Lecture 10
Multiple View Geometry 4

Davide Scaramuzza

http://rpg.ifi.uzh.ch/
### This afternoon: Intermediate VO Integration

<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
<th>Instructor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.11.2019</td>
<td>Lecture 10 - Multiple-view geometry 3 (Part 2)</td>
<td>Davide Scaramuzza</td>
</tr>
<tr>
<td></td>
<td>Exercise session: Intermediate VO Integration</td>
<td>Daniel &amp; Mathias Gehrig</td>
</tr>
<tr>
<td></td>
<td>Exercise 08 - Lucas-Kanade tracker</td>
<td>Daniel &amp; Mathias Gehrig</td>
</tr>
<tr>
<td>05.12.2019</td>
<td>Lecture 12 - Place recognition and 3D Reconstruction</td>
<td>Davide Scaramuzza</td>
</tr>
<tr>
<td></td>
<td>Exercise session: Deep Learning Tutorial</td>
<td>Daniel &amp; Mathias Gehrig</td>
</tr>
<tr>
<td></td>
<td>Exercise 09 - Bundle Adjustment</td>
<td>Daniel &amp; Mathias Gehrig</td>
</tr>
<tr>
<td>19.12.2019</td>
<td>Lecture 14 - Event based vision</td>
<td>Davide Scaramuzza</td>
</tr>
<tr>
<td></td>
<td>After the lecture, we will visit Scaramuzza's lab. Departure from lecture room at 12:00 via tram 10.</td>
<td>Daniel &amp; Mathias Gehrig</td>
</tr>
<tr>
<td></td>
<td>Exercise session: Final VO Integration</td>
<td></td>
</tr>
</tbody>
</table>
How can we evaluate VO/SLAM algorithms?

This problem is known as “Benchmarking”
### Popular Datasets for VO/SLAM Benchmarking

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devon Island</td>
<td>[Furgale’11]</td>
<td>Stereo + D-GPS + inclinometer + sun sensor</td>
</tr>
<tr>
<td>EuRoC</td>
<td>[Burri’16]</td>
<td>MAV with synchronized IMU and stereo</td>
</tr>
<tr>
<td>Blackbird</td>
<td>[Antonini’18]</td>
<td>MAV indoor aggressive flight with rendered images and real dynamics + IMU</td>
</tr>
<tr>
<td>MVSEC</td>
<td>[Zhu’18]</td>
<td>Events, frames, lidar, GPS, IMU from cars, drones, and motorcycles</td>
</tr>
<tr>
<td>UZH Drone Racing</td>
<td>[Delmerico’19]</td>
<td>MAV aggressive flight, standard + event cameras, IMU, indoors and outdoors</td>
</tr>
</tbody>
</table>
What metrics should be used?

- **Accuracy**
- **Efficiency** (speed, memory, and CPU load)
- **Robustness** (HDR, motion blur, low texture)

Accuracy vs. Efficiency and Robustness: Linear relationship

Accuracy: Higher values are preferred.

Efficiency: Lower values are preferred.

Robustness: Robust systems are preferred.
Evaluation is a non-trivial task...

Maybe align the first poses and measure the end-pose error?

- **How many poses** should be used for the alignment?
- **Not robust:**
  - **Most VIOs are non-deterministic** (e.g., RANSAC, multithreading) → every time you run your VIO on the same dataset, you get different results
  - **Not meaningful:**
    - too sensitive to the trajectory shape
    - does not capture the error statistics
Metric 1: Absolute Trajectory Error (ATE)

**Absolute Trajectory Error**
RMSE of the aligned estimate and the groundtruth.

- Single number metric
- Many parameters to specify

**Step 1:** Align the trajectory

$$\arg\min_{R, T, s} \sum_{i=0}^{N} \| \hat{t}_i - s R t_i - T \|^2$$

Alignment parameters

- groundtruth positions
- estimated positions

**Step 2:** Root mean squared errors between the aligned estimate and the groundtruth.

$$\sqrt{\frac{\sum_{i=1}^{N} \| \hat{t}_i - s R t_i - T \|^2}{N}}$$

- Sturm et al., "A benchmark for the evaluation of RGB-D SLAM systems." IROS 2012.
- Zhang et al., "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." IROS’18. [PDF]
Metric 2: Relative Trajectory Error (RTE)

Relative Error (Odometry Error)
Statistics of sub-trajectories of specified lengths.

