Historically, French physicist, Hippolyte Fizeau [18] was the first to precisely measure the speed of light. To give an idea about the time scale of science, Fizeau's success was reported nearly two hundred years after Rømer's pioneering effort. At the turn of the century in early 1900's, more interesting observations were made regarding the physical nature of light and its constituent properties. One of the most celebrated results is due to Albert Einstein who was recognised for the discovery of the law of the photoelectric effect [19]. What followed this revolution was a series of contributions towards electronic image acquisition and sensing. Electron tubes were replaced by solid state devices. As of now, the optoelectronic landscape has benefitted extensively from the development of CMOS[1] technology. When compared with its predecessor—the Charged Couple Device (CCD)—CMOS is not only cheaper but also allows for on-chip digital functionality.

### 2.1.2 ToF Sensing Meets 3D Imaging

3D or three dimensional imaging [20] refers to any imaging technique that can acquire 3D information about the scene of interest. For example, in Fig. 2-2, we show some examples of 3D images acquired by a ToF camera. Unlike the conventional cameras that produce photographs, 3D imaging systems provide an intensity image together with a depth image. While the photograph provides the spatial information of an object/scene in x-y co-ordinates, the depth information provides information about the object along the z–dimension. The art of contact-less 3D imaging mainly relies on electromagnetic radiation based interaction of a probing signal with 3D object. Such measurements mainly fall into three categories [20].

- **Microwave Imaging** The wavelength of operation is in range of 3–30 mm. Equivalently, the frequency ranges from 10–100 MHz. This is well suited for large scale 3D

---

[1] Complementary Metal Oxide Semiconductor.

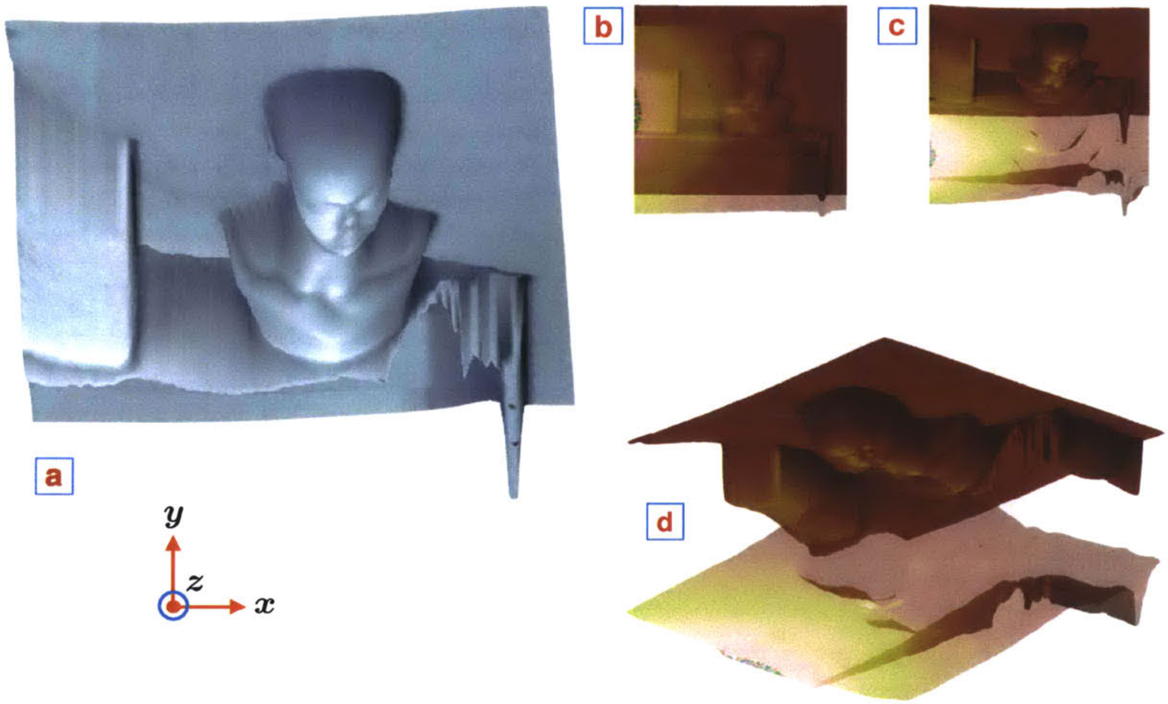


Figure 2-2: Example of 3D image. (a) 3D image of a mannequin. (b)–(d) 3D image of a mannequin behind a transparent sheet with different viewpoints.

shape inference.

- **Lightwave Imaging** The wavelength of operation is in range of 0.5–1 $\mu$m. Equivalently, the frequency ranges from 300–600 THz. The wavelength being shorter, this is well suited for depth measurements of objects where high lateral resolution is useful.
- **Ultrasonic Imaging** The wavelength of operation is in range of 0.1–1 mm. Equivalently, the frequency ranges from 0.3–3 MHz. Due to the nature of such waves, ultrasound based measurements (unlike lightwaves) tend to be sensitive to variations in temperature and pressure.

In this work, we will consider measurements based on electromagnetic radiation. In particular, we will discuss lightwave based imaging. Broadly speaking, optical 3D imaging techniques rely on three principles: **Triangulation, Time–of–Flight** and **Interferometry**. A taxonomical overview of the methods is presented in Fig. 2-3. We will briefly discuss the main principles before we start with our discussion on ToF imaging.

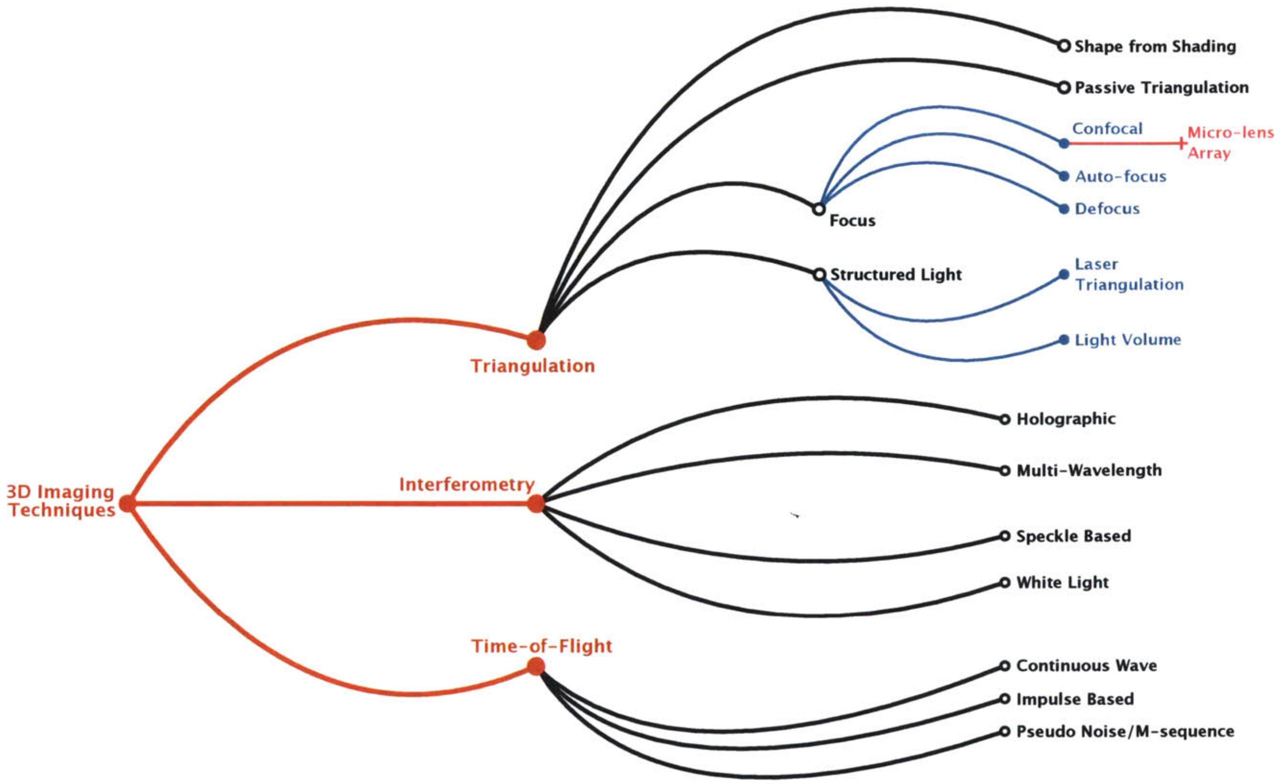


Figure 2-3: Overview of 3D imaging techniques.

**Triangulation** is essentially a trigonometric approach where by the distance to an unknown point is computed by measuring the respective angles to the point from the either side of a triangle [21].

**Time–of–Flight** methods as the name suggests, uses echoes of light to measure the distance. With speed of light fixed at $c = 3 \times 10^8$ m/s, the object distance is computed by using the relation,

$$t = \frac{2d}{c},$$

where $t$ is the Time–of–Flight and $2d$ is the round trip distance to the object. From an instrumental standpoint, the optical signal need not be coherent. In fact, for most known settings, incoherent light is used for this modality of 3D imaging. ToF imaging [22] is accomplished via continuous–wave signal modulation [23], impulse based methods and pseudo–random sequence based modulation. In coming chapters, we will discuss ToF imaging in details.

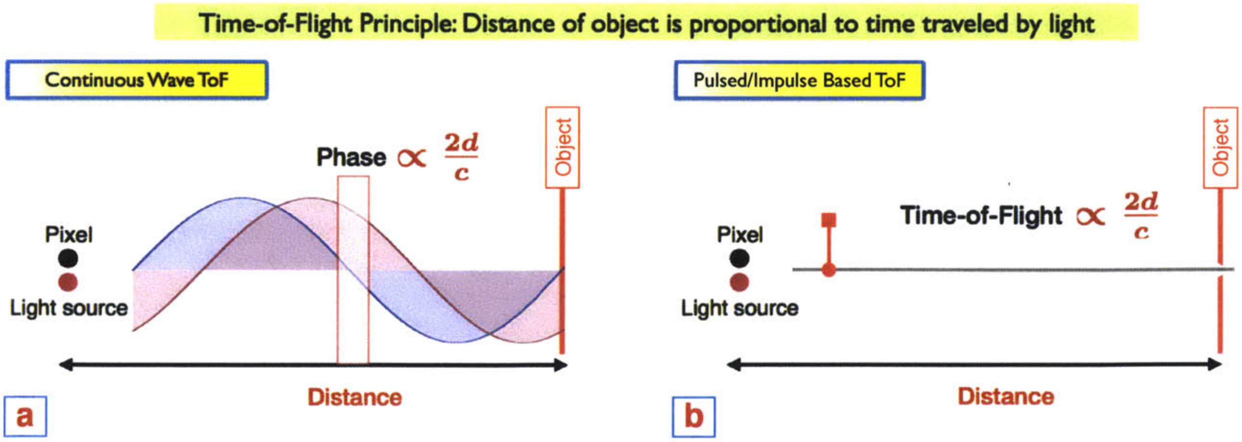


Figure 2-4: Continuous wave and impulse based ToF imaging for single depth estimation.

**Interferometry** uses ToF principle for depth imaging. The main difference being, the waveforms are required to be coherent.

## 2.2 Depth Sensing with Time–of–Flight Principle

As seen in the previous subsection, ToF imaging can broadly be classified as amplitude modulated continuous-wave (AMCW) ToF imaging and impulse based imaging. Before characterizing the inverse problems linked with the topic, we will present a light overview of the two methods. This will serve as a mathematical preliminary for the discussion.

### 2.2.1 Amplitude modulated continuous-wave (AMCW) ToF Imaging

We show the experimental setup in Fig. 2-4(a). The ToF pixel emits a sinusoidal waveform,

$$p(t) = \cos(\omega_0 t),$$

with modulation frequency $\omega_0$ (usually in MHz). Upon reflection from an object at depth $d$ meters from the pixel/camera, the emitted waveform $p$ undergoes a reflection and assumes form of,

$$\begin{aligned} r(t) &= \mathrm{DC} + \alpha \cos\left(\omega_0\left(t - \frac{2d}{c}\right)\right) \\ &= \mathrm{DC} + \alpha \cos\left(\omega_0 t - \phi\right), \end{aligned}$$

where DC denotes the constant bias component and,

$$\phi = \omega_0 \frac{2d}{c}$$

is the phase shift due to depth $d$. The ToF camera pixel is a lock–in sensor [24] that cross–correlates the reflected waveform $r$ with the reference waveform $p$ to compute measurements of form,

$$\begin{aligned} m(t) &= \int p(\tau)\, r(t+\tau)\, d\tau \\ &= \frac{\alpha}{2}\cos\left(\omega_0 t + \phi\right) + \mathrm{DC}. \end{aligned}$$

The TOF camera pixel samples the measurements at least four times per period to discretize measurements,

$$m[k] = m\left(\frac{k\pi}{2\omega_0}\right), \qquad k = 0, \ldots, 3.$$

Ignoring the constant bias, it now remains to compute intensity and phase $\alpha$ and $\phi$, respectively from the measurements,

$$\left\{\alpha_{\mathrm{opt}}, \phi_{\mathrm{opt}}\right\} = \arg\min_{\alpha,\phi} \sum_{k=0}^{k=3} \left(m[k] - \left(\frac{\alpha}{2}\cos\left(\omega_0 t + \phi\right)\right)\right)^2.$$

Thankfully, there is a computationally efficient way to estimate the parameters of the sinusoid [25]. Now since,

$$\begin{bmatrix} m[0] & m[1] \\ m[2] & m[3] \end{bmatrix} = \frac{1}{2}\begin{bmatrix} +\alpha\cos(\phi) & -\alpha\sin(\phi) \\ -\alpha\cos(\phi) & +\alpha\sin(\phi) \end{bmatrix},$$

Table 2.1: Comparison Between Impulse and AMCW Based ToF

| | AMCW Based | Impulse Based |
|---|---|---|
| Advantages | High signal-to-noise ratio.<br>Relaxed peak light power requirements.<br>Eye safe (diffused illumination). | Direct computation of depth. |
| Disadvantages | Distance ambiguity due to phase wrapping.<br>Indirect computation of depth (but real time). | High dynamic range.<br>Limited frames per second.<br>High bandwidth requirements |

it turns out that,

$$\alpha_{\mathsf{opt}} = \sqrt{\left(m\left[3\right] - m\left[1\right]\right)^2 + \left(m\left[0\right] - m\left[2\right]\right)^2}$$
$$\phi_{\mathsf{opt}} = \tan^{-1}\left(\frac{m\left[3\right] - m\left[1\right]}{m\left[0\right] - m\left[2\right]}\right).$$

As a result, the camera is able to compute the depth and intensity for each pixel which yields the 3D information about the scene.

The "Four Bucket Trick" is what makes the ToF sensors real time devices [23, 24, 25]. This method is used by almost all the AMCW–ToF cameras.

### 2.2.2 Impulse Based ToF Imaging

The experimental setup for impulse based ToF imaging is shown in Fig. 2-4(b). The ToF pixel emits a time localized pulse. At the same time, a high precision clock for each pixel is initiated which keeps track of the time it takes for the impulse to return back. As soon as the pixel reports backscattered signal, the clock is reset and hence, the camera is able to compute the time–of–flight.

To give the reader an idea about the scale of time precision it takes for this modality to work, it must be noted that for the light to travel one meter, it takes,

$$t = \frac{2}{c} = \frac{2}{3} \times 10^{-8} \approx 6.667\text{ns}.$$

Figure 2-5: Consumer and industrial grade ToF cameras. (a) PMD CamCube (photonics mixer device) (b) Mesa (also known as the Swiss Ranger) (c) Microsoft Kinect (d) PMD Nano.

While the ToF computation is relatively simpler for this approach, the computational load has to be compensated by the electronics in the hardware. Impulse based ToF systems demand high precision clock for accurate timing of the returning signal. Also, physical properties of the medium resulting in dispersion and attenuation smear the reflected signal which results in timing uncertainty.

We compare and contrast the key differences between the two forms of ToF imaging modalities in Table 2.1.

## 2.3 Industrial and Consumer ToF Cameras

As noted in the previous section, AMCW method has considerable advantages over the impulse based ToF imagers. This has led to the rapid development of a number of consumer and industrial grade ToF cameras. Almost all of them deploy the "4–Bucket Sampling Trick" [24]. In figure Fig. 2-5 we show some of the ToF cameras. These cameras offer competitive performances. The key deciding factors are,

- Maximum modulation frequency
- Wavelength of operation
- Resolution
- Frame rate
- Measurement accuracy
- Maximum range.

Considering the above mentioned metrics, a comparison of various camera models is presented in Fig. 2-6

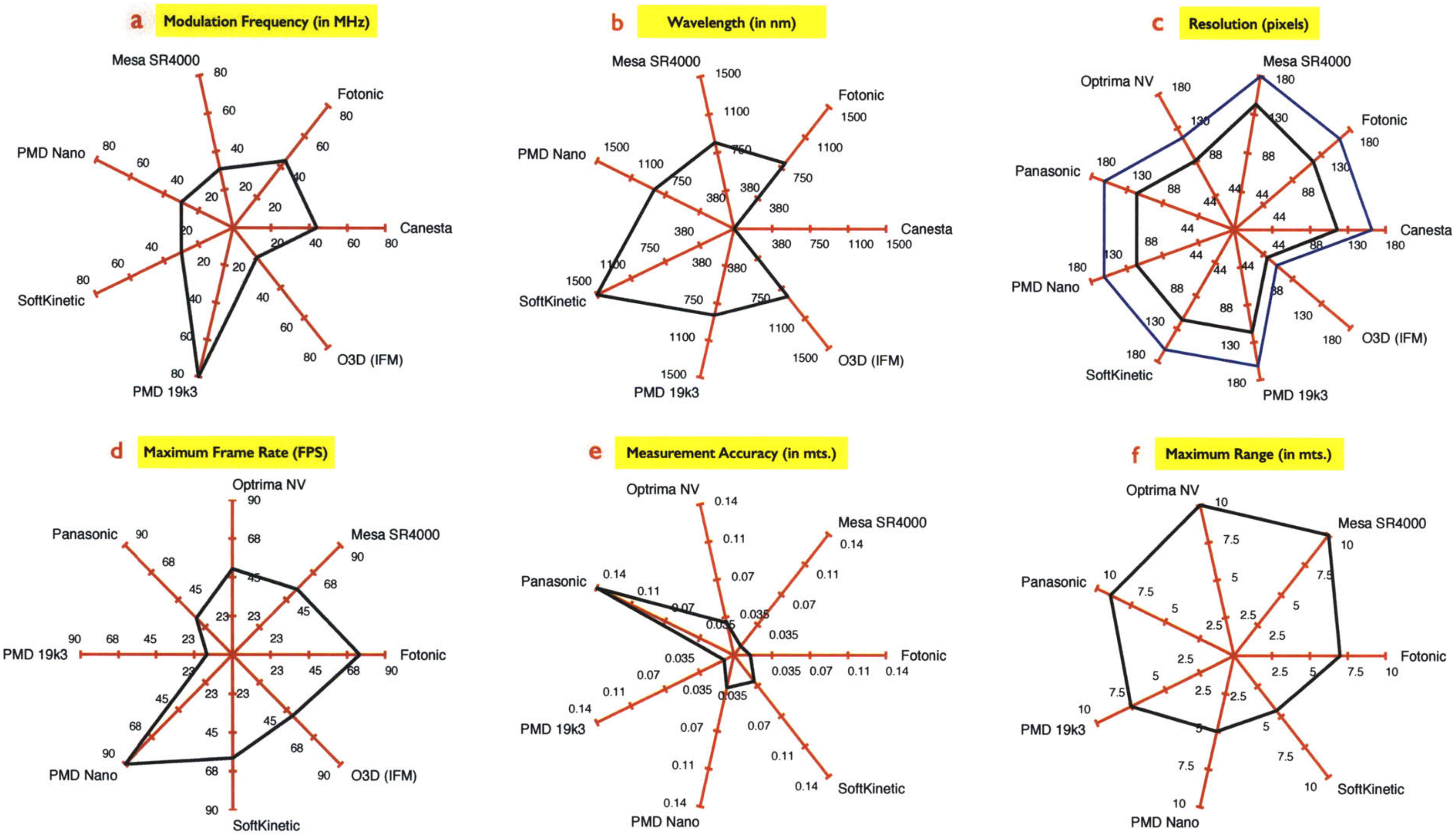


Figure 2-6: Comparison of various industrial and consumer ToF cameras. Comparison parameters (a) modulation frequency (b) wavelength (c) resolution (d) frame rate (e) measurement accuracy (f) maximum range.

