Computer Vision Group
TUM Department of Informatics
Technical University of Munich

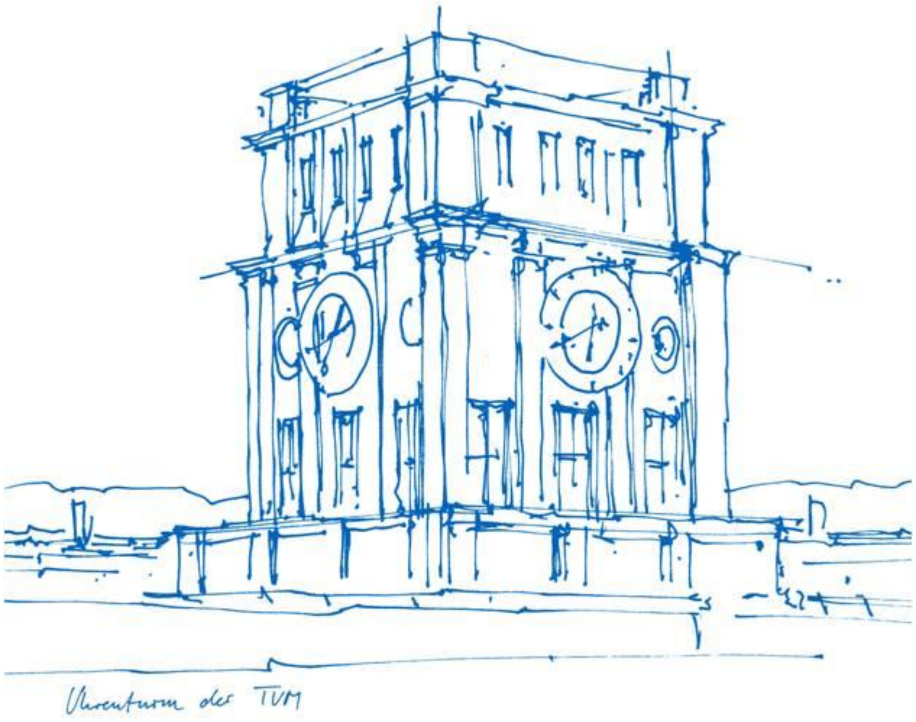
# Integration of Deep Optical Flow in Visual-Inertial Odometry

Semester Thesis

**Jingkun Feng**

Advisor: Mariia Gladkova

January 31st , 2022



# Outline

- Introduction and Motivation
- Preliminaries
  - Optical Flow
  - Basalt VIO
- Integration and Outlier Removal
- Evaluation
- Discussion
- Summary

# Introduction and Motivation

- Before, handcrafted optical flow
- Recently, deep optical flow with rise of deep learning
- Inspired by DF-VO from Zhang et al [2]
- **Aim** to explore probability of leveraging deep optical flow to improve the **accuracy** and **robustness** of a state-of-the-art VIO system.

# Preliminaries

## Optical Flow

- A displacement vector describes apparent motion of the same pixel in consecutive frames.

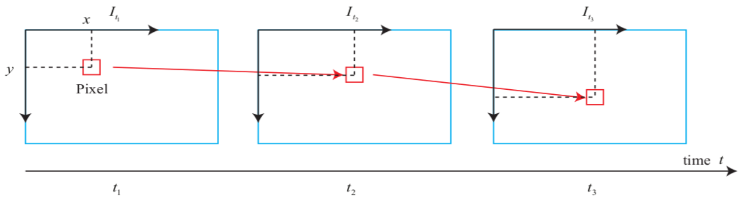



Fig 1. Optical flow for a single pixel. Constant intensity is assumed: $\boldsymbol{I}(x_1, y_1, t_1) = \boldsymbol{I}(x_1, y_1, t_1) = \boldsymbol{I}(x_1, y_1, t_1)$

- Useful for feature tracking
- Assumptions:
  - Brightness constancy
  - Constant motion in a local neighborhood (Lucas-Kanade method [5])
  - Spatially smooth motion (Horn-Schunck method [6])
- Sparse or dense vector field

Fig 2. Sparse optical flow

Fig 3. Color coded dense optical flow

# Preliminaries

## Basalt VIO [1]

- Consists of **visual-inertial odometry** and visual-inertial mapping
- Algorithm framework of Basalt VIO

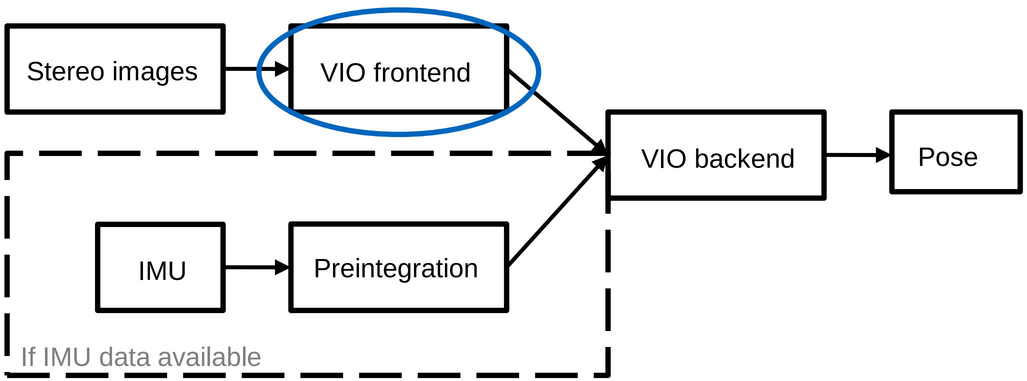



Fig 4. Basalt VIO framework

- Patch-based KLT for tracking
  - Locally-scaled sum of squared differences (LSSD)
  - Coarse-to-fine optimization using pyramidal approaches

