SVO: Fast Semi-Direct Monocular Visual Odometry

Christian Forster, Matia Pizzoli, Davide Scaramuzza
Robotics and Perception Group, University of Zurich, Switzerland
ICRA 2014

Presenter: You-Yi Jau, M.S. in UCSD

Motivation and problem description

- Why visual odometry?
  - Micro Aerial Vehicles (MAVs) need localization systems
- Why semi-direct method?
  - Feature-based method suffers in textureless scenes
  - Efficient: no feature extraction and matching on pixels
  - Robust: in repetitive, or high-frequency textures
  - Use features (120 per image) and small patches
- Problem description:
  - Input: Image frames
  - Output: Camera pose, and semi-dense depth

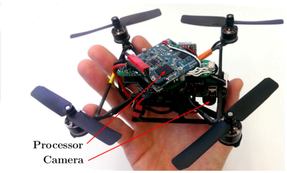
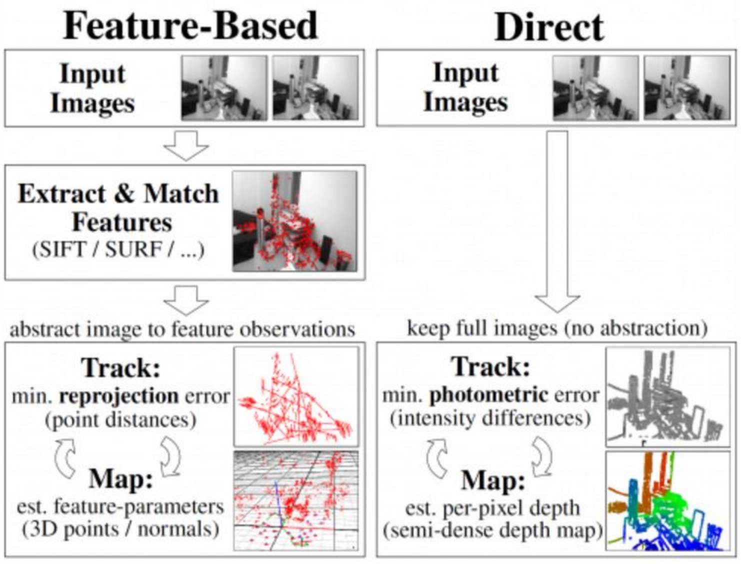


Fig. 17: "Nano+" by KMel Robotics, customized with embedded processor and downward-looking camera. SVO runs at 55 frames per second on the platform and is used for stabilization and control.

Prior work

- Visual Motion Estimation Methods
  - Feature-based method
  - Direct method
- Parallel Tracking and Mapping for Small AR Workspaces (PTAM)
  - 2007
- Monocular Vision for Long-term Micro Aerial Vehicle State Estimation: A Compendium
  - 2013



Medium: LSD-slam and ORB-slam2, a literature based explanation

Method overview

- Semi-direct
  - Motion Estimation Thread
  - Mapping Thread

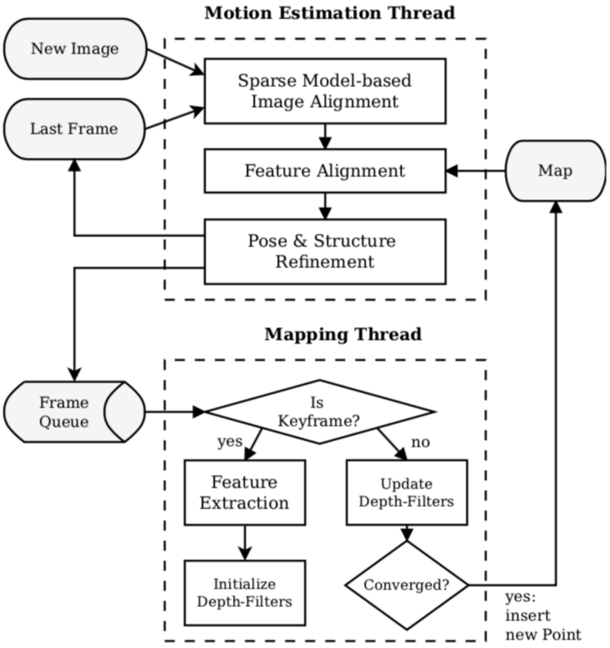


Fig. 1: Tracking and mapping pipeline

Method details and analysis

- motion-estimation (optimization, cost)

Image alignment

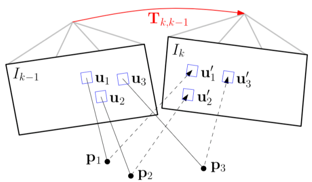


Fig. 2: Changing the relative pose $\mathbf{T}_{k,k-1}$ between the current and the previous frame implicitly moves the position of the reprojected points in the new image $\mathbf{u}'_i$. Sparse image alignment seeks to find $\mathbf{T}_{k,k-1}$ that minimizes the photometric difference between image patches corresponding to the same 3D point (blue squares). Note, in all figures, the parameters to optimize are drawn in red and the optimization cost is highlighted in blue.

