SAMUEL BRYNER

Event-based, Direct Camera Tracking from a Photometric Map using Nonlinear Optimization

Supervision:
Henri Rebecq, Dr. Guillermo Gallego, Mathias Gehrig, Prof. Dr. Davide Scaramuzza

Master Thesis

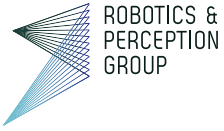
Robotics and Perception Group
Department of Informatics,
University of Zurich

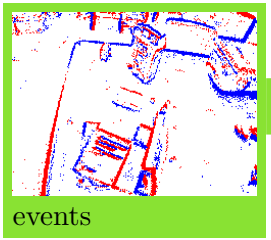
August 2018

University of Zurich[UZH]

ETH zürich

Institute of Neuroinformatics





Motion Tracking/SLAM with Event Cameras

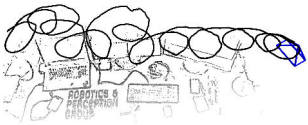
- "EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping in Real-Time"
  Rebecq et al., RA-L'17
- "Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera"
  Kim et al., ECCV'16

Motion Tracking/SLAM with Event Cameras

- "EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping in Real-Time"
  Rebecq et al., RA-L'17
- "Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera"
  Kim et al., ECCV'16
- "Event-based, 6-DOF Camera Tracking from Photometric Depth Maps"
  Gallego et al., PAMI'17

Motion Tracking/SLAM with Event Cameras

- "EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping in Real-Time"
  Rebecq et al., RA-L'17
- "Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera"
  Kim et al., ECCV'16
- "Event-based, 6-DOF Camera Tracking from Photometric Depth Maps"
  Gallego et al., PAMI'17
- "Asynchronous, Photometric Feature Tracking using Events and Frames"
  Gehrig et al., ECCV'18



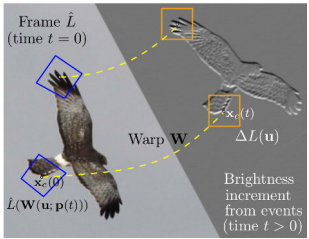
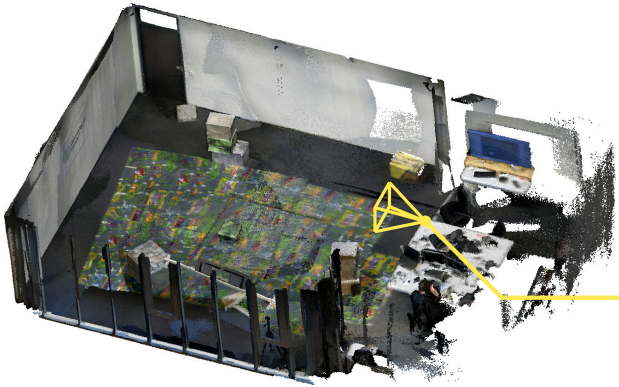


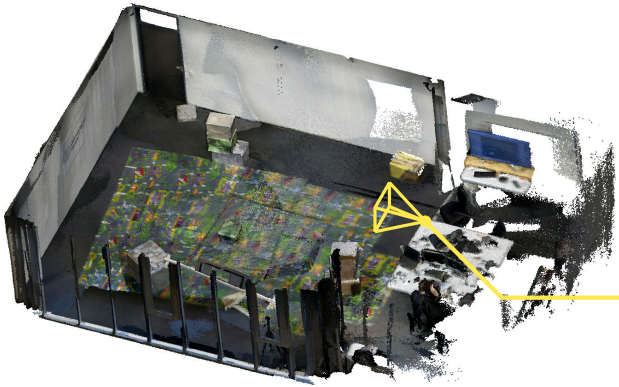
Image matching: Find view that most closely resembles the captured image.

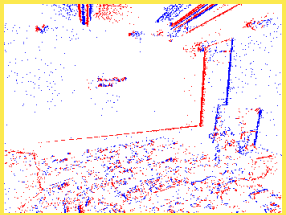


camera image

map view

Image matching: Find view that most closely resembles the captured image.

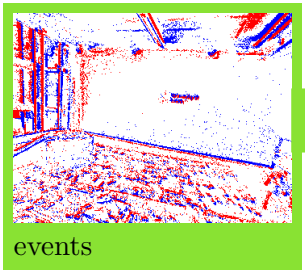


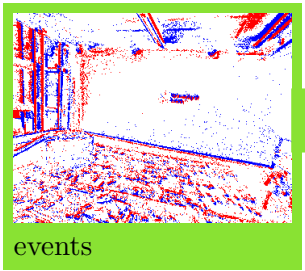


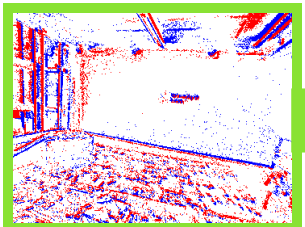
camera image

???

map view


events
measured $\Delta I(\boldsymbol{u})$
error calculation
map
predicted $\Delta \hat{I}(\boldsymbol{u})$


events
measured $\Delta I(\boldsymbol{u})$
error calculation
map
predicted $\Delta \hat{I}(\boldsymbol{u})$
$\dot{\boldsymbol{T}}$ velocity

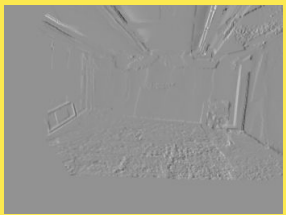


events

measured $\Delta I(\boldsymbol{u})$

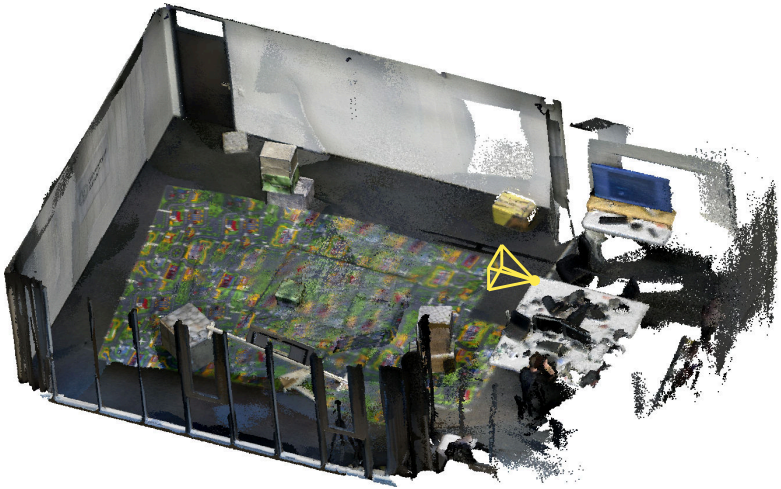
error calculation

map



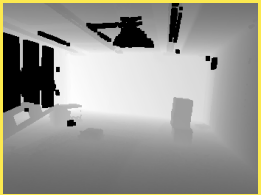
predicted $\Delta \hat{I}(\boldsymbol{u})$

