Doppler Time-of-Flight Imaging
Supplemental Material

1 Harmonic ToF Camera Model

In this section we expand on the sinusoidal model derived in the to allow for arbitrary periodic modulation functions for both the illumination and the camera reference signal.

As before, we assume periodically modulated illumination with an angular frequency ω. In the general case, this illumination signal can be described as a superposition of all the harmonics of this frequency at different phases θ_k:

$$g(t) = \sum_{k=0}^{\infty} g_k \, \cos(k\omega t + \theta_k)$$

Assuming a stationary object, the illumination arriving at the sensor can then be expressed as

$$s(t) = \sum_{k=0}^{\infty} s_k \, \cos(k\omega t + \phi_k)$$

where s_0 includes ambient illumination, and s_k for $k > 0$ includes geometric terms and the surface albedo as before. The phase shifts are given as

$$\phi_k = \theta_k + k\omega \frac{2d}{c}$$

Assuming a zero-mean modulation function f, the modulated signal at the sensor is then

$$\widetilde{i}_\psi(t) = f_\psi(t) \cdot s(t) \tag{1}$$

$$= \sum_{j=1}^{\infty} f_j \, \cos(j\omega t + \psi_j) \cdot \sum_{k=0}^{\infty} s_k \, \cos(k\omega t + \phi_k) \tag{2}$$

$$= \frac{1}{2} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} f_j s_k \cos((k-j)\omega t + \phi_k - \psi_j) +$$
$$\frac{1}{2} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} f_j s_k \cos((k+j)\omega t + \phi_k + \psi_j) \tag{3}$$

$$= \frac{1}{2} \sum_{k=0}^{\infty} f_k s_k \cos(\phi_k - \psi_k) +$$
$$\frac{1}{2} \sum_{j=1}^{\infty} \sum_{k \neq j} f_j s_k \cos((k-j)\omega t + \phi_k - \psi_j) +$$
$$\frac{1}{2} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} f_j s_k \cos((k+j)\omega t + \phi_k + \psi_j). \tag{4}$$

Low-pass filtering due to the finite exposure time T eliminates all but the first term in Equation 4:

$$i_\psi(t) = \widetilde{i}_\psi(t) * \mathrm{rect}(Tt) \approx \frac{1}{2} \sum_{k=0}^{\infty} f_k s_k \cos(\phi_k - \psi_k). \tag{5}$$

In order to recover depth and albedo from Eq. 5, the camera needs to be calibrated for different depths, either as a dense lookup table of depth, or by estimating the coefficients f_k and s_k from sparser measurements.

For moving objects, we can again analyze the Doppler shift, and obtain

$$s(t) = \sum_{k=0}^{\infty} s_k \, \cos(k(\omega + \Delta\omega)t + \phi_k)$$

and

$$i_\psi(t) \approx \frac{1}{2} \sum_{k=0}^{\infty} f_k s_k \cos(\Delta\omega t + \phi_k - \psi_k). \tag{6}$$

Finally, we can use Equation 6 to analyze inter-carrier interference in the case of general periodic functions. As in Section 4.1 of the main article, we define the ratio image of a heterodyne and a homodyne image, and obtain:

$$r \approx \frac{\int_0^T \frac{1}{2} \sum_{k=0}^{\infty} f_k s_k \cos((\omega_f - \omega_g - \Delta\omega)t + \psi - \phi) \, dt}{\int_0^T \frac{1}{2} \sum_{k=0}^{\infty} f_k s_k \cos(-\Delta\omega t + \psi - \phi) \, dt} \tag{7}$$

Due to the multiple terms in both numerator and the denominator this equation cannot be simplified such that the dependencies on the phases ψ and ϕ cancel out. Unlike the sinusoidal model, general periodic functions thus introduce a phase- or depth-dependency to the velocity measurements. These need to be calibrated (see Section 6 in the main text).

