Perception and Robotics Group

Fast Active Monocular Distance Estimation from Time-to-Contact

Levi Burner, Nitin J. Sanket, Cornelia Fermüller, Yiannis Aloimonos

Department of Electrical and Computer Engineering, A. James Clark School of Engineering, University of Maryland, College Park

MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

Introduction

- Distance estimation is required for a variety of robotic applications including navigation, manipulation and planning
- Inspired by the mammal's visual system, which gazes at specific objects and estimates when the object will reach it [1], we develop a novel constraint between time-to-contact, acceleration, and distance
- The result allows a monocular camera and IMU to accurately measure distance to a visual patch while ignoring the rest of the incoming image

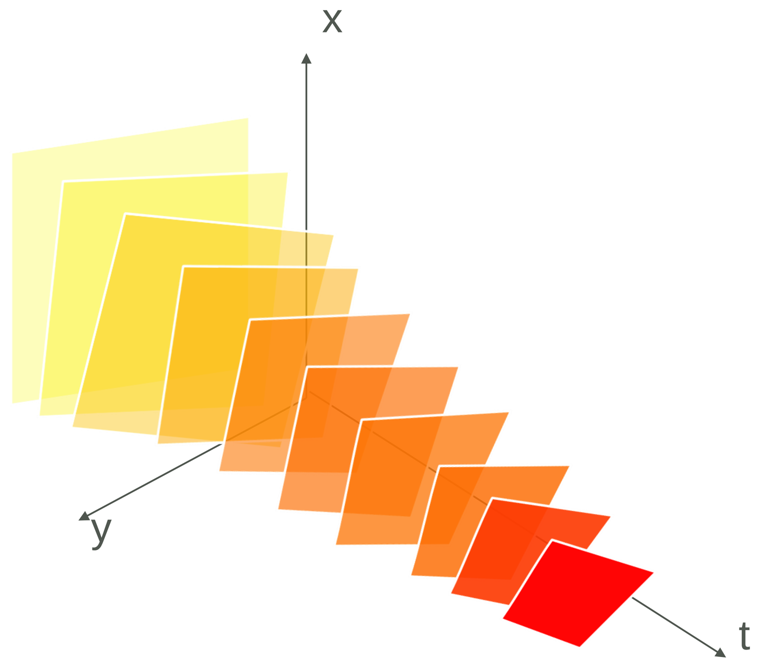

Figure 1. A visual patch being tracked over time changes in shape and location in a way that can be modelled by an affine transformation. The parameters of the transformation determines time-to-contact directly. The tau-constraint uses time-to-contact to efficiently estimate the distance to the tracked patch.

Contributions

- A closed form solution for estimating the 3D position of the camera given a short history of time-to-contact and acceleration that we call the tau-constraint
- A computationally efficient pipeline for 3D position estimation using the tau-constraint, which demonstrates viability with inexpensive sensors
- Comparisons against the popular VIO methods, VINS-Mono [2] and ROVIO [3], as well as the fiducial marker based posed estimation method, AprilTag 3 [4]

Results

We redefine time-to-contact as frequency-of-contact with the relation and generalized to 3 dimensions using by defining the quantity:

$$\mathbf{F}(t) := \frac{\dot{\mathbf{X}}(t)}{Z(t)}$$

X is the relative position between the camera and Z is distance along the optical axis.

Theorem (tau-constraint). If **F**(t) and $\ddot{\mathbf{X}}$(t) are known for all t in a closed interval, and Z(t) ≥ ε > 0, then the following constraint between depth Z(0), frequency-of-contact **F**, and acceleration $\ddot{\mathbf{X}}$ holds for each point on the scene patch.

$$Z(0)\mathbf{E}(t) = \mathbf{\Delta}\{\ddot{\mathbf{X}}\}(t)$$

Where,

$$\mathbf{E}(t) := \begin{bmatrix} \int_0^t \begin{bmatrix} F_X(\lambda) \\ F_Y(\lambda) \end{bmatrix} \Phi_{F_Z}(\lambda) d\lambda \\ \Phi_{F_Z}(t) - 1 \end{bmatrix} - t\mathbf{F}(0)$$

$$\Phi_{F_Z}(t) := \exp\left(\int_0^t F_Z(\lambda) d\lambda\right)$$

$$\mathbf{\Delta}\{\mathbf{f}\}(t) := \int_0^t \left(\int_0^\lambda \mathbf{f}(\lambda_2) d\lambda_2\right) d\lambda$$