Feature alignment

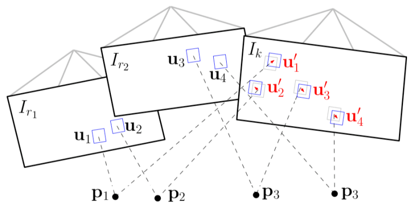


Fig. 3: Due to inaccuracies in the 3D point and camera pose estimation, the photometric error between corresponding patches (blue squares) in the current frame and previous keyframes r_i can further be minimised by optimising the 2D position of each patch individually.

Pose alignment

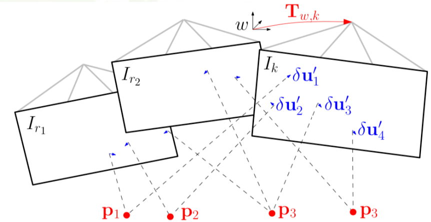


Fig. 4: In the last motion estimation step, the camera pose and the structure (3D points) are optimized to minimize the reprojection error that has been established during the previous feature-alignment step.

- mapping
  - Depth-filter
  - Update with correlation

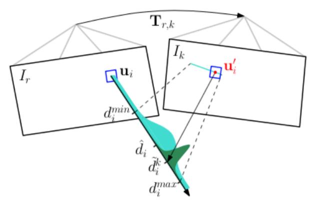


Fig. 5: Probabilistic depth estimate $\hat{d}_i$ for feature i in the reference frame r. The point at the true depth projects to similar image regions in both images (blue squares). Thus, the depth estimate is updated with the triangulated depth $\tilde{d}_i^k$ computed from the point $\mathbf{u}'_i$ of highest correlation with the reference patch. The point of highest correlation lies always on the epipolar line in the new image.

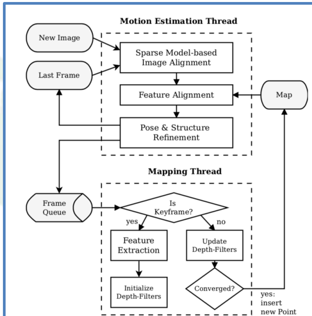


Fig. 1: Tracking and mapping pipeline

Experiments

- Baseline
  - 2013: Modified PTAM
- Settings
  - Fast or accurate method

	Fast	*Accurate*
Max number of features per image	120	200
Max number of keyframes	10	50
Local Bundle Adjustment	no	yes

TABLE I: Two different parameter settings of SVO.

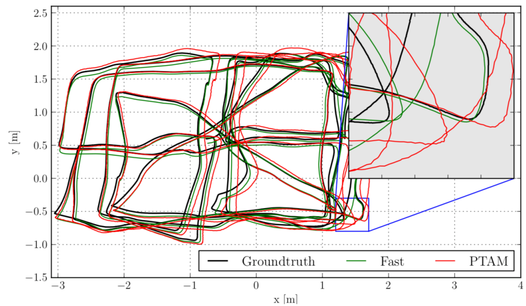


Fig. 7: Comparison against the ground-truth of SVO with the *fast* parameter setting (see Table I) and of PTAM. Zooming-in reveals that the proposed algorithm generates a smoother trajectory than PTAM.

	Pos-RMSE [m/s]	Pos-Median [m/s]	Rot-RMSE [deg/s]	Rot-Median [deg/s]
fast	0.0059	0.0047	0.4295	0.3686
accurate	0.0051	0.0038	0.4519	0.3858
PTAM	0.0164	0.0142	0.4585	0.3808

TABLE II: Relative pose and rotation error of the trajectory in Figure 7

  - Dataset: outdoor
- Speed

	Laptop (fps)	**Embedded (fps)**
Fast	>300	55
PTAM	91	27

Future work and discussion

- Discussion
  - Can we only use step(1): image alignment? – more drift
  - Can we skip step(1), and work directly on feature alignment and pose alignment? – outliers
- Future work
  - Unknown scale: visual-inertial

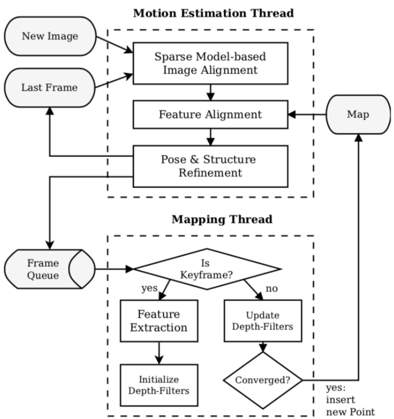


Fig. 1: Tracking and mapping pipeline

Questions?

- Failure cases?

Thank you

ALLPPT.com

Backup

ALLPPT.com

Prior work

- Visual Motion Estimation Methods
  - Feature-based method
    - feature detectors and descriptors that allow matching between images even at large inter-frame movement
    - the neccessity for robust estimation techniques to deal with wrong correspondences
  - Direct method
    - estimate struc- ture and motion directly from intensity values in the image
    - outperform feature-based methods in terms of robustness in scenes with little texture [14] or in the case of camera- defocus and motion blur
    - he computation of the photometric error is more intensive than the reprojection error

Prior work

- Monocular VO algorithm
- PTAM
- DTAM

Method details and analysis

- motion-estimation
  - pose initialisation through sparse model-based image alignment
    - minimizing the photometric error
- mapping
- Features → bundle adjustment