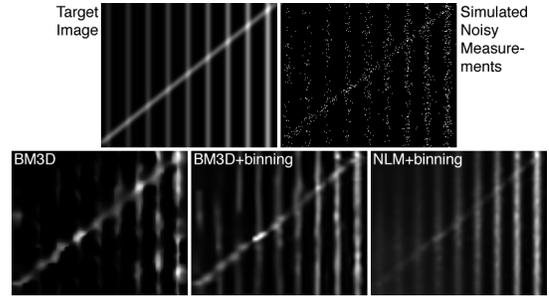## 2 Additional Information for Velocity Map Denoising

Figure 2: *Synthetic test pattern (top left) and simulated measurement that is heavily corrupted by Poisson noise (top right). Various modern denoising strategies applied, including BM3D (bottom left), BM3D with binning (bottom center), and non-local means (NLM) with binning (bottom right). To account for the high level of Poisson noise in the measurements, we apply the latter denoising strategy to all captured velocity data.*

With our system, we capture an extremely small frequency shift (in the Hz range) relative to the modulation frequency (the MHz range). Additionally, the quantum efficiency of emerging time-of-flight sensors is still far from that of modern solid state sensors [Erz and Jähne 2009], and the demodulation contrast drops significantly for high frequencies [Shcherbakova 2013]. Under these conditions,
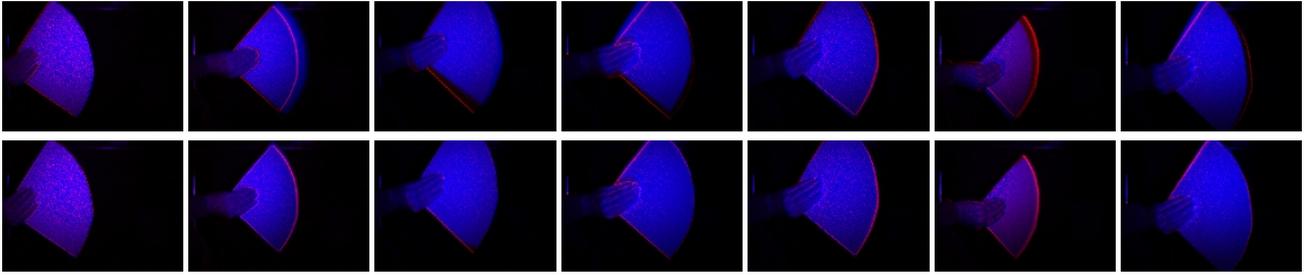
Figure 1: *Alignment of raw frames: Each individual image has the heterodyne capture encoded in the red color channel and the homodyne capture in the blue channel. The top row shows frames before alignment. The bottom row are the respective frames after alignment.*

the few photons collected during a measurement are strongly affected by Poisson noise. To account for this, we apply a denoising strategy to all captured velocity maps.

We test several denoising strategies for a synthetic example in Figure 2. The ground truth test pattern (top left) is synthetically corrupted by Poisson noise (top right). Standard denoising approaches, such as BM3D fail in these conditions, even when variance-stabilizing transforms are applied (bottom left). As recently discussed by Salmon et al. [2014], a simple improvement of BM3D results in significant improvements. First, one accumulates the photon counts locally in bins (for example with a size of 3×3 pixels) that now have much higher photon counts (but lower spatial resolution). The denoiser is subsequently applied to the binned counts. Finally, the image is upsampled. We show simulated results using this method in Figure 2 (bottom center). Unfortunately, we observe completely missing structures in the image, which is undesirable. Using the binning strategy [Salmon et al. 2014] but replacing BM3D by a non-local means denoiser results in slightly blurrier but overall favorable results (bottom right). We apply this denoising strategy to all of the presented results.

3   Subframe Alignment

In our hardware implementation we did not have access to the FPGA controlling the capture protocol of our sensor. Therefore, we were limited to capturing the individual subframes in an alternating fashion. Note that if frame-level synchronization can be implemented, the required heterodyne and homodyne shots could be captured simultaneously using multi-sensor configurations. For our time-sequential implementation we align the frames post-capture using SIFT flow on the raw data. The resulting flow fields are used to warp each homodyne sub-frame to the raw heterodyne sub-frame of each respective capture. Figure 1 shows results for this alignment procedure. While not being perfect, the flow estimates are reasonable for most scenarios.

4   Additional Results

This section includes additional results which have not been presented in the paper. Results for a variety of different scenes and velocities are shown in the Figures 3, 6, 8 , 4 and 5. We show linear trajectories with Figure 3, rotating movement orthogonal to the optical axis in Figure 6, arced trajectories in Figure 8, periodic movement along the optical axis in Figure 4 and slow rotating movements in Figure 5.