# 3

# Resolving Multi-path Interference in Time-of-Flight Imaging

## 3.1 Overview of Results

Time-of-flight (ToF) cameras calculate depth maps by reconstructing phase shifts of amplitude-modulated signals. For broad illumination or transparent objects, reflections from multiple scene points can illuminate a given pixel, giving rise to an erroneous depth map. We report here a sparsity regularized solution that separates $K$ interfering components using multiple modulation frequency measurements. The method maps ToF imaging to the general framework of spectral estimation theory and has applications in improving depth profiles and exploiting multiple scattering.

## 3.2 Motivation

Optical ranging and surface profiling have widespread applications in image-guided surgery [26], gesture recognition [27], remote sensing [28], shape measurement [29], and novel phase imaging [30]. Generally, the characteristic wavelength of the probe determines the resolution of the image, making time-of-flight (ToF) methods suitable for macroscopic scenes[24, 31, 22]. Although ToF sensors can be implemented with impulsive sources, commercial ToF cameras rely on the continuous wave approach: the source intensity is modulated at radio frequencies (~10s of MHz), and the sensor reconstructs the phase shift between the reflected and emitted signals. Distance is calculated by scaling the phase by the modulation frequency (Fig. 3-1 (a)). This method, amplitude modulated continuous wave (AMCW) ToF, offers high SNR in real time.

However, AMCW ToF suffers from multipath interference (MPI) [32, 33, 34, 35, 36, 37, 38, 39, 40]. Consider, for example, the scenes in Figs. 3-1 (b,c). Light rays from multiple reflectors scatter to the observation point. Each path acquires a different phase shift, and the measurement consists of the sum of these components. The recovered phase, therefore, will be incorrect. Such "mixed" pixels contain depth errors and arise whenever global lighting effects exist. In some cases (Fig. 3-1 (d)), the measurement comprises a continuum of scattering paths. This can be improved with structured light or mechanical scanning [41, 42], but these are limited by the source resolution. Computational optimization [43, 44] schemes rely on radiometric assumptions and have limited applicability.

Here, we resolve MPI via sparse regularization of multiple modulation frequency measurements. The formulation allows us to recast this problem into the general framework of spectral estimation theory [45]. This contribution generalizes the two-component, dual-frequency approach [34, 38, 39], beyond which the two-component optimization methods fail. Thus, our method here has two significant benefits. First, we separate MPI from direct illumination to produce improved depth maps. Second, we resolve MPI into its components, so that we can characterize and exploit multiple scattering phenomena. The procedure has two steps: (1) record a scene with multiple modulation frequencies and (2) reconstruct the MPI components using a sparsity constraint.

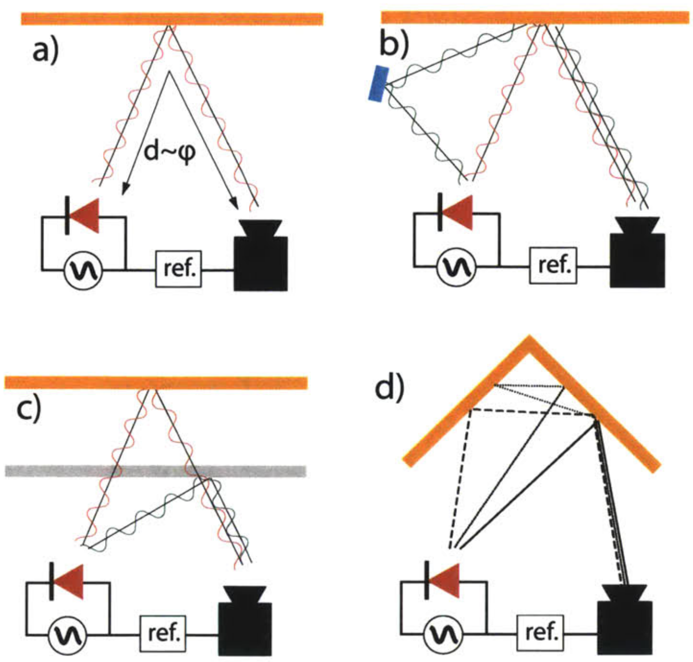


Figure 3-1: (a) ToF principle: the phase delay of an emitted AMCW wave proportionally encodes the distance of the reflecting object. (b) Mirror-like and (c) semi-transparent reflections produce MPI at a given camera pixel and yields an incorrect phase. (c) A complicated scene with severe MPI.

## 3.3 Problem Formulation

Consider first the single-component case. Mathematically, the camera emits the normalized time-modulated intensity $s(t)^1$ and detects a signal $r(t)$:

$$s\left(t\right) = 1 + s_0 \cos\left(\omega t\right), t \in \mathbb{R} \tag{3.1a}$$

$$r\left(t\right) = \Gamma(1 + s_0 \cos\left(\omega t - \phi\right)). \tag{3.1b}$$

Here, $s_0$ and $\Gamma \in [0, 1]$ are the signal modulation depth and the reflection amplitude, respectively, $\omega$ is the modulation frequency, and $\phi$ is the phase delay between the reference waveform $s\left(t\right)$ and the delayed version $r\left(t\right)$. For a co-located source and detector, the distance to the object from the camera is given by the relation $d = c\phi/2\omega$, where $c$ is the speed

---

[1] Here, we consider continuous wave imaging and hence the sinusoidal model, but the discussion is generally applicable to any periodic function.

of light.

Electronically, each pixel acts as a homodyne detector, measuring the cross-correlation between the reflected signal and the reference. Denoting the complex conjugate of $f \in \mathbb{C}$ by $f^*$, the cross-correlation of two functions $f$ and $g$ is

$$\mathsf{C}_{f,g}(\tau) \overset{\text{def}}{=} \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f^*(t+\tau)\, g(t)\, dt. \tag{3.2}$$

Note that infinite limits are approximately valid when the integration window $2T$ is such that $T \gg \omega^{-1}$. A shorter time window produces residual errors, but this is easily avoidable in practice. The pixel samples the cross-correlation at discrete times $\tau_q$:

$$\begin{aligned} m_\omega[q] &\overset{(1.3)}{=} \lim_{T\to\infty} \frac{1}{2T} \int_{\Omega_\psi} \psi(t,z)\, r(z)\, dz \bigg|_{t=\frac{\pi q}{2\omega}}, \quad \Omega_\psi = [-T,T] \\ &\overset{\text{def}}{=} \mathsf{C}_{s,r}(\tau_q) \overset{(3.2)}{=} \Gamma\left(1 + \frac{s_0^2}{2}\cos(\omega\tau_q + \phi)\right) \end{aligned} \tag{3.3}$$

where in view of 1.3, $\psi(t,z) = \bar{s}(t-z)$.

Using the "4 Bucket Sampling" technique [24], we calculate the estimated reflection amplitude and the phase, $\widetilde{\Gamma}, \widetilde{\phi}$, using four samples $\tau_q = \pi q/2\omega$ with $q = 0, ..., 3$:

$$\widetilde{\Gamma} = \sqrt{(m_\omega[3] - m_\omega[1])^2 + (m_\omega[0] - m_\omega[2])^2}/s_0^2, \tag{3.4a}$$

$$\tan\widetilde{\phi} = \left(\frac{m_\omega[3] - m_\omega[1]}{m_\omega[0] - m_\omega[2]}\right). \tag{3.4b}$$

Therefore, we associate a complex value, $z_\omega$, with a pixel measurement:

$$z_\omega = \widetilde{\Gamma} e^{\jmath\widetilde{\phi}(\omega)}. \tag{3.5}$$

Note that these results are formally equivalent to wavefront reconstruction via phase-shifting digital holography [46]. Also, since we have,

$$\widetilde{\Gamma} e^{\jmath\widetilde{\phi}(\omega)} = \widetilde{\Gamma} e^{\jmath\left(\widetilde{\phi}(\omega)+2\pi\right)},$$

the computed distance is unambiguous provided,

$$\widetilde{\phi}(\omega) \in [0, 2\pi],$$

failing which the distance that is estimated will be,

$$\widetilde{d} = \frac{c}{2\omega}\left(\widetilde{\phi} + 2\pi k\right), k \in \mathbb{Z}.$$

The problem of phase ambiguity can be solved using multiple frequency based estimation [47, 36].

When multiple reflections contribute to a single measurement, the return signal comprises a sum. In phasor notation, for $K$ components,

$$r(t) = C_0 + \sum\nolimits_{k=0}^{K-1} \Gamma_k e^{\jmath(\omega t - \phi_k(\omega))}, \tag{3.6}$$

where $C_0$ is a constant, $\phi_k(\omega) = 2d_k\omega/c$, and $\{d_k\}_{k=0}^{K-1}$ are $K$ depths at which the corresponding reflection takes place. The reflection amplitude of the $k^{\text{th}}$ surface is $\Gamma_k$. Each pixel records

$$m_\omega^K[q] = C_0 + \frac{s_0^2}{2} e^{\jmath\omega\tau_q} \sum\nolimits_{k=0}^{K-1} \Gamma_k e^{\jmath\phi_k(\omega)}. \tag{3.7}$$

Importantly, for a given modulation frequency $\omega_0$ (ignoring a constant DC term), $m_{\omega_0}^K[\tau_q] \propto \exp \jmath\omega_0\tau_q$, i.e., there is no variation with respect to individual depth components $\{\Gamma_k(\omega), \phi_k\}_{k=0}^{K-1}$ [32], regardless of the sampling density. Equivalently, the camera measurement,

$$z_\omega^{(K)} = \widetilde{\Gamma}(\omega) e^{\jmath\widetilde{\phi}(\omega)} = \sum\nolimits_{k=0}^{K-1} \Gamma_k(\omega) e^{\jmath\phi_k(\omega)} \tag{3.8}$$

is now a complex sum of $K$ reflections, which cannot be separated without independent measurements. Thus, at a given frequency, the measured phase, and hence the depth, is a nonlinear mixture of all interefering components.

Our method separates these components by recording the scene with equi-spaced frequencies $\omega = n\omega_0$ ($n \in \mathbb{N}$) and acquiring a set of measurements $\mathbf{z}$:

$$\mathbf{z} = \left[z_{\omega_0}^{(K)}, z_{2\omega_0}^{(K)}, \ldots, z_{N\omega_0}^{(K)}\right]^{\mathsf{T}}. \tag{3.9}$$

The forward model can be written compactly in vector-matrix form as $\mathbf{z} = \boldsymbol{\Phi}\boldsymbol{g} + \boldsymbol{\sigma}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{N\times K}$ is identified as a Vandermonde matrix,

$$\boldsymbol{\Phi} = \begin{bmatrix} e^{\jmath\phi_0(\omega_0)} & e^{\jmath\phi_1(\omega_0)} & \ldots & e^{\jmath\phi_{K-1}(\omega_0)} \\ e^{\jmath\phi_0(2\omega_0)} & e^{\jmath\phi_1(2\omega_0)} & \ldots & e^{\jmath\phi_{K-1}(2\omega_0)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\jmath\phi_0(N\omega_0)} & e^{\jmath\phi_1(N\omega_0)} & \ldots & e^{\jmath\phi_{K-1}(N\omega_0)} \end{bmatrix}, \tag{3.10}$$

$\boldsymbol{g} = [\Gamma_0, \ldots, \Gamma_{K-1}]^{\mathsf{T}} \in \mathbb{R}^{K\times 1}$, and $\boldsymbol{\sigma}$ represents zero-mean Gaussian i.i.d. noise, which controls the error $\varepsilon_0$ in our reconstruction algorithm. Our goal is to estimate the phases $\boldsymbol{\phi} = [\phi_0, \ldots, \phi_{K-1}]^{\mathsf{T}} \in \mathbb{R}^{K\times 1}$ and the reflection amplitude vector $\boldsymbol{g}$.

To recover these quantities, first note the similarity between $\boldsymbol{\Phi}$ and an oversampled $N \times L$ discrete Fourier transform (DFT) matrix $\boldsymbol{\Psi}$, with elements $\Psi_{nl} = \exp(\jmath nl/L)$. If $L \gg K$, the discretization of $\boldsymbol{\Psi}$ is small enough to assume that the columns of $\boldsymbol{\Phi}$ are contained in $\boldsymbol{\Psi}$. We can also define a vector $\boldsymbol{g}' \in \mathbb{R}^{L\times 1}$, whose elements are zero except for $K$ reflection amplitudes $\{\Gamma_k\}_{k=0}^{K-1}$, such that $\mathbf{z} = \boldsymbol{\Psi}\boldsymbol{g}'$. We use the ($K$-)sparsity of $\boldsymbol{g}'$ to regularize the problem:

$$\underbrace{\|\mathbf{z} - \boldsymbol{\Psi}\boldsymbol{g}'\|_{\ell_2}^2}_{\textsf{Data-Fidelity}} < \varepsilon_0 \quad \text{such that} \quad \underbrace{\|\boldsymbol{g}'\|_{\ell_0} = K}_{\textsf{Sparsity}}, \tag{3.11}$$

where the $\ell_p$–norm as $\|\mathbf{x}\|_{\ell_p}^p \overset{\text{def}}{=} \sum_n |x_n|^p$. The case of $p \to 0$ is used to define $\|\boldsymbol{g}'\|_{\ell_0}$ as the number of nonzero elements of $\boldsymbol{g}'$. Eq. 3.11 demands a least-squares solution to the data-fidelity problem $\|\mathbf{z} - \boldsymbol{\Psi}\boldsymbol{g}'\|_{\ell_2}^2$ up to some error tolerance $\varepsilon_0$, with the constraint that we accommodate up to $K$ nonzero values of $\boldsymbol{g}'$.

The sparsity of $\boldsymbol{g}'$ arises from two underlying assumptions. First, we do not consider the case of volumetric scattering, which would preclude discrete reflections and require a different parametrization (e.g., through the diffusion coefficient). Second, we ignore the con-

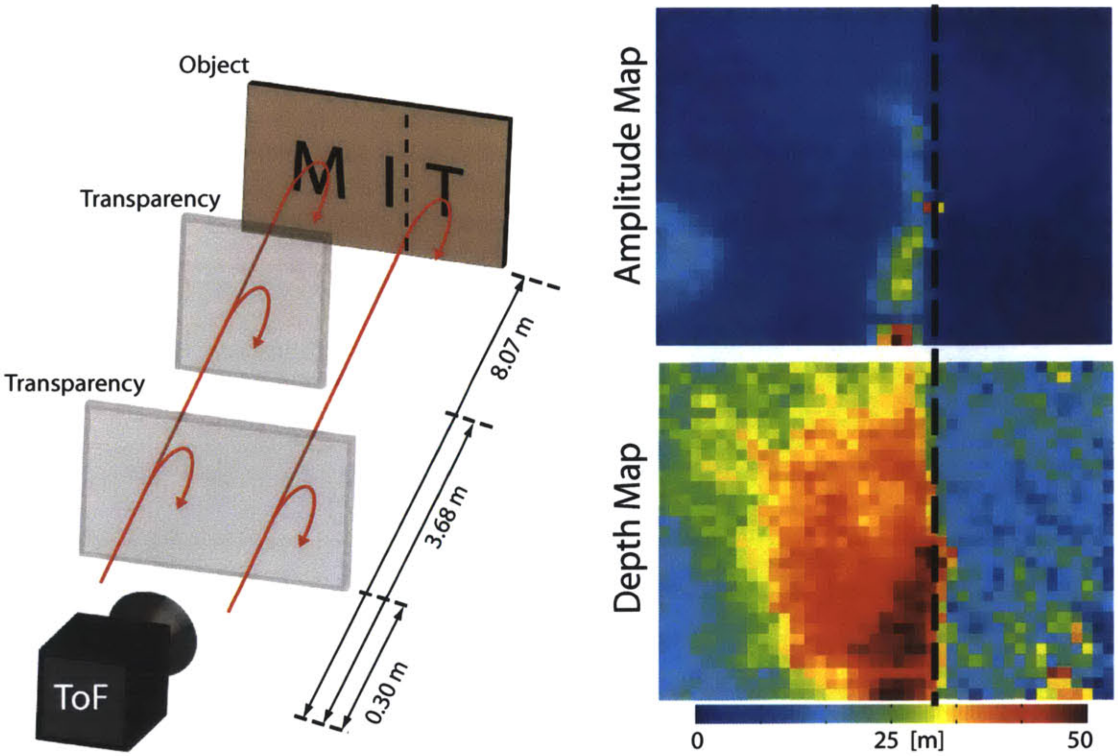


Figure 3-2: Left: experimental setup. Two transparencies block the left side of the camera (for a three-component measurement), and one transparency blocks the right (two-component measurement). Right: measured amplitude and depth at $\omega = 3\omega_0$. Dashed line indicates edge of second transparency.

tributions of inter-reflections between scattering layers, as their amplitudes fall off quickly. They could be incorporated, into our formulation, with the result of changing the sparsity of $\boldsymbol{g}'$ from $K$ to $K'$, where $K' - K$ is the number of inter-reflections considered.

We solve Eq. 3.11 via orthogonal matching pursuit (OMP), which is an iterative algorithm that searches for the best-fit projections (in the least-squares sense) of the coefficients onto an over-complete dictionary. We input $\boldsymbol{\Psi}$ and measurements $\mathbf{z}$ into the algorithm. The outputs are the set of reflection coefficients $\Gamma_k$ and their positions in $\boldsymbol{g}'$. With the position of each $\Gamma_k$ reconstructed, the corresponding phases $\phi_k$ are recovered through the elements of $\boldsymbol{\Psi}$: $\phi_k = (\jmath n)^{-1} \log(\Psi_{nl_k}) = l_k/L$, where $l_k$ is the location of $\Gamma_k$ in $\boldsymbol{g}'$.

## 3.4 Experimental Verification

We verify this theory with the experimental setup shown in Fig. 3-2. A PMD19k-2 160 × 120 sensor array is controlled by a Stratix III FPGA. Analog pixel values are converted to 16-bit unsigned values by an ADC during the pixel readout process. Eight 100 mW Sony SLD 1239JL-54 laser diodes illuminate the scene. The lasers are placed symmetrically around the detector for a coaxial configuration. The base frequency modulation is $f_0 = \omega_0/(2\pi) =$ 0.7937 MHz, and the integration time is 47 ms. The scene consists of three layers. Farthest, at 8.1 m, is an opaque wall with gray-scale text ("MIT") printed on it. Closest, at 0.3 m is a semi-transparent sheet. Between the two layers is another semi-transparent sheet that covers only the left half of the field of view. Therefore, the left-hand side records three bounces and the right only two. All three layers are within the depth of field of the camera to avoid mixed pixels from blurring.

Depth and amplitude maps acquired at a specific frequency are shown in Fig. 3-2. Due to MPI, the measured depths do not correspond to any physical layer in the scene. All depth and amplitude information from the three scene layers is mixed nonlinearly into a set of composite measurements (pixels) and cannot be recovered.

We repeat the acquisition 77 times, with modulation frequencies spaced 0.7937 MHz apart and input these data into the OMP algorithm with $K = 3$. The reconstruction, shown in Fig. 3-3, shows each depth correctly recovered. The closest depth map (Fig. 3-3 (a), first transparency) is constant. The second map (Fig. 3-3 (b)) contains two depths: the second transparency on the LHS and the wall on the RHS. The third depth map contains the wall depth on the LHS (Fig. 3-3 (c)). The third-bounce amplitude (Fig. 3-3 (f)) is zero where there are only two layers (RHS). The depth here is therefore undefined, though we set the distance to be 10 m to avoid random fluctuations. Further, the text is recovered properly in the amplitude maps corresponding to the correct depths (Figs. 3-3 (e,f)). Note that accurate depths are recovered even in the presence of strong specularity (Fig. 3-3 (e)).