# Preliminaries

## Basalt VIO

- Locally-scaled sum of squared differences (LSSD)
  - Patch $\Omega$
  - Desired transformation $\boldsymbol{T} \in SE(2)$ between two matching patches in adjacent images
  - Average intensity of all pixels in the patch $\overline{I}$
  - Residual $r$ of an increment $\xi$

$$r_i(\xi) = \frac{I_{t+1}(\boldsymbol{T}\boldsymbol{x}_i)}{\overline{I_{t+1}}} - \frac{I_t(\boldsymbol{x}_i)}{\overline{I_t}}$$

  - Minimize LSSD over patches to obtain $\boldsymbol{T}$

$$\underset{\mathbf{T}\epsilon\ \mathrm{SE}(2)}{argmin} \sum_{\boldsymbol{x}_i \in \Omega} \left(r_i(\xi)\right)^2$$

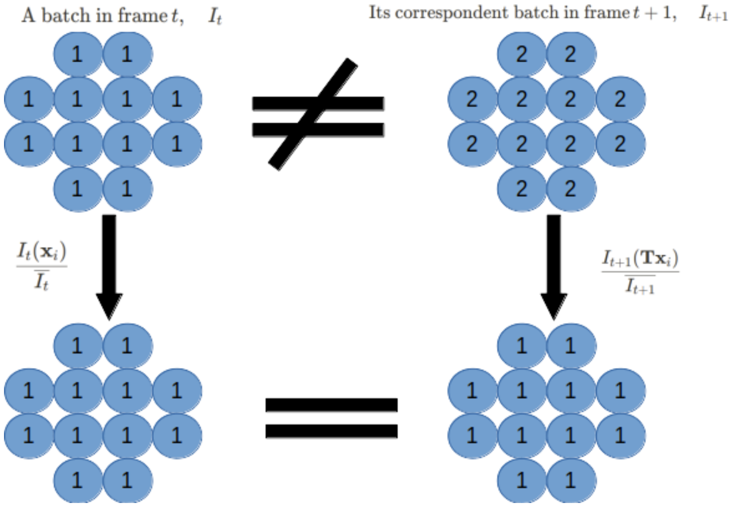



Fig 5. Main concept of LSSD

- Coarse-to-fine optimization using pyramidal approaches
  - Achieve robustness to large displacements in the image
  - The pyramid level is fixed

    → only robust to large displacements in certain degree

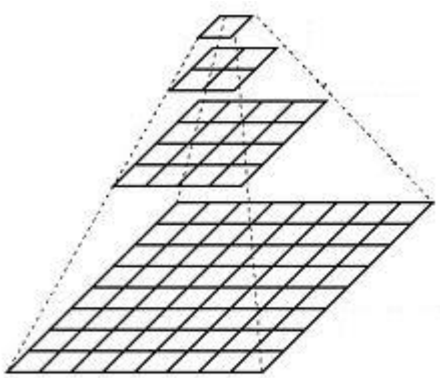
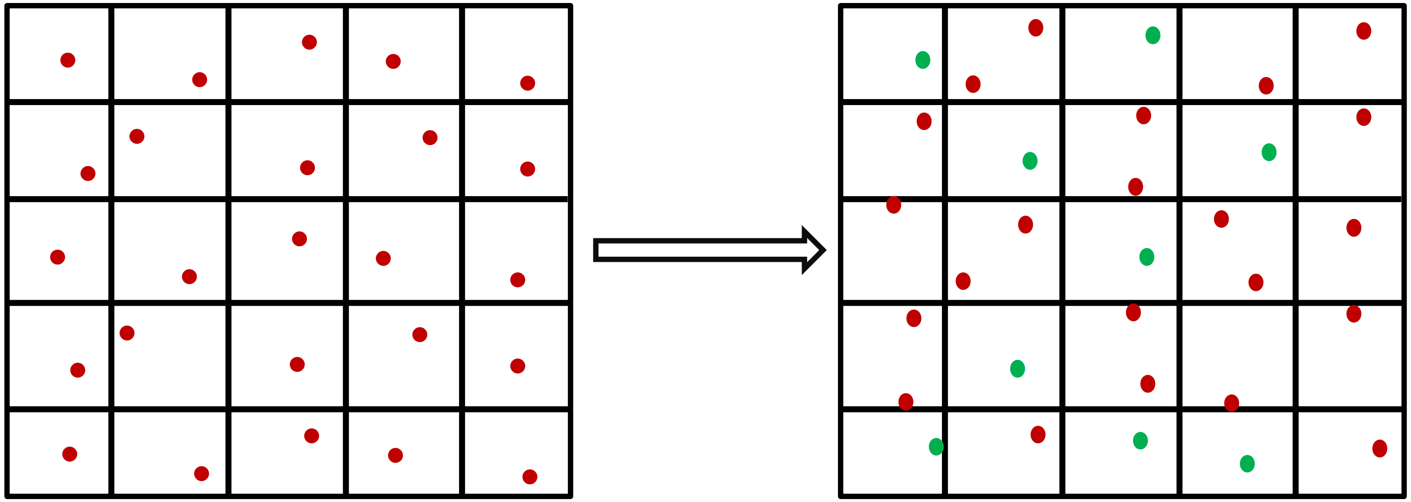