$\dot{\boldsymbol{T}}$ velocity



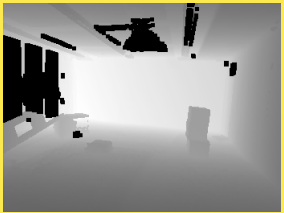
$\mathcal{M}$

$I(\boldsymbol{u})$ brightness



$Z(\boldsymbol{u})$ depth

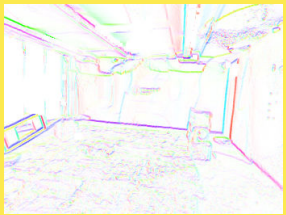
$I(\boldsymbol{u})$ brightness



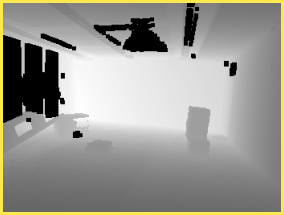
$Z(\boldsymbol{u})$ depth

$I(\boldsymbol{u})$ brightness

∇



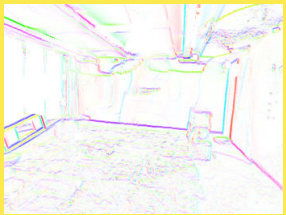
$\nabla I(\boldsymbol{u})$ gradient



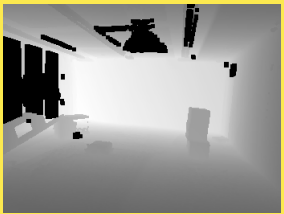
$Z(\boldsymbol{u})$ depth

$I(\boldsymbol{u})$ brightness

∇

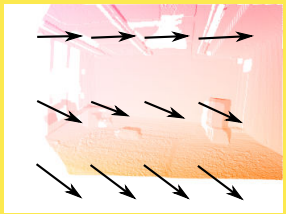


$\nabla I(\boldsymbol{u})$ gradient



$Z(\boldsymbol{u})$ depth

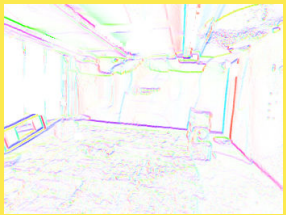
$J_{\boldsymbol{p}}$



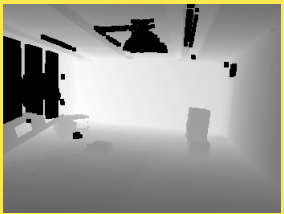
$\dot{\boldsymbol{u}}$ optic flow

$I(\boldsymbol{u})$ brightness

∇

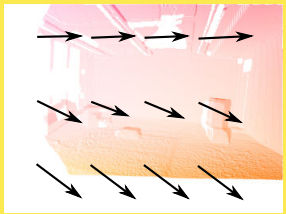


$\nabla I(\boldsymbol{u})$ gradient



$Z(\boldsymbol{u})$ depth

$J_{\boldsymbol{p}}$

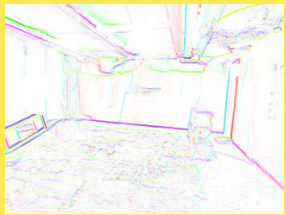


$\dot{\boldsymbol{u}}$ optic flow

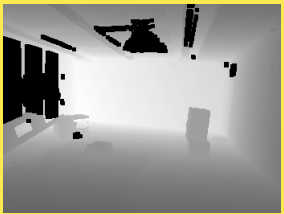
$\dot{\boldsymbol{T}}$ velocity

$I(\boldsymbol{u})$ brightness

∇

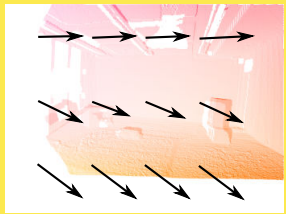


$\nabla I(\boldsymbol{u})$ gradient



$Z(\boldsymbol{u})$ depth

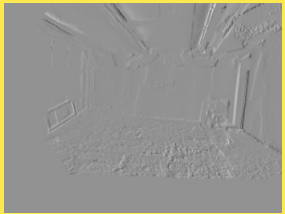
$J_{\boldsymbol{p}}$



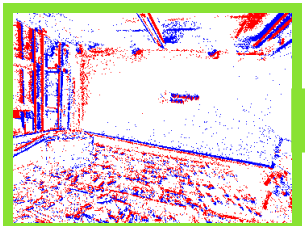
$\dot{\boldsymbol{u}}$ optic flow

$\nabla I \cdot \dot{\boldsymbol{u}} \Delta t$

$\dot{\boldsymbol{T}}$ velocity



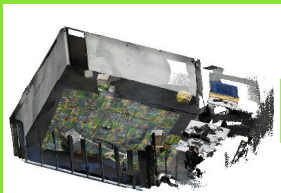
$\Delta \hat{I}(\boldsymbol{u})$ expected brightness change



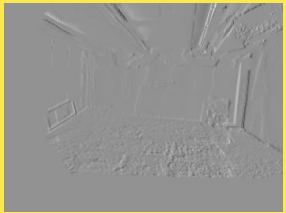
events

measured $\Delta I(\boldsymbol{u})$

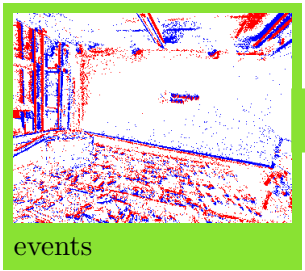
error calculation



map



predicted $\Delta \hat{I}(\boldsymbol{u})$

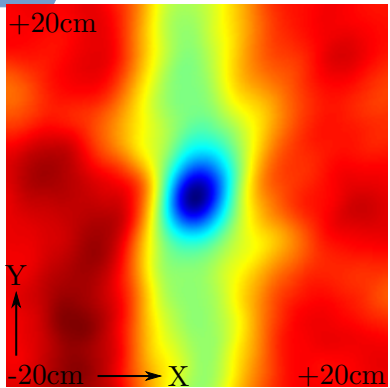

events
measured $\Delta I(\boldsymbol{u})$
error calculation
map
predicted $\Delta \hat{I}(\boldsymbol{u})$

Estimated brightness change

$$\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}) = -\nabla I(\tau(\boldsymbol{u}, \boldsymbol{T}), \mathcal{M}) \cdot \dot{\boldsymbol{u}}(\dot{\boldsymbol{T}})\Delta t$$

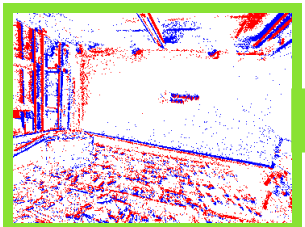
Measured brightness change

$$\Delta I(\boldsymbol{u}, \{e_k\}) = \sum_{t_k \in \Delta t} p_k C \delta(\boldsymbol{u} - \boldsymbol{u}_k)$$



Maximum likelihood

$$p(\{e_k\}|\boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}) \propto \exp\left(-\frac{1}{2\sigma^2}\int_{\mathcal{U}} (\Delta I(\boldsymbol{u}) - \Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}))^2 d\boldsymbol{u}\right)$$

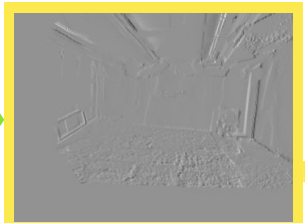


events

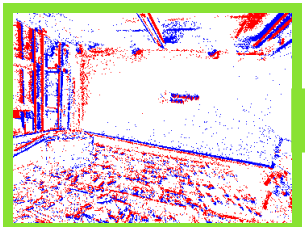
measured $\Delta I(\boldsymbol{u})$

error calculation

map



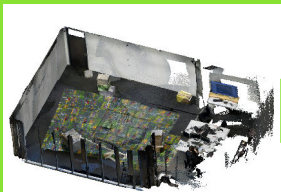
predicted $\Delta \hat{I}(\boldsymbol{u})$



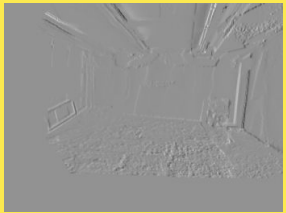
events

measured $\Delta I(\boldsymbol{u})$

error calculation



map

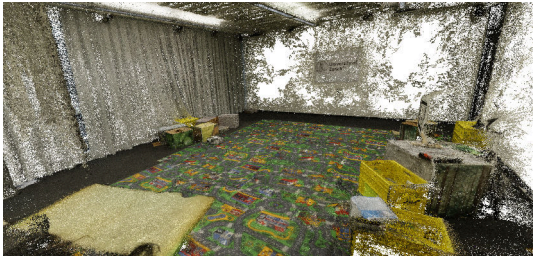


predicted $\Delta \hat{I}(\boldsymbol{u})$

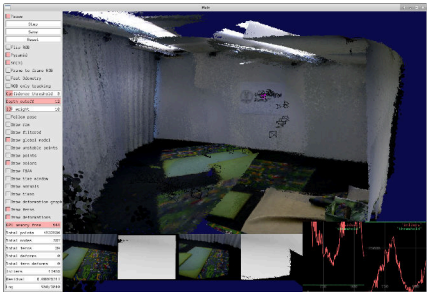
We need a map! And it should use affordable technology.

Software	Method	Sensors	Usage	Quality
ROVIOLI, voxblox	VI-SLAM & BA	Camera, IMU, depth	-	-
Agisoft PhotoScan	SfM	Camera	++	++
ElasticFusion	RGB-D-SLAM	Camera, depth	+	++

ROVIOLI + voxblox

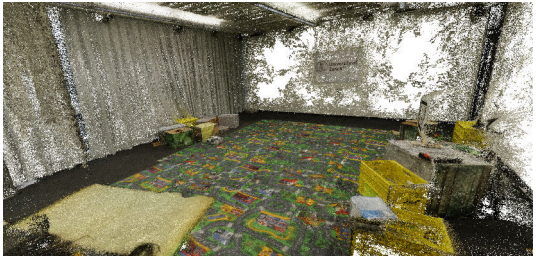


PhotoScan

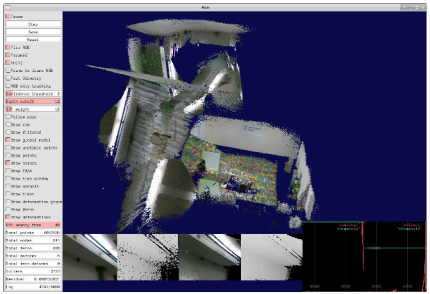


ElasticFusion

ROVIOLI + voxblox

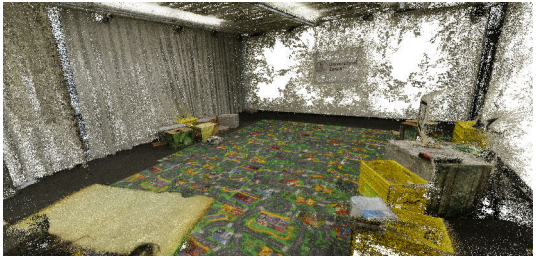


PhotoScan

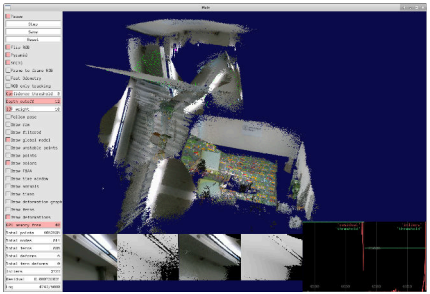


ElasticFusion

ROVIOLI + voxblox



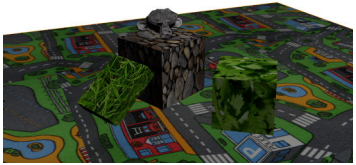
PhotoScan



ElasticFusion

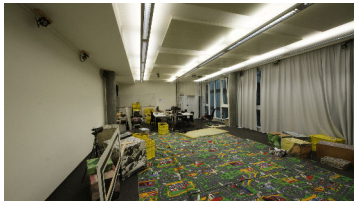
ElasticFusion + OptiTrack

Evaluation on both **simulated** and **real** data:

