Resolving Multipath Interference in Kinect: An Inverse Problem Approach

Resolving Multipath Interference in Kinect: An Inverse Problem Approach

Ayush Bhandari*, Micha Feigin*, Shahram Izadi†, Christoph Rhemann†, Mirko Schmidt† and Ramesh Raskar*
*Massachusetts Institute of Technology, Cambridge MA, USA
Email: {ayush,michaf,raskar}@mit.edu
†Microsoft research, Cambridge, UK
Email: {shahram,christoph,mirko}@microsoft.com

***Abstract*—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 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.**

I. Introduction

Amplitude modulated continuous wave (AMCW) Time–of–flight (ToF) imaging cameras [1], [2] measure at each pixel both amplitude and optical travel time (depth), thus capturing three dimensional scene information. Fig. 3(e) and 3(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. They then measure the time delay between the transmission and the reflection arriving back from the scene (Fig. 1(a)), similar in principle to LIDAR.

A. 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 optics and geophysics domains as the Born approximation [3]). 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. 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 [4]–[15] To that end, almost all existing solutions consider up to two interfering paths of light [8]–[10]. The case of K–interfering paths was first discussed in [4] and later in [5]. Both of these papers rely on a sparsity formulation which leads to computationally intensive and iterative algorithms.

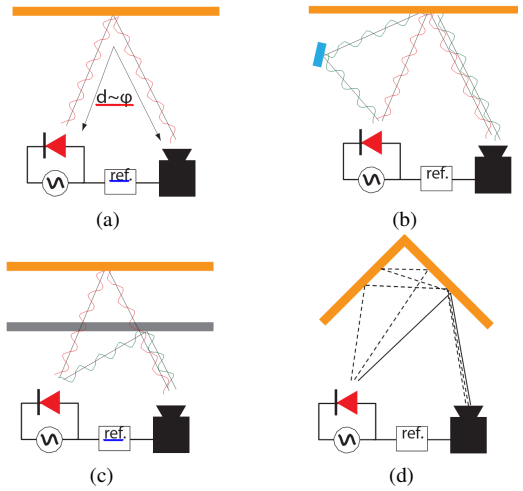

Fig. 1: (a) The ToF camera emits a reference signal. Delay in time of arrival of the reflection encodes the depth d. (b) Specular or mirror like and, (c) Semi–transparent reflections cause multiple light paths to mix at the sensor which leads to multipath interference or the mixed–pixel problem (MPP). (d) Case of continuous multipath reflections.

B. Contribution and Organization of this Paper

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 [4], [5].

In Section II, 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 [16], [17]. We leverage on previous ideas [6], [16] to solve this problem in closed–form. In Section III, we discuss some first results linked with MPI cancellation in Microsoft Kinect One. Finally, we conclude with some future directions.

978-1-4799-0162-3/14/$31.00 ©2014 IEEE

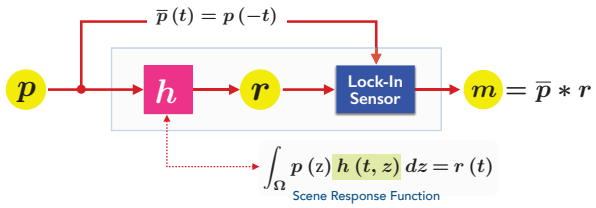

Fig. 2: Time–of–flight imaging system pipeline for solving inverse problems. p is the transmitted signal, h the system (scene) response, r the measured signal and m the correlation measurement.

II. ToF Image Formation Model

The ToF imaging pipeline is shown in Fig. 2. 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 by transmitting a T–periodic function of form $p(t) = p(t+T)$. This is achieved by amplitude modulating a light source, usually co–located at the sensor. In current cameras this signal is, $p(t) = 1 + \cos(\omega_0 t)$, $\omega_0 = 2\pi/T$.
2) **Scene Interaction** The probing function p interacts with the scene resulting with the reflected signal s. This interaction is defined mathematically using scene response function h

$$r(t) = \int_\Omega p(z)\, h(t-z)\, dz \quad \text{(Reflected Signal)}. \tag{1}$$

In the standard multi path case, the scene interaction function is symmetric. With more complex inverse problems such as Fluorescence-lifetime imaging microscopy (FLIM) this function can be non symmetric, i.e $h(z,t)$ and the response function is then described by a more general Fredholm integral [18]
3) **Cross–Correlation** The ToF lock–in sensor measures a discretized cross correlation function of the reference signal against the reflected signal for some sampling step Δ

$$m(t)\Big|_{t=k\Delta} = (p \otimes r)(t),\ k \in \mathbb{Z} \tag{2}$$

here $\otimes$ denotes the correlation function.

