Macroscopic Interferometry
Supplemental Material

Achuta Kadambi, Jamie Schiel, and Ramesh Raskar

MIT Media Lab

{achoo,schiel,raskar}@mit.edu

Supplemental Contents

This supplement is organized as follows:

- Section 1 describes algorithms for frequency estimation (main paper, line 509)
- Section 2 provides additional details about the resolution bound on multipath (main paper, lines 532 and 574)
- Section 3 describes aliasing in the context of long-range depth sensing (main paper, line 858).
- Section 4 contains implementation details: source code, datasets and hardware schematics.

1. Algorithms for Frequency Estimation

A comprehensive overview of tone estimation can be found in Stoica and Moses [6]. We review only the facets of tone estimation that are useful for reproducing our FD-ToF paper.

Our paper employs the Discrete Fourier Transform (DFT), which converts equally spaced primal-domain samples into a a coefficient representation of complex sinusoid. The accuracy to which frequencies can be localized depends on the sampling frequency and the number of samples (i.e, the frequency bin width is f_s/N, where f_s is the sampling rate and N is the number of temporal samples). The DFT is very general, able to handle many simultaneous tones within the received signal, but can be less accurate than frequency estimators that assume a fixed number of frequencies.

Two such estimators are the Quinn-Fernandes (QF) and Matrix Pencil (MP) techniques.[1] The QF estimator uses a regression model to fit a single frequency to a smoothed parametric representation of the data [5]. For a single tone, the QF algorithm is robust to noise and poor initial estimates. In comparison, the MP technique can be used to separate multiple returns [3]. Unlike the DFT, standard implementations of MP require the number of frequencies to be known *a priori*.

2. Resolution Bound on Multipath

Numerical approximation of Proposition 1

Recall from *Proposition 1* that the sampling function in the primal-domain is a boxcar, from a minimum frequency of f_{M}^{-} to a maximum frequency of f_{M}^{+}. In the dual-domain, the sampling function is written as

$$\begin{aligned}\mathcal{F}\left[\Pi\left(f_{\mathrm{M}}\right)\right](\kappa) &= \Delta f_{\mathrm{M}} \frac{\sin \Delta f_{\mathrm{M}} \kappa}{\kappa}, \\ &\approx \Delta f_{\mathrm{M}}\left[\Delta f_{\mathrm{M}} - \frac{\Delta f_{\mathrm{M}}^{3}\kappa^{2}}{3!} + \frac{\Delta f_{\mathrm{M}}^{5}\kappa^{4}}{5!} - \frac{\Delta f_{\mathrm{M}}^{7}\kappa^{6}}{7!} + \frac{\Delta f_{\mathrm{M}}^{9}\kappa^{8}}{9!} - \ldots\right],\end{aligned} \tag{1}$$

[1] MP can also be used for phase ToF [2, 1].

where the approximation follows from Taylor expansion. Note that the maximum value of $\mathcal{F}\left[\Pi\left(f_{\mathrm{M}}\right)\right](\kappa)$ occurs at the origin, where $\lim_{\kappa\to 0}\Delta f_{\mathrm{M}}\frac{\sin\Delta f_{\mathrm{M}}\kappa}{\kappa}=\Delta f_{\mathrm{M}}^{2}$. Then, the FWHM is obtained by finding the roots of the polynomial:

$$p(\kappa)=\frac{\Delta f_{\mathrm{M}}^{2}}{2}-\frac{\Delta f_{\mathrm{M}}^{4}\kappa^{2}}{3!}+\frac{\Delta f_{\mathrm{M}}^{6}\kappa^{4}}{5!}-\frac{\Delta f_{\mathrm{M}}^{8}\kappa^{6}}{7!}+\frac{\Delta f_{\mathrm{M}}^{10}\kappa^{8}}{9!}-\ldots \tag{2}$$

Using, for example, Newton's bisection method to obtain the polynomial roots, the FWHM is calculated as approximately $1.206\pi/\Delta f_{\mathrm{M}}$, rounded to $1.2\pi/\Delta f_{\mathrm{M}}$ in the main paper. Following Equation 25 from the main paper, the FWHM is expressed in terms of the depth pseudospectrum, yielding an axial resolution of $0.6c/\Delta f_{\mathrm{M}}$, the result of *Proposition 1*.

Simulation noise model

The noise model for simulations follows that of a typical imaging system. For each subframe, the noise variance follows an additive model, i.e., $\sigma^2=\sigma_p^2+\sigma_d^2+\sigma_r^2$, where σ_p^2, σ_d^2 and σ_r^2 is the variance due to photon shot noise, dark noise, and read noise, respectively. The simulation assumes that dark noise is mitigated (e.g. by cooling the sensor) and that sufficient light is measured, such that the imaging system is photon shot noise limited, i.e., $\sigma^2\approx\sigma_p^2$. Then, the SNR in dB is expressed as $\mathrm{SNR}=20\log_{10}\frac{N}{\sqrt{N}}=10\log_{10}N$, where N is the number of electrons. As illustrated in Figure 4 of the main paper, the typical SNR range is from 15 to 40 dB.