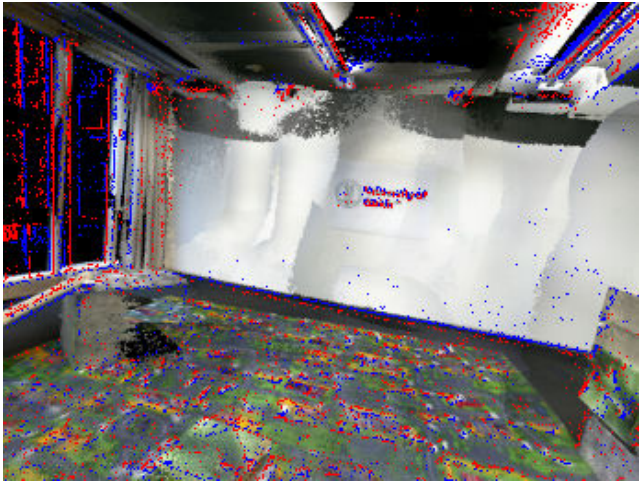


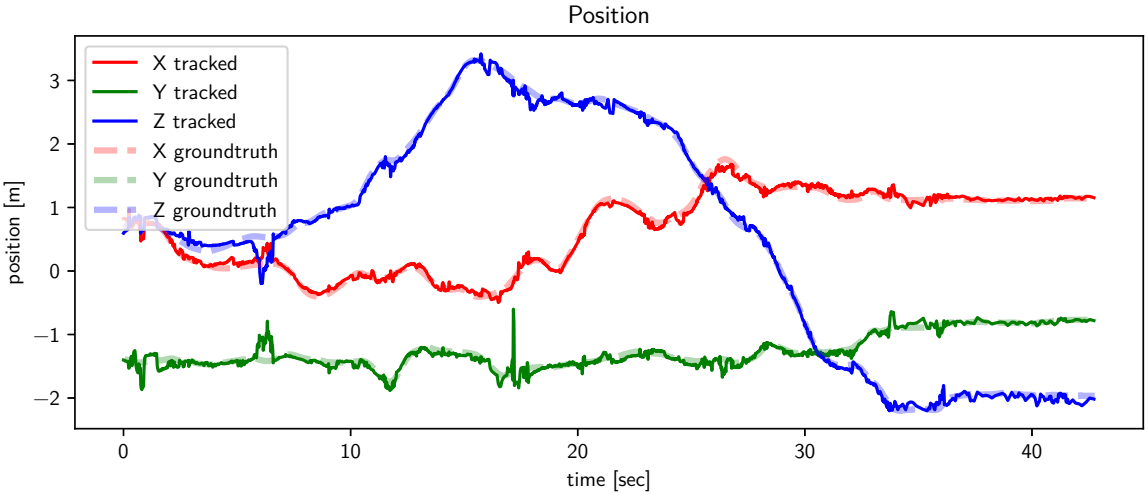
Boxes

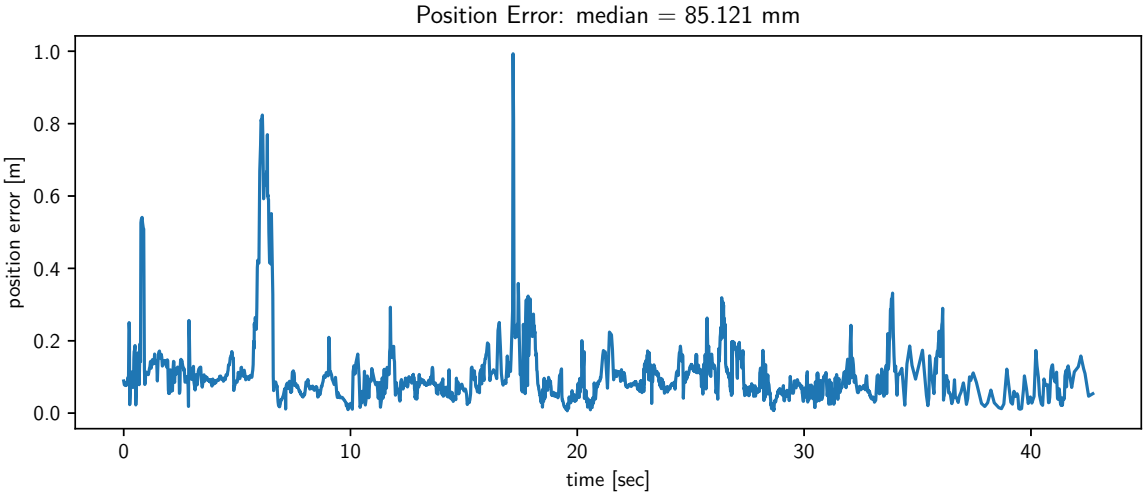
Sponza



Flying Room




Position
X tracked
Y tracked
Z tracked
X groundtruth
Y groundtruth
Z groundtruth
position [m]
3
2
1
0
−1
−2
0
10
20
30
40
time [sec]


Position Error: median = 85.121 mm
position error [m]
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
time [sec]

Scene	Mean Depth (mm)	Pose	
		Position (mm)	Rotation (°)
Boxes	2000	4.5	0.20
Sponza	10000	21	0.14
Sponza*	10000	45	0.36
Flying Room	3400	85	3.84

* downsampled, i.e. using lower resolution

		Processing		
	Length (s)	Time (s)	Slowdown	Events/sec
Boxes	20	10 772	539	3160
Sponza	10	638	64	4131
Sponza*	10	229	23	11 532
Flying Room, track	43.2	52 094	1206	1016

* downsampled, i.e. using lower resolution

Gallego, PAMI'17
Accuracy: 2-6%
Speed: 31k events/sec

This Work
Accuracy:
- Real data: 4.2%
- Simulation: 0.3%

Speed: 1-13k events/sec

Accuracy is given as RMS relative to scene depth.

CONCLUSION

- Photometric tracking works, if...
  - there is enough texture
  - no close objects that dominate events/flow

CONCLUSION

- Photometric tracking works, if...
  - there is enough texture
  - no close objects that dominate events/flow
- accuracy of about 4% for real data and 0.2% in simulation
- not realtime

CONCLUSION

- Photometric tracking works, if...
  - there is enough texture
  - no close objects that dominate events/flow
- accuracy of about 4% for real data and 0.2% in simulation
- not realtime
- There are good and cheap methods to create accurate 3D scans.

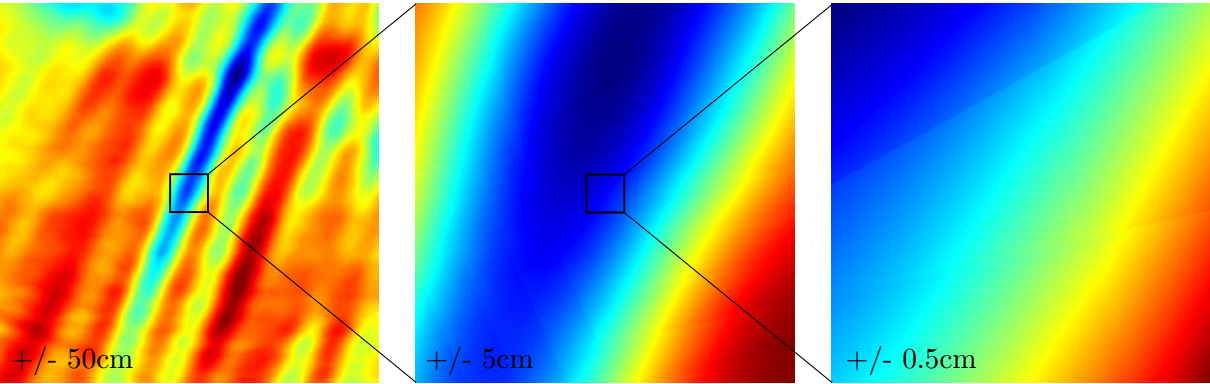
APPENDIX

OUTLOOK

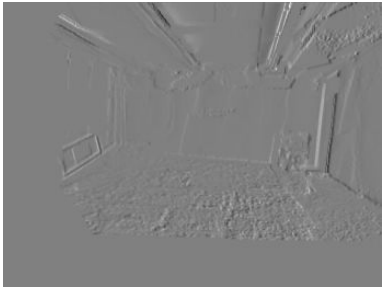
- Optimize implementation (GPU)
- Filter output, tune optimization
- IMU integration

ERROR FUNCTION

Translation in X/Y (not centered around minima)

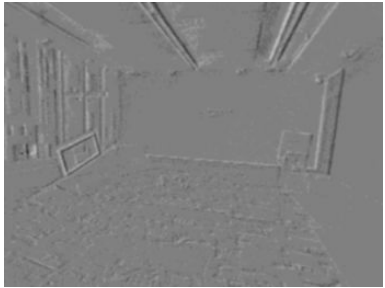


ERROR FUNCTION



Estimated brightness change

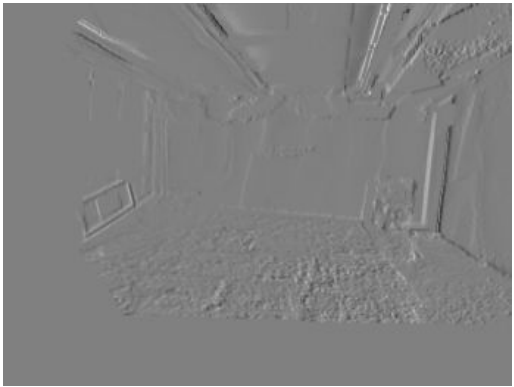
$$\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}) = \nabla I(\tau(\boldsymbol{u}, \boldsymbol{T}), \mathcal{M}) \cdot \dot{\boldsymbol{u}}(\dot{\boldsymbol{T}})\Delta t$$



