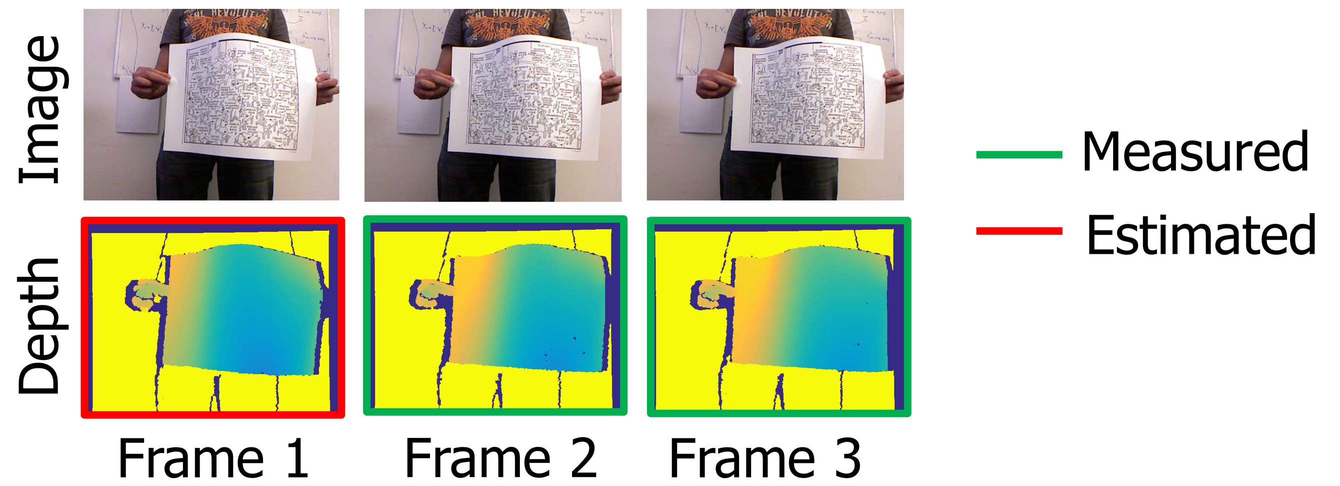
Depth Estimation of Non-Rigid Objects for Time-of-Flight Imaging

James Noraky, Vivienne Sze
Massachusetts Institute of Technology

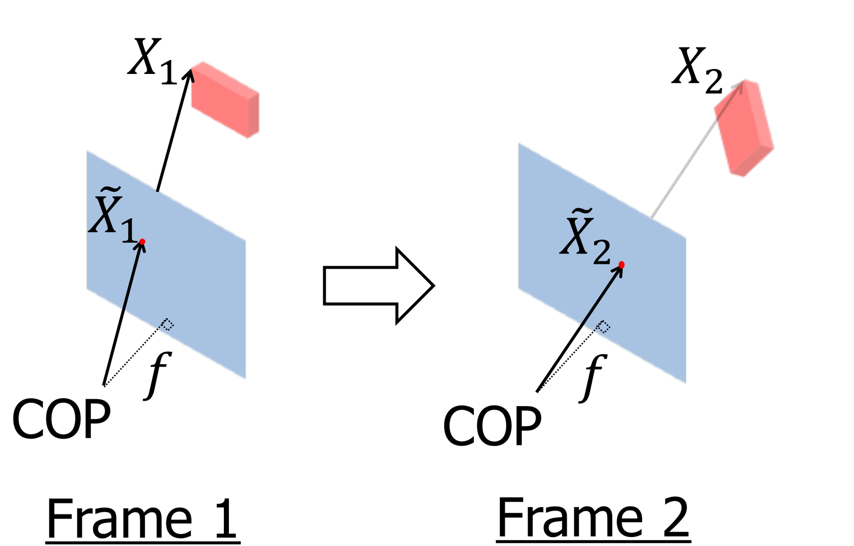
Motivation

- Time-of-flight (TOF) cameras are useful for many applications
- Due to system power constraints or multi-camera inference, TOF cameras cannot always acquire depth

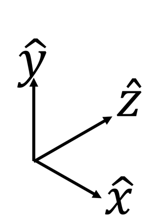


Our algorithm extends our work in [1] to estimate the depth for **non-rigid objects** by assuming that they are locally rigid

Rigidity Assumption



From Frame 1 to 2, the patch undergoes rotation, R, and translation, T



COP = Center of Projection

Images are formed by **perspective projection**:

$$\tilde{X}_1 = \frac{f}{\hat{z}\cdot X_1} X_1 \qquad \tilde{X}_2 = \frac{f}{\hat{z}\cdot X_2} X_2 = \frac{f}{\hat{z}\cdot X_2}(RX_1 + T)$$