- Calculate errors for all the sub-trajectories of certain lengths.

✓ Informative statistics
✗ Complicated to compute and rank

- Zhang et al., "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." IROS’18. PDF
Metric 2: Relative Trajectory Error (RTE)

**Relative Error (Odometry Error)**
Statistics of sub-trajectories of specified lengths.

- Calculate errors for all the sub-trajectories of certain lengths.

✓ Informative statistics
✗ Complicated to compute and rank

- Zhang et al., "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." IROS’18. [PDF]
Metric 2: Relative Trajectory Error (RTE)

**Relative Error (Odometry Error)**
Statistics of sub-trajectories of specified lengths.

- Informative statistics
- Complicated to compute and rank

➢ Calculate errors for all the subtrajectories of certain lengths.

---

Ground Truth

![Ground Truth](image1.png)

Estimate

![Estimate](image2.png)

- Zhang et al., "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." IROS’18. [PDF](#)
Metric 2: Relative Trajectory Error (RTE)

Relative Error (Odometry Error)
Statistics of sub-trajectories of specified lengths.

➤ Calculate errors for all the subtrajectories of certain lengths.

✓ Informative statistics
✗ Complicated to compute and rank

- Zhang et al., "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." IROS’18. PDF
Trajectory Accuracy: Error Metrics

- Both ATE and RTE are widely used in practice, but:
  - Many details need to be specified which are often omitted in papers
    - Number of poses used for the alignment (also, frames or keyframes?)
    - Type of transformation used for the alignment:
      - \( \text{SE}(3) \) for stereo VO
      - \( \text{Sim}(3) \) for monocular VO
      - 4DOF for VIO
    - Sub-trajectory lengths in RTE
  
- Results are not directly comparable with different settings
  - Report the evaluation settings in detail.
  - Is there a publicly available evaluation tool to facilitate reproducible evaluation? Yes: Trajectory Evaluation Toolbox

- White: Normal frames (used for real time pose update)
- Green: Keyframes (usually updated after BA)
Trajectory Evaluation Toolbox

- Designed to make trajectory evaluation easy!
  - Implements **different alignment methods** depending on the sensing modalities: \( SE(3) \) for stereo, \( sim(3) \) for monocular, \( 4DOF \) for VIO.
  - Implements **Absolute Trajectory Error** and **Relative Error**.
  - Automated evaluation of different algorithms on multiple datasets (for N runs).

- Code: [https://github.com/uzh-rpg/rpg_trajectory_evaluation](https://github.com/uzh-rpg/rpg_trajectory_evaluation) [Zhang, IROS’18]

---

**Zhang et al.,"A tutorial on quantitative trajectory evaluation for visual (inertial) odometry." IROS’18.** [PDF](https://example.com/pdf)
What metrics should be used?

- **Accuracy**
- **Efficiency** (speed, memory, and CPU load)
- **Robustness** (HDR, motion blur, low texture)
Benchmarking Efficiency

- Different computational resources
  - Memory
  - CPU load
  - Processing time

Depends not only on algorithm design, but also implementation, platforms, etc.

- There are **different definitions of processing time** in SLAM systems.

  - Processing time for real-time pose: 
    \[ t_{\text{pose output}} - t_{\text{image arrival}} \]

  - Processing time for asynchronously executed threads (e.g., bundle adjustment)

    - White: Normal frames (used for real time pose update)
    - Green: Keyframes (usually updated after BA)
Case study: VIO for Flying Robots [ICRA’18]

- **Algorithms:** MSCKF, OKVIS, ROVIO, VINS-Mono, SVO+MSF, SVO+GTSAM, VINS-Mono w/ and w/o loop closure

- **Hardware:** consider the limitation of flying robots

- **Evaluation**
  - **Absolute Trajectory Error (ATE)** – RMSE after sim(3) trajectory alignment (7DoF)
  - **Relative Trajectory Error (RTE)** – Error distribution of the subtrajectories
  - **CPU usage** – total load of CPU
  - **Memory usage** – total percentage of available RAM
  - **Time per frame** – from input until pose is updated

Delmerico, Scaramuzza, A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots, ICRA’18. [PDF](#). [Video](#).
Case study: VIO for Flying Robots [ICRA’18]

No free lunch: more computation → better accuracy

Delmerico, Scaramuzza, A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots, ICRA’18. [PDF]. [Video]
What metrics should be used?