In Figure 3 of the main paper, the noiseless bound is compared to the observed resolution bound in the presence of photon shot noise. The red line is the theoretical value (calculated from *Proposition 1*) and the black line is the measured FWHM in the presence of shot noise. The deviation from the bound is on average 1cm. Note that the y-axis of Figure 3 is the minimum multipath separation in depth (as described in text).

3. Aliasing of Long-range Depths

Although the proposed technique is immune to wrapping, long-range depths create high-frequency tones that can alias to lower frequencies. However, we show here that this is not a practical issue. Indicate the cutoff depth as $d_{\max}$ and substitute into Equation 24 to obtain

$$c\left(f_{\mathrm{M}}\right)=\frac{\alpha}{2}\cos\left(\frac{4\pi d_{\max}}{c}f_{\mathrm{M}}\right)+\beta. \tag{3}$$

The Nyquist rate is then $4d_{\max}/c$, in units of samples per Hz. The sampling period in the primal-domain of frequency is $T=c/4d_{\max}$, expressed in units of Hz. The cutoff depth is then:

$$d_{\max}=\frac{c}{4T}. \tag{4}$$

In practical settings, the sampling period can be made small. For instance at $T=1$KHz, the cutoff depth $d_{\max}$ is 300 kilometers.

4. Implementation Details

For reproducibility, a sketch of implementation is provided for each figure.

Figure 2

Figure 2 of the main paper shows a plot of the primal-domain and the dual-domain for simulated FD-ToF measurements from 10 and 20 meters, in the noiseless case. The signal is sampled at a bandwidth of 1 GHz. To reproduce the figure, the pseudocode roughly follows this structure (shown only for the object at 10 meters):

```
% primal-domain signal for 10 meter reflection.
depth = 10; % depth in meters
c = 3e8; % speed of light
alpha = 50000; % signal amplitude
fM = linspace(minF,maxF,N); % determines bandwidth
y = alpha/2*cos(2*pi*2*depth*(1/c)*fM)+alpha/2;

% dual-domain signal
yfft = 1/length(y)*fft(y-mean(y)),nfft); % subtract DC
```

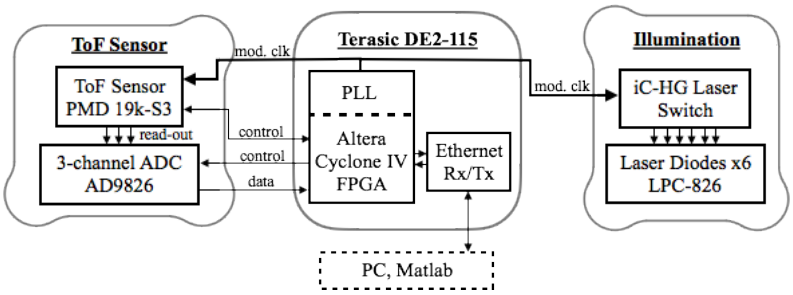

Figure 1. Block diagram of hardware architecture

```
P1 = abs(yfft(1:nfft/2+1));
P1(2:end-1) = 2*P1(2:end-1);

F = fs*(0:(nfft/2))/nfft;
F = c/2*F; % depth pseudospectrum
plot(F,P1);
```

Figure 3

Figure 3 of the main paper compares the theoretical bound with the observed resolution bound in the context of shot noise (SNR is 40 dB). The red curve is the theoretical bound, obtained through direct plotting of *Proposition 1*. The black curve is the empirical resolution in context of shot noise. The empirical resolution is obtained by numerical approximation of the FWHM (through Newton's method) in the dual-domain. Note that the plot in Figure 3 is in context of depth, i.e., it divides the axial bound by 2.

Figure 4

Figure 4 of the main paper renders a complex scene with multipath. To perform renderings a modified version of the research software Mitsuba is used to provide multipath corrupted depth maps (by setting the bounce index in the renderer).

The Phase ToF result is shown with and without MPI correction (using Bhandari's method [1]). The Bhandari method uses the data in multiple frequencies and phases to structure a generalized eigenvalue problem. To motivate a fair comparison, the code and optimal tuning parameters were sourced directly from the authors of [1].

Prototype Camera: Figures 5 to 8

Hardware Architecture

An overview diagram of the prototype camera used to collect frequency-domain data is shown in Figure 1 of the supplemental material. The camera is an evolution of the design in [4]. There are three primary components (circuit boards): the ToF sensor board, illumination board, and FPGA control board.