A phase histogram is shown in Fig. 3-4. The histogram from the single frequency measurement in Fig. 3-1 varies from 0.6 to 1.8 rad. Recovered phases are centered around the ground truth values. The third-phase variance is wider because OMP computes the first

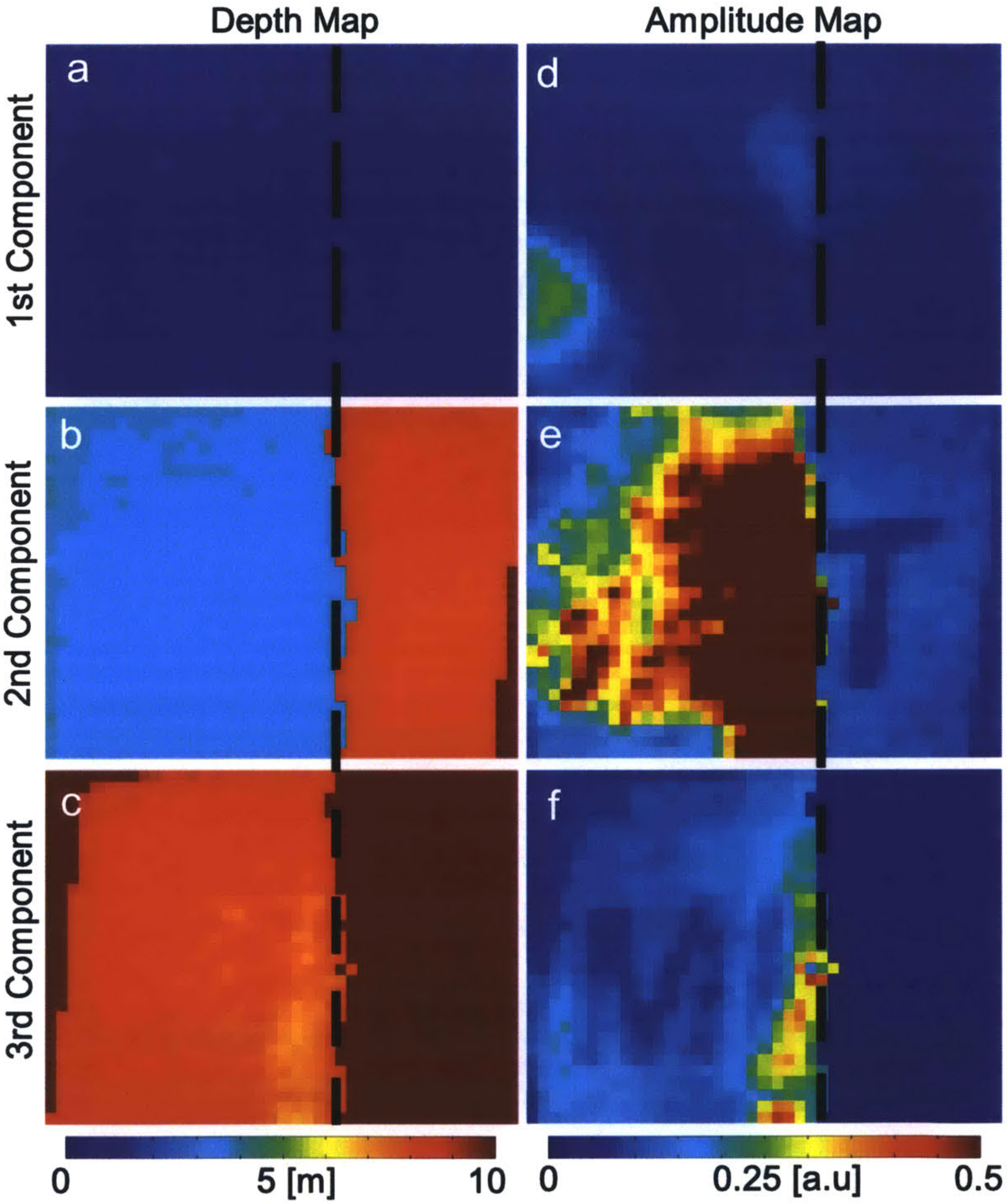


Figure 3-3: Reconstructed amplitudes and depths via sparse regularization. Dashed lines indicate edge of second transparency.

two components, leaving little residual energy, so that several columns in $\Psi$ can minimize the least-squares error.

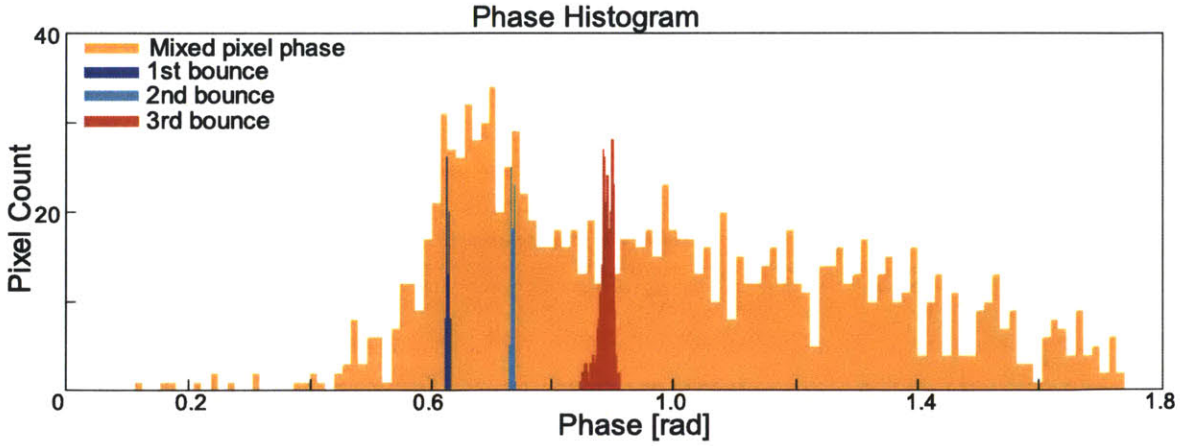


Figure 3-4: Phase histogram for reconstructed and measured depth maps. Reconstructed phases cluster around the correct depths, whereas the measured depth map has a wide variance across the entire range.

In principle, the technique can be extended to any number of bounces, provided enough modulation frequencies are used (though a first-principles derivation is beyond the scope of this contribution). In practice, however, the reflected amplitudes decrease with increasing component number, so that higher-order components diminish in importance. Furthermore, OMP need not assume a number of components that is the same as that of the physical implementation. If the assumed number is greater than the physical number, OMP will reconstruct all the physical components, with higher-order ones having an amplitude on order of the system noise. Conversely, if the assumed number is less than the physical number, OMP will recover the strongest reflections.

Therefore, the method is a generalization of global/direct illumination separation and can decompose different elements of global lighting. This is useful not only for improved depth accuracy, but also imaging in the presence of multiple scatterers such as diffuse layers, sediment, turbulence, and turbid media, as well as in places where third-component scattering must be extracted [48]. Furthermore, because it is based on phase measurements, this technique can be mapped to multiple scattering in holography [49] by substituting optical frequency for the modulation frequency.

## 3.5 Concluding Remarks

In conclusion, we implemented a multi-frequency approach for decomposing multiple depths for a ToF camera. The result is general and holds for any number of bounces, and it can be extended to non-harmonic signals [33]. Future work includes calculating bounds on measurements and resolution. The method can be incorporated with structured illumination and pixel correlations and for edge detection, and refocusing. The result holds promise for mitigating and exploiting multipath for a wide variety of scenes.

# 4

# Sparse Linear Operator Identification for Time–of–Flight Imaging

## 4.1 Overview of Results

We consider the problem of Sparse Linear Operator identification which is also linked with the topic of Sparse Deconvolution. In its abstract form, the problem can be stated as follows: Given a well behaved probing function, is it possible to identify a Sparse Linear Operator from its response to the function? We present a constructive solution to this problem. Furthermore, our approach is devoid of any sparsity inducing penalty term and explores the idea of parametric modeling. Consequently, our algorithm is non–iterative by design and circumvents tuning of any regularization parameter. Our approach is computationally efficient when compared the $\ell_0/\ell_1$–norm regularized counterparts.

Our work addresses a problem of industrial significance: decomposition of mixed–pixels in Time–of–Flight/Range imaging. In this case, each pixel records range measurements from multiple contributing depths and the goal is to isolate each depth. Practical experiments corroborate our theoretical set–up and establish the efficiency of our approach, that is, speed-up in processing with lesser mean squared error. We also derive Cramér–Rao Bounds for performance characterization.

## 4.2 Introduction and Motivation

### 4.2.1 From Sampling Functions to Sensing Operators

Since Shannon's introduction of the topic, sampling theory has been at the heart of signal processing [50]. The field was revitalized by advancements in wavelet/approximation theory, and sampling spaces were extended to a much broader class of finite–energy subspaces: the Shift–Invariant Space [51, 2], its extensions [52, 1] and non–subspace models (cf. [53, 54, 55]). In crux, most of the approximation theoretic ideas in literature master the art of approximation of functions from their equidistant or uniform samples.

Recent studies have raised an interesting question: What is the analog of sampling theory for operators? In simple terms, the key idea is to identify an operator from its response to a probing function or the identifier. In order to formulate the problem, some assumptions are made on the nature of operator. Common examples include smoothness/bandlimited and sparsity priors [56].

Within the framework of bandlimited hypothesis, Pfander discussed "Sampling of Operators" in [57]. This result inherits the non–local flavor that is central to Shannon's sampling theorem—localized reconstruction requires knowledge of all the samples. Krahmer and Pfander then take up the task of localized approximation in [58].

Departing from the bandlimited hypothesis [57], only recently, Heckel and Bölcskei discuss the problem of identification of *Sparse Linear Operators* (SLO) in [59, 60]. This work considers identification of SLO. We begin the discussion with the problem of identifying sparse operators in context of a practical application, that is, mixed–pixel problem in Time–of–

Flight (or ToF) imaging.

### 4.2.2 Motivation: Mixed–Pixel Problem in ToF Imaging

#### Time–of–Flight (ToF) Imaging from first principles

Time–of–flight (ToF) imaging is a recent imaging industry development [23] which offers an effective alternative to triangulation and stereo vision based methods for acquiring depth maps. This modality has found a breadth of applications in several areas such as non line–of–sight (NLOS) imaging [48], gesture recognition [27] and computer graphics [31, 61], to name a few. A number of ToF camera manufacturers including Microsoft (Kinect), Mesa, SoftKinetic and PMD provide competitive performance. To an entry level reader, we refer to the book [22] or the survey article [24]. In this area, the **mixed–pixel problem** (MPP) is critical to ToF cameras. There has been a surge of research to solve this problem [22, 62, 35, 63, 32, 33, 34, 36, 37]. We introduce the problem starting from first principles.

**Extracting Single Depth:** For a fixed frequency $\omega$, the ToF camera probes the scene with illuminant $p(t) = \mathrm{DC} + \alpha\cos(\omega t)$—an Amplitude Modulated Continuous Wave (AMCW) where the DC term ensures that optical function $p > 0$. For simplicity, we will ignore the DC term. After reflection from an object at depth $d$ meters from the camera, the probing function assumes form of,

$$r(t) = \beta\cos(\omega(t - 2d/c)) \equiv \underbrace{\beta\cos(\omega t - \phi_\omega)}_{\text{Reflected Signal}}, \tag{4.1}$$

where $c = 3 \times 10^8$ m/s is the speed of light and $\beta$ is the reflection coefficient. The relative delay (or phase $\phi_\omega = \frac{2\omega d}{c}$) with respect to the reference signal encodes the depth information of the scene. Fig. 4-1. (a) explains the physical setup.

Let $\langle p, r\rangle_I = \int_I p(t)\, r^*(t)\, dt$ denote the standard $L_2$ inner–product between functions $p, r \in L_2$ where $r^*(t)$ is the complex–conjugate of $r(t)$. Following this, cross–correlation between functions $p$ and $r$ is defined as, $\mathsf{C}_{p,r}(\tau) = \langle p(t+\tau), r(t)\rangle_{[-\Delta,\Delta]}$. The ToF lock–in pixel [24] then decodes the depth information using the **Four Bucket Principle** which is as follows. Starting with,

$$m(\tau) = \lim_{\Delta \to \infty} \frac{C_{p,r}(\tau)}{2\Delta} = \frac{1}{2}\alpha\beta\cos(\omega\tau + \phi_\omega), \tag{4.2}$$

each lock–in pixel computes four discrete measurements, $m[k] = m(k\pi/2\omega)$, $k = 0, \ldots, 3$. Let $m_{k,l} = m[k] - m[l]$. The reflection coefficient $\beta$ and the corresponding phase $\phi$ are estimated by using, $\widetilde{\beta} = \left(m_{3,1}^2 + m_{0,2}^2\right)^{1/2}/\alpha$ and $\widetilde{\phi} = \arctan(m_{3,1}/m_{0,2})$.

This methodology makes ToF camera a real–time sensing device because the depth estimation is computationally efficient. However, this is a fragile result which only holds for a **strong** assumption. If two or more depths correspond to the same lock–in pixel, the measurements are corrupted. For example, a transparent sheet between the scene and the camera will disrupt the hypothesis much in the same way a photograph gets corrupted during imaging through a reflective surface. Fig. 4-1. (b) shows an exemplary setting which results in corrupted measurements. Assuming negligible inter–reflections, in case of $K$–depths the reflected function is modeled as, $r(t) = \sum_{k=0}^{K-1} \beta_k \cos(\omega t - \phi_{k,\omega})$ and the measurements take form of,

$$m_K(\tau) \overset{(4.2)}{=} \lim_{\Delta \to \infty} \frac{C_{p,r}(\tau)}{2\Delta} = \alpha \sum_{k=0}^{K-1} \beta_k \cos(\omega\tau + \phi_{k,\omega}). \tag{4.3}$$

This is the **mixed–pixel problem** (MPP) in context of ToF imaging as it is impossible to decode the $K$–depths from $m_K(\tau)$ (cf. Fig. 4-1 and Fig. 4-3). This problem has received a lot of attention in the recent past [62, 35, 63, 32, 33, 34, 36, 37] due to difficulty of extracting $\phi_{k,\omega} = \frac{2d_k\omega}{c}$ from $m_K(\tau)$. An inspection of (4.3) reveals an interesting operator identification problem. Let '$*$' denote convolution operation. Notice that,

$$\sum_{k=0}^{K-1} \beta_k \cos(\omega t - \phi_{k,\omega}) \overset{(4.1)}{=} \cos(\omega t) * \sum_{k=0}^{K-1} \beta_k \delta(t - 2d_k/c).$$

Indeed this problem is associated with a sparse linear operator as,

$$m_K : p \to p * \underbrace{\sum_{k=0}^{K-1} \beta_k \delta\left(\cdot - \tfrac{2d_k}{c}\right)}_{\text{Sparse Linear Operator}}, \quad p = \cos(\cdot).$$

However, this is a **special case** that uses AMCW [36, 34] with $p = \cos(\cdot)$. One need not restrict to the class of cosine probing functions.

### 4.2.3 Contributions and Organization of this chapter

The MPP has been considered in number of papers (cf.[62, 35, 63, 32, 33, 34, 36, 37] and references there in). Existing approaches can resolve up to two components and use AMCW/sinusoidal model. There is no mathematical framework that can explain solutions in a general setting. In Section 4.3.1, we establish a concrete link between mixed pixel problem (MPP) and the sparse linear operator identification (SLO) problem. In Section 4.3.3, we outline a closed–form solution to recover the SLO without using any sparsity regularized technique. In Section 4.4.1, we discuss experimental results. We resolve multiple components of mixed–pixels acquired using a ToF camera.

- **Novelty**: Our approach to identify the sparse linear operator is devoid of any sparsity inducing penalty term. We solve the problem using parameter estimation method for which we derive Cramér–Rao Bounds. The method is easily implementable (cf. Section 4.4).

- **Computational Efficiency**: Our solution is closed–form/non–iterative by design. It avoids the computation of regularization parameter which is central to optimization based methods. The complexity of implementation is discussed in Section 4.3.4.

- **Practicability**: The approach is verified on practical experiments linked with ToF imaging. We report speed–up over OMP and LASSO as well as lesser mean squared errors (cf. Fig. 4-2, 4-3).

## 4.3 The Sparse Operator Identification Problem

We define the Fourier Transform of $p$ as $\hat{p}(\omega) = \mathcal{F}[p] \overset{\text{def}}{=} \langle p, e^{\jmath \omega t} \rangle$. Let $\mathrm{B}_\Omega$ be the set of bandlimited functions, $\mathrm{B}_\Omega = \{p : \|p\|_{L_2} < \infty, \hat{p}(\omega) = 0, |\omega| > \Omega\}$. In this chapter, we will consider bandlimited probing functions, that is $p \in \mathrm{B}_\Omega$.

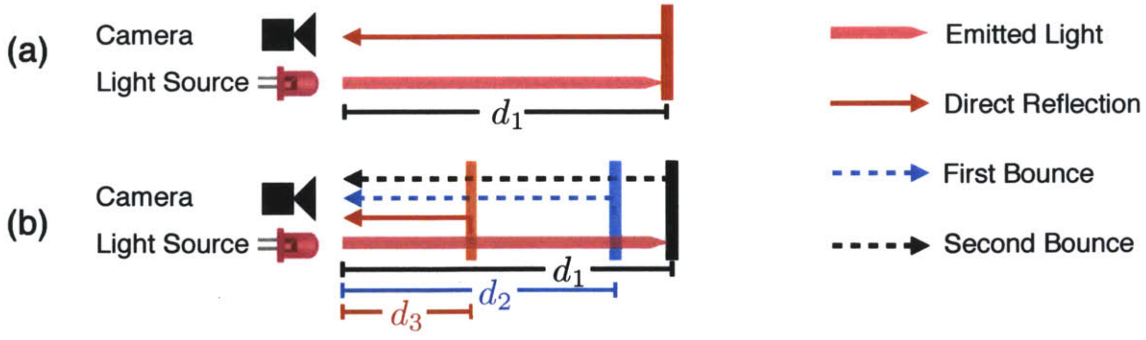


Figure 4-1: (a) The ToF camera emits a reference signal. Time delay of arrival from direct reflection encodes the depth $d_1$. (b) Demonstration of mixed–pixel problem (MPP): two or more paths contribute to the pixel. The inter–reflections between sheets are assumed to be negligible.

### 4.3.1 Link with the Mixed–Pixel Problem in ToF Imaging

Let $s = p \in \mathrm{B}_\Omega$. The link between (4.5) and the mixed pixel problem can be established by developing $m_K(\tau) \overset{(4.2)}{=} \lim_{\Delta\to\infty} \mathsf{C}_{s,r}(\tau)/2\Delta$ or,

$$\begin{aligned} m_K(\tau) &= \langle p(t+\tau), r(t)\rangle_{t\in\mathbb{R}} \\ &= \int p(t+\tau) \sum_{k=0}^{K-1} \beta_k p\left(t - \frac{2d_k}{c}\right) dt \\ &\overset{(a1)}{=} \sum_{k=0}^{K-1} \beta_k \int p(z)\, p\left(z - \left(\tau + \frac{2d_k}{c}\right)\right) dz \\ &\overset{(a2)}{=} \underbrace{\mathsf{C}_{p,\overline{p}} * \sum_{k=0}^{K-1} \beta_k \delta(\tau - t_k)}_{2K \text{ Sparse Linear Operator}}, \quad t_k \overset{\text{def}}{=} -\tfrac{2d_k}{c} \end{aligned} \tag{4.4}$$

where (a1) is due to $t + \tau \mapsto z$ and (a2) is due to $2d_k/c \mapsto -t_k$. Next, we define the sparse linear operator.

**Sparse Linear Operator** *Let* $p \in B_\Omega$, $\overline{p}(t) \overset{\text{def}}{=} p(-t)$ *and* $\mathsf{C}_{p,\overline{p}} = (p * \overline{p})$. *We define a Sparse Linear Operator of* $2K$ *parameters as:*

$$\mathcal{O}_K[p] : p \to \mathsf{C}_{p,\overline{p}} * \sum_{k=0}^{K-1} \beta_k \delta(\cdot - t_k). \tag{4.5}$$

*Parameters* $\{\beta_k, t_k\}_{k=0}^{K-1}$ *completely characterize the operator* $\mathcal{O}_K$.