A. Modeling Single Depth

Most ToF cameras including the Kinect for the X–Box One probe the scene using an amplitude modulated sine wave

$$p(t) = 1 + \cos(\omega t). \tag{3}$$

At some depth d, the probing function undergoes a reflection as well as attenuation as a function of the scene albedo at the current pixel $0 \leq \Gamma \leq 1$. The total time–of–flight is, $t_d = 2d/c$ where $c = 3 \times 10^8 m/s$ is the speed of light.

Simple computation [1], [2] shows that the measured correlation response at the lock–in–pixel is

$$m_\omega(t) = \Gamma(1 + \cos(\omega(t + t_d))). \tag{4}$$

An equivalent conclusion can be reached be setting the scene response function to

$$h(t) = \Gamma\delta(t - t_d), \tag{5}$$

where δ denotes Dirac's Delta function.

It remains to estimate $\{\Gamma, d\}$ from the measurements m in Eq. (2). Most ToF cameras are so called 4-tap devices, where 4 correlation measurements are used to compute the approximations $\{\widetilde{\Gamma}, \widetilde{d}\}$ at sampling steps $\Delta = \pi/2\omega, k = 0 \cdots 3$ [2],

$$\begin{aligned}\widetilde{\Gamma} &= p_0^{-2}\sqrt{(m_\omega(3\Delta) - m_\omega(\Delta))^2 + (m_\omega(0) - m_\omega(2\Delta))^2} \\ \widetilde{d} &= \frac{c}{2\omega}\tan^{-1}\left(\frac{m_\omega(3\Delta) - m_\omega(\Delta)}{m_\omega(0) - m_\omega(2\Delta)}\right).\end{aligned} \tag{6}$$

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

$$Z_\omega = \widetilde{\Gamma}\exp\left(\jmath\left(2\widetilde{d}\omega\right)/c\right) \triangleq \widetilde{\Gamma}\exp(\jmath\omega t_d), \quad Z_\omega \in \mathbb{C}. \tag{7}$$

One thing to note is that Z_ω is the Fourier Transform of $\Gamma\delta(t - t_d)$.

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

B. Multipath Interference Problem

Consider the case discussed in Fig. 1(b) and 1(c). When multiple bounces of light or optical paths arrive at the sensor, the depth measurements are corrupted [4], [8], [14]. The system response function, modeled based on Eq. (5), is

$$h_K(t) = \sum_{k=0}^{K-1}\Gamma_k\delta\left(t - \frac{2d_k}{c}\right) \equiv \sum_{k=0}^{K-1}\Gamma_k\delta(t - t_k). \tag{8}$$

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

$$h_K(t) \xrightarrow{\text{Fourier}} \widehat{h}_K(\omega) = \sum_{k=0}^{K-1}\Gamma_k e^{-\jmath\omega t_k}. \tag{9}$$

Indeed, for the multipath case, the Fourier domain representation of the system response function in (8) is a sum of K complex exponential functions. Using the phasor formulation (7), the measurement can be represented using this complex value $Z_\omega = \widehat{h}_K^*(\omega)$. From this we get the following formulas for $\{\widetilde{\Gamma}, \widetilde{d}\}$

$$\widetilde{\Gamma}_\omega = \left|\sum_{k=0}^{K-1}\Gamma_k e^{\jmath\omega t_k}\right| \text{ and } \widetilde{\phi}_\omega = \angle\left(\sum_{k=0}^{K-1}\Gamma_k e^{\jmath\omega t_k}\right), \tag{10}$$

a consequence of which is that the camera cannot resolve the interference corrupted measurements.