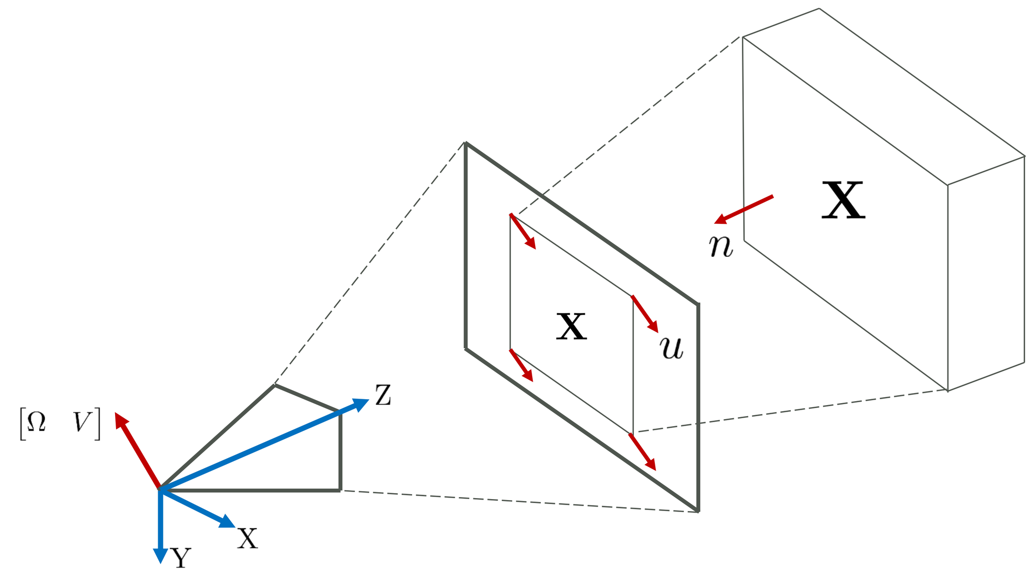

Figure 2. Camera coordinate system used to define frequency-of-contact.