Fig 6. Image pyramid

Computer Vision Group
TUM Department of Informatics
Technical University of Munich

# Integration and Outlier Removal
## Integration

- Extract **FAST** keypoints
  - Split the image into regular cells
  - Extract and track the keypoint with strongest response in each cell
  - Resample if no keypoint remains in the cell



Initialization Resample

# Integration and Outlier Removal

## Integration

- Extract **FAST** keypoints
  - Split the image into regular cells
  - Extract and track the keypoint with strongest response in each cell
  - Resample if no keypoint remains in the cell

- Deep optical flow for temporal feature tracking
  - Predict forward optical flow using **Recurrent All-Pairs Field Transforms** (RAFT) # [3]
  - Use deep optical flow as prior to warp patches
  - **Refine by minimizing LSSD**

- Pyramidal KLT for stereo matching

# The model we used is the pretrained model released in the official repo of RAFT.

# Integration and Outlier Removal

## Outlier Removal

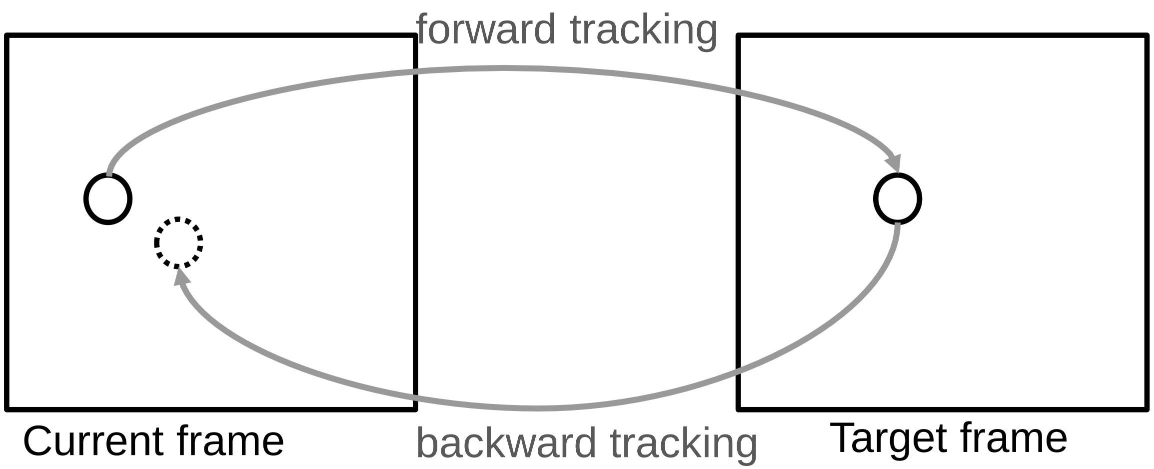
1. Forward-backward flow inconsistency
   - To remove outliers in temporal feature tracking

2. Epipolar constraint
   - To remove outliers in stereo matching

# Integration and Outlier Removal
## Outlier Removal

Forward-backward flow inconsistency

- Predict backward optical flow
- Track points from the current frame to the target frame and back
- Calculate distance between initial position and position after the second tracking
- Large distance denotes high inconsistency → to remove



Computer Vision Group
TUM Department of Informatics
Technical University of Munich

# Integration and Outlier Removal
## Outlier Removal

Epipolar constraint

- Check epipolar geometry of correspondences on stereo images
- Calibration → Fundamental matrix $\boldsymbol{F}$
- $x' \boldsymbol{F} \, x = 0$

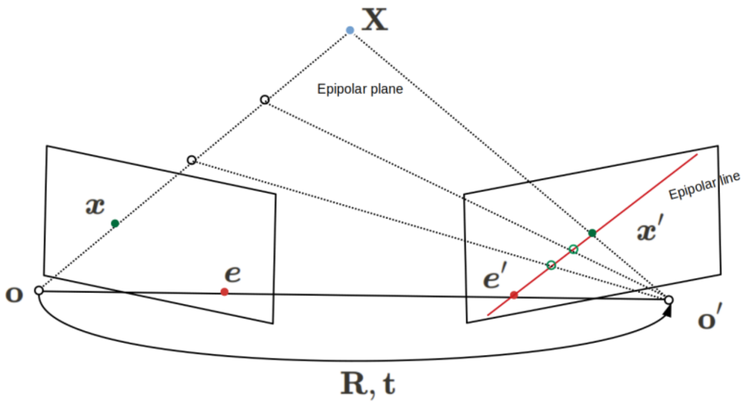



Fig 7. Epipolar geometry

- Remove points on the right frame if constraint is violated
- Keep points on the left frame

# Evaluation
## Dataset

1. KITTI Odometry [4]
   - 11 stereo sequences of various driving scenarios with ground-truth
   - Due to storage limitation, long sequences (02, 05, 08) are excluded
   - Grayscale and color images
   - **No IMU data**

2. EuRoC MAV [9]
   - 11 sequences of different difficulties with accurate motion ground-truth
   - Collected on-board a drone (6 DoF)
   - Grayscale images
   - **IMU measurements**

Computer Vision Group
TUM Department of Informatics
Technical University of Munich