From a signals and system perspective, the input/probing function and the output are related as follows:

$$p \to \boxed{\mathcal{O}_K} \to \sum\nolimits_{k\in\mathcal{K}} \beta_k \mathsf{C}_{p,\overline{p}}\left(\cdot - t_k\right), \quad \mathsf{C}_{p,\overline{p}} = \left(p * \overline{p}\right).$$

Consequently, we conclude that $m_K\left(\tau\right) \overset{(4.4)}{=} \mathcal{O}_K\left[p\right]\left(\tau\right)$.

**Inverse Problem:** Given $m_K\left(\tau\right), \tau \in \mathbb{R}$, how can we characterize the $2K$–Sparse Linear Operator, or $\mathcal{O}_K$ in (4.5), with $\left\{\beta_k, t_k\right\}_{k=0}^{K-1}$ ?

**Remark:** We use (4.4) for notational simplicity and to introduce the problem. It is noteworthy to mention that **more generally**, the measurements can be written as inner–products in sense of distributions,

$$m_K\left(\tau\right) = C_0 + \left\langle \mathsf{C}_{p,\overline{p}}\left(t+\tau\right), \omega_0 \beta\left(t\right) \delta\left(\mathsf{H}\left(e^{-\jmath\omega_0 t}\right)\right)\right\rangle, \omega_0 = \tfrac{2\pi}{T},$$

where $\mathsf{H}\left(z\right), z \in \mathbb{C}$ is a proxy polynomial to be defined in (4.14). This formalism is *backward compatible* with (4.4) and its special case (4.3). The equivalence is obtained by invoking the function composition property of Dirac distributions (cf. pg. 184 [64]) which states:

$$\left\langle f_1, \delta\left(f_2\right)\right\rangle = \sum\nolimits_k c_k f_1\left(t_k\right),$$

with $t_k = \left\{t : f_2\left(t\right) = 0\right\}$, and $c_k^{-1} = \left|\partial_t f_2\left(t_k\right)\right|$. In our case, $c_k$'s are identically equal to $\omega_0^{-1}$ and $\beta_k = \beta\left(t_k\right)$. With $K$ distinct roots, other choices of $\mathsf{H}$ are possible.

The identification problem in (4.5) is central to the mixed pixel problem. For the case of $K$ depths, we have,

$$r\left(t\right) \overset{\text{def}}{=} 1 + \sum\nolimits_{k=0}^{K-1} \beta_k p\left(t - \frac{2d_k}{c}\right). \tag{4.6}$$

### 4.3.2 Related Work

In general, (4.5) is a multi–path/sparse deconvolution problem which has been a topic of several papers starting [65]. Almost all papers use a sparsity inducing penalty term [66, 67]. From a modern perspective, Santosa and Symes [68] introduced the notion of $\ell_2/\ell_1$–

optimization of cost–function to solve the problem (P1),

$$\boxed{\mathsf{P1}} \qquad J_\lambda(\mathbf{x}) = \|\mathbf{Tx}-\mathbf{m}\|_{\ell_2}^2 + \lambda\|\mathbf{x}\|_{\ell_1} \tag{4.7}$$

where **T** is a convolution/Toeplitz matrix. The first row/column of **T** is the auto–correlation of the probing signal and **m** are the measurements. Our problem can be cast as variant of P1 and solved using optimization methods such as Orthogonal Matching Pursuit (OMP) and LASSO [56, 69]. For instance, in [33], we used LASSO and OMP. However, the technique is computationally intensive which is inherent to the sparsity induced optimization recipe. The problem scales with the number of pixels, size of the probing function and the sparsity level. Depth–sensing with compressive sensing methods was used in [70]. Within super–resolution framework, this problem was recently discussed in [71]. The solution is based on optimization.

With the bandlimited hypotheses, our work comes close to the topic of sparse sampling [53, 54, 55] with marked differences. The fact that we can design the function $p$, and hence $C_{p,\overline{p}}(t)$ is a huge degree of freedom which is not the case in sampling theory. In fact, not being able to design $p$ is a restriction in context of sampling theory. All the more, we are *not sampling* a stream of Dirac impulses. We are *interested in system identification* (4.3.1). A similar idea was pursued in [72], however, the solution was based on $\ell_1$ penalty term.

### 4.3.3 Sparse Linear Operator Identification without Sparse Regularization: Forward Model and Solution to Inverse Problem

Consider a probing function $p \in \mathrm{B}_\Omega$ which is $T$–periodic. In practice, this periodization is a standard trick. It goes by the name of **cyclic prefixing** in OFDM related literature [73]. We use an M–sequence for our practical set up (4.15). Consequently, $p$ is defined as,

$$p(t) = \sum\nolimits_{|m|\leqslant\Omega} \hat{p}_m e^{\jmath m\omega_0 t}, \quad \omega_0 = 2\pi/T \tag{4.8}$$

where $\hat{p}_m = \langle p, e^{\jmath m\omega_0 t}\rangle$ are the Fourier series coefficients of $p(t)$. Let $\widehat{\phi}_m = \widehat{p}_m\widehat{p}^*_m$. It is not difficult to show that $\mathsf{C}_{p,\bar{p}}(t) = \phi(t)$,

$$\mathsf{C}_{p,\bar{p}}(t) \overset{(4.8)}{=} \sum_{|m|\leqslant\Omega} \hat{\phi}_m e^{\jmath m\omega_0 t}. \tag{4.9}$$

In this setting, we will show that it is possible to reframe problem P1 in (4.7) as a parameter estimation problem. Observe that the response of the Sparse Linear Operator to the probing function is modeled by,

$$\begin{aligned}
\mathcal{O}_K[p](t) &\overset{(4.5)}{=} \mathsf{C}_{p,\bar{p}} * \sum_{k\in K} \beta_k \delta(\cdot - t_k) \\
&= \sum_{k=0}^{K-1} \beta_k \underbrace{\langle \mathsf{C}_{p,\bar{p}}(\tau), \delta(t-\tau-t_k)\rangle_{\tau}}_{\mathsf{C}_{p,\bar{p}}(t-t_k)\equiv\phi(t-t_k)} \\
&\overset{(4.9)}{=} \sum_{|m|\leqslant\Omega} \hat{\phi}_m \underbrace{\sum_{k=0}^{K-1} \beta_k e^{-\jmath\omega_0 m t_k}}_{b_m(t)} e^{\jmath\omega_0 m t}.
\end{aligned} \tag{4.10}$$

Also, $b_m(\boldsymbol{t})$ is parameterized by the vector $\boldsymbol{t} = [t_0, \ldots, t_{K-1}]^{\top}$. On discretization, the above assumes a compact form of,

$$\mathcal{O}[p](n) \overset{(4.10)}{=} \mathbf{V}_{\mathsf{IDFT}}\mathbf{D}_{\hat{\phi}}\mathbf{V}_{t}\boldsymbol{\beta} \overset{\text{def}}{=} \mathbf{y}, \tag{4.11}$$

or simply, $\mathbf{V}_{\mathsf{IDFT}}\mathbf{D}_{\hat{\phi}}\mathbf{b} = \mathbf{y}$ with $n = 0, \ldots, N-1$ and where,

- $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{K-1}]^{\top}, \boldsymbol{\beta} \in \mathbb{R}^{K\times 1}$ is the coefficient vector.
- $\mathbf{V}_{\boldsymbol{t}} \in \mathbb{C}^{(2\Omega+1)\times(K)}$ is a $\boldsymbol{t}$–parameterized Vandermonde matrix with elements, $[\mathbf{V}_{\boldsymbol{t}}]_{m,k} = \exp(\jmath m\omega_0 t_k), \quad m = -\Omega, \ldots, +\Omega$.
- $\mathbf{D}_{\hat{\phi}} \in \mathbb{C}^{(2\Omega+1)\times(2\Omega+1)}$ is the diagonal matrix $\mathbf{D}_{\hat{\phi}} = \operatorname{diag}\left(\hat{\boldsymbol{\phi}}\right)$, with $\widehat{\boldsymbol{\phi}} = [\hat{\phi}_{-\Omega}, \ldots, \hat{\phi}_{+\Omega}]^{\top} \equiv \left[\hat{p}_{-\Omega}\hat{p}^*_{-\Omega}, \ldots, \hat{p}_{\Omega}\hat{p}^*_{\Omega}\right]^{\top}$ (4.9).
- $\mathbf{V}_{\mathsf{IDFT}} \in \mathbb{C}^{(N\times 1)\times(2\Omega+1)}$ is the usual inverse DFT/Vandermonde matrix with matrix elements, $\mathbf{V}_{\mathsf{IDFT}} = \left[e^{\jmath\omega_0 nm}\right]_{n,m}$.

We then define $\mathbf{b} = \mathbf{V}_t \boldsymbol{\beta}$. Next, we will outline the solution of the inverse problem of retrieving parameters $\boldsymbol{\beta}$ and $\boldsymbol{t}$ from $\mathbf{y}$.

**Solving the Inverse Problem: Retrieving $\boldsymbol{\beta}$ and $\boldsymbol{t}$**

In practice, we have pixel-wise measurements, $m_K(\tau) = \mathcal{O}_K[p](\tau)$ (4.4). Given $p$, $\mathbf{D}_{\hat{\phi}}$ is fixed by design. Let the discrete measurement vector be defined as,

$$\mathbf{y} = \mathbf{m} = [m_K(0), \ldots, m_K(N-1)]^\top,$$

and let $\mathbf{A}^+$ be the pseudo–inverse of matrix $\mathbf{A}$. Under the **conditions** that $N \geq 2\Omega + 1$ and $\Omega \geq K$ [74], we have,

$$\mathbf{b} = \mathbf{D}_{1/\hat{\phi}} \mathbf{V}_{\mathsf{IDFT}}^{+} \mathbf{y}, \quad \mathbf{b} \in \mathbb{R}^{N \times 1}. \tag{4.12}$$

Having computed $\mathbf{b}$, it remains to estimate parameters $\boldsymbol{t}$ and $\boldsymbol{\beta}$ from,

$$\mathbf{b} = \mathbf{V}_t \boldsymbol{\beta} \Leftrightarrow b_m = \sum\nolimits_{k=0}^{K-1} \beta_k e^{-\jmath \omega_0 m t_k}. \tag{4.13}$$

This is the classic line spectrum estimation/Prony's problem [7, 75, 76, 77, 74]. Starting with a polynomial, *Stoica*1997$\mathsf{H}(z), z \in \mathbb{C}$, of form,

$$\mathsf{H}(z) = \sum\nolimits_{m=0}^{K-1} h_m z^{-m} \equiv \prod\nolimits_{k=0}^{K-1} \left(1 - e^{-j\omega_0 t_k} z^{-1}\right), \tag{4.14}$$

it is not difficult to show that $h * b = 0 \Leftrightarrow z_k = e^{-j\omega_0 t_k}$, that is, the roots of $\mathsf{H}(z)$ encode the locations $\{e^{-j\omega_0 t_k}\}_{k=0}^{K-1}$ [7, 75, 76, 77, 74, 78, 53]. Since $\mathbf{h}$ is in the null–space of Toeplitz matrix $\mathbf{T}$ constructed from $\mathbf{b}$, the singular value decomposition of $\mathbf{T}$ leads to $\mathbf{h}$ which in turn results in $\mathsf{H}(z)$. From the roots of $\mathsf{H}(z)$, that is $z_k = e^{-j\omega_0 t_k}$, we are able to compute $t_k = -\jmath \arg(z_k)/\omega_0$ and hence $\mathbf{V}_t$ [78]. Given $\mathbf{V}_t$ and $\mathbf{b}$, we finally compute $\boldsymbol{\beta} = \mathbf{V}_t^+ \mathbf{b}$. This solves the problem.

### 4.3.4 Note on Enforcing Sparsity and Computational Complexity

As opposed to $\ell_0$ or $\ell_1$ penalty based methods, sparsity manifests as a rank–constraint on Hankel/Topelitz matrix used for solving (4.13). This is an in–built feature of algorithms such as Matrix Pencils [76] or [75], which solve (4.13). Note that $\ell_0/\ell_1$ based methods start with a dictionary/matrix $[\mathbf{T}]_{m,n} = \mathsf{C}_{p,\overline{p}}(m-n)$ in (4.7). Hence they are limited in resolution of shifts $t_k$ upto the grid of $\mathbf{T}$. Our method uses (4.13). In noiseless setting, it is **not limited** by resolution and $t_k$'s can be arbitrarily close upto machine precision in simulation. This is a limitation for problem setup in (4.7). On the other hand, we must remark that (4.12) must be stabilized in noisy setting. We use Matrix Pencils [76] for the case of noise and [75] for model mismatch.

While the computational complexity of OMP [79] with $\mathbf{T}^{N\times N}$ (4.7) and for $K$–sparse signal scales as, $2KN^2+2K^2N+4KN+K^3$, our method uses $(2K+1)\,2^{\sqrt{2\log(2K+1)}}\log(2K+1)$ for diagonal matrix multiplication (pointwise multiplication) and $N\log N$ for DFT in (4.12). What remains is due to Matrix Pencils [76] which is known to be efficient in implemention.

## 4.4 From Theory to Practice

### 4.4.1 Design of Practical Experiments with ToF camera

We demonstrate the efficiency of our method on two practical experiments. To set up the experiment, we first calibrate the probing function to obtain $\mathsf{C}_{p,\overline{p}}$. For this purpose, we use cyclic–prefixing of a code based on M–sequence which is prescribed by,

$$\mathsf{M}_{\text{seq}} = 0101110110001111100110100100001. \tag{4.15}$$

We omit the details of conversion of $\mathsf{M}_{\text{seq}}$ into $p(t)$ and eventually $\mathsf{C}_{p,\overline{p}}(t)$. In Fig. 4-2. (b), we plot $\mathsf{C}_{p,\overline{p}}(t)$ with its Fourier Series approximation in (——) as well as the Fourier Series coefficients, $\widehat{\phi}_m$ in the inset. For this case $\Omega = 29$, $N = 59$. Each camera pixel measures a mixture of light paths from 3 objects ($K = 3$): a glass unicorn at depth $z = 0$, a plexiglass sheet at depth $z = 2.00$ meters and a wall at depth $z = 3.875$ meters. To resolve the mixed pixels (MPP), we use the SLO formulation in Section 4.3. Starting with measurements

$\mathbf{y} \in \mathbb{R}^{4464\times 1}$ in Fig. 4-2. (b), we estimate $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{t}}$. Unfortunately, it is hard to obtain the ground truth for $\boldsymbol{\beta}$ however, the ground truth for $\boldsymbol{t} = [\texttt{2.0139 2.1553 2.2813}]^{\top}$ ps or picoseconds. When we solve the problem using (P1) in (4.7) using OMP, we obtain $\widetilde{\boldsymbol{t}}_{\mathsf{OMP}} = [\texttt{2.020 2.156 2.257}]^{\top}$ ps. For our case, we first obtain $\mathbf{b}$ (4.12) and then use Matrix–Pencils [76] for our problem. Our estimate is reported as $\widetilde{\boldsymbol{t}} = [\texttt{2.014 2.159 2.2289}]^{\top}$ ps. The estimates are plotted in Fig. 4-2. (c). We denote the Mean Squared Error or MSE between vectors $\mathbf{x}$ and $\mathbf{y}$ by $\mathsf{MSE}(\mathbf{x}, \mathbf{y})$. Our estimation is better than OMP in MSE sense. For our experiments, $\mathsf{MSE}(\boldsymbol{t}, \widetilde{\boldsymbol{t}}) = 5.5 \times 10^{-3}$ while, for OMP, $\mathsf{MSE}(\boldsymbol{t}, \widetilde{\boldsymbol{t}}_{\mathsf{OMP}}) = 4.7 \times 10^{-2}$. Our solution is non–iterative. For the same computing resources on `Matlab`, our solution offers an average speed–up of 16.7x compared to OMP. This is tested across a patch of $31 \times 31 = 961$ mixed–pixels. The computational time is reported in Fig. 4-2. (d). In Fig. 4-2. (a), we compare the result of using Matrix–Pencils [76] and Cadzow's algorithm [75] and plot MSE (in dB) versus Signal–to–Noise Ratio (SNR) for SNR range of $-10$ to 38 (dB) for different values of oversampling factor $\eta$ such that $N = 2\eta K + 1$. We consider $\eta = 1, \ldots, 5$. Oversampling results in improvement of results. For experiments, we used $K = 3$ with uniformly distributed $\boldsymbol{t} \in [0, 1]$ and the results are an average over 5000 trials/SNR value. In another experiment, we consider the case of $K = 2, \Omega = 28, N = 57$. A diffusive sheet covers a placard reading: "Time Of Flight," which is 2 meters away. The camera measurements are shown in Fig. 4-3. (a). Indeed it is impossible to read anything. However, since each measurement corresponds to a mixed–pixel, we can use the sparse operator identification method to decouple the measurements. Consequently, in Fig. 4-3. (b), we show the strong reflection from the translucent sheet, that is $\{\beta_0, t_0\}$. Since we have estimated $\{\beta_1, t_1\}$, it is possible to read through the translucent/diffusive sheet as shown in Fig. 4-3. (c). In Fig. 4-3. (d), we show the measurements $\mathbf{y}$ and estimates $\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{t}}$.

### 4.4.2 Cramér–Rao Lower Bounds (CRB): Overview of Results

With zero–mean, Gaussian noise assumption on $e_n$ and covariance matrix $\Sigma = \mathcal{E}[\mathbf{e}\mathbf{e}^{\top}] = \sigma^2\mathbf{I}$ we model $N$ noisy measurements (4.11) as, $y_n = \mathcal{O}_K[p](n) + e_n \Leftrightarrow \mathbf{y} = \mathbf{m} + \mathbf{e}$. The parameter vector of interest is $\widetilde{\boldsymbol{\theta}} = [\widetilde{t}_0, \ldots, \widetilde{t}_{K-1} | \widetilde{\beta}_0, \ldots, \widetilde{\beta}_{K-1}]^{\top}$. Within this framework, the CRB of an unbiased estimator of the parameter vector $\boldsymbol{\theta}$ holds such that, $\mathcal{V}(\widehat{\boldsymbol{\theta}}) \leqslant \mathbf{J}^{-1}(\boldsymbol{\theta})$

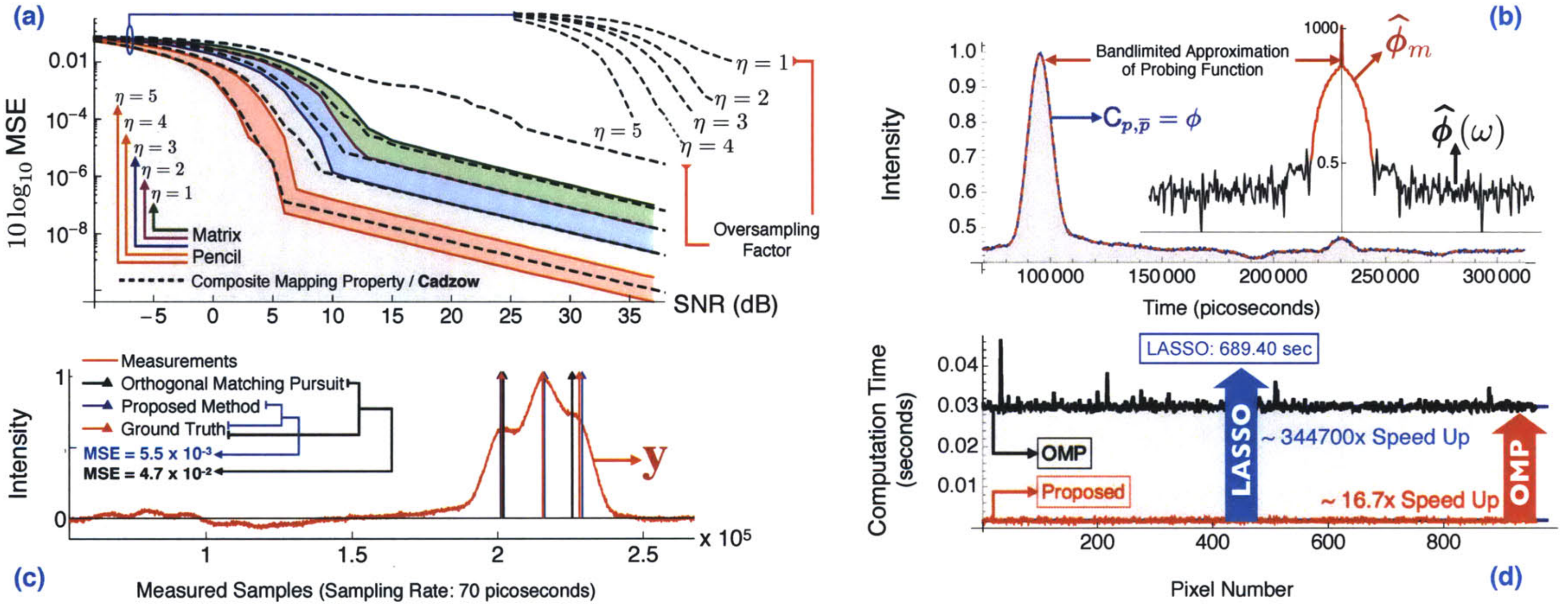


Figure 4-2: (a) Performance analysis of methods used for solving $b_m = \sum_{k=0}^{K-1} \beta_k e^{-\jmath\omega_0 m t_k}$ (4.13) with different oversampling factors $N = 2\eta K + 1$. We compare Matrix Pencil Method [76] with Cadzow's algorithm [75] for the case of $K = 3$. The parameters are chosen from a uniform distribution. For each SNR value, we average the result over 5000 realizations. (b) Starting with M–sequence in (4.15) we obtain the calibrated function $\mathsf{C}_{p,\overline{p}}$ in —- and its $\Omega = 29$–term Fourier Series approximation, $\phi \in \mathrm{B}_\Omega$ in – –. The inset shows the Fourier Series coefficients (4.9). (c) Mixed–Pixel Problem for $K = 3$ mixing paths. The measurements $\mathbf{y}$ are marked in —-. Ground truth and estimates using OMP in (4.7) and our method are also marked. (d) We report the computation time involved with using LASSO, OMP and our non–iterative method. OMP requires $\sim 0.03$ seconds per pixel while our method (using Matrix Pencils) requires $\sim 0.002$ seconds. This result is consistent over a set of 961 pixels. LASSO is computationally intensive. For one pixel, the requirement is 689.40 seconds.

