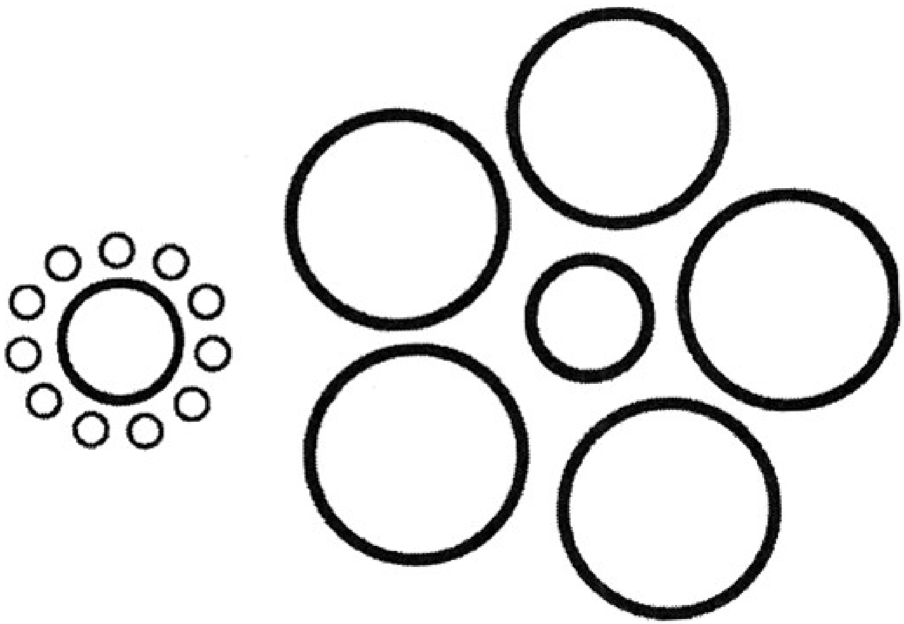
Dealing with Scale

Jet Propulsion Laboratory
California Institute of Technology

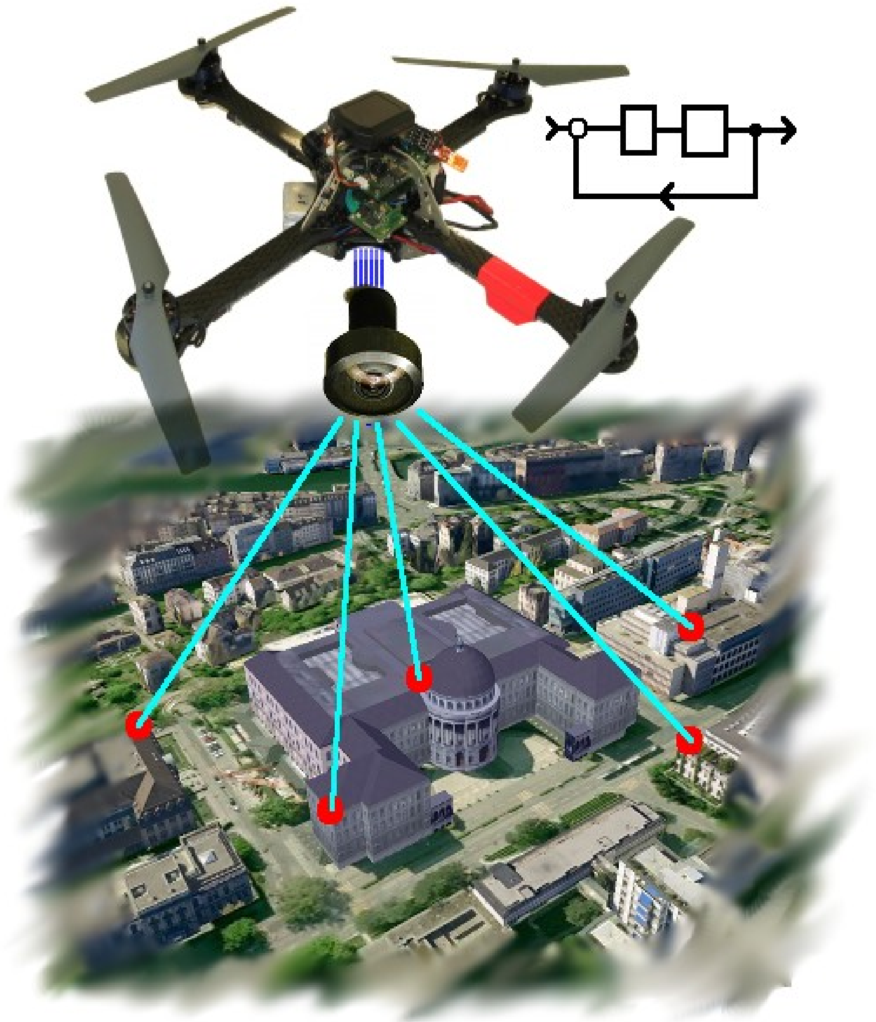
Stephan Weiss
Computer Vision Group
NASA-JPL / CalTech

Stephan.Weiss@ieee.org



Outline

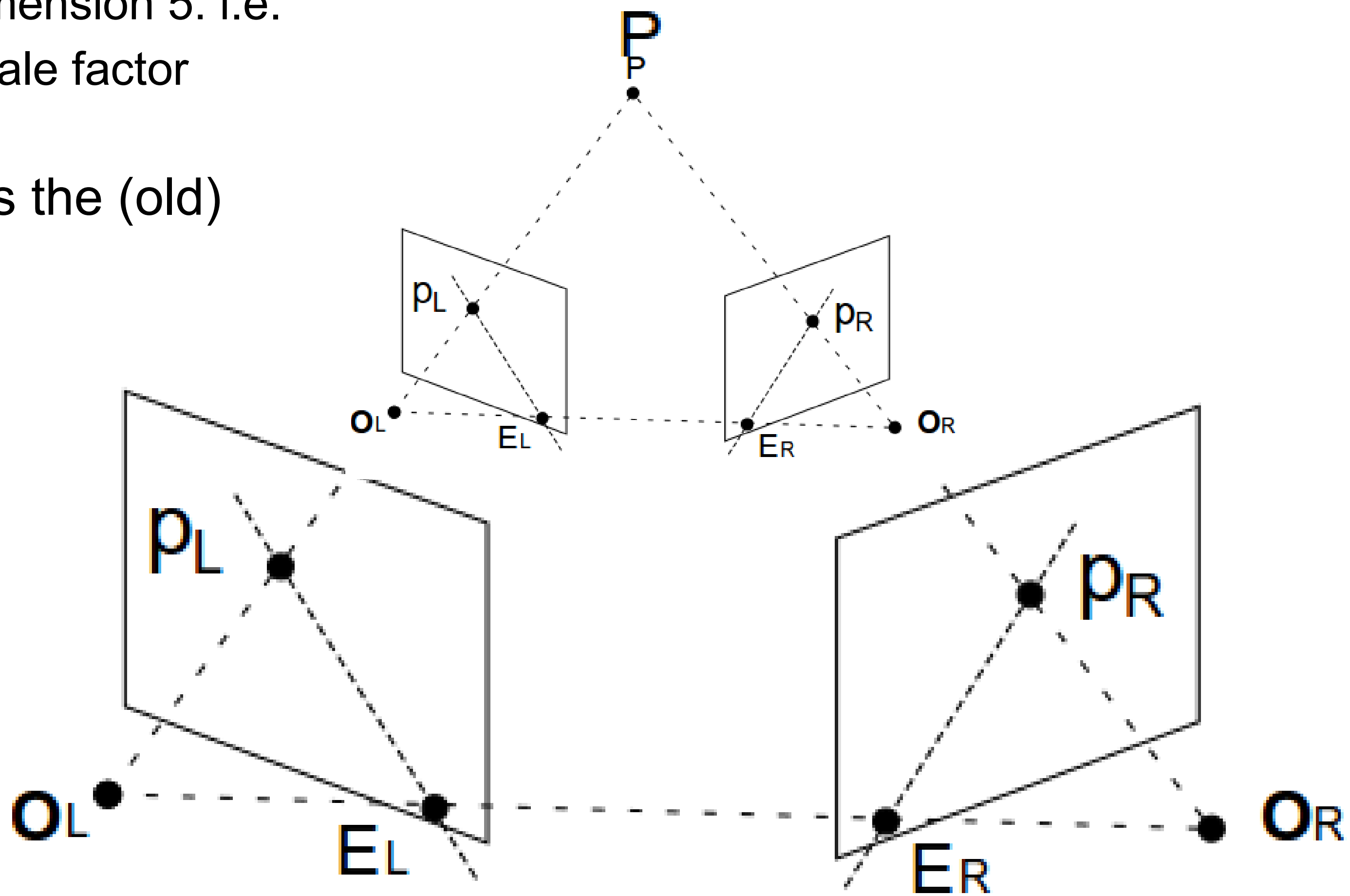
- Why care about size?
- The IMU as scale provider:
  - The idea: closed form
  - Tightly vs. Loosely coupled
  - Variable scale in optical flow