Measured brightness change

$$\Delta I(\boldsymbol{u}, \{e_k\}) = \sum_{t_k \in \Delta t} p_k C \delta(\boldsymbol{u} - \boldsymbol{u}_k)$$

ERROR FUNCTION



Estimated brightness change

$$\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}) = \nabla I(\tau(\boldsymbol{u}, \boldsymbol{T}), \mathcal{M}) \cdot \dot{\boldsymbol{u}}(\dot{\boldsymbol{T}})\Delta t$$

ERROR FUNCTION

Estimated brightness change

$$\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}) = \nabla I(\tau(\boldsymbol{u}, \boldsymbol{T}), \mathcal{M}) \cdot \dot{\boldsymbol{u}}(\dot{\boldsymbol{T}})\Delta t$$

Measured brightness change

$$\Delta I(\boldsymbol{u}) = \sum_{t_k \in \Delta t} p_k C \delta(\boldsymbol{u}, \boldsymbol{u}_k)$$

Final minimization problem

$$\underset{\boldsymbol{T}, \dot{\boldsymbol{T}}}{\operatorname{argmin}} \sum \left(\frac{\Delta I(\boldsymbol{u})}{\|\Delta I(\boldsymbol{u})\|} - \frac{\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}})}{\|\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}})\|} \right)^2$$

ERROR FUNCTION

Estimated brightness change

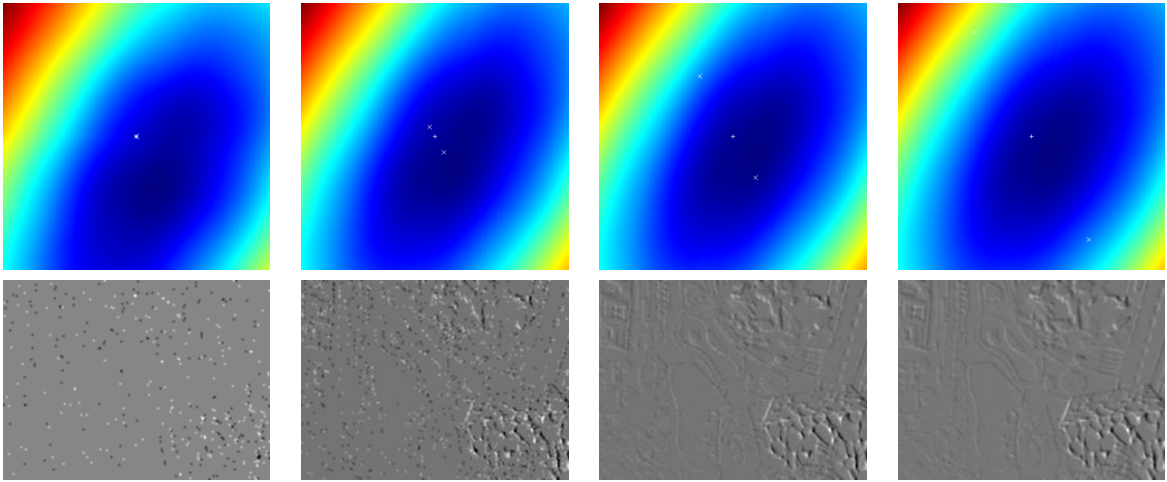
$$\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}}, \mathcal{M}) = \nabla I(\tau(\boldsymbol{u}, \boldsymbol{T}), \mathcal{M}) \cdot \dot{\boldsymbol{u}}(\dot{\boldsymbol{T}}) \Delta t$$

Measured brightness change

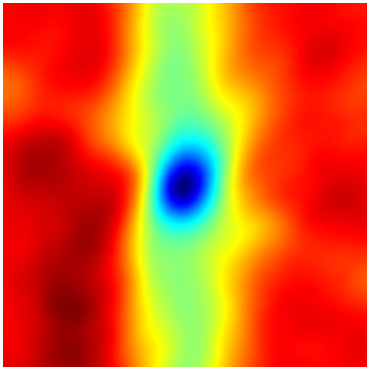
$$\Delta I(\boldsymbol{u}) = \sum_{t_k \in \Delta t} p_k C \delta(\boldsymbol{u}, \boldsymbol{u}_k)$$

Final minimization problem

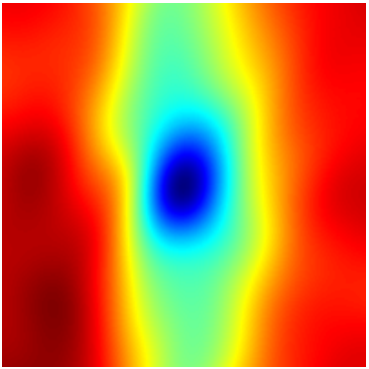
$$\underset{\boldsymbol{T}, \dot{\boldsymbol{T}}}{\operatorname{argmin}} \sum \left(\frac{\Delta I(\boldsymbol{u})}{\|\Delta I(\boldsymbol{u})\|} - \frac{\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}})}{\|\Delta\hat{I}(\boldsymbol{u}, \boldsymbol{T}, \dot{\boldsymbol{T}})\|} \right)^2$$



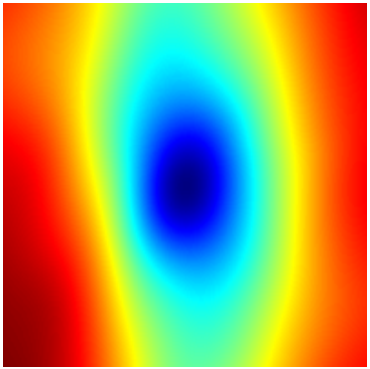
0.01 events/px	0.05 events/px	0.25 events/px	0.5 events/px
259 events	1296 events	6480 events	12960 events



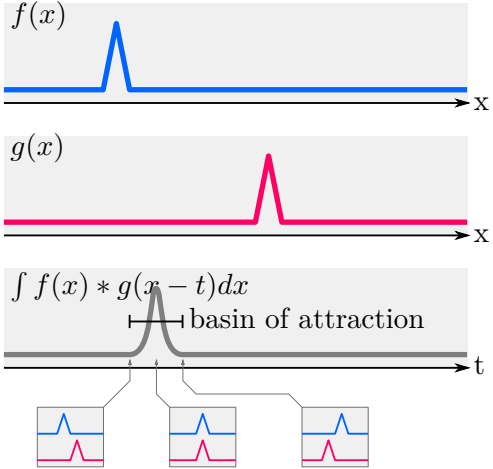
no blur

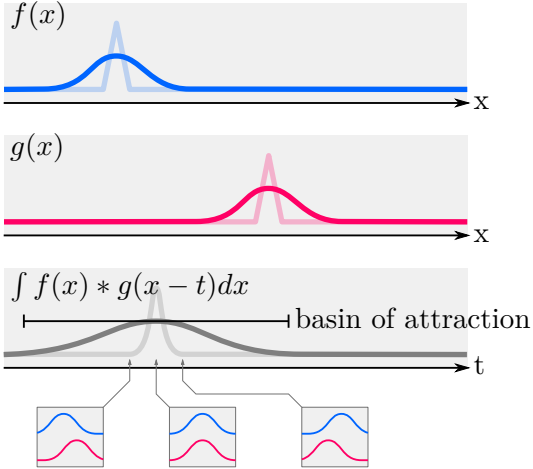


9px



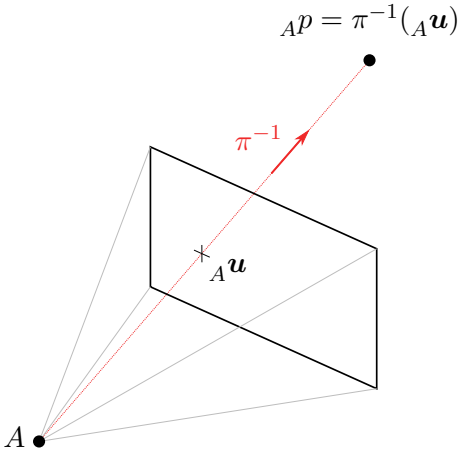
16px


$f(x)$
x
$g(x)$
x
$\int f(x) * g(x-t)dx$
basin of attraction
t


$f(x)$
x
$g(x)$
x
$\int f(x) * g(x-t)dx$
basin of attraction
t

back-projection

$$\pi^{-1}(\boldsymbol{u}, z) := z \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix}$$