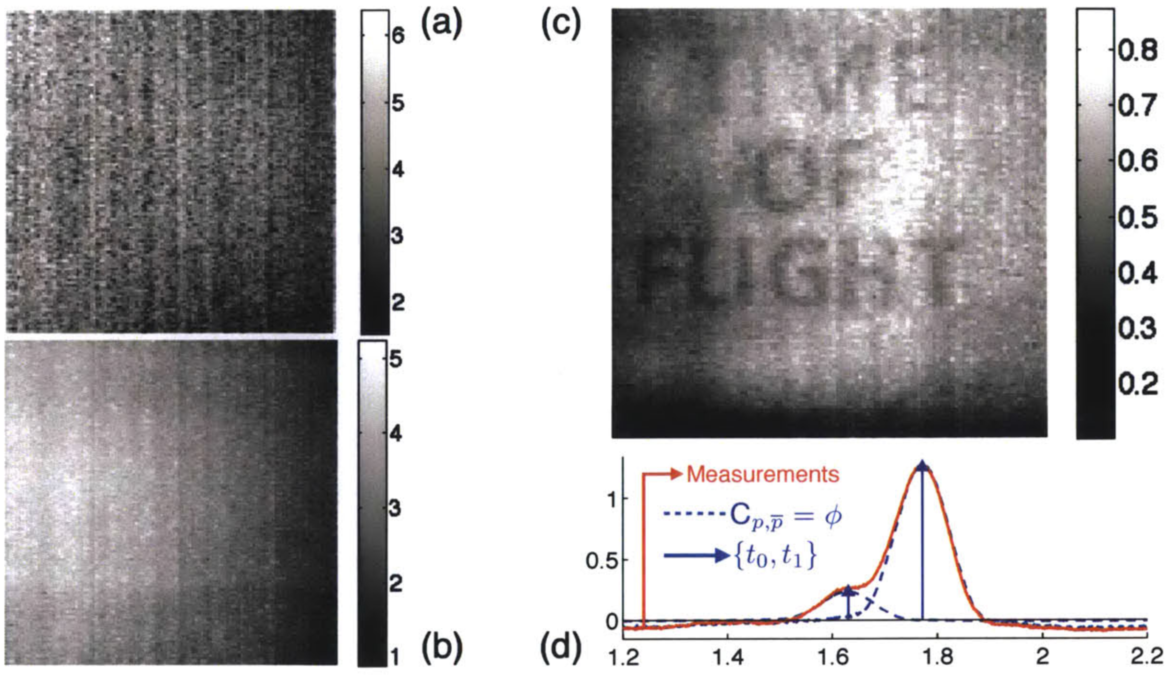


Figure 4-3: (a) Measurements (—) of a scene where a translucent/diffusing sheet $\{\beta_0, t_0\}$ hides a placard $\{\beta_1, t_1\}$. The mixed–pixels correspond to the case of $K = 2$. Using our method, the mixed–pixels in (a) are decomposed into the diffusing sheet in (b) and the placard that reads, "Time of Flight" in (c). (d) We show measurements $\mathbf{y}$ for one of the pixel of $120 \times 120$ image. These measurements are decomposed into shifted and amplitude scaled versions of $\mathsf{C}_{p,\overline{p}} = \phi$ marked in - - - -. Depths $\propto \{t_0, t_1\}$ are marked with →.

where $\mathbf{J}$ is the Fisher Information Matrix and $\mathcal{V}$ is the variance operator. Following the CRB simplification for Gaussian distribution (cf. Appendix 3C [80]), we have, $\mathbf{J}(\boldsymbol{\theta}) = \sigma^{-2}\boldsymbol{\Upsilon}^{\top}\boldsymbol{\Upsilon}$. For simplicity we consider $K = 1$, then,

$$\boldsymbol{\Upsilon}^{\top} = \left[\frac{-\beta_0\phi'(1-t_0) \quad \cdots \quad \beta_0\phi'(N-t_0)}{\phi(1-t_0) \quad \cdots \quad \phi(N-t_0)}\right].$$

Since the anti–diagonal of $\mathbf{J}(\boldsymbol{\theta})$ is zero, the variance $\mathcal{V}(\widetilde{\boldsymbol{\theta}}) \geqslant \sigma^2 \mathrm{diag}\left(\beta_0^2 N \omega_0^2 S_m, N S_1\right)$ with $S_m = \sum_m m^2 \hat{\phi}_m^2$. Set $\mathsf{SNR} = \sigma^2/\beta_0^2$, we obtain the spread in estimation of $t_0$ and $\beta_0$ as follows:

$$\frac{\Delta t_0}{T} \geqslant \frac{1}{2\pi\sqrt{S_m N \cdot \mathsf{SNR}}} \quad \text{and} \quad \frac{\Delta \beta_0}{|\beta|} \geqslant \frac{1}{\sqrt{S_1 N \cdot \mathsf{SNR}}}.$$

### Derivation of Cramér–Rao Bounds for Sparse Estimation

Let $e_n$ be the zero–mean, Gaussian noise element with covariance matrix $\Sigma = \mathcal{E}[\mathbf{e}\mathbf{e}^\top] = \sigma^2\mathbf{I}$. Given measurements,

$$y_n = \underbrace{\sum_{k \in \mathcal{K}} \beta_k \mathsf{C}_{p,\overline{p}}\left(n - t_k\right)}_{\mathcal{O}_K[p](n)} + e_n,$$

where $\mathsf{C}_{p,\overline{p}}\left(t\right) \triangleq \phi\left(t\right)$ is the probing function, we seek to establish the Cramér–Rao Bound on the variance of parameter vector,

$$\widehat{\boldsymbol{\theta}} = \left[\widehat{t}_0, \ldots, \widehat{t}_K | \widehat{a}_0, \ldots, \widehat{a}_K\right]^\top.$$

Within the Cramér–Rao framework, the CRB for an unbiased estimator $\widehat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$ holds such that,

$$V\left(\widehat{\boldsymbol{\theta}}\right) \leqslant \frac{1}{\mathbf{J}\left(\boldsymbol{\theta}\right)}$$

where $\mathbf{J}$ is the Fisher Information Matrix. Moreover, for the Gaussian distribution case, the element wise Fisher Information matrix has a simpler expression which takes form of,

$$[\mathbf{J}\left(\boldsymbol{\theta}\right)]_{k,l} = \frac{1}{2}\mathrm{Tr}\left\{\Sigma^{-1}\left(\boldsymbol{\theta}\right)\frac{\partial\Sigma\left(\boldsymbol{\theta}\right)}{\partial\theta_k}\Sigma^{-1}\left(\boldsymbol{\theta}\right)\frac{\partial\Sigma\left(\boldsymbol{\theta}\right)}{\partial\theta_l}\right\} + \left[\frac{\partial\mathbf{y}\left(\boldsymbol{\theta}\right)}{\partial\theta_k}\right]^\top \Sigma^{-1}\left(\boldsymbol{\theta}\right)\left[\frac{\partial\mathbf{y}\left(\boldsymbol{\theta}\right)}{\partial\theta_l}\right],$$

where $\mathrm{Tr}\{\cdot\}$ is the trace operator. Since $\frac{\partial\Sigma(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = 0$, our problem is simplified and boils down to computing,

$$[\mathbf{J}\left(\theta\right)]_{k,l} = \left[\frac{\partial\mathbf{y}\left(\theta\right)}{\partial\theta_k}\right]^\top \underbrace{\Sigma^{-1}\left(\theta\right)}_{=\frac{1}{\sigma^2}\mathbf{I}}\left[\frac{\partial\mathbf{y}\left(\theta\right)}{\partial\theta_l}\right] = \frac{1}{\sigma^2}\left[\frac{\partial\mathbf{y}\left(\theta\right)}{\partial\theta_k}\right]^\top\left[\frac{\partial\mathbf{y}\left(\boldsymbol{\theta}\right)}{\partial\theta_l}\right].$$

In vector–matrix notation, we have,

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\boldsymbol{\Upsilon}^{\top}\boldsymbol{\Upsilon}, \qquad \text{(Fisher Information Matrix)}$$

where,

$$\boldsymbol{\Upsilon} = \underbrace{\left[\begin{array}{ccc|ccc} -\beta_0\phi'(1-t_0) & \cdots & \beta_{K-1}\phi'(1-t_{K-1}) & \phi(1-t_0) & \cdots & \phi(1-t_{K-1}) \\ -\beta_0\phi'(2-t_0) & \cdots & \beta_{K-1}\phi'(2-t_{K-1}) & \phi(2-t_0) & \cdots & \phi(2-t_{K-1}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\beta_0\phi'(N-t_0) & \cdots & \beta_{K-1}\phi'(N-t_{K-1}) & \phi(N-t_0) & \cdots & \phi(N-t_{K-1}) \end{array}\right]}_{N\times 2K}.$$

For $K = 1$,

$$\boldsymbol{\Upsilon} = \left[\begin{array}{c|c} -\beta_0\phi'(1-t_0) & \phi(1-t_0) \\ -\beta_0\phi'(2-t_0) & \phi(2-t_0) \\ \vdots & \vdots \\ -\beta_0\phi'(N-t_0) & \phi(N-t_0) \end{array}\right],$$

and consequently,

$$\begin{aligned} \boldsymbol{\Upsilon}^{\top}\boldsymbol{\Upsilon} &= \left[\begin{array}{cccc} -\beta_0\phi'(1-t_0) & \beta_0\phi'(2-t_0) & \cdots & \beta_0\phi'(N-t_0) \\ \hline \phi(1-t_0) & \phi(2-t_0) & \cdots & \phi(N-t_0) \end{array}\right] \left[\begin{array}{c|c} -\beta_0\phi'(1-t_0) & \phi(1-t_0) \\ -\beta_0\phi'(2-t_0) & \phi(2-t_0) \\ \vdots & \vdots \\ -\beta_0\phi'(N-t_0) & \phi(N-t_0) \end{array}\right] \\ &= \left[\begin{array}{cc} \beta_0^2 \sum\limits_{n=0}^{N-1} \left|\phi'(n-t_0)\right|^2 & -\sum\limits_{n=0}^{N-1} \beta_0\phi'(n-t_0)\,\phi(n-t_0) \\ -\sum\limits_{n=0}^{N-1} \beta_0\phi'(n-t_0)\,\phi(n-t_0) & \beta_0^2 \sum\limits_{n=0}^{N-1} \left|\phi(n-t_0)\right|^2 \end{array}\right]. \qquad (4.16) \end{aligned}$$

Given the choice of probing function,

$$\mathsf{C}_{p,\overline{p}}(t) \triangleq \phi(t) = \sum_{m=-M}^{m=+M} \phi_m e^{\jmath m \omega_0 t},$$

we will now show that,

$$\Upsilon^{\top}\Upsilon = \begin{bmatrix} N\omega_0^2\beta_0^2 \sum\limits_{|m|\leqslant M} m^2\phi_m\phi_m^* & \\ & N \sum\limits_{|m|\leqslant M} |\phi_m|^2 \end{bmatrix}.$$

- Derivation of $\left[\Upsilon^{\top}\Upsilon\right]_{0,0}$.

  Since, $\partial_t\phi\left(t\right) = \sum\limits_{m=-M}^{m=+M} \jmath m\omega_0\phi_m e^{\jmath m\omega_0 t}$ and $\partial_t\phi\left(t\right)|_{t=n-t_0} = \sum\limits_{m=-M}^{m=+M} \jmath m\omega_0\phi_m e^{\jmath m\omega_0(n-t_0)}$, it readily follows,

$$\begin{aligned}\sum_{n=0}^{N-1} \left|\phi'\left(n-t_0\right)\right|^2 &= \sum_{n=0}^{N-1}\sum_{m=-M}^{m=+M} \jmath m\omega_0\phi_m e^{\jmath m\omega_0(n-t_0)} \sum_{l=-M}^{l=+M} -\jmath l\omega_0\phi_l^* e^{-\jmath l\omega_0(n-t_0)} \\ &= \sum_{m=-M}^{m=+M}\sum_{l=-M}^{l=+M} ml\omega_0^2\phi_m\phi_l^* e^{\jmath m\frac{2\pi}{L}t_0(l-m)} \underbrace{\sum_{n=0}^{N-1} e^{\jmath\frac{2\pi}{L}n(m-l)}}_{=N\delta_{m-l},qL=N,q\in\mathbb{Z}} \\ &= N\omega_0^2 \sum_{|m|\leqslant M} m^2|\phi_m|^2.\end{aligned}$$

- Derivation of $\left[\Upsilon^{\top}\Upsilon\right]_{0,1} = \left[\Upsilon^{\top}\Upsilon\right]_{1,0} = 0$.

  Note that, $\left[\Upsilon^{\top}\Upsilon\right]_{0,1} = \left[\Upsilon^{\top}\Upsilon\right]_{1,0}$. This follows from (4.4.2). It remains to show,

$$\begin{aligned}\sum_{n=0}^{N-1} \phi'\left(n-t_0\right)\phi\left(n-t_0\right) &= \sum_{n=0}^{N-1}\sum_{m=-M}^{m=+M} \jmath m\omega_0\phi_m e^{\jmath m\omega_0(n-t_0)} \sum_{|l|\leqslant M} \phi_l e^{\jmath l\omega_0(n-t_0)} \\ &= \sum_{|m|\leqslant M}\sum_{|l|\leqslant M} \jmath m\omega_0\phi_m\phi_l e^{-\jmath\frac{2\pi}{L}t_0(l+m)} \underbrace{\sum_{n=0}^{N-1} e^{\jmath\frac{2\pi}{L}n(l+m)}}_{=N\delta_{m+l},qL=N,q\in\mathbb{Z}} \\ &= \jmath\omega_0 \sum_{|m|\leqslant M} m\phi_m^2 \\ &= \jmath\omega_0 \underbrace{\left(\sum_{-M\leqslant m<0} m\phi_m^2 + \sum_{0<m\leqslant M} m\phi_m^2\right)}_{=0,\phi(t)\in\mathbb{R}\Leftrightarrow|\phi_m|=|\phi_{-m}|} = 0\end{aligned}$$

where the last step is due to the fact that $\phi$ is a real function and its Fourier Series coefficients have the property, $|\phi_m| = |\phi_{-m}|$.

- Derivation of $\left[\Upsilon^{\top}\Upsilon\right]_{1,1} = N \sum_{|m|\leqslant M} |\phi_m|^2$.

  This is fairly simple to establish.

$$\begin{aligned}
\sum_{n=0}^{N-1} |\phi(n-t_0)|^2 &= \sum_{n=0}^{N-1} \phi(n-t_0)\,\phi^*(n-t_0) \\
&= \sum_{n=0}^{N-1} \sum_{|m|\leqslant M} \phi_m e^{\jmath m\omega_0(n-t_0)} \sum_{|l|\leqslant M} \phi_l^* e^{-\jmath l\omega_0(n-t_0)} \\
&= \sum_{|m|\leqslant M} \sum_{|l|\leqslant M} \phi_m \phi_l^* e^{\jmath \frac{2\pi}{L} t_0 (l-m)} \underbrace{\sum_{n=0}^{N-1} e^{\jmath \frac{2\pi}{L} n(m-l)}}_{=N\delta_{m-l},qL=N,q\in\mathbb{Z}} \\
&= N \sum_{|m|\leqslant M} |\phi_m|^2
\end{aligned}$$

Following the above simplifications, we are not set to bound the variance of the estimator. The inverse of the $2 \times 2$, diagonal matrix is given by,

$$\left(\Upsilon^{\top}\Upsilon\right)^{-1} = \begin{bmatrix} \left(N\omega_0^2\beta_0^2 \sum_{|m|\leqslant M} m^2|\phi_m|^2\right)^{-1} & \\ & \left(N \sum_{|m|\leqslant M} |\phi_m|^2\right)^{-1} \end{bmatrix} = \begin{bmatrix} \frac{S_m}{N\omega_0^2\beta_0^2} & \\ & \frac{S_1}{N} \end{bmatrix}$$

where $S_m = 1 \Big/ \sum_{|m|\leqslant M} m^2|\phi_m|^2$. Finally, we have,

$$\mathcal{V}\left(\widehat{\theta}\right) \geqslant \frac{1}{\sigma^2}\left(\Upsilon^T\Upsilon\right)^{-1} \Rightarrow \mathsf{CRLB}\left(\widehat{\theta}\right) = \begin{bmatrix} \frac{\sigma^2}{\beta_0^2}\frac{S_m}{N\omega_0^2} & \\ & \frac{\sigma^2}{N}S_1 \end{bmatrix}.$$

Let us define peak–signal–to–noise ratio by $\mathsf{PSNR} = \frac{\beta_0^2}{\sigma^2}$. As a result, the uncertainty in time and amplitude is given by,

$$\frac{\Delta t_0}{L} \geqslant \frac{1}{2\pi}\sqrt{\frac{S_m}{N}}\frac{1}{\sqrt{\mathsf{PSNR}}} \quad \text{(for time)},$$

and,

$$\frac{\Delta\beta_0}{|\beta_0|} \geqslant \sqrt{\frac{S_1}{N}} \frac{1}{\sqrt{\mathsf{PSNR}}} \quad \text{(for amplitude).}$$

These results are in consensus with the bounds in [53].

## 4.5 Conclusions

As a follow–up on the recent investigations regarding identification of operators from its response to a probing function [57, 58, 59, 60], we discuss a non–iterative method for the problem of sparse linear operator identification [60]. Our method is devoid of any sparsity inducing penalty term. Our work finds application in context of resolving mixed pixels in Time–of–Flight imaging for which our theoretical set corroborates with practical experiments. We report a speed up in computation time over previously used methods. Our method leads to lesser mean squared error.