# Evaluation
## Evaluation Metrices

1. Root mean squared absolute trajectory error: $ATE$
2. Relative pose error: translational $RPE_{tran}$ and rotational $RPE_{rot}$
3. Average translational and rotational error: $t_{err}$ and $r_{err}$

Notation:

- Estimated camera pose: $\boldsymbol{Q} \in SE(3)$
- Ground-truth camera pose: $\mathbf{P} \in SE(3)$
- Translation and rotation part of a rigid body transformation $\boldsymbol{T}$: $trans(\boldsymbol{T}), rot(\boldsymbol{T})$

Computer Vision Group
TUM Department of Informatics
Technical University of Munich

# Evaluation

## Evaluation Metrices – Root Mean Squared Absolute Trajectory Error ($ATE$)

- Evaluate global consistency
- Align the estimated and the ground-truth trajectory with a transformation matrix $\boldsymbol{S}$ (Horn method [])
- Absolute trajectory error matrix at time step $i$

$$\boldsymbol{E}_i \coloneqq \boldsymbol{Q}_i^{-1}\boldsymbol{S}\boldsymbol{P}_i$$

- Compute the root mean squared error over all time indices

$$ATE \coloneqq \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left|\left|trans(\boldsymbol{E}_i)\right|\right|^2}$$

# Evaluation

## Relative Pose Error ($RPE_{rot},\ RPE_{tran}$ )

- Evaluate local consistency
- Relative pose error matrix $\boldsymbol{F}_{i:\Delta} \coloneqq \left(\boldsymbol{Q}_i^{-1}\boldsymbol{Q}_{i+\Delta}\right)^{-1}\left(\boldsymbol{P}_i^{-1}\boldsymbol{P}_{i+\Delta}\right)$
- Translational part

$$RPE_{trans} \coloneqq \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left|\left|trans(\boldsymbol{F}_i)\right|\right|^2} \quad \text{for} \quad \mathrm{i} = 1, \dots, \mathrm{n}$$

- Rotational part

$$RPE_{rot} \coloneqq \frac{1}{m}\sum_{i=1}^{m}\angle\boldsymbol{F}_i \quad \text{for} \quad \mathrm{i} = 1, \dots, \mathrm{n} \quad \text{where} \quad \angle\boldsymbol{F}_i \coloneqq \arccos\left(\frac{tr\big(rot(\boldsymbol{F}_i)\big) - 1}{2}\right)$$

# Evaluation

## Average translational and rotational error

- Specific metric adopted to evaluation on KITTI Odometry
- Measures errors as function of the trajectory length

# Evaluation

## Evaluation Results

- On KITTI Odometry

| Method | Metric | 01 | 03 | 04 | 05 | 06 | 07 | 09 | 10 | Avg. excl. 01 |
|---|---|---|---|---|---|---|---|---|---|---|
| Original | $t_{err}$ | 4.5239 | 0.9962 | 1.1921 | 0.7646 | 1.0605 | 0.8625 | 1.0590 | **0.5892** | 0.9320 |
| | $r_{err}$ | 0.1713 | 0.2293 | **0.1922** | 0.2276 | **0.2313** | 0.4851 | 0.1937 | 0.2652 | 0.2606 |
| | $ATE$ | 30.7334 | 1.3648 | 1.2690 | 2.7245 | 2.5591 | 1.5547 | 4.3127 | 0.9834 | 2.1098 |
| | $RPE_{tran}$ | 0.6737 | 0.0143 | 0.0267 | 0.0136 | 0.0183 | 0.0113 | 0.0213 | 0.0139 | 0.0171 |
| | $RPE_{rot}$ | 0.0469 | 0.0328 | 0.0237 | 0.0309 | 0.0239 | 0.0281 | 0.0332 | 0.0383 | 0.0301 |
| Ours | $t_{err}$ | **1.7562** | **0.9033** | **0.9665** | **0.6996** | **0.9144** | X | **0.9602** | 0.6122 | **0.8427** |
| | $r_{err}$ | **0.1258** | **0.2144** | 0.2342 | **0.2262** | 0.2432 | X | **0.1819** | **0.2459** | **0.2243** |
| | $ATE$ | **5.1679** | **1.0309** | **1.0170** | **2.2426** | **2.4629** | X | **3.7208** | **0.9023** | **1.8961** |
| | $RPE_{tran}$ | **0.2653** | **0.0133** | **0.0240** | **0.0118** | **0.0146** | X | **0.0188** | **0.0131** | **0.0159** |
| | $RPE_{rot}$ | **0.0324** | **0.0325** | **0.0226** | **0.0303** | **0.0228** | X | **0.0322** | **0.0380** | **0.0297** |