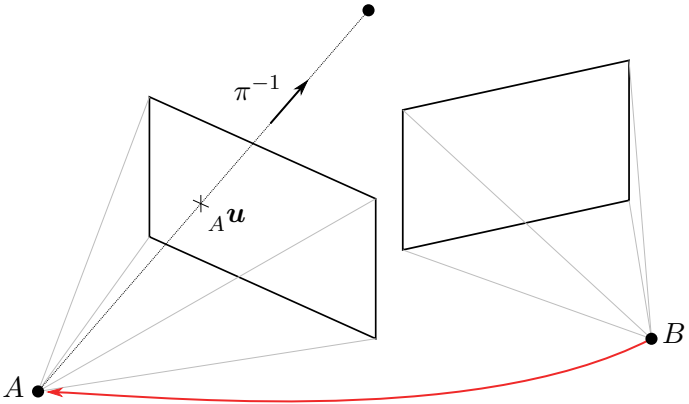


back-projection

$$\pi^{-1}(\boldsymbol{u}, z) := z \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix}$$

$$_A p = \pi^{-1}(_A \boldsymbol{u})$$

$$_B p = \boldsymbol{R}_{BA} \cdot {}_A \boldsymbol{p} + \boldsymbol{t}_{BA}$$



back-projection

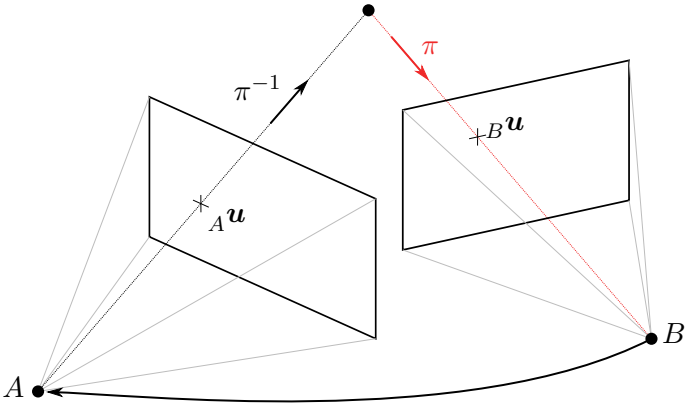
$$\pi^{-1}(\boldsymbol{u}, z) := z \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix}$$

projection

$$\pi(\boldsymbol{p}) := \frac{1}{p_z} \begin{bmatrix} p_x \\ p_y \end{bmatrix}$$

$$_A p = \pi^{-1}(_A \boldsymbol{u})$$

$$_B p = \boldsymbol{R}_{BA} \cdot {}_A \boldsymbol{p} + \boldsymbol{t}_{BA}$$



$$\boldsymbol{T}_{BA} = \boldsymbol{R}_{BA}, \boldsymbol{t}_{BA}$$

WARPING

back-projection

$$\pi^{-1}(\boldsymbol{u}, z) := z \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix}$$

projection

$$\pi(\boldsymbol{p}) := \frac{1}{p_z} \begin{bmatrix} p_x \\ p_y \end{bmatrix}$$

warping

$$\tau(\boldsymbol{u}) := \pi(\boldsymbol{R}\pi^{-1}(\boldsymbol{u}, z) + \boldsymbol{t})$$

$I(\boldsymbol{u})$ brightness

τ

$I(\tau(\boldsymbol{u}))$ warped bright.

image warping

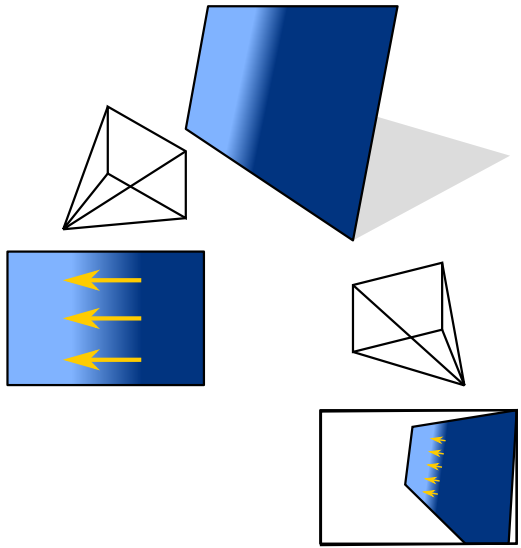
$$I({}_A\boldsymbol{u}) = I(\tau_{AB}({}_B\boldsymbol{u}))$$

$$\nabla I({}_A\boldsymbol{u}) = \nabla I(\tau_{AB}({}_B\boldsymbol{u})) \cdot \left.\frac{\partial \tau}{\partial \boldsymbol{u}}\right|_{{}_B\boldsymbol{u}}$$

correction factor

$$\left.\frac{\partial \tau}{\partial \boldsymbol{u}}\right|_{{}_B\boldsymbol{u}} = \left.\frac{\partial}{\partial \boldsymbol{u}} \pi(\boldsymbol{R}\pi^{-1}(\boldsymbol{u}) + \boldsymbol{t})\right|_{{}_B\boldsymbol{u}}$$

$$= D_\pi \left|_{{}_A\boldsymbol{p}} \cdot \boldsymbol{R}_{AB} \cdot D_{\pi^{-1}}\right|_{{}_B\boldsymbol{u}}$$



It depends on your exact defintion of $\pi^{-1}(\boldsymbol{u})$ ...

correction factor

$$\left.\frac{\partial \tau}{\partial \boldsymbol{u}}\right|_{B\boldsymbol{u}} = D_{\pi}\Big|_{A\boldsymbol{p}} \cdot \boldsymbol{R}_{AB} \cdot \left. D_{\pi^{-1}}\right|_{B\boldsymbol{u}}$$

Constant Z

$$\pi^{-1}(\boldsymbol{u}) = \pi_z^{-1}(\boldsymbol{u})$$

$$D_u \pi_z^{-1} = \begin{bmatrix} z & 0 \\ 0 & z \\ 0 & 0 \end{bmatrix}$$

Independent Z

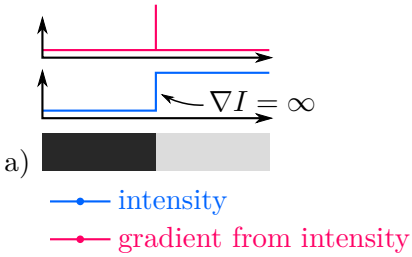
$$\pi^{-1}(\boldsymbol{u}) = \pi^{-1}(\boldsymbol{u}, z)$$

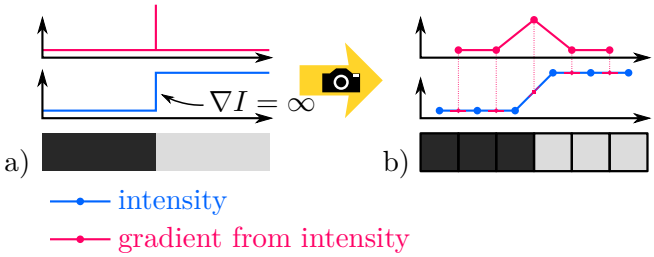
$$D_u \pi^{-1} = \begin{bmatrix} z & 0 & u_x \\ 0 & z & u_y \\ 0 & 0 & 1 \end{bmatrix}$$

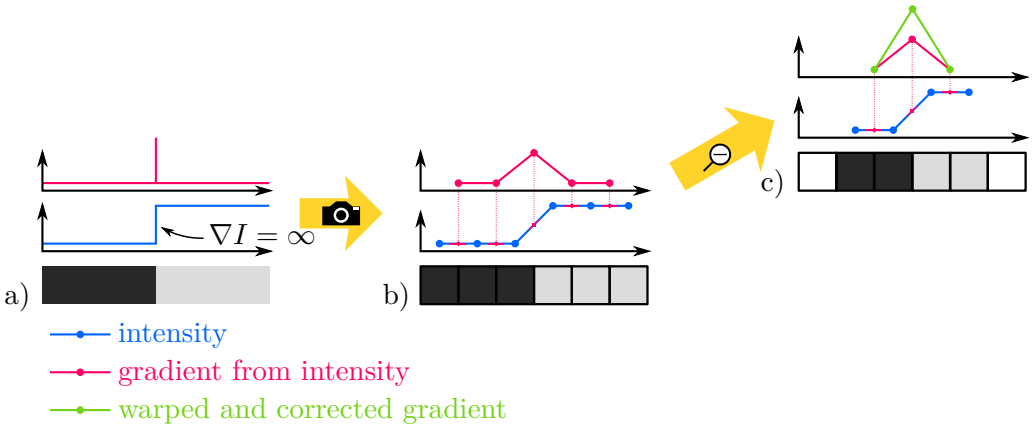
Dependent Z

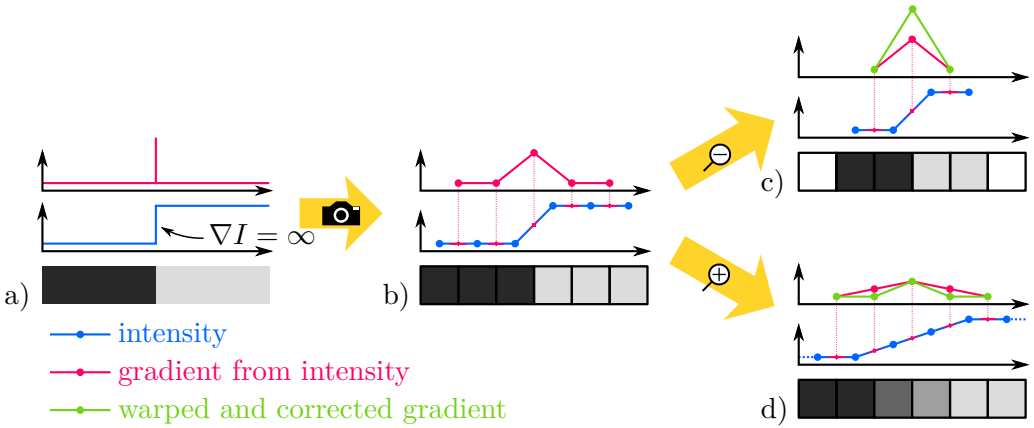
$$\pi^{-1}(\boldsymbol{u}) = Z(\boldsymbol{u}) \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix}$$