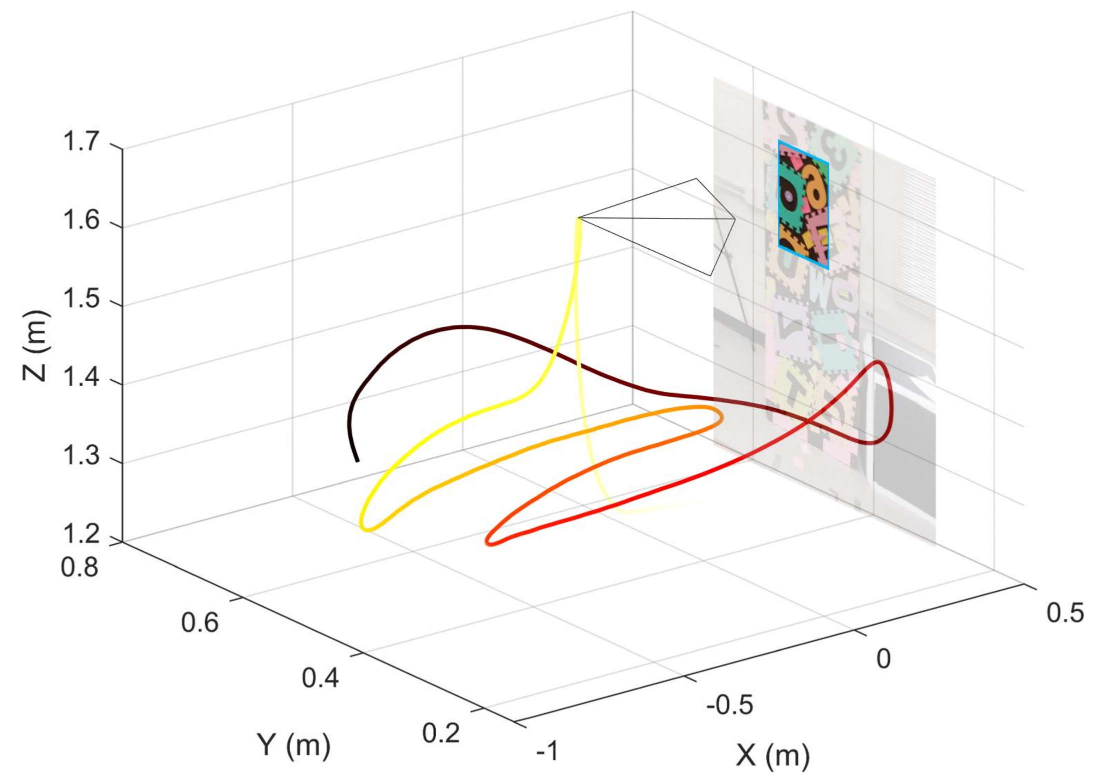

Figure 3. Five seconds of a typical sequence along with the patch fixated on. Ten sequences were recorded with ground from a Vicon system.

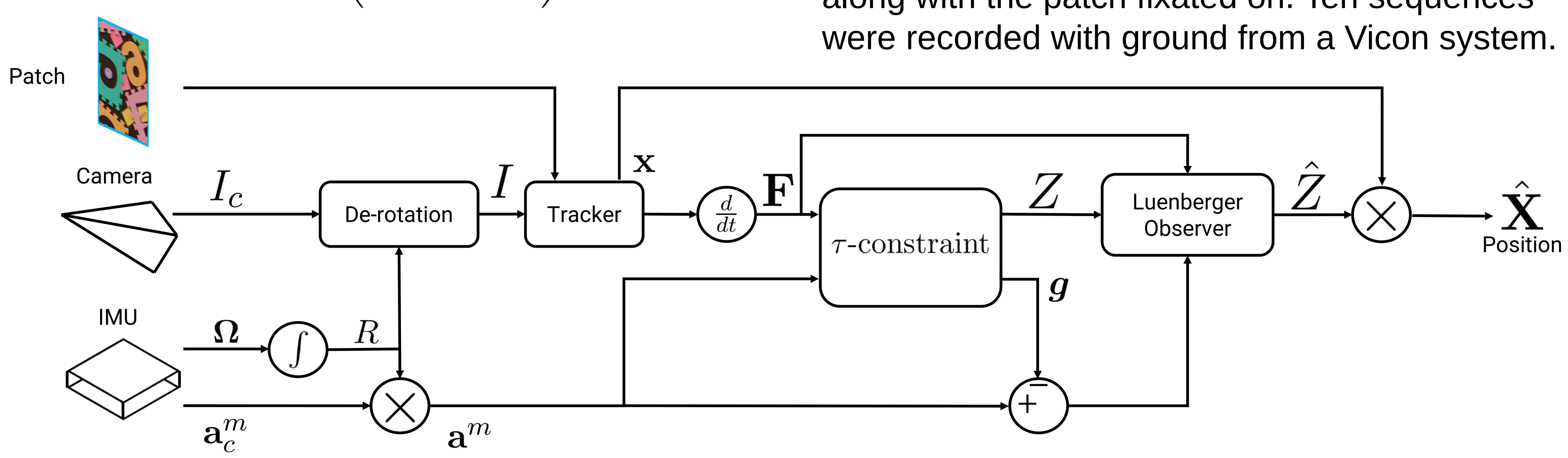

Figure 4. System overview of our method to estimate camera position using the tau-constraint. The output of the tau-constraint block is the solution to: $\underset{Z(0),g_Z}{\operatorname{argmin}} \|E_Z Z(0) + \Delta\{a_Z^m\} + \Delta\{\mathbf{1}_{[0,T]}\} g_Z\|_2^2$