**Table 1** Evaluation results on KITTI Odometry (Seq. 01, 03-07, 09, 10).

- On EuRoC MAV

| Method | Metric | MH_01 | MH_02 | MH_03 | MH_04 | MH_05 | V1_01 | V1_02 | V1_03 | V2_01 | V2_02 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Original | $ATE$ | 0.09081 | **0.05387** | 0.08488 | 0.10852 | 0.12732 | **0.04284** | 0.05636 | 0.07201 | 0.05636 | 0.06414 | 0.07650 |
| | $RPE_{tran}$ | 0.00138 | 0.00180 | 0.00374 | 0.00509 | 0.00370 | 0.00229 | 0.00295 | 0.00508 | 0.00118 | 0.00988 | 0.00371 |
| | $RPE_{rot}$ | 0.00040 | 0.00043 | 0.00055 | 0.00069 | 0.00054 | 0.00068 | 0.00086 | 0.00107 | 0.00067 | 0.00098 | 0.00069 |
| Ours | $ATE$ | **0.08618** | 0.05395 | **0.07096** | **0.10008** | **0.10767** | 0.04322 | **0.04114** | **0.04876** | **0.03777** | **0.03974** | **0.06295** |
| | $RPE_{tran}$ | **0.00136** | **0.00138** | **0.00353** | **0.00485** | **0.00357** | **0.00228** | **0.00265** | **0.00348** | **0.00110** | **0.00298** | **0.00272** |
| | $RPE_{rot}$ | **0.00038** | **0.00041** | **0.00054** | **0.00066** | **0.00051** | **0.00067** | **0.00082** | **0.00104** | **0.00065** | **0.00091** | **0.00066** |

**Table 2** Evaluation results on EuRoC MAV (V2_03 is excluded).

- Outperforms the original in terms of global and local accuracy
- However, our system fails at a single frame on KITTI 07.

# Discussion

## Failure Case


Fig 8. Failure case. Time step above is t and below is t+1.

# Discussion

## Ablation Study

1. Optical flow inference: Grayscale vs. color images

2. With or without refinement using LSSD

3. … (for other studies please refer to the paper)

# Discussion

## Ablation Study – Grayscale vs. RGB

- Color images are more informative than grayscale images
- Most existing datasets(e.g., Flyingthings [9] and Sintel [8]) contain merely color images.
- Currently proposed deep-learning-based methods mainly train on color images.

- Evaluated on KITTI Odometry

# Discussion

## Ablation Study – Grayscale vs. RGB

- Using color images for optical flow inference can boost performance in pose estimation.

| Method | Metric | 03 | 04 | 05 | 06 | 09 | 10 | Avg. |
|---|---|---|---|---|---|---|---|---|
| Grayscale | $t_{err}$ | 0.9033 | **0.9665** | 0.6996 | **0.9144** | **0.9602** | 0.6122 | 0.8427 |
| | $r_{err}$ | **0.2144** | 0.2342 | 0.2262 | 0.2432 | **0.1819** | 0.2459 | 0.2243 |
| | $ATE$ | 1.0309 | **1.0170** | 2.2426 | 2.4629 | **3.7208** | 0.9023 | 1.8961 |
| | $RPE_{tran}$ | **0.0133** | **0.0240** | 0.0118 | **0.0146** | 0.0188 | 0.0131 | 0.0159 |
| | $RPE_{rot}$ | 0.0325 | **0.0226** | 0.0303 | 0.0228 | 0.0322 | 0.0380 | 0.0297 |
| RGB | $t_{err}$ | **0.8827** | 1.0082 | **0.6946** | 0.9170 | 0.9644 | **0.5622** | **0.8382** |
| | $r_{err}$ | 0.2249 | **0.2282** | **0.2234** | **0.2420** | 0.1849 | **0.2278** | **0.2219** |
| | $ATE$ | **1.0049** | 1.0668 | **2.1745** | **2.3706** | 3.7324 | **0.8626** | **1.8686** |
| | $RPE_{tran}$ | 0.0134 | 0.0241 | 0.0118 | 0.0147 | 0.0188 | **0.0130** | 0.0159 |
| | $RPE_{rot}$ | **0.0324** | 0.0228 | 0.0303 | 0.0228 | 0.0322 | 0.0380 | 0.0297 |

**Table 3** Evaluation results of ablation study about the image format used for inference on KITTI Odometry (Seq.03, 04, 05, 06, 09, 10).

- But
  - the improvement is not significant, about 1% in average ATE.
  - only some of the datasets provide RGB images.

# Discussion
## Ablation Study – Refinement

- In general, refinement helps achieve more accurate trajectory estimation
- System with refined optical flow has obvious larger drift in KITTI 03 and 06

| Method | Metric | 03 | 04 | 05 | 06 | 09 | 10 | Avg. |
|---|---|---|---|---|---|---|---|---|
| Refined | $t_{err}$ | 0.9033 | **0.9665** | **0.6996** | **0.9144** | **0.9602** | **0.6122** | **0.8427** |
| | $r_{err}$ | **0.2144** | **0.2342** | **0.2262** | **0.2432** | **0.1819** | **0.2459** | **0.2243** |
| | $ATE$ | 1.0309 | 1.0170 | **2.2426** | 2.4629 | **3.7208** | **0.9023** | **1.8961** |
| | $RPE_{tran}$ | **0.0133** | **0.0240** | **0.0118** | **0.0146** | **0.0188** | **0.0131** | **0.0159** |
| | $RPE_{rot}$ | **0.0325** | **0.0226** | **0.0303** | **0.0228** | **0.0322** | **0.0380** | **0.0297** |
| Not refined | $t_{err}$ | **0.6714** | 1.0386 | 0.8153 | 1.0155 | 1.0376 | 0.6348 | 0.8689 |
| | $r_{err}$ | 0.2595 | 0.4916 | 0.2730 | 0.3138 | 0.2466 | 0.3638 | 0.3247 |
| | $ATE$ | **0.6747** | **0.9074** | 3.3607 | **2.0447** | 4.4613 | 1.1232 | 2.0953 |
| | $RPE_{tran}$ | 0.0147 | 0.0385 | 0.0154 | 0.0225 | 0.0242 | 0.0165 | 0.0220 |
| | $RPE_{rot}$ | 0.0333 | 0.0287 | 0.0327 | 0.0279 | 0.0352 | 0.0410 | 0.0331 |