C. 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 ω.

Problem Statement: Given a vector $\mathbf{m}$ of N measurements at equispaced modulation frequencies where

$$m_n = \sum_{k=0}^{K-1}\Gamma_k u_k^n,\ n \leqslant N \text{ with } u_k = \exp(jt_k) \tag{11}$$

extract $\{\Gamma_k, d_k\}_{k=0}^{K-1}$.

This is a classic inverse problem in line spectrum estimation [17] and is a special case of the solution discussed in context of Fractional Fourier transforms [16].

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

$$\mathsf{Q}(z) = \prod_{\ell=0}^{K-1} \left(1 - u_k z^{-1}\right) \equiv \underbrace{\sum_{\ell=0}^{K} q_\ell z^{-\ell}}_{\text{Factorized Polynomial}}, \quad z \in \mathbb{Z}. \tag{12}$$

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 (12)). Hence the null space of the polynomial contains the decoupled phase information for the different paths. In vector–matrix notation, this is equivalent to,

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

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 $\text{rank} = K$. We need at least $N = 2K+1$ measurements to construct $\mathbf{M}$ and hence, $2K+1$ modulated frequencies in m_ω to recover K optically interfering paths. Once we compute $\mathbf{q}$, we factorize it to compute the roots $\{u_k\}$ in (12). This, in turn leads to,

$$\widetilde{d}_k = (2\jmath)^{-1} c \log(u_k) \qquad \text{where} \qquad c = 3 \times 10^8 m/s. \tag{13}$$

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

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

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 $\mathbf{U}$.

D. Remarks and Discussion

- **Improvement over previous work.** To begin with, our work generalizes the results in [9], [10] 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 [9]. Also, unlike [9] 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 [4] 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_ω 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 [17] 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 [19] to solve to problem.

III. Results

For experiments we use an off the shelf, Microsoft Kinect for the X–Box 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 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. 3(j). In this example, we set $K = 2$. The amplitude and phase measurements at 52 MHz are shown in Fig. 3(a) and 3(b) respectively. We use 21 measurements from $52 - 72$ MHz for our processing. For the experiment in Fig. 3(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 background. Fig. 3(c) and 3(d) shows the amplitude/depth pair corresponding to the contribution due to acrylic sheet. As can be seen in Fig. 3(c), we observe a specular reflection due to the reflective acrylic sheet. In Fig. 3(e) and 3(f), we show the contribution due to the background. The depth of the table is estimate to be 1.67 meters. The amplitude image in Fig. 3(e) is devoid of any multipath effects due to the acrylic sheet, including the specular reflection.

Fig 3(g) and 3(g) show the 3d reconstruction (amplitude embedded on the depth data) at 52 MHz and 72 MHz. The error in amplitude recovery is a tell tale sign of multi path interference. Fig. 3(i) shows amplitude embeded onto the 3d information for the 2 bounces.

IV. 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.

References

[1] M. Hansard, S. Lee, O. Choi, and R. Horaud, *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer, 2013.

[2] R. Lange, "3d time–of–flight distance measurement with custom solid-state image sensors in cmos/ccd-technology," Ph.D. dissertation, Department of Electrical Engineering and Computer Science at Unversity of Siegen, 2000.