Recap: Monocular visual odometry is up to scale

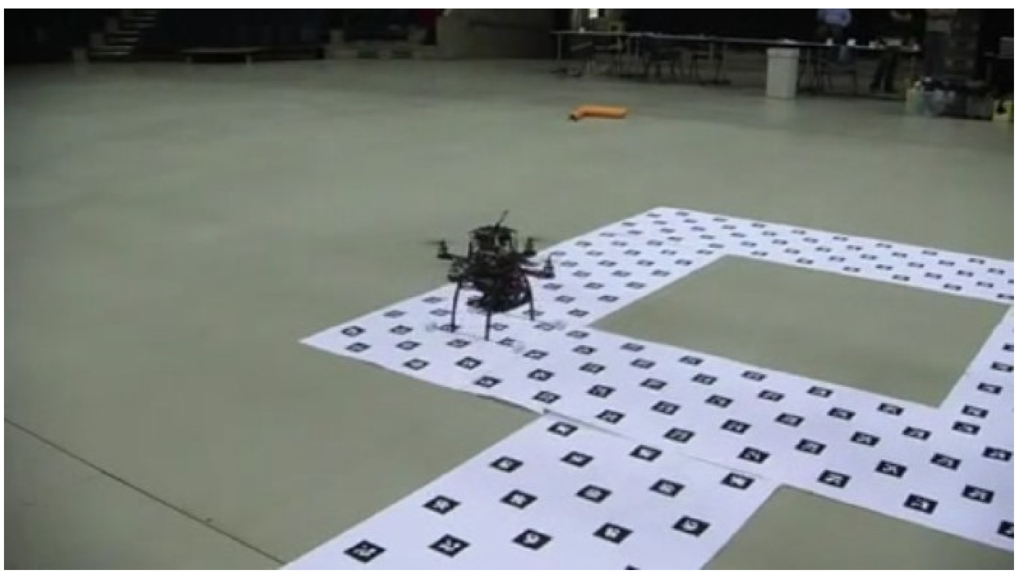
- If the 3D coordinates of P are unknown and we measure only is projection in 2 images:
  - Compute the essential matrix using 5 or more points
  - Problem is of dimension 5: i.e. up to a global scale factor

- This is the same as the (old) Hollywood effect

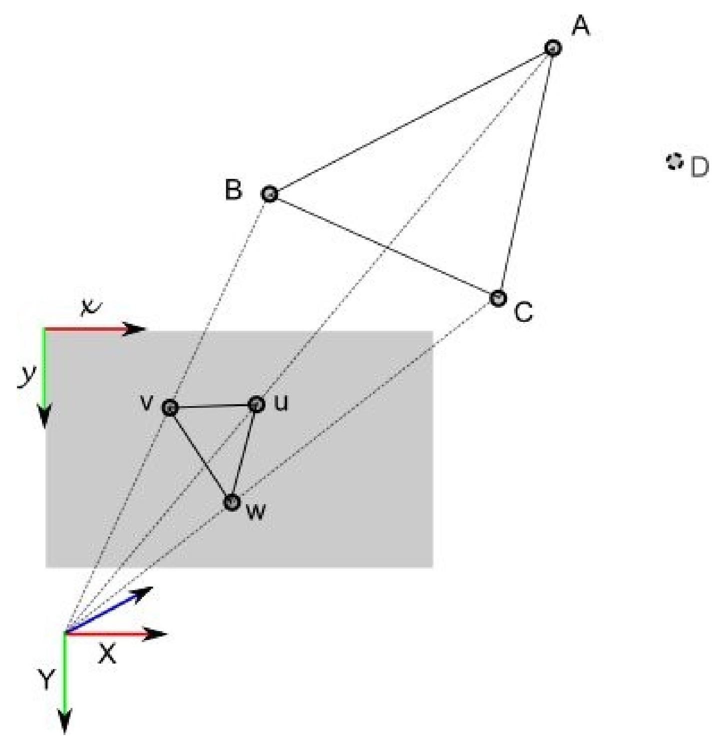


Recap: Monocular visual odometry is up to scale

- Recovering scale if the 3D points are known
  - p3p problem (e.g. stereo vision)
  - Requires known markers
  - Popular: AR Toolkit, APRIL Tags



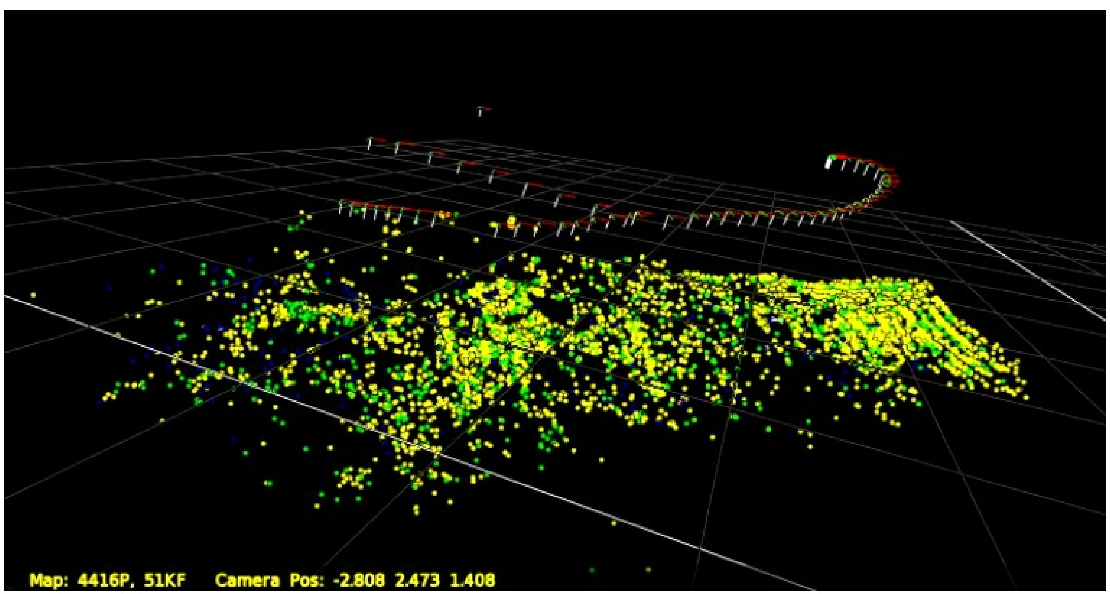
[Meier et al. EMAV 2010]