The ToF sensor board houses a 19k-S3, time-of-flight CMOS image sensor from PMD Technologies. Rapid read-out of the 120x160 sensor array is achieved with simultaneously output of 3 pixels as an analog voltage, which is cast to a digital reading by a high-speed analog-to-digital converter (ADC). The critical feature of the ToF board is the ability to precisely control the electronic exposure from an external trigger, labeled as mod. clk.

Scene illumination is controlled by six LPC-826 650nm, diffuse laser diodes. Diffuse, visible light is used for eye-safety reasons; operators can know when the camera is on. However, IR light could also be used, as is common practice in consumer devices.

Image sensor configuration, pixel array readout, ADC polling, and modulation of the illumination and sensor exposure are controlled by a Cyclone IV FPGA. We utilize the FPGAs dynamically reconfigurable phased-locked loops (PLL) to synchronously modulate the image sensor and illumination, and alter frequency on-the-fly. Alteras soft-core microprocessor, Nios, implements an ethernet stack for half-duplex communication with a PC. Image data can be read directly into Matlab, or control signals sent to the camera, via this connection.

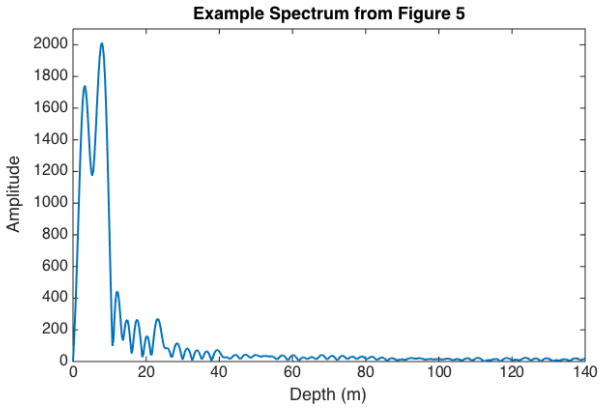

Figure 2. Dual-domain of a single pixel of Figure 5 from the main paper. Note that the two reflections—from the glass transparency and far wall—are visible as two distinct peaks at 4.71 and 8.86 meters.

Capture Process and Depth Estimation

An image is acquired at each desired frequency, with invariant exposure times. The result is a datacube of size $120 \times 160 \times F$, where F is the number of frequencies captured. Note that there will be a phase-offset induced by electronic delays in the datapaths through the camera, which results in erroneous depth estimates. However, this phase-offset is constant for all pixels, and at all frequencies, so can be calibrated away using a scene-point of known depth.

Matlab code for processing a single pixel across the cameras frequency bandwidth is shown below.

```
% Input parameters
f = 50; % Input data captured at 50Mhz bandwidth (for example)
N = 2048; % 2048-point fft

% FFT on raw_data input
fft_ = 1/length(raw_data) * fft(raw_data, N); % Perform N-point fft

% Sort frequency components
spec = abs(fft_(0:N/2+1)); % one-sided power spectrum
[spec_sorted, I] = sort(spec, 'descend');
f_axis = f*(0:N/2)/N;

% Convert to depth
fundamental = f_axis(I(2)); % Exclude DC
depth = (fundamental./(4*pi))*3e8;
```

We show the example of using an N-point FFT for depth estimation, as used in the paper to obtain simulated and experimental datasets. The example code processes a single scene point (pixel) captured with a frequency bandwidth of 50Mhz. It is assumed that the dominant frequency corresponds to the true object depth. Specifically, this means that the optical return signal from the object is of higher intensity than any multi-path returns. Figure 2 of the supplement plots the received signal in the dual-domain, for a single pixel of Figure 5 from the main paper. The two return signals from the scene and glass sheet can clearly be seen in Figure 2 of the supplement as distinct frequency peaks.

Specular highlights As shown in Figure 5 of the main paper, a scar is visible in the upper left of the recovered depth maps. This occurs because of a specular highlight on the transparency, which drowns out the reflection from the object. This limitation occurs with any multipath separation technique; for example, Bhandari's method also suffers from the specular highlight.

References

[1] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar. Resolving multipath interference in kinect: An inverse problem approach. In *IEEE Sensors*, 2014. 1, 3

[2] 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 *ICASSP*, 2014. 1

[3] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(5):814–824, 1990. 1

[4] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics*, 2013. 3

[5] B. Quinn and J. Fernandes. A fast efficient technique for the estimation of frequency. *Biometrika*, 1991. 1

[6] P. Stoica and R. L. Moses. *Introduction to spectral analysis*, volume 1. Prentice hall Upper Saddle River, NJ, 1997. 1