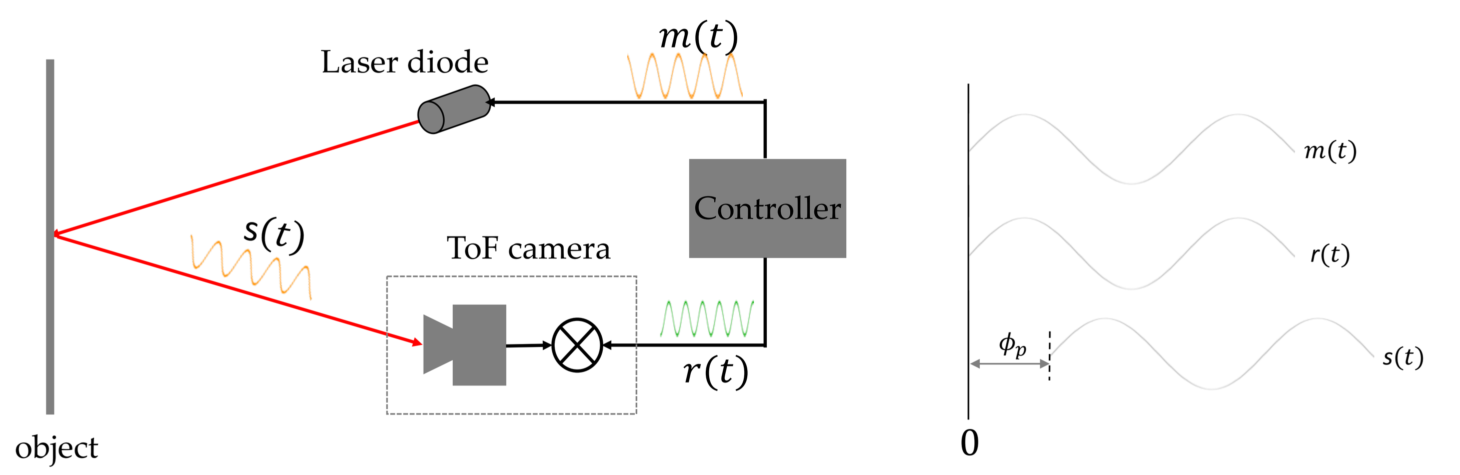
MULTI-FRAME SUPER-RESOLUTION FOR TIME-OF-FLIGHT IMAGING

Fengqiang Li[1,*], Pablo Ruiz[2], Oliver Cossairt[1], Aggelos K. Katsaggelos[2]
[1]Department of Computer Science, [2]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, 60208
*Email: fengqiang.li@u.northwestern.edu

INTRODUCTION & MOTIVATION

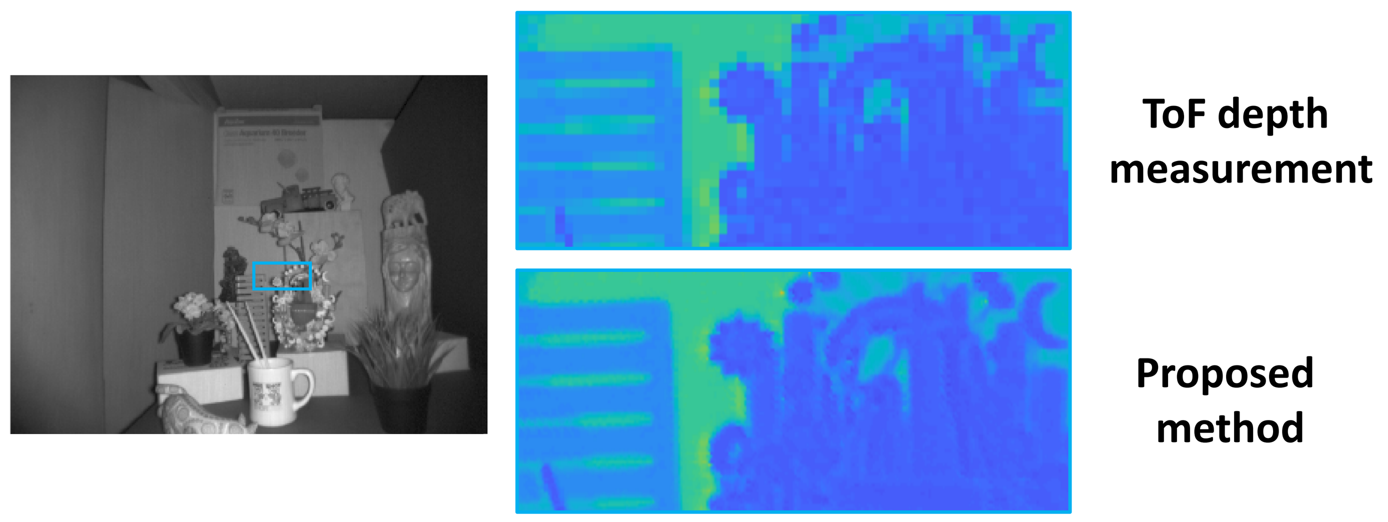
Time-of-flight (ToF) camera is an active 3D imaging technique which measures the phase delay between the reflection $s(t)$ and the reference $r(t)$ for depth estimation.



