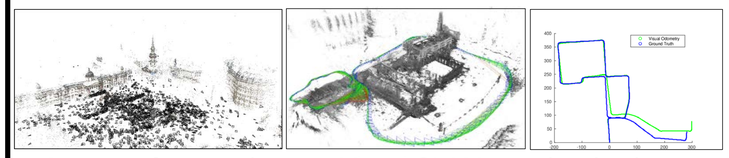
Carnegie Mellon

Monocular and Stereo Visual Odometry

Jai Prakash Utkarsh Sinha
Robotics Institute

Introduction

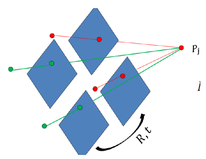
- Augmented Reality on Smartphone.
- Hololens globally consistent.
- No need to be globally consistent many a times.



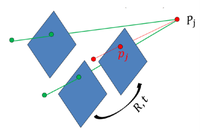
Structure from Motion Visual SLAM Visual Odometry

Background

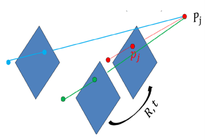
- 3D – 3D
- Reconstruction using stereo
- Register: Rigid body transformation



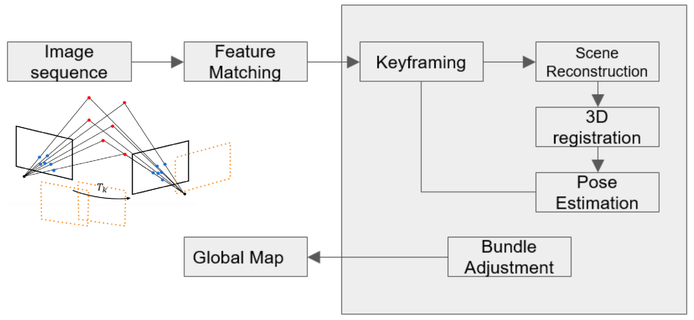
- 3D – 2D
- 3D points and 2D correspondence
- Use PnP to get pose



- 2D-2D
- Monocular camera
- 3D reconstruction using Key-frames
- Find pose using PnP

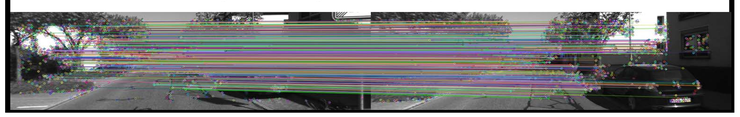


System

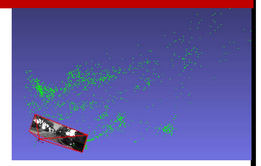


Feature Matching

- ORB and AKAZE features
- Ratio test, Symmetry Test, Epi-polar/RANSAC test



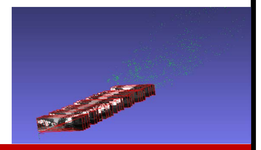
Key-frames and 3D reconstruction



- Key-frame generation heuristics
  - Every n[th] frame
  - Homography based
  - number of matches
- 3D reconstruction using Stereo images
- 3D reconstruction using previous Keyframe
- 2-view triangulation
- Scale ambiguity

Point cloud registration

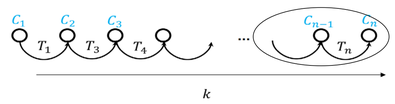
- Find scale of 2 point clouds using common features
- Register them together using common 3D points
- Minimize $\sum_{i=1}^{n} w_i \left\| (R\mathbf{p}_i + \mathbf{t}) - \mathbf{q}_i \right\|^2$
- Centroid $\bar{\mathbf{p}} = \frac{\sum_{i=1}^{n} w_i \mathbf{p}_i}{\sum_{i=1}^{n} w_i}, \quad \bar{\mathbf{q}} = \frac{\sum_{i=1}^{n} w_i \mathbf{q}_i}{\sum_{i=1}^{n} w_i} \quad \mathbf{x}_i := \mathbf{p}_i - \bar{\mathbf{p}}, \quad \mathbf{y}_i := \mathbf{q}_i - \bar{\mathbf{q}}, \quad i = 1, 2, \ldots, n$
- SVD of covariance $S = XWY^\top \qquad S = U\Sigma V^\top$
- R and t $R = V \begin{pmatrix} 1 & & \\ & \ddots & \\ & & \det(VU^\top) \end{pmatrix} U^\top \qquad \mathbf{t} = \bar{\mathbf{q}} - R\bar{\mathbf{p}}$
- Recover pose using PnP