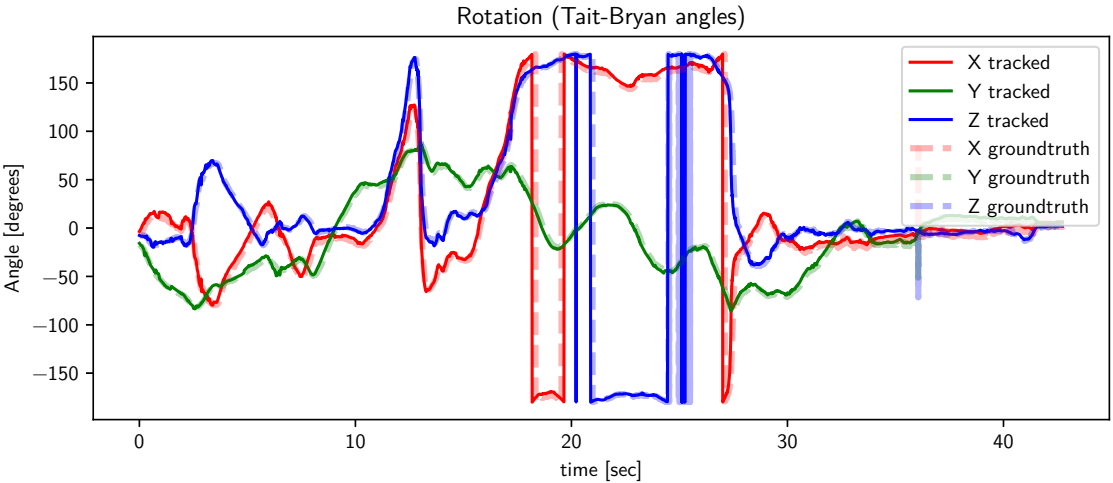
$$D_{\pi^{-1}} = \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix} \nabla Z(\boldsymbol{u})^{\top} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} Z(\boldsymbol{u})$$

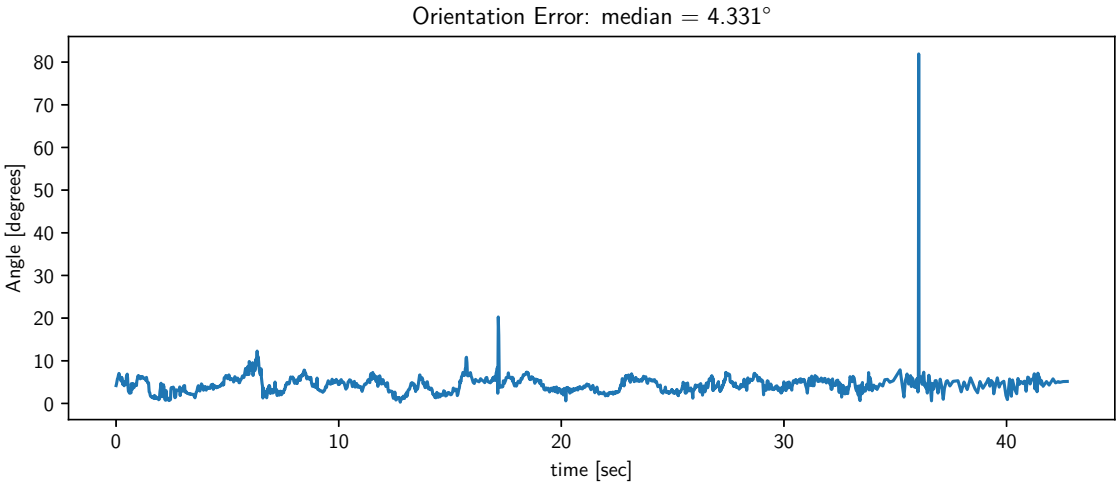

$\nabla I = \infty$
a)
intensity
gradient from intensity

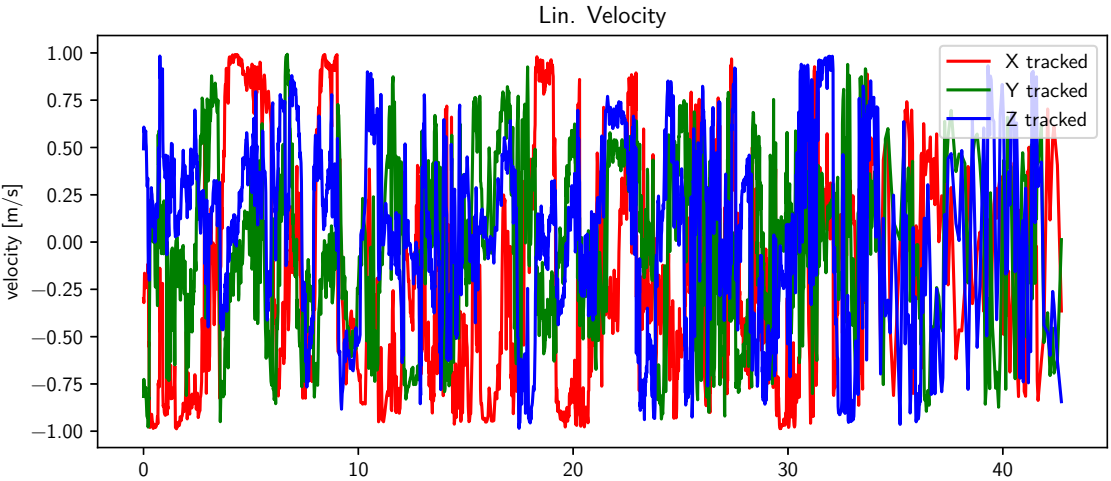

$\nabla I = \infty$
a)
b)
intensity
gradient from intensity

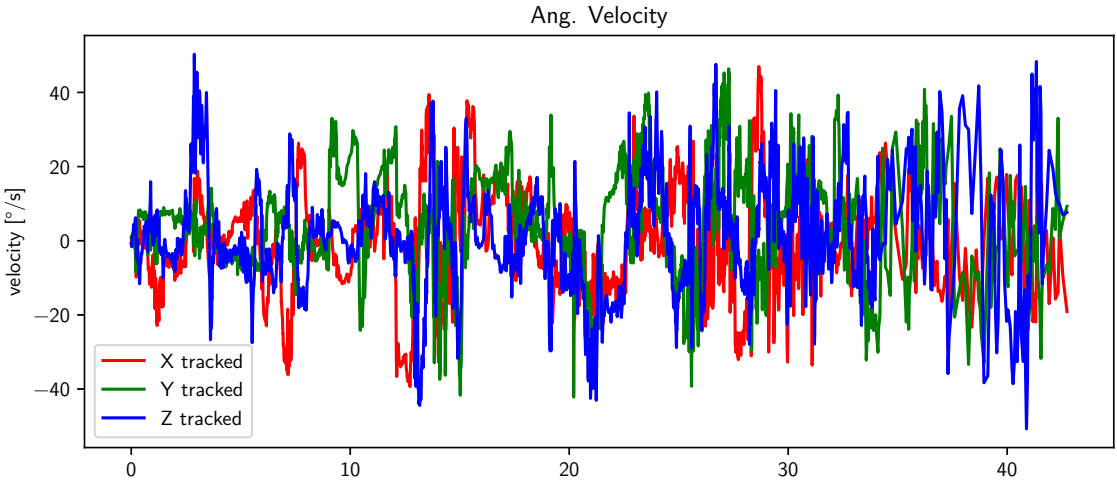

$\nabla I = \infty$
a)
b)
c)
intensity
gradient from intensity
warped and corrected gradient

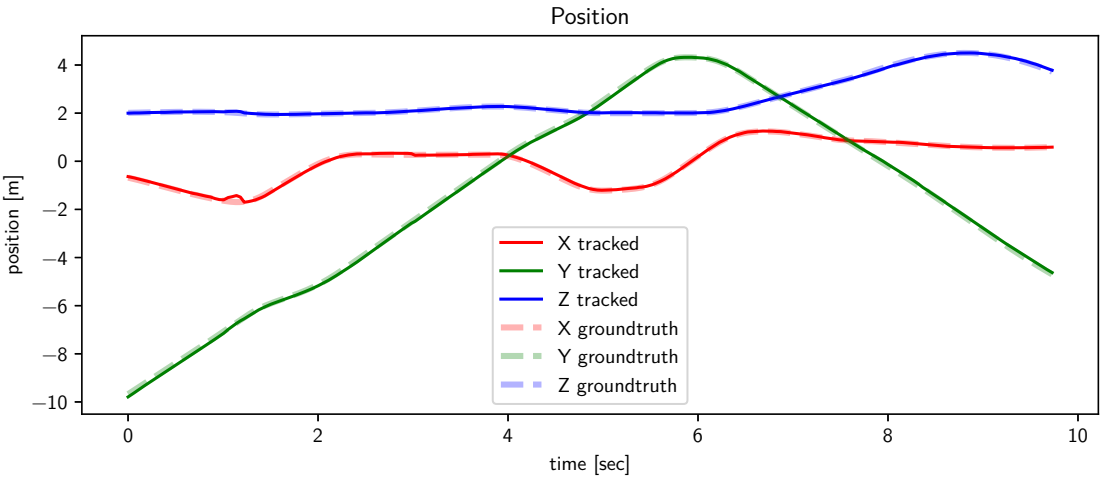

$\nabla I = \infty$
a)
b)
c)
d)
intensity
gradient from intensity
warped and corrected gradient