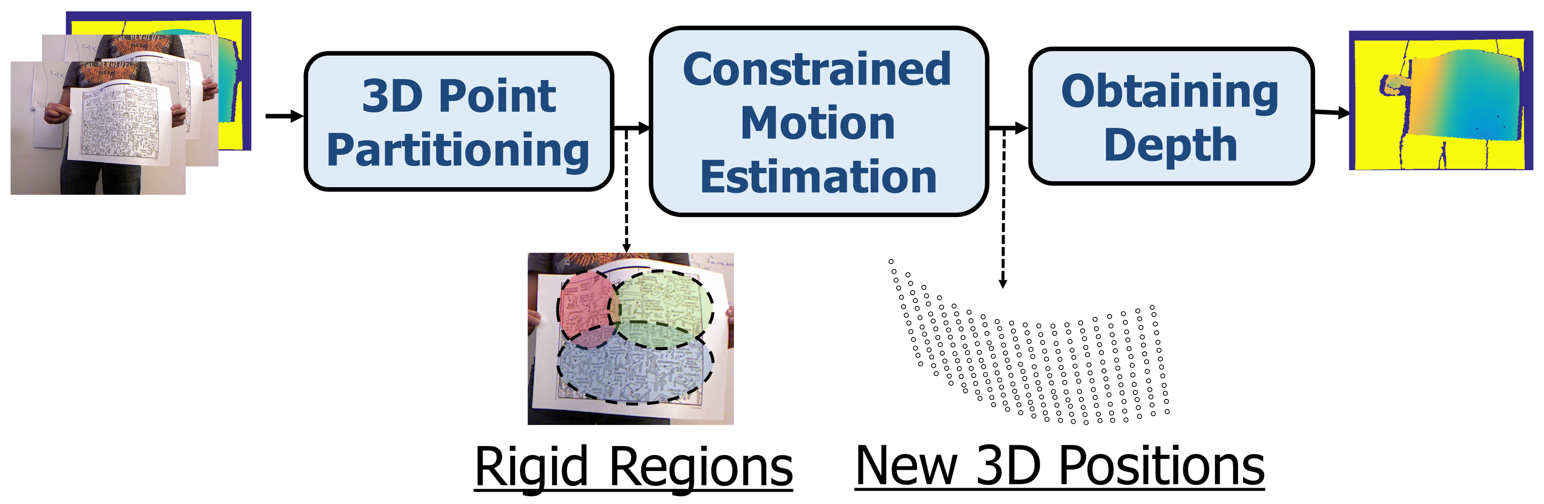
Approximate R with **angular velocity**, ω, because the time between frames is small:

$$\tilde{X}_2 = \frac{f}{\hat{z}\cdot X_2}(RX_1 + T) \approx \frac{f}{\hat{z}\cdot X_2}(X_1 + \omega \times X_1 + T)$$

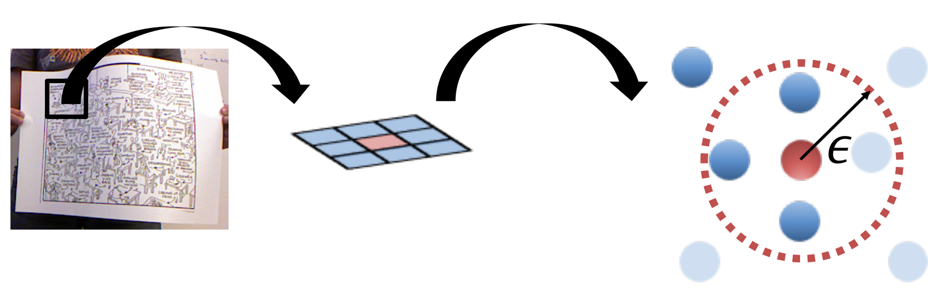
Exploit **collinearity**: $\boxed{\tilde{X}_2 \times (X_1 + \omega \times X_1 + T) = 0}$

The pixel wise motion of a locally rigid patch must follow the **rigidity assumption**