Accuracy

Efficiency
(speed, memory, and CPU load)

Robustness
(HDR, motion blur, low texture)
Robustness is the greatest challenge for SLAM today!

How to cope & quantify robustness to:

- Low texture
- High Dynamic Range (HDR) scenes
- Motion blur
- Dynamically changing environments
- Algorithmic randomness

How can we quantify the robustness of algorithms to such situations?

UZH-FPV Drone Racing Dataset

Contains data recorded by a drone flying up to over 20m/s indoors and outdoors flown by a professional pilot. Contains frames, events, IMU, and Ground Truth from a Robotic Total Station: http://rpg.ifi.uzh.ch/uzh-fpv.html

Delmerico et al. "Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset" ICRA’19
PDF. Video. Datasets.
UZH-FPV Drone Racing Dataset

- Recorded with a drone flown by a professional pilot up to over 20m/s
- Contains images, events, IMU, and ground truth from a robotic total station: [http://rpg.ifi.uzh.ch/uzh-fpv.html](http://rpg.ifi.uzh.ch/uzh-fpv.html)

Robustness to high speed motion

Can be quantified in terms of “optical flow” (see Lecture 10 for def. of optical flow)

Delmerico et al. "Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset" ICRA’19
PDF. Video. Datasets.
Dataset Bias

Typical workflow of developing VO/VIO/SLAM algorithms:

- Development
- Evaluation on public datasets

As a community, we are overfitting the public dataset.

Potential problems:

- **Generalizability**: Performance on one does not guarantee to generalize to others
  - E.g., KITTI $\rightarrow$ low frame rate, not friendly for direct methods

- **Old datasets (e.g., KITTI) are already saturated**:
  - It becomes more and more difficult to tell whether we are making real progress or just overfitting the datasets.
  - E.g., does 1 or 2 cm improvement in RMSE over a 100 meter trajectory really mean something?
Dataset Bias

We need more datasets to evaluate the performance of SLAM algorithms along different axes

**Robustness**
(HDR, motion blur, low texture)
- BlackBird [Antonini’18]
- UZH-FPV dataset [Delmerico’18]
- Event Camera [Mueggler’17]
- MVSEC [Zhu’18]
- ...

**Efficiency**
(speed, memory, and CPU load)
- Devon Island [Furgale’13]
- KITTI [CVPR’12]
- EuRoC [Burri’16]
- TUM-RGBD [Sturm’12]
- TUM VI Benchmark [Schubert’18]
- ...

**Accuracy**

**Realistic simulators:**
- AirSim
- FlightGoggles [Guerra’19]
- ESIM [Rebecq’18]
- ......
Overview of the last 30 years of Visual Inertial Odometry & SLAM

  - Accuracy
  - Efficiency (speed and CPU load)
  - Robustness (HDR, motion blur, low texture)

- **Feature + Direct methods (from 2000)**
  - Efficiency (speed and CPU load)

- **+IMU (Lecture 13)**
  - (10x accuracy)

- **+Event Cameras (Lecture 14)**
Open Research Opportunities
Actively Control Camera Exposure Time to achieve Robustness to High Dynamic Range (HDR) scenes

ORB-SLAM with Standard Built-in Auto-Exposure

ORB-SLAM with Our Active Exposure Control

Zhang, et al., Active Exposure Control for Robust Visual Odometry in HDR Environments, ICRA’17. PDF. Video
“Ultimate SLAM”: Frames + Events + IMU (Lecture 14)

85% accuracy gain over standard visual-inertial SLAM in HDR and high speed scenes

Standard camera

Event camera

Estimated trajectory

Rosinol et al., *Ultimate SLAM?* IEEE RAL’18 best Paper Award Honorable Mention [PDF](#). [Video](#). [IEEE Spectrum](#).
Understanding Check

Are you able to answer the following questions:

- How do we benchmark VO/SLAM algorithms?
- Along which axes can we evaluate them?
- Benchmarking accuracy: Can we use the end pose error? What are ATE and RTE?
- How can we quantify Efficiency? And Robustness?
- What are open research opportunities?