Technische Universität München

Model-Based Tracking at 300Hz using Raw Time-of-Flight Observations

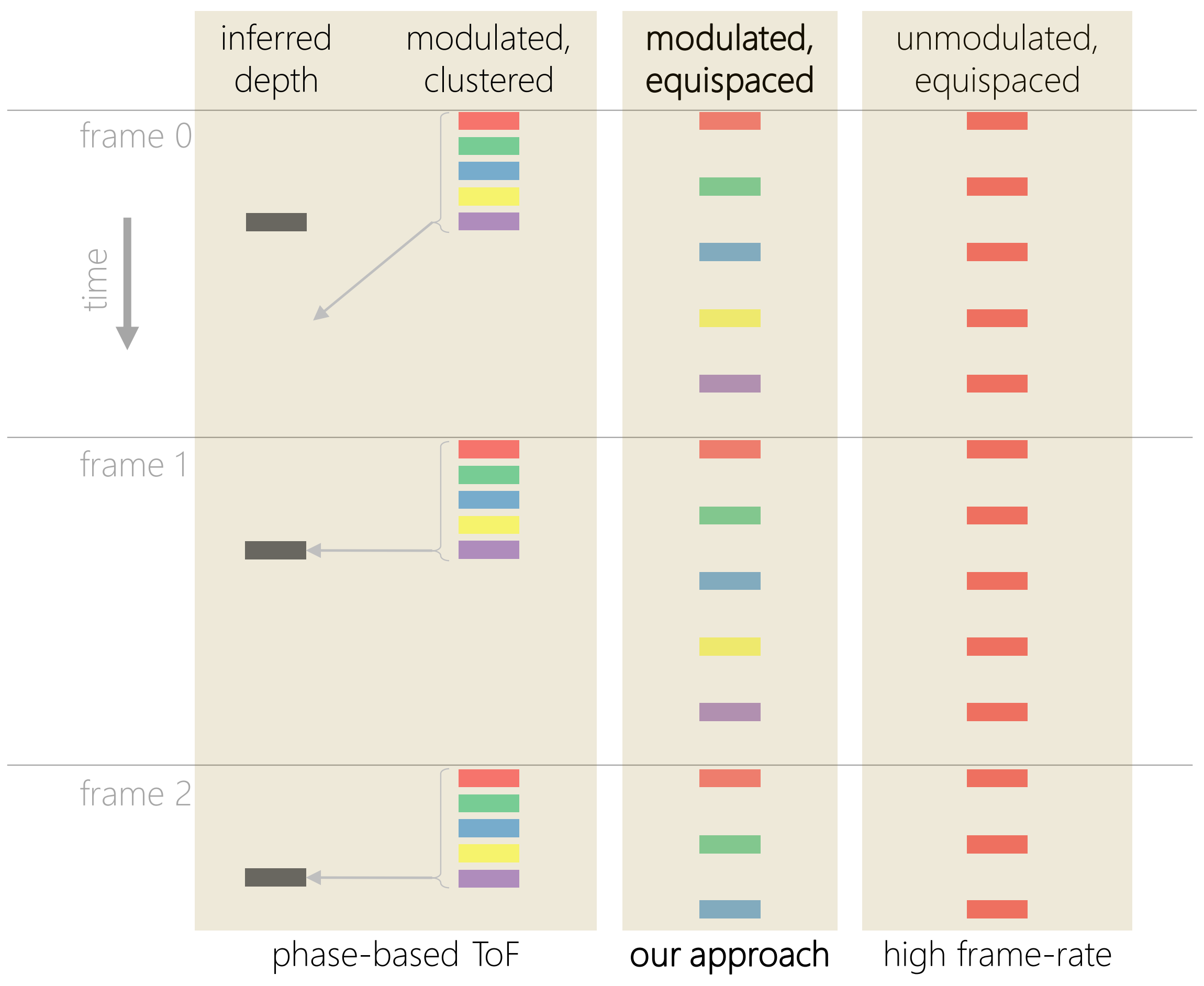
Jan Stühmer, Sebastian Nowozin, Andrew Fitzgibbon, Richard Szeliski,
Travis Perry, Sunil Acharya, Daniel Cremers and Jamie Shotton

Microsoft Research

Introduction

We show how to perform model-based object tracking directly on the raw infrared captures of a time-of-flight (Tof) camera.

We focus on phase-based time-of-flight sensing, which reconstructs each low frame-rate depth image from a set of short exposure `raw' infrared captures. These raw captures are taken in quick succession near the beginning of each depth frame, and differ in the modulation of their active illumination.



Contributions

First, we detail how to perform model-based tracking against these raw captures.

Second, we show that by taking the raw captures uniformly in time, we obtain a 10x higher frame-rate, and improve the ability to track fast-moving objects.