Besides accurate velocity in all shown scenarios, we also demonstrate with Figures 6, 8 that our system is capable of producing reasonable results outdoors under strong ambient illumination.
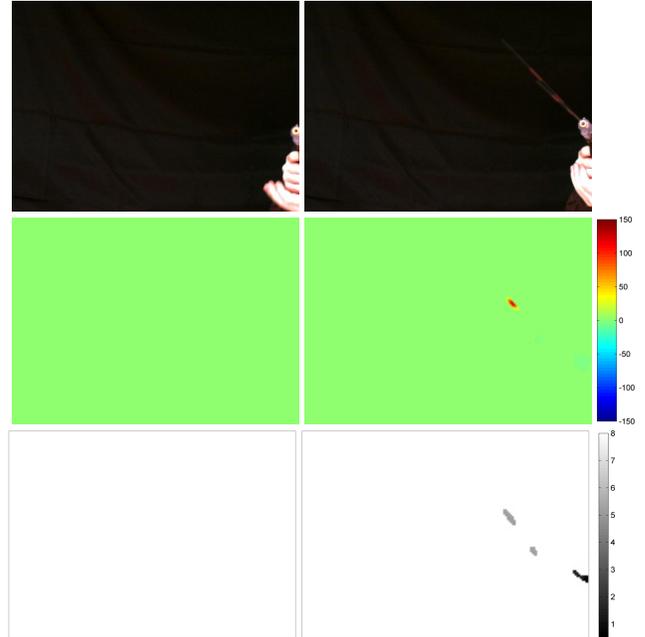

Figure 3: *Additional airsoft gun example including depth. Top: Color frames, Center: Measured and reconstructed velocity, Bottom: Reconstructed depth.*
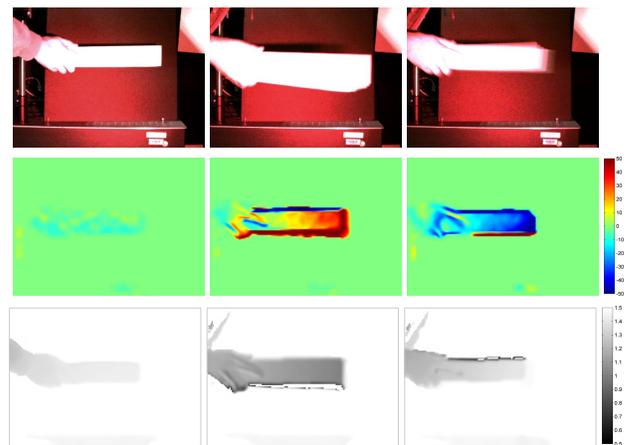

Figure 4: *Velocity reconstruction for a long stick moved back and forth along the optical axis.*
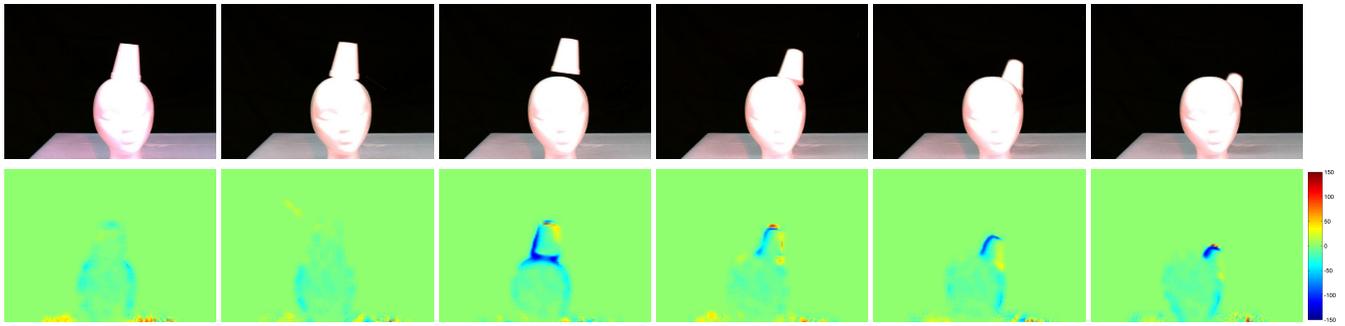
Figure 5: *Multiple color (top row) and velocity (bottom row) frames of a video. The styrofoam head balances a cup that is shot by the spring airsoft gun.*
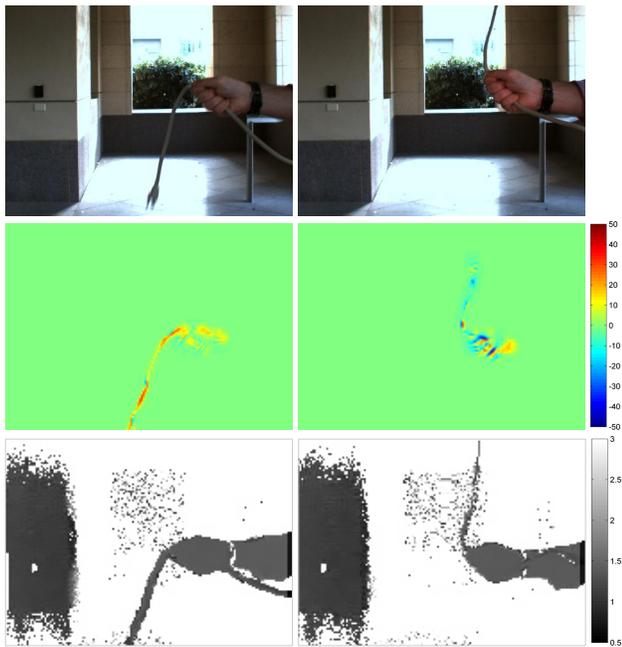

