lamar.ethz.ch

# LaMAR Tutorial
# 2. The Dataset

Paul-Edouard Sarlin

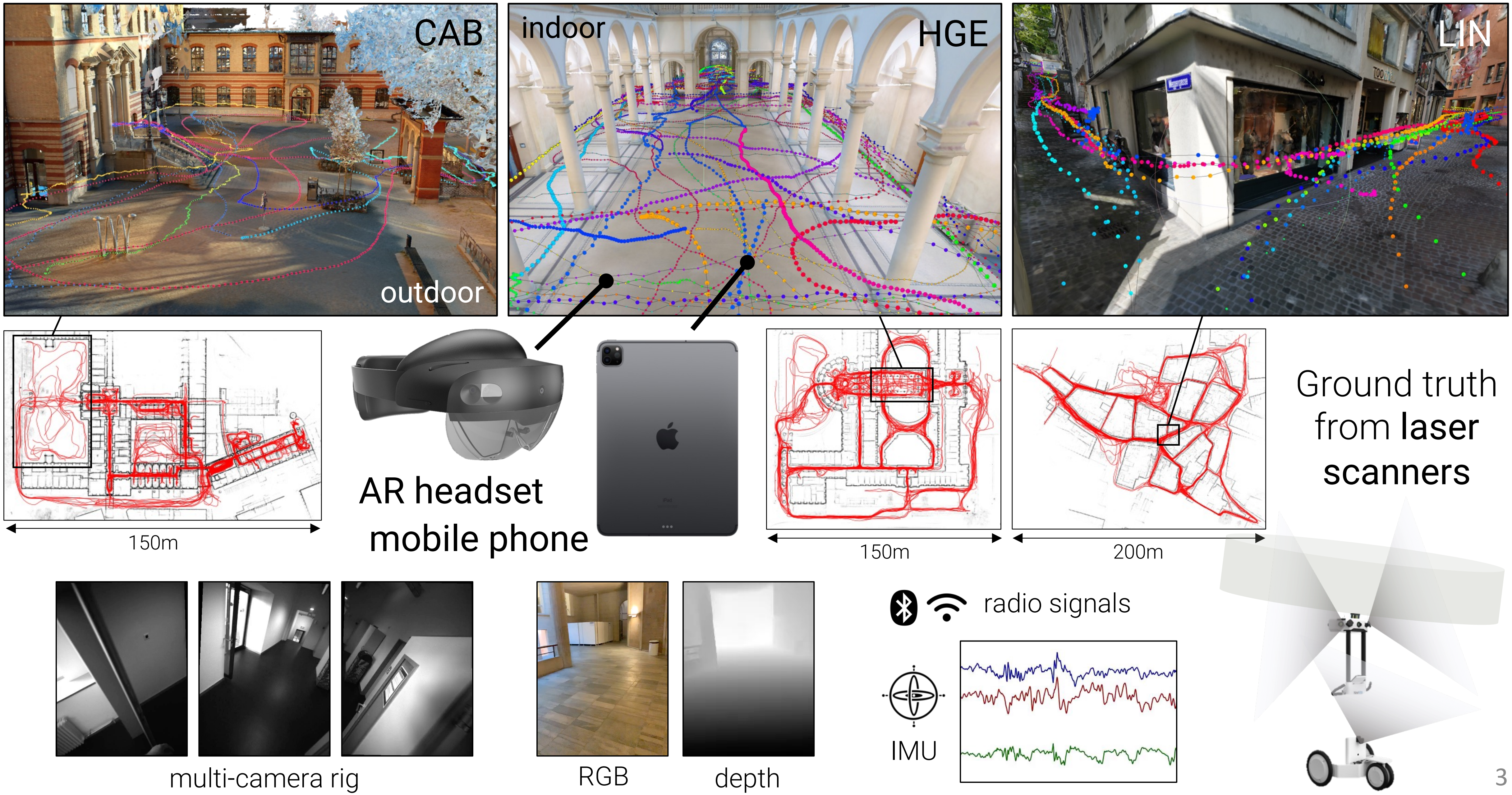
ETH zürich

ECCV TEL AVIV 2022

Microsoft

# GOAL

capturing realistic but challenging AR data and conditions


CAB
outdoor
indoor
HGE
LIN
150m
AR headset
mobile phone
150m
200m
Ground truth
from **laser**
**scanners**
multi-camera rig
RGB
depth
radio signals
IMU

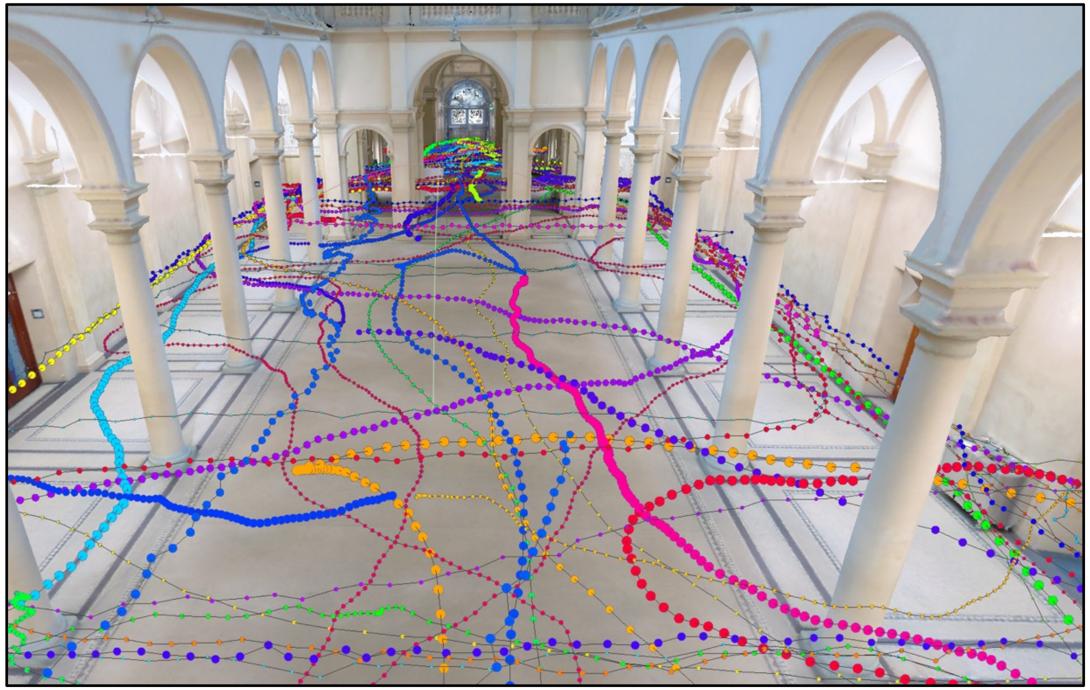
# Outline

a) Raw data
b) Processed data
c) Use cases
d) Outlook

# a) Raw data

# Capture process - crowdsourcing

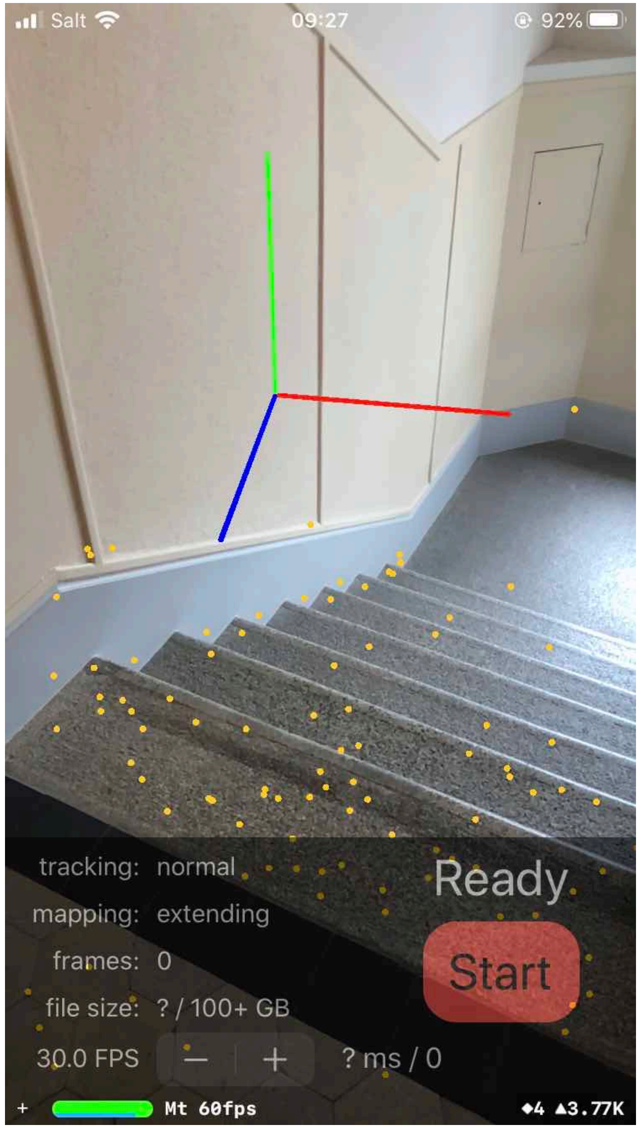
- Give AR devices to ~20 non-expert users
- Asked to explore the capture area given a map of it
  - Mimics navigation, exploration, inspection
  - No AR interactions
  - No specific instruction

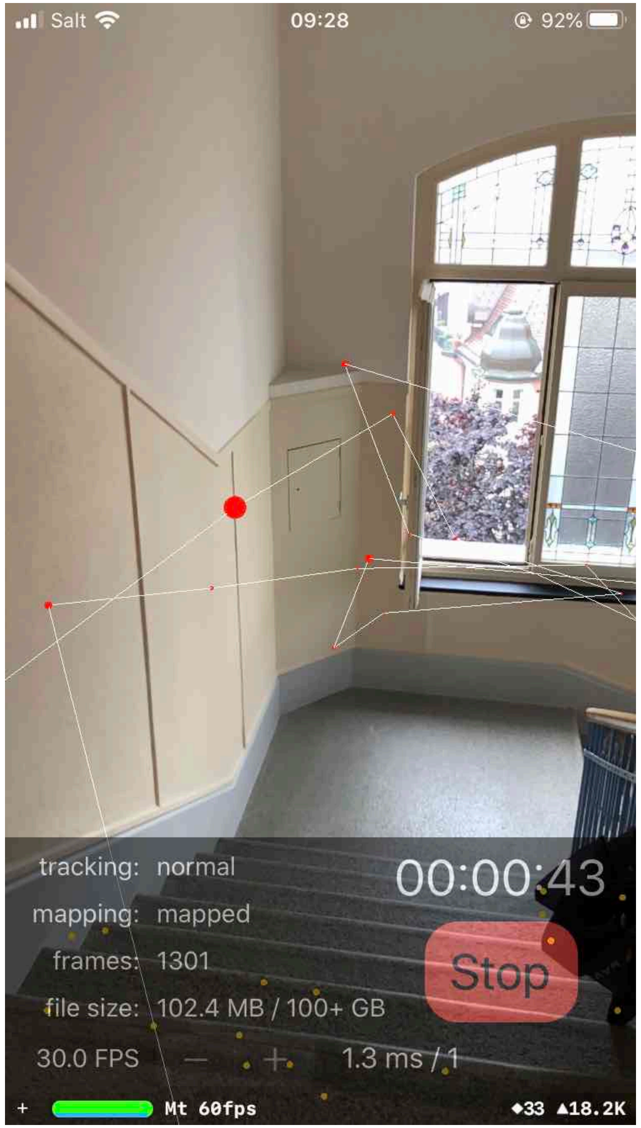


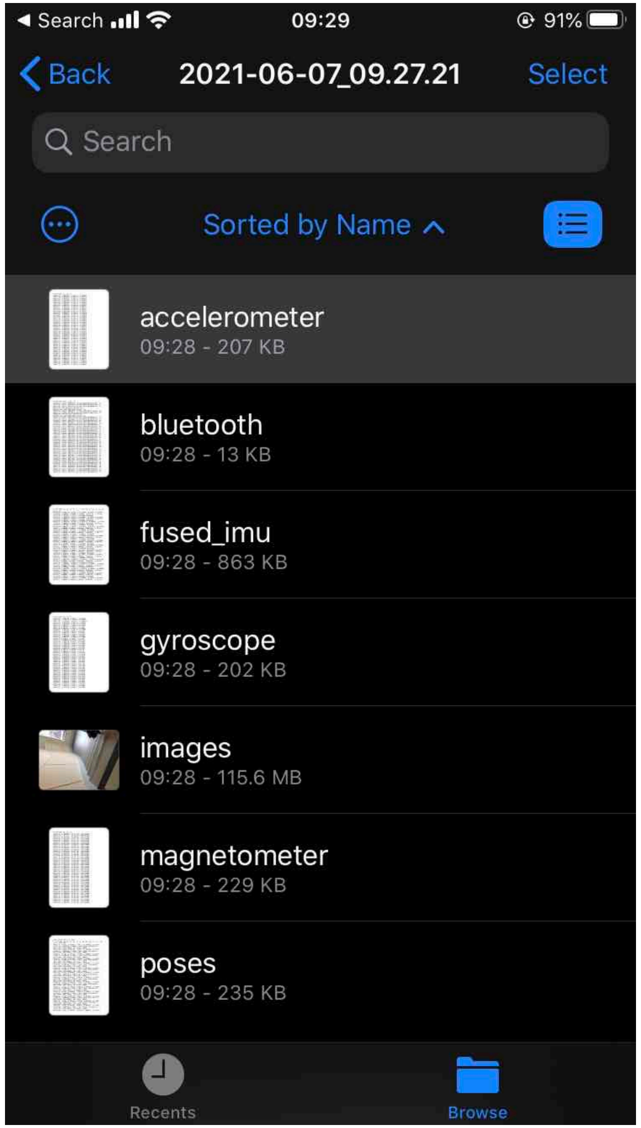
# AR devices

| | device | **HoloLens 2** | **iPhone 8 & iPad Pro** |
|---|---|---|---|
| | motion | head-mounted | hand-held |
| camera | # | 4x @ 30 Hz | 1x @ 10 Hz |
| | FOV | 83° | 64° |
| | resolution | VGA | 1080p |
| | specs | gray, GS | RGB, RS, AF |
| | depth | ToF+IR @ 1Hz | lidar @ 10 Hz |
| | other | IMU, gravity, , | IMU, gravity, GNSS, |
| | poses+calibration | head-tracking | ARKit |
| | | <u>async cameras with partial trigger</u> | <u>time-varying intrinsics</u> |

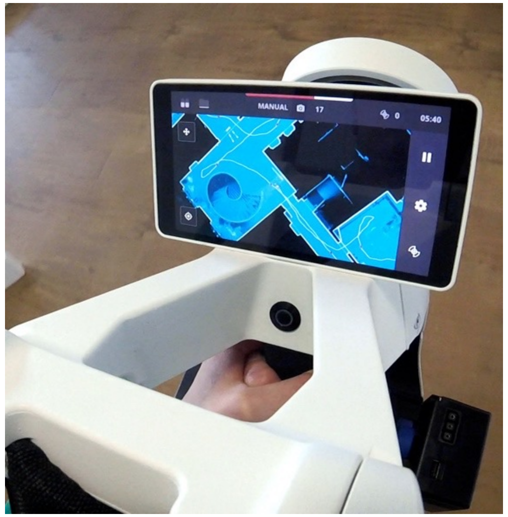
# Recording app



# Lidar reference



- Scan the space with NavVis devices
  - Commercial scanning rig: cameras, lidar, screen
  - Licensed processing software:
    SLAM + point cloud merge
- Large scenes are captured in multiple sessions of 1.5h

**<u>NavVis M6</u>**
trolley,
6 cameras,
1x long-range laser

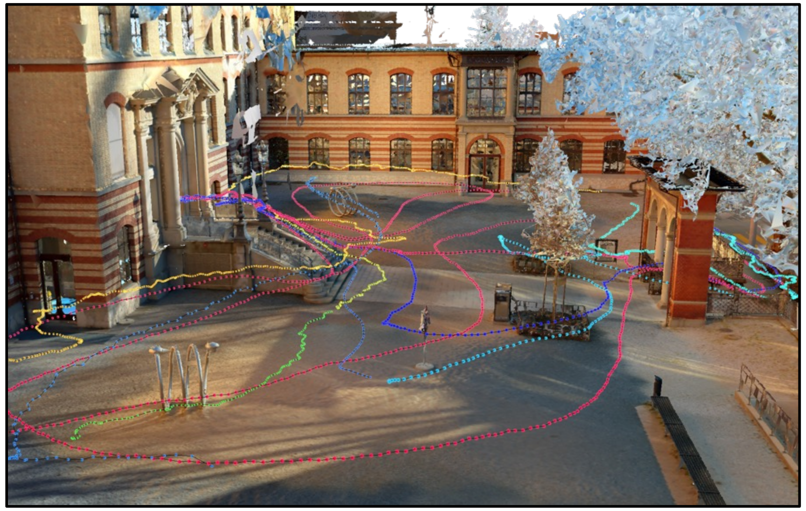
**<u>NavVis VLX</u>**
backpack,
5 cameras,
2x long-range lasers

# Lidar reference - outputs

- Registered & calibrated images
  - manually triggered every 2-3m
- Point cloud
  - 2cm res, 100m range, no dynamic
  - normals, colors, sensor positions

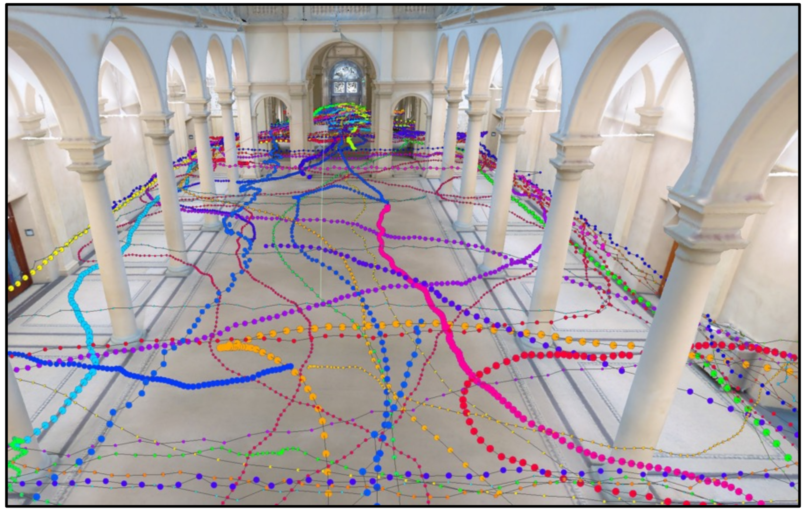
# Large spatial extent

## 3 locations that are difficult to navigate



**CAB**

office building at ETH, indoor + outdoor, multi-floor



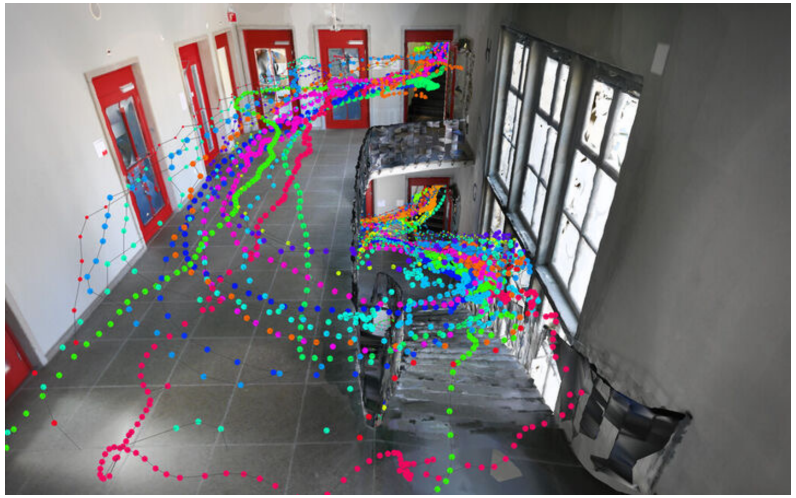
**HGE**

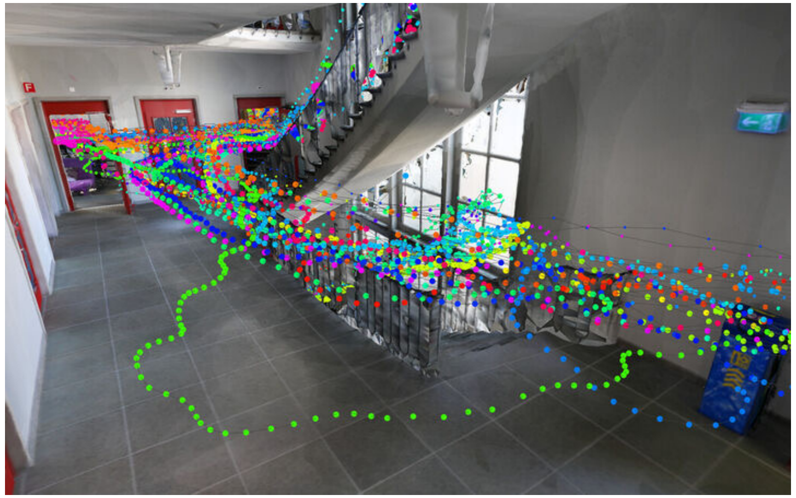
ground floor of the main ETH building, indoor + outdoor