Our method is efficiently implemented on the GPU and allows to track an object with 300 frames per second.

Model-based Tracking

We use a model-based tracking approach with a generative observation model to relate the tracked position to the observation over time and a temporal model for tracking stability [1,2]. While our Motion model is standard, the observation model for raw ToF capture is a novel contribution:

Observation Model

We propose an observation model which removes the dependence on a ToF depth reconstruction and instead compute observation likelihoods directly against the raw ToF captures:

Object hypothesis

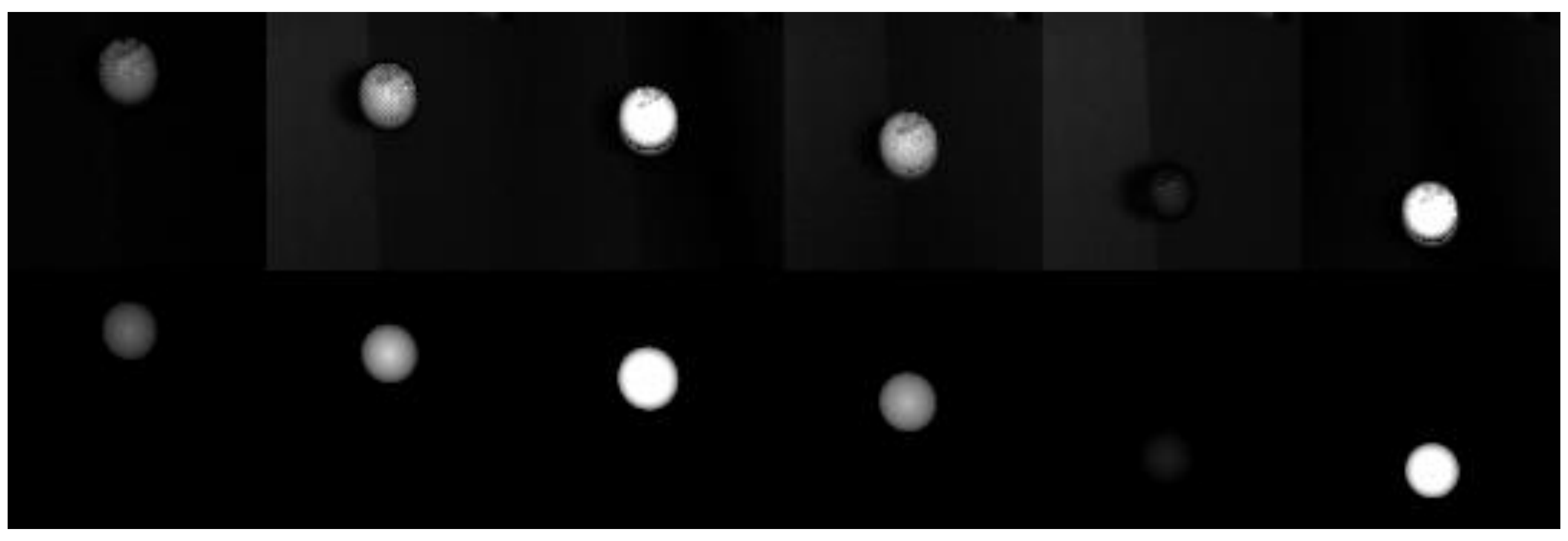
Measured response

Expected ideal object response

$$\log P(Y_t|X_t) = -\sum_{u:M(u)=1}\left[\frac{(\bar{R}_i(u)-R_{\text{obj}}(u))^2}{2\sigma_{\text{obj}}^2}+\log\sigma_{\text{obj}}\right]$$
$$-\sum_{u:M(u)=0}\left[\frac{(\bar{R}_i(u)-\hat{\mu}_i(u))^2}{2\sigma_{\text{bg}}^2}+\log\sigma_{\text{bg}}\right]+C$$

Raw ToF observation

Background model



Model Based Tracking: Depending on the frequency and phase configuration of the individual exposures, the object appears with different illuminations in the raw ToF captures. Our generative forward model allows to synthesize the appearance of the object for theses different illuminations (Columns correspond to these different exposures)

First row: Observed raw ToF image

Second row: Rendered image of the best hypothesis

Phase-Modulation Time-of-Flight: Modern ToF cameras operate based on the principle of phase modulation: a modulated light source emits a sinusoidal light signal at a specific frequency, and a special sensor images the light's reflection, gain-modulated at the same frequency.