Figure 6: *Rotational movement on a circular trajectory. The forward motion on the left (towards the camera) is measured in the lower part of the trajectory. Backward motion on the right (away from the camera) is measured in the upper part of the trajectory. Additionally we show depth estimates.*
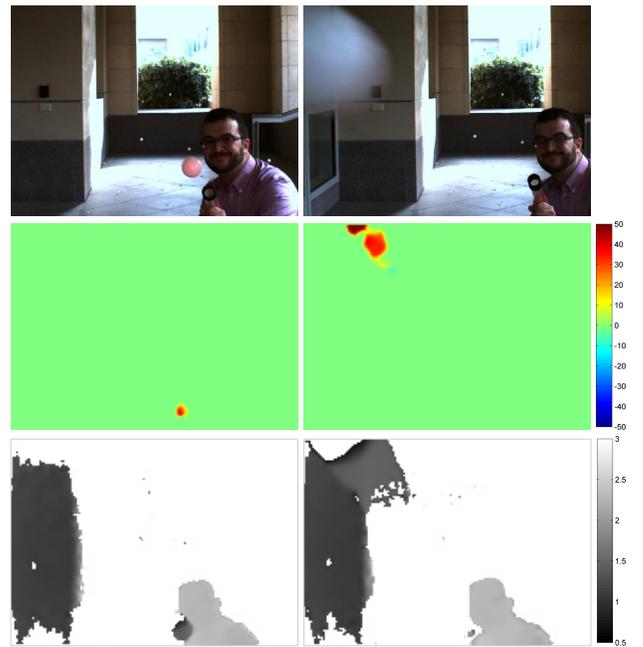

Figure 8: *Additional results of the ping-pong gun. Because of the baseline between the two cameras, objects very close to the camera are not perfectly aligned as seen on the right.*
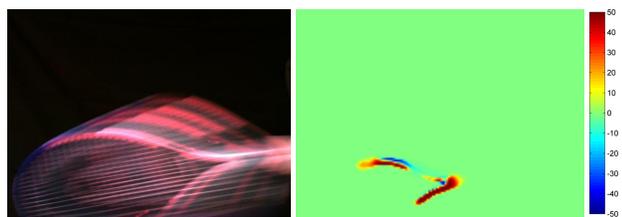
References

ERZ, M., AND JÄHNE, B. 2009. Radiometric and spectrometric calibrations, and distance noise measurement of tof cameras. In *Dynamic 3D Imaging*. Springer, 28–41.

SALMON, J., HARMANY, Z., DELEDALLE, C.-A., AND WILLETT, R. 2014. Poisson noise reduction with non-local pca. *Journal of mathematical imaging and vision 48*, 2, 279–294.

SHCHERBAKOVA, O. 2013. *3D Camera Based on Gain-Modulated CMOS Avalanche Photodiodes*. PhD thesis, University of Trento.

Figure 7: *Very fast rotating movement of a tennis bat. This is a failure case for the alignment. The thin structures of the mesh inside of the bat fail to be aligned properly and therefore do not result in a sensor response. Only the thicker handle is properly reconstructed.*