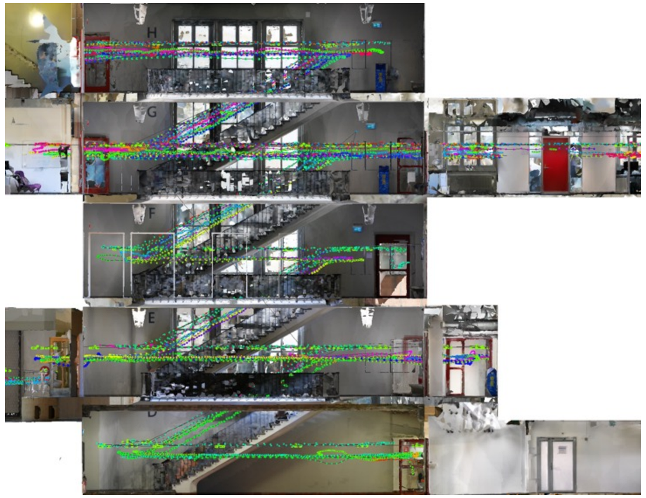
**LIN**

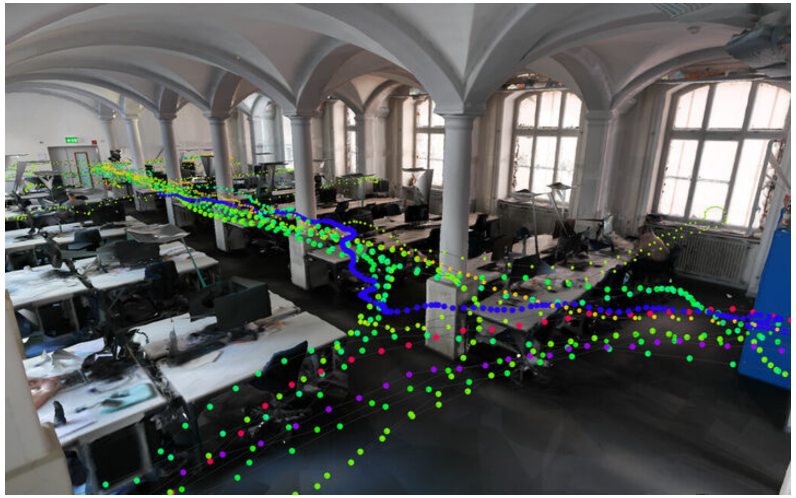
old city of Zurich, outdoor-only

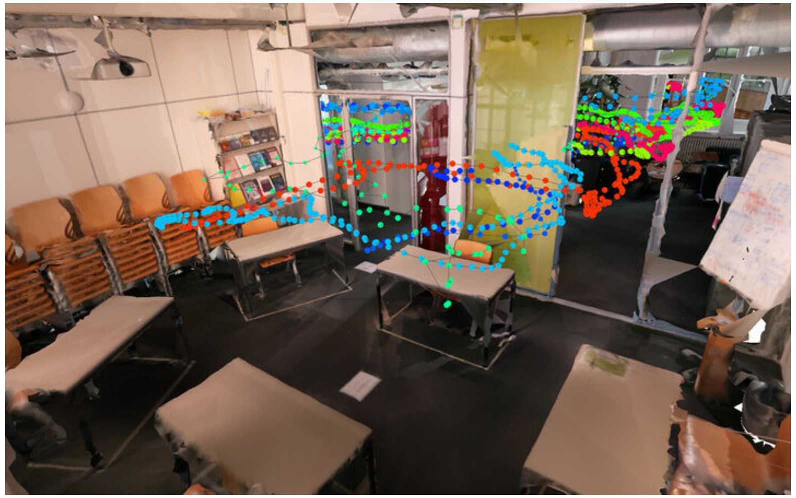
# Locations - CAB

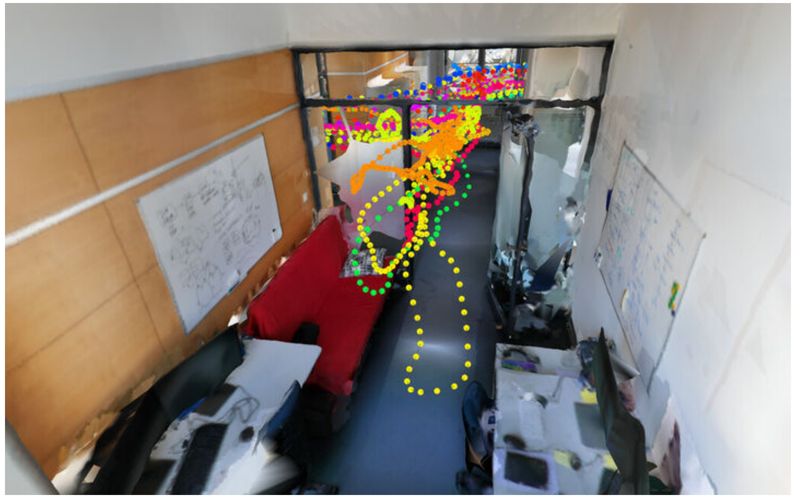




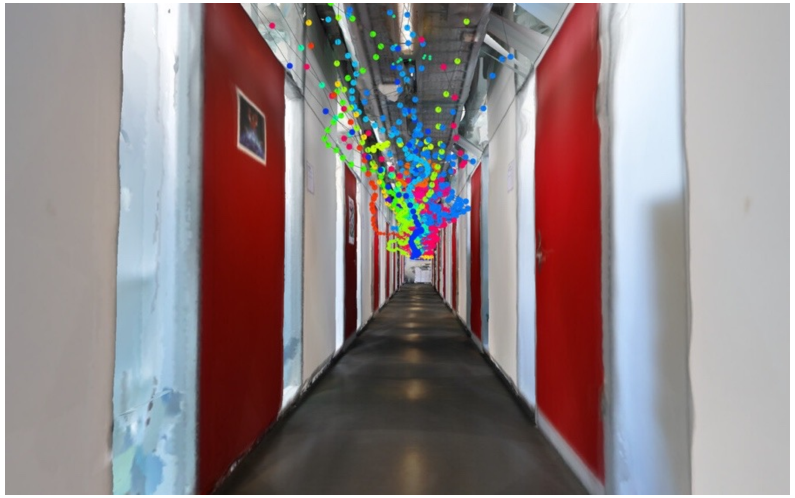


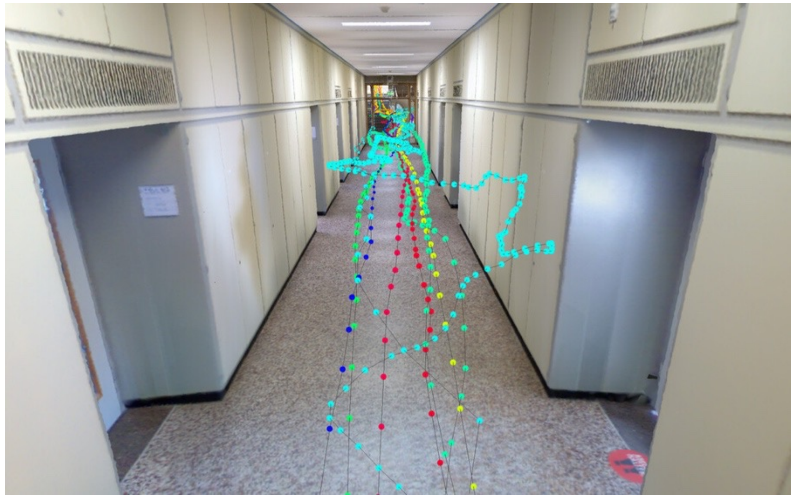


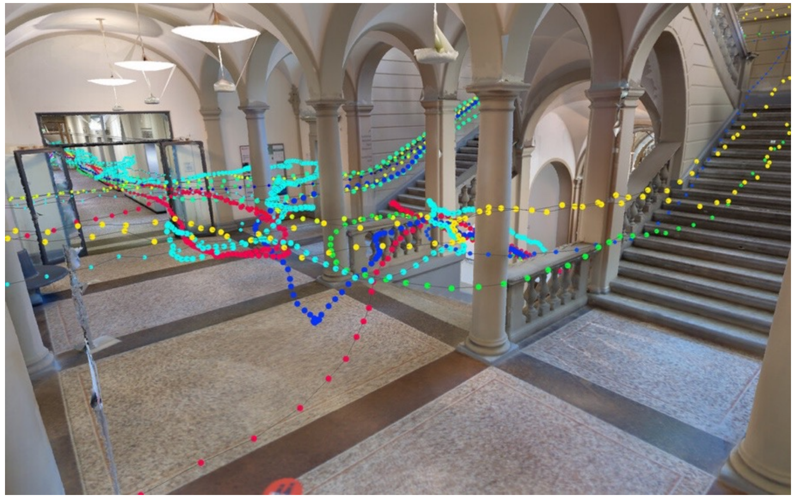


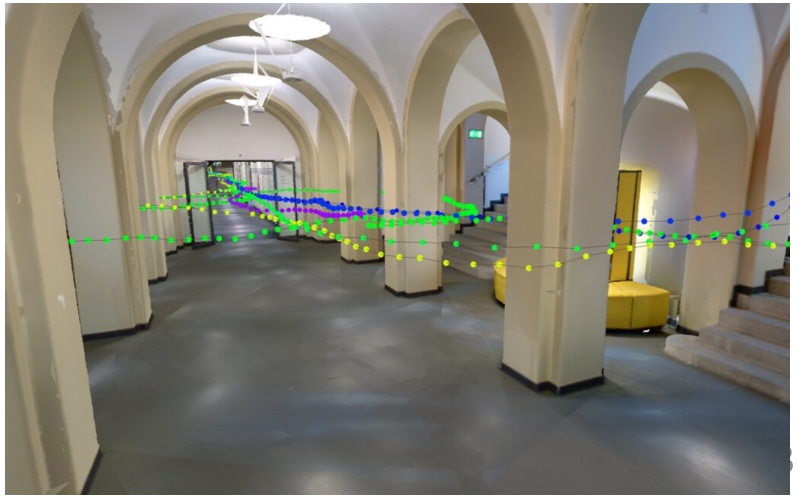


# Locations - CAB

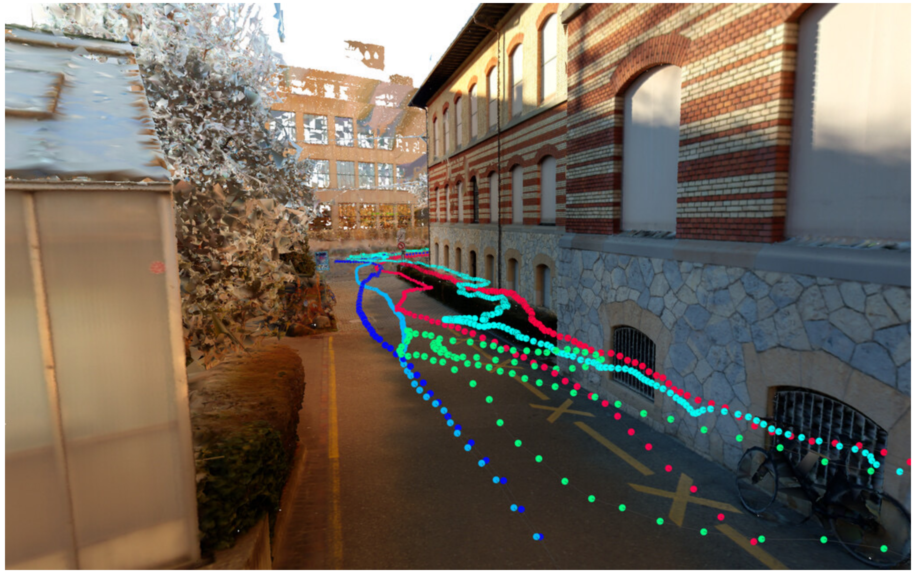


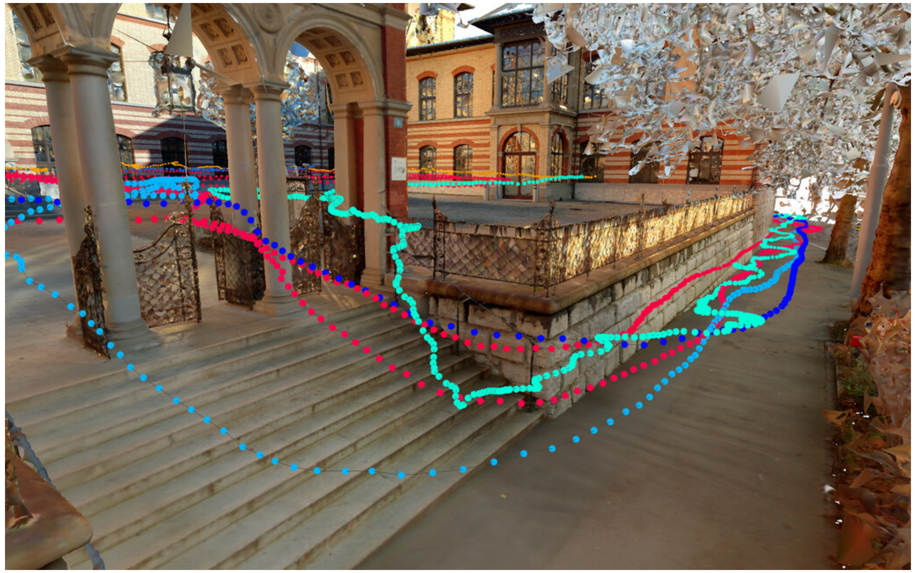


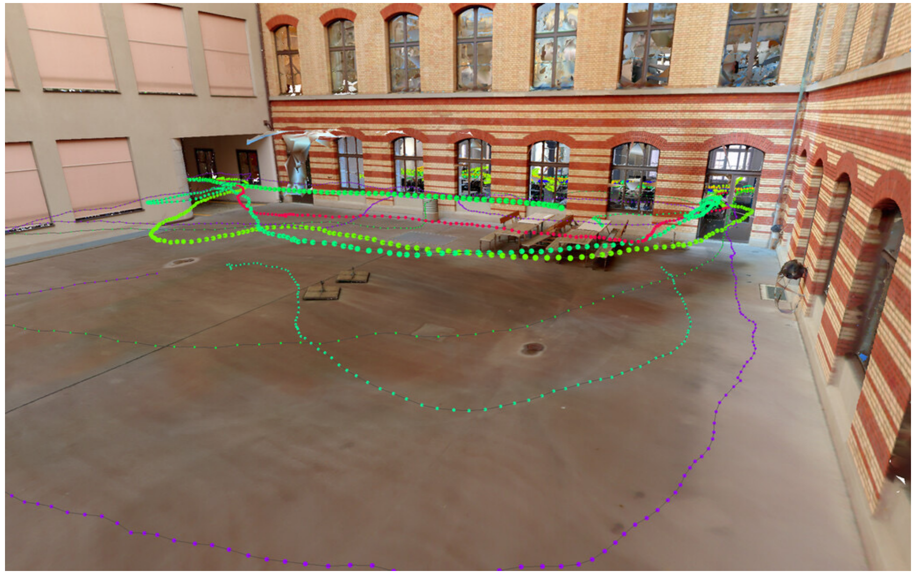