Why ToF camera for 3D imaging:

- ❑ Large baseline for comparable depth resolution (Triangulation based 3D imaging e.g. structure light camera).
- ❑ Compact size for ToF cameras.

Problems with current ToF camera and our motivation:



- ❑ Improve the lateral resolution of ToF camera measurements.

MULTI-FRAME SR FOR TOF IMAGING

Linear forward model for ToF:

- ToF camera outputs amplitude (a) and phase/depth (ϕ).
- Assume two ToF pixels p_1 and p_2 with corresponding amplitude and phase of (a_1, ϕ_1) and (a_2, ϕ_2).
- Represent ToF output and the object as a phasor

$$y(p) = a_p e^{i\phi_p}$$
$$x(P) = A_P e^{i\Phi_P}$$

where $\boldsymbol{y} \in \mathbb{C}^{M\times N}$ is the low resolution (LR) ToF measurement, $\boldsymbol{x} \in \mathbb{C}^{kM\times kN}$ is the high resolution depth image to be reconstructed, k is the magnification factor, p is one pixel of the LR ToF sensor, and P is one pixel of the HR reconstruction.

Multi-frame SR for ToF imaging:

- A sequence of L LR frames acquired by ToF camera as column vectors

$$\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_L \in \mathbb{C}^{MN\times 1}$$

- Assume the HR frame we want to recover (column vector) is

$$\boldsymbol{X} \in \mathbb{C}^{k^2MN\times 1}$$

- We perform SR on the real and imaginary parts separately

$$\boldsymbol{y}_l = \boldsymbol{D}\boldsymbol{W}_l\mathbf{x} + \boldsymbol{n}_l$$
$$\boldsymbol{y}_l \in \{\mathrm{Re}\,(\boldsymbol{Y}_l), \mathrm{Im}(\boldsymbol{Y}_l)\}$$
$$\mathbf{x} \in \{\mathrm{Re}(\boldsymbol{X}), \mathrm{Im}(\boldsymbol{X})\}$$

where $\boldsymbol{D} \in \mathbb{R}^{MN\times k^2MN}$ is a known downsampling matrix. $\boldsymbol{W}_l \in \mathbb{R}^{k^2MN\times k^2MN}$ is a warping matrix which models the movement between the HR image $\mathbf{x}$ and the LR version of l-th frame $\boldsymbol{y}_l$. $\boldsymbol{n}_l \in \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I})$ is additive noise term. We assume the warping matrices are not known and they are estimated jointly with the HR image.

BAYESIAN MODELING

- The probability distribution on observations is as

$$\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{W}, \beta) \propto \beta^{\frac{LMN}{2}} \exp\left\{-\frac{\beta}{2}\sum_{l=1}^{L} ||\boldsymbol{y}_l - \boldsymbol{D}\boldsymbol{W}_l\mathbf{x}\,\|^2\right\}$$

where $\boldsymbol{y} = \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_L$ and $\mathbf{W} = \boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_L$

- The prior is as

$$\mathrm{p}(\mathbf{x}|\alpha) \propto \alpha^{NM/2}\exp\{-\alpha\mathrm{TV}(\mathbf{x})\}$$

where $\mathrm{TV}(\mathbf{x}) = \sum_{j=1}^{MN}\sqrt{\left(\Delta_j^h\mathbf{x}\right)^2 + \left(\Delta_j^v\mathbf{x}\right)^2}$.

- The posterior distribution of the un-knowns given the observations is as

$$\mathrm{p}(\mathbf{x}, \mathbf{W}, \alpha, \beta|\mathbf{y}) = \frac{\mathrm{p}(\boldsymbol{y}|\mathbf{W}, \beta)\mathrm{p}(\mathbf{x}|\alpha)\prod_{l=1}^{L}\mathrm{p}(\boldsymbol{W}_l)\mathrm{p}(\alpha)\mathrm{p}(\beta)}{\mathrm{p}(\mathbf{y})}$$

where $\mathrm{p}(\boldsymbol{W}_l)$, $\mathrm{p}(\alpha)$, and $\mathrm{p}(\beta)$ are non-informative flat priors.