# 5

# Multi–bounce Decomposition of Light with Microsoft Kinect Xbox One

## 5.1 Problem Statement

Multipath interference (MPI) is one of the major sources of both depth and amplitude measurement errors in Time–of–Flight (ToF) cameras. This problem has seen a lot of attention recently. In this work, we discuss the MPI problem within the framework spectral estimation theory and multi–frequency measurements. As compared to previous approaches that consider up to two interfering paths, our model considers the general case of $K$–interfering paths. In the theoretical setting, we show that for the case of $K$–interfering paths of light, $2K + 1$ frequency measurements suffice to recover the depth and amplitude values corresponding to each of the $K$ optical paths. What singles out our method is the that our

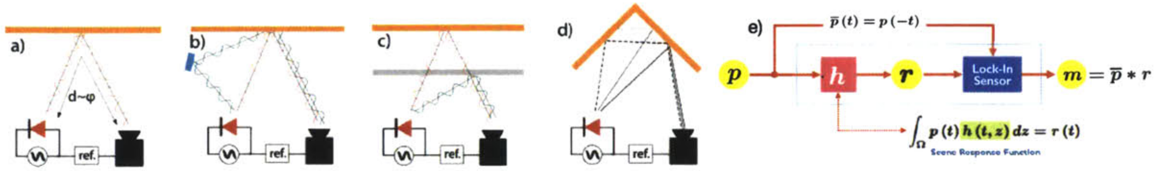


Figure 5-1: (a) The ToF camera emits a reference signal. Time delay of arrival from direct reflection encodes the depth $d$. (b) Specular or mirror like and, (c) semi–transparent reflections cause multiple light paths to mix at the sensor which leads to multipath interference or the mixed–pixel problem (MPP). (d) Case of continuous multipath reflections. (e) Time–of–flight imaging system pipeline for solving inverse problems.

algorithm is non–iterative in implementation. This leads to a closed–form solution which is computationally attractive. Also, for the first time, we demonstrate the effectiveness of our model on an off–the–shelf Microsoft Kinect for the X–Box one.

## 5.2 Introduction

Amplitude modulated continuous wave (AMCW) Time–of–flight (ToF) imaging cameras [22] measure at each pixel both amplitude and optical travel time (depth), thus capturing three dimensional scene information. Fig. 5-2(e) and 5-2(f) show an example of such intensity and depth images as produced by a ToF camera.

These cameras work on the principle of emitting a coded light signal (generally a sine wave) by amplitude modulating a light source and measuring the time delay between the transmission and the reflection arriving back from the scene (Fig. 5-1(a)), similar in principle to LIDAR.

With recent advances in solid state sensor design and on–chip computational capabilities, the last decade has seen a tremendous growth in 3D sensor design with a number of consumer grade ToF sensors are available in the market. Microsoft (Kinect), SoftKinetic, PMD and Mesa among others provide robust solutions for 3D imaging. Availability of ToF cameras has catalyzed research effort in context of 3D sensing with applications in human–computer–interaction (HCI) [27], computer graphics [31] and computational imaging [11, 81, 82]. In fact, in a recent paper we demonstrated the application of ToF cameras in Fluorescence Lifetime Estimation [5].

### 5.2.1 Multipath Interference in ToF Cameras

All of the existing ToF cameras work under the hypothesis that each given pixel observes one optical path. Another way to state this, the assumption is that the scene is only illuminated directly with no inter reflections (known in the optic and geophysics domains as the Born approximation). This however is not the case in many practical cases of interest such as inter reflections in the scene due to multiple objects or corners, in the presence of transparencies such as windows, or sub surface scattering. Some of these cases are presented in Fig. 5-1.

When multiple optical paths combine at a given pixel, the depth measurements are corrupted. This is known as the **multipath interference problem** or the **mixed pixel problem (MPI)**. This is one of the major sources of errors in ToF sensors [5, 83, 6, 33, 34, 84, 39, 62, 36, 38, 63]. To that end, almost all existing solutions consider up to two interfering paths of light [34, 84, 39]. The case of $K$–interfering paths was first discussed in [5] and later in [83]. Both of these papers rely on a sparsity formulation which leads to computationally intensive and iterative algorithms.

### 5.2.2 Contribution and Organization of this Chapter

In this work, we report a model for the case of $K$–path interference for which the solution is non–iterative in implementation which makes it computationally attractive when compared to the sparsity based, iterative schemes discussed in [5, 83].

In Section 5.3, we provide a general description of the ToF image formation model. Within our inverse–problem framework, we set–up the mathematical model for the multipath interference problem. As will be seen, our reformulation of this problem shows that the MPI is intrinsically linked with parametric spectral estimation theory [55, 45]. We leverage on previous ideas [6, 55] to solve this problem in closed–form. In Section 5.4, we discuss some first results linked with MPI cancellation in Microsoft Kinect One. Finally, we conclude with some future directions.

## 5.3 ToF Image Formation Model

The ToF imaging pipeline is shown in Fig. 5-1(e). Pixelwise image formation model for ToF cameras can be disseminated into 3 steps.

1. **Probe the Scene:** ToF cameras are active sensors. The ToF camera probes the scene with some $T$–periodic function of form $p(t) = p(t+T), t \in \mathbb{R}, T > 0$. For example, $p(t) = 1 + \cos(\omega_0 t), \omega_0 = 2\pi/T$.

2. **Scene Interaction** Next, the probing function $p$ interacts with the scene $h$ and this interaction results in the reflected signal,

$$r(t) = \underbrace{\int_{\Omega} p(z)\, h(t,z)\, dz}_{\textsf{Fredholm Integral}} \qquad \text{(Reflected Signal)}, \tag{5.1}$$

   where $h$ is the continuously defined kernel or the scene response.

3. **Cross–Correlation** Let $\overline{p}(t) \overset{\text{def}}{=} p(t)$ and let $*$ denote the convolution operation, that is, $(f * g)(t) = \int f(z) g(t-z) dz$. For some sampling step $\Delta$, the ToF lock–in sensor filters the reflected signal $r$ with $\bar{p}$ and then descritizes it. Quantitatively, we have,

$$m(t)\big|_{t=k\Delta} = (\overline{p} * r)(t) \quad \text{with} \quad k \in \mathbb{Z}, \Delta > 0. \tag{5.2}$$

**Remark:** Note that for a special class of scene response functions/kernels, namely, Shift–invariant kernels, which satisfy the property, $h_{\mathsf{SI}}(t,z) = h(t-z)$, the measurements (5.2) assume form of,

$$m(t) = (\overline{p} * p * h)(t) = (\mathsf{C}_{p,p} * h)(t), \tag{5.3}$$

where $\mathsf{C}_{f,g}(t) = (\overline{f}^* * g)(t)$ is the cross–correlation of complex functions $f, g \in \mathbb{C}$ and where $f^*$ is the complex conjugate of $f$.

### 5.3.1 Modelling Single Depth

Within the context of our imaging model, we first show how to recover amplitude and depth from discrete measurements $m$. Most ToF cameras including the Kinect One probe the scene with,

$$p(t) = 1 + p_0 \cos(\omega t) \quad \text{and} \quad 0 < p_0 < 1.$$

At some depth $d$, the probing function undergoes a reflection. The total time–of–flight is, $t_d = 2d/c \quad \text{where} \quad c = 3 \times 10^8$. The reflected signal is modelled as,

$$r(t) = \Gamma\left(s\left(t - t_d\right)\right) \quad \text{where} \quad 0 < \Gamma < 1.$$

An equivalent conclusion can be reached by setting, $h(t, z) = \Gamma\delta(t - z - t_d)$, where we use Dirac's Delta function, that is $\delta$, to model the time–delay or the time spent by the probing function $p$ as it arrives at the ToF sensor. In this case, a simple computation [5, 22] shows that,

$$m_\omega(t) = \Gamma\left(1 + \tfrac{p_0^2}{2}\cos\left(\omega\left(t + t_d\right)\right)\right)$$

and it remains to estimate $\{\Gamma, d\}$ from measurements $m$ in (5.2). The ToF camera uses sampling step $\Delta = \pi/2\omega$, $k = 0 \cdots 3$ to estimate,

$$\begin{aligned} \widetilde{\Gamma} &= p_0^{-2}\sqrt{\left(m_\omega(3\Delta) - m_\omega(\Delta)\right)^2 + \left(m_\omega(0) - m_\omega(2\Delta)\right)^2} \quad \text{and,} \\ \widetilde{d} &= \frac{c}{2\omega}\tan^{-1}\left(\tfrac{m_\omega(3\Delta) - m_\omega(\Delta)}{m_\omega(0) - m_\omega(2\Delta)}\right). \end{aligned} \tag{5.4}$$

Note that for any given modulation frequency $\omega$, and estimated parameters $\{\Gamma, d\}$, we can associate a complex number/phasor,

$$Z_\omega = \widetilde{\Gamma}\exp\left(\jmath\left(2\widetilde{d}\omega\right)/c\right) \triangleq \widetilde{\Gamma}\exp\left(\jmath\omega t_d\right), \quad Z_\omega \in \mathbb{C}. \tag{5.5}$$

An attentive reader must have noticed that that $Z_\omega$ happens to be the Fourier Transform of $\Gamma\delta(t - t_d)$. Indeed, in view of the ToF pipeline in Fig. 5-1(e), single–depth estimation problem is linked with the estimation of scene response function, $h(t) = \Gamma\delta(t - t_d)$, which is a parametric function with 2 unknowns, reflection coefficient $\Gamma$ and depth $d$ which is encoded in variable $t_d$. This is one instantiation where $h$ or the scene response function is parametric. Another interesting parametric choice of $h$ discussed in [8] is linked with

Fluorescence–lifetime imaging microscopy (FLIM) where,

$$h(t) = \delta(t - t_d) * (\Gamma\delta(t) + \rho \exp(-t/\tau)), \quad t \geq 0. \tag{5.6}$$

We now revert our attention to the case of multipath interference.

### 5.3.2 Multipath Interference Problem

Consider the case discussed in Fig. 5-1(b–c). When multiple bounces of light or optical paths arrive at the sensor, the depth measurements are corrupted [6, 5, 32, 39]. As shown in [6], the system response function is modelled as,

$$h_K(t) = \sum\nolimits_{k=0}^{K-1} \Gamma_k \delta\left(t - \frac{2d_k}{c}\right) \equiv \sum\nolimits_{k=0}^{K-1} \Gamma_k \delta(t - t_k). \tag{5.7}$$

Let us denote the Fourier Transform of $h(t)$ by $\widehat{h}_K(\omega)$ with,

$$h_K(t) \xrightarrow{\text{Fourier}} \widehat{h}_K(\omega) = \sum\nolimits_{k=0}^{K-1} \Gamma_k e^{-\jmath\omega t_k} \tag{5.8}$$

Indeed, for the multipath case, the Fourier domain representation of the system response function in (5.7) is a sum of $K$–complex exponential functions. Without loss of generality, if we assume, $p(t) = p_0 e^{\jmath\omega t}$, then,

$$m_\omega(t) = C_0 e^{\jmath\omega t} \widehat{h}_K^*(\omega) \quad \text{where} \quad C_0 = p_0^2/2. \tag{5.9}$$

Note that the multipath component $\widehat{h}_K^*(\omega)$ appears in the measured signal and is purely an argument of modulation frequency, $\omega$. For $\omega = \omega_0$, the ToF camera uses (5.4) to record the measurement, $Z_{\omega_0} = \widehat{h}_K^*(\omega_0)$. Now since, $\widetilde{\Gamma}_\omega = |Z_{\omega_0}|$ and $\widetilde{\phi}_\omega = \angle Z_{\omega_0}$, or equivalently,

$$\widetilde{\Gamma}_\omega = \left|\sum\nolimits_{k=0}^{K-1} \Gamma_k e^{\jmath\omega t_k}\right| \text{ and } \widetilde{\phi}_\omega = \angle\left(\sum\nolimits_{k=0}^{K-1} \Gamma_k e^{\jmath\omega t_k}\right), \tag{5.10}$$

the camera is unable to resolve the interference corrupted measurements in (5.10).

### 5.3.3 Closed Form Solution for Retrieving $\{\Gamma_k, d_k\}_{k=0}^{K-1}$

We start with measurements $m_\omega(t)$. Since the multipath component is independent of $t$, we will only consider measurements in $\omega$. Let us define the vector of measurements, $\mathbf{m} = [m_0, m_1, \ldots, m_{N-1}]^\top$ where $(\cdot)^\top$ is the matrix–transpose operation.

**Problem Statement:** Given a vector of measurements $\mathbf{m}$ for $N$ modulation frequencies, extract $\{\Gamma_k, d_k\}_{k=0}^{K-1}$.

This is a classic inverse problem in line spectrum estimation [45] and is a special case of the solution discussed in context of Fractional Fourier transforms [55]. To solve this inverse problem, we begin with $N$ measurements, $m_n = \sum_{k=0}^{K-1} \Gamma_k u_k^n, n \leqslant N$ with $u_k = \exp(\jmath t_k)$.

Let us define a Laurent polynomial such that its roots are $u_k$,

$$\mathsf{Q}(z) = \prod_{\ell=0}^{K-1} \left(1 - u_k z^{-1}\right) \equiv \underbrace{\sum_{\ell=0}^{K} q_\ell z^{-\ell}}_{\text{Factorized Polynomial}} \quad , \quad z \in \mathbb{Z}. \tag{5.11}$$

We leverage from an interesting factorization property of complex–exponentials which is $q_\ell * m_\ell = m_\ell \mathsf{Q}(u_k) = 0$,

$$q_\ell * m_\ell = \sum_{n=0}^{K} q_n m_{\ell-n} = \underbrace{\sum_{k=0}^{K-1} \Gamma_k u_k^\ell}_{m_\ell} \underbrace{\sum_{n=0}^{K} q_n u_k^{-n}}_{\mathsf{Q}(u_k)=0} = 0.$$

This is because $\mathsf{Q}(z)|_{z=u_k} = 0$ (see (5.11)). Hence the sequence $q_\ell$ that filters measurements $m_\ell$ to zeros also encodes the information of $u_k$ in form of the roots of the polynomial $\mathsf{Q}$. In vector–matrix, this is equivalent to,

$$\underbrace{\begin{bmatrix} m_K & \cdots & m_0 \\ \vdots & \ddots & \vdots \\ m_{2K} & \ldots & m_K \end{bmatrix}}_{\text{Toeplitz Matrix} \quad \mathbf{M}^{(K+1)\times(K+1)}} \underbrace{\begin{bmatrix} q_0 \\ \vdots \\ q_K \end{bmatrix}}_{\mathbf{q}^{K+1}} = \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}^{K+1}} \Leftrightarrow \mathbf{Mq} = \mathbf{0},$$

where $\mathbf{M}$ is identified as the convolution/Toeplitz matrix and vector $\mathbf{q} \in \text{Null}(\mathbf{M})$ is a vector in the null–space of rank–deficit matrix $\mathbf{M}$ with rank $= K$. We need at least $N = 2K + 1$ measurements to construct $\mathbf{M}$ and hence, $2K + 1$ modulated frequencies in $m_\omega$ to recover

$K$ optically interfering paths. Once we compute $\mathbf{q}$, we factorize it to compute the roots $\{u_k\}$ in (5.11). This, in turn leads to,

$$\widetilde{d}_k = (2\jmath)^{-1} c \log(u_k) \qquad \text{where} \qquad c = 3 \times 10^8.$$

Let us define a Vandermonde matrix $\mathbf{U}^{N \times K}$ with matrix elements,

$$[U]_{n,k} = u_k^n, \quad k = 0, \ldots, K-1, \quad n = 0, \ldots, N-1.$$

To compute $\{\Gamma\}_k$, we solve the system of equations $\mathbf{U\Gamma} = \mathbf{m} \Rightarrow \mathbf{\Gamma} = \mathbf{U}^{+}\mathbf{m}$ where $\mathbf{U}^{+}$ is pseudo–inverse of $\mathbf{U}$.

### 5.3.4 Remarks and Discussion

- **Improvement over previous work.** To begin with, our work generalizes the results in [39, 84] in that our theoretical results are applicable to general setting of $K$–interfering paths. For example, setting $K = 2$ leads to $N = 5$ frequency measurements which is the result discussed in [84]. Also, unlike [84] that requires measurements both in frequency and phase, our formulation does not require phase sampling.

  The idea of using multiple modulation frequencies to resolve *multiple interfering paths* ($K > 2$) in context of TOF was first presented in [5] however, the choice greedy algorithm for sparse optimization makes the solution inefficient. In this work, we show that it is possible to retrieve $\{\Gamma_k, d_k\}_{k=0}^{K-1}$ from $m_\omega$ by using a closed–form solution.

- **Reconstruction Guarantee.** In absence of model mismatch and noise, we show that $N = 2K+1$ frequency measurements are necessary for de–mixing $K$–interfering paths. To the best of our knowledge, this result has not been reported in literature (in context of ToF imaging). Also, for our setting, Cramér–Rao bounds [45] may be used to provide reconstruction guarantees.

- **Implementation Details.** As is well known, the Vandermonde matrix $\mathbf{U}^{+}$ is highly unstable in presence of noise and oversampling is an efficient counter–measure. Many variants have been proposed in literature to alleviate this problem. For implementation purposes, we use Matrix Pencils [76] to solve to problem.

## 5.4 Results

For experiments we use a single, off the shelf, Microsoft "Kinect One" ToF camera with modified firmware that allows custom modulation frequencies. The Kinect is at a distance of about 1.5 meters from a scene comprising of a mannequin on a table with a wall in the background. We further place a transparent acrylic sheet about 15 cm in front of the camera. The acrylic reflects part of the emitted infrared light directly back to the camera, resulting in mixed measurements (also known as multipath effect). We proceed with recording of the amplitude/phase measurements of the Kinect at varying modulation frequencies starting from 50 MHz.

The experimental setup is shown in Fig. 5-2(j). In this example, we set $K = 2$. The amplitude and phase measurements at 52 MHz are shown in Fig. 5-2(a) and Fig. 5-2(b), respectively. We use 20 measurements from $52 - 72$ MHz for our processing. The out put of our algorithm provides $2K$ images, in form of amplitude and phase pair. For the experiment in Fig. 5-2(j), our goal is to decompose the mixture of images into 2 components corresponding to two optical paths, one due to the acrylic sheet and another due to the what remains in the background. Fig. 5-2(c)–(d) shows the amplitude/depth (in meters) pair corresponding to the contribution due to acrylic sheet. As can be seen in Fig. 5-2(c), we observe a specular reflection due to the reflective acrylic sheet. In Fig. 5-2(e)–(f), we show the remaining contribution due to all other optical paths. The depth of the table is estimate to be 1.67 meters. The amplitude image in Fig. 5-2(e) is devoid of any multipath effects due to the acrylic sheet.

To demonstrate the effectivity of our algorithm, we show the 3-dimensional or 3D reconstruction in Fig. 5-2(g),(h),(i). We show the 3D amplitude/depth map of measurements at 52 MHz and 55MHz in Fig. 5-2(g) and Fig. 5-2(h), respectively. Clearly, the measurements are distorted. We are able to decompose the corrupted images in Fig. 5-2(g),(h). Our processed 3D image is shown in Fig. 5-2(i).