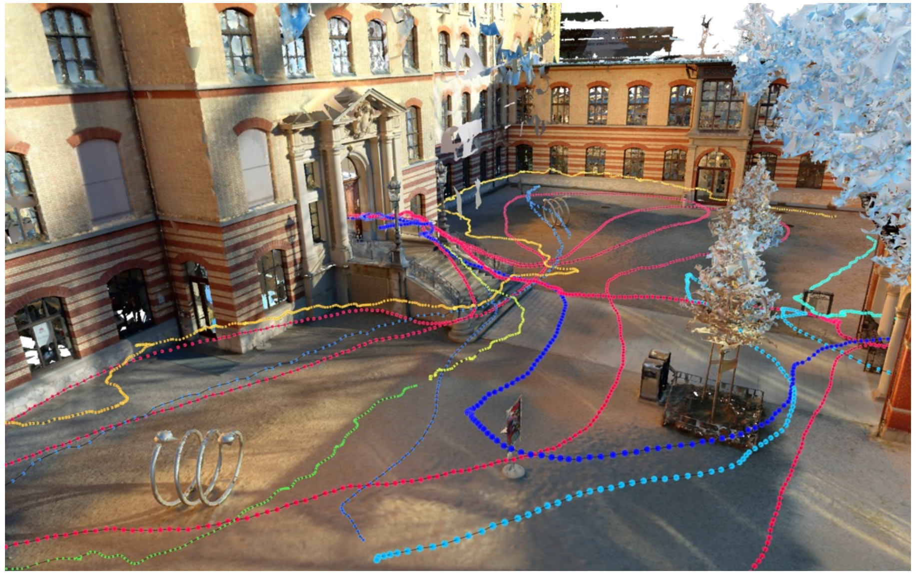


# Locations - CAB

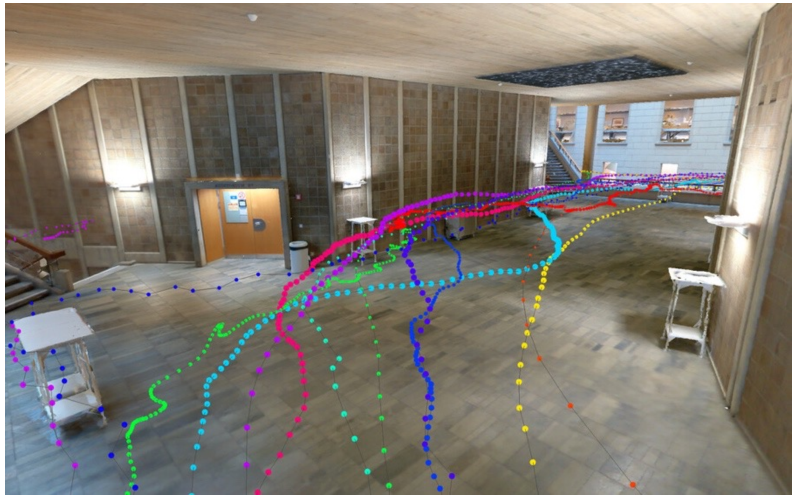


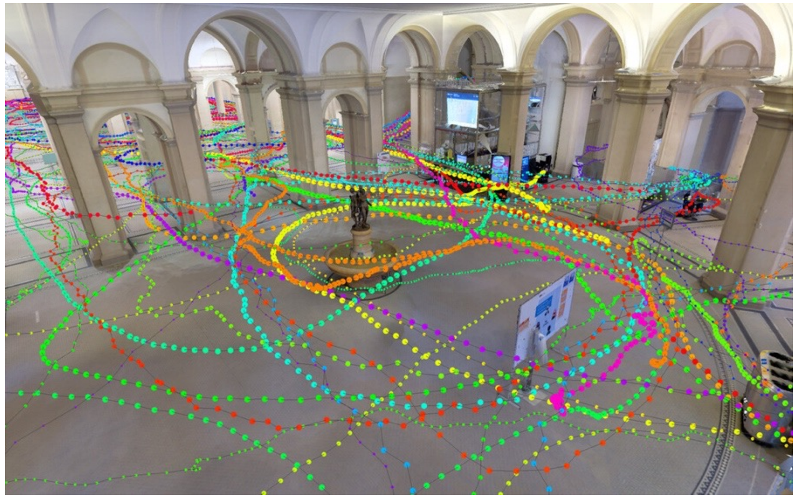


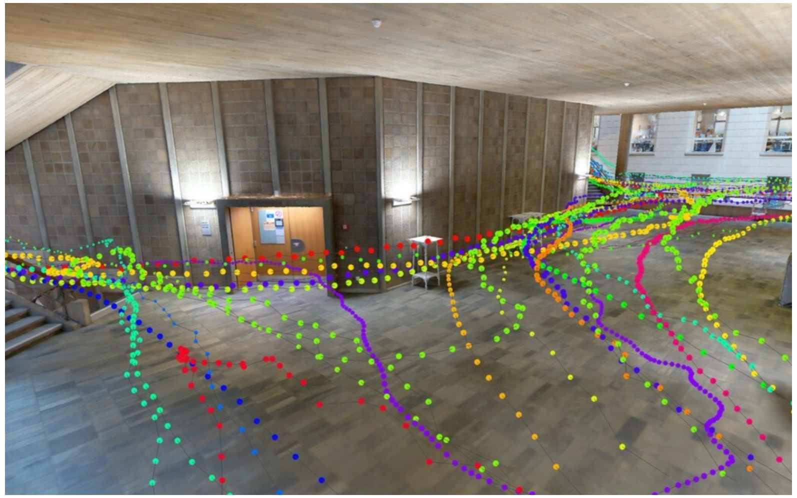


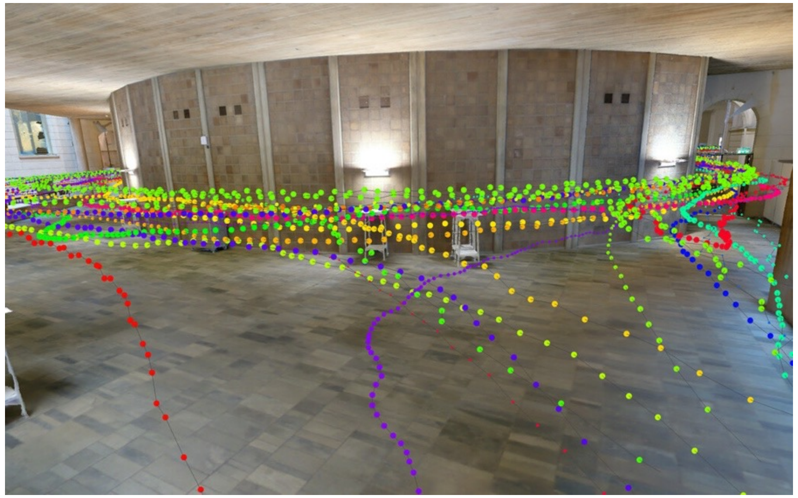


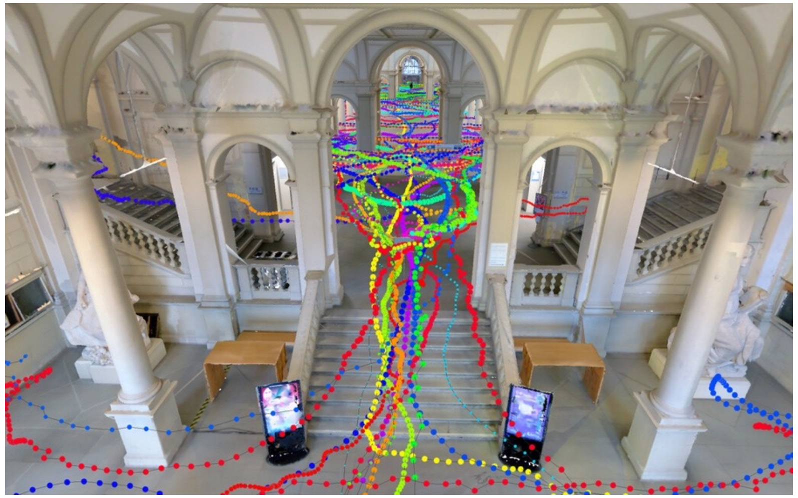
# Locations - HGE



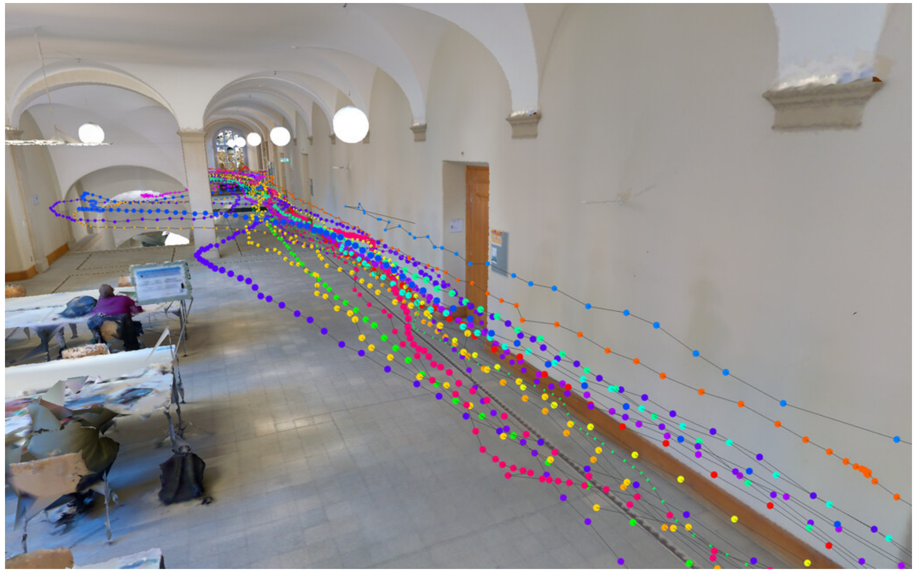


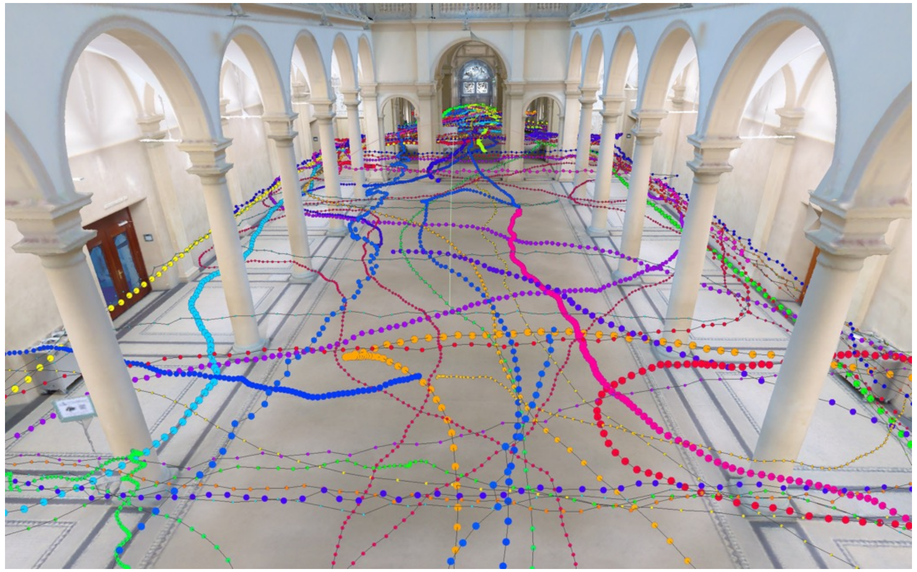


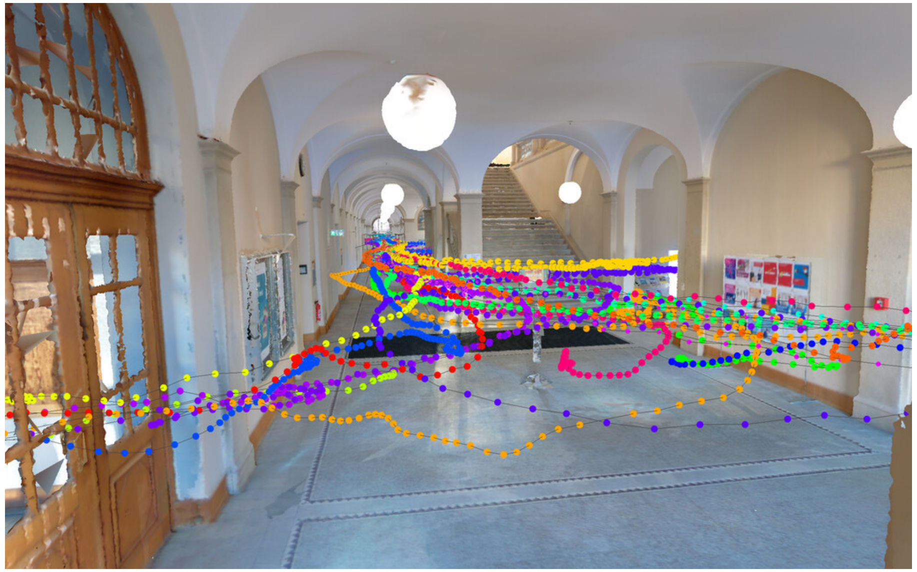


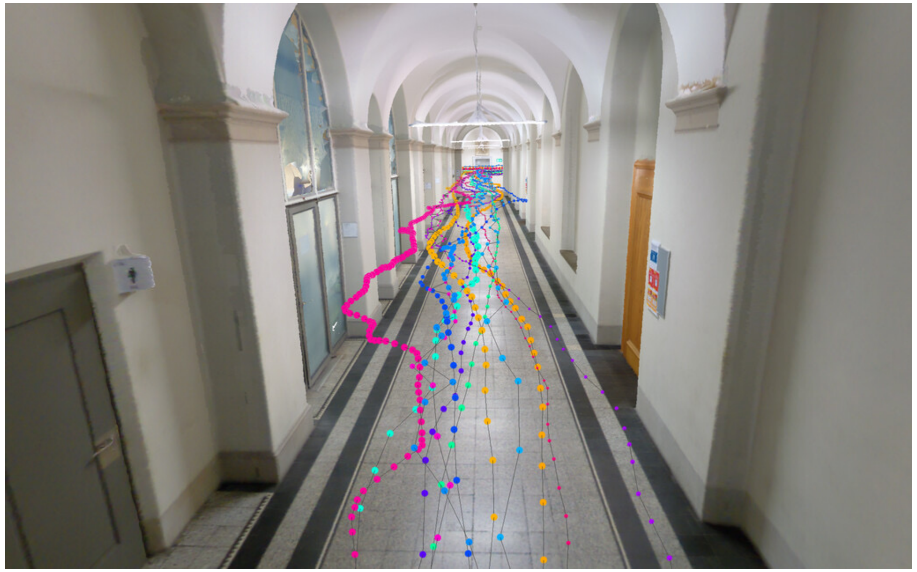


# Locations - HGE

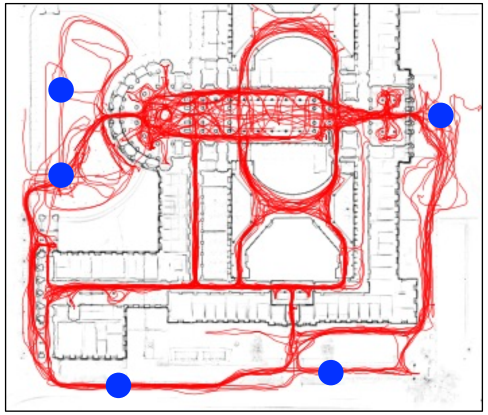