Non-Rigid Depth Estimation Algorithm

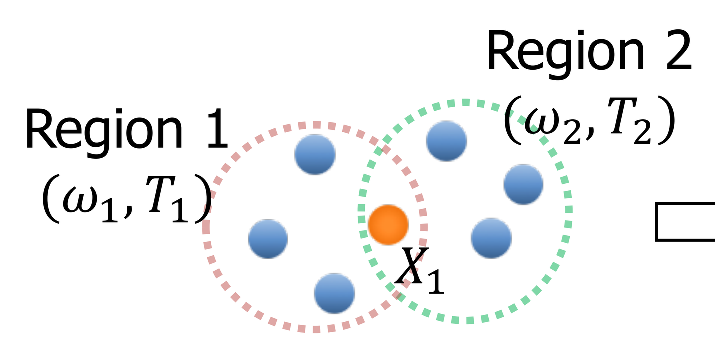


3D Point Partitioning: Group all rigid points together



Use the **rigidity assumption** along with RANSAC to identify rigid regions

Constrained Motion Estimation: Estimate the pose of each rigid region

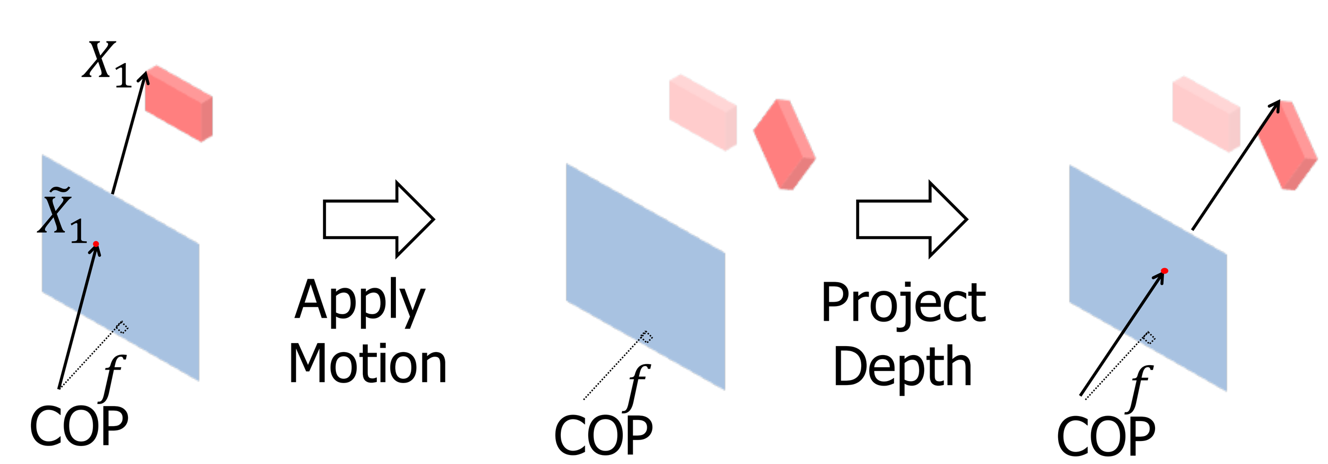


Consistency Constraint

$$\omega_1 \times X_1 + T_1 = \omega_2 \times X_1 + T_2$$

Solve **least squares** formulation that minimizes the rigidity assumption while maintaining the consistency constraint

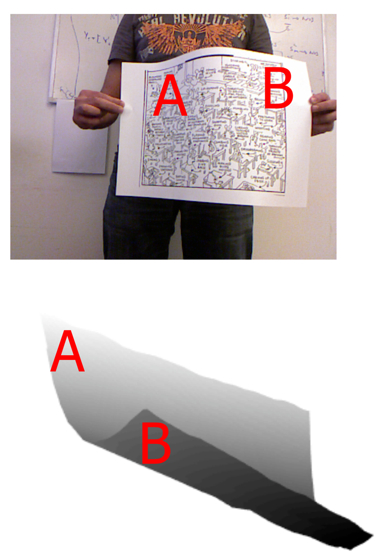
Obtaining Depth: Reproject each point and interpolate



Simplify depth estimation for non-rigid objects by using previous depth measurements

Algorithm Evaluation

- Sequentially estimate depth for our synthetic sequences and those in [2]; Evaluate with percent mean relative error (MRE)



Sequence	Frame 2	Frame 3	Frame 4	Mean
syn_bend	0.27	0.25	0.24	0.26
syn_crease	0.27	0.27	0.27	0.27
kinect_paper [2]	0.19	0.43	0.23	0.28
kinect_tshirt [2]	0.35	0.52	1.16	0.68
Mean	0.27	0.37	0.47	0.37

Fig. 1: *kinect_paper*

- Achieves **MRE of 0.48%** for [2], outperforming NRSFM approaches surveyed in [3] (MRE of 3.71%)
- Run time on standard laptop: **0.06 seconds** vs minutes for techniques in [3]

Fig. 2: Using RANSAC to partition points preserves the structure in the depth map of *syn_crease*

Key Contribution: Estimate depth maps with a mean relative error of 0.37% (0.48% for sequences in [2])

References

[1] James Noraky and Vivienne Sze, "Low power depth estimation for time-of-flight imaging," in 2017 IEEE International Conference on Image Processing, 2017, pp. 2114–2118.

[2] Aydin Varol, Mathieu Salzmann, Pascal Fua, and Raquel Urtasun, "A constrained latent variable model," in Conference on Computer Vision and Pattern Recognition, 2012, pp. 2248–2255.

[3] Suryansh Kumar, Yuchao Dai, and Hongdong Li, "Monocular Dense 3D Reconstruction of a Complex Dynamic Scene from Two Perspective Frames," in International Conference on Computer Vision, 2017, pp. 4649– 4657.

We thank Analog Devices for funding this work