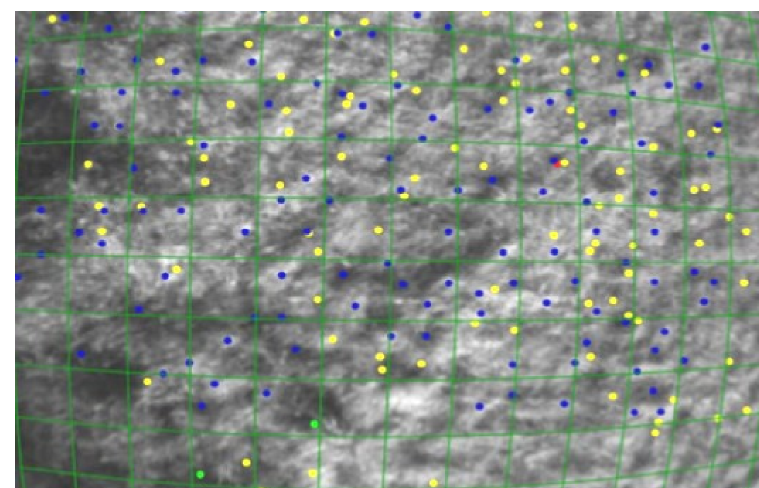
Recap: Monocular visual odometry is up to scale

- Recovering scale if the 3D points are known
  - p3p problem (e.g. stereo vision)
  - Requires known markers
  - Popular: AR Toolkit, APRIL Tags

- Typically, no known markers in the real world



[Weiss et al. JFR 2013]



typical outdoor scene

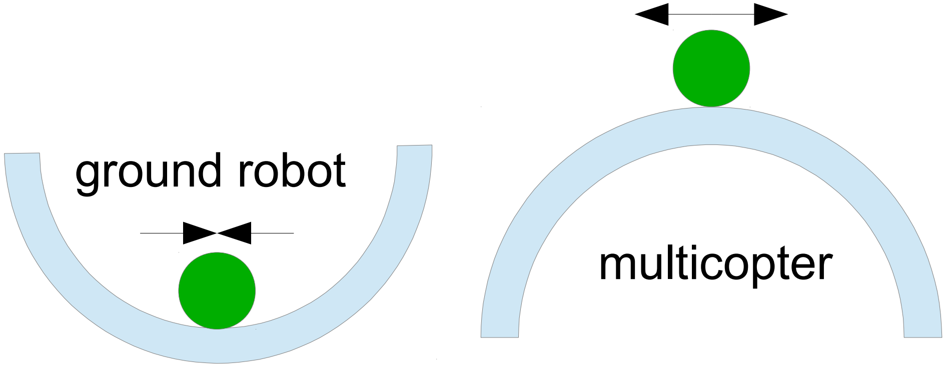
Why Care About Size?

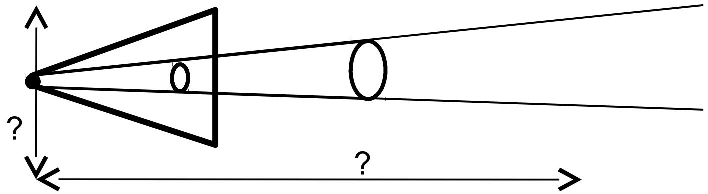
- Metric information used for robot control
- Parameter tuning necessary if no scale available
  - Can change from run to run arbitrarily
  - Can be constraint up to certain level (e.g. Mahony et al)
  - Usually impractical for fast deployable platforms

[Herisse T-RO 2012]

Why Care About Size?

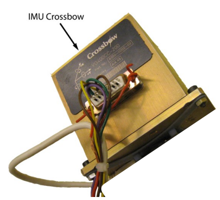
- Metric information used for robot control
- Parameter tuning necessary if no scale available
  - Can change from run to run arbitrarily
  - Can be constraint up to certain level (e.g. Mahony et al)
  - Usually impractical for fast deployble platforms
- Ground robots less affected than aerial/unstable platforms
  - Can hold actuators to enter stable regime





IMU as a Scale Provider

- Accelerometers have metric information and provide information at high rate
- Pure IMU integration drifts quadratically in time
  - High grade IMUs can be integrated over long periods
  - Cost effective robots usually carry MEMS IMUs:
    very noisy and bias drifts relatively fast
- VO only drifts spatially (when observing new features)
  - Intelligent combination of IMU and visual odometry (VO) can provide scale
    (and more!)



Inertial Sensor (IMU)

Visual Sensor (camera)