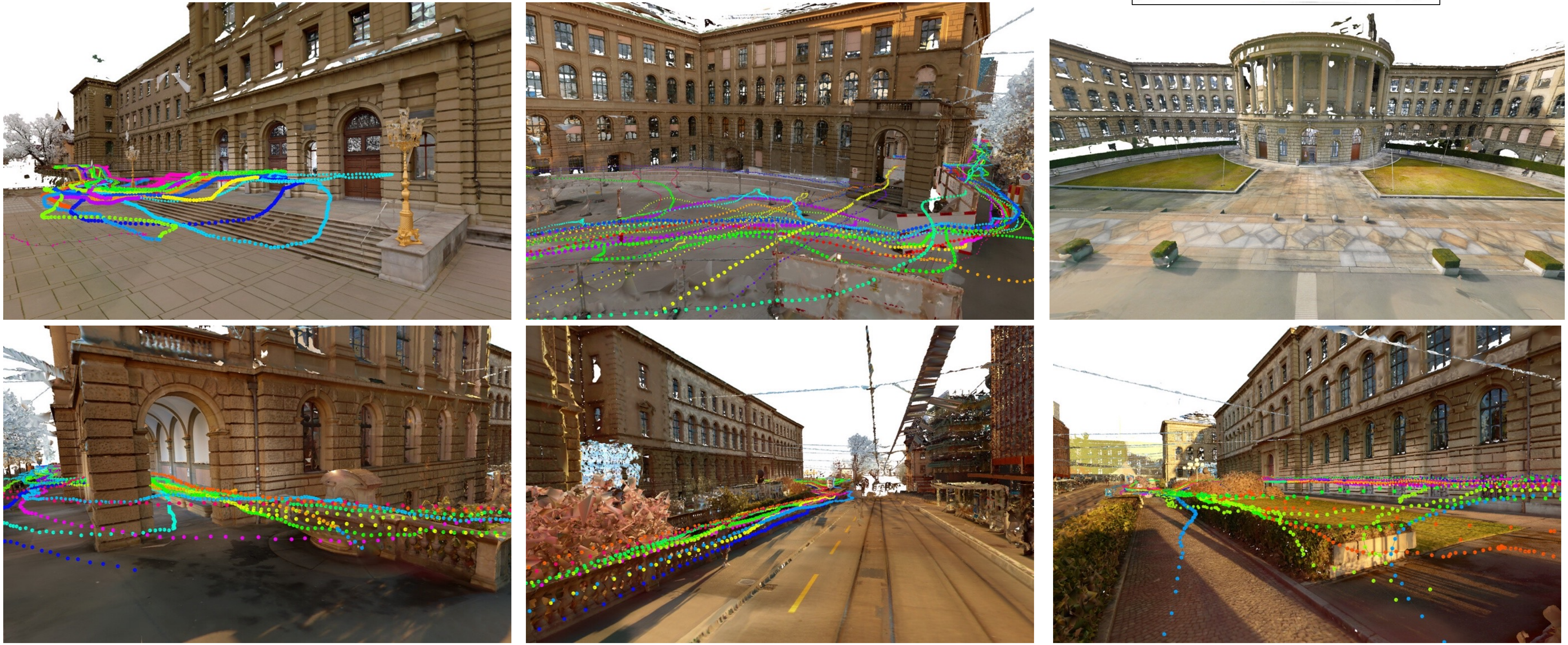




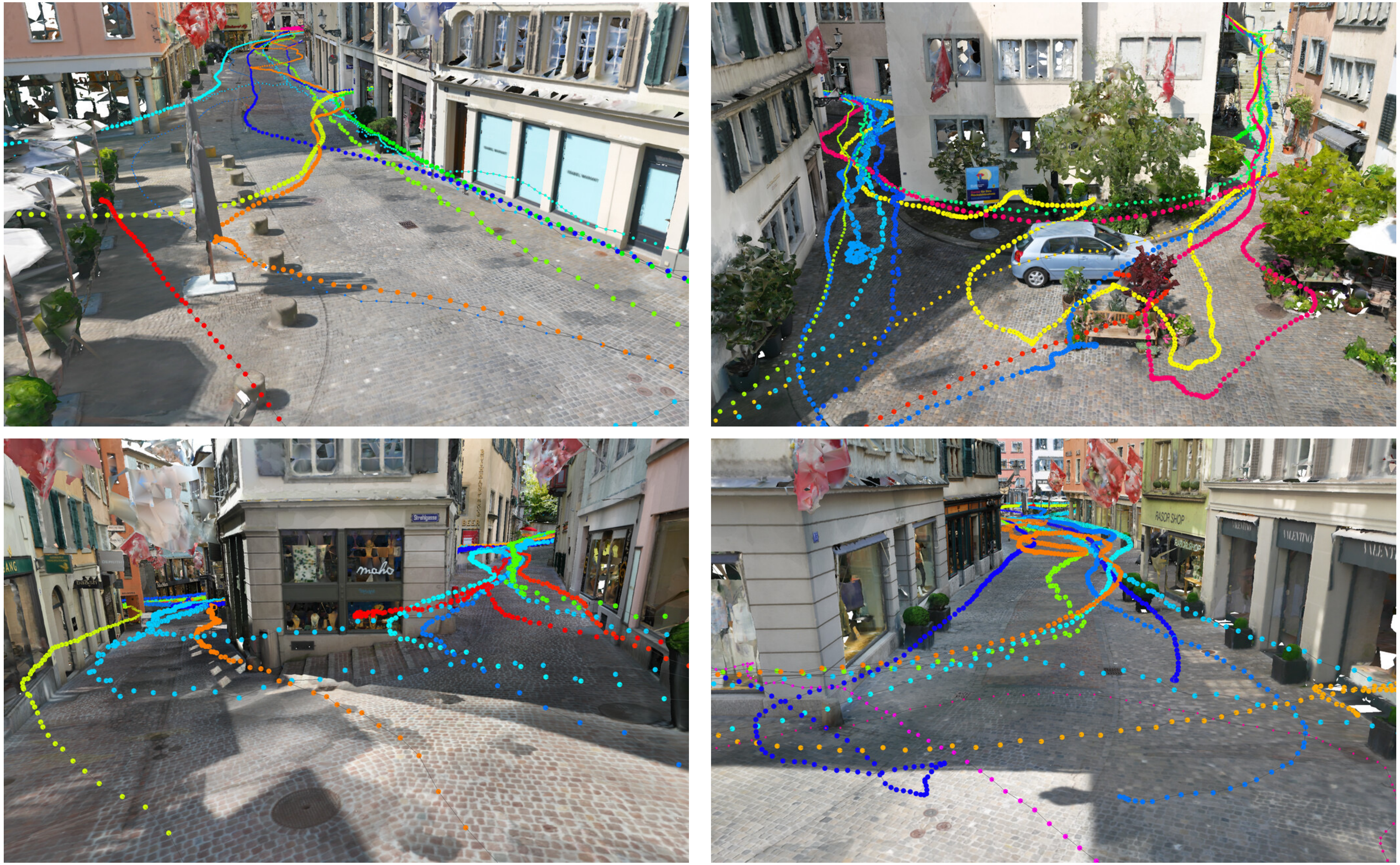


# Locations - HGE

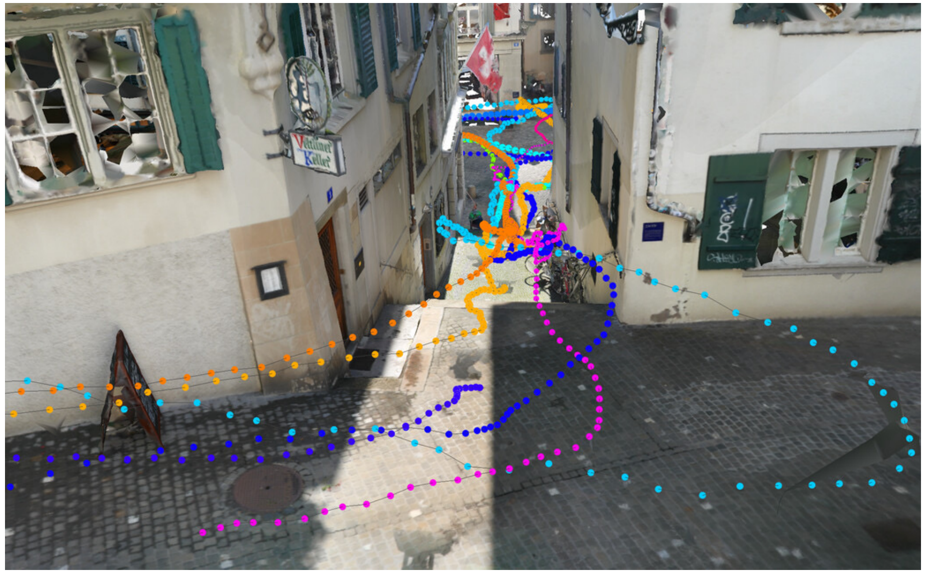


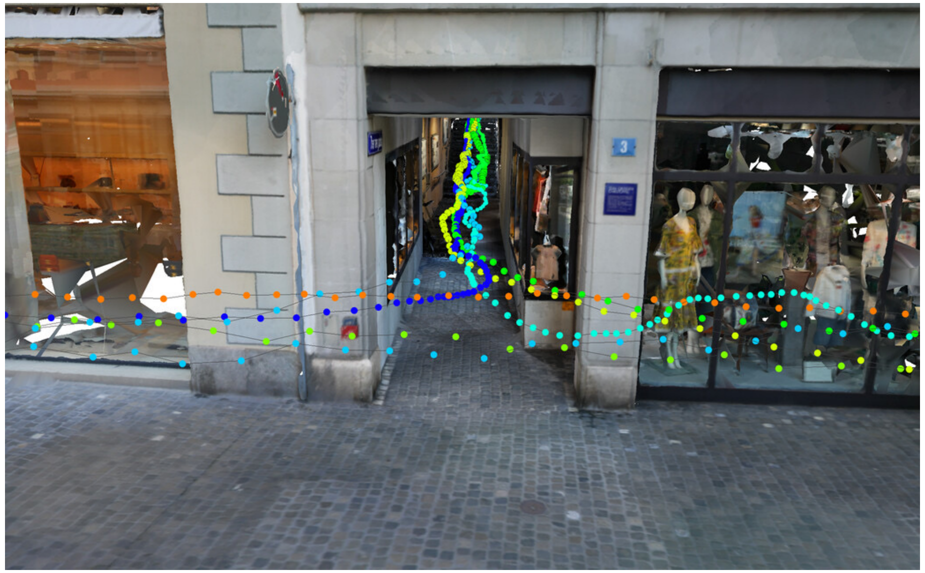


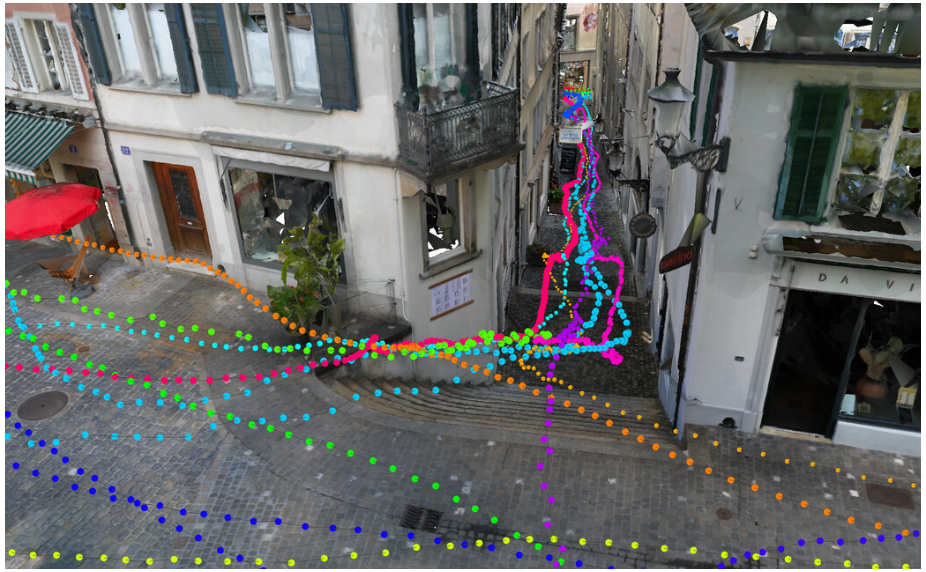
# Locations - LIN



# Locations - LIN

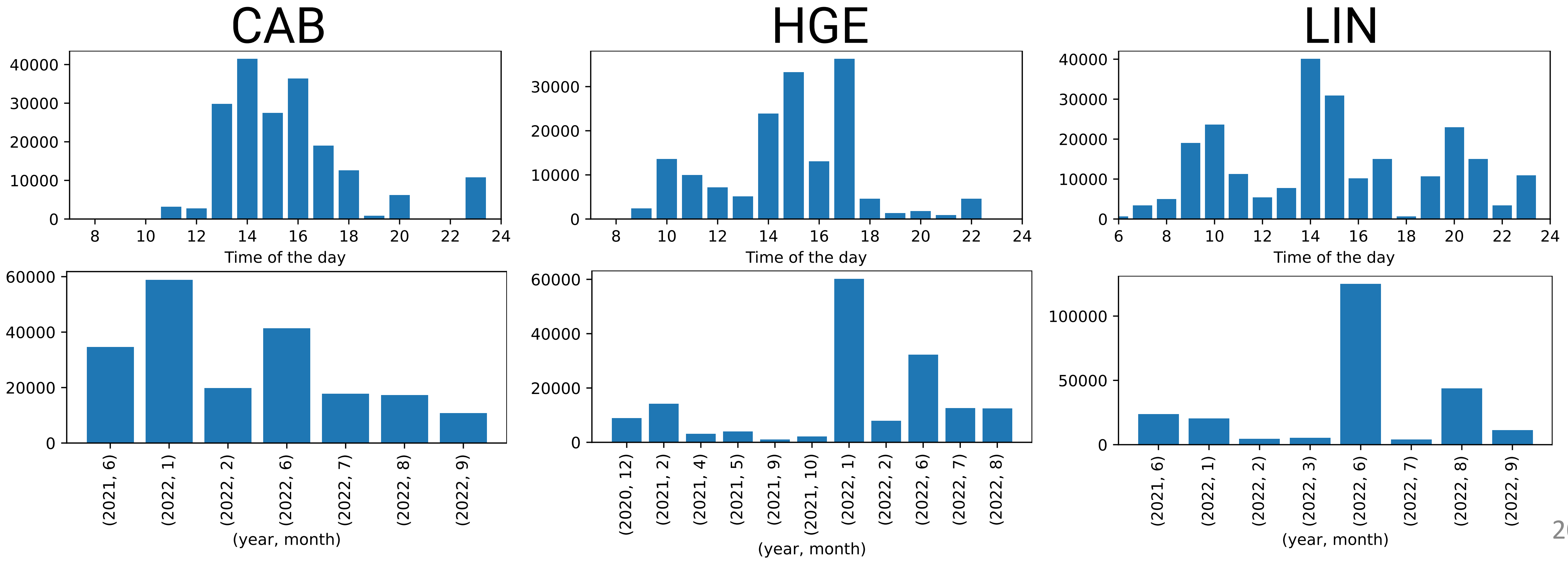




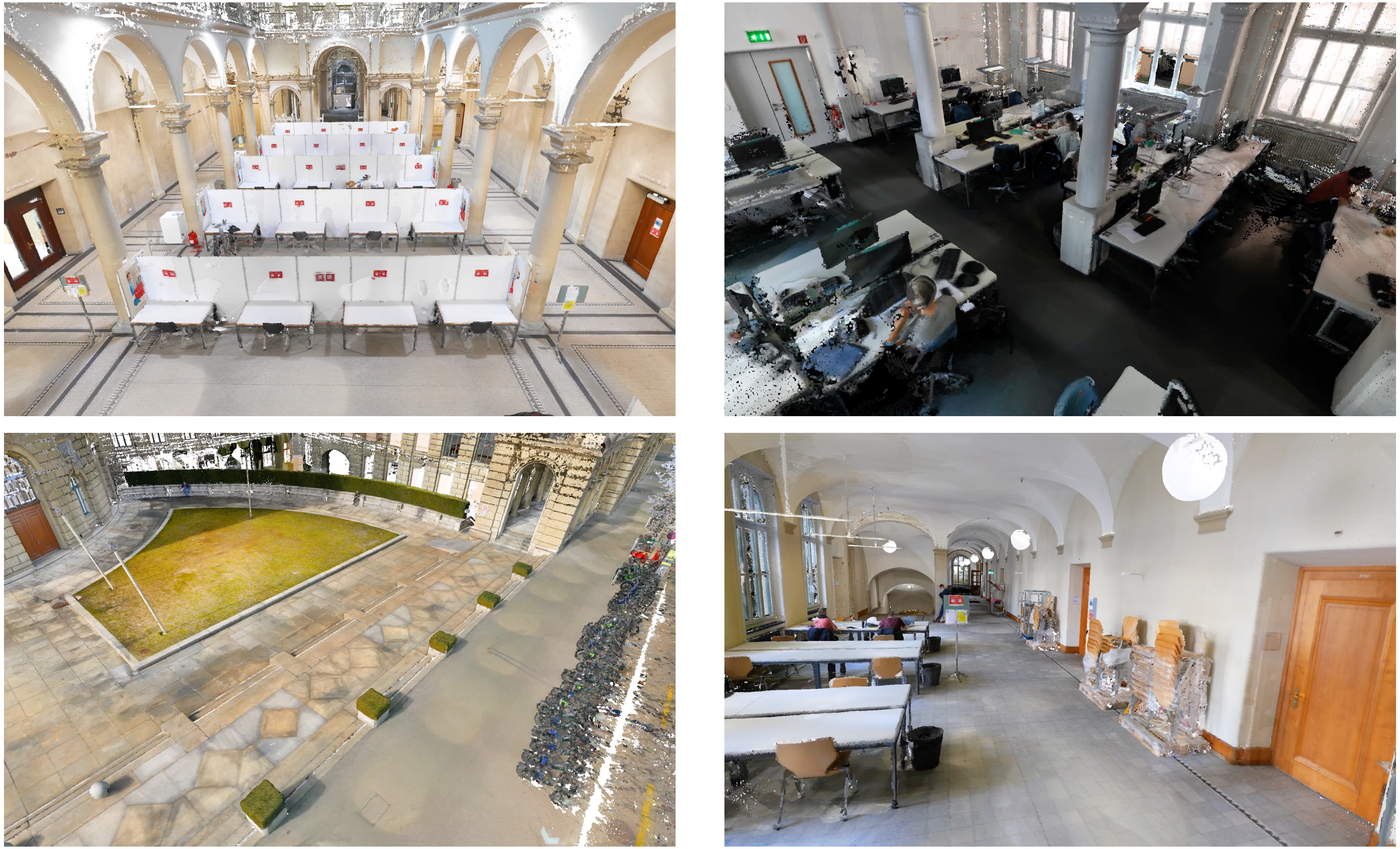


# Large temporal extent

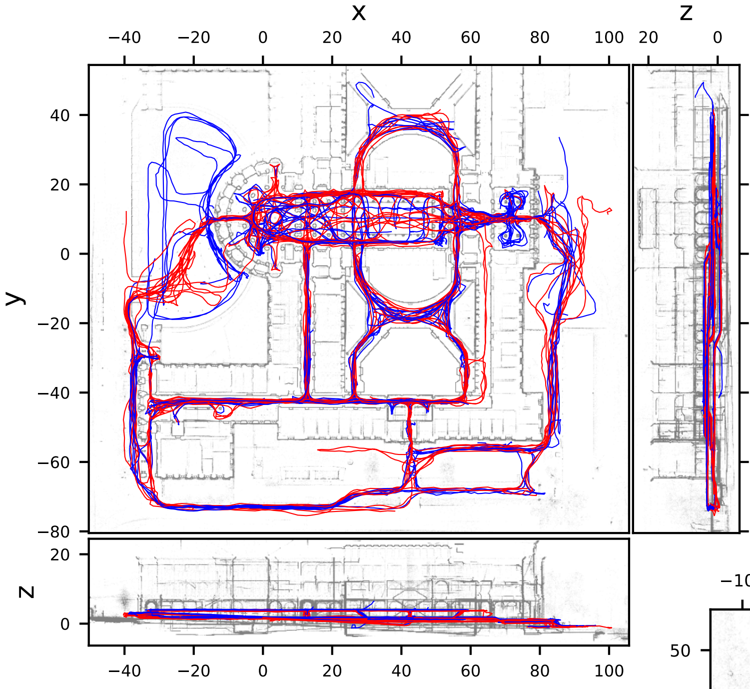
- Data captured over 1.5-2.5 years (varies by location)
- Mobile sequences captured by burst every few months
- Reference scans every year



