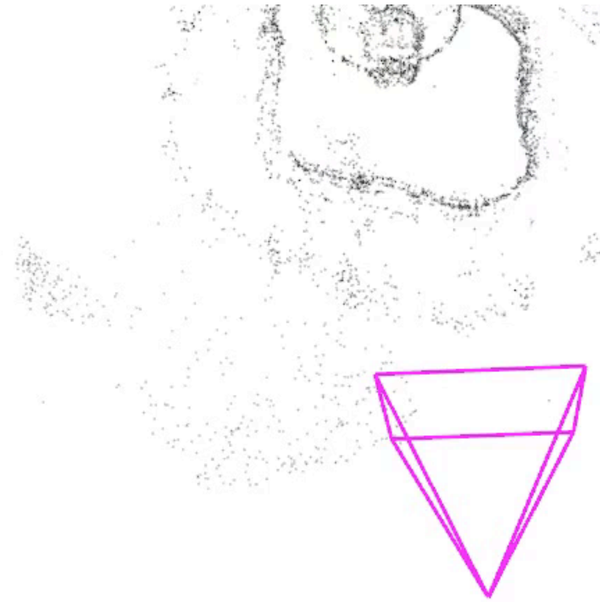
Scene Representations for Visual Localization

Torsten Sattler

Czech Institute of Informatics, Robotics and Cybernetics
Czech Technical University in Prague

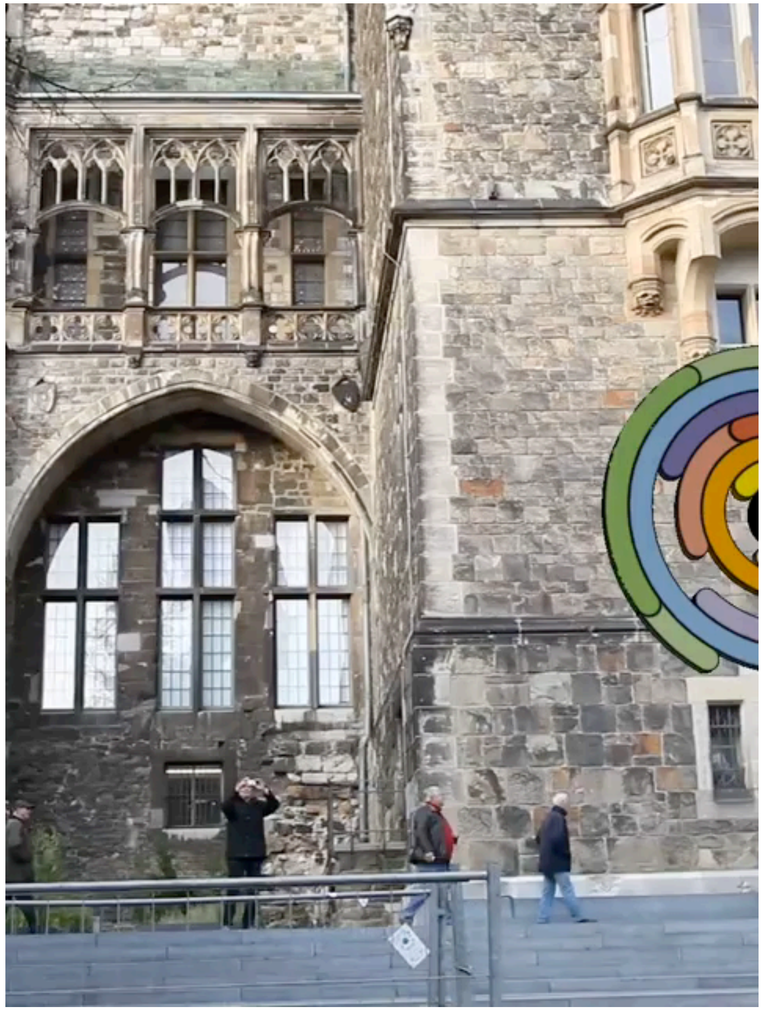
CZECH INSTITUTE OF INFORMATICS ROBOTICS AND CYBERNETICS CTU IN PRAGUE

The Visual Localization Problem

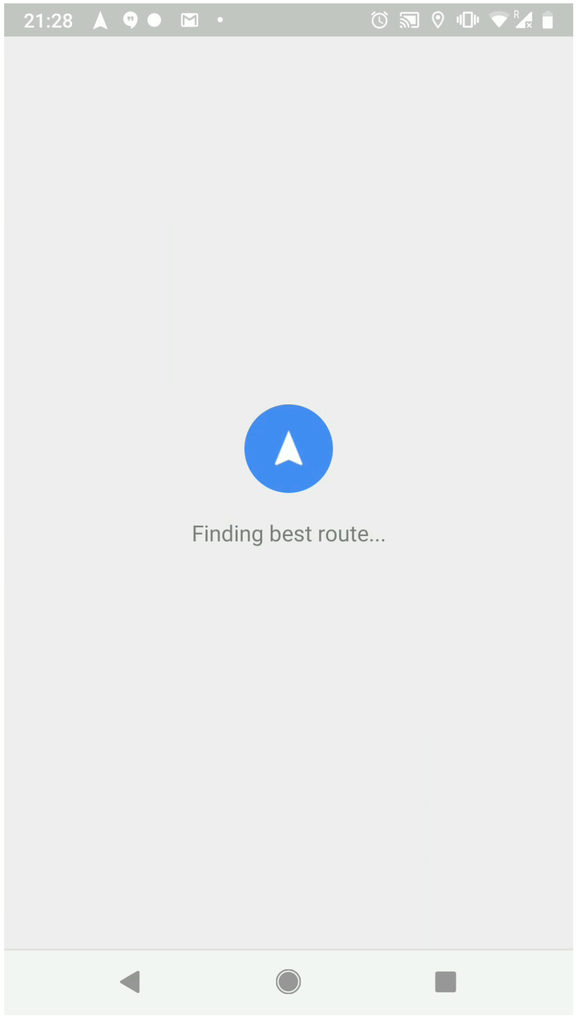


Compute **exact position and orientation** of query image

Applications: Augmented Reality

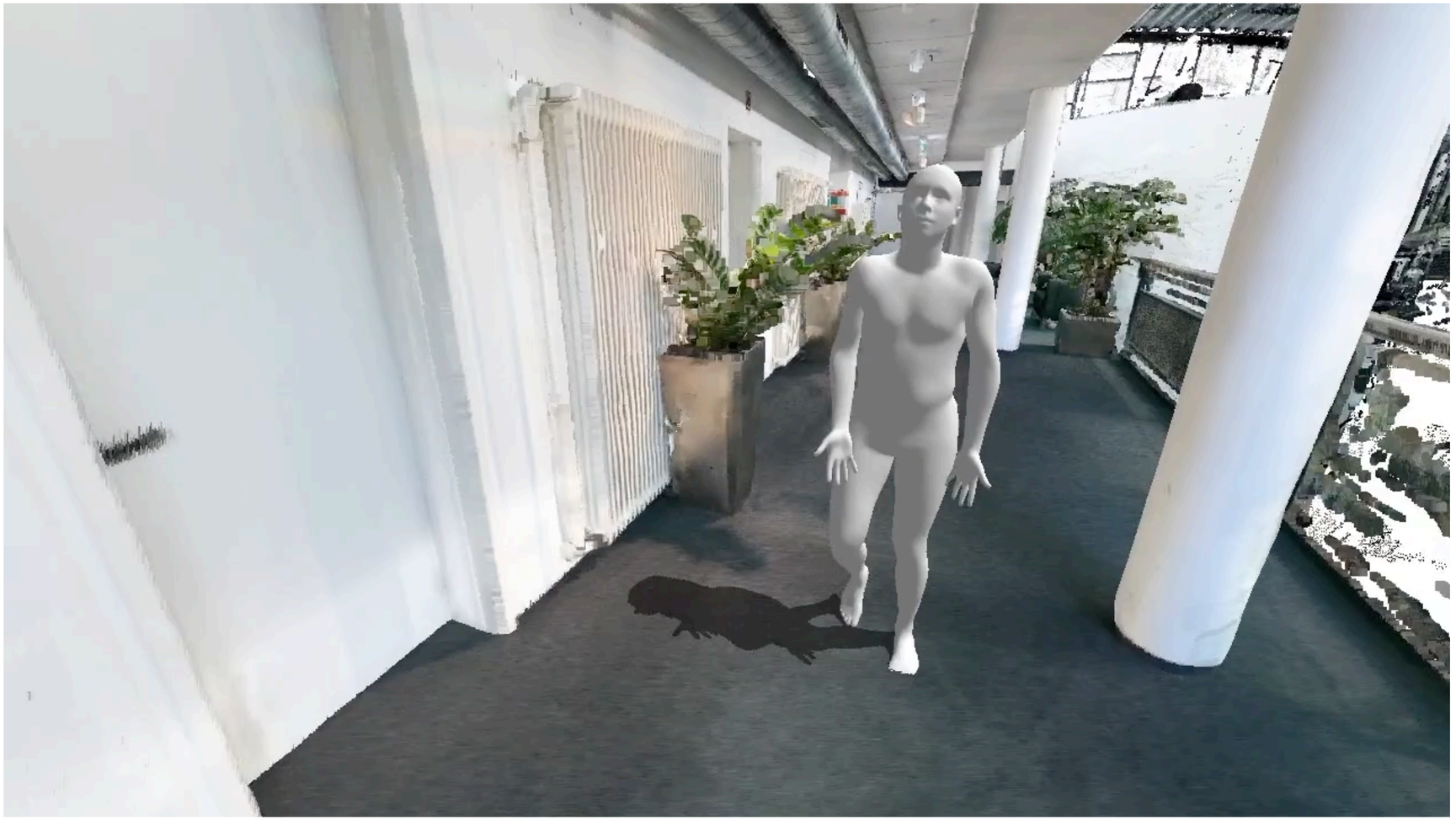


[Middelberg, Sattler, Untzelmann, Kobbelt, Scalable 6-DOF Localization on Mobile Devices, ECCV 2014]



AR navigation in Google Maps

Applications: Performance Capture



slide credit: Vladimir Guzov, Aymen Mir

[Guzov*, Mir*, Sattler, Pons-Moll, Human POSEitioning System (HPS): 3D Human Pose Estimation and Self-localization in Large Scenes from Body-Mounted Sensors, CVPR 2021]

Applications: Visual Localization for Modeling Interactions

slide credit: Vladimir Guzov

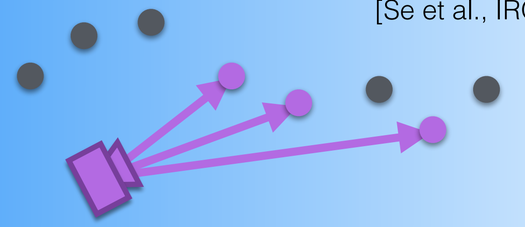
[Guzov, Chibane, Marin, He, Sattler, Pons-Moll, Interaction Replica: Tracking human–object interaction and scene changes from human motion, arXiv 2023]

Visual Localization - A Taxonomy

3D structure-based representation

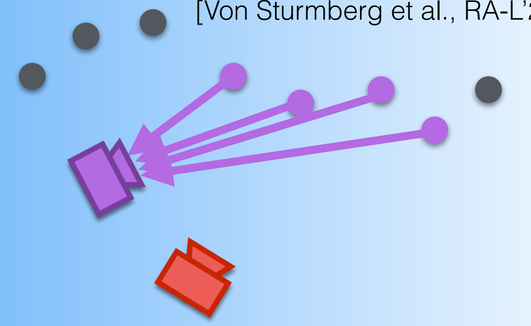
feature-based localization

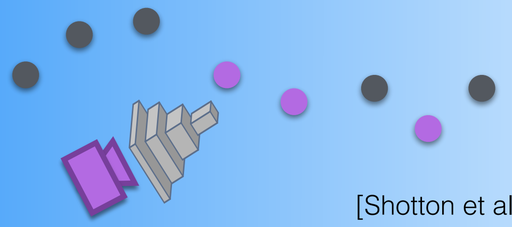
[Se et al., IROS'02]



pose refinement

[Von Sturmberg et al., RA-L'20]





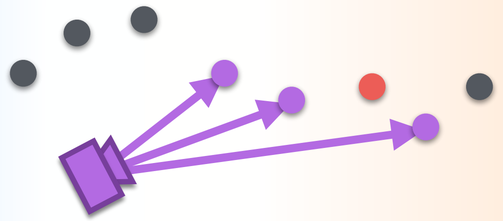
[Shotton et al., CVPR'13]

scene coordinate regression

hybrid

hierarchical localization

[Irschara et al., CVPR'09]



hybrid pose estimation

[Camposeco et al., ICCV'21]

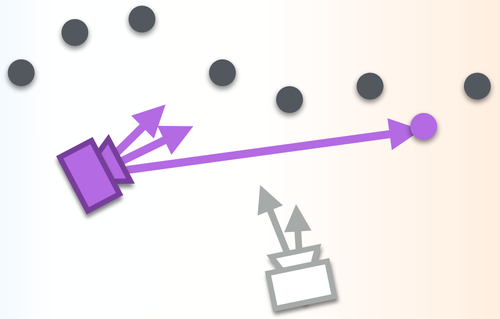


image-based representation

pose triangulation

[Zhang & Kosecka, 3DPVT'06]

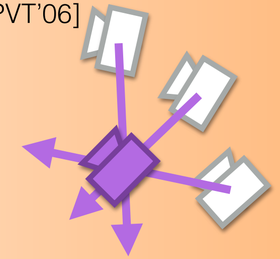
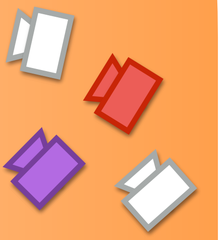
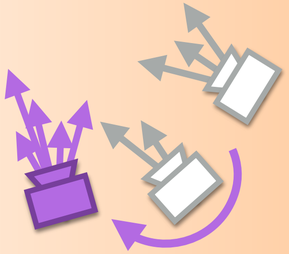


image retrieval



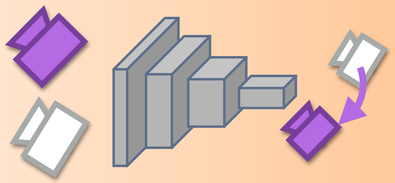


semi-generalized relative pose / homography

[Zheng & Wu, ICCV'15] [Bhayani et al., ICCV'21]

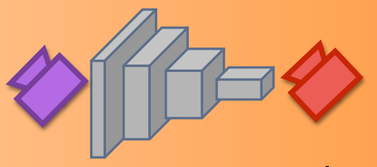
pose interpolation

[Torii et al., ICCVW'11]



relative pose regression

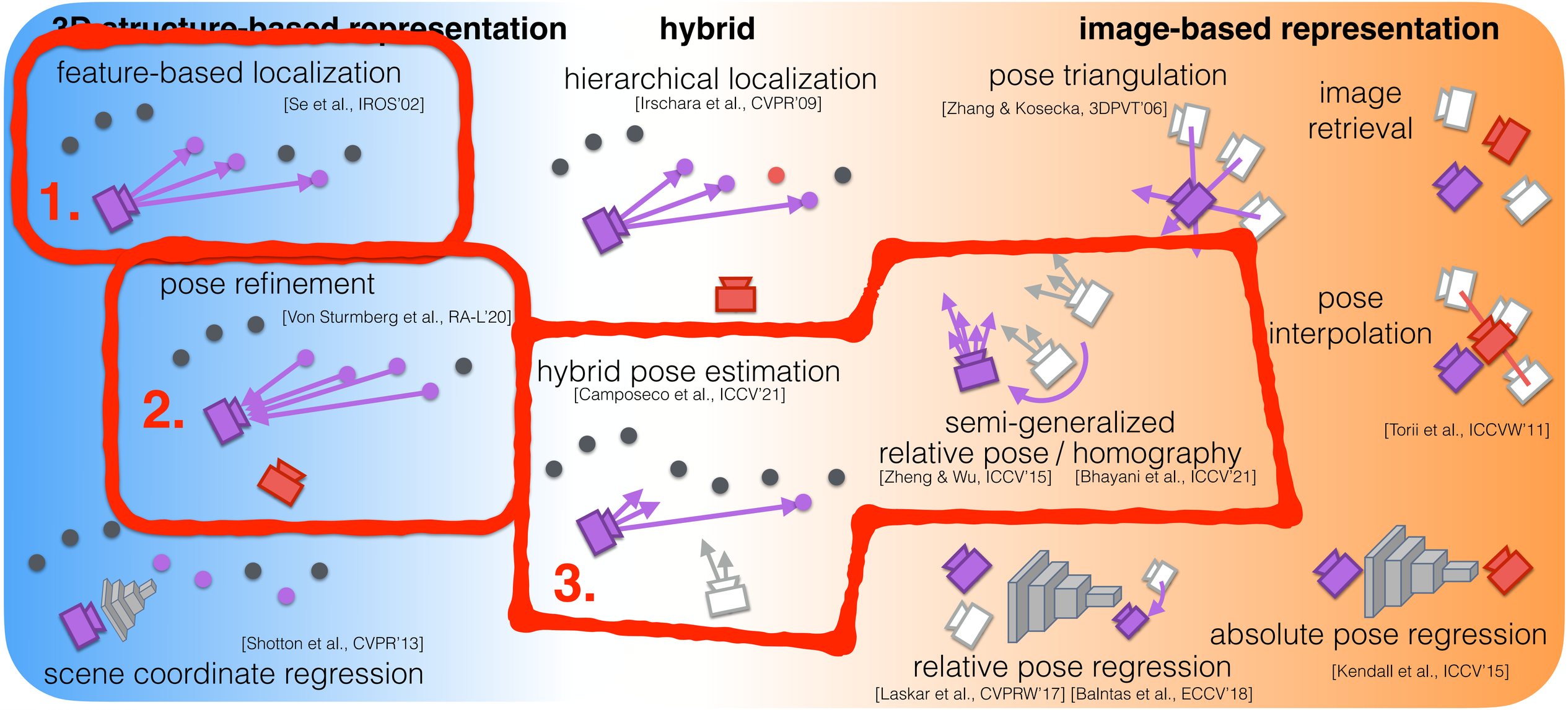
[Laskar et al., CVPRW'17] [Balntas et al., ECCV'18]



absolute pose regression

[Kendall et al., ICCV'15]

Visual Localization - A Taxonomy

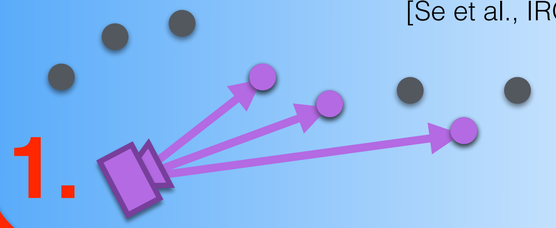


Visual Localization - A Taxonomy

3D structure-based representation

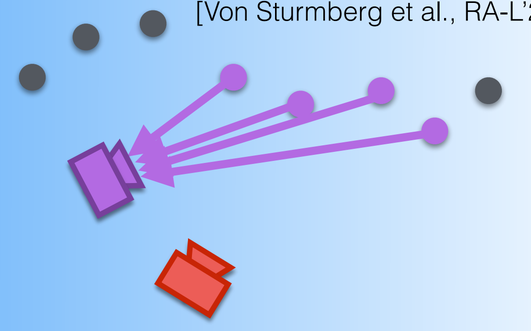
feature-based localization

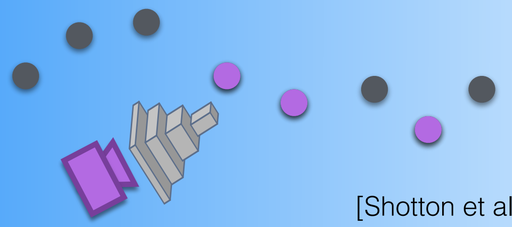
[Se et al., IROS'02]



pose refinement

[Von Sturmberg et al., RA-L'20]





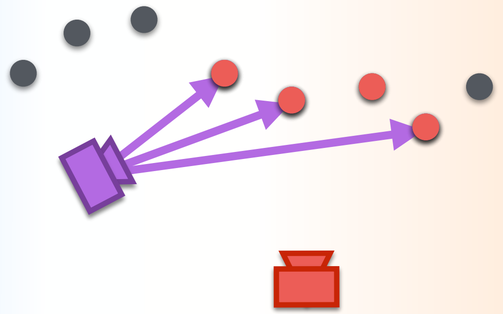
[Shotton et al., CVPR'13]

scene coordinate regression

hybrid

hierarchical localization

[Irschara et al., CVPR'09]



hybrid pose estimation

[Camposeco et al., ICCV'21]

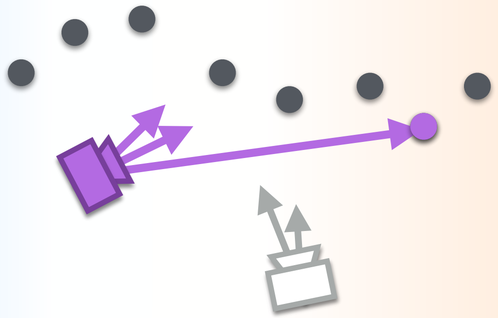


image-based representation

pose triangulation

[Zhang & Kosecka, 3DPVT'06]

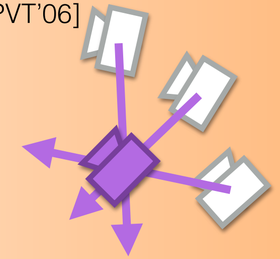
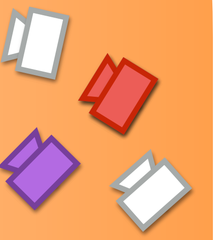
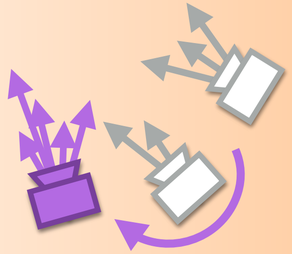


image retrieval

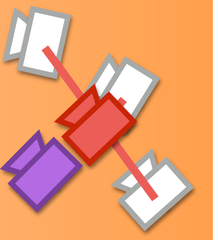




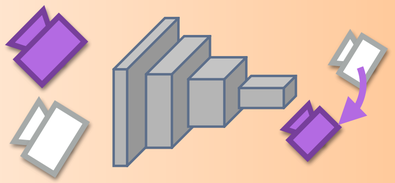
semi-generalized relative pose / homography

[Zheng & Wu, ICCV'15] [Bhayani et al., ICCV'21]

pose interpolation

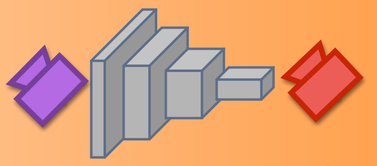


[Torii et al., ICCVW'11]



relative pose regression

[Laskar et al., CVPRW'17] [Balntas et al., ECCV'18]



absolute pose regression

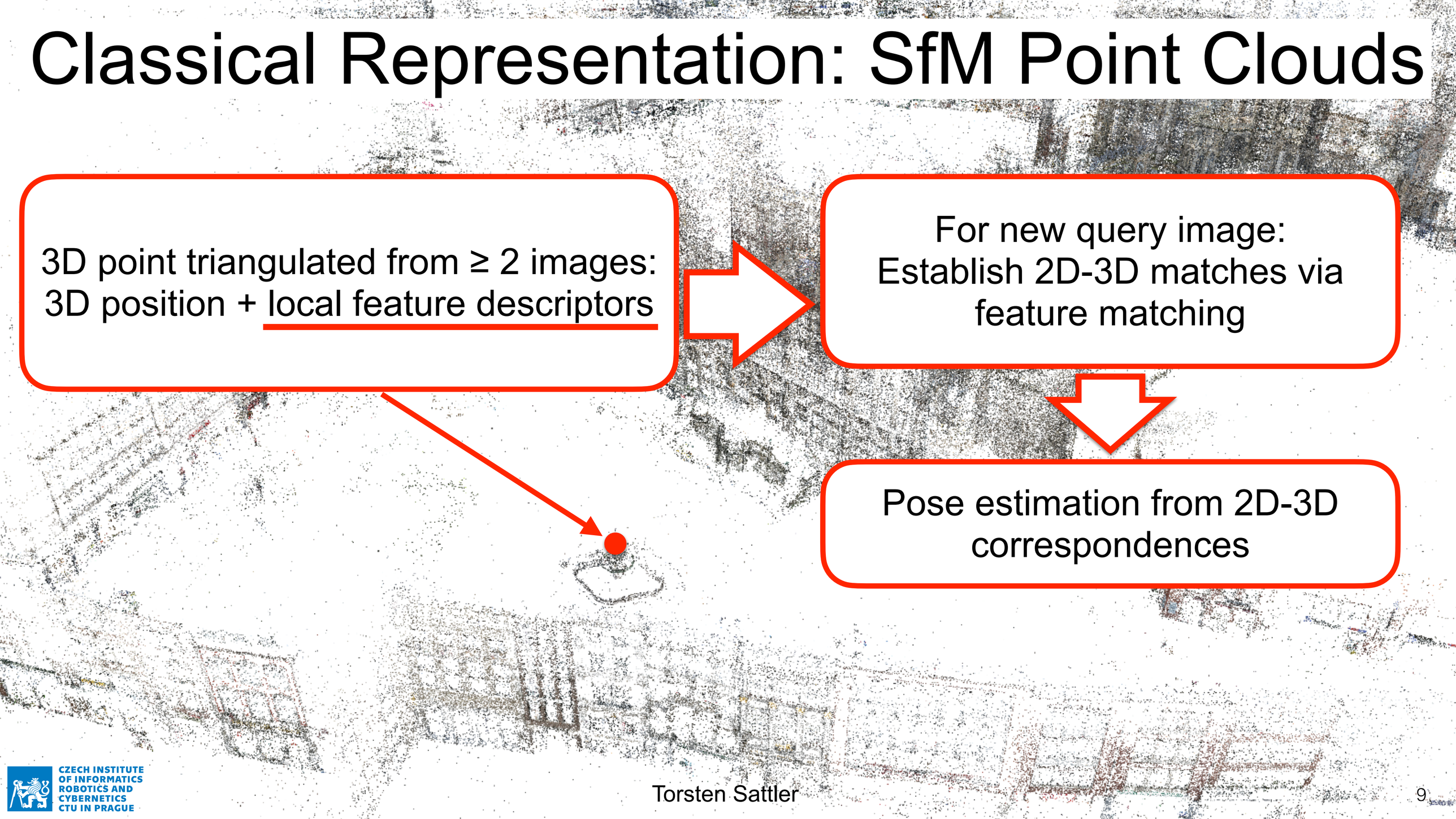
[Kendall et al., ICCV'15]

Classical Representation: SfM Point Clouds

3D point triangulated from ≥ 2 images:
3D position + local feature descriptors

For new query image:
Establish 2D-3D matches via feature matching

Pose estimation from 2D-3D correspondences



Classical Representation: SfM Point Clouds

SIFT (1999)

3D point triangulated from ≥ 2 images:
3D position + local feature descriptors

kd-trees (1975), BoW (2003/4)

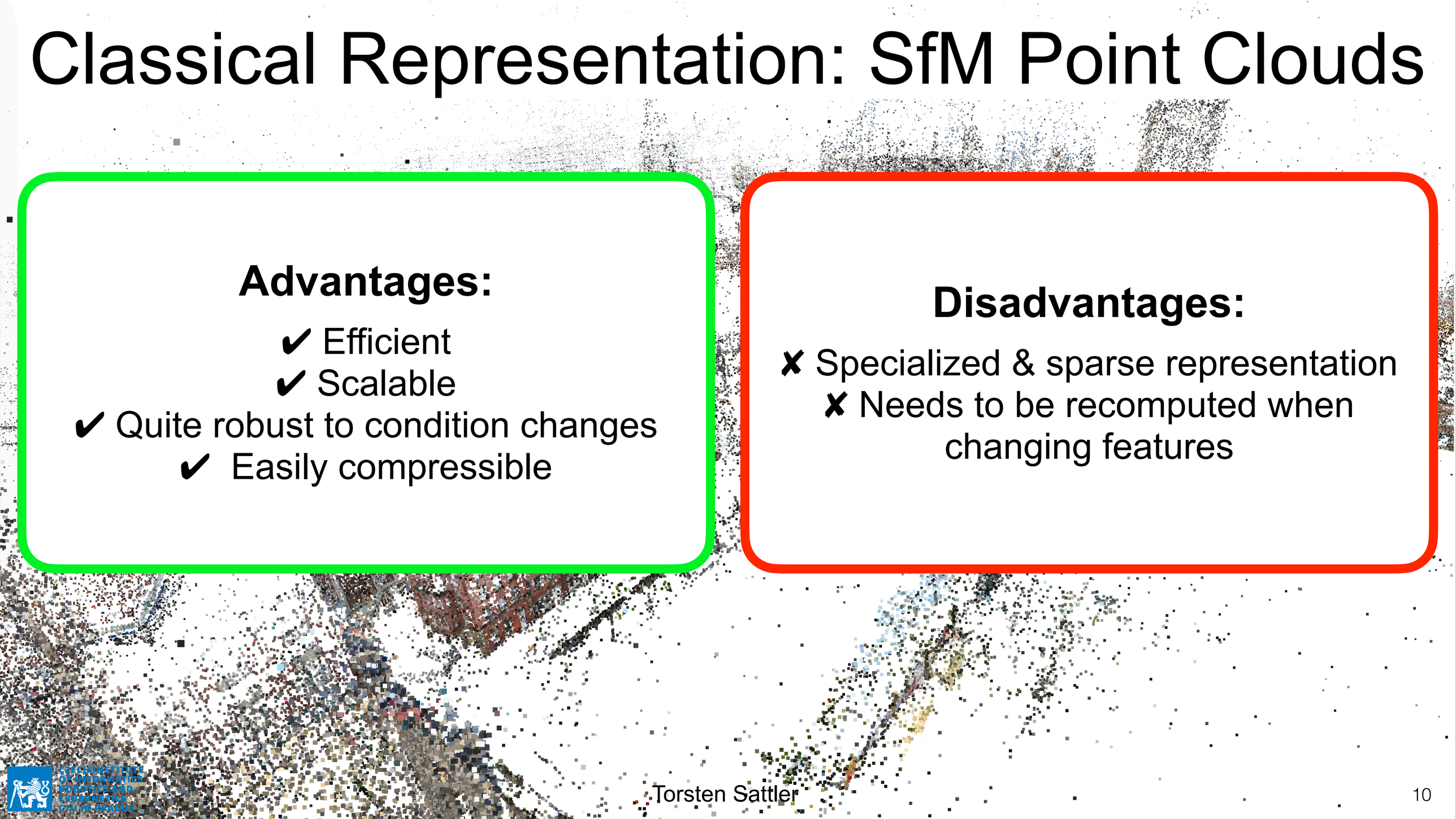
For new query image:
Establish 2D-3D matches via feature matching

Pose estimation from 2D-3D correspondences

P3P (≤1773)
RANSAC (1981)

Classical Representation: SfM Point Clouds

Classical Representation: SfM Point Clouds

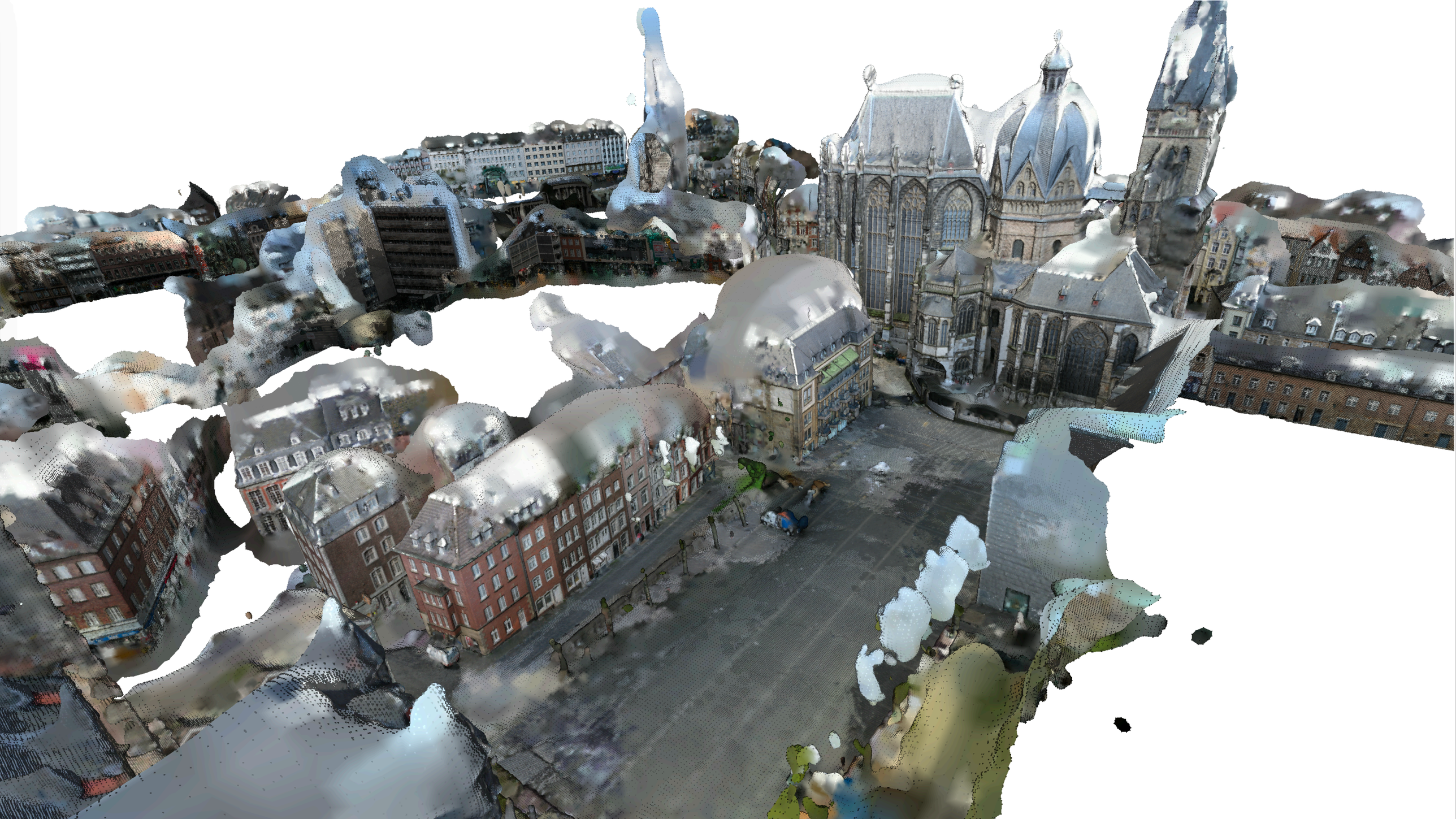


Advantages:

✔ Efficient
✔ Scalable
✔ Quite robust to condition changes
✔ Easily compressible

Disadvantages:

✘ Specialized & sparse representation
✘ Needs to be recomputed when changing features

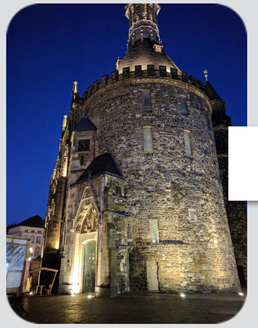


SfM-based vs. Mesh-based Localization

offline (input: posed images)

Extract features from database images → Match features between database images → Triangulate scene structure

online

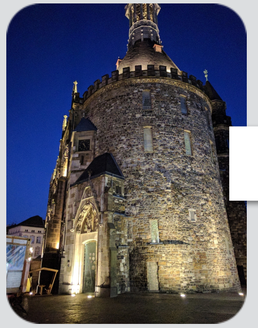


Find N most similar database images → Feature matching between query and top-N database images → 2D-3D matches from associations between database features and 3D points → Pose estimation from 2D-3D matches (P3P-LO-RANSAC)

offline (input: posed images)

3D reconstruction

online



Find N most similar database images → Feature matching between query and top-N database images → 2D-3D matches from depth maps → Pose estimation from 2D-3D matches (P3P-LO-RANSAC)

[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

SfM-based vs. Mesh-based Localization

offline (input: posed images)

Extract features from database images

Match features between database images

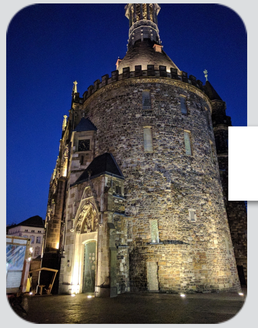
Triangulate

offline (input: posed images)

nstruction

Any renderable representation works, including NeRFs and Gaussian Splatting

online



Find N most similar database images

Feature matching between query and top-N database images

2D-3D matches from associations between database features and 3D points

Pose estimation from 2D-3D matches (P3P-LO-RANSAC)

similar database images

Feature matching between query and top-N **re-rendered** database images

2D-3D matches from depth maps

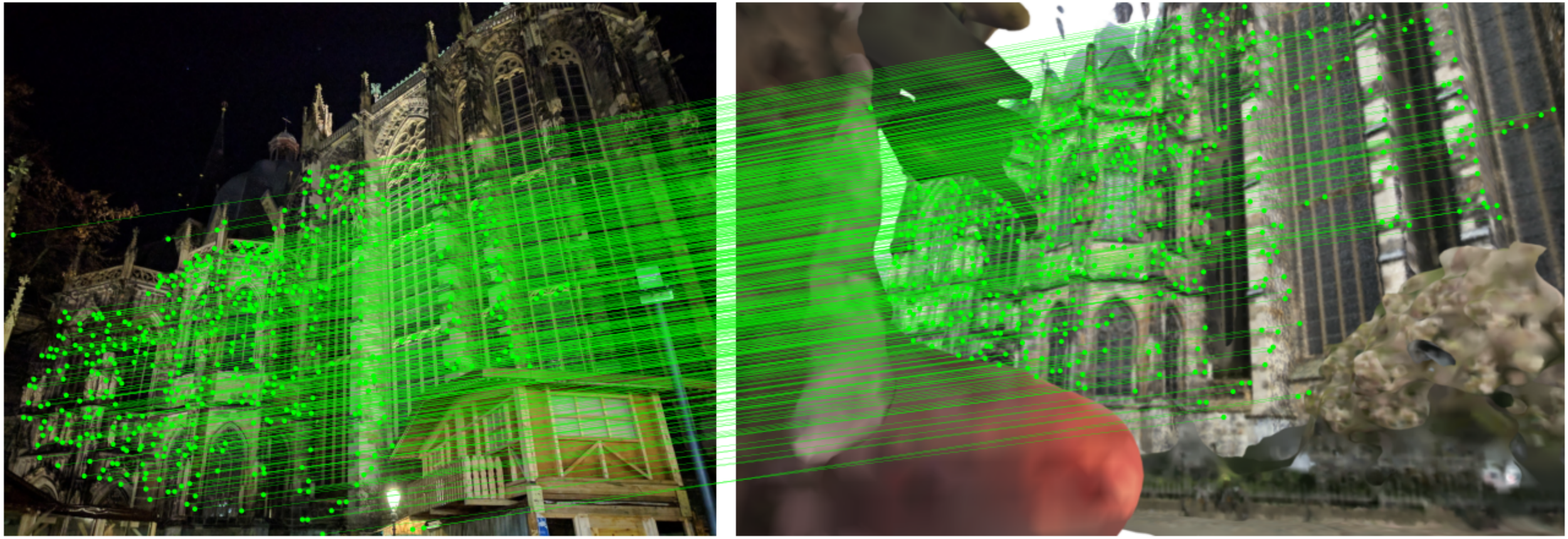
Pose estimation from 2D-3D matches (P3P-LO-RANSAC)

[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

Matching Against Rendered Images

[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

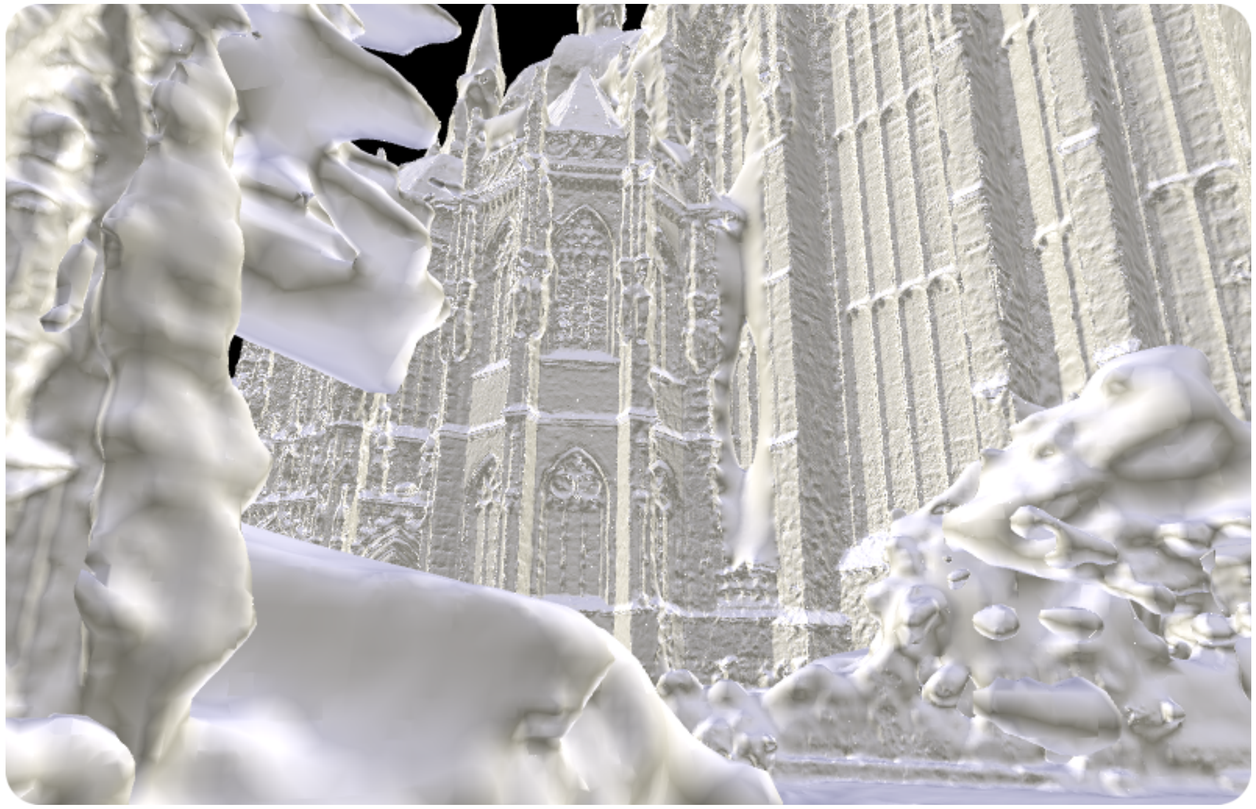
Matching Against Rendered Images



ALIKED features [Zhao et al., IEEE TIM 2023] with LightGlue [Lindenberger et al., ICCV 2023]
matcher not trained on renderings

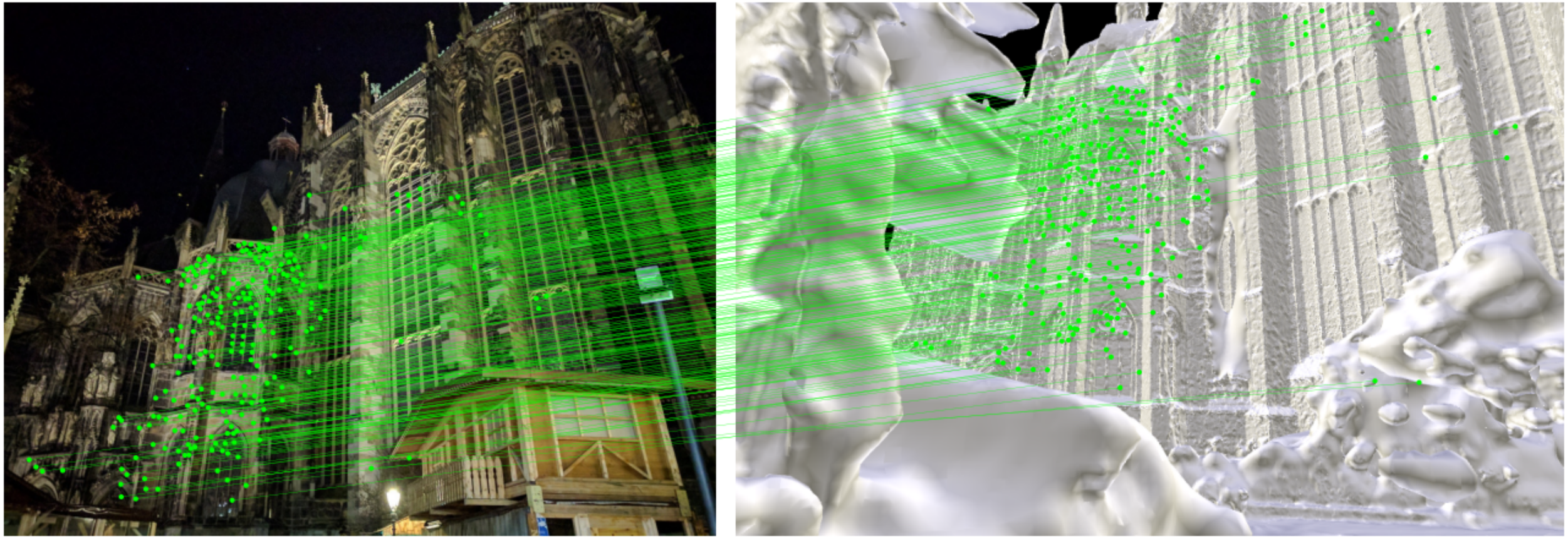
[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

Matching Against Raw Geometry



[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

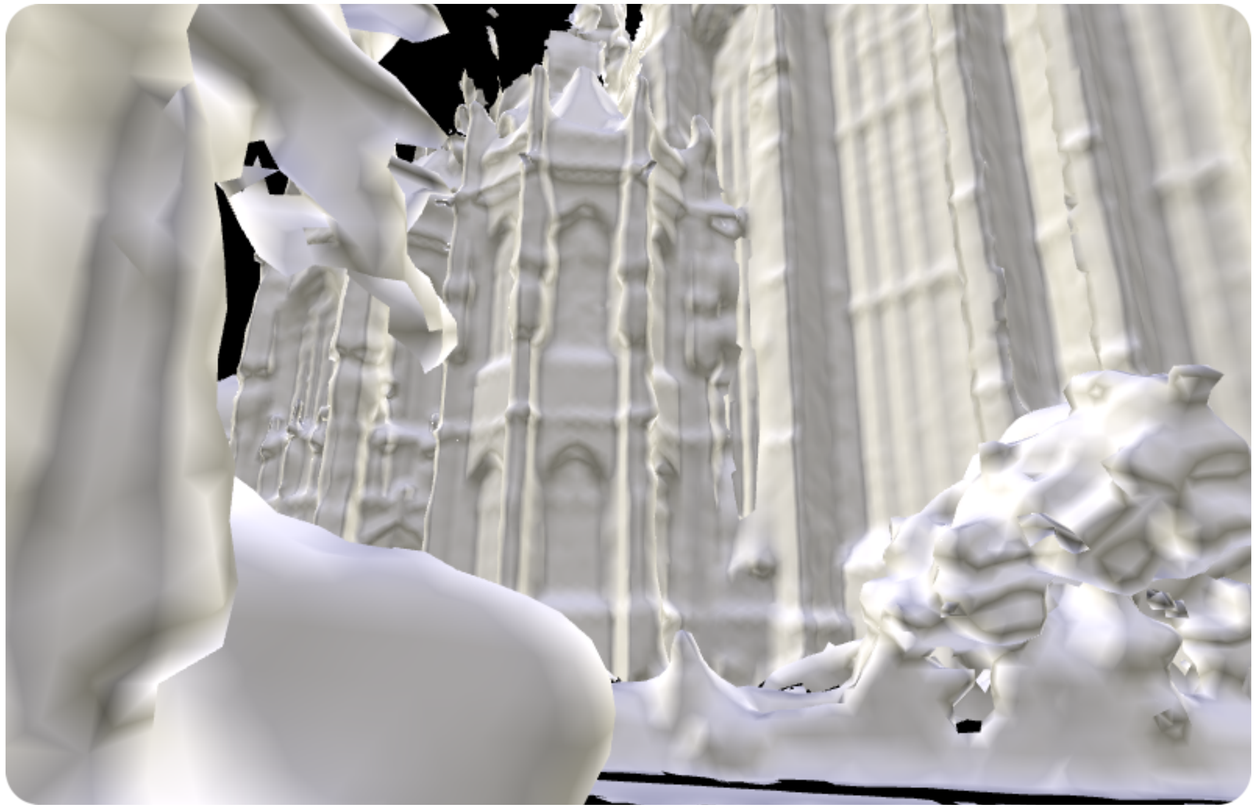
Matching Against Raw Geometry



ALIKED features [Zhao et al., IEEE TIM 2023] with LightGlue [Lindenberger et al., ICCV 2023]
matcher not trained on renderings

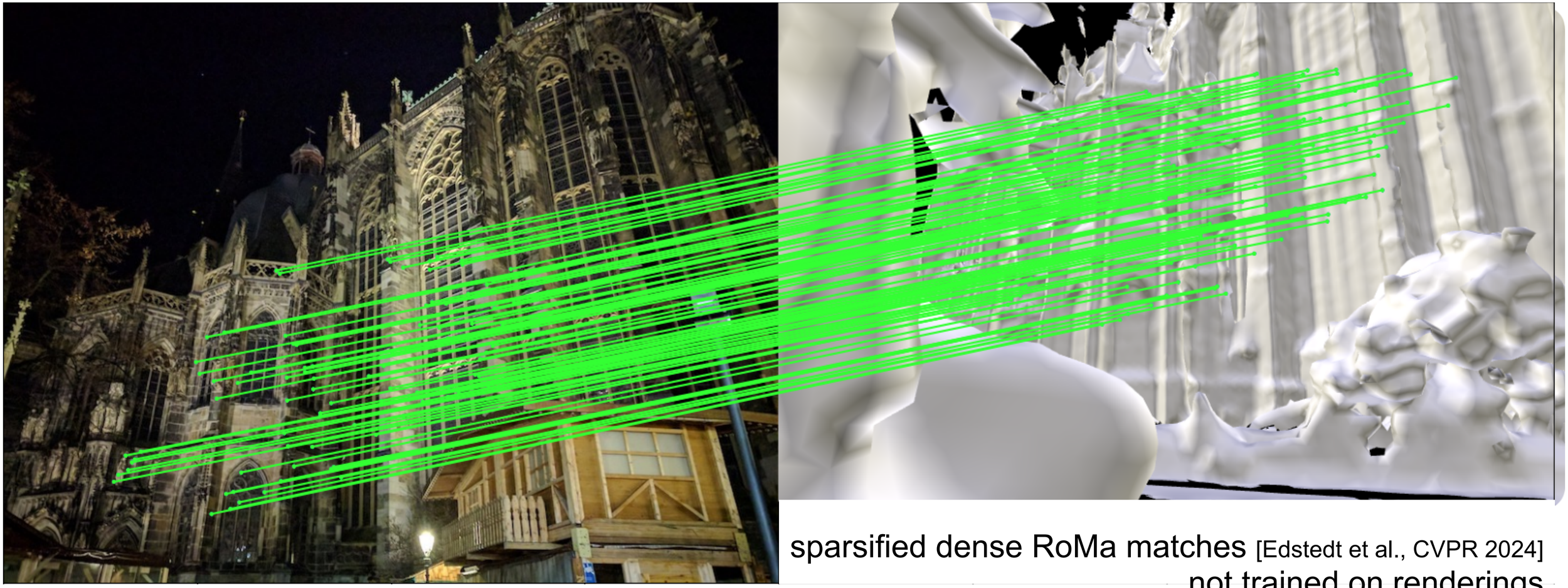
[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

Matching Against Raw Geometry



[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

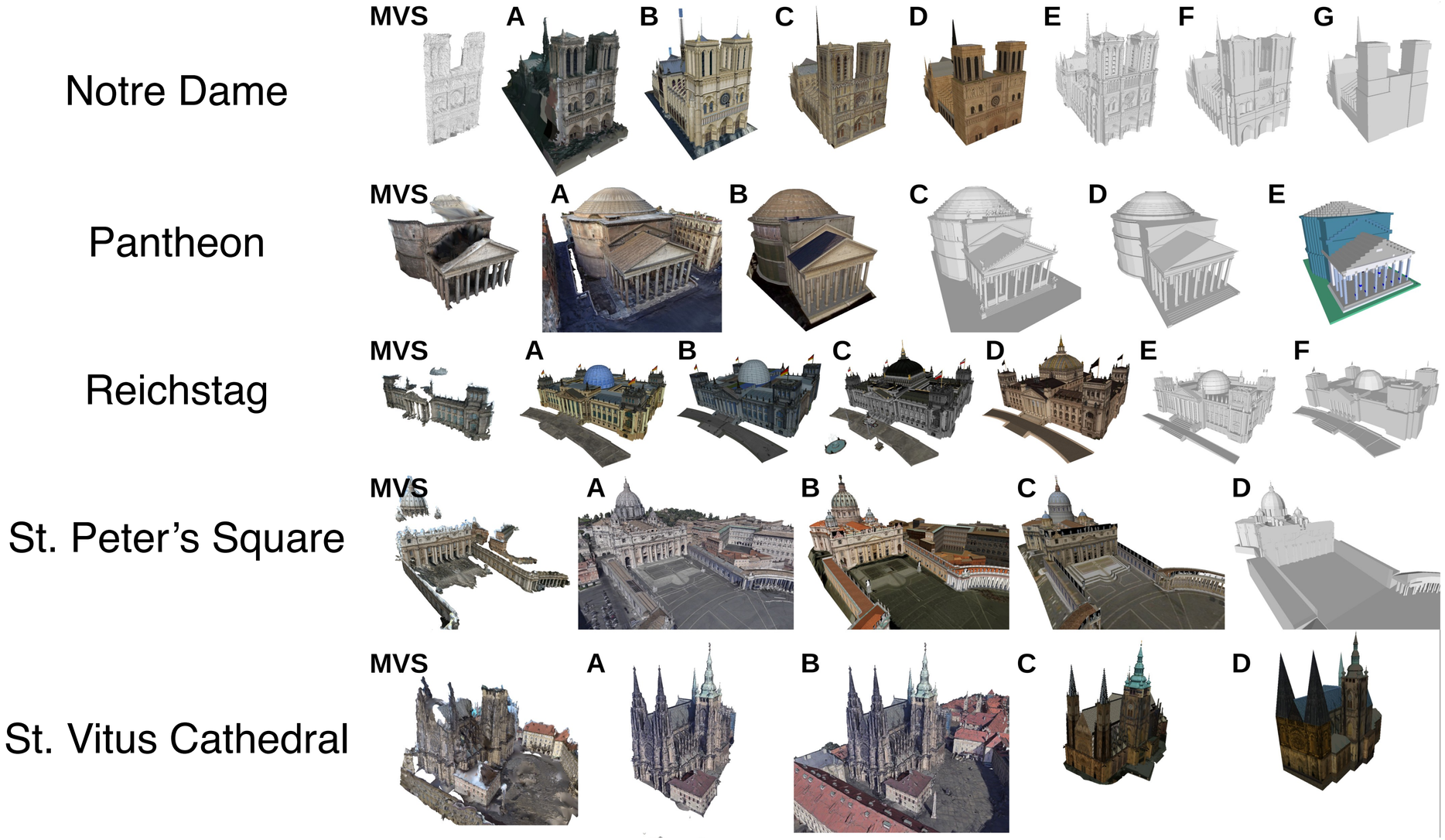
Matching Against Raw Geometry



sparsified dense RoMa matches [Edstedt et al., CVPR 2024]
not trained on renderings

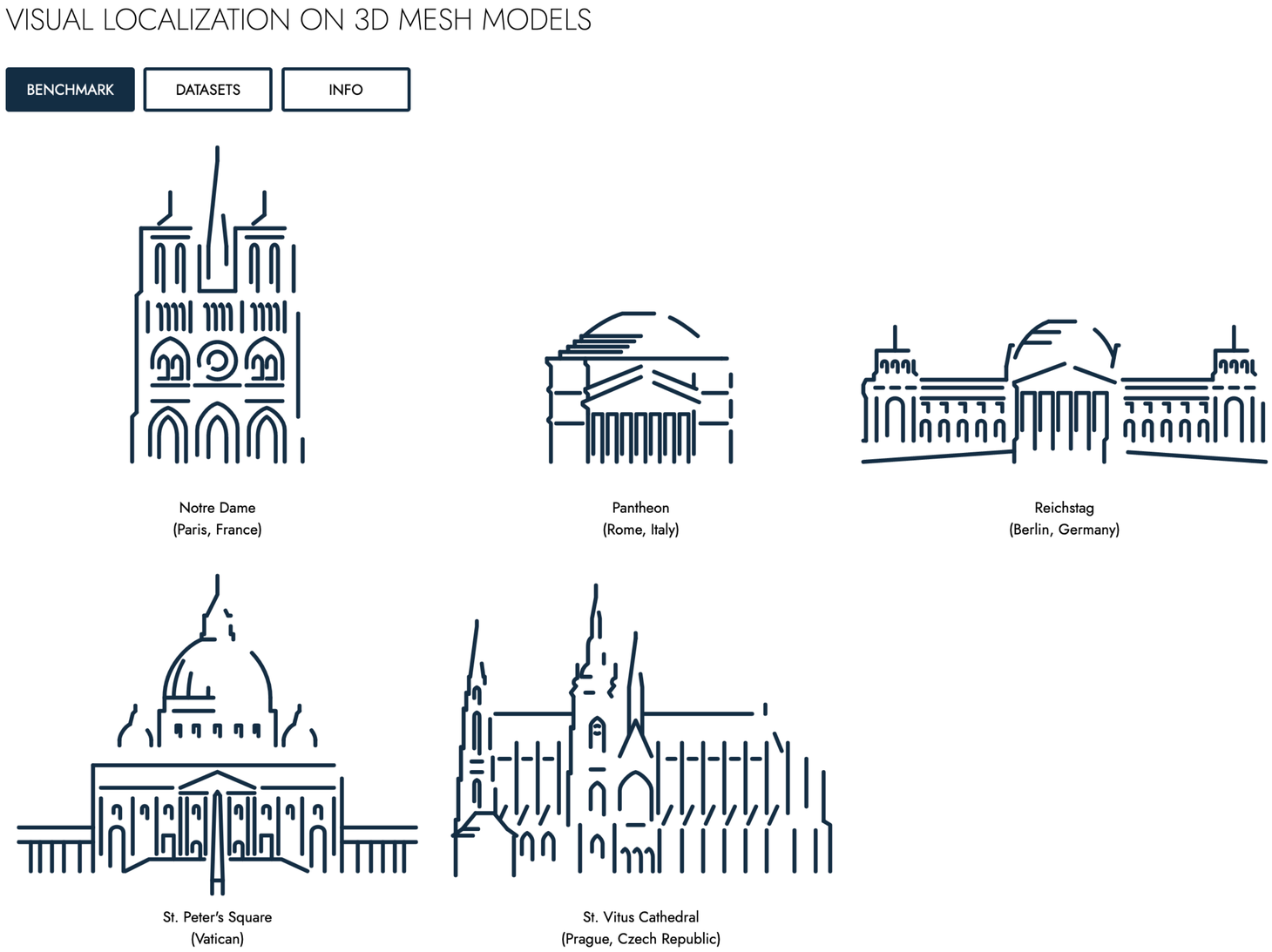
[Panek, Kukelova, Sattler, MeshLoc: Mesh-Based Visual Localization, ECCV 2022]

Visual Localization using Internet Models



[Panek, Kukelova, Sattler, Visual Localization using Imperfect 3D Models from the Internet, CVPR 2023]

Benchmarking Visual Localization using Internet Models



v-pnk.github.io/cadloc

slide credit: Vojtech Panek

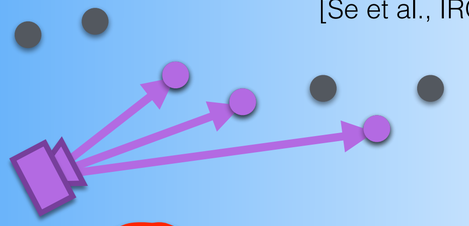
[Panek, Kukelova, Sattler, Visual Localization using Imperfect 3D Models from the Internet, CVPR 2023]

Visual Localization - A Taxonomy

3D structure-based representation

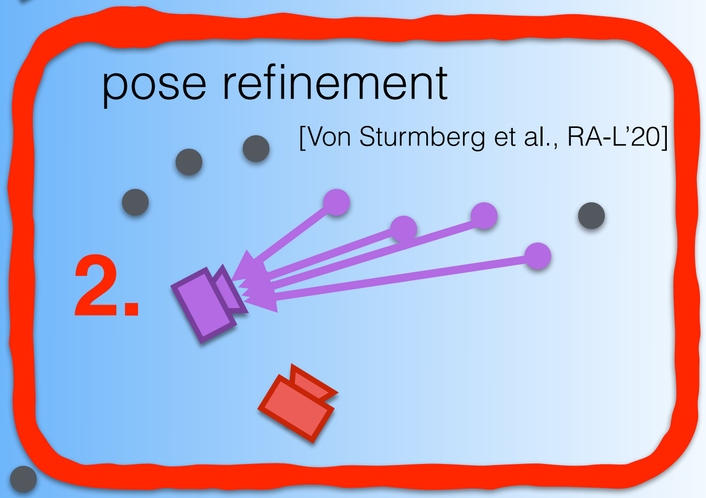
feature-based localization

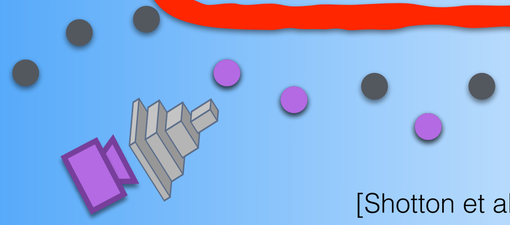
[Se et al., IROS'02]



pose refinement

[Von Sturmberg et al., RA-L'20]





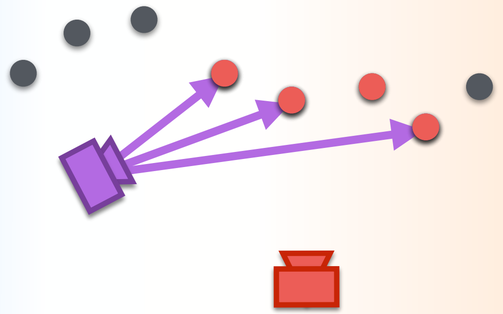
[Shotton et al., CVPR'13]

scene coordinate regression

hybrid

hierarchical localization

[Irschara et al., CVPR'09]



hybrid pose estimation

[Camposeco et al., ICCV'21]

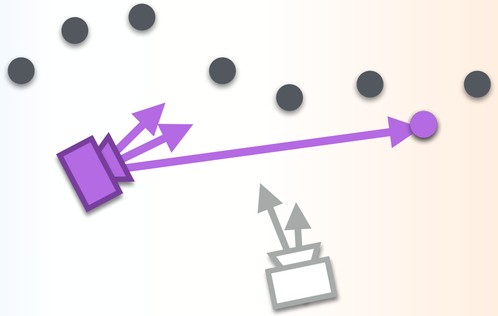


image-based representation

pose triangulation

[Zhang & Kosecka, 3DPVT'06]

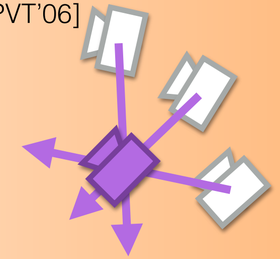
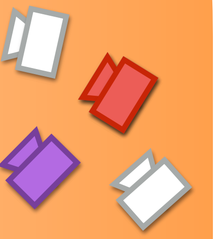
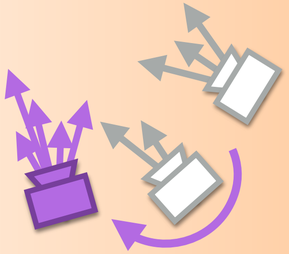


image retrieval

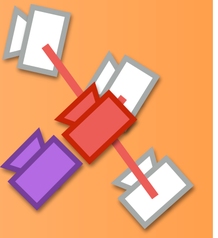




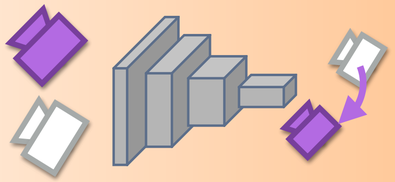
semi-generalized relative pose / homography

[Zheng & Wu, ICCV'15] [Bhayani et al., ICCV'21]

pose interpolation

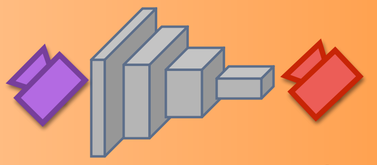


[Torii et al., ICCVW'11]



relative pose regression

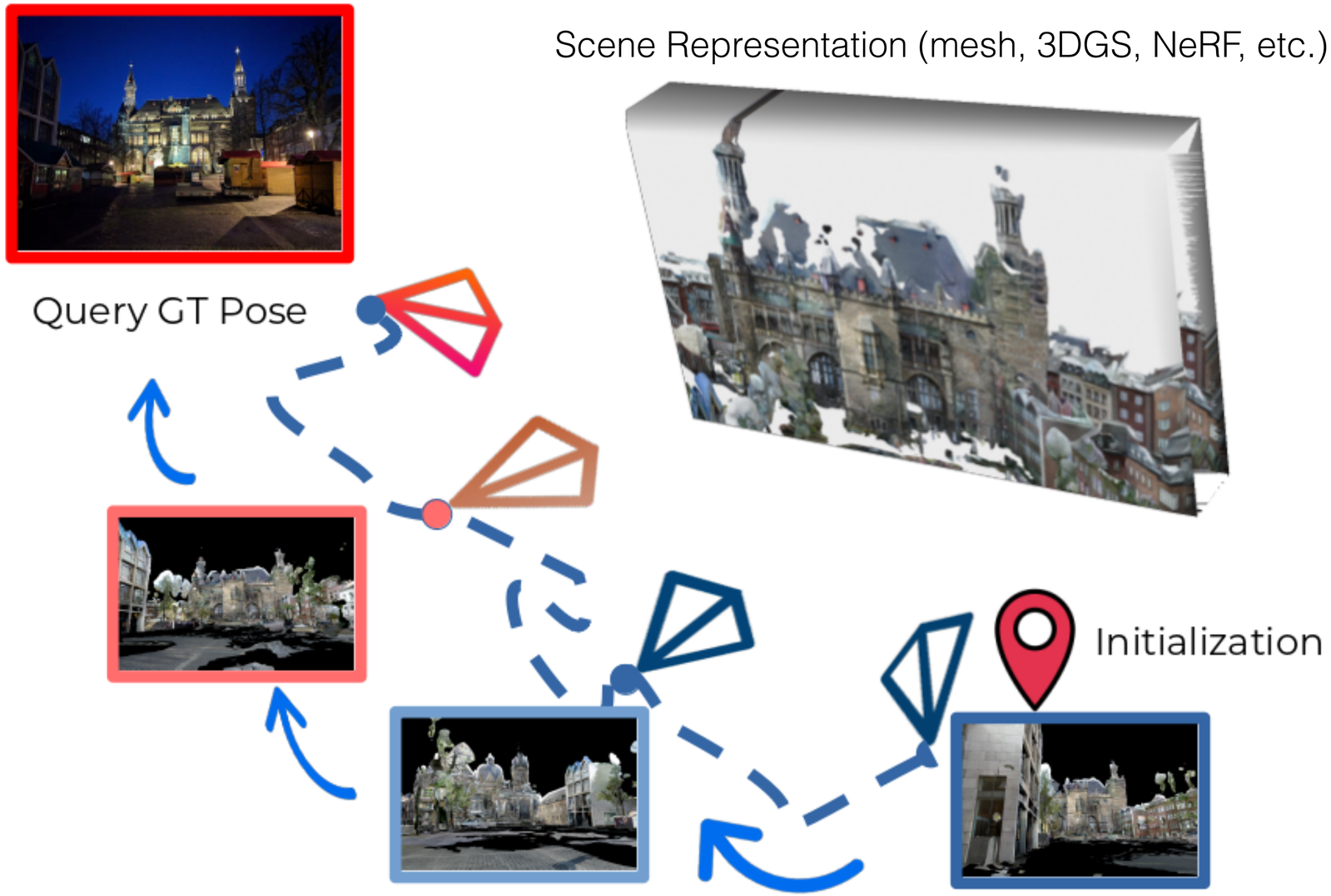
[Laskar et al., CVPRW'17] [Balntas et al., ECCV'18]



absolute pose regression

[Kendall et al., ICCV'15]

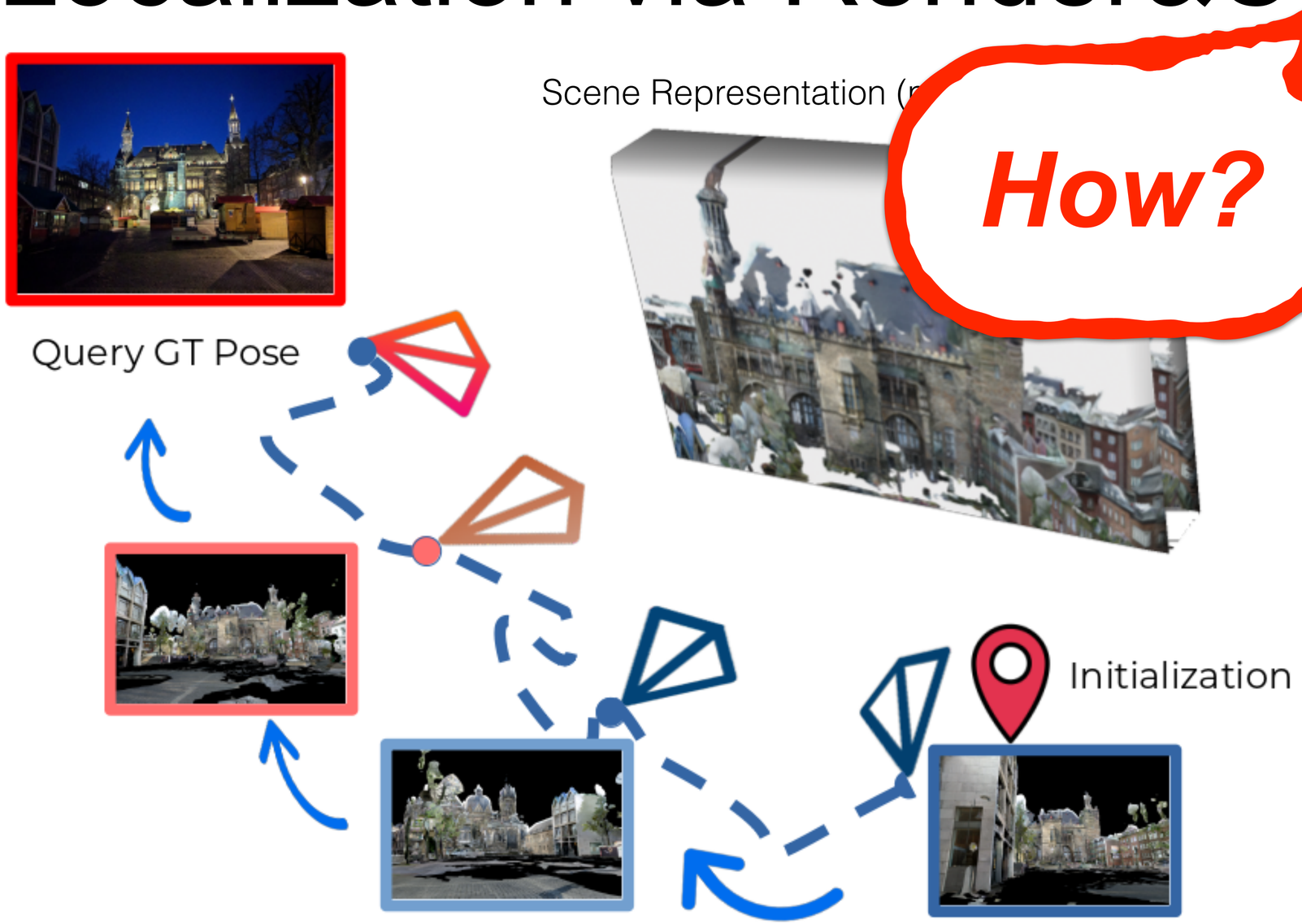
Visual Localization via Render&Compare



slide credit: Gabriele Trivigno

[Gabriele Trivigno, Carlo Masone, Barbara Caputo, Sattler, The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement, CVPR 2024] (highlight)

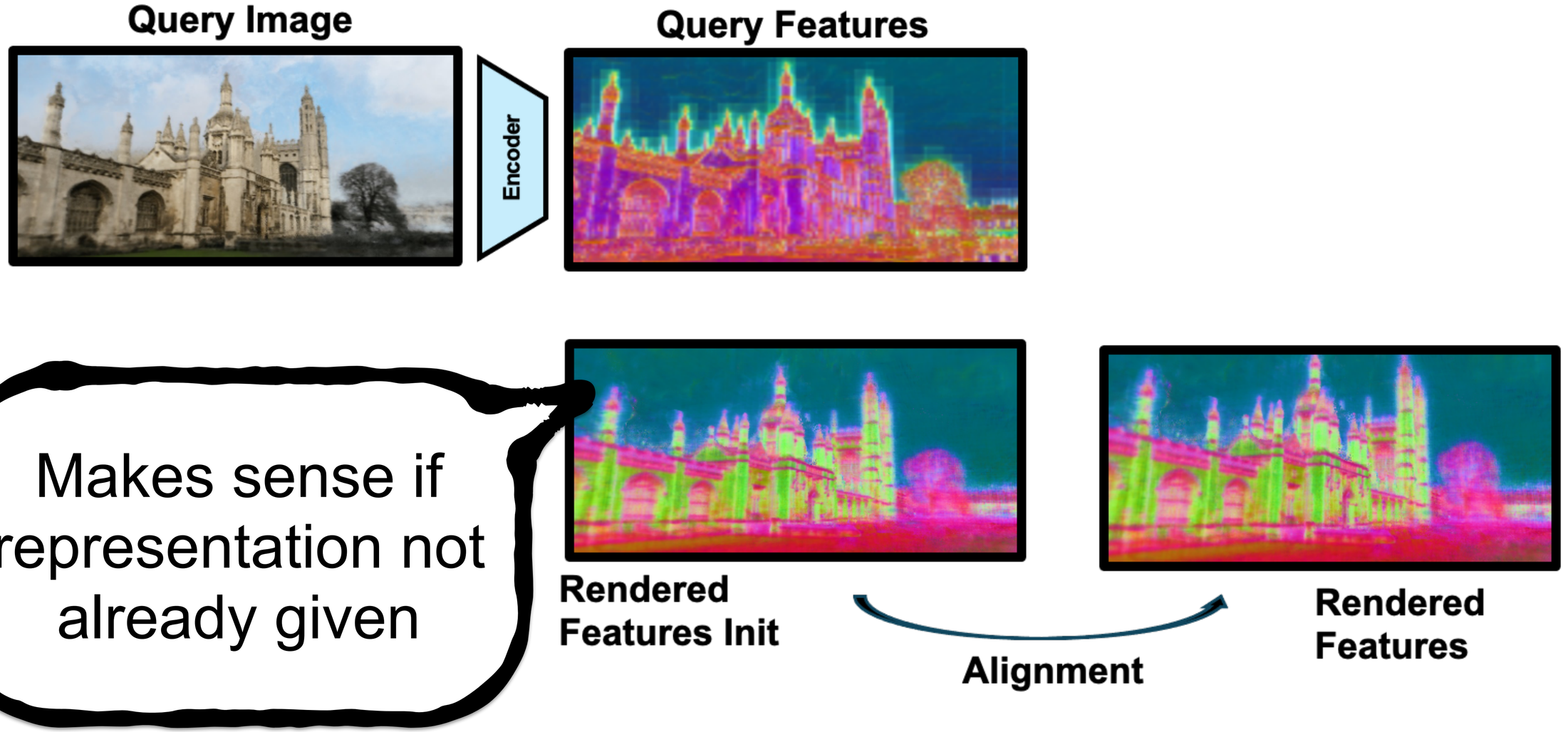
Visual Localization via Render&Compare



slide credit: Gabriele Trivigno

[Gabriele Trivigno, Carlo Masone, Barbara Caputo, Sattler, The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement, CVPR 2024] (highlight)

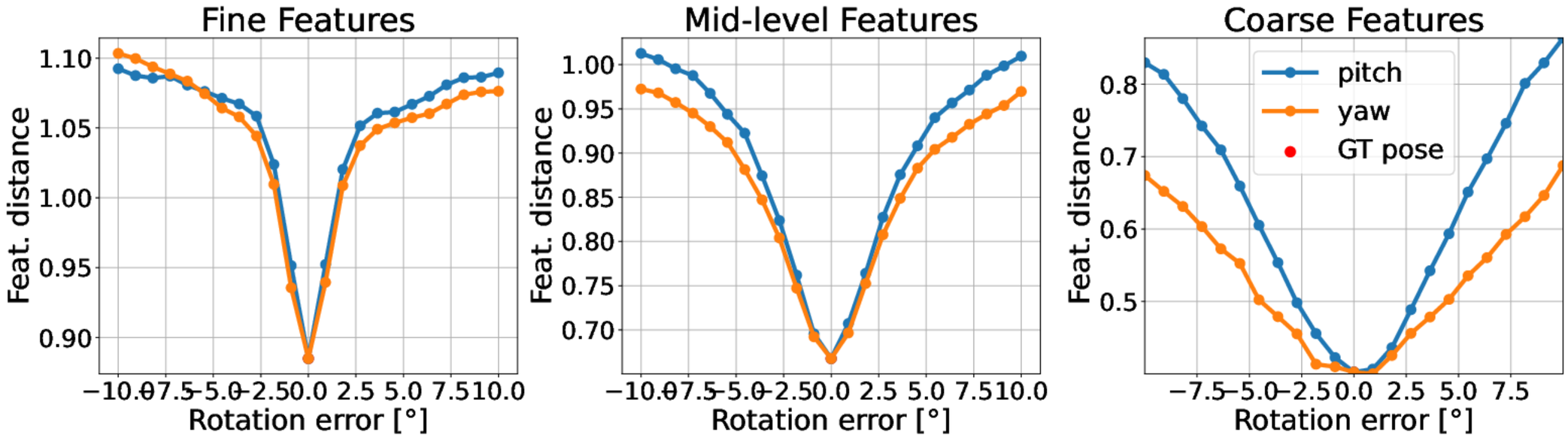
Jointly Training Representation and Features



[Maxime Pietrantoni, Gabriela Csurka, Martin Humenberger, Sattler, Self-supervised learning of Neural implicit Feature Fields for Camera Pose Refinement, 3DV 2024]

The Unreasonable Effectiveness of Pre-Trained Features

- Dense deep features are known to be good estimators of **perceptual similarity**
- This property can be exploited to measure **pose similarity** as well
- **Feature depth** is correlated with the **sensitivity** → hierarchical scheme

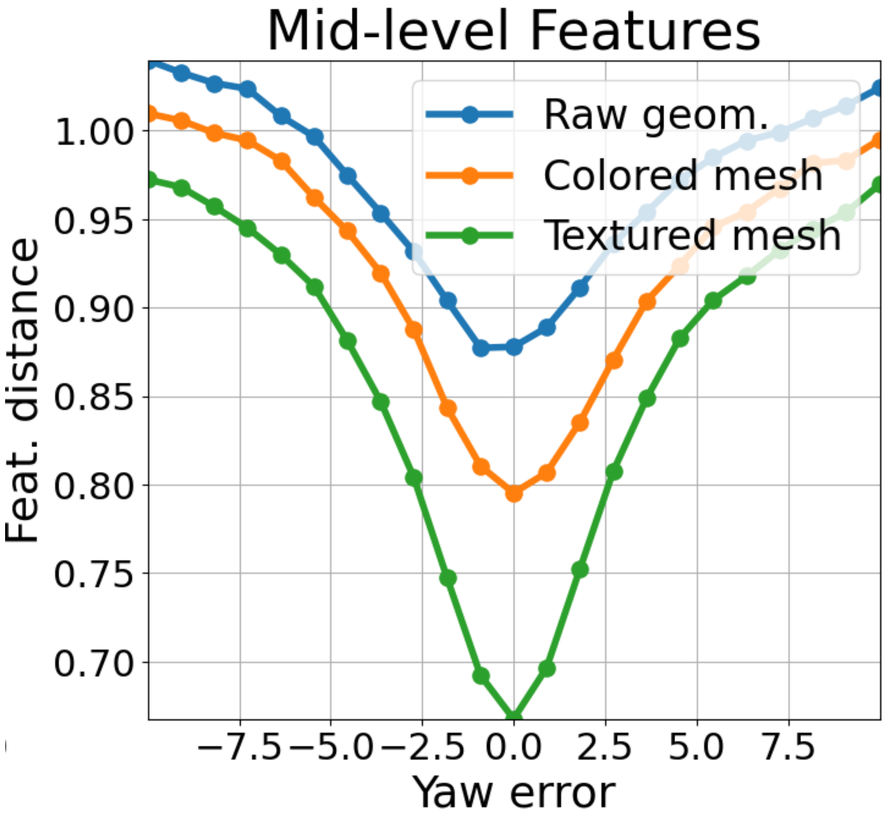


slide credit: Gabriele Trivigno

[Gabriele Trivigno, Carlo Masone, Barbara Caputo, Sattler, The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement, CVPR 2024]

The Unreasonable Effectiveness of Pre-Trained Features

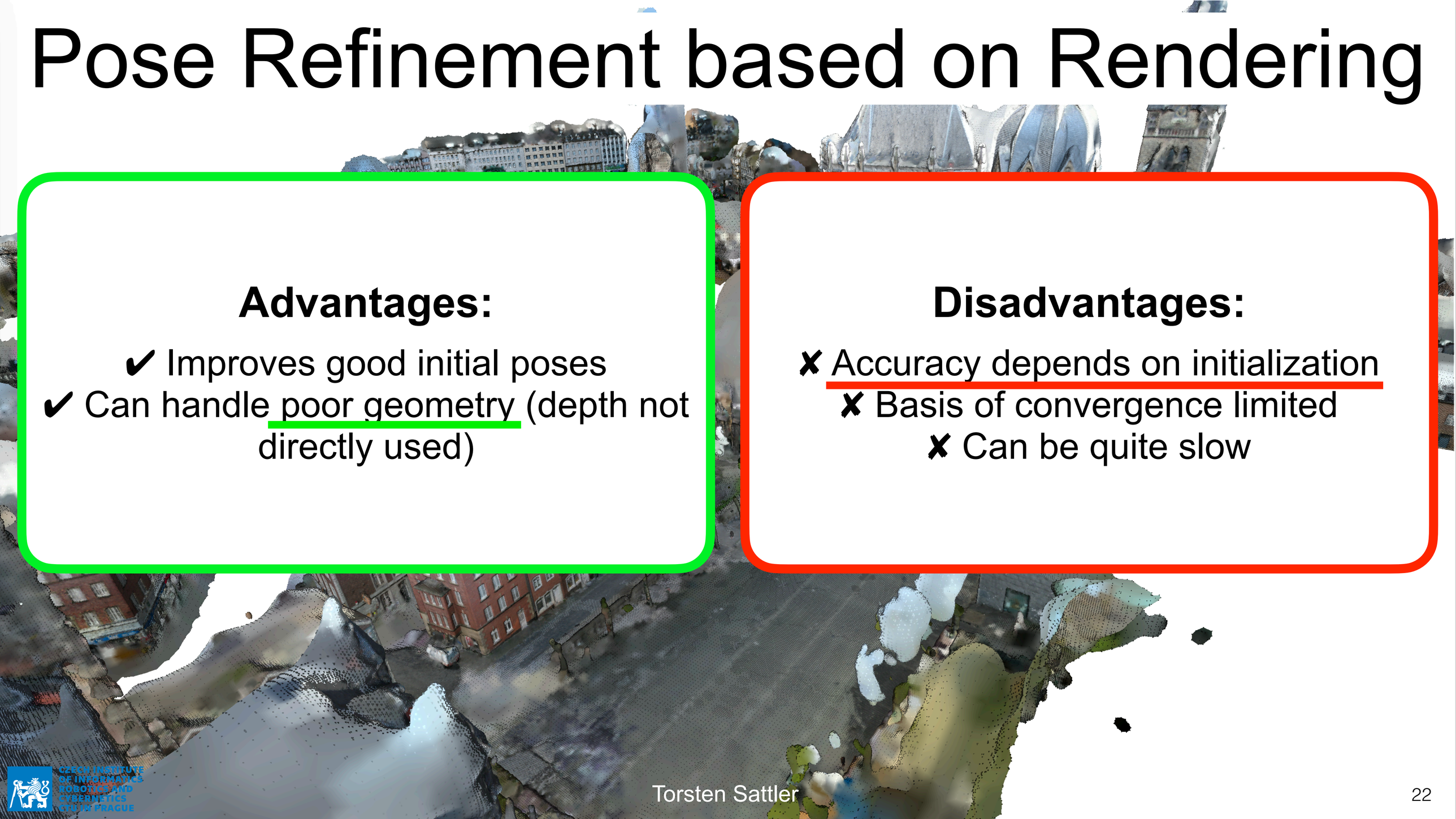
Dense deep features are quite robust to domain changes



slide credit: Gabriele Trivigno

[Gabriele Trivigno, Carlo Masone, Barbara Caputo, Sattler, The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement, CVPR 2024]

Pose Refinement based on Rendering



Advantages:

✔ Improves good initial poses
✔ Can handle poor geometry (depth not directly used)

Disadvantages:

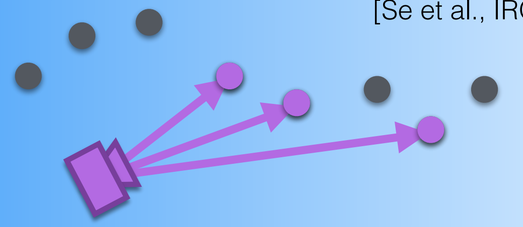
✘ Accuracy depends on initialization
✘ Basis of convergence limited
✘ Can be quite slow

Visual Localization - A Taxonomy

3D structure-based representation

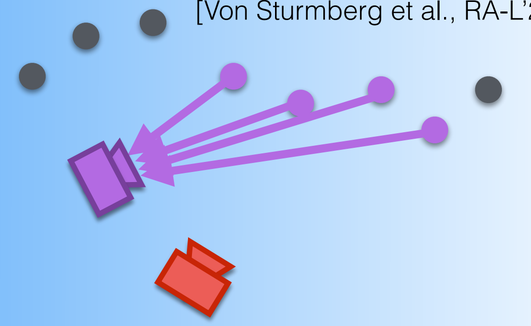
feature-based localization

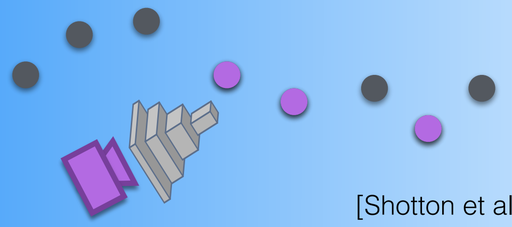
[Se et al., IROS'02]



pose refinement

[Von Sturmberg et al., RA-L'20]





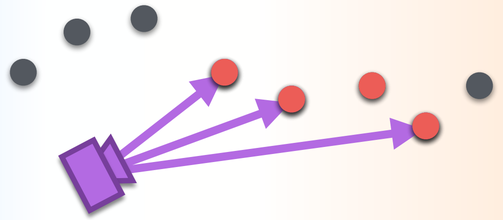
[Shotton et al., CVPR'13]

scene coordinate regression

hybrid

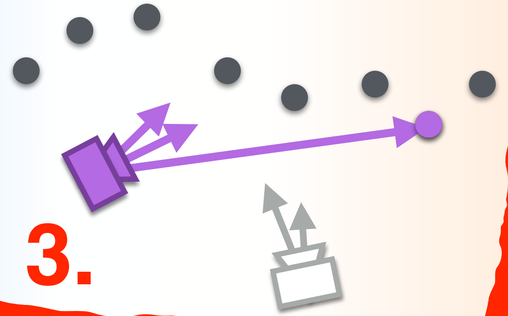
hierarchical localization

[Irschara et al., CVPR'09]



hybrid pose estimation

[Camposeco et al., ICCV'21]



3.

image-based representation

pose triangulation

[Zhang & Kosecka, 3DPVT'06]

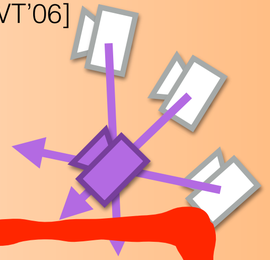
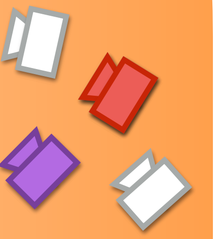
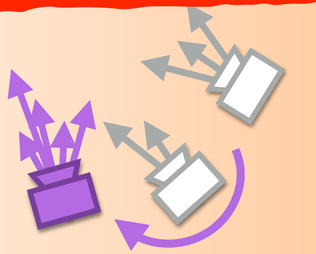


image retrieval

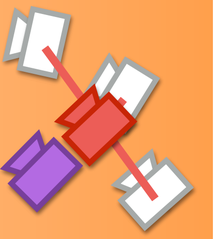




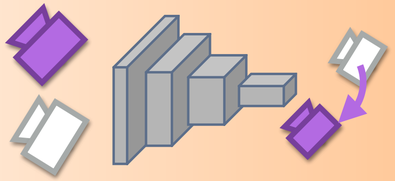
semi-generalized relative pose / homography

[Zheng & Wu, ICCV'15] [Bhayani et al., ICCV'21]

pose interpolation

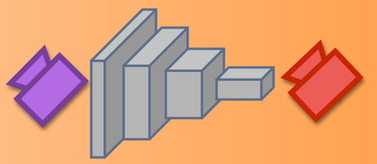


[Torii et al., ICCVW'11]



relative pose regression

[Laskar et al., CVPRW'17] [Balntas et al., ECCV'18]



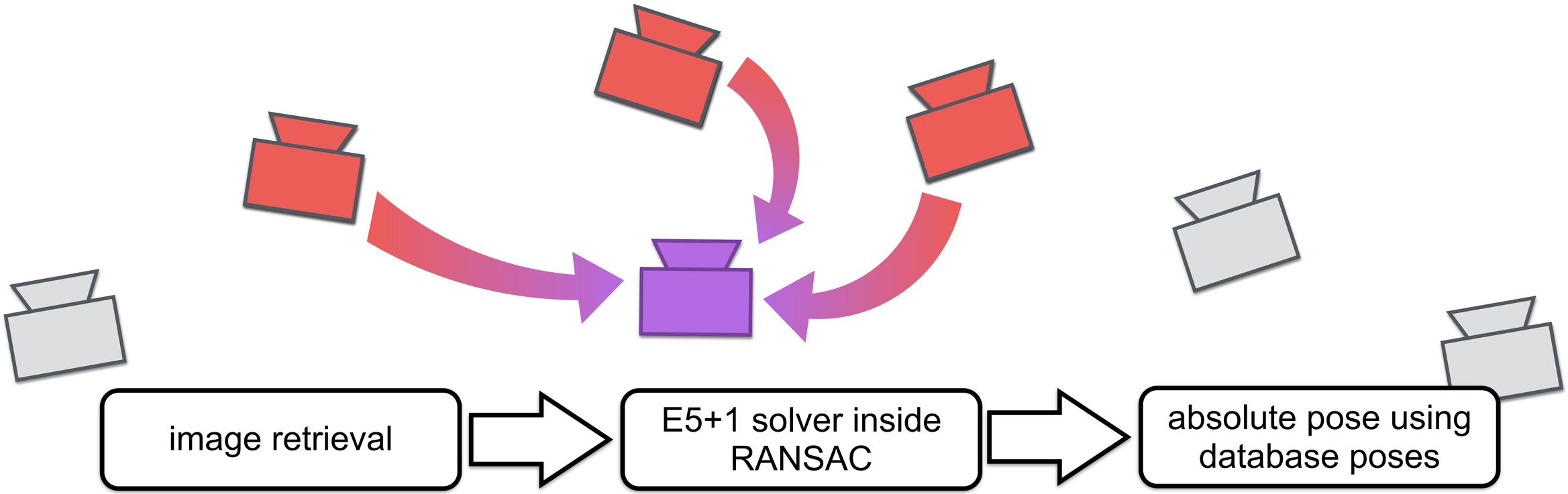
absolute pose regression

[Kendall et al., ICCV'15]

Structure-Less Visual Localization

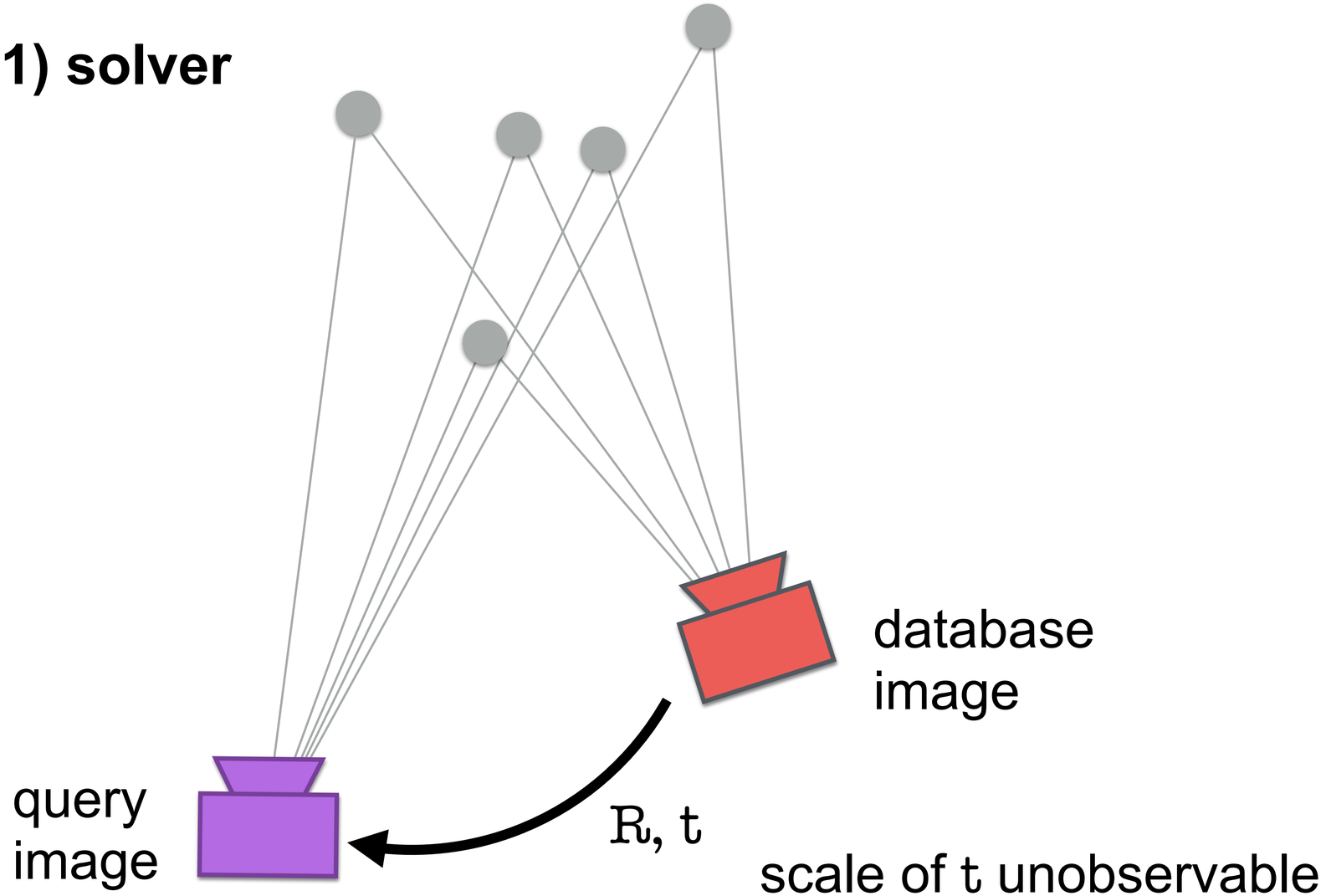
Scene representation: images with known poses, **no 3D points**

- ✓ Easy to update: just add / remove image and pose from database
- ✓ Extract features on the fly, easy to use new feature type



Semi-Generalized Relative Pose Estimation

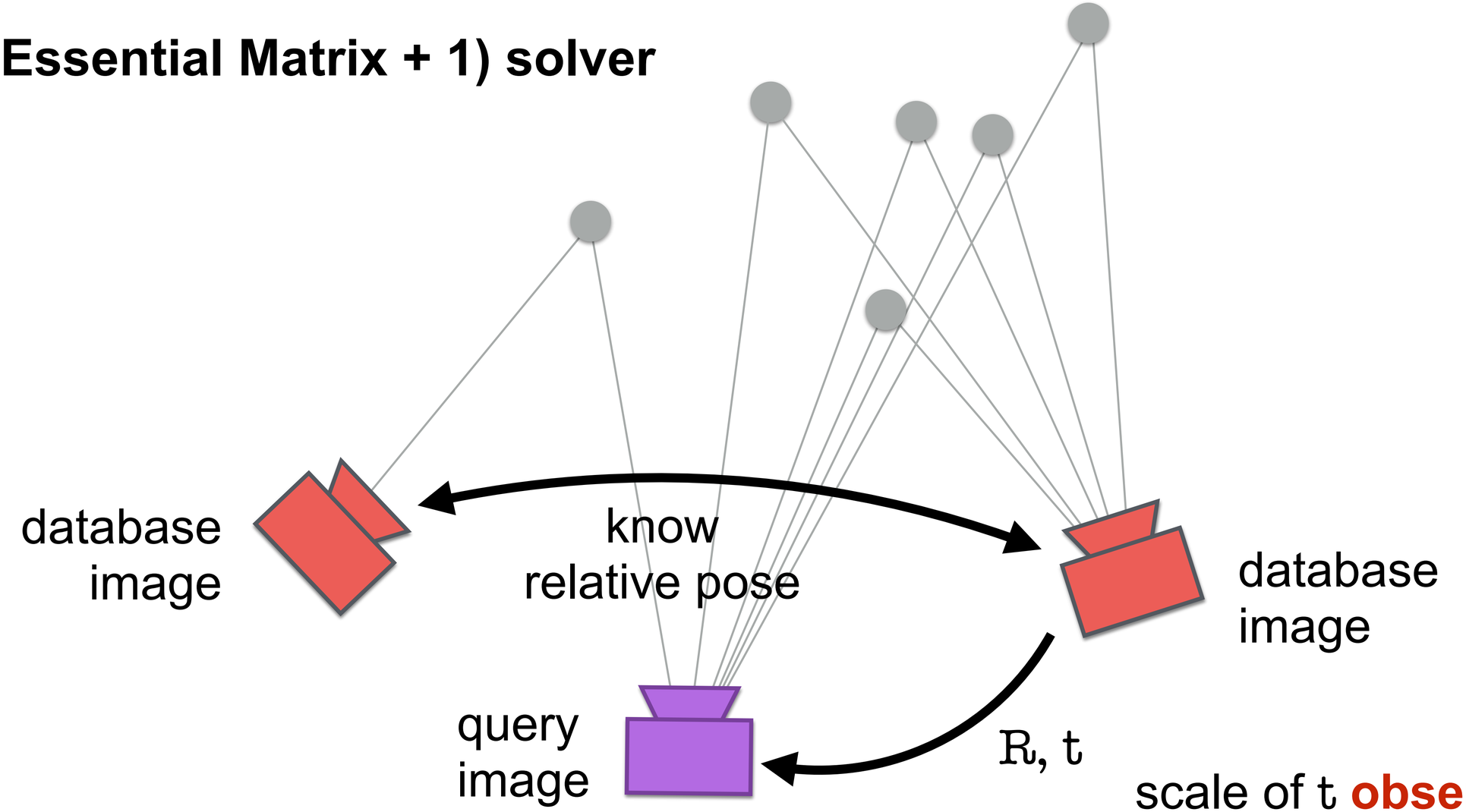
E5+1 (Essential Matrix + 1) solver



[Zheng, Wu, Structure from Motion Using Structure-less Resection, ICCV 2015]
[Bhayani, Sattler, Barath, Beliansky, Heikkila, Kukelova, Calibrated and Partially Calibrated Semi-Generalized Homographies, ICCV 2021]

Semi-Generalized Relative Pose Estimation

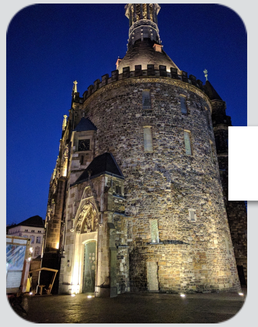
E5+1 (Essential Matrix + 1) solver



[Zheng, Wu, Structure from Motion Using Structure-less Resection, ICCV 2015]
[Bhayani, Sattler, Barath, Beliansky, Heikkila, Kukelova, Calibrated and Partially Calibrated Semi-Generalized Homographies, ICCV 2021]

Structure-Based vs. Structure-Less Localization

Structure-Based Localization



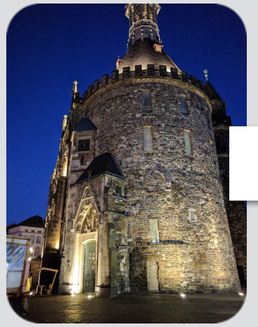
Find N most similar database images

Feature matching between query and top-N database images

2D-3D matches from associations between database features and 3D points

Pose estimation from 2D-3D matches (P3P-LO-RANSAC)

Structure-Less Localization