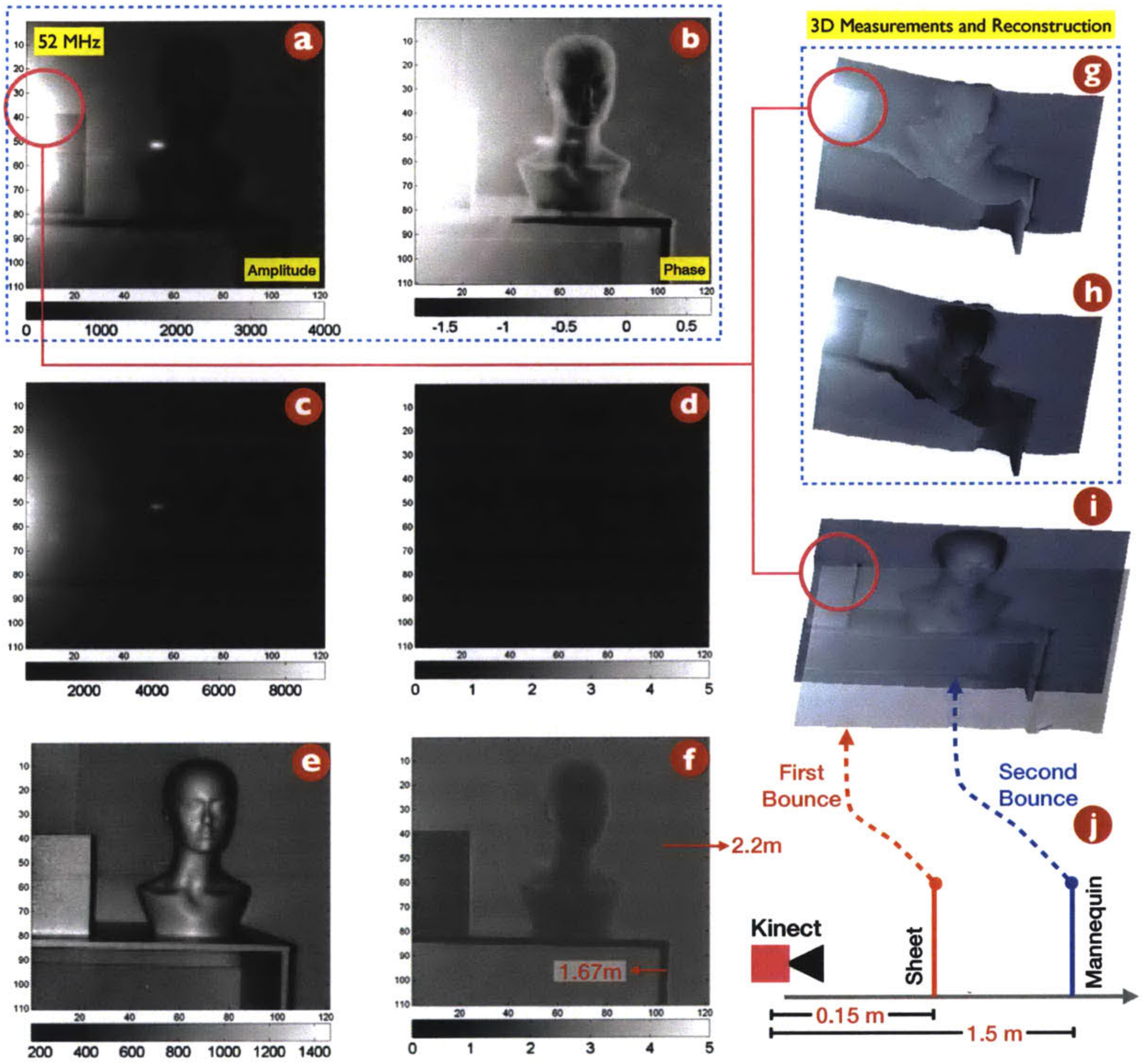


Figure 5-2: Images (a) and (b) show one of the input amplitude and phase images while (g) and (h) show the embedding of the intensity image onto the depth data. The scene, depicted in (j) is constructed from a mannequin on a table with a transparency between the camera and the scene. As can be seen, there is a strong flashback from the transparency as well as corruption of the depth image due to multi bounce. (c) and (d) show the intensity and depth reconstruction of the transparency while (e) and (f) show that of the scene. (i) Shows the embedding of the intensity image onto the depth reconstruction. Not only have the intensity and depth information been recovered correctly for both the scene and transparency, we have also recovered both amplitude and depth information that was completely occluded by the flashback (highlighted).

## 5.5 Conclusion

Our theoretical development is general in that it considers the general case of $K$−optical paths and is computationally efficient in that the algorithm is non-iterative (compared to the recent work). Our first results demonstrate the practicability of our modelling with the ease of computation.

# 6

# Fluorescence Lifetime Estimation via Coded Time–of–Flight Cameras

We present a novel, single-shot, calibration-free, framework within which Time-of-Flight/depth cameras can be used to estimate lifetimes of fluorescent samples. Our technique relaxes the high time resolution or the multi-frequency measurement requirements in conventional systems.

## 6.1 Overview of Problem Statement and Contributions

Fluorescence lifetime imaging (FLI) is a well–known problem with the notable example of Fluorescence lifetime imaging microscopy (FLIM). FLI methods rely on either time domain (or pulsed) processing (Fig. 6-1.c ) or frequency domain processing which is a phase based method (Fig. 6-1.d). Each method has its advantages and disadvantages.

This work combines the merits of time and frequency based processing. We describe a technique for FLI which is **cost–effective** in that it is based on consumer grade Time–of–Flight (ToF) cameras. ToF cameras can not match up with the time resolution of de-facto equipment used for time-domain FLIM. Also, multiple frequency measurements require sweeping of frequencies over a given bandwidth. To overcome these limitations, we prescribe a **time coded illumination** based technique which results in **single–shot, multi–frequency capture**. Also, our technique is **reference free**. No prior information of the depth of sample with respect to the camera is assumed.

## 6.2 Inverse Problem Framework for FLI–Time-of-Flight Imaging

A Time-of-flight (ToF) camera, such as the Microsoft Kinect, is a low-cost, real-time instrument that acquires the three-dimensional scene structure by encoding simultaneously the intensity and depth information of a scene. Such cameras have been very successful in consumer applications.

We provide a general model of ToF sensing using the Fredholm formulation. First, generalizing conventional ToF cameras, which use (approximately) sinusoidal illumination, we illuminate a scene with a time-periodic probing signal $p(t) = p(t+T)$ ($T \in \mathbb{R}^+$, $t \in \mathbb{R}$). This signal interacts with a scene, described by a response kernel $h(t,z)$, and generates a return signal $y(t)$, which is then cross-correlated with the illumination to yield measurements $m(t)$. Defining $\overline{p} = p(-t)$ and letting $\mathsf{C}_{y,\overline{p}}(t) = (y * \overline{p})(t)$ be the cross-correlation of two functions, we have

$$p \to \underbrace{\boxed{h}}_{\text{Scene}} \xrightarrow{\text{(a)}} \underbrace{y(t) = \overbrace{\int_{\Omega} p(z)\, h(t,z)\, dz}^{\text{Fredholm Integral}}}_{\text{Interaction of the scene and probing function}} \xrightarrow{\text{(b)}} \underbrace{\boxed{\overline{p} * y}}_{\text{Sensor}} \xrightarrow{\text{(c)}} \underbrace{m(t)}_{\text{Measurements}} \quad \text{(ToF Pipeline)}. \tag{6.1}$$

For a shift-invariant operators, $\Omega \in \mathbb{R}$, $h(t,z) \to h(t-z)$, so that $m(t) = \mathsf{C}_{y,\bar{p}}(t) = (p * \bar{p} * h)(t)$. This assumption is not restrictive, as it applies to many systems, including FLI and depth imaging. For example, for conventional depth sensing, $p(t) = 1 + cos(\omega_0 t)$, and $h(t - z) = \rho_0 \delta(t - 2d/c - z)$, where $\rho_0$ is the reflection coefficient, $c$ is the speed of light, and $d$ is the distance to the scene. For multiple depths $\{d_k\}_{k=0}^{K-1}$, $h(t,z) = \sum_{k=0}^{K-1} \rho_k \delta(t - 2d_k/c - z)$ [5, 61].

## 6.3 Forward Model for FLI using ToF Sensors

For scenes containing fluorescent samples, the system kernel becomes more complex, as fluorescence emmission decays exponentially for impulsive excitation. To describe this system, we apply Eq. 6.1 to a scene containing fluorescent pixels $(x_f, y_f)$ and non-fluorescent background $(x_b, y_b)$. For the latter pixel type, the kernel is just the standard delta-function depth kernel above. We can describe this using operator notation by defining the time delay operator $\mathcal{T}_d[\phi] = \phi(t - 2d/c)$. Defining $\mathcal{I}$ as the identity operator and denoting $\mathcal{P} \circ \mathcal{Q} = \mathcal{P}(\mathcal{Q}[\cdot])$ for composition, we have

$$m_{(x_b,y_b)}(t) \overset{(6.1)}{=} (\mathcal{T}_d \circ \rho\mathcal{I})[p] * \bar{p} \Leftrightarrow (\mathcal{T}_d \circ \rho\mathcal{I})[\phi], \qquad \mathsf{C}_{p,\bar{p}} = \phi. \tag{6.2}$$

Note here that we can write the depth kernel as

$$h(t,z) = (\mathcal{T}_d \circ \rho\mathcal{I})[\delta(t-z)] \Leftrightarrow \rho\delta(t - 2d/c - z).$$

To model fluorescence (Fig. 6-1a), we recognize that the associated kernel is $h_f(t-z) = \mu \exp(t-z)/\tau \cdot \mathbf{1}_{(t-z)\geqslant 0}(t-z)$, which we associate with a well-defined Green's function $\mathsf{G}(t,z) = h_f(t-z)$. Therefore, we may model this interaction with a linear, constant-coefficient, first order operator $\mathcal{L}_\tau = \partial_t + 1/\tau$ acting on the illumination function $p(t)$ to yield the fluorescence output $u(t)$. We have

$$u(t) = \mu\mathcal{L}_\tau^{-1}[p \cdot 1_{t\geqslant 0}] \Leftrightarrow \int_{t-z\geqslant 0} \mathsf{G}(t,z)\, p(z)\, dz$$

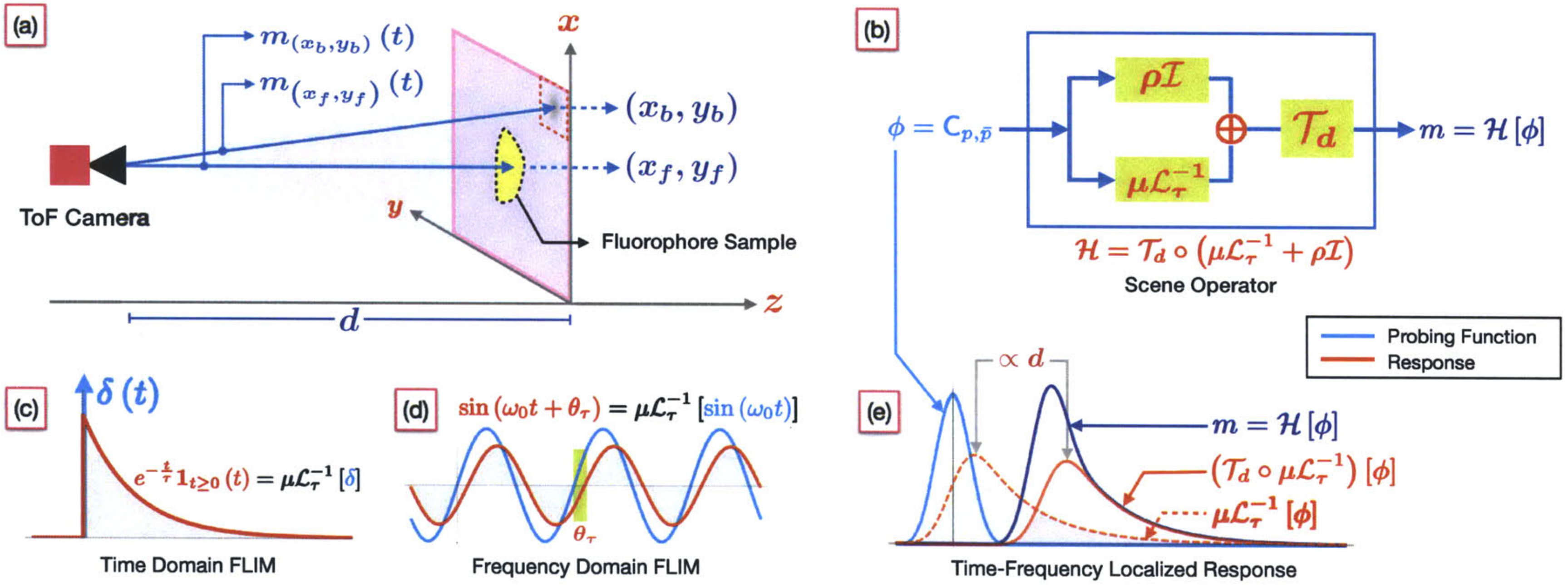


Figure 6-1: (a) Setup for fluorescence lifetime estimation using ToF cameras. Pixels at $(x_b, y_b)$ measure depth by using time–of–flight principle in accordance with (6.2). Pixels $(x_f, y_f)$ correspond to fluorescent sample. Impinging light signal $p$ undergoes two transformations: first, amplitude scaling and delay $((\mathcal{T}_d \circ \rho\mathcal{I})[p])$ due to the depth $d$ and second, due to its interaction with the fluorophore sample which is in effect $y = \mathcal{T}_d \circ \mu\mathcal{L}_\tau^{-1}[p]$. (b) Mathematical notation for the physical set up described in part (a). The system response in context of (6.1) is given by $\mathcal{H} = \mathcal{T}_d \circ (\mu\mathcal{L}_\tau^{-1} + \rho\mathcal{I})$ and the measurements read, $m(t) = \mathcal{H}[\phi], \phi = \mathsf{C}_{p,\bar{p}}$. (c) Time–domain FLIM is modeled by $m(t) = \mathcal{H}[\phi]$ with $\phi = p = \delta$ and $d = 0$. (d) Frequency–domain FLIM is modeled by $m(t) = \mathcal{H}[\phi]$ with $p = \delta$ and some calibrated $d$. (e) ToF–FLIM uses $m(t) = \mathcal{H}[\phi]$ where $\phi$ is some band–limited function with utmost $M_0$ Fourier components. We show $\mu\mathcal{L}_\tau^{-1} \rightarrow \mathcal{T}_d \circ \mu\mathcal{L}_\tau^{-1} \rightarrow \mathcal{H}$—the sequence of transformations on $\phi$.

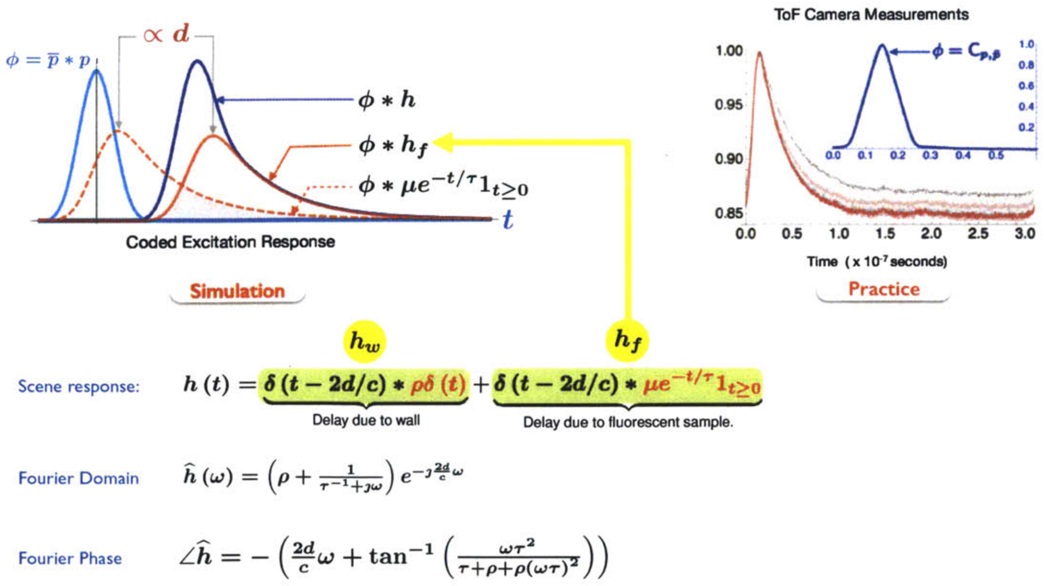


Figure 6-2: Overview of forward model for coded FLIM imaging.

with,

$$\mathsf{G}(t, z) = \mu \exp\left(-\left(t - z\right)/\tau\right) \mathbf{1}_{t \geqslant z}\left(t - z\right).$$

In the context of Eq. (6.1), the combined effect of scaling/delay from the depth and fluorescence decay leads to,

$$y = \mathcal{T}_d \circ \left(\rho \mathcal{I} + \mu \mathcal{L}_\tau^{-1}\right)[p] = \mathcal{H}[p]. \tag{6.3}$$

In Fig. 6-1(b) we show the schematic for the forward model. Simulated signal and measurements are shown in Fig. 6-2. With $h(t, z) = h_{\mathsf{SI}}(t - z)$, the system response is identified as,

$$h_{\mathsf{SI}}(t) = \delta\left(t - \frac{2d}{c}\right) * \left(\rho\delta(t) + \mu e^{-(t)/\tau} \mathbf{1}_{t \geqslant 0}(t)\right) \tag{6.4}$$

The measurement becomes $m(t) = \mathcal{H}[\phi], \phi = \mathsf{C}_{p,\bar{p}}$. Our model is backward compatible with the FLIM setup [85] in that time domain FLIM uses $p(t) = \delta(t)$ (Fig. 6-1.c). Its frequency domain counterpart in Fig. 6-1d uses $p(t) = \sin(\omega t)$, and calibration is used for $d$.

## 6.4 Coded Illumination for FLI

Continuous-wave ToF cameras cannot attain the time resolution of standard pulse-based methods. To overcome this, we engineer the $T$-periodic signal, $p(t)$, within the window of length $T$. Because $p$ is periodic, it is expressible as a Fourier Series,

$$p\left(t\right)=T^{-1}\sum\nolimits_{|m|\leqslant M_0}\widehat{p}_m \exp\left(\jmath\omega_0 mt\right),$$

where $\hat{p}_m$ are the Fourier Series coefficients, with $\omega_0 T=2\pi$, and $M_0$ is the bandwidth. We thus write,

$$\mathsf{C}_{p,\bar{p}}\left(t\right)=\phi\left(t\right)=T^{-1}\sum\nolimits_{|m|\leqslant M_0}\widehat{\phi}_m \exp\left(\jmath\omega_0 mt\right),$$

with $\forall m, \widehat{\phi}_m=\widehat{p}_m\widehat{p}_m^*$. It is well-known that for a given budget of $M_0$, a Maximal Length Sequence (M-sequence) provides the best spectral response for $\phi$. Because complex exponentials are eigenfunctions of time-invariant operators such as $\mathcal{H}$, the ToF measurements for FLI are given by $m=\mathcal{H}\left[\phi\right]$, which simplifies to,

$$m\left(t\right)=T^{-1}\sum\nolimits_{|m|\leqslant M_0}\widehat{\phi}_m\widehat{h}_m e^{\jmath\omega_0 mt}, \tag{6.5}$$

where

$$\widehat{h}_m=\left(\rho+\frac{1}{1/\tau+\jmath\omega_0 m}\right)e^{-\jmath\omega_0\left(\frac{2d}{c}\right)m} \quad \text{(Fourier Transform of }\mathcal{H}\text{)}.$$

For discrete camera measurements, we have, in vector–matrix notation, $\mathbf{m}=\mathbf{V}\mathbf{w}$ where $\mathbf{m}\in\mathbb{R}^N$ is the measurement vector, $\mathbf{V}\in\mathbb{C}^{N\times 2M_0+1}$ is the inverse DFT matrix and $\mathbf{w}\in\mathbb{C}^{2M_0+1}$ is the vector such that $\mathbf{w}=\mathbf{D}_{\hat{\phi}}\hat{\mathbf{h}}$ and $\mathbf{D}_{\hat{\phi}}\in\mathbb{C}^{2M_0+1\times 2M_0+1}$ is a diagonal matrix with entries $\hat{\phi}_k$. Given measurements $\mathbf{m}$ and known $\hat{\phi}_k$, our task is to estimate parameters $\tau$ and $d$. We first estimate $\hat{\mathbf{h}}$ using $\hat{\mathbf{h}}=\mathbf{D}_{1/\hat{\phi}}\mathbf{V}^{+}\mathbf{m}$ where $\mathbf{V}^{+}$ is the pseudo–inverse. From Eq. (6.5), we note that

$$\angle\widehat{\mathbf{h}}\left(\omega_0 m\right)=-\left(\varphi_\tau\left(\omega_0 m\right)+\varphi_d\left(\omega_0 m\right)\right),$$

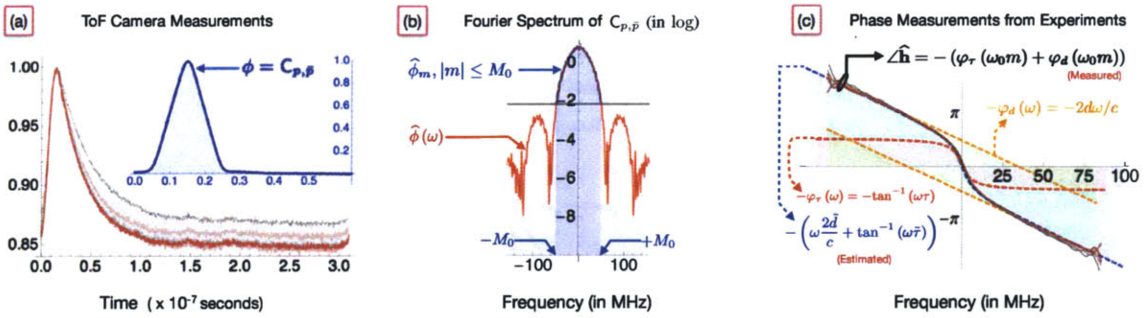


Figure 6-3: (a) Measurements for a quantum dot with $\tau = 32$ ns acquired by a ToF camera that uses a time code $\phi$ (in the inset). (b) Fourier Transform of $\phi$ and its band-limited approximation via Fourier Series consisting of $M_0 = 51$ components.(c) We estimate lifetime and depth from measurements using $\angle\widehat{\mathbf{h}} = \angle\mathbf{D}_{\widehat{\phi}}^{-1}\mathbf{V}^{+}\mathbf{m} \approx -\left(\varphi_\tau(m\omega_0) + \varphi_d(m\omega_0)\right)$. Curves in —— show measurements and the expected phase function is marked with - - - .

where $\varphi_\tau(\omega) = \tan^{-1}\left(\frac{\omega\tau^2}{\tau+\rho+\rho\omega^2\tau^2}\right)$ is nonlinear in lifetime ($\tau$), and $\varphi_d(\omega) = 2d\omega/c$ is linear depth ($d$). Experimentally, we optically filter out the contribution due to $\rho$, so that with $\rho = 0$, we are left to minimize the least-squares cost function $\mathrm{J}(\omega)$ that leads to the depth and lifetime estimates $\widetilde{d}$ and $\widetilde{\tau}$, respectively, as $(\widetilde{d}, \widetilde{\tau}) = \arg\min_{d,\tau} \ \mathrm{J}(\omega)$, where $\mathrm{J}(\omega) = ||\angle\hat{\mathbf{h}} + \omega\frac{2d}{c} + \tan^{-1}(\omega\tau)||_{\ell_2}^2$. Being able to extract $d$ without its prior knowledge makes our approach calibration free.