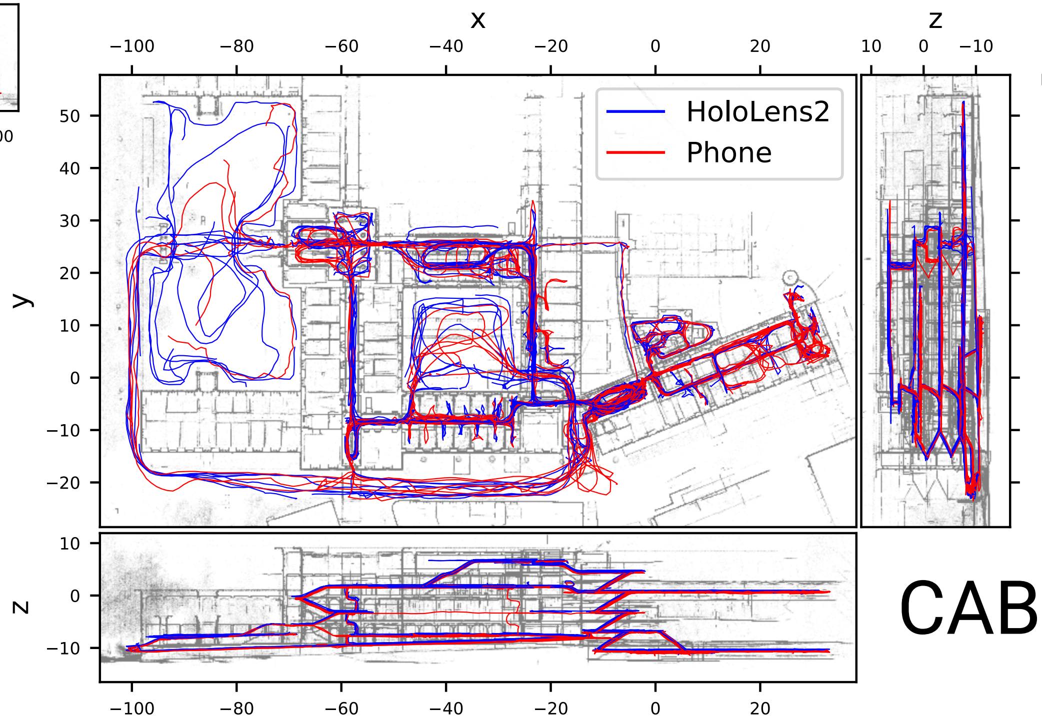
# Large temporal extent



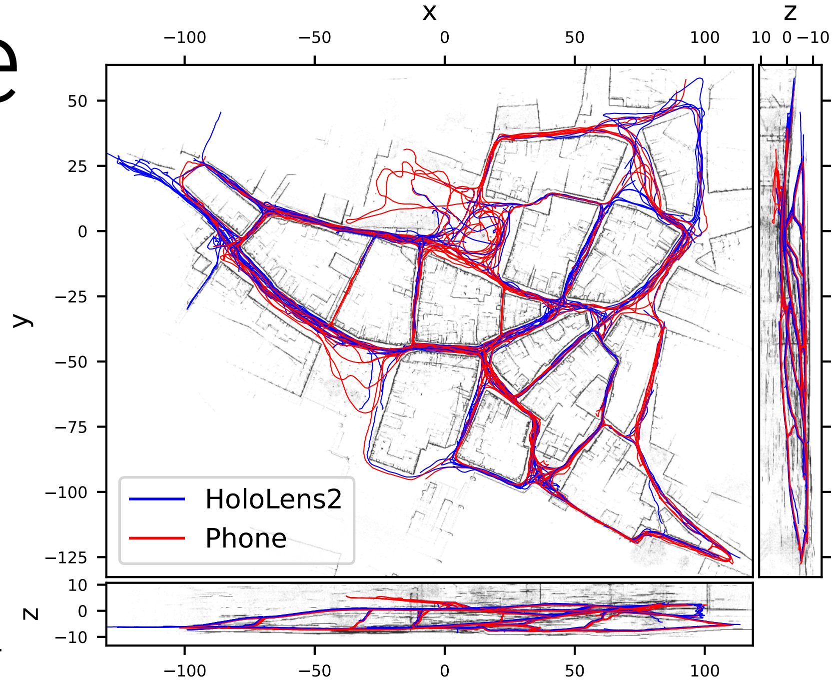
# Device coverage




HGE




CAB



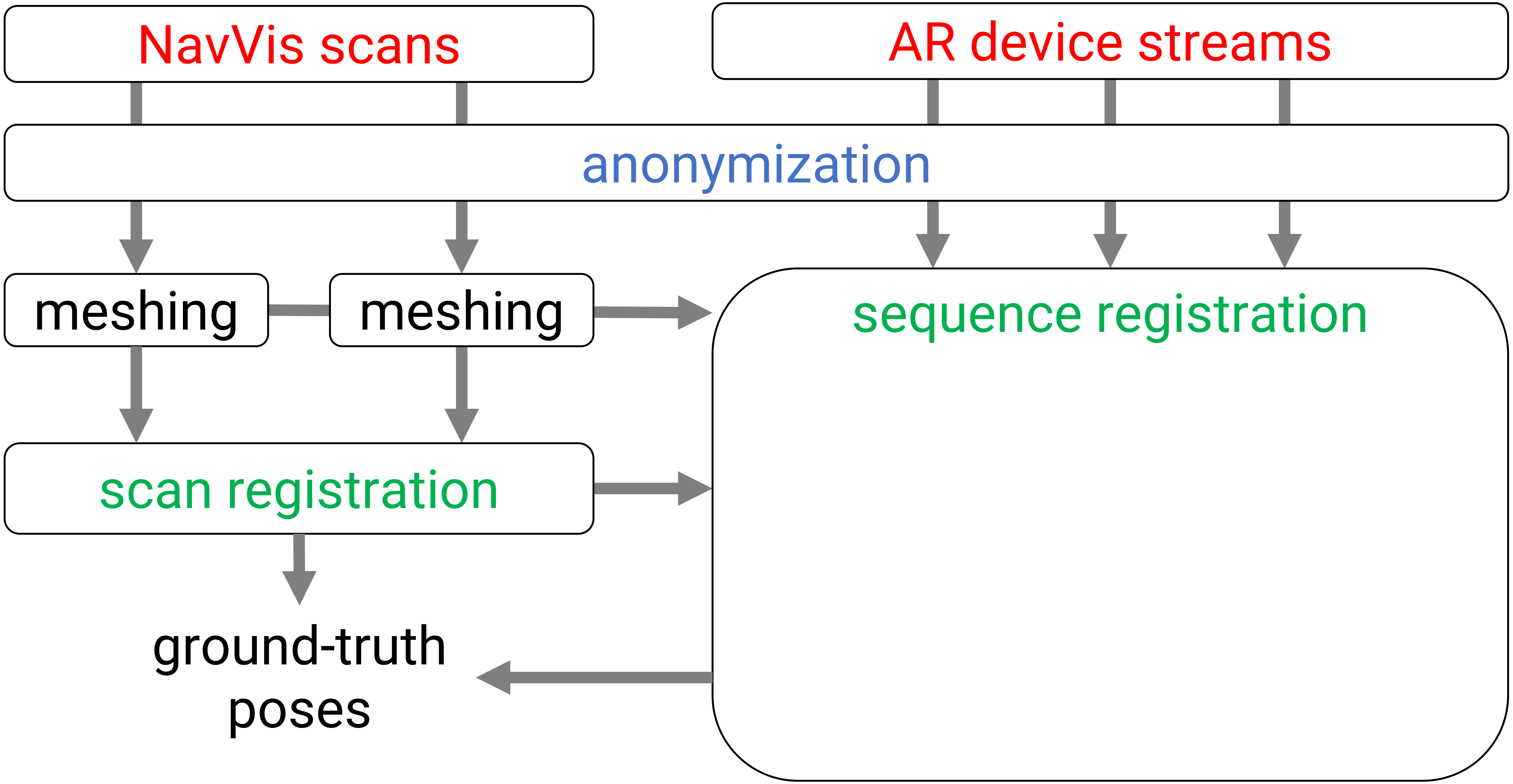

LIN

# Comparison with existing datasets

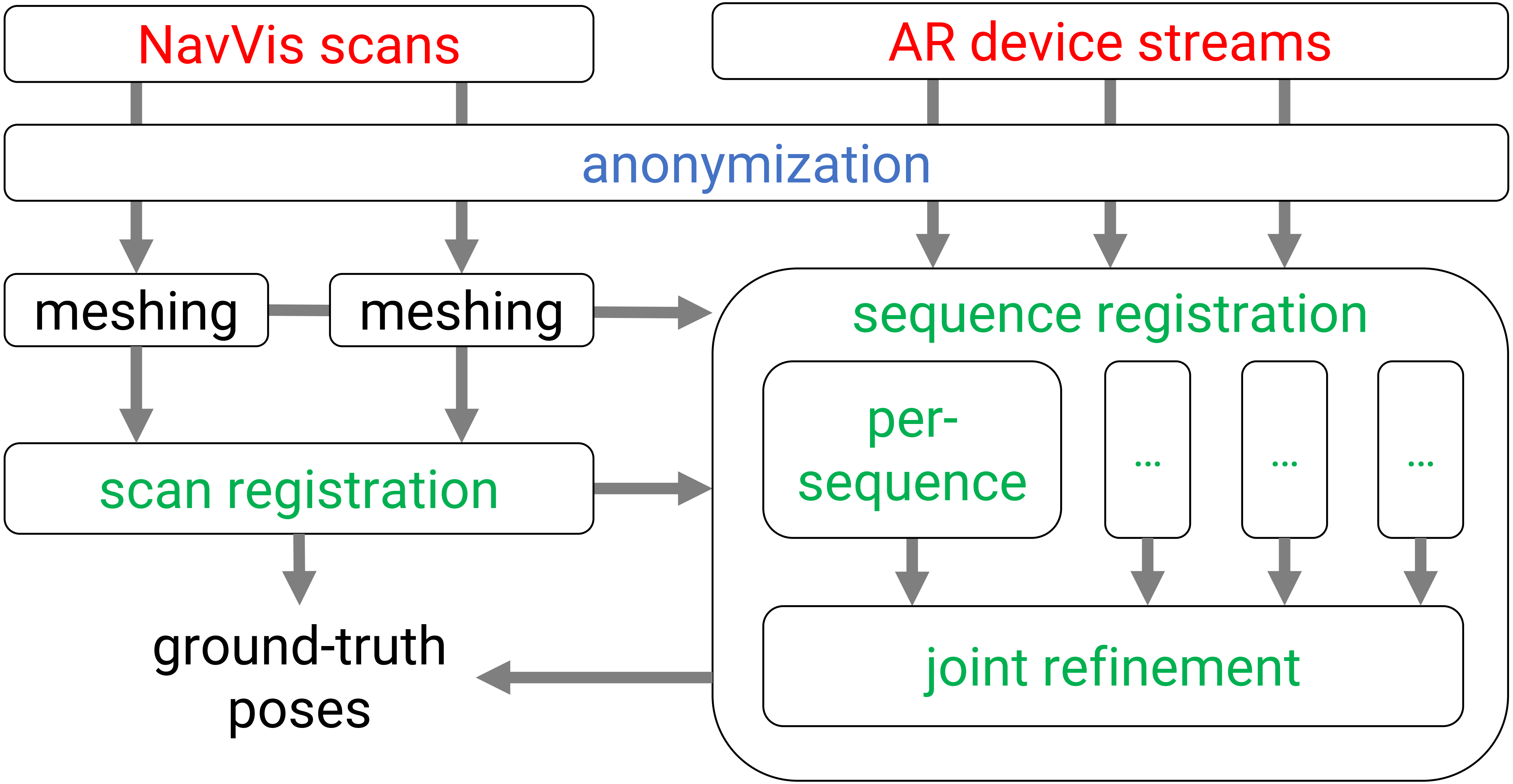
| dataset | out/indoor | changes | scale | density | camera motion | imaging devices | additional sensors | ground truth | accuracy |
|---|---|---|---|---|---|---|---|---|---|
| Aachen [67,66] | ✅❌ | 🌗🏗️ | ★★⯪ | ★★☆ | still images | DSLR | ❌ | SfM | >dm |
| Phototourism [34] | ✅❌ | 🏃🏗️ | ⯪☆☆ | ★★★ | still images | DSLR, phone | ❌ | SfM | ∼m |
| San Francisco [14] | ✅❌ | 🏃🏗️ | ★★★ | ★☆☆ | still images | DSLR, phone | GNSS | SfM+GNSS | ∼m |
| Cambridge [37] | ✅❌ | 🏃🌧️ | ⯪☆☆ | ★★☆ | handheld | mobile | ❌ | SfM | >dm |
| 7Scenes [73] | ❌✅ | ❌ | ⯪☆☆ | ★★★ | handheld | mobile | depth | RGB-D | ∼cm |
| RIO10 [84] | ❌✅ | 🪑 | ⯪☆☆ | ★★★ | handheld | Tango tablet | depth | VIO | >dm |
| InLoc [77] | ❌✅ | 🪑 | ★⯪☆ | ⯪☆☆ | still images | panoramas, phone | lidar | manual+lidar | >dm |
| Baidu mall [76] | ❌✅ | 🏃 | ★⯪☆ | ★★☆ | still images | DSLR, phone | lidar | manual+lidar | ∼dm |
| Naver Labs [40] | ❌✅ | 🏃🪑 | ★★☆ | ★★☆ | robot-mounted | fisheye, phone | lidar | lidar+SfM | ∼dm |
| NCLT [12] | ✅✅ | 🌧️🪑 | ★★☆ | ★★☆ | robot-mounted | wide-angle | lidar, IMU, GNSS | lidar+VIO | ∼dm |
| ADVIO [57] | ✅✅ | 🏃 | ★★☆ | ⯪☆☆ | handheld | phone, Tango | IMU, depth, GNSS | manual+VIO | ∼m |
| ETH3D [71] | ✅✅ | ❌ | ⯪☆☆ | ★★☆ | handheld | DSLR, wide-angle | lidar | manual+lidar | ∼mm |
| **LaMAR (ours)** | ✅✅ | 🏃🌧️🌗 🪑🏗️ | ★★⯪ 3 locations 45'000 $m^2$ | ★★★ 100 hours 40 km | handheld head-mounted | phone, headset backpack, trolley | lidar, IMU, 📶 ᛒ depth, infrared | lidar+SfM+VIO automated | ∼cm |

# b) Processed data

# Processing pipeline



# Processing pipeline



# Ground-truth visual overlap

[Predicting Visual Overlap of Images Through Interpretable Non-Metric Box Embeddings, Rau et al, ECCV 2020]

# Ground-truthing

- Basic recipe: fuse multiple constraints:
  - Image matching 2D/3D-3D with GT laser data
  - SLAM poses: NavVis (rigid) and mobile (non-rigid)

# Ground-truthing

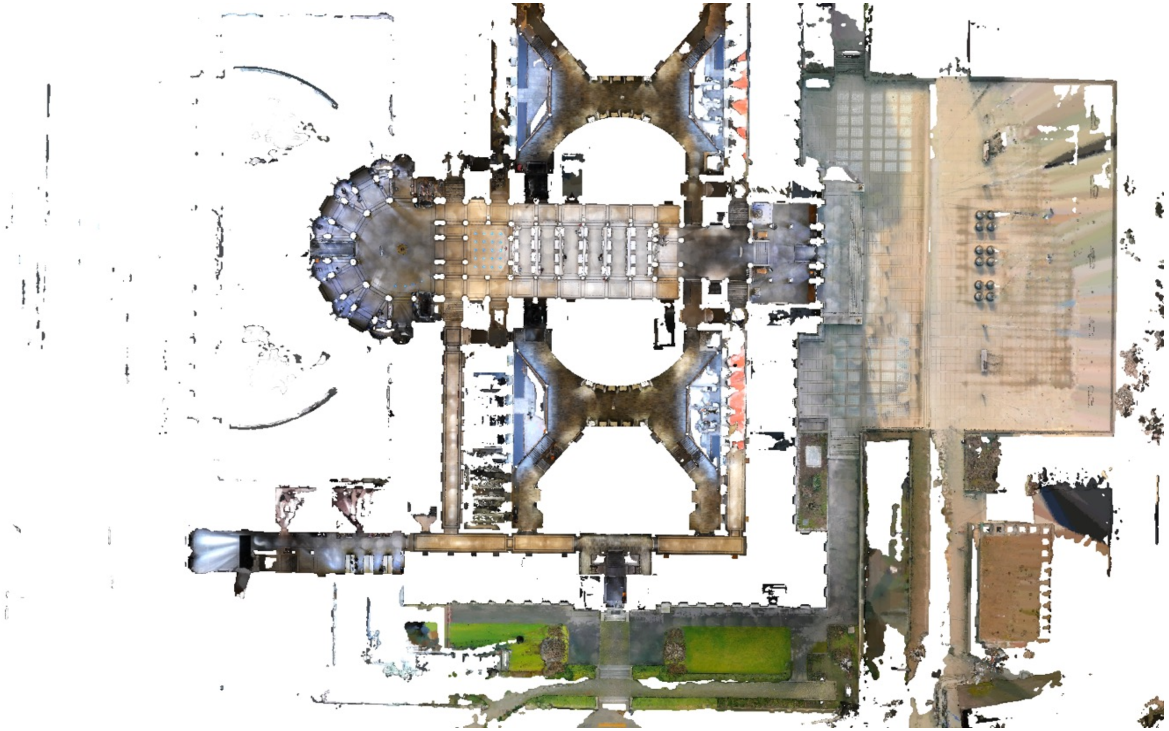
- Basic recipe: fuse multiple constraints:
  - Image matching 2D/3D-3D with GT laser data
  - SLAM poses: NavVis (rigid) and mobile (non-rigid)
- Assumptions:
  - Lidar-inertial poses = small, negligible drift
  - VI poses = larger drift but locally reliable

# Ground-truthing

- Basic recipe: fuse multiple constraints:
  - Image matching 2D/3D-3D with GT laser data
  - SLAM poses: NavVis (rigid) and mobile (non-rigid)
- Assumptions:
  - Lidar-inertial poses = small, negligible drift
  - VI poses = larger drift but locally reliable
- Entirely automated
  - No manual annotations
  - No fiducial markers

# Scan & sequence registration

scans → [image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] pairwise registration

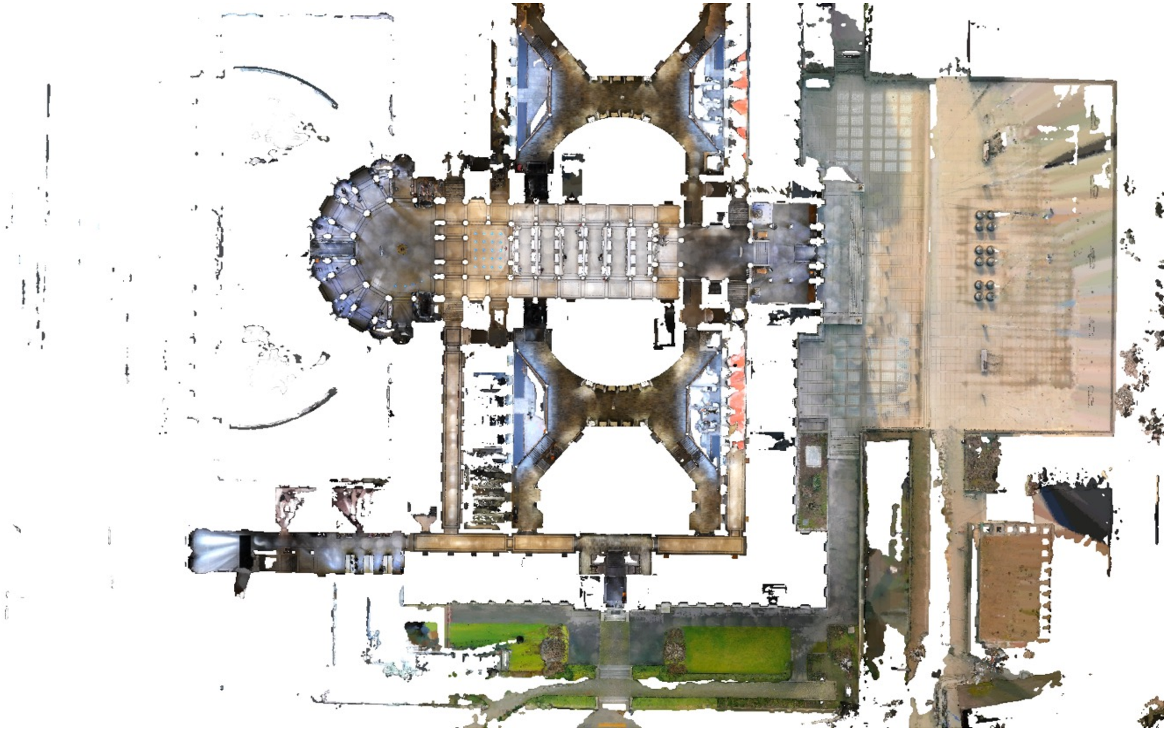


28

Lidar scans
~180M points each

# Scan & sequence registration

scans → [pairwise registration: image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] → Pose graph optimization → registered NavVis mesh + images

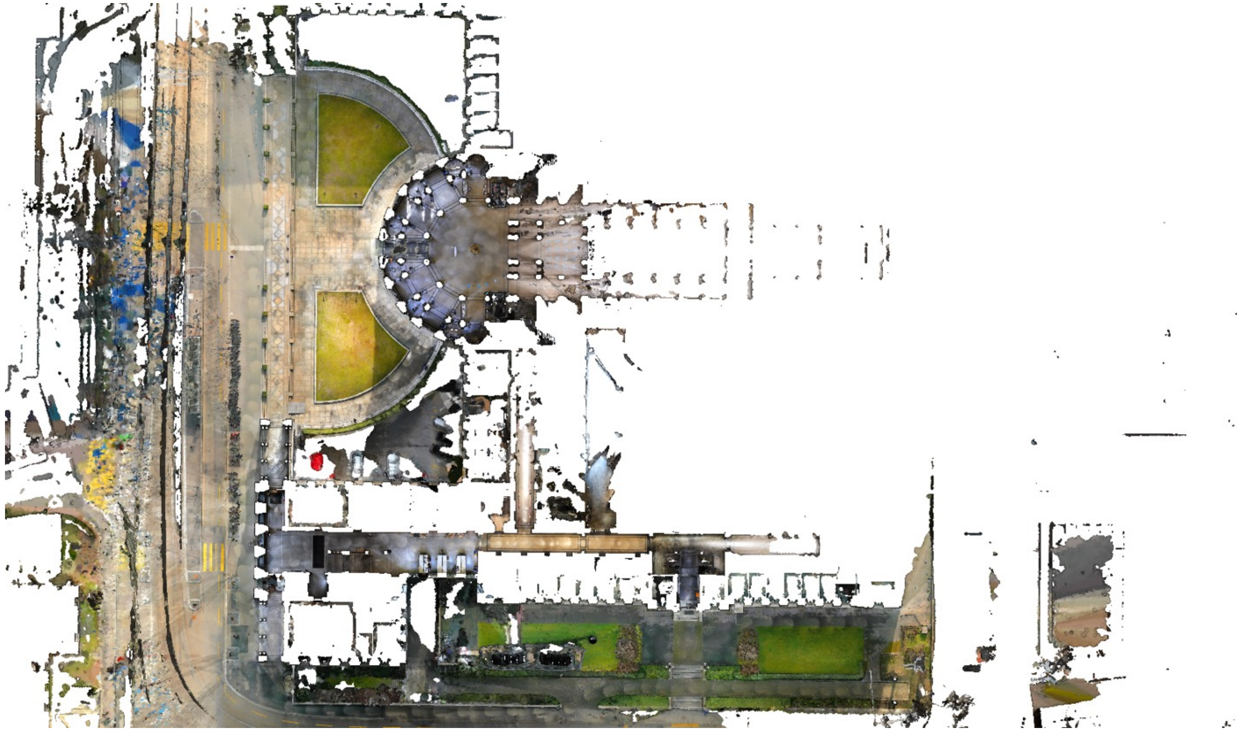


Lidar scans
~180M points each

28

# Scan & sequence registration

scans → [pairwise registration: image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] → Pose graph optimization → registered NavVis mesh + images

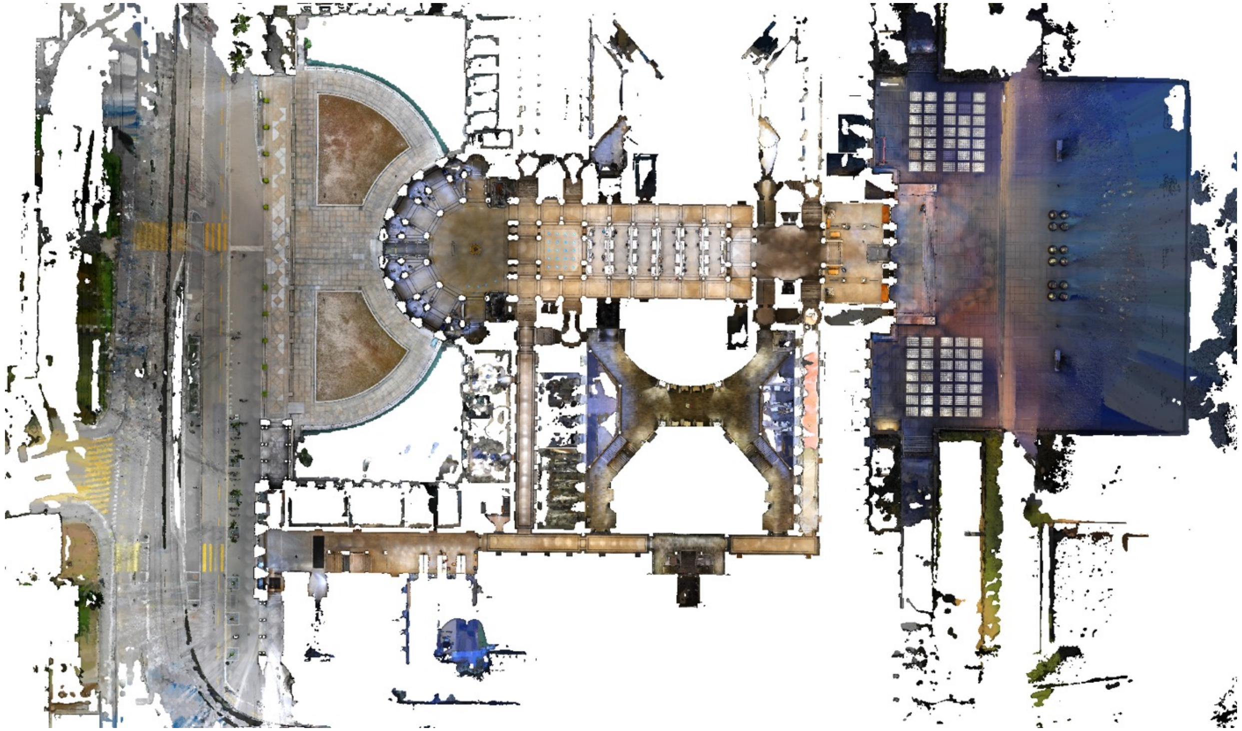


Lidar scans
~180M points each

28

# Scan & sequence registration

scans → [ image retrieval → image matching → 3D-3D lifting → RANSAC + ICP ] (pairwise registration) → Pose graph optimization → registered NavVis mesh + images



28

Lidar scans
~180M points each

# Scan & sequence registration

scans → [ image retrieval → image matching → 3D-3D lifting → RANSAC + ICP ] (pairwise registration) → Pose graph optimization → registered NavVis mesh + images

Lidar scans
~180M points each

28

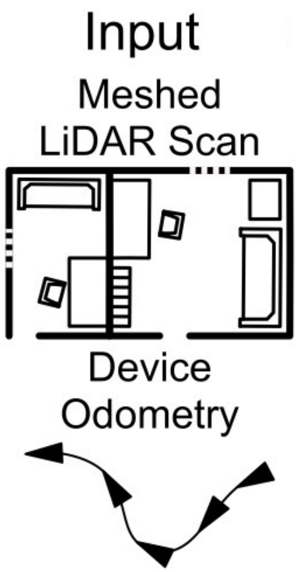
# Scan & sequence registration

scans

image retrieval → image matching → 3D-3D lifting → RANSAC + ICP → Pose graph optimization