Seq.	1	2	3	4	5	6	7	8	9	10
Sequence Duration (s)	14.99	26.14	32.22	36.23	16.35	16.27	8.02	32.15	26.75	40.07
Distance Traveled (m)	15.71	29.65	22.14	34.75	15.56	15.78	7.30	26.75	21.40	35.36
Method					ATE (cm) ↓					
AprilTag 3 [20]	**2.88**	-	**2.66**	-	**3.77**	-	**0.66**	-	2.55	-
VINS-Mono [27]	5.39	8.79	14.21	15.37	-	6.11	1.15	18.45	13.07	4.34
ROVIO [7, 8]	7.55	9.89	11.86	33.23	29.96	**2.84**	0.69	**3.93**	16.62	3.79
τ-constraint (ours)	8.06	6.91	11.96	**10.27**	16.80	7.21	10.70	4.34	**2.38**	**3.24**
τ-constraint (ours) (30 Hz LK output)	7.75	**6.59**	12.43	15.04	17.40	8.18	5.91	4.75	3.63	3.49
τ-constraint (ours) (15 Hz LK output)	9.81	7.59	16.17	19.29	14.96	8.70	5.86	5.37	3.47	3.48

Table 1. Sequence duration, path length (meters), and centimeters of Average Trajectory Error (ATE). Overall average ATE: **tau-constraint 8.5cm**, VINS-MONO 12.2 cm, ROVIO 16.9 cm

Conclusions

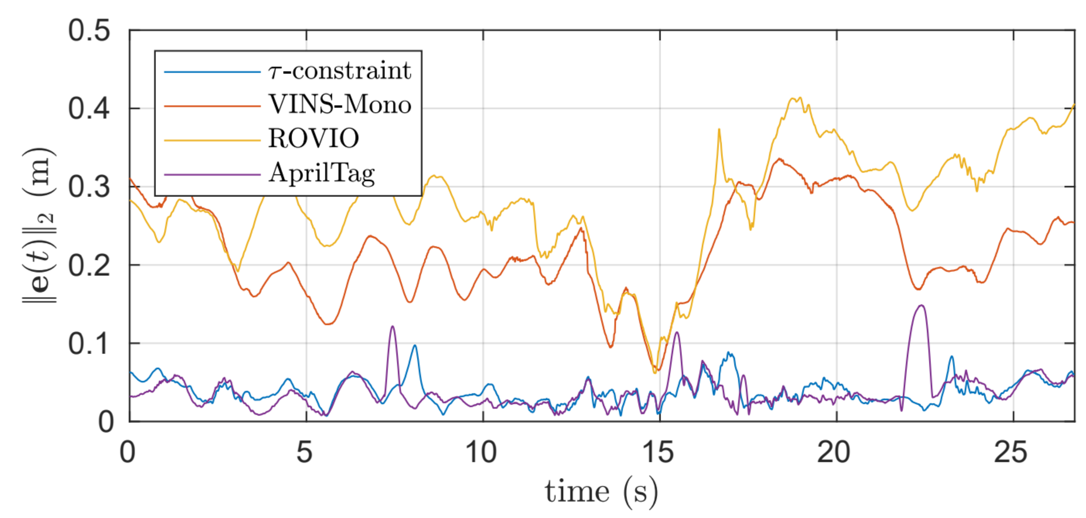

Figure 5. L2 error of all methods on for Seq. 9. On this sequence our accuracy exceeds even AprilTag 3's.

- Our method ran 6.8x to 27x faster than two state-of-the art VIO methods that would typically be used for the same task, while achieving 30% to 50% less ATE
- This speedup is largely due to using only a small portion (initially 2.5%) of the image
- Development of the tau-constraint, in theory and practice, is a promising direction for VIO, VI-SLAM, active perception, and robotics

References

[1] Ruzena Bajcsy, Yiannis Aloimonos, and John K. Tsotsos. Revisiting active perception. Autonomous Robots, 42:177–196, 2018.

[2] Tong Qin, Peiliang Li, and Shaojie Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics, 34(4):1004–1020, 2018*

[3] Michael Bloesch, Sammy Omari, Marco Hutter, and Roland Siegwart. Robust visual inertial odometry using a direct EKF-based approach. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 298–304, 2015.

[4] Maximilian Krogius, Acshi Haggenmiller, and Edwin Olson. Flexible layouts for fiducial tags. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* pages 1898–1903, 2019

[5] Levi Burner, Nitin J Sanket, Cornelia Fermüller, and Yiannis Aloimonos. Fast active monocular distance estimation from time-to-contact. arXiv preprint arXiv:2203.07530, 2022

Acknowledgments

The support of the NSF under awards DGE-1632976 (COMBINE) and OISE 2020624 (AccelNet) is gratefully acknowledged.

COMBINE Computation and Mathematics for Biological Networks