+ Fast sampling
+ Scaled units

– Biased, noisy data
– No vel. nor pos.
– Large drift

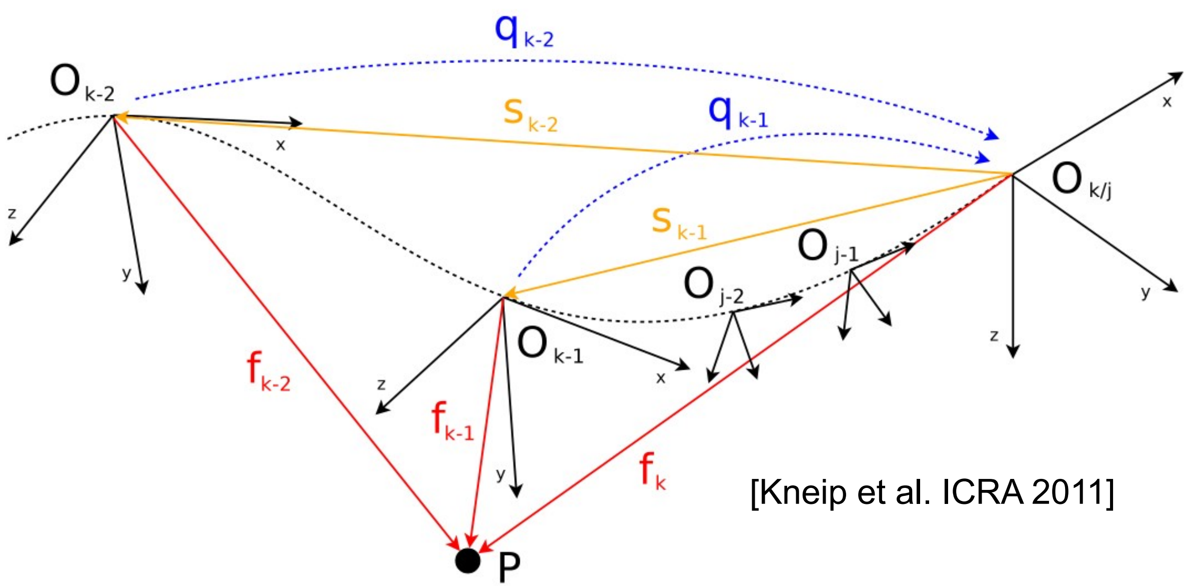
– Slow sampling
– unscaled units

+ position and attitude
+ slow spatial drift

IMU as a Scale Provider: The Idea

- Idea for combining IMU and VO to retrieve scale
  - Compare integrated IMU data to VO: problem with drift
  - Compare differentiated VO data to IMU: increased noise
- Methods:
  - Closed form least squares solution over short periods
    - Noisy, fails if motion is not sufficient
    - No scale propagation
    - Good for initialization
  - Probabilistic approach (e.g. GTSAM, iSAM, EKF)

$$\Lambda \leftarrow\rightarrow \frac{\int acclerometer}{\frac{d}{dt} VO\ position}$$

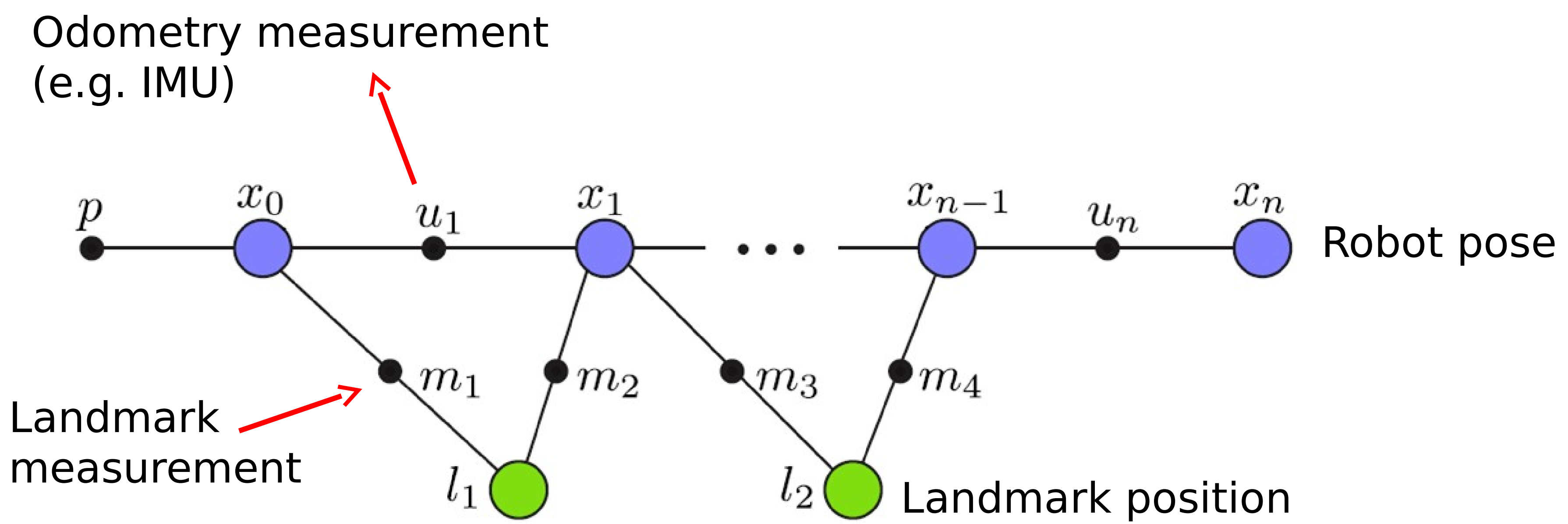


[Kneip et al. ICRA 2011]

IMU as a Scale Provider
A Probabilistic approach

Jet Propulsion Laboratory
California Institute of Technology

- BA setup using IMU readings as edges in a graph: GTSAM, iSAM



- Problem complexity can grow rabidly with high rate IMU readings
  - Use pre-integrated IMU terms (Indelman et al. RAS2013)
  - Use continuous time batch optimization (Furgale ICRA 2012)

IMU as a Scale Provider
An EKF Approach

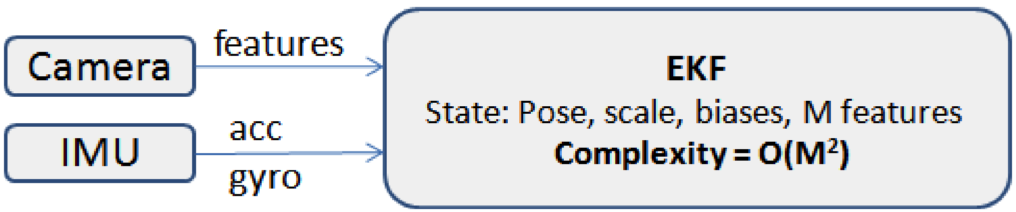
- EKF setup using the IMU in the motion model
- Do not use IMU in as a measurement/update
  - Would require an EKF update (i.e. matrix inversion) at IMU rate
  - Motion model without IMU is difficult: cannot model external disturbances
  - Process noise of motion model is difficult to assess

$$
\begin{aligned}
\dot{p}_w^i &= v_w^i \\
\dot{v}_w^i &= C_{(q_w^i)}^T (a_m - b_a - n_a) - g \\
\dot{q}_w^i &= \frac{1}{2}\Omega(\omega_m - b_\omega - n_\omega) q_w^i
\end{aligned}
$$

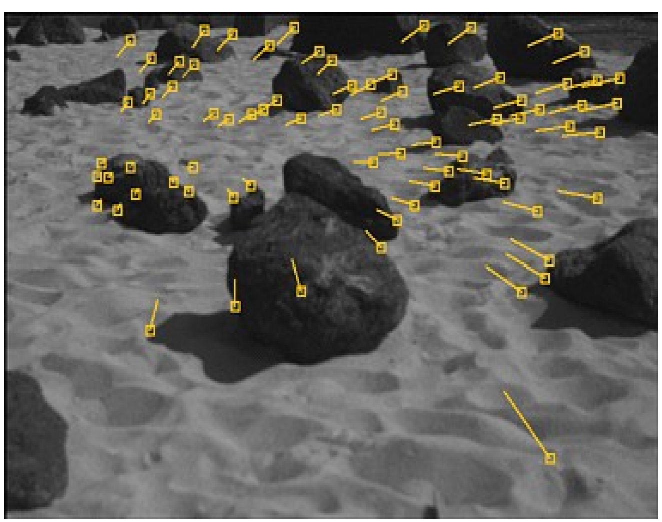
- How do we implement the update?