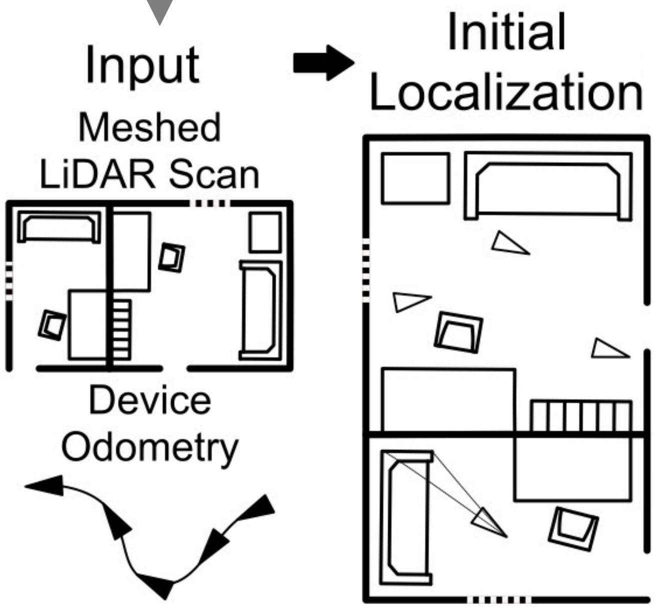
pairwise registration

registered NavVis mesh + images



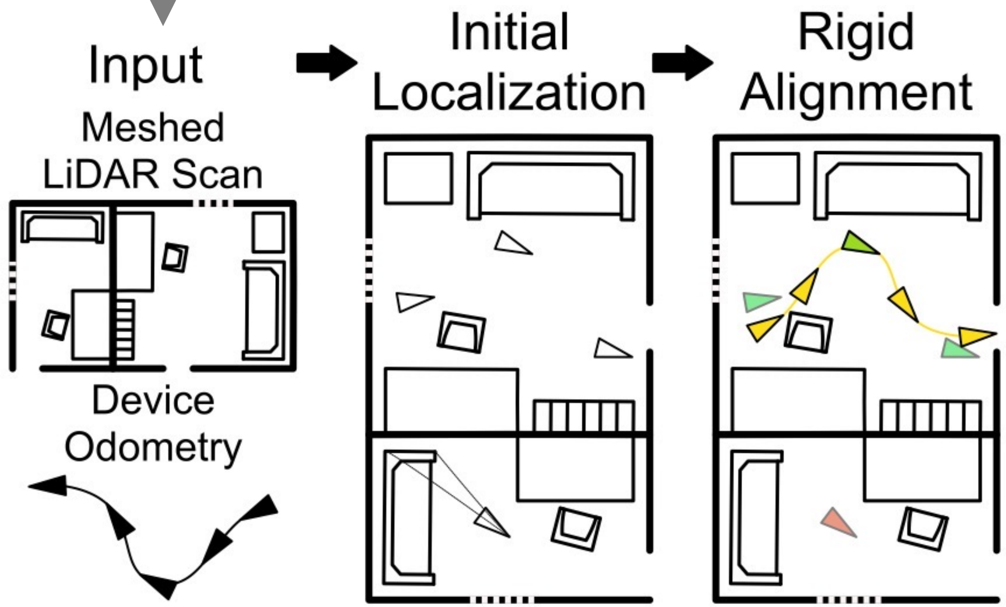
# Scan & sequence registration

scans → [pairwise registration: image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] → Pose graph optimization → registered NavVis mesh + images



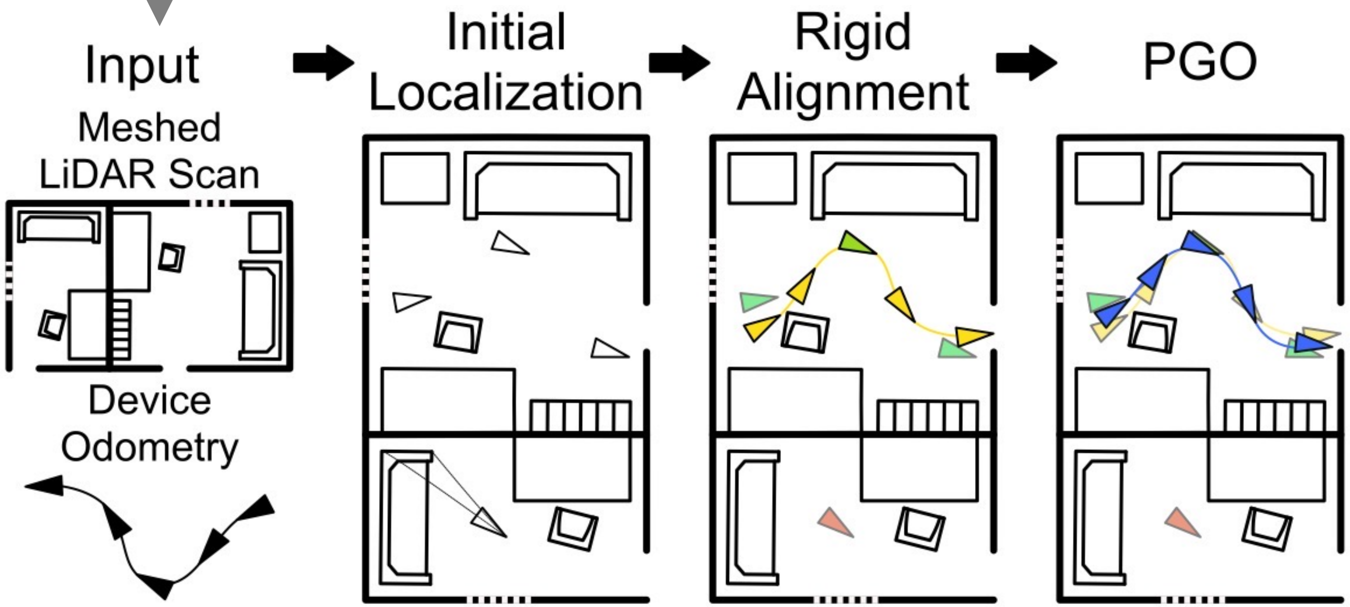
# Scan & sequence registration

scans → [pairwise registration: image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] → Pose graph optimization → registered NavVis mesh + images



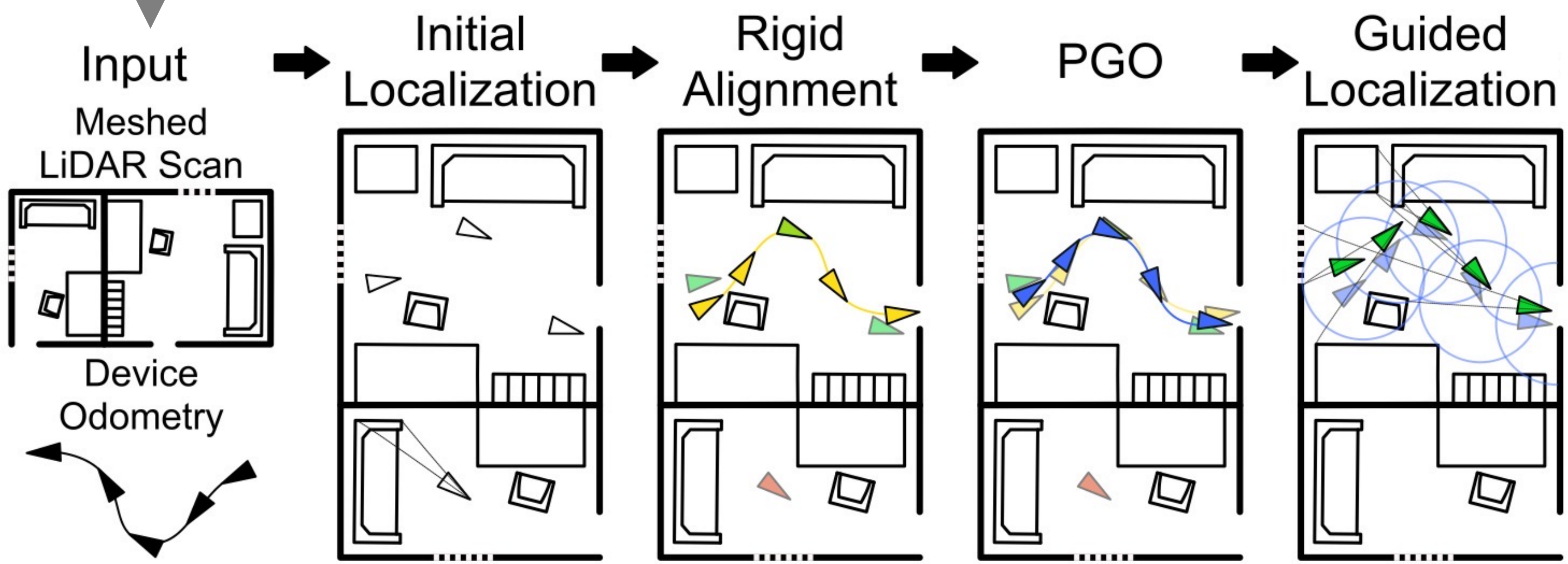
# Scan & sequence registration

scans → [image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] (pairwise registration) → Pose graph optimization → registered NavVis mesh + images



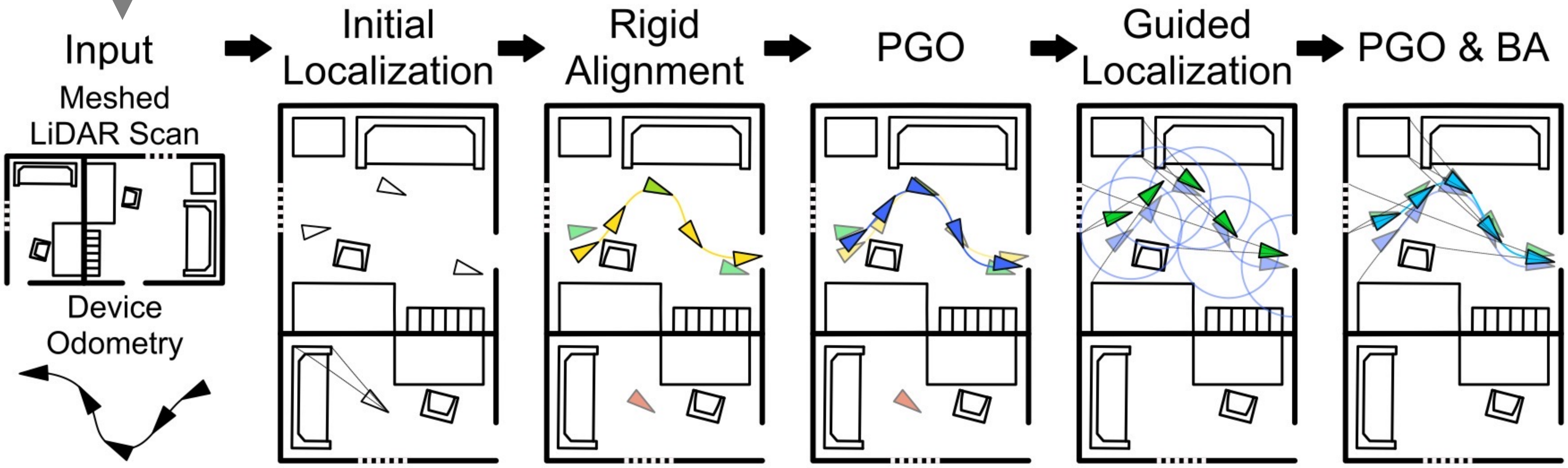
# Scan & sequence registration

scans → [pairwise registration: image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] → Pose graph optimization → registered NavVis mesh + images



# Scan & sequence registration

scans → [pairwise registration: image retrieval → image matching → 3D-3D lifting → RANSAC + ICP] → Pose graph optimization → registered NavVis mesh + images