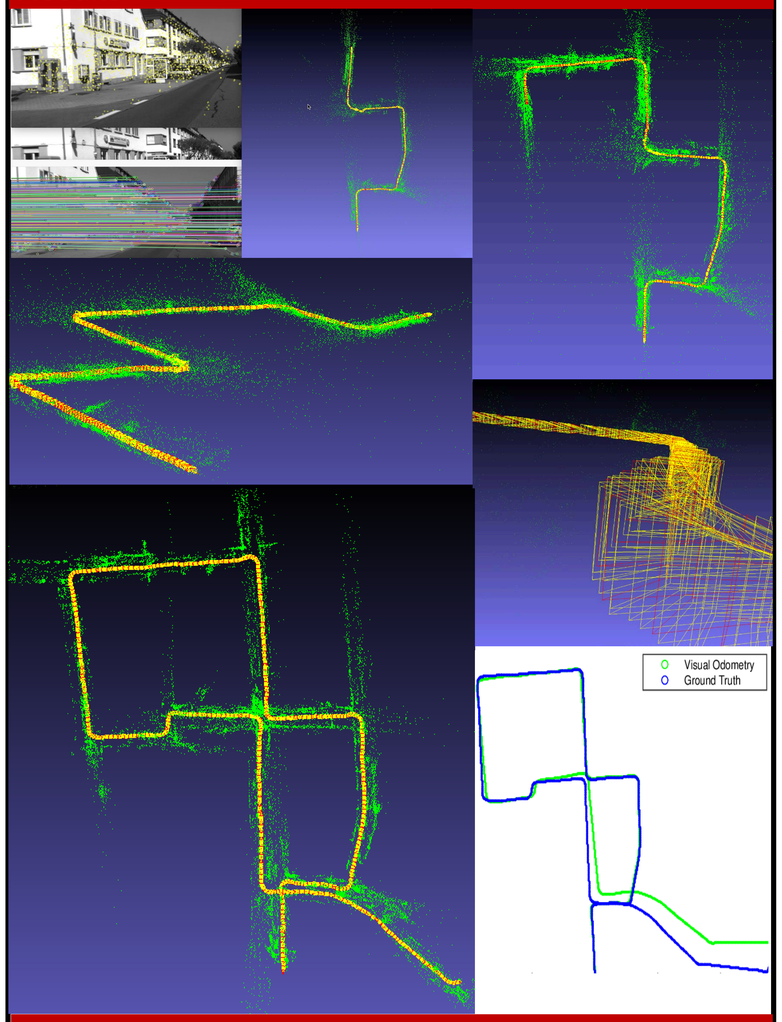
Bundle Adjustment

- Non linear optimization over last m-frames
- Globally consistent map and visibility in current frame
- Minimize re-projection error
- Dense schur
- OpenCV and Ceres solver

$$\min_{\hat{\mathbf{P}}^i, \hat{\mathbf{X}}_j} \sum_{ij} d(\hat{\mathbf{P}}^i \hat{\mathbf{X}}_j, \mathbf{x}_j^i)^2$$



Results



Future work

- 3-point algorithm for feature matching
- Key-frame based on covariance
- Integration with inertial sensor for pose
- Efficient data structures for correspondences
- Probabilistic modelling for robustness

References

1. Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós. ORB-SLAM: A Versatile and Accurate Monocular SLAM System
2. D. Nister, O. Naroditsky and J. Bergen. Visual Odometry. CVPR 20114.
3. Visual Odometry tutorials, CVPR 2014. http://frc.ri.cmu.edu/~kaess/vslam_cvpr14/