**Table 4** Evaluation results of ablation study about refinement of the deep optical flow on KITTI Odometry (Seq.03, 04, 05, 06, 09, 10).

| Method | Metric | MH_01 | MH_02 | MH_03 | MH_04 | MH_05 | V1_01 | V1_02 | V1_03 | V2_01 | V2_02 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Refined | $ATE$ | **0.0862** | **0.0540** | **0.0710** | **0.1001** | **0.1077** | **0.0432** | **0.0411** | **0.0488** | **0.0378** | **0.0397** | **0.06295** |
| | $RPE_{tran}$ | **0.0014** | **0.0014** | **0.0035** | **0.0048** | **0.0036** | **0.0023** | **0.0026** | **0.0035** | **0.0011** | **0.0030** | **0.00272** |
| | $RPE_{rot}$ | **0.0004** | **0.0004** | **0.0005** | **0.0007** | **0.0005** | **0.0007** | **0.0008** | **0.0010** | **0.0007** | **0.0009** | **0.00066** |
| Not refined | $ATE$ | 0.2238 | 0.1707 | 0.1429 | 0.4098 | 0.3747 | 0.0543 | 0.0549 | 0.0508 | 0.0397 | 0.0530 | 0.15746 |
| | $RPE_{tran}$ | 0.0022 | 0.0027 | 0.0044 | 0.0082 | 0.0063 | 0.0024 | 0.0030 | 0.0035 | 0.0014 | 0.0026 | 0.00368 |
| | $RPE_{rot}$ | 0.0005 | 0.0005 | 0.0006 | 0.0008 | 0.0006 | 0.0007 | 0.0009 | 0.0011 | 0.0007 | 0.0009 | 0.00074 |

**Table 5** Evaluation results of ablation study about refinement of the deep optical flow on EuRoC MAV.

# Discussion

## Ablation Study – Refinement

- In general, refinement helps achieve more accurate trajectory estimation
- System with refined optical flow has obviously larger drift on KITTI 03 and 06

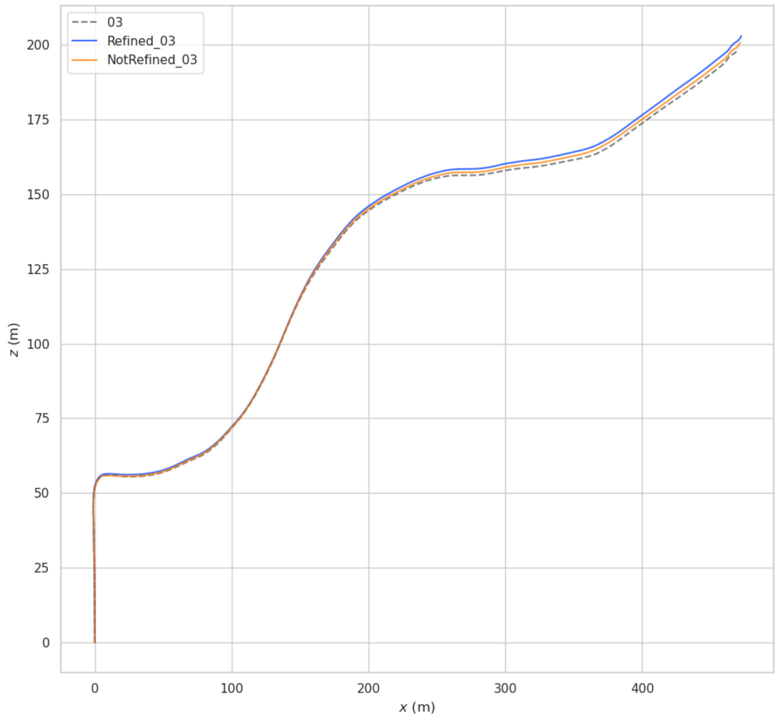



Fig 9. Estimated trajectory of KITTI 03

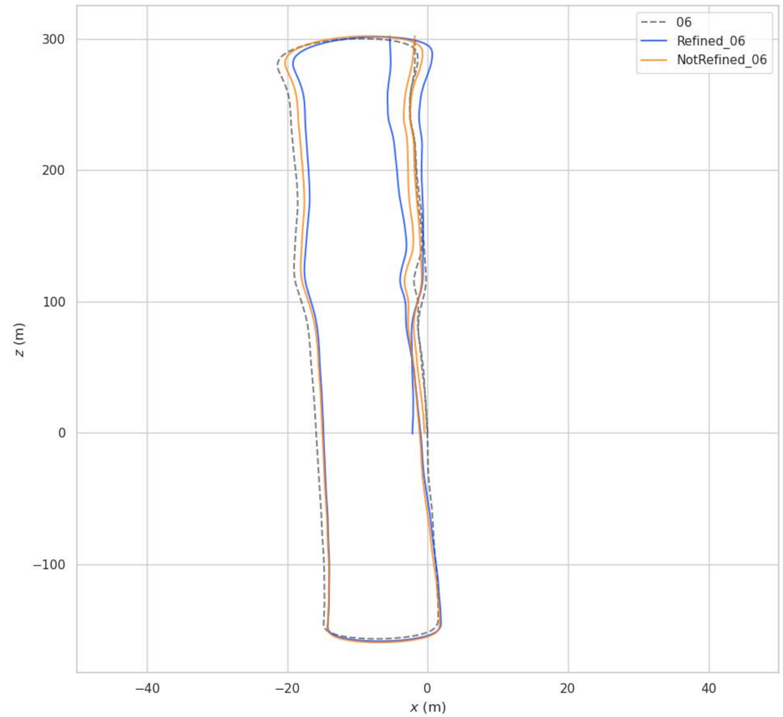



Fig 10. Estimated trajectory of KITTI 06

# Discussion
## Timing and Efficiency

- **Not efficient**
  - A huge part of available information is not in use.
    - About 300 pixels out of (370×1226) pixels