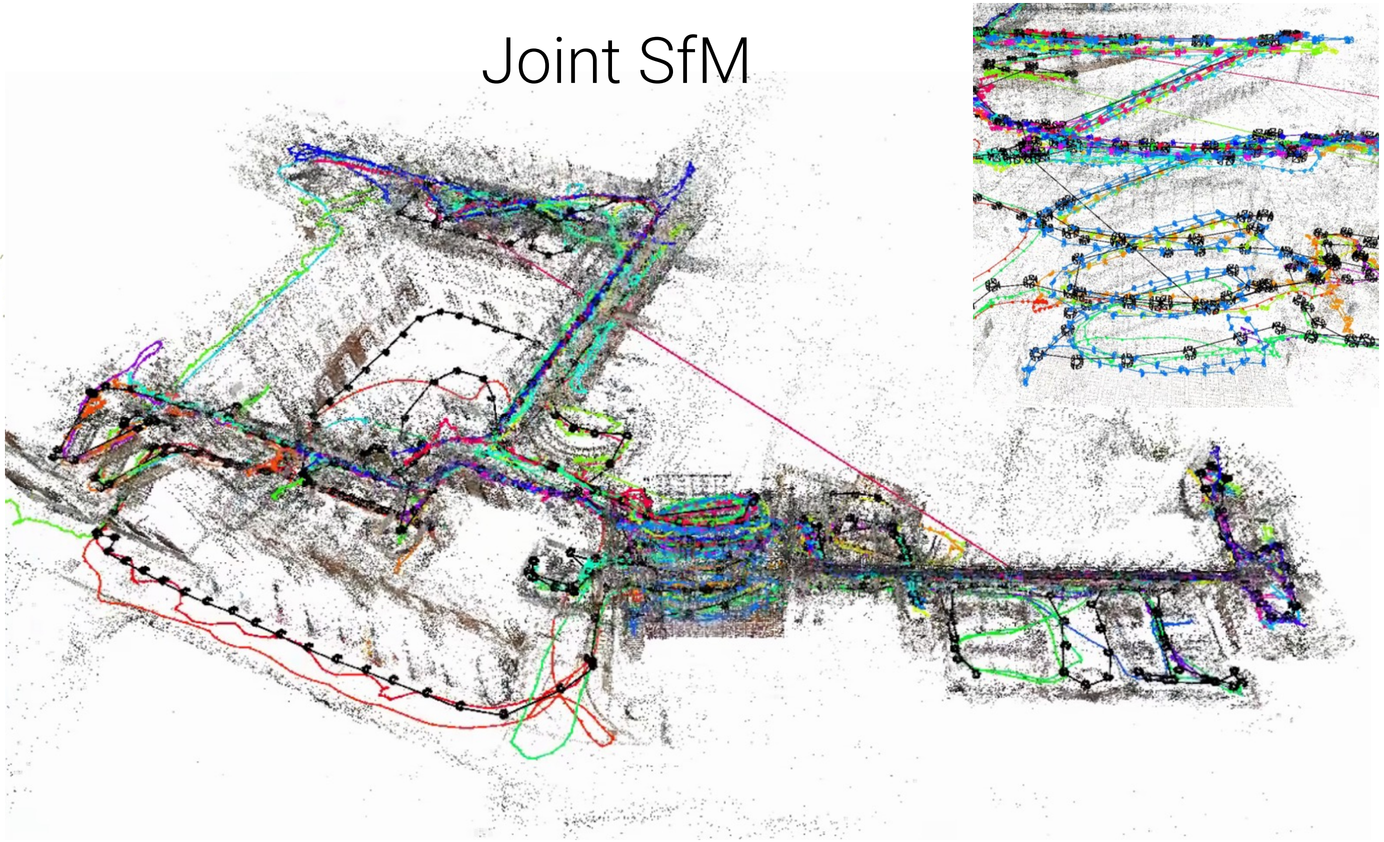


# Joint refinement

## Bundle adjustment with relative pose constraints from SLAM

Lidar scans ~300M points

Joint SfM

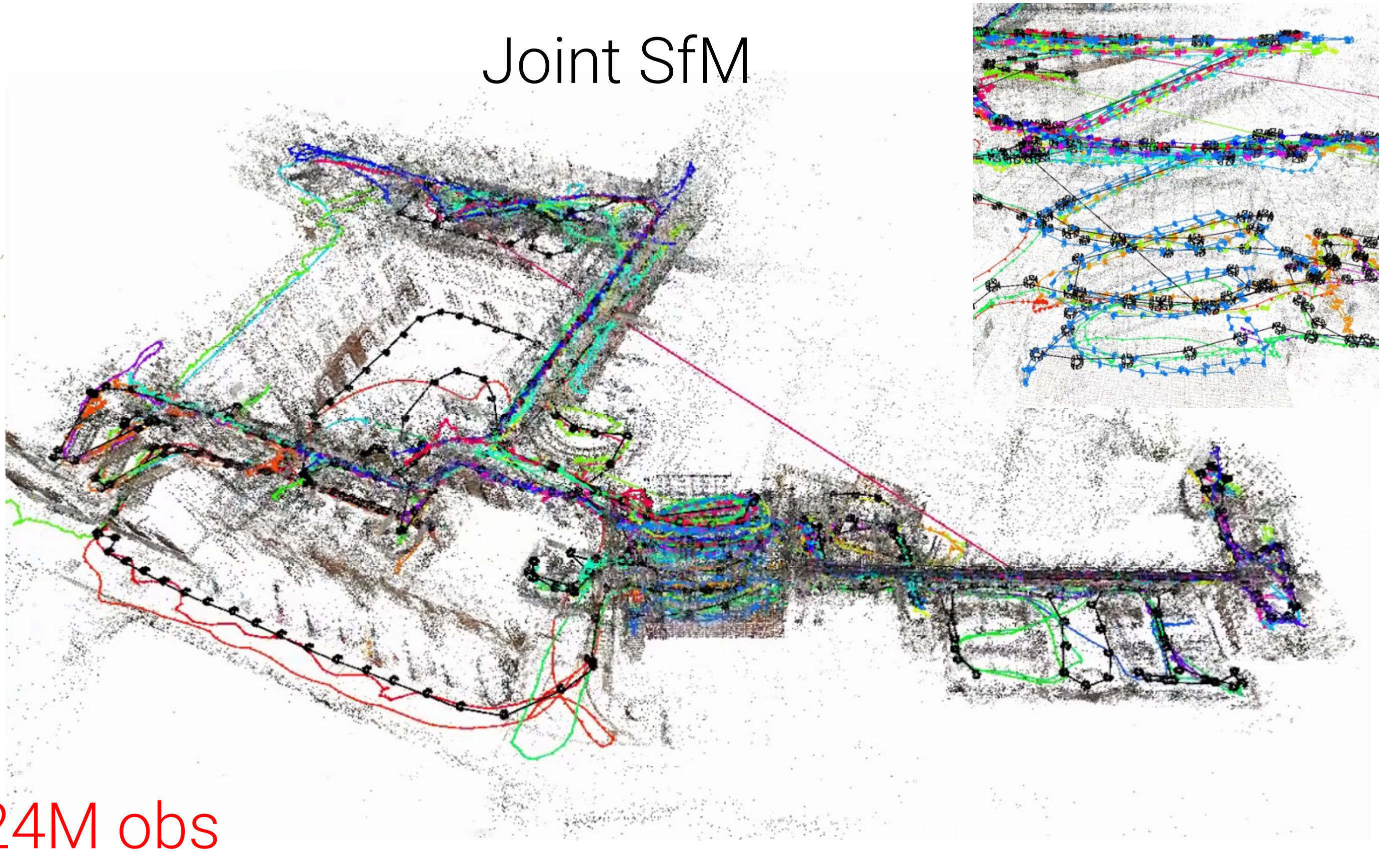


# Joint refinement

## Bundle adjustment with relative pose constraints from SLAM

Lidar scans ~300M points

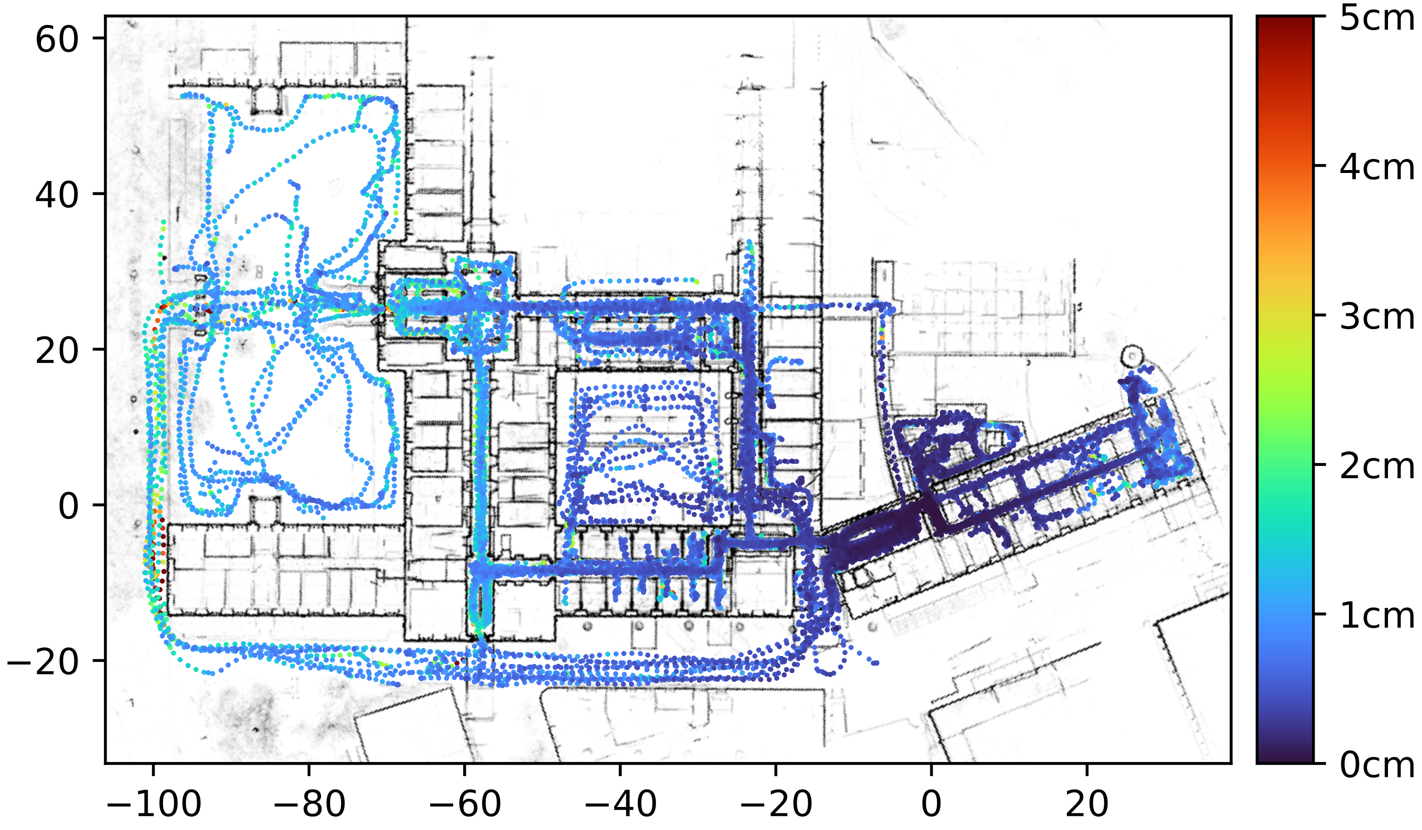
Joint SfM

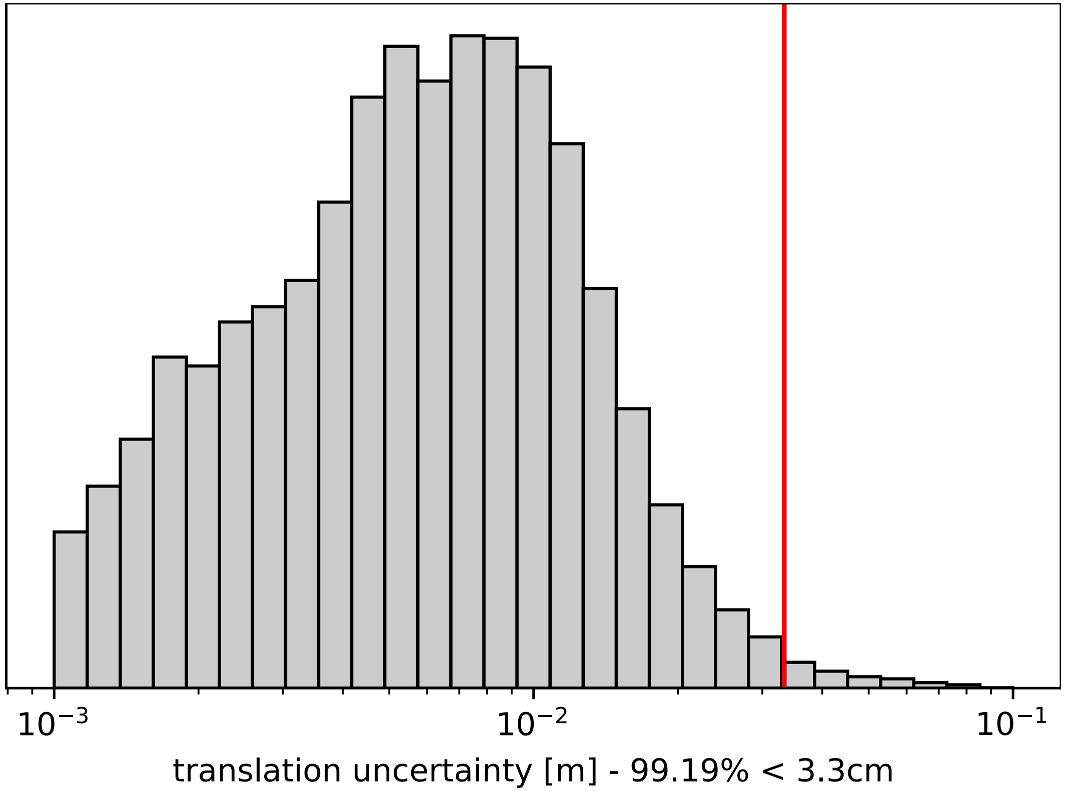


CAB: 53k images, 6.2M points, 24M obs
LIN: 49k images, 7.7M points, 37M obs
HGE: 50k images, 5.3M points, 29M obs

# Ground-truthing - accuracy

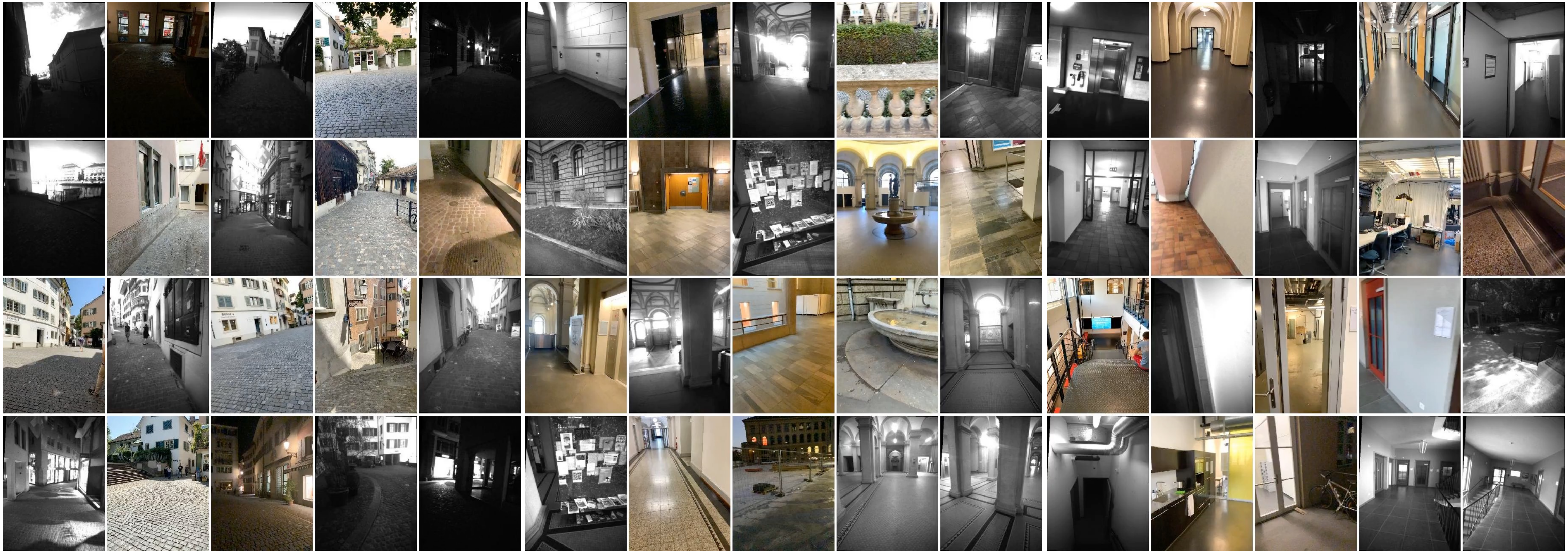
## Estimate covariances from the joint optimization



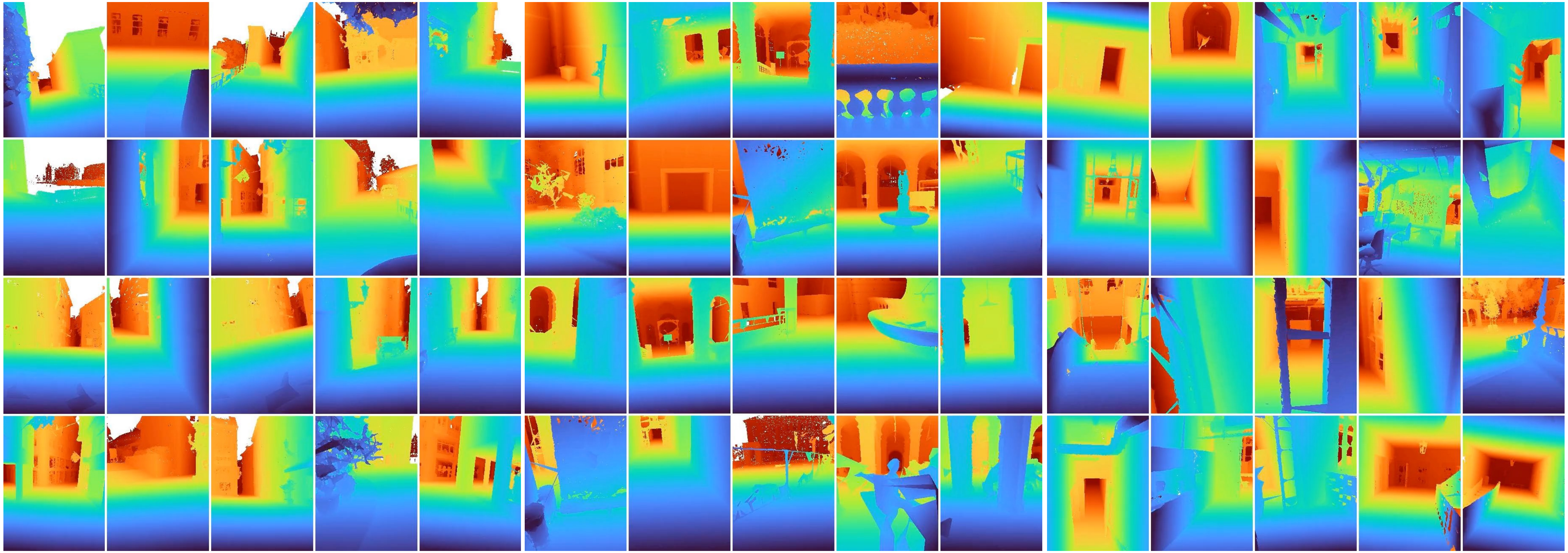

translation uncertainty

# Ground truth at scale



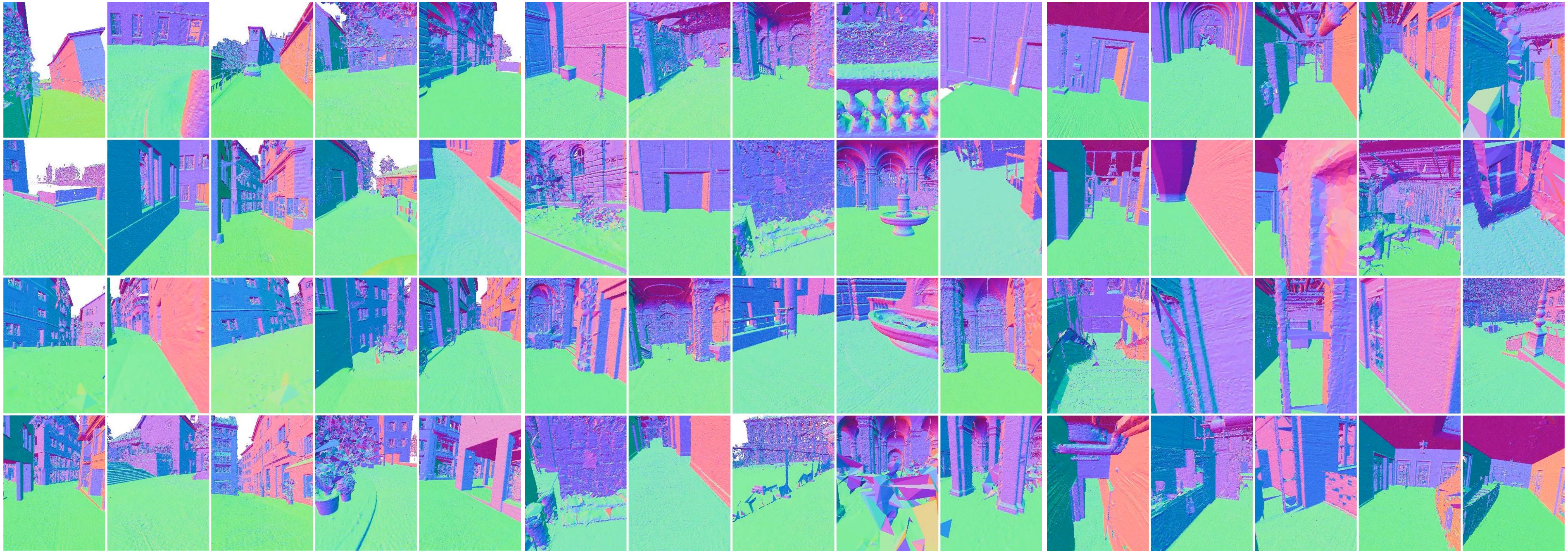
**Automated ground truthing:** no manual annotation!

# Ground truth at scale



**Automated ground truthing:** no manual annotation!

# Ground truth at scale

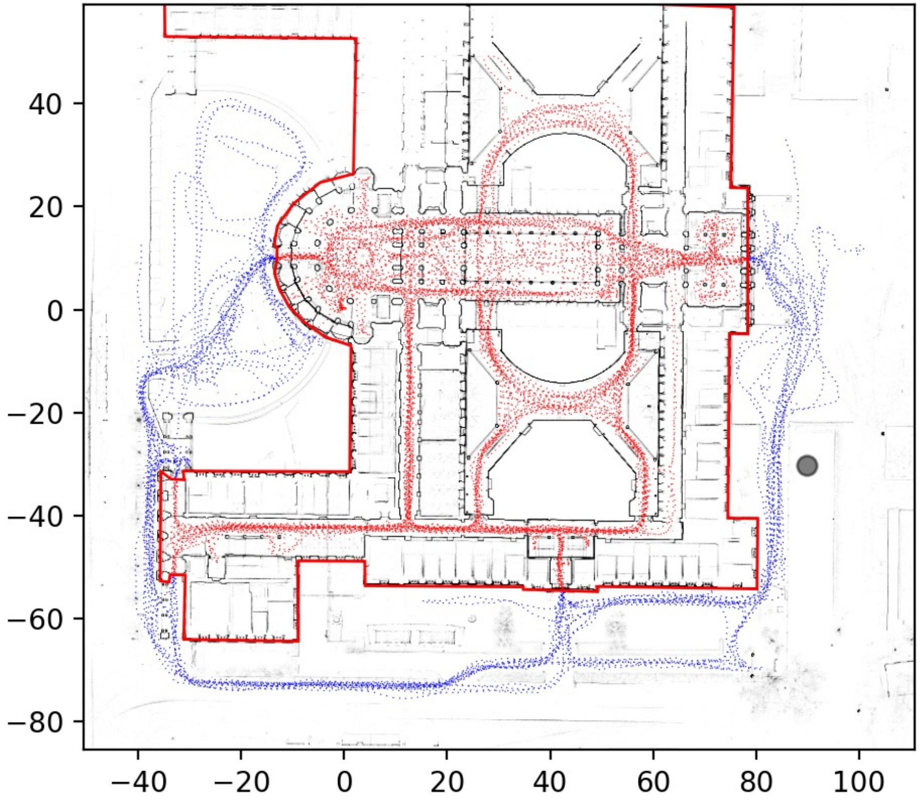


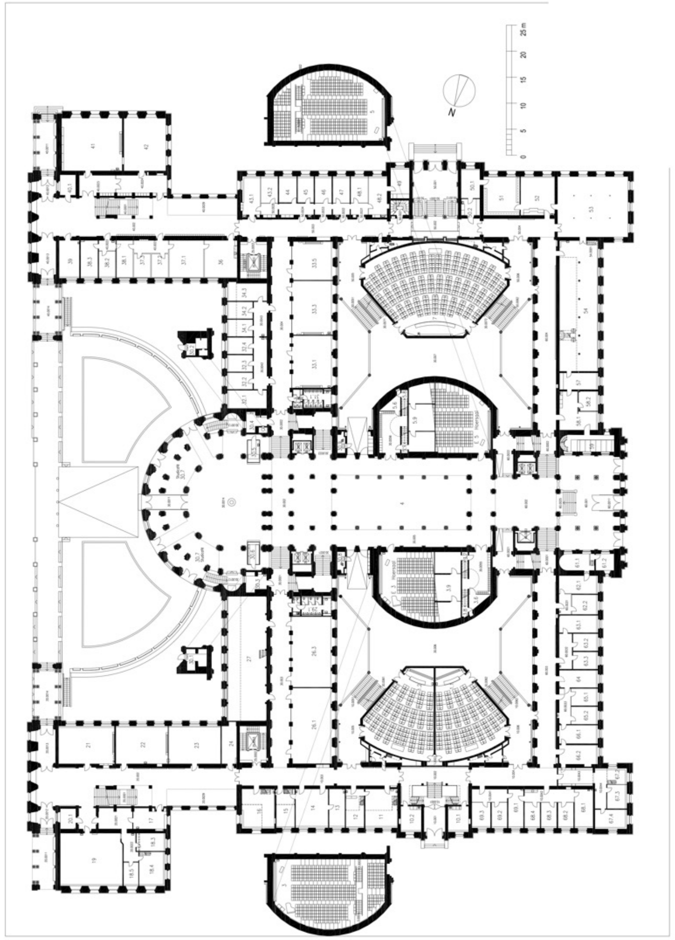
**Automated ground truthing:** no manual annotation!