EKF Approach: Tightly Coupled

- Tightly coupled approach
  - Very high computational complexity
  - 3D features in the state need special handling



$$\mathbf{r}_i^{(j)} = \mathbf{z}_i^{(j)} - \hat{\mathbf{z}}_i^{(j)}$$



$$\hat{\mathbf{z}}_i^{(j)} = \frac{1}{{}^{C_i}\hat{Z}_j} \begin{bmatrix} {}^{C_i}\hat{X}_j \\ {}^{C_i}\hat{Y}_j \end{bmatrix}$$

$$\begin{bmatrix} {}^{C_i}\hat{X}_j \\ {}^{C_i}\hat{Y}_j \\ {}^{C_i}\hat{Z}_j \end{bmatrix} = \mathbf{C}({}^{C_i}_{G}\hat{\bar{q}})({}^{G}\hat{\mathbf{p}}_{f_j} - {}^{G}\hat{\mathbf{p}}_{C_i})$$

- Idea: reduce complexity by *not* including the 3D feature positions in the filter
  - Mourikis & Roumeliotis ICRA 2007

EKF Approach: Tightly Coupled

- Tightly coupled approach: remove feature positions from the state
  - Corrupts Gaussian noise assumption of EKF
  - Works well in practice
- Residuals $\mathbf{r}_i^{(j)} = \mathbf{z}_i^{(j)} - \hat{\mathbf{z}}_i^{(j)}$

$$\hat{\mathbf{z}}_i^{(j)} = \frac{1}{{}^{C_i}\hat{Z}_j} \begin{bmatrix} {}^{C_i}\hat{X}_j \\ {}^{C_i}\hat{Y}_j \end{bmatrix}$$

$$\begin{bmatrix} {}^{C_i}\hat{X}_j \\ {}^{C_i}\hat{Y}_j \\ {}^{C_i}\hat{Z}_j \end{bmatrix} = \mathbf{C}({}^{C_i}_{G}\hat{\bar{q}})({}^{G}\hat{\mathbf{p}}_{f_j} - {}^{G}\hat{\mathbf{p}}_{C_i})$$

- Multiplying with A = left nullspace of H_f

$$\mathbf{r}^{(j)} \simeq \mathbf{H}_{\mathbf{X}}^{(j)}\widetilde{\mathbf{X}} + \mathbf{H}_f^{(j)}\,{}^{G}\widetilde{\mathbf{p}}_{f_j} + \mathbf{n}^{(j)}$$

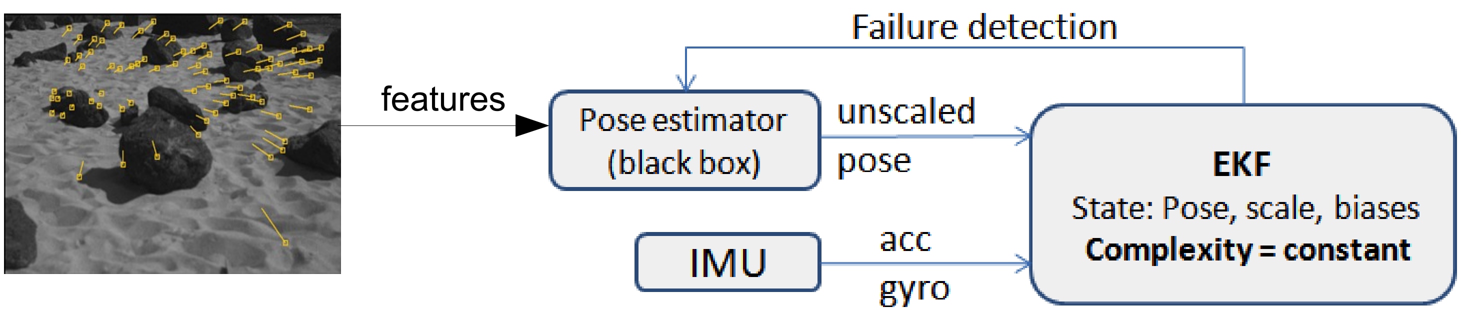
$$\mathbf{r}_o^{(j)} \simeq \mathbf{A}^T\mathbf{H}_{\mathbf{X}}^{(j)}\widetilde{\mathbf{X}} + \mathbf{A}^T\mathbf{n}^{(j)}$$

EKF Approach: Tightly Coupled

- Runs at more than 10Hz on a mobile phone processor
- Accurate scaling
- Works well even though large trajectories with low acceleration
  - Closed form solution would be difficult to use

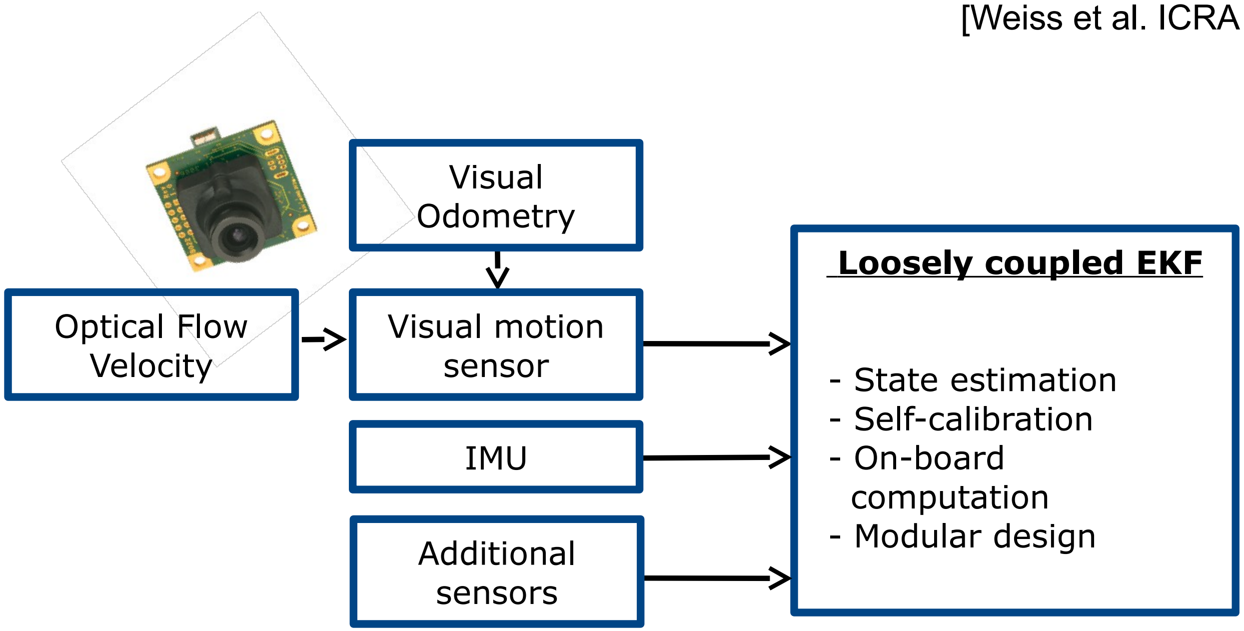
EKF Approach: Loosely Coupled

- Loosely coupled: faster, less complex, modular; but needs failure detection
  - Modularity: Use previously discussed VO/SLAM modules as black boxes
  - At the cost of filter consistency



Update includes a 6x6 matrix inversion (camera pose measurement)

[Weiss et al. ICRA 2011]



[Weiss PhD 2012]