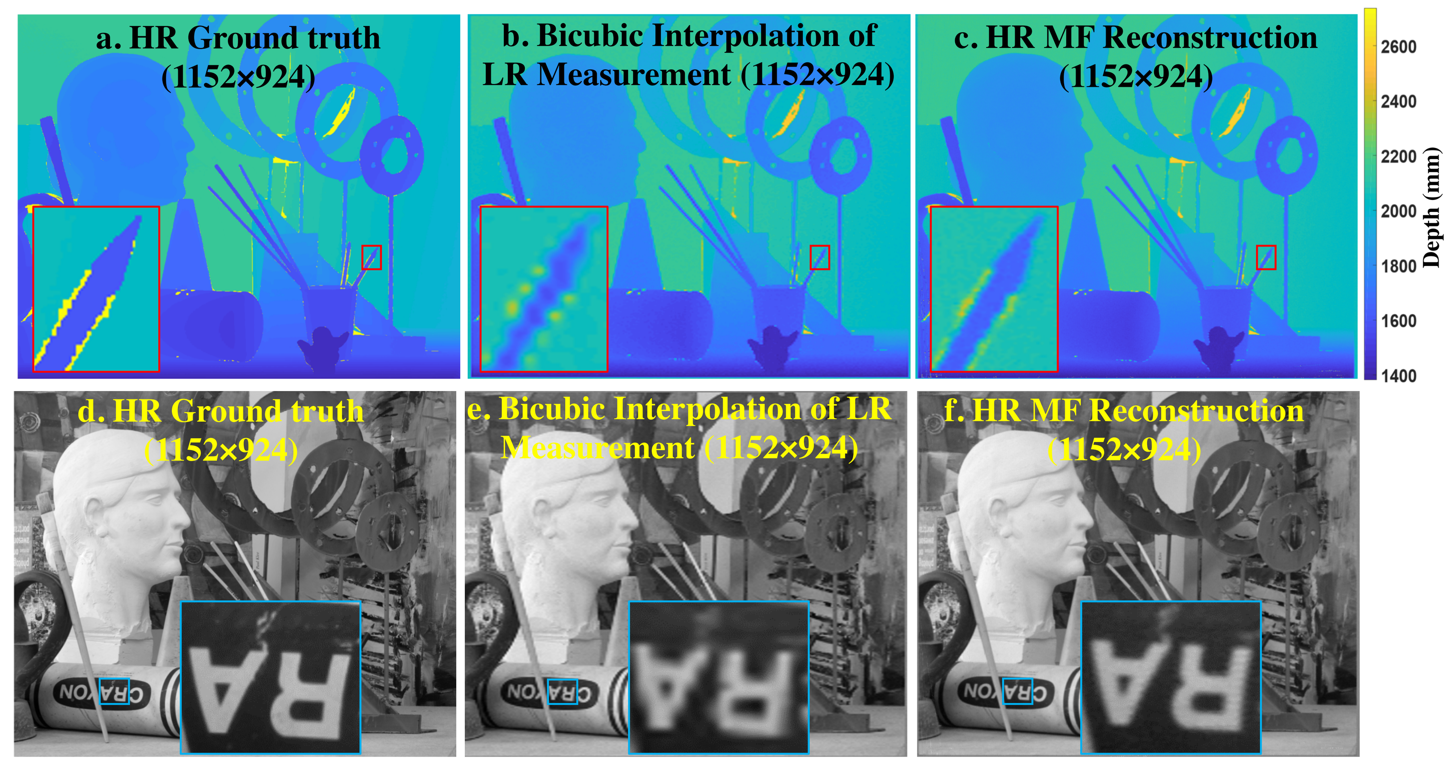
- Since we cannot explicitly calculate $\mathrm{p}(\mathbf{y})$, we do not have access to the posterior distribution. We apply the variational Bayesian analysis to approximate posterior distribution $\mathrm{p}(\Theta|\mathbf{y})$ by a tractable distribution $\mathrm{q}(\Theta)$. This approximating distribution is estimated by minimizing the Kullback-Leibler (KL) divergence between $\mathrm{p}(\Theta|\mathbf{y})$ and $\mathrm{q}(\Theta)$ as

$$\mathrm{p}(\mathbf{x}, \mathbf{W}, \alpha, \beta|\mathbf{y}) \approx \mathrm{q}(\mathbf{x})\prod_{l=1}^{L}\mathrm{q}\,(\boldsymbol{W}_l)\mathrm{q}(\alpha)\mathrm{q}(\beta)$$

- Finally, the HR image $\mathbf{x}$ is estimated as the mean of $\mathrm{q}(\mathbf{x})$.

SIMULATION

- GT: 1154 × 912 pixels. LR measurements: 288 × 231 pixels
- Warping matrix is acquired by shifting and rotating
- Gaussian noise of 30 dB is added



	Bicubic	Proposed Method
Intensities (PSNR)	17.9 dB	21.1 dB
Depths (RMSE)	30.3 cm	20.6 cm
Time	< 1 minute	15 minutes

REAL EXPERIMENT

- Texas Instruments ToF camera: 320 × 240 pixels
- 25 frames acquired by ToF camera are used for reconstruction