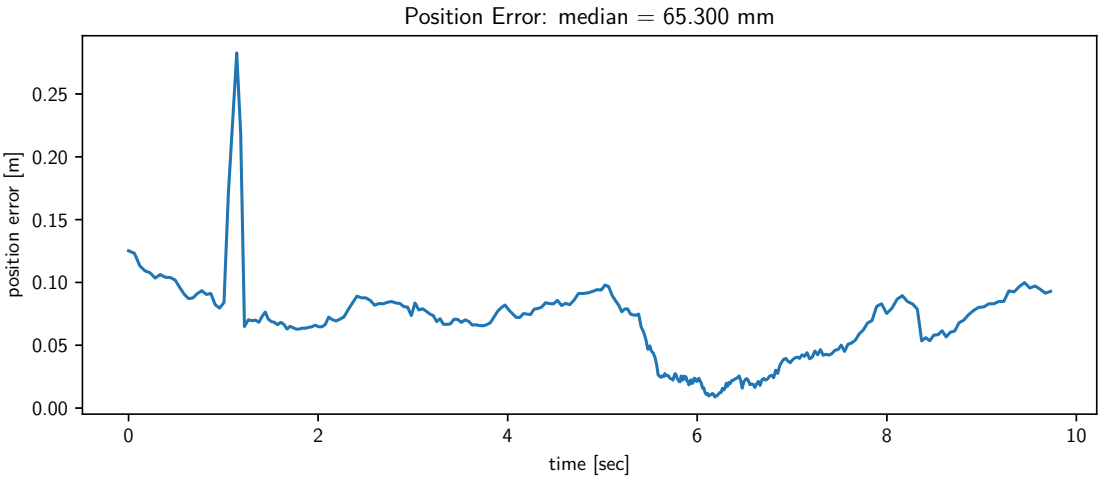

Rotation (Tait-Bryan angles)
X tracked
Y tracked
Z tracked
X groundtruth
Y groundtruth
Z groundtruth
Angle [degrees]
150
100
50
0
−50
−100
−150
0
10
20
30
40
time [sec]

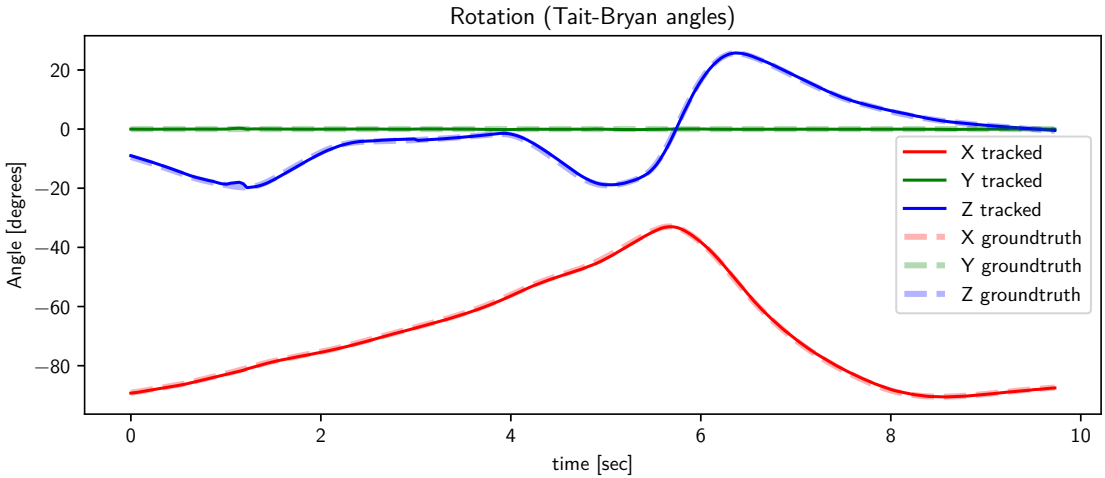

Orientation Error: median = 4.331°
Angle [degrees]
0
10
20
30
40
50
60
70
80
time [sec]
0
10
20
30
40

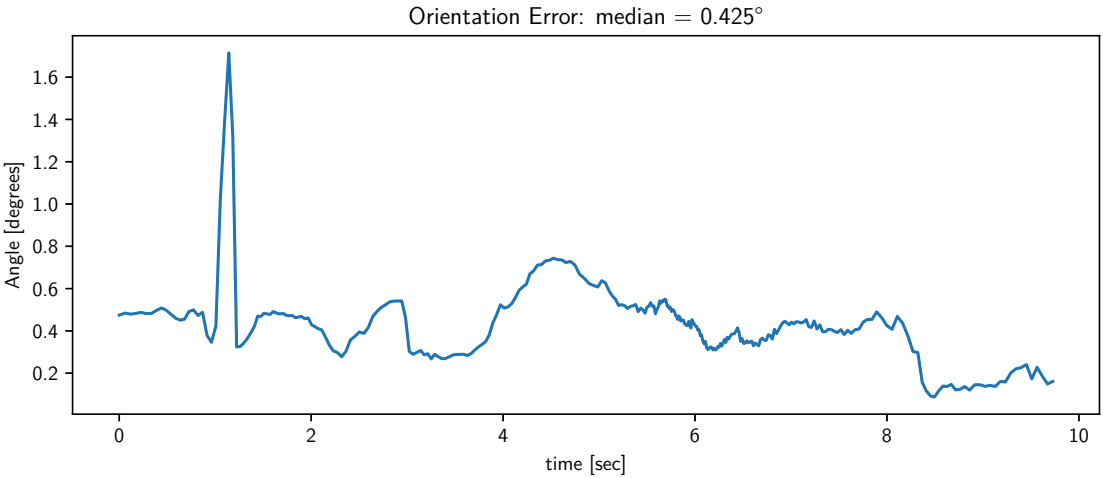

Lin. Velocity
X tracked
Y tracked
Z tracked
velocity [m/s]
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40

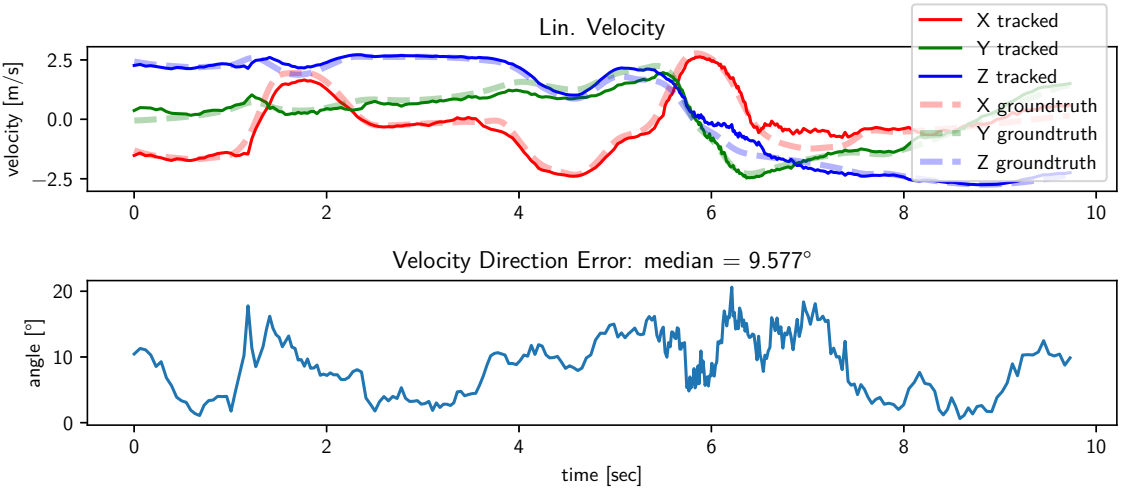

Ang. Velocity
velocity [°/s]
40
20
0
−20
−40
0
10
20
30
40
X tracked
Y tracked
Z tracked

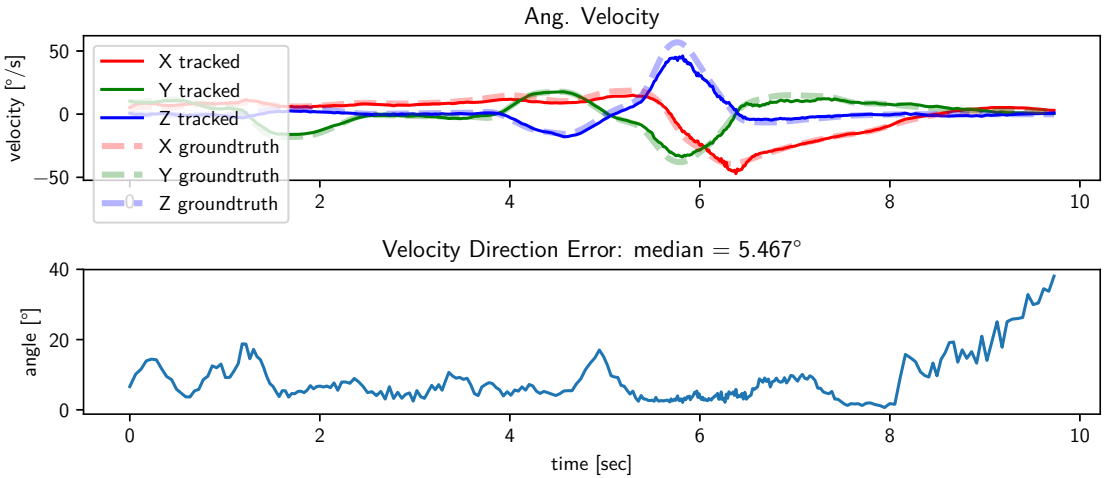

Position
position [m]
time [sec]
X tracked
Y tracked
Z tracked
X groundtruth
Y groundtruth
Z groundtruth
4
2
0
−2
−4
−6
−8
−10
0
2
4
6
8
10