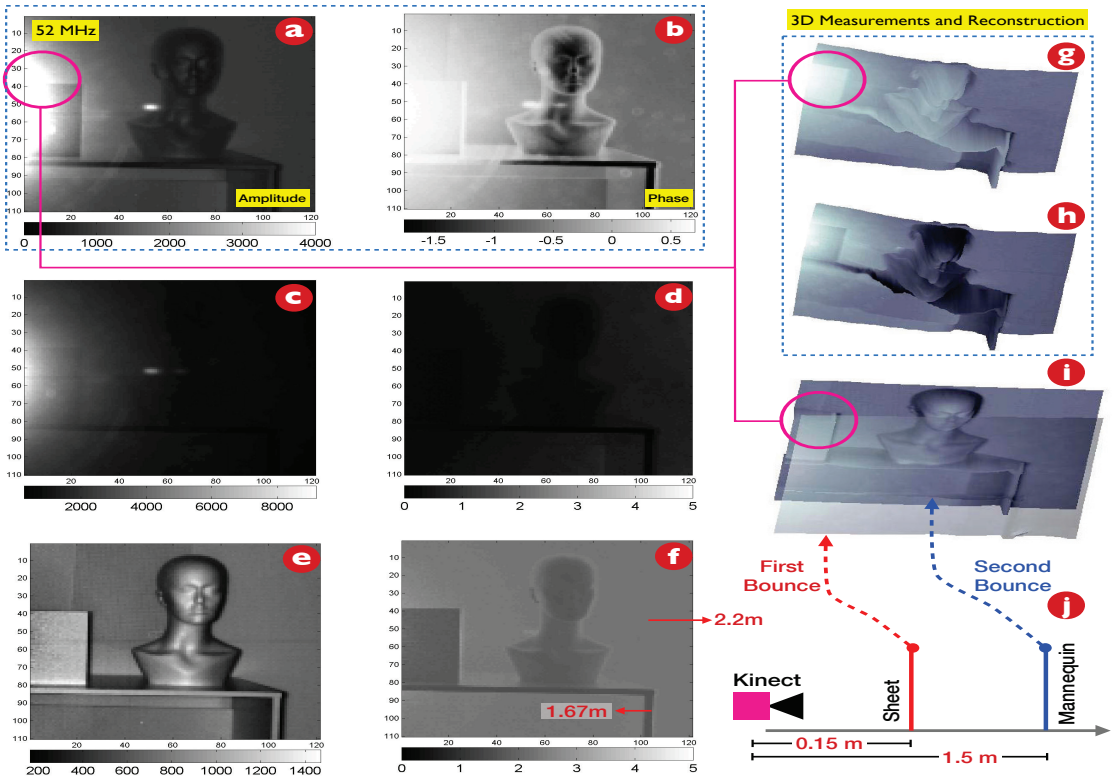

Fig. 3: Images (a) and (b) show one of the input amplitude and phase images while (g) and (h) show the embedding of the intensity image onto the depth data. The scene, depicted in (j) is constructed from a mannequin on a table with a transparency between the camera and the scene. As can be seen, there 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).

[3] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. Cambridge University Press, 1999.

[4] 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*, vol. 39, no. 7, 2014.

[5] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for ToF sensors," *arXiv:1403.5919*.

[6] A. Bhandari, A. Kadambi, and R. 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. (ICASSP)*, 2014, pp. 365–369.

[7] A. Kadambi and R. Whyte and A. Bhandari and L. Streeter and C. Barsi and A. Dorrington and R. Raskar, "Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles," *ACM Trans. on Graphics (TOG)*, vol. 32, no. 6, p. 167, 2013.

[8] 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*. Springer, 2013, pp. 91–116.

[9] A. Kirmani, A. Benedetti, and P. 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)*, 2013, pp. 1–6.

[10] J. P. Godbaz, M. J. Cree, and A. Dorrington, "Closed-form inverses for the mixed pixel/multpath interference problem in amcw lidar," in *IS&T/SPIE Electronic Imaging*, 2012.

[11] 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)*, 2012, pp. 893–900.

[12] A. Jongenelen, D. Bailey, A. Payne, A. Dorrington, and D. Carnegie, "Analysis of errors in ToF range imaging with dual-frequency modulation," *IEEE Trans. on Inst. and Meas.*, vol. 60, no. 5, pp. 1861–1868, 2011.

[13] 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," vol. 7864. SPIE, 2011, pp. 786 404–786 404–10.

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

[15] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," in *ACM Transactions on Graphics*, 2013.

[16] A. Bhandari and P. Marziliano, "Sampling and reconstruction of sparse signals in fractional Fourier domain," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 221–224, 2010.

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

[18] A. Bhandari, C. Barsi, R. Whyte, A. Kadambi, A. J. Das, A. Dorrington, and R. Raskar, "Coded time-of-flight imaging for calibration free fluorescence lifetime estimation," in *Imaging Systems and Applications*. Optical Society of America, 2014, pp. IW2C–5.

[19] H. Yingbo and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise." *IEEE Trans. Acoust., Speech, and Signal Proc.*, pp. 814–824, 1990.