- **Not real-time capable**
  - Original Basalt VIO is around 4 times faster than real-time
    - Frame rate of EuRoC is 30 fps (0.03s per frame)
    - About 7.5 ms per frame on EuRoC
  - However, Optical flow inference is very "time consuming".
    - 0.4 s per frame on EuRoC using RAFT

# Summary

- We extended the Basalt VIO by integrating deep optical flow
  - replace the pyramid KLT tracker in BASALT VIO with refined deep optical flow
  - remove outliers using forward-backward flow inconsistency and epipolar constraint

- According to the evaluation, our system outperforms the original Basalt VIO w.r.t accuracy of trajectory estimation.

- However, our integration has drawbacks
  - less robust to dynamic objects
  - inefficient in terms of the usage of available information
  - not real time capable

Computer Vision Group
TUM Department of Informatics
Technical University of Munich

# Thank you!

# Reference

[1] V. Usenko, N. Demmel, D. Schubert, J. Stuckler, and D. Cremers, “Visual-Inertial Mapping With Non-Linear Factor Recovery,” IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 422–429, Apr. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8938825

[2] H. Zhan, C. S. Weerasekera, J.-W. Bian, R. Garg, and I. Reid, “DF-VO: What Should Be Learnt for Visual Odometry?” arXiv:2103.00933 [cs], Mar. 2021, arXiv: 2103.00933. [Online]. Available: http://arxiv.org/abs/2103.00933

[3] Z. Teed and J. Deng, “RAFT: Recurrent All-Pairs Field Transforms for Optical Flow,” Available: http://arxiv.org/abs/2003.12039

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” The International Journal of Robotics Research, vol. 32, no. 11, pp. 1231–1237, Sep. 2013. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364913491297

[5] B. D. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision,” p. 10.

[6] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” Artificial Intelligence, vol. 17, no. 1, pp. 185–203, 1981. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0004370281900242

# Reference

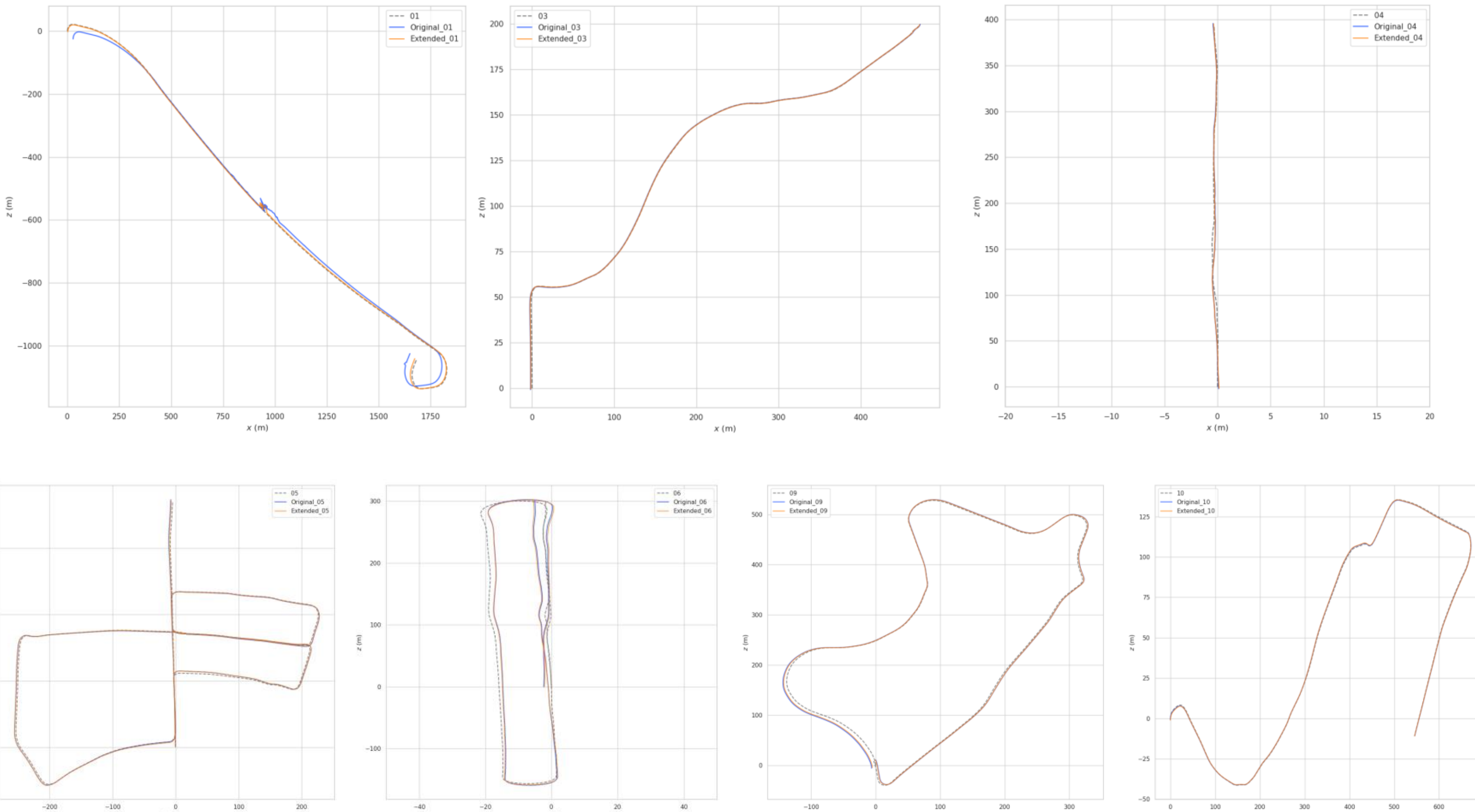
[7] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura-Algarve, Portugal: IEEE, Oct. 2012, pp. 573–580. [Online]. Available: http://ieeexplore.ieee.org/document/6385773