For each pixel we obtain a sequence of nine measurements $R_1, \ldots, R_9$ (3 frequencies × 3 phases) via

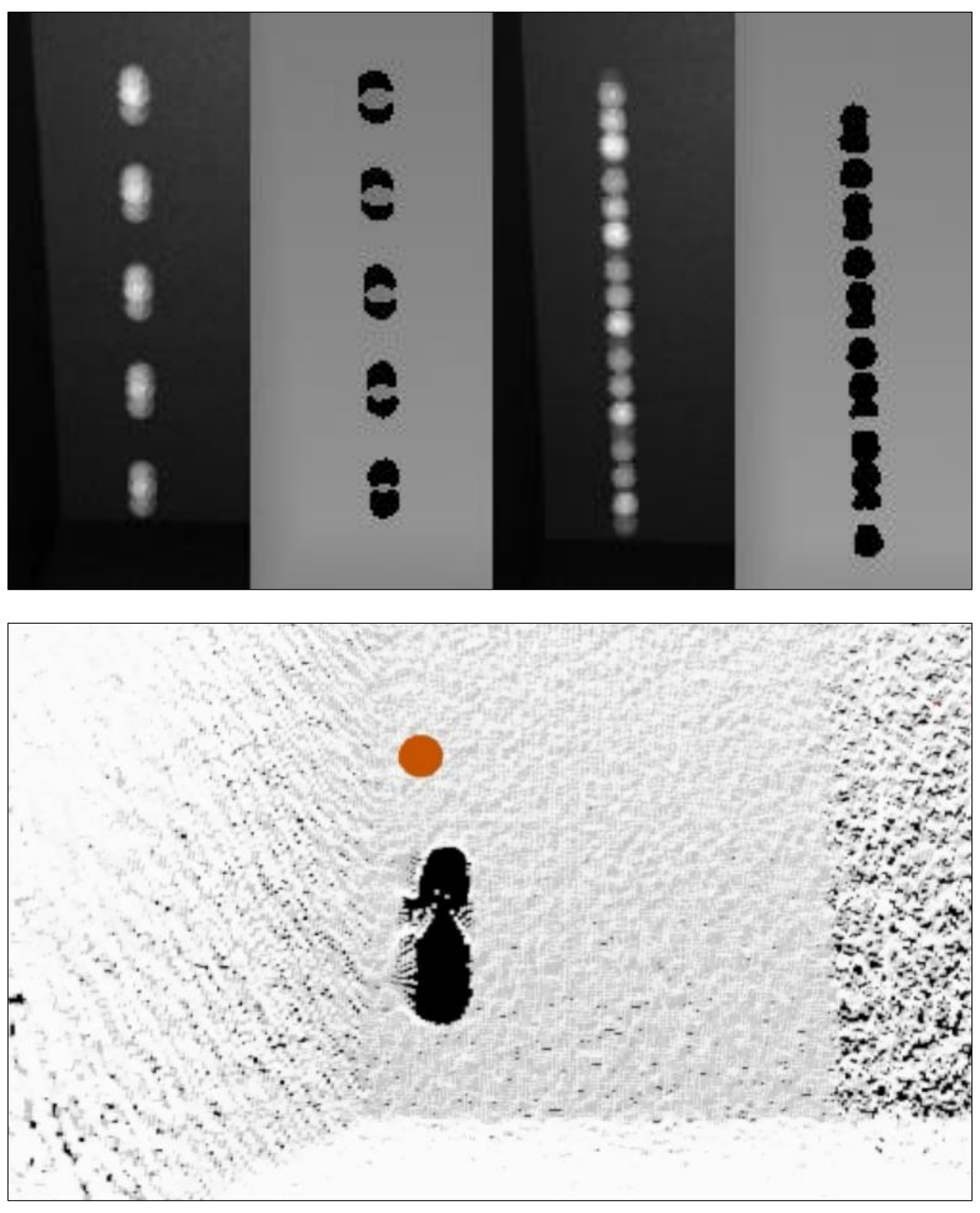
$$R_i = \frac{\rho}{d^2}\,S_i(d) + \epsilon_i\ ,$$

where $d > 0$ is the depth of the imaged surface and $\rho > 0$ is the surface *albedo*.

The ideal responses are dependent on modulation frequency and phase delay and are given by an idealized calibrated response curve [3]:

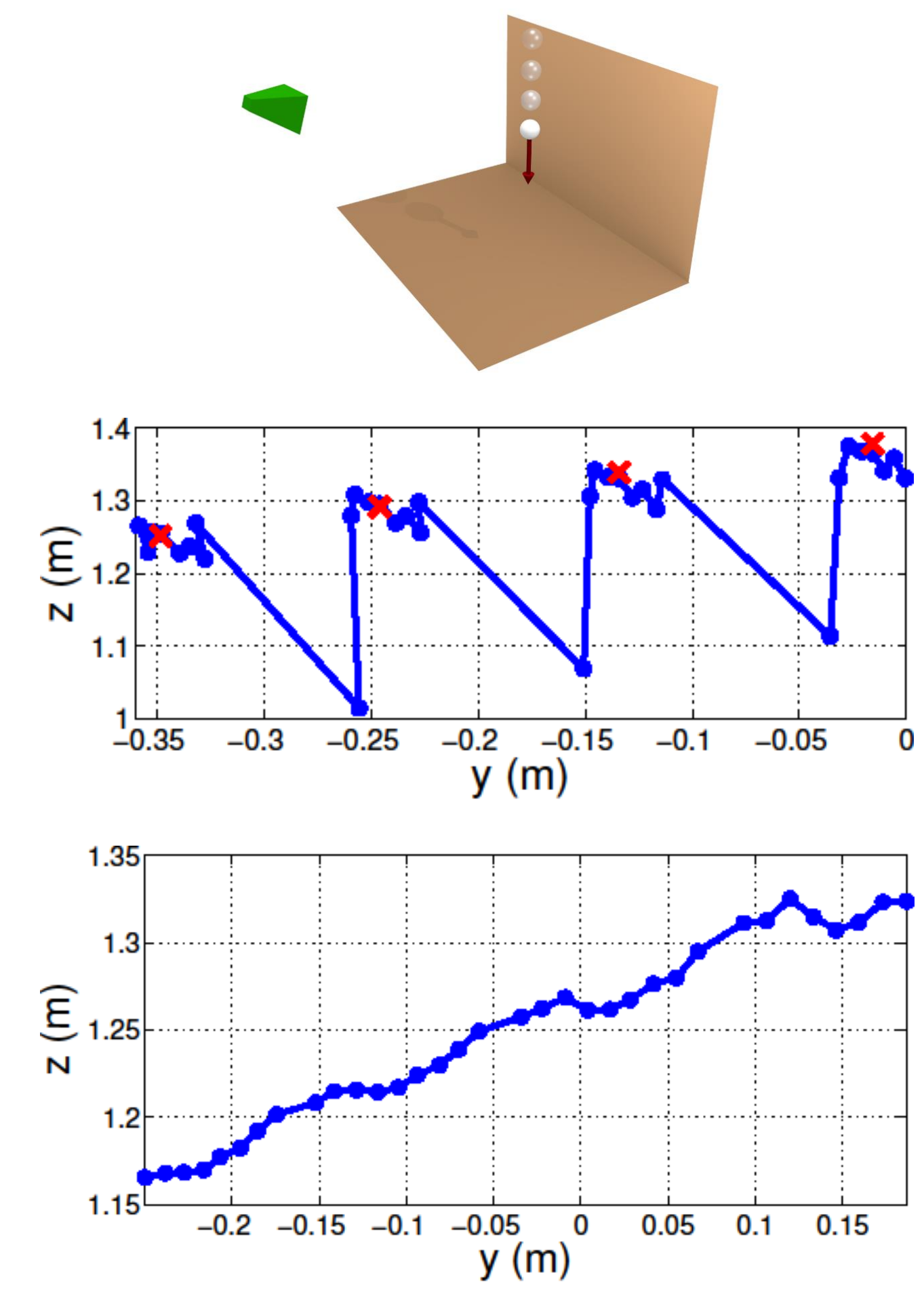
$$S_i : [d_{min}, d_{max}] \rightarrow \{-I_{min}, I_{max}\}$$

Failure of Kinect Depth Tracking



The model-based tracking approach allows to track the object even when the standard depth reconstruction fails due to fast motion.

Benefit of Equispacing



Equispaced exposure timing (bottom) leads to a more stable depth estimate when tracking from raw time of flight captures in comparison to standard exposure timing (top).
Red crosses correspond to depth values from the time of flight depth reconstruction engine and demonstrate the validity of our approach.

Comparison with a Commercial Motion Capture System

Exposure Timing	Object speed	RMSE	RMSE [mm]		
			x	y	z
clustered	1,33 km/h	16,3 mm	4,3	2,3	15,5
	2,12 km/h	19,3 mm	8,4	9,1	14,7
equidistant	1,01 km/h	15,5 mm	4,5	3,3	14,4
	2,32 km/h	23,7 mm	6,4	6,9	21,8

References

[1] M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. IJCV, 1998.

[2] N. J. Gordon, D. J. Salmond, and A. F. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings F, 1993.

[3] C. S. Bamji et al. A 0.13 μm CMOS system-on-chip for a 512 × 424 time-of-flight image sensor with multi-frequency photo-demodulation up 130 MHz and 2 GS/s ADC. J. Solid-State Circuits, 2015.

ICCV – International Conference on Computer Vision 2015

research.microsoft.com, vision.in.tum.de