# Additional data

- Per-timestamp tag: day/night, indoor/outdoor
- Floorplans for indoor: walls, doors, stairs

indoor / outdoor

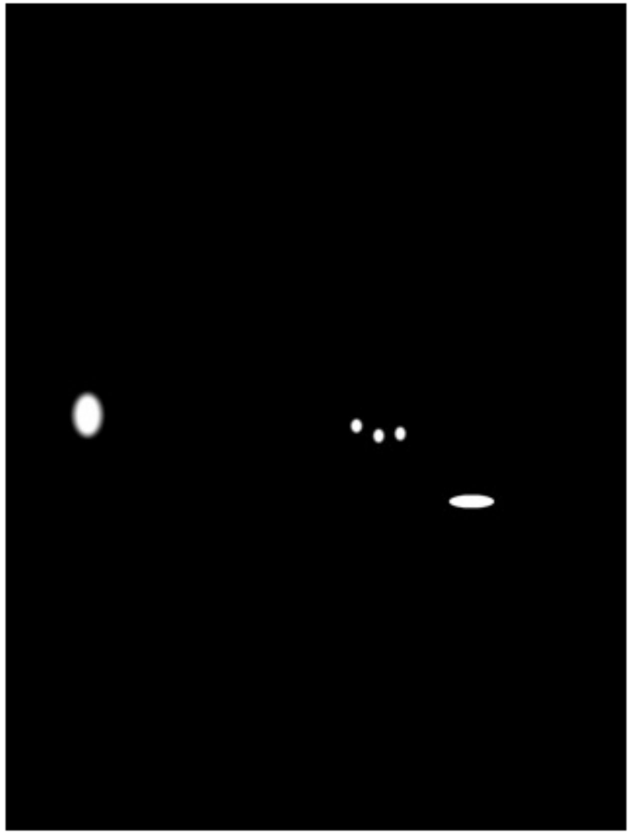


[Dynamic and scalable large scale image reconstruction, Stretcha et al, CVPR 2010]

# Anonymization

- Images: faces + license plates
- Radio endpoint identifiers



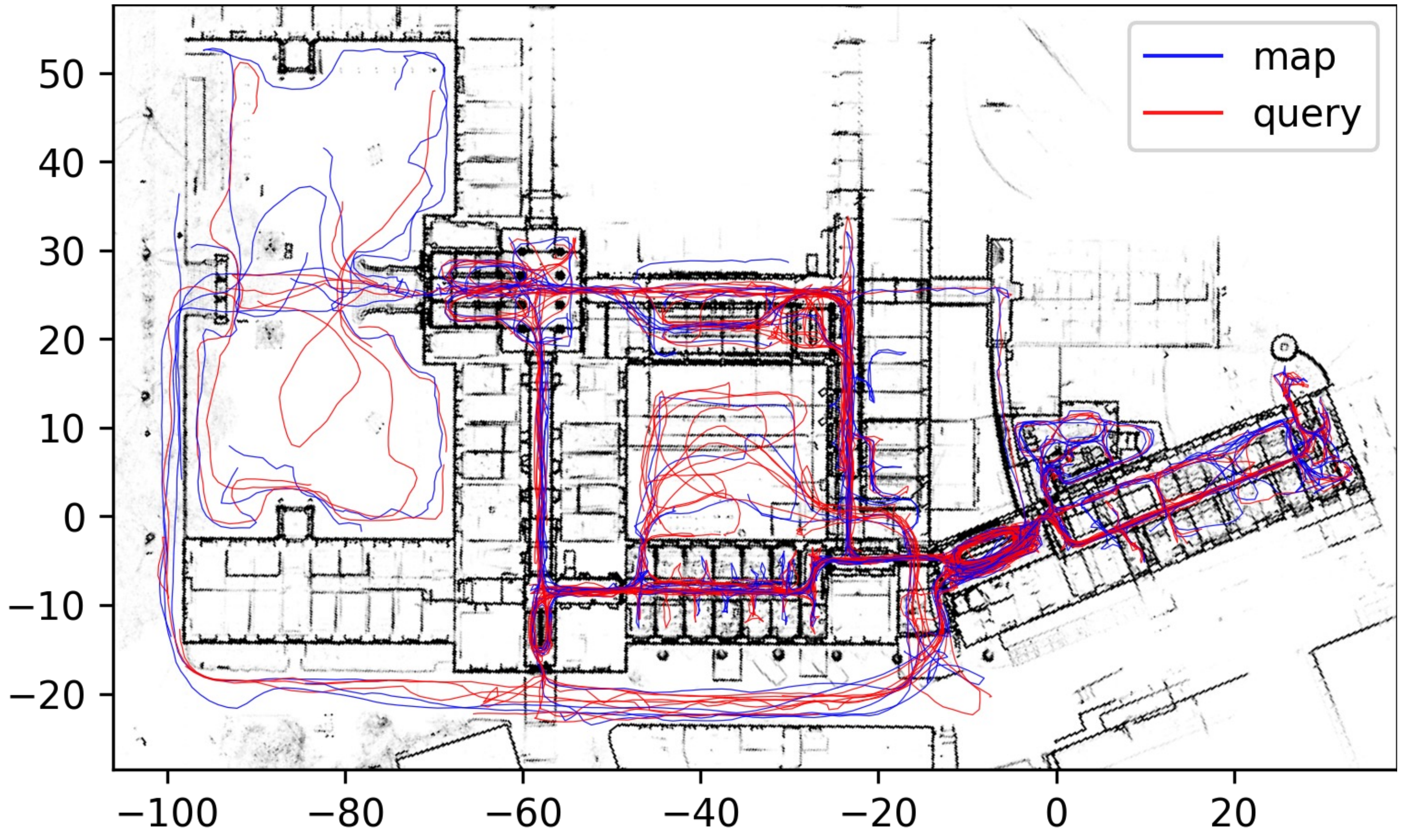
# c) Use cases

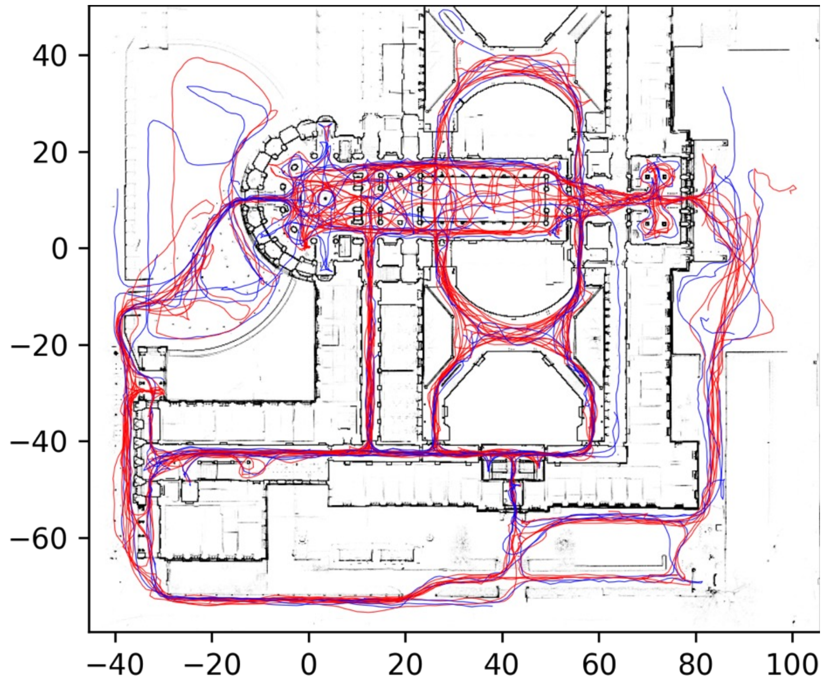
# Localization & mapping

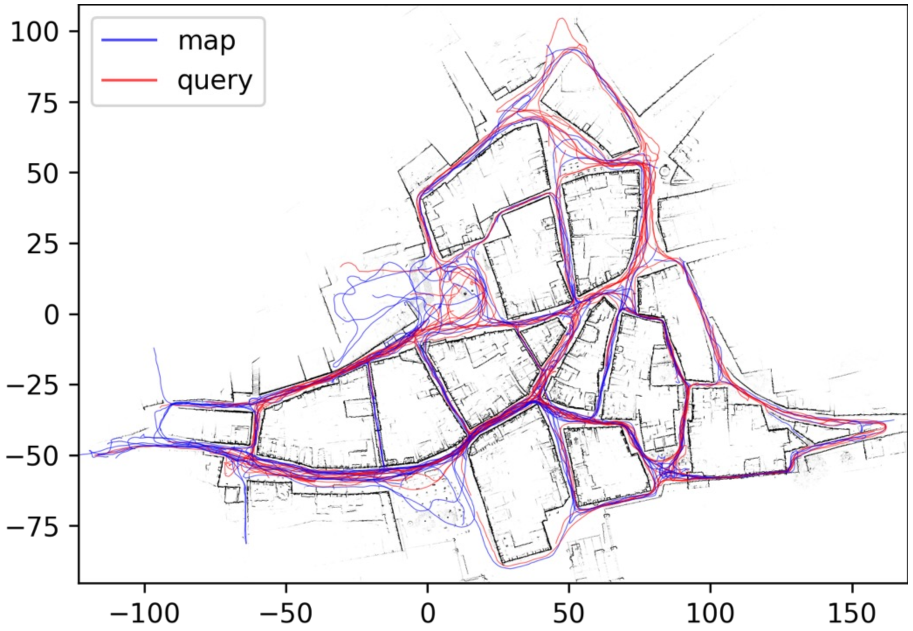
- Simulate crowd-sourced mapping: select some AR sequences
- Full flexibility in adjusting the difficulty
  - Easier maps with increased map-query overlap
- Algorithm to automatically compute the split
  - Minimize the coverage between map sequences
  - Ensure minimum coverage of all queries
- Keyframe mapping & query data at 2.5FPS/50cm & 1FPS/1m

| dataset | CAB | HGE | LIN | Aachen v1.1 | InLoc |
|---|---|---|---|---|---|
| # mapping images | **34k** | **26k** | **38k** | 6.7k | 10k |

# Localization & mapping

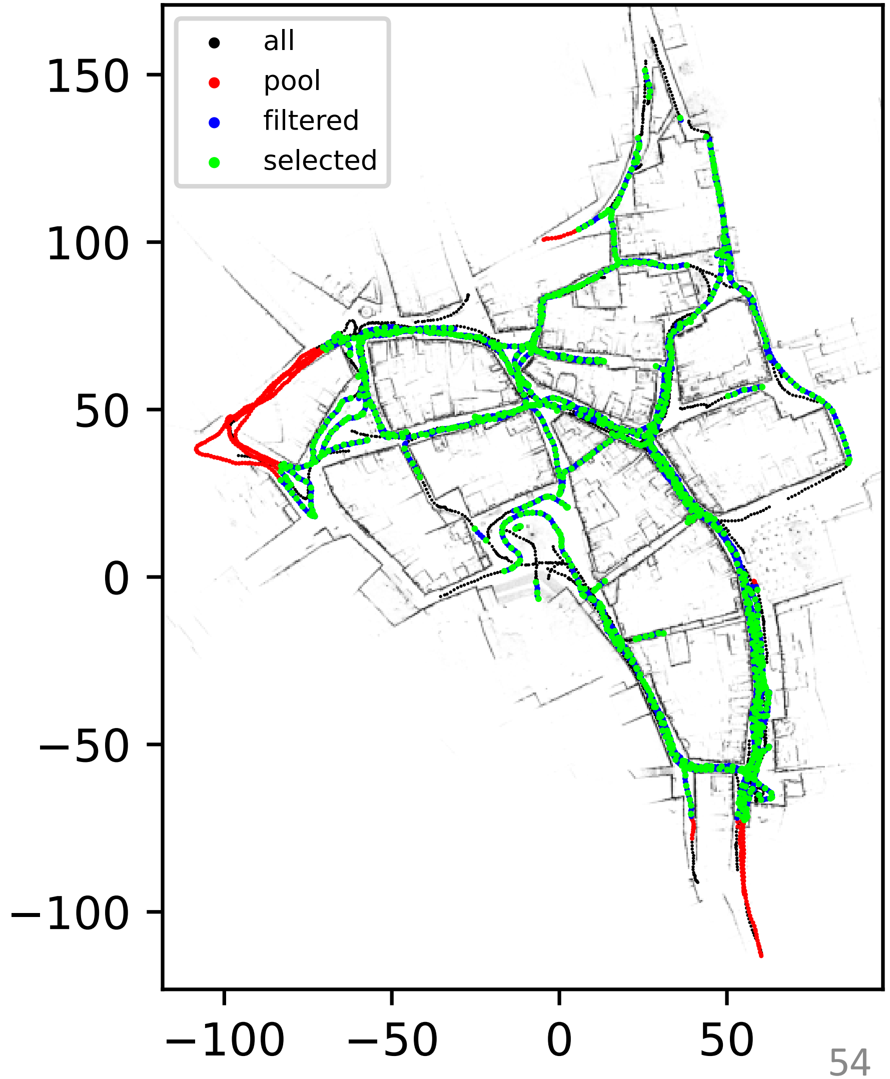


# Localization & mapping

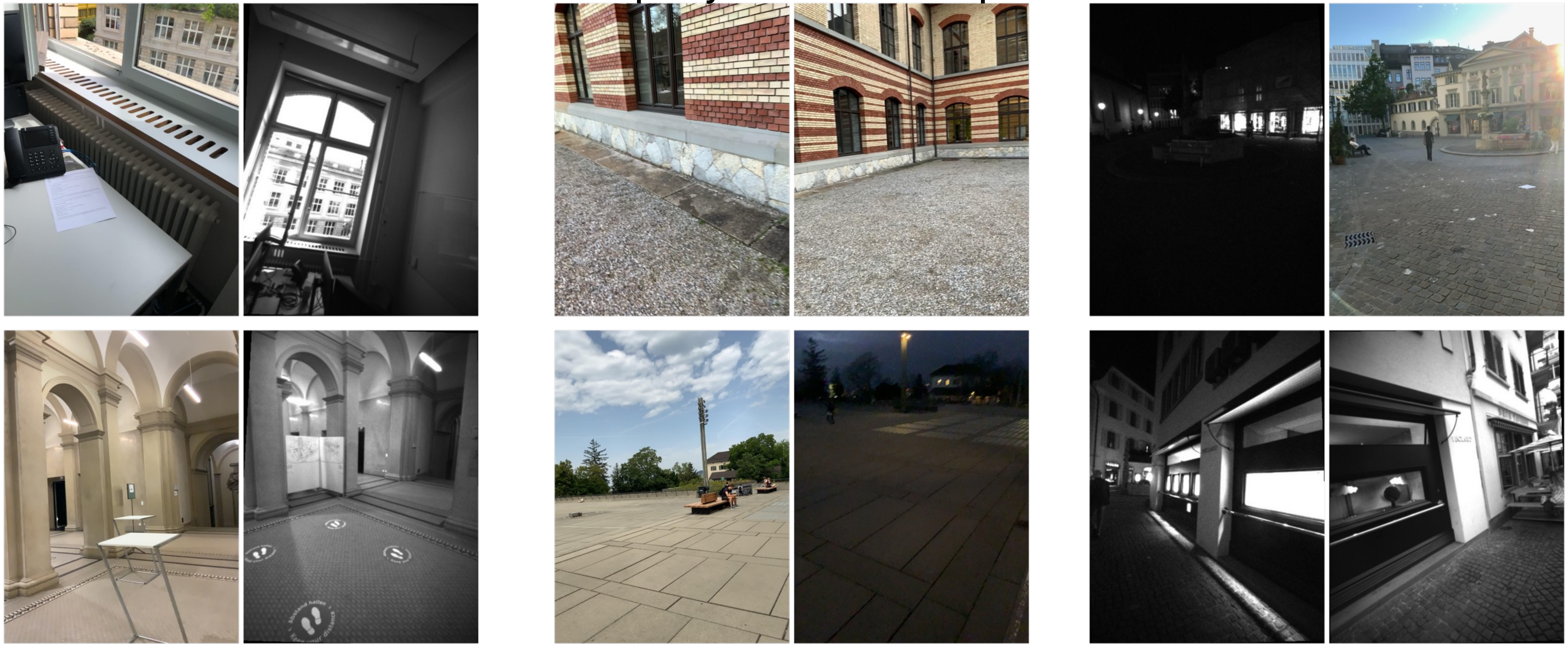
- Randomly sample query images
  - well distributed on the map
  - 1k per scene per device
  - Pad with 20s for sequence evaluation
- Reject queries with high pose uncertainty



# Localization & mapping

## Examples of hard queries with best map overlap

query best map



# Learning-based localization

- We did not try to train PoseNet/ESAC on LaMAR
  - Because past datasets of similar scale failed to make them work
- Good test bed to learn priors of the dynamic world
- Much larger & harder than 7Scenes and Cambridge Landmarks

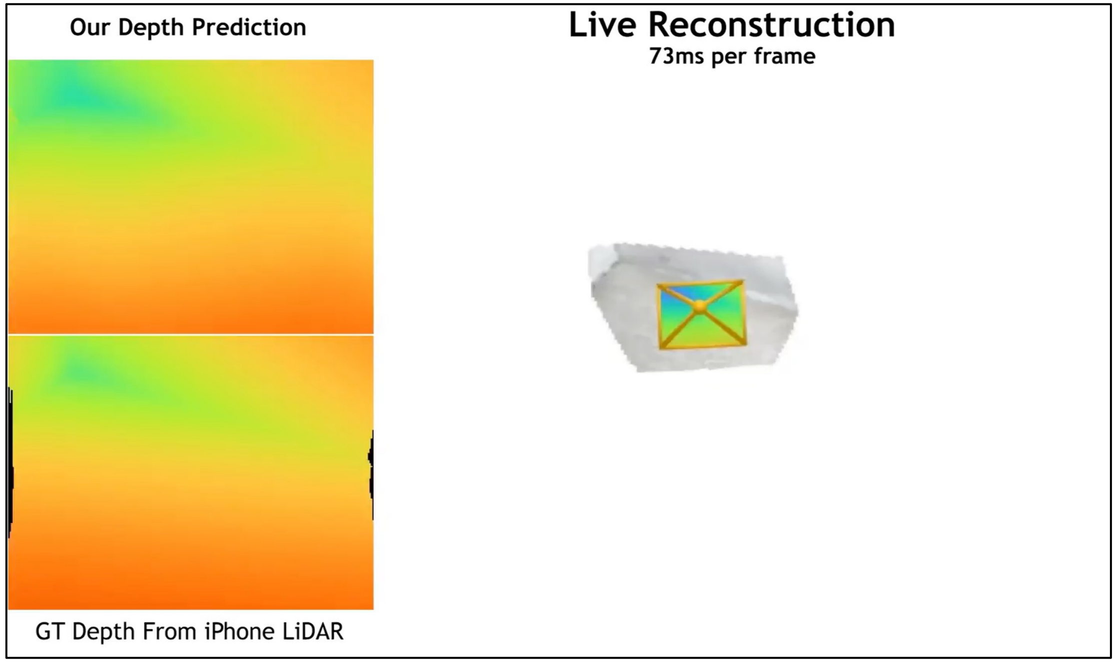
[Beyond Controlled Environments: 3D Camera Re-Localization in Changing Indoor Scenes, Wald et al, ECCV 2020]

# Odometry & SLAM

- Sequences of calibrated images with GT poses
- Calibrated depth from sensors / GT
- IMU @ 100Hz
    - but not calibrated on phones (IMU-camera & intrinsics)
    - could be estimated within the GT pipeline

# Dense 3D reconstruction

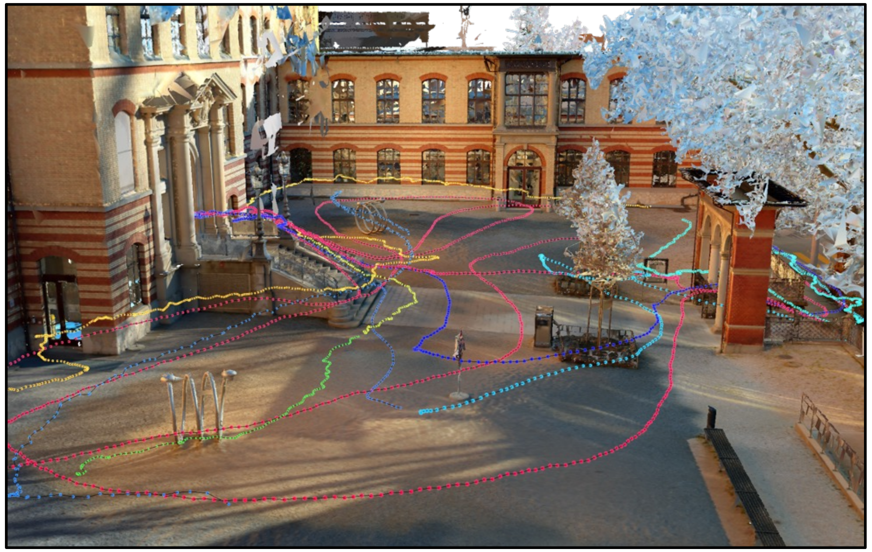
- Posed & calibrated RGB(-D) images
- GT depth maps from the mesh
- Evaluate MVS or RGB-D fusion
  - Across devices and time
  - Handle different cameras and conditions (day vs night)
- Pitfall: not perfect & consistent GT




[SimpleRecon: 3D Reconstruction Without 3D Convolutions, Sayed et al, ECCV 2022]

# View synthesis, rendering

- High-res colored mesh, posed & calibrated images
- Building-scale indoor NeRF
- Pitfall:
  - Mesh is not perfect and has artefacts
  - but could easily be improved
  - Lidar point cloud is much cleaner



[NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, Mildenhall et al, ECCV 2020]

# d) Outlook

# Public release

- Evaluation data released this week
  - Mapping + query sets
  - Keyframed
- Full data will be released later: lidar, full-FPS data, floorplans, etc.
- License
  - CC-BY-SA for all raw data → allow commercial use
  - But: GT pipeline includes SuperPoint+SuperGlue

# Limitations

- iOS does not expose full radios:
  - no WiFi, anonymized BT, no BT beacon
- GTing
  - NavVis is assumed perfect but is not always
    - Blackbox software, no multi-session optimization
  - no tight IMU integration, no tracking covariances
  - uncertainty is likely underestimated: camera-only covariances
  - But have plans to improve it
- Mesh is not perfect

# Community contributions

- Your company develops AR devices?
- Consider exposing raw sensor data via a “**research mode**”


Microsoft HoloLens 2 / Apple iPad / Meta Project Aria / Magic Leap 2

# Q&A

Next: benchmarking localization & mapping