[8] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A Naturalistic Open Source Movie for Optical Flow Evaluation," in Computer Vision – ECCV 2012, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 7577, pp. 611–625, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-642-33783-3_44

[9] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," The International Journal of Robotics Research, vol. 35, no. 10, pp. 1157–1163, Sep. 2016, publisher: SAGE Publications Ltd STM. [Online]. Available: https://doi.org/10.1177/0278364915620033

Computer Vision Group
TUM Department of Informatics
Technical University of Munich

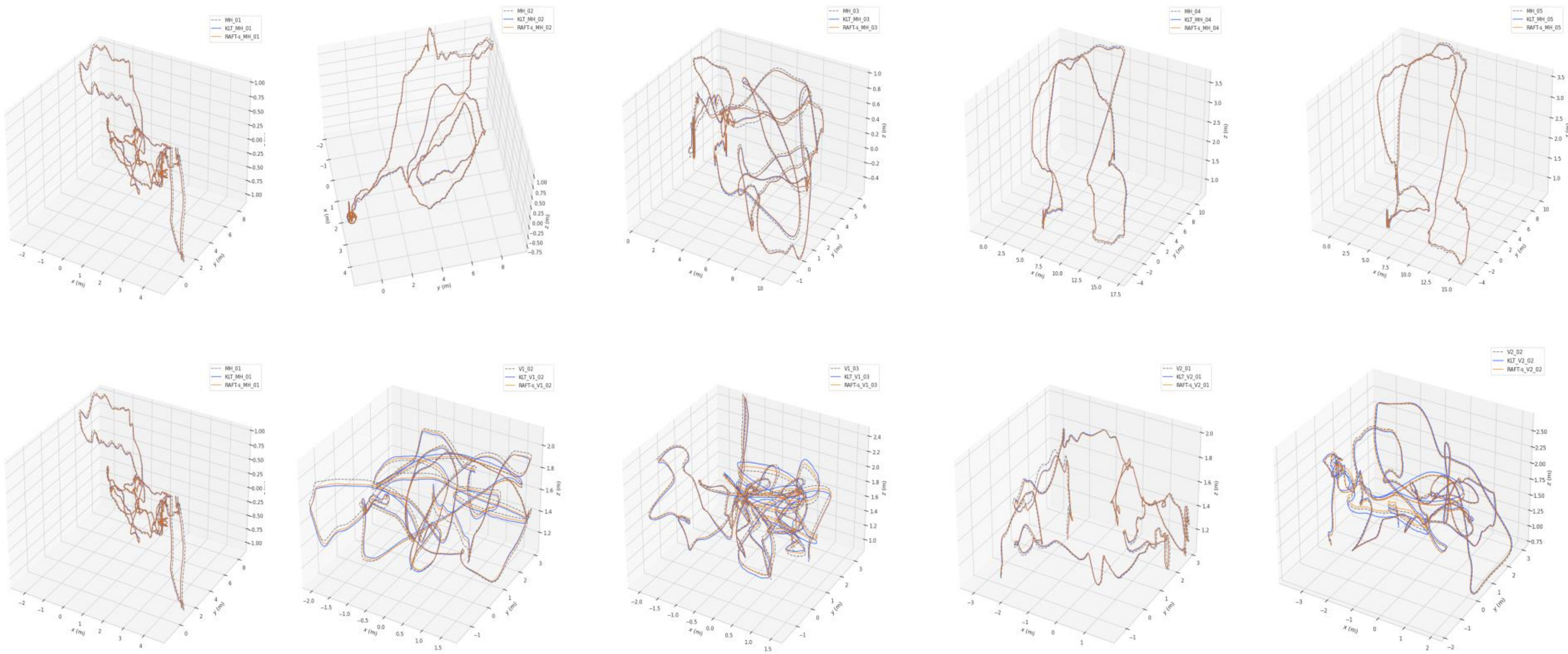
# Appendix
## Qualitative Evaluation Results – KITTI Odometry




Qualitative evaluation results on KITTI Odometry Seq. 01, 03-06, 09, 10

# Appendix

## Qualitative Evaluation Results – EuRoC MAV



Qualitative evaluation results on EuRoC MAV (V2_03 is excluded)