EKF Approach: Loosely Coupled

Jet Propulsion Laboratory
California Institute of Technology

- Augment system model
  - Include visual scale as separate state $\dot{\lambda} = 0$
  - Has constant dynamics for VO, VSLAM...
  - ...scale drifts spatially not temporally with keyframe based VO
  - Not constant in optical flow approaches (see later)

- Previous dynamics remain unchanged

$$\begin{aligned}
\dot{p}_w^i &= v_w^i \\
\dot{v}_w^i &= C_{(q_w^i)}^T (a_m - b_a - n_a) - g \\
\dot{q}_w^i &= \frac{1}{2}\Omega(\omega_m - b_\omega - n_\omega) q_w^i
\end{aligned}$$

- State vector $X = \{ {p_w^i}^T \ {v_w^i}^T \ {q_w^i}^T \ b_\omega^T \ b_a^T \ \lambda \}$

EKF Approach: Loosely Coupled

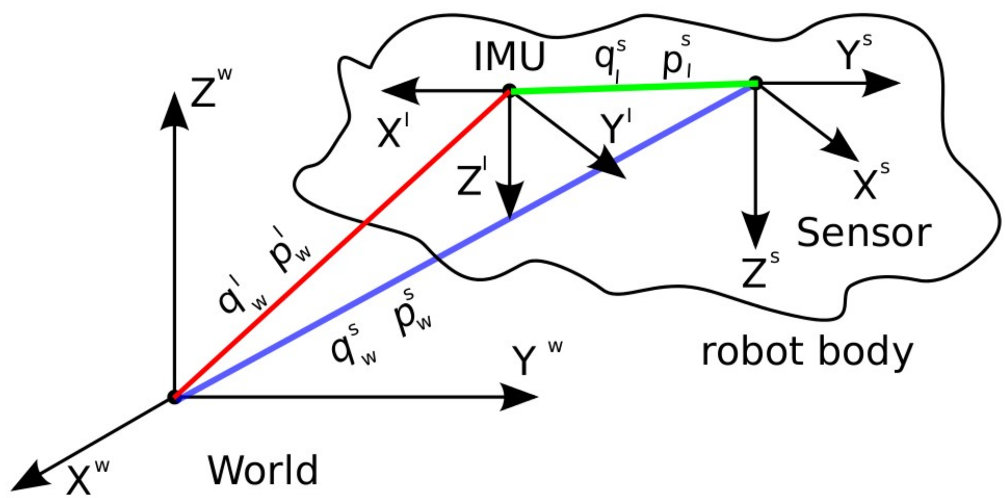
- Simple, constant size update, independent of number of features
  - Complexity is distributed between vision module and EKF module
  - Can be implemented in different architectures (or run in different cores)

position:

$$\tilde{z}_p = z_p - \hat{z}_p \qquad \boldsymbol{z}_p = \boldsymbol{p}_w^s = (\boldsymbol{p}_w^i + C_{(q_w^i)}^T \boldsymbol{p}_i^s)\lambda + \boldsymbol{n}_p$$

attitude (multiplicative):

$$\tilde{z}_q = z_q \otimes \hat{z}_q^{-1} \qquad \boldsymbol{z}_q = q_w^s = q_i^s \otimes q_w^i$$

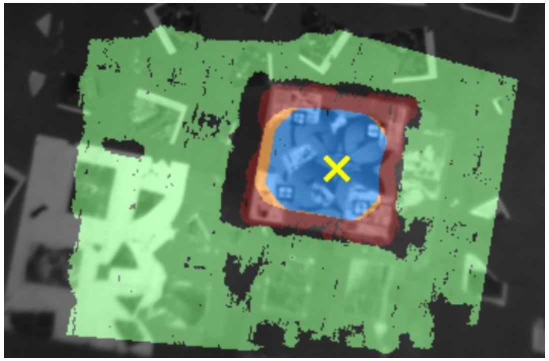


EKF Approach: Loosely Coupled

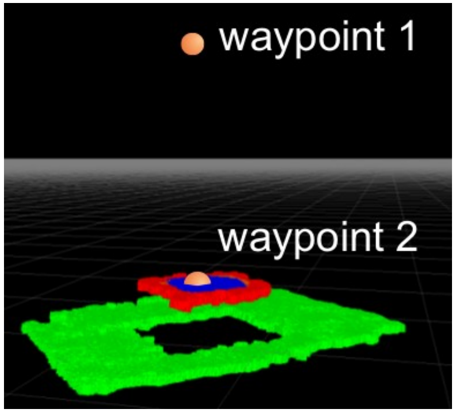
- Implementation:
  - EKF module uses minimal computation power
  - Bottleneck is the covariance propagation, not the update step:

$$P_{k+1|k} = F_d P_{k|k} F_d^T + Q_d$$

→ use block-sparse methods to reduce complexity

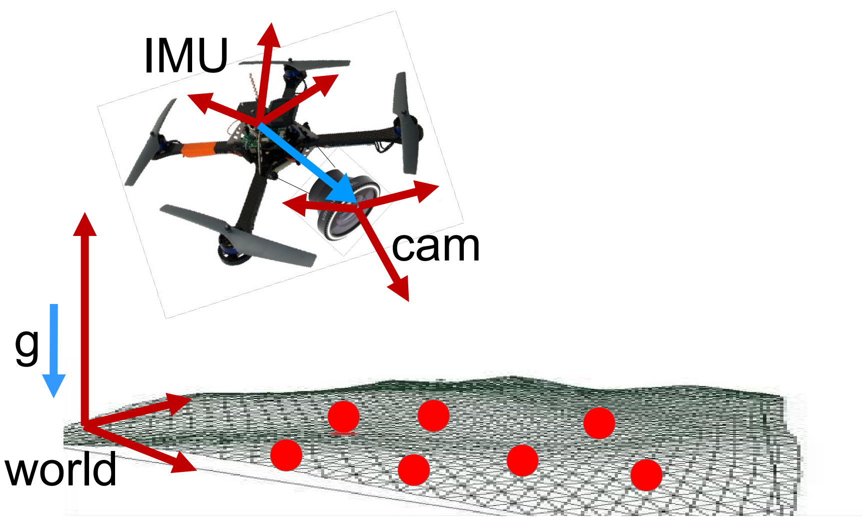


- Setup on cell phone processor 1.7GHz quad core:
  - 55Hz Visual SLAM: uses 1.5 cores (ethzasl_ptam: wiki.ros.org/ethzasl_ptam)
  - 100Hz EKF module: uses 0.2 cores (ethzasl_sensor_fusion: wiki.ros.org/ethzasl_sensor_fusion)
  - Sufficient computation power free for high level task: e.g. autonomous, safe landing @ 2Hz (Brockers et al. CVPR 2014)



Power-On-And-Go
From scale estimation to self-calibrating platforms

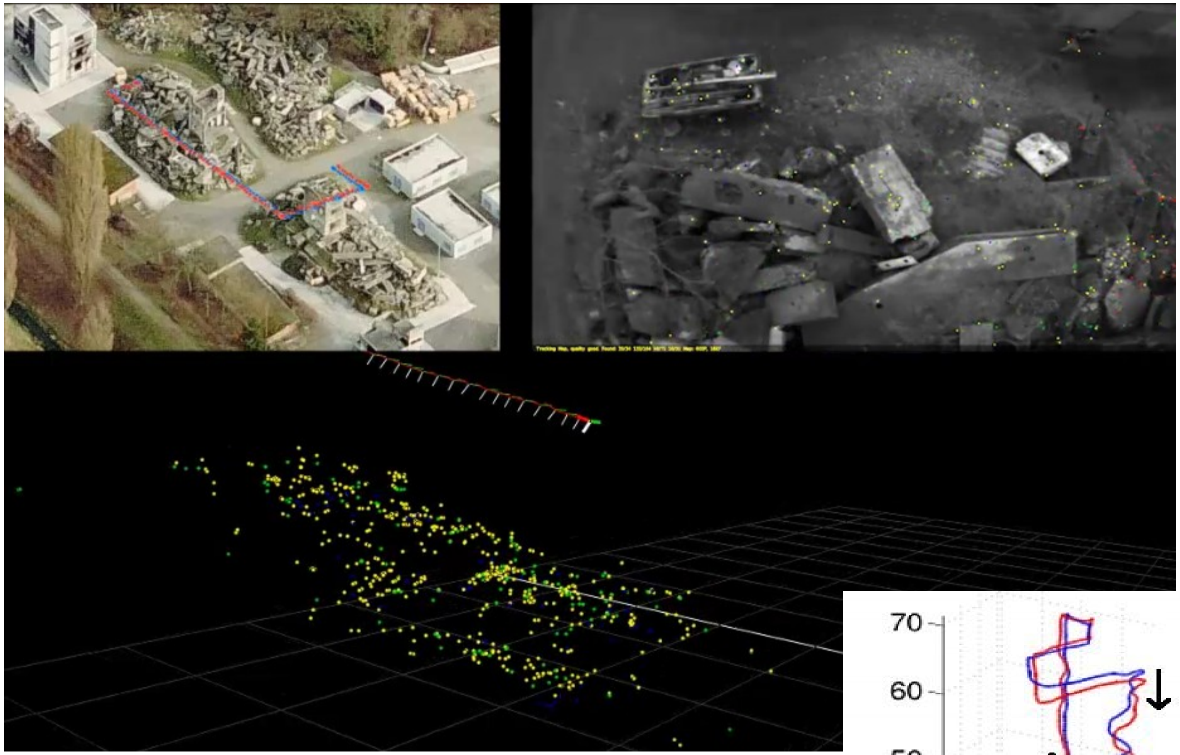
- Do not stop at scale: use full capabilities of fusing IMU with vision
  - Particularly: gravity aligned navigation frame
  - Self-calibrating sensor suite
  - Yields **power-on-and-go** robots



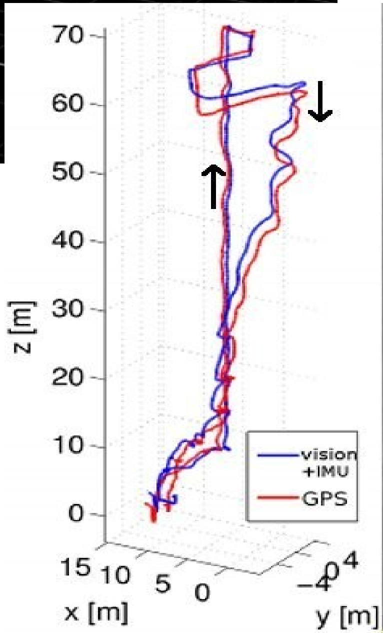
[Weiss et al. JFR 2013]

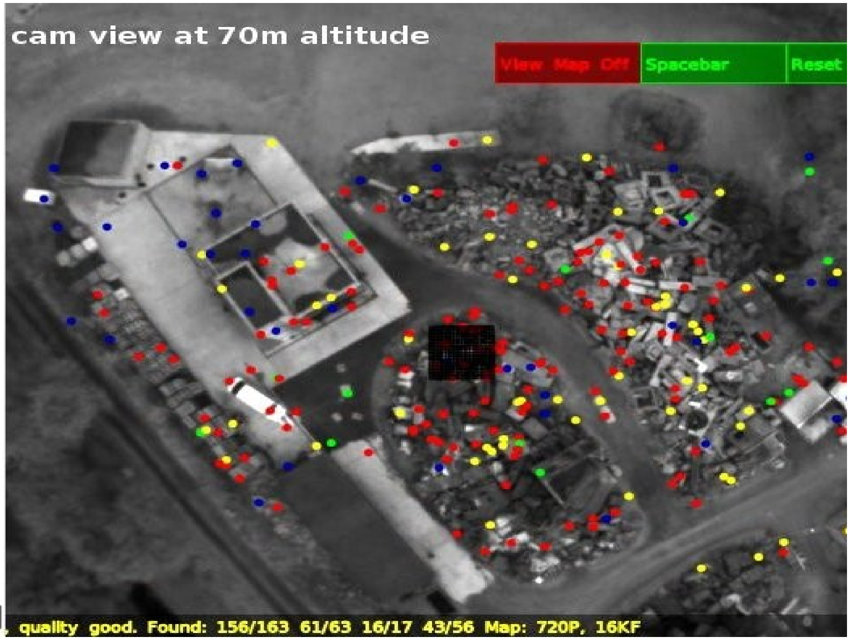
MAV Control States			IMU Intrinsics		Sensor Extrinsics		Visual Drifts		
pos.	vel.	att.	b_acc	b_gyr	trans.	rot.	scale	drift_r	drift_p
metric control			continuous self-calibration						

Power-On-And-Go
From scale estimation to self-calibrating platforms



[Weiss et al. JFR 2013]





Variable Scale in Optical Flow

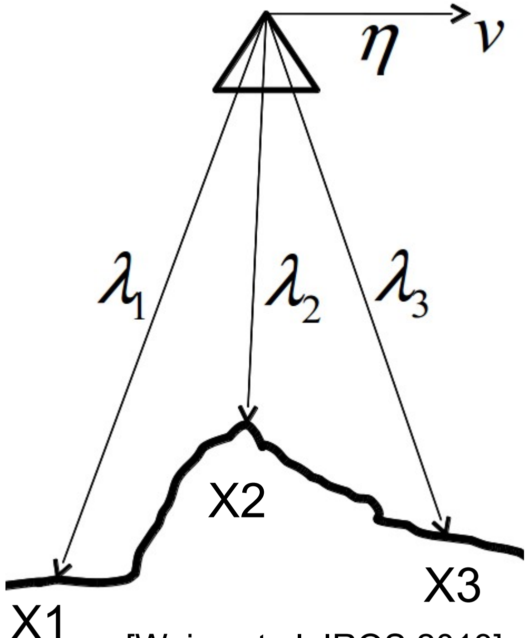
- Back to the basics: frame to frame motion estimation
  - Epipolar constraints yields R, T between two camera poses

$$\mathbf{x}_2^T E \mathbf{x}_1 = 0 \qquad\qquad E = T_\times R$$

- With high framerate we are entering the time-continuous domain
  - Continuous epipolar constraint yields linear and angular velocities of the current camera frame

$$\dot{\mathbf{X}}(t) = \cancel{\lfloor\vec{\omega}(t)\rfloor\mathbf{X}(t)} + \vec{V}(t)$$

$$\dot{\vec{x}}^T\lfloor\vec{v}(t)\rfloor\vec{x} + \vec{x}^T\cancel{\lfloor\vec{\omega}(t)\rfloor\lfloor\vec{v}(t)\rfloor}\vec{x} = 0$$



[Weiss et al. IROS 2013]

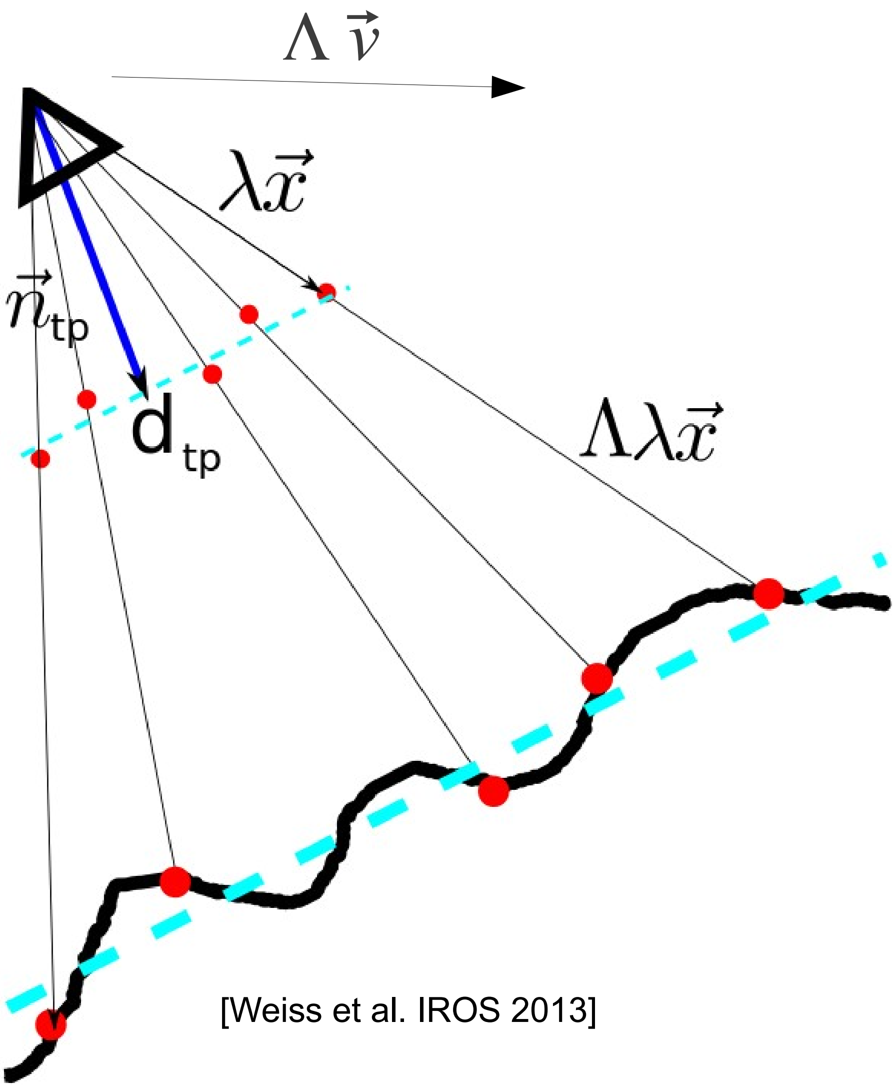
- 5-dimensional problem. Reduce to 2D:
  - Use IMU readings for ω
  - De-rotate features and set $\omega = 0$

Variable Scale in Optical Flow

- Optical flow (OF) is proportional to the ratio of velocity and feature distance

$$OF \propto \frac{v}{\lambda}$$

- Fusion of IMU and OF disambiguates scale and velocity
  - Scale is proportional to the scene distance
- Motion model for scale propagation can be applied
  - Allows fast scale tracking in agile motion

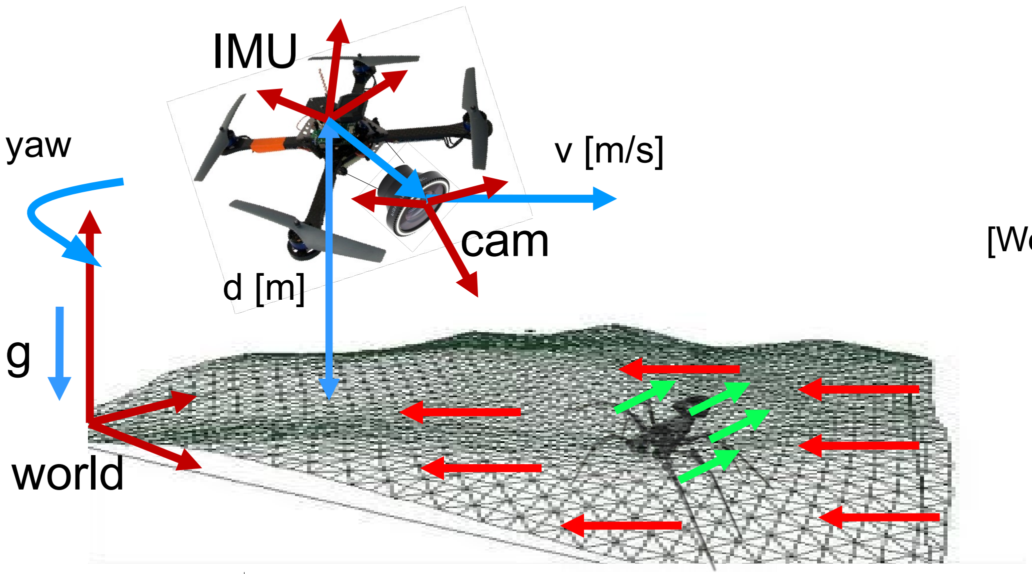


[Weiss et al. IROS 2013]

Variable Scale in Optical Flow

- Optical flow benefits
  - Only need two consecutive frames, no feature history, scale propagation, or map
  - Use IMU effectively to reduce problem to 2 dimensions: need only 2 features
  - Very fast computation and RANSAC outlier rejection

- BUT: no position estimation (other than scene distance)



[Weiss et al. IROS2013]

MAV Control States			IMU Intrinsics		Sensor Extrinsics		Visual Drifts
Plane dist	velocity	attitude	b_acc	b_gyr	trans.	rot.	scale
metric control			continuous self-calibration				

Throw-And-Go
From scale estimation to fail-safe self-calibrating platforms

- Loosely coupled implementation: 50Hz on 1 core of 1.7GHz quad core CPU
- Inherently fail safe: only uses 3 feature matches in 2 consecutive images
- No feature history nor local map

[Weiss et al. IROS 2013]

Modularity: other sources to estimate scale and drift

- Additional sensors
  - Other cameras: stereo vision
  - Visual patterns
  - GPS
  - UWB range sensing
  - Air pressure

- Use more than just scale information
  - Observability analysis for additional states
  - Self-Calibration is crucial for long-term opertation
  - Literture: Martinelli FTR 2013

- Modular Multi-Sensor Fusion available online:
  - wiki.ros.org/ethzasl_sensor_fusion
  - Theory: Weiss PhD Thesis 2012
  - To start: wiki.ros.org/ethzasl_sensor_fusion

UWB range sensing module:
accuracy 2cm (TimeDomain)

Q & A