Position Error: median = 65.300 mm
position error [m]
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
time [sec]


Rotation (Tait-Bryan angles)
Angle [degrees]
20
0
−20
−40
−60
−80
0
2
4
6
8
10
time [sec]
X tracked
Y tracked
Z tracked
X groundtruth
Y groundtruth
Z groundtruth


Orientation Error: median = 0.425°
Angle [degrees]
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
2
4
6
8
10
time [sec]


Lin. Velocity
X tracked
Y tracked
Z tracked
X groundtruth
Y groundtruth
Z groundtruth
velocity [m/s]
2.5
0.0
−2.5
0
2
4
6
8
10
Velocity Direction Error: median = 9.577°
angle [°]
20
10
0
time [sec]


Ang. Velocity
velocity [°/s]
50
0
−50
X tracked
Y tracked
Z tracked
X groundtruth
Y groundtruth
Z groundtruth
2
4
6
8
10
Velocity Direction Error: median = 5.467°
angle [°]
40
20
0
0
2
4
6
8
10
time [sec]

Scene	Mean Depth (mm)	Pose Pos. (mm)	Pose Rot. (°)	Velocity Direction Linear (°)	Velocity Direction Angular (°)
Carpet 2D	2109	7.264	0.158 78	23.561	17.760
Carpet 2D*	—”—	11.223	0.205 34	20.235	16.168
Boxes	1993	4.460	0.203 04	17.537	62.686
Boxes*	—”—	8.889	0.284 81	17.532	61.172
Sponza	10 600	65.300	0.425 14	9.5767	5.4670
Sponza*	—”—	69.040	0.520 06	11.171	6.0394
Flying Room 1	3200	99.480	3.0794	n/a	n/a
Flying Room 2A†	3415	96.410	3.9917	n/a	n/a
Flying Room 2A*†	—”—	92.691	3.6311	n/a	n/a
Flying Room 2B	—”—	88.232	3.8392	n/a	n/a

* downsampled, i.e. using lower resolution

† tracking failed after about 18 seconds

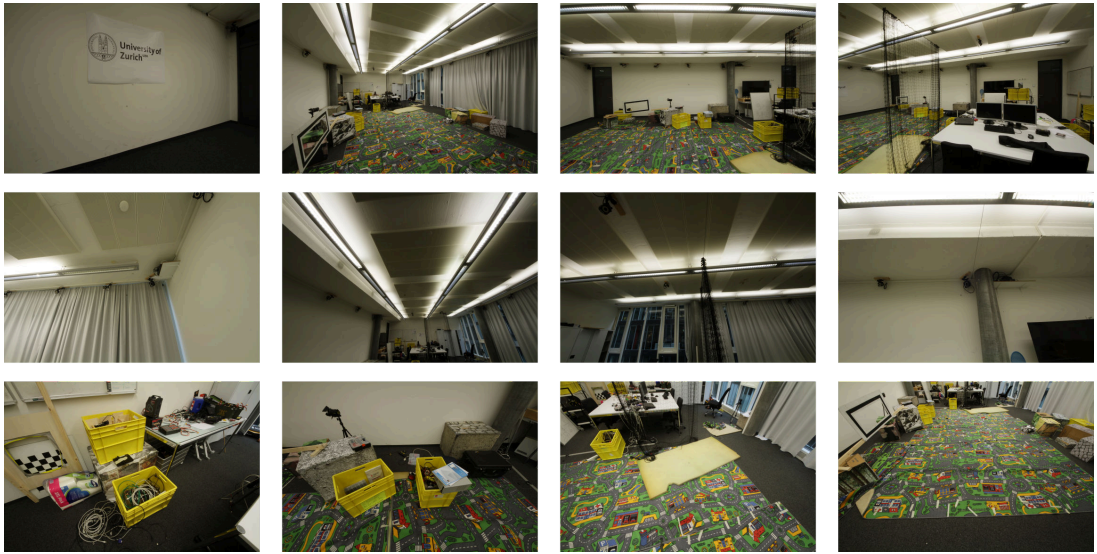
		Processing		Full Tracking Step		
Scene	Length	Time	Events/sec	Iter./step	sec/iter.	total
Carpet 2D	20s	8771.6s	2857.2	24.656	120.06ms	3024.7ms
Carpet 2D downsampled	20s	1957.0s	12789	17.005	38.516ms	675.743ms
Boxes	20s	10772s	3160.1	25.942	132.26ms	3418.5ms
Boxes downsampled	20s	5274.7s	6433.0	27.586	59.585ms	1679.3ms
Sponza	10s	637.97s	4130.5	12.234	166.07ms	2098.6ms
Sponza downsampled	10s	229.25s	11532	17.344	41.012ms	751.63ms
Flying Room, track 1	8.32s	10058s	994.99	64.634	359.54ms	22704ms
Flying Room, track 2B	43.24s	52094s	1016.2	90.307	315.08ms	22139.2ms

Systematic Errors

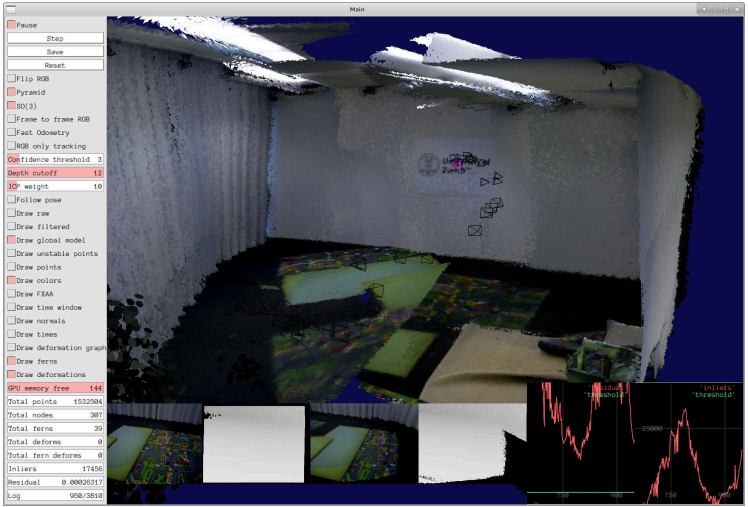
- Linearization
- Gradient Warping
- Occlusions
- Event Integration

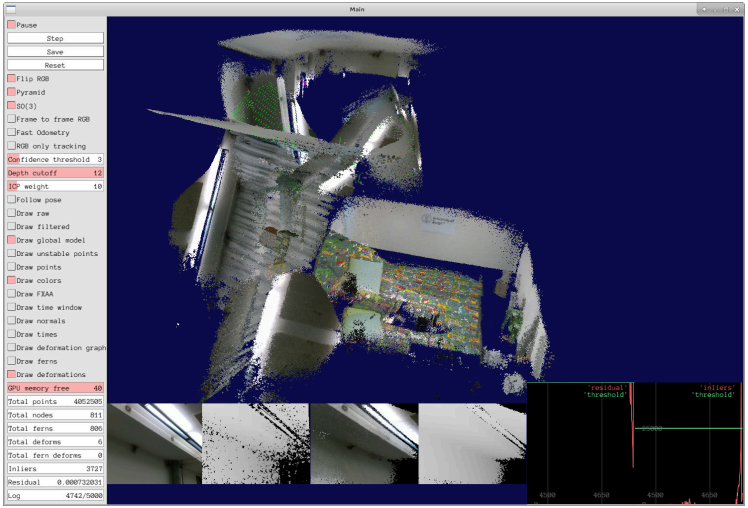
Real World Error Sources

- Hand-eye calibration of event camera
- Hand-eye calibration of RGB-D-camera for map
- Noise in map
- Timing noise in ground-truth recording


University of Zurich


University of
Zurich


Main
Pause
Step
Save
Reset
Flip RGB
Pyramid
SO(3)
Frame to frame RGB
Fast Odometry
RGB only tracking
Confidence threshold 3
Depth cutoff 12
ICP weight 10
Follow pose
Draw raw
Draw filtered
Draw global model
Draw unstable points
Draw points
Draw colors
Draw FXAA
Draw time window
Draw normals
Draw times
Draw deformation graph
Draw ferns
Draw deformations
GPU memory free 144
Total points 1532504
Total nodes 307
Total ferns 39
Total deforms 0
Total fern deforms 0
Inliers 17456
Residual 0.00026317
Log 950/3810
'Residual' 'threshold'
'Inliers' 'threshold'
25000
750
700
750


Main
Pause
Step
Save
Reset
Flip RGB
Pyramid
SO(3)
Frame to frame RGB
Fast Odometry
RGB only tracking
Confidence threshold 3
Depth cutoff 12
ICP weight 10
Follow pose
Draw raw
Draw filtered
Draw global model
Draw unstable points
Draw points
Draw colors
Draw FXAA
Draw time window
Draw normals
Draw times
Draw deformation graph
Draw ferns
Draw deformations
GPU memory free 40
Total points 4052505
Total nodes 811
Total ferns 806
Total deforms 6
Total fern deforms 0
Inliers 3727
Residual 0.000732031
Log 4742/5000
'residual' 'threshold'
'inliers' 'threshold'
45000
4500
4550
4600
4650

MAPPING ELASTIC FUSION + OPTITRACK