## 6.5 Experimental Verification

The experimental setup is depicted in Fig. 6-1.a. The fluorescent pixels consist of a glass slide that contains CdSe-CdS quantum dot material ($\tau = 32$ ns) obtained by dissolving it in hexane. The ToF camera is a PMD 19K-2 lock-in sensor with custom FPGA programming to define and precompute $p$ and hence $\phi$. The normalized measurements corresponding to the quantum dot sample are shown in Fig. 6-3a. The data are acquired with sampling rate $T = 78.12$ ps. We also plot the function $\phi$ in the inset. For data processing, we set $M_0 = 31$ (see Fig. 6-3b). From data samples $\mathbf{m}$, we compute

$$\angle\widehat{\mathbf{h}} = \angle\mathbf{D}_{\widehat{\phi}}^{-1}\mathbf{V}^{+}\mathbf{m} \approx -\left(\varphi_\tau(m\omega_0) + \varphi_d(m\omega_0)\right), \omega_0 = 2\pi/T.$$

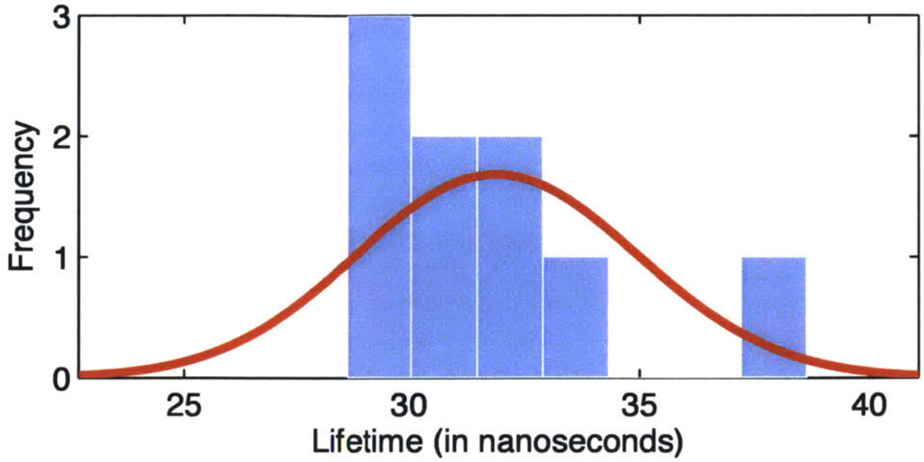


Figure 6-4: Histogram of estimated lifetimes.

The phase measurements are shown in Fig. 6-3c. Note that measurements corresponding to $m > 50$ are noisy. Having obtained $\angle\widehat{\mathbf{h}}$, we use a nonlinear fitting algorithm using the least absolute residual criterion. A histogram of estimated lifetimes is plotted in Fig. 6-4. The expected lifetime for our computation is $\underline{31.89 \times 10^{-9}}$ sec.

In summary, we have implemented a calibration-free, single-shot FLI system using a ToF camera. The method combines a low-cost implementation with a generalized model of ToF sensing. Future work includes extending the technique for FLIM systems for, e.g., FRET analysis, multi-exponential decays, and diffusive quenching.

# 7

# Conclusion

In this work we formalize the ToF imaging model which unifies continuous wave and impulse based ToF solutions. Within this framework, we tackle inverse problems linked with ToF imaging. Unlike conventional cameras, the ToF cameras encode both amplitude and depth. The main theme of this work is that the association of depth/phase with range is just a specific case of modeling. In a general setting, more interesting phenomenon can be encoded in the phase component of the ToF cameras.

The first direction attempts to solve the multi–path problem for the general case of $K$ light paths. For this purpose a closed–form solution is presented. The solution is linked with the spectral estimation problem. In noiseless setting, $2K$ modulated frequency measurements suffice to recover the parameters of the multi–path scene response function.

The second line of attack deals with super–resolution ToF imaging. It is shown that coded ToF cameras facilitate multi–frequency measurements. Furthermore, we leverage from the fact that the broadband code has a finite Fourier series. This leads to a parametric solution to the sparse deconvolution problem.

Finally, we investigate the problem of fluorescence lifetime estimation using ToF cameras. In this context, we show that by using a broadband code, it is possible to recover both the distance of the sample from the camera as well as its lifetime. Hence our solution leads to a calibration-free, single-shot FLI system.

# Bibliography

[1] A. Bhandari and A. I. Zayed. Shift–invariant and sampling spaces associated with the fractional Fourier domain. *IEEE Trans. Signal Proc.*, 60(4):1627–1637, 2012.

[2] M. Unser. Sampling–50 years after Shannon. *Proc. IEEE*, 88(4):569–587, 2000.

[3] David C. Rife and Robert R. Boorstyn. Single tone parameter estimation from discrete-time observations. *IEEE Transactions on Information Theory*, 20(5):591–598, 1974.

[4] Steve Tretter. Estimating the frequency of a noisy sinusoid by linear regression (corresp.). *Information Theory, IEEE Transactions on*, 31(6):832–835, 1985.

[5] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar. Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization. *Optics Letters*, 39(7), 2014.

[6] Ayush Bhandari, Achuta Kadambi, and Ramesh Raskar. Sparse linear operator identification without sparse regularization? Applications to mixed pixel problem in time-of-flight / range imaging. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pages 365–369, 2014.

[7] P. Stoica and R.L. Moses. *Introduction to spectral analysis*, volume 89. Prentice Hall Upper Saddle River, NJ, 1997.

[8] Ayush Bhandari, Christopher Barsi, Refael Whyte, Achuta Kadambi, Anshuman J Das, Adrian Dorrington, and Ramesh Raskar. Coded time-of-flight imaging for calibration free fluorescence lifetime estimation. In *Imaging Systems and Applications*, pages IW2C–5. Optical Society of America, 2014.

[9] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: Capturing and visualizing the propagation of light. *ACM Transactions on Graphics (TOG)*, 32(4):44, 2013.

[10] Di Wu, Andreas Velten, Matthew O?Toole, Belen Masia, Amit Agrawal, Qionghai Dai, and Ramesh Raskar. Decomposing global light transport using time of flight imaging. *International Journal of Computer Vision*, 107(2):123–138, 2014.

[11] Felix Heide, Matthias B. Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Trans. Graph. (Proc. SIGGRAPH 2013)*, 32(4):45:1–45:10, 2013.

[12] James A Simmons and Roger A Stein. Acoustic imaging in bat sonar: echolocation signals and the evolution of echolocation. *Journal of Comparative Physiology*, 135(1):61–84, 1980.

[13] Björn M Siemers and Hans-Ulrich Schnitzler. Echolocation signals reflect niche differentiation in five sympatric congeneric bat species. *Nature*, 429(6992):657–661, 2004.

[14] Lasse Jakobsen, John M Ratcliffe, and Annemarie Surlykke. Convergent acoustic field of view in echolocating bats. *Nature*, 2012.

[15] Lawrence D Rosenblum, Michael S Gordon, and Luis Jarquin. Echolocating distance by moving and stationary listeners. *Ecological Psychology*, 12(3):181–206, 2000.

[16] M. Romer and I. Bernard Cohen. Roemer and the first determination of the velocity of light (1676). *Isis*, pages 327–379, 1940.

[17] Patricia Daukantas. Ole Rømer and the speed of light. *Optics and Photonics News*, 20(7):42–47, 2009.

[18] Thomas Hockey, Katherine Bracher, Marvin Bolt, Virginia Trimble, Richard Jarrell, JoAnn Palmeri, Jordan D Marché, Thomas Williams, and F Jamil Ragep. *Biographical encyclopedia of astronomers*. Springer, 2007.

[19] The Nobel Prize in physics 1921. `http://www.nobelprize.org/nobel_prizes/physics/laureates/1921/`.

[20] Bernd Jähne, Horst Haussecker, and Peter Geissler. *Handbook of computer vision and applications*, volume 1. Academic Press, 1999.

[21] Richard I Hartley and Peter Sturm. Triangulation. *Computer vision and image understanding*, 68(2):146–157, 1997.

[22] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Horaud. *Time-of-flight cameras: principles, methods and applications*. Springer, 2013.

[23] Robert Lange and Peter Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37(3):390–397, 2001.

[24] Sergi Foix, Guillem Alenya, and Carme Torras. Lock-in time-of-flight (ToF) cameras: a survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.

[25] Robert Lange. *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. PhD thesis, University of Siegen, 2006.

[26] D. M. Cash, T. K. Sinha, W. C. Chapman, H. Terawaki, B. M. Dawant, R. L. Galloway, and M. I. Miga. Incorporation of a laser range scanner into image-guided liver surgery: surface acquisition, reigsration, and tracking. *Med. Phys.*, 30:1671–1682, 2003.

[27] P. Breuer, C. Eckes, and S. Müller. Hand gesture recognition with a novel IR time-of-flight range camera–a pilot study. In *Comp. Vis./Comp. Graphics Collab. Tech.*, pages 247–260. Springer, 2007.

[28] M. C. Amann, T. Boch, R. Myllyla, M. Rioux, and M. Lescure. Laser ranging: a critical review of usual techniques for distance measurement. *Opt. Eng*, 40:10–19, 2001.

[29] Y. Cui, S. Schoun, D. Chan, S. Thrun, and C. Theobalt. 3d shape scanning with a time-of-flight camera. In *Proc. Computer Vision and Pattern Recognition*, 2010.

[30] J. C. Halimeh and M. Wegener. Time-of-flight imaging of invisibility cloaks. *Opt. Express*, 20:63–74, 2012.

[31] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight sensors in computer graphics. In *Proc. Eurographics*, pages 119–134, 2009.

[32] A. Bhandari, A. Kadambi, R. Whyte, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Multifrequency time of flight in the context of transient renderings. In *ACM SIGGRAPH 2013 Posters*, number 46, pages 46:1–46:1. ACM, 2013.

[33] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. Customizing time of flight modulation codes to resolve mixed pixels. In *ACM SIGGRAPH 2013 Posters*, number 29, pages 29:1–29:1. ACM, 2013.

[34] J. P. Godbaz, A. A. Dorrington, and M. J. Cree. Understanding and ameliorating mixed pixels and multipath interference in amcw lidar. In *TOF Range-Imaging Cameras*, pages 91–116. Springer, 2013.

[35] M. Frank, M. Plaue, H. Rapp, U. Köthe, B. Jähne, and F. A. Hamprecht. Theoretical and experimental error analysis of continuous-wave Time-of-Flight range cameras. *Proc. SPIE Conf. on Vis. Commun. and Image Proc.*, 48(1):013602—013602, 2009.

[36] A. P. Jongenelen, D. G. Bailey, A. D. Payne, A. A. Dorrington, and D. A. Carnegie. Analysis of errors in tof range imaging with dual-frequency modulation. *IEEE Trans. on Instrumentation and Measurement*, 60(5):1861–1868, 2011.

[37] D. Droeschel, D. Holz, and S. Behnke. Multi-frequency phase unwrapping for time-of-flight cameras. In *IEEE/RSJ Intl. Conf. on Intell. Robots and Systems*, pages 1463–1469, 2010.

[38] A.A. Dorrington, J.P. Godbaz, M.J. Cree, A.D. Payne, and L.V. Streeter. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *IS&T/SPIE Electronic Imaging*, pages 786404–786404, 2011.

[39] J.P. Godbaz, M.J. Cree, and A.A. Dorrington. Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar. In *IS&T/SPIE Electronic Imaging*, pages 829618–829618, 2012.

[40] John P Godbaz, Michael J Cree, and Adrian A Dorrington. Understanding and ameliorating non-linear phase and amplitude responses in amcw lidar. *Remote Sensing*, 4(1):21–42, 2011.

[41] S. Y. Chen, Y. F. Li, and J. W. Zhang. Vision processing for realtime 3d data acquisition based on coded structured light. *IEEE Trans. Image Proc.*, 17:167–176, 2008.

[42] Active Sensor Plnanning for Multiview Vision Tasks. *Chen, S. Y.* Springer, 2008.

[43] S. Fuchs. Multipath interference compensation in time-of-flight camera images. In *Proc. Computer Vision and Pattern Recognition*, 2010.

[44] D. Jimenez, D. Pizarro, M. Mazo, and S. Palazuelos. Modelling and correction of multipath interference in time of flight cameras. In *Proc. Computer Vision and Pattern Recognition*, 2012.

[45] P. Stoica and R. L. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[46] T. Yamaguchi, I.and Zhang. Phase shifting digital holography. *Opt. Lett.*, 22:1268, 1997.

[47] Adrian PP Jongenelen, Dale A Carnegie, Andrew D Payne, and Adrian A Dorrington. Maximizing precision over extended unambiguous range for TOF range imaging sys-

tems. In *IEEE Intl. Conf. on Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1575–1580. IEEE, 2010.

[48] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012.

[49] J. J. Barton. Removing multiple scattering and twin images from holographic images. *Phys. Rev. Lett.*, 67:3106–3109, 1991.

[50] A.J. Jerri. The Shannon sampling theorem–its various extensions and applications: A tutorial review. *Proc. IEEE*, 65(11):1565–1596, Nov. 1977.

[51] C. D. Boor, R. A. Devore, and A. Ron. Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$. *Trans. Amer. Math. Soc.*, 341(2):787–806, 1994.

[52] Y. M. Lu and M. N. Do. A theory for sampling signals from a union of subspaces. *IEEE Trans. Signal Proc.*, 56(6):2334–2345, 2008.

[53] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. *IEEE Signal Proc. Mag*, 25(2):31–40, March 2008.

[54] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Proc.*, 50(6):1417–1428, 2002.

[55] A. Bhandari and P. Marziliano. Sampling and reconstruction of sparse signals in fractional Fourier domain. *IEEE Signal Proc. Letters*, 17(3):221–224, 2010.

[56] M. Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010.

[57] G. Pfander. Sampling of operators. *Journal of Fourier Analysis and Applications*, pages 1–39, 2013.

[58] F. Krahmer and G. Pfander. Local sampling and approximation of operators with bandlimited Kohn–Nirenberg symbols. *arXiv*, (1211.6048), 2013.

[59] R. Heckel and H. Bölcskei. Compressive identification of linear operators. In *Proc. of IEEE Intl. Sym. on Info. Theory (ISIT)*, pages 1412–1416, August 2011.

[60] R. Heckel and H. Bölcskei. Identification of sparse linear operators. *IEEE Trans. Info. Theory (to appear)*, 2013.

[61] A. Kadambi, A. Bhandari, R. Whyte, A. Dorrington, and R. Raskar. Multiplexing illumination via low cost sensing and nanosecond coding. In *IEEE International Conference on Computational Photography (ICCP)*, March 02-04 2014.

[62] D. Jiménez, D. Pizarro, M. Mazo, and S. Palazuelos. Modelling and correction of multipath interference in time of flight cameras. In *Comp. Vis. and Patt. Rec. (CVPR)*, pages 893–900, 2012.

[63] S. Fuchs. Multipath interference compensation in time-of-flight camera images. In *Intl. Conf. Patt. Rec. (ICPR)*, pages 3583–3586, 2010.

[64] I. M. Gel'fand and G. E. Shilov. *Generalized Functions. Volume I: Properties and Operations*, volume 1. Academic Press, 1964.

[65] J. F. Claerbout and Muir. F. Robust modelling with erratic data. *Geophysics*, 38(1):826–844, 1973.

[66] H. L. Taylor, S. C. Banks, andJ. F. McCoy. Deconvolution with the $\ell_1$ norm. *Geophysics*, 44(1):39–52, 1979.

[67] J. J. Fuchs. Multipath time-delay detection and estimation. *IEEE Trans. Signal Proc.*, 47(1):237–243, 1999.

[68] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. on Sci. and Stat. Comp.*, 7(4):1307–1330, 1986.

[69] Special issue on compressed sensing. *IEEE Signal Processing Magazine*, 25(2), 2008.

[70] P. T. Boufounos. Depth sensing using active coherent illumination. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, March 25-30 2012.

[71] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. on Pure and Appl. Math.*, 2013.

[72] S. Levy and P. K. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46(9):1235–1243, 1981.

[73] B. Muquet, Z. Wang, G. B. Giannakis, M. De Courville, and P. Duhamel. Cyclic prefixing or zero padding for wireless multicarrier transmissions? *IEEE Trans. Commun.*, 50(12):2136–2148, 2002.

[74] G. Plonka and M. Wischerhoff. How many Fourier samples are needed for real function reconstruction? *Jour. of Appl. Math. and Comp.*, 42(1-2):117–137, 2012.

[75] J. Cadzow. Signal enhancement–A composite property mapping algorithm. *IEEE Trans. Signal Proc.*, 36(1):49–62, 1988.

[76] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Signal Proc.*, 38(5):814–824, 1990.

[77] Special issue on spectral estimation. *Proc. IEEE*, 70, Sept. 1982.

[78] Thomas Peter and Gerlind Plonka. A generalized prony method for reconstruction of sparse sums of eigenfunctions of linear operators. *Inverse Problems*, 29(2):025001, 2013.

[79] R. Rubinstein, M. Zibulevsky, and M. Elad. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *CS Technion*, 2008.

[80] S. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[81] Jingyu Lin, Yebin Liu, Matthias B. Hullin, and Qionghai Dai. Fourier analysis on transient imaging by multifrequency time-of-flight camera. In *IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, page to appear, June 2014.

[82] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B. Hullin. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *IEEE Conf. on Comp. Vis. and Patt. Rec. (CVPR)*, page to appear, June 2014.

[83] Daniel Freedman, Eyal Krupka, Yoni Smolin, Ido Leichter, and Mirko Schmidt. SRA: Fast removal of general multipath for ToF sensors. *arXiv:1403.5919*.

[84] Ahmed Kirmani, Arrigo Benedetti, and Philip A Chou. Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods. In *IEEE Intl. Conf. on Multimedia and Expo (ICME)*, pages 1–6, 2013.

[85] Joseph R Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2007.