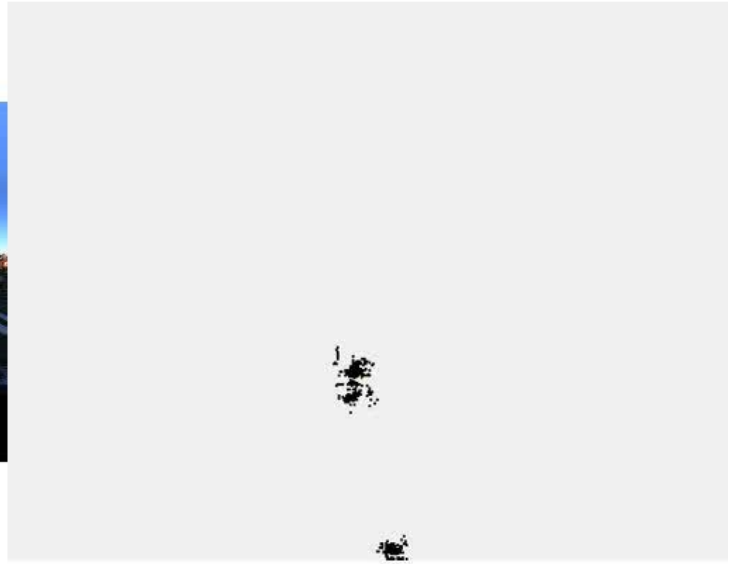
Perception: Visual Odometry

Kostas Daniilidis

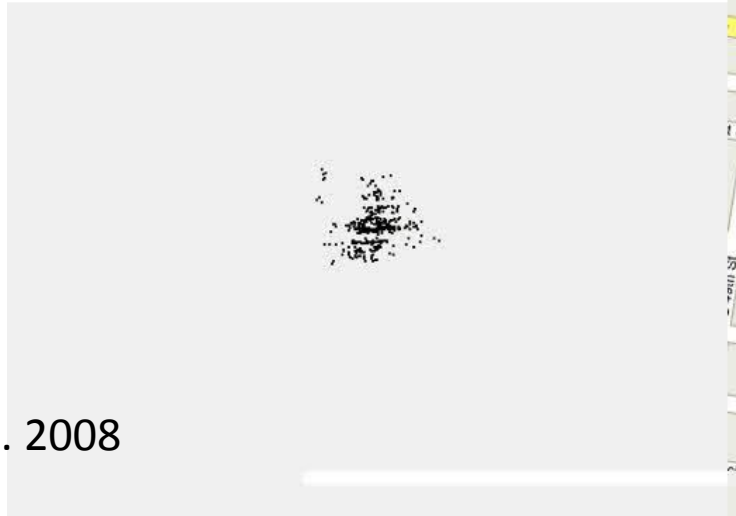
Extract camera trajectory from video

Panoramic image (from 6 cameras)

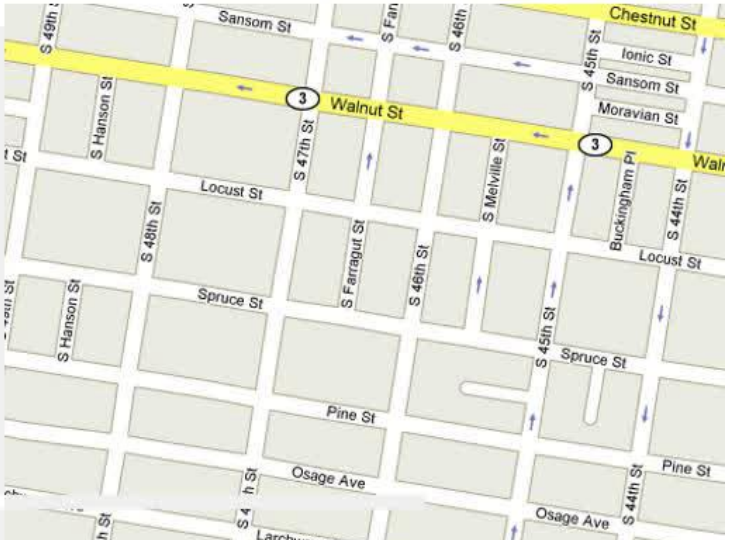
Reconstruction (global view)



Reconstruction (close-up)



Google Map



UPenn,
Tardif et al. 2008

What is Odometry ?

- Measuring how far you go by counting wheel rotations or steps.
- Known as “path integration” in biological perception.
- More general, integration of velocity or acceleration measurements: inertial odometry.

What is Visual Odometry ?

The process of incrementally estimating your position and orientation with respect to an initial reference frames by tracking visual features.

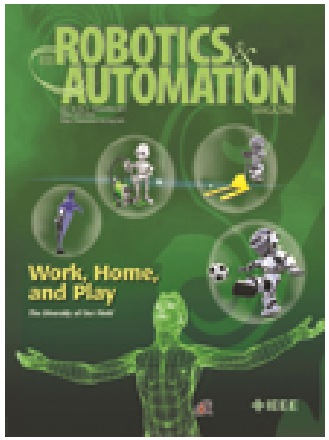
What is the difference to multiple views/bundle adjustment in the last lecture?

- Bundle adjustment can have large baselines, different cameras, sparse viewpoints
- Visual odometry is based on video and needs to be incremental
- Video allows a motion model

What is the difference to visual SLAM (Simultaneous Localization and Mapping) ?

- Used interchangeably but visual SLAM produces also map of features while visual odometry focuses on the camera trajectory.

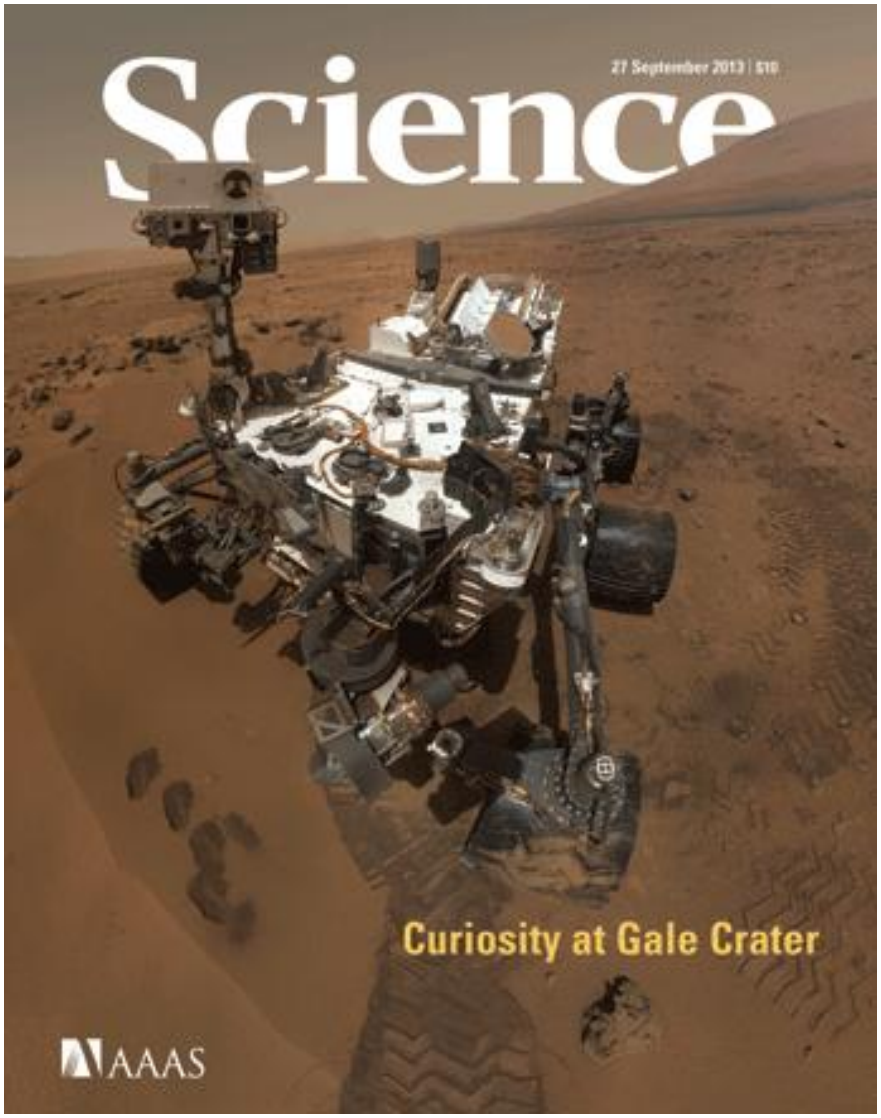
New field, not in textbooks, but good reference tutorial



- Scaramuzza, D., Fraundorfer, F., Visual Odometry: Part I - The First 30 Years and Fundamentals, IEEE Robotics and Automation Magazine, Volume 18, issue 4, 2011.
- Fraundorfer, F., Scaramuzza, D., Visual Odometry: Part II - Matching, Robustness, and Applications, IEEE Robotics and Automation Magazine, Volume 19, issue 1, 2012.

The Future of Real-Time SLAM: 18th December 2015 (ICCV Workshop)

Visual odometry on the MARS



Dyson 360 (Andrew Davison)

Multiple views setting:

Given calibrated point projections of $p = 1 \ldots N$ points in $f = 1 \ldots F$ frames (x_p^f, y_p^f)

Find the 3D rigid transformation R^f, T^f and the 3D points $\mathbf{X}_p = (X_p, Y_p, Z_p)$ that best satisfy the projection equations

Multiple views setting:

Given calibrated point projections of $p = 1 \dots N$ points in $f = 1 \dots F$ frames (x_p^f, y_p^f)

Find the 3D rigid transformation R^f, T^f and the 3D points $\mathbf{X}_p = (X_p, Y_p, Z_p)$ that best satisfy the projection equations

Visual Odometry:

Given **an estimate** R_k, T_k of the current camera pose as well as the 3D points $\mathbf{X}_p = (X_p, Y_p, Z_p)$ and correspondences to calibrated point projections in frame $(k+1)$ (x_p^{k+1}, y_p^{k+1})

Update to the pose R_{k+1}, T_{k+1}

Monocular visual odometry will leave an unknown global scale.

Update step for rotation:

- Find correspondences from view k to view $k+1$ using RANSAC and 5-point algorithm.
- Solve for epipolar geometry between two views k and $k+1$ using all inliers
- Use the rotation estimate ${}^{k}R_{k+1}$ to update the rotational pose

$$R_{k+1} = R_k \; {}^{k}R_{k+1}$$

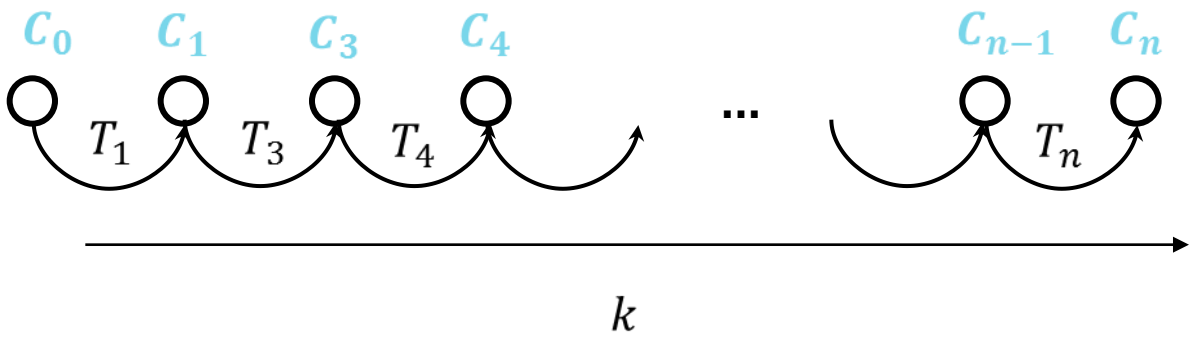
- If we use this for translation, we do not know the scale of ${}^{k}T_{k+1}$:

$$T_{k+1} = T_k + R_k \; {}^{k}T_{k+1}$$

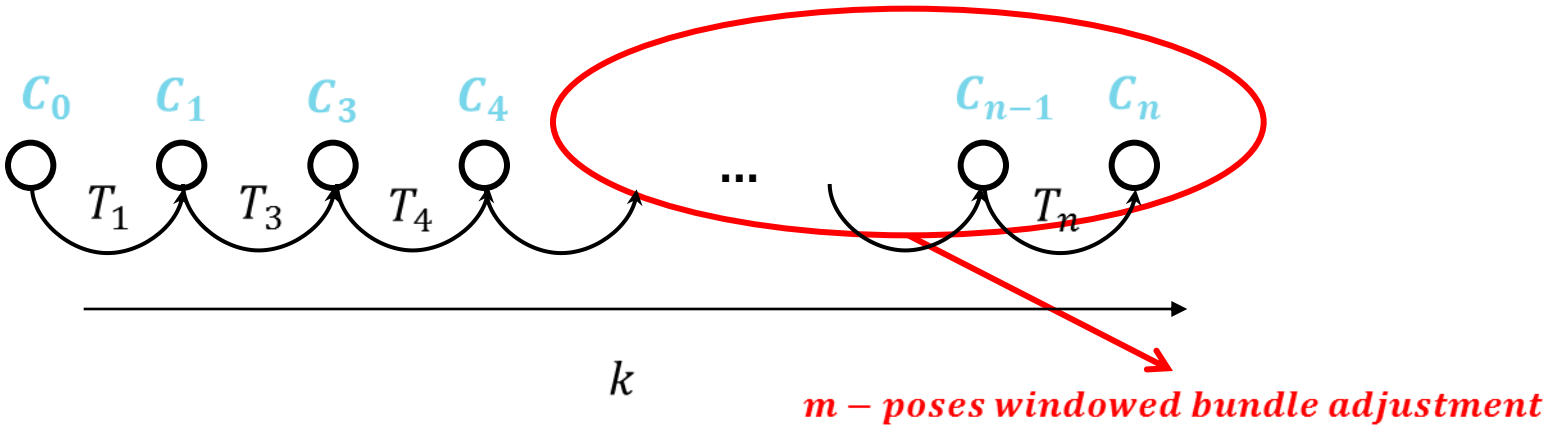
Update step for translation and structure:

- Use the estimated 3D points $\mathbf{X}_p$ and their projection correspondences (x_p^{k+1}, y_p^{k+1}) to update translational pose R_{k+1}, T_{k+1} using 2D-3D pose algorithms (usually only the translation is updated).
- Update the estimates of the 3D points $\mathbf{X}_p$

Main cycle of visual odometry



To minimize drift we run bundle adjustment over a window

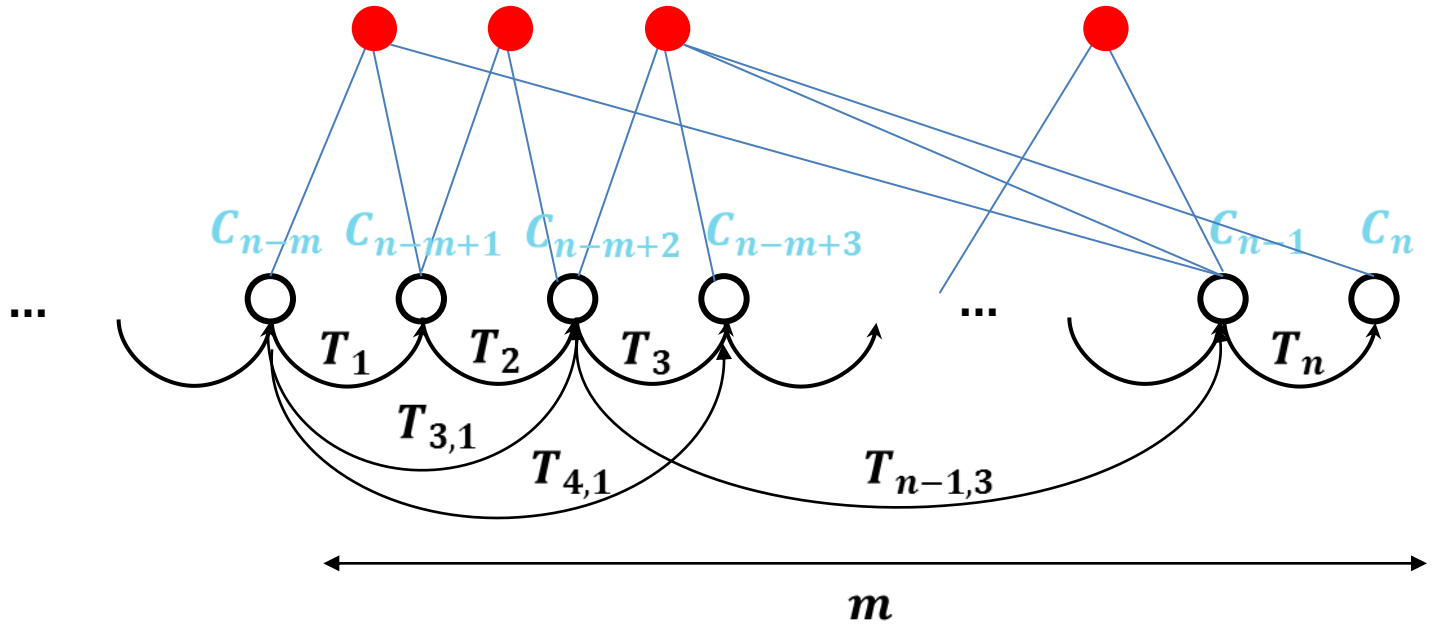


Use the Bundle Adjustment equations?

$$\arg\min_{X^i, C_k} \sum_{i,k} \| p_k^i - g(X^i, C_k) \|^2$$

$$x_p^f = \frac{R_{11}^f X_p + R_{12}^f Y_p + R_{13}^f Z_p + T_x}{R_{31}^f X_p + R_{32}^f Y_p + R_{33}^f Z_p + T_z}$$

$$y_p^f = \frac{R_{21}^f X_p + R_{22}^f Y_p + R_{23}^f Z_p + T_y}{R_{31}^f X_p + R_{32}^f Y_p + R_{33}^f Z_p + T_z}$$



State vector consists of all points which remain fixed in the global frame (some approaches use the projections and inverse depths as state) as well as the poses and velocities .

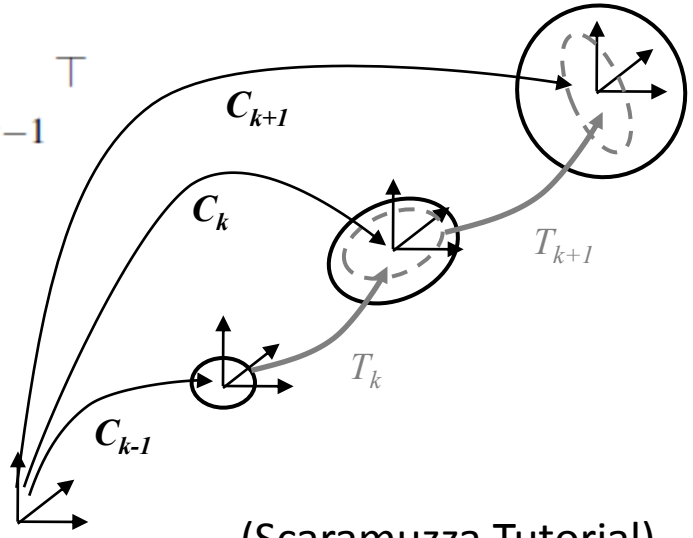
$$
\begin{aligned}
\mathbf{X}_p^{k+1} &= \mathbf{X}_p^k \\
R^{k+1} &= e^{\hat{\omega}^k} R^k \\
\omega^{k+1} &= \omega^k \\
T^{k+1} &= T^k + R^k v^k \\
v^{k+1} &= v^k
\end{aligned}
$$

Error Propagation

State Covariance is an estimate of its uncertainty. If uncertainty is Gaussian it can be visualized as an ellipsoid.

Its update depends on the previous uncertainty Σ_{k-1}, the measurement uncertainty $\Sigma_{k,k-1}$, and the Jacobian with respect to state J.

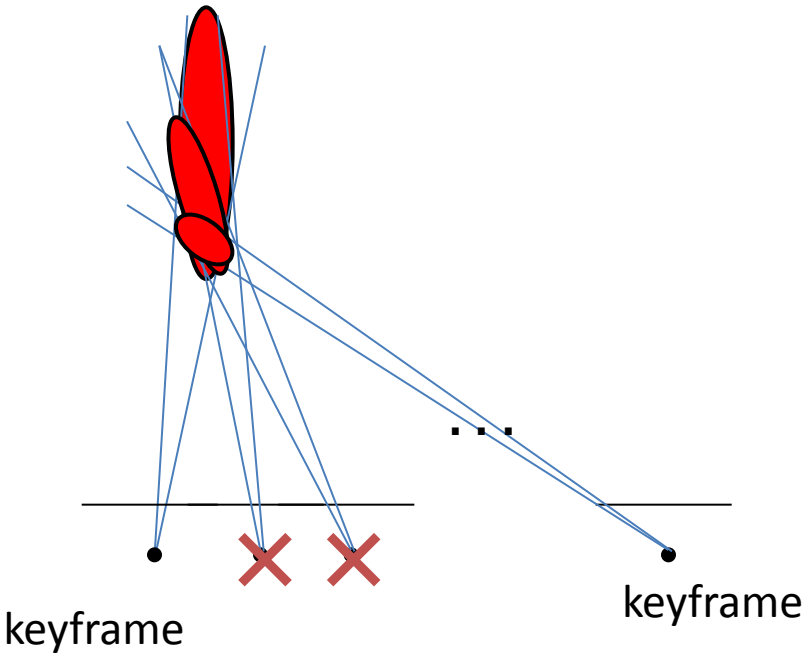
$$\begin{aligned}\Sigma_k &= J \begin{bmatrix} \Sigma_{k-1} & 0 \\ 0 & \Sigma_{k,k-1} \end{bmatrix} J^\top \\ &= J_{\vec{C}_{k-1}} \Sigma_{k-1} {J_{\vec{C}_{k-1}}}^\top + J_{\vec{T}_{k,k-1}} \Sigma_{k,k-1} {J_{\vec{T}_{k,k-1}}}^\top\end{aligned}$$



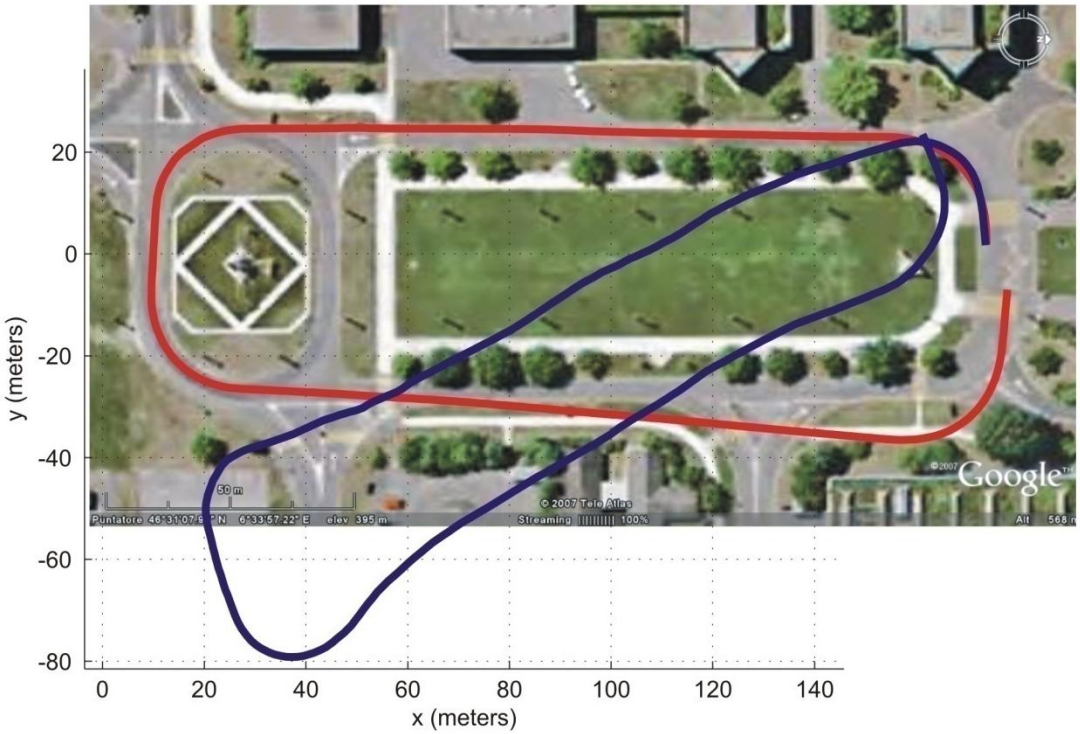
(Scaramuzza Tutorial)

Triangulation and Keyframe Selection

- Pose (translation) update depends on triangulated points whose error depends on baseline and distance.
- Wait until error in 3D triangulation decreases and then update pose: keyframe



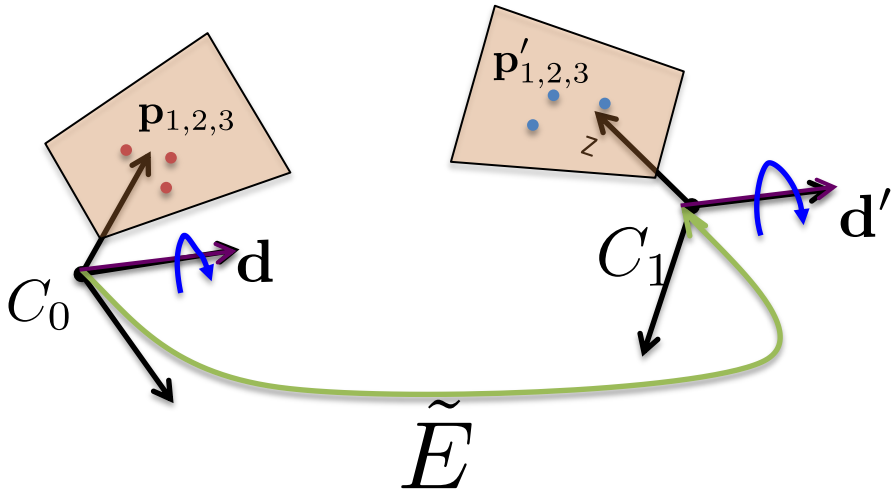
Outliers in VO: before and after



- While keyframe selection reduces drift, a large factor is the good inlier selection in point correspondences (Nister 2004, 5-point algorithm)

The 3-pt algorithm

- Gravity or a point at infinity constrains 2 DOF of the rotation, 3 DOF remaining ("yaw" and two for translation)
- First align each directional vector with the y-axis
  - Only rotation around the y-axis remains
- New 3-DOF epipolar constraint is $\mathbf{p}_i'^{\top} \tilde{E} \mathbf{p}_i = 0$



Formulating the 3pt Problem

- Parameterize the essential matrix

$$\tilde{E} = \hat{\tilde{\mathbf{t}}}(I + \sin\theta\hat{\mathbf{e}}_2 + (1-\cos\theta)\hat{\mathbf{e}}_e^2)$$

- 4 unknowns $\hat{\tilde{\mathbf{t}}} = [x, y, 1]^\top, \sin\theta$ and $\cos\theta$
- To make a polynomial system, let

$$c = \cos\theta \qquad s = \sin\theta$$

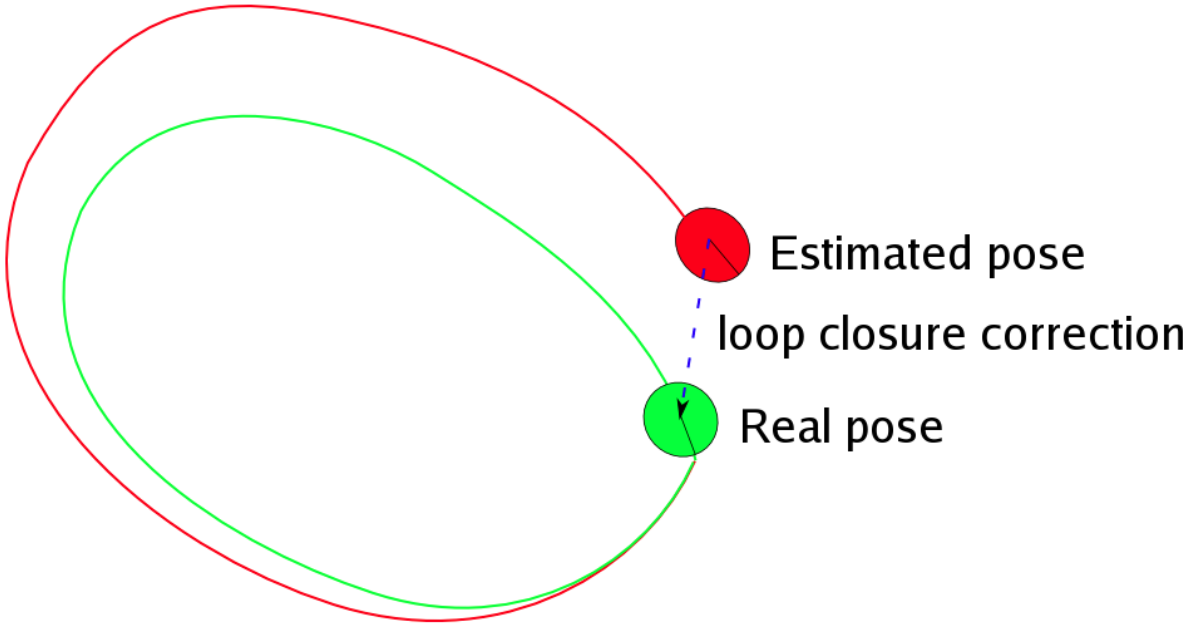
- Add the trigonometric constraint

$$s^2 + c^2 - 1 = 0$$

- Result: 4 polynomial equations in 4 unknowns

Naroditsky, et. al, *Two Efficient Solutions for Visual Odometry Using Directional Correspondence*, PAMI 2012

Visual loop closing



Angelis et al. 2008

TWO STEPS:

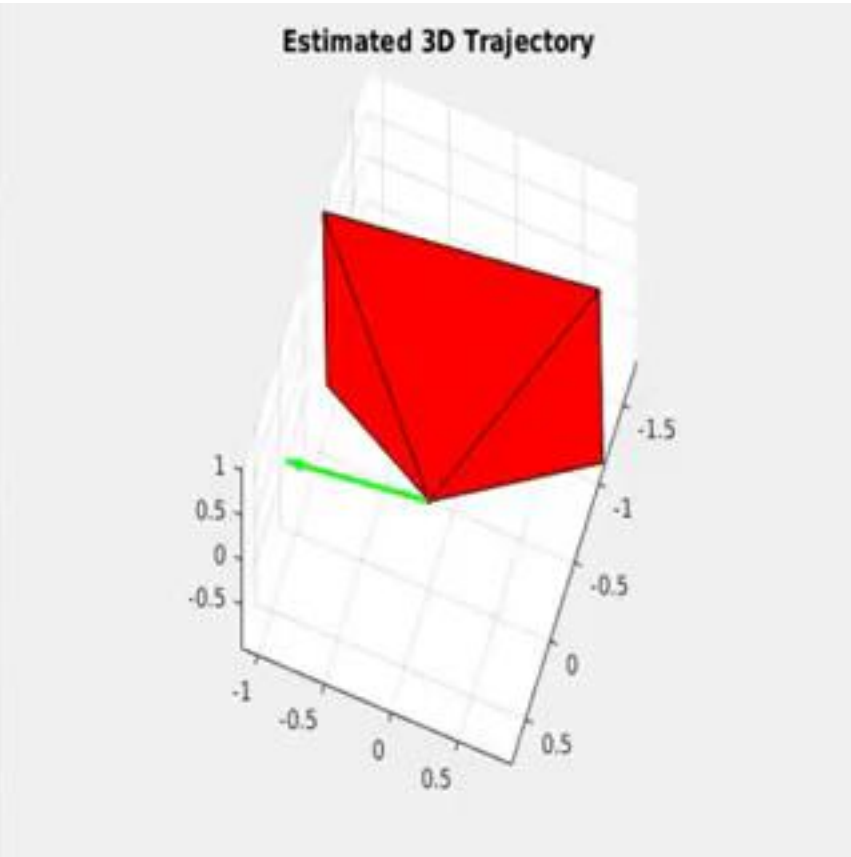
- Search for the closest VISITED image using feature retrieval (vocabulary trees)
- Geometric consistency with epipolar constraint.

Summary of Visual Odometry Tools

- Bundle Adjustment over a window
- Keyframe Selection
- RANSAC for 5-points or reduced minimal problem with 3 points.
- Visual Closing to produce unique trajectories when places are revisited

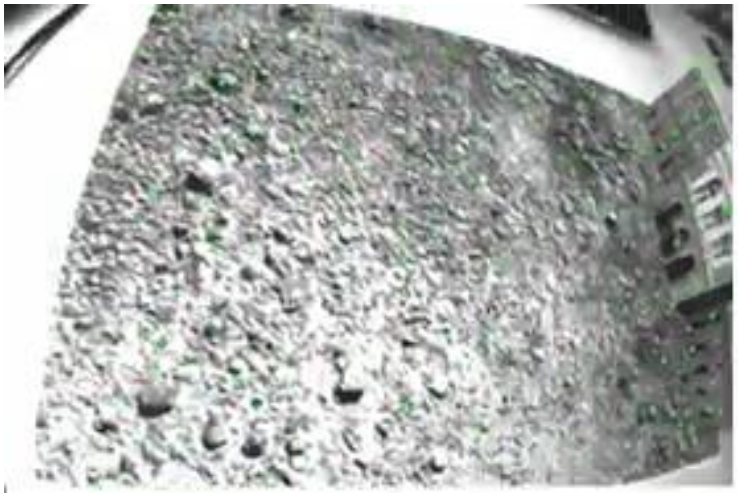
Integration with IMU (Inertial Measurement Unit)

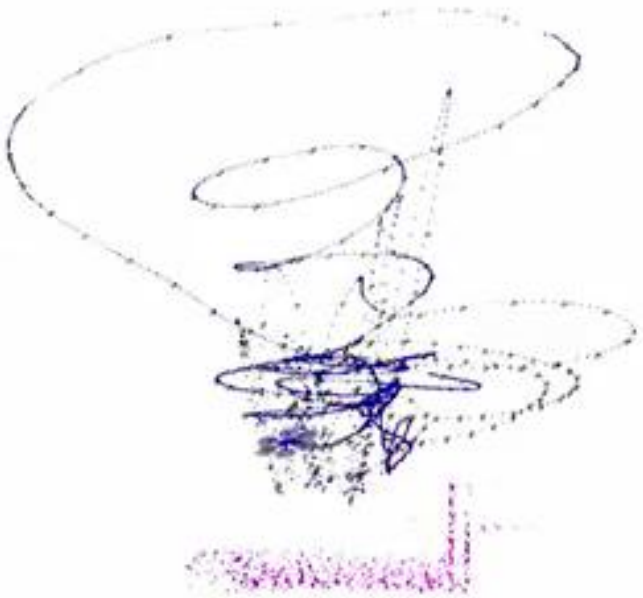
- Acceleration measurements make the unknown monocular scale observable!
- State vector is augmented with the unknown bias in the acceleration and angular velocity IMU measurements.



Libviso (Geiger at al. 2012)

Semi Direct Visual Odometry (Forster et al. 2014)





Realtime
Camera at 70fps

Google Project Tango

The Future of visual SLAM: semantic visual inertial navigation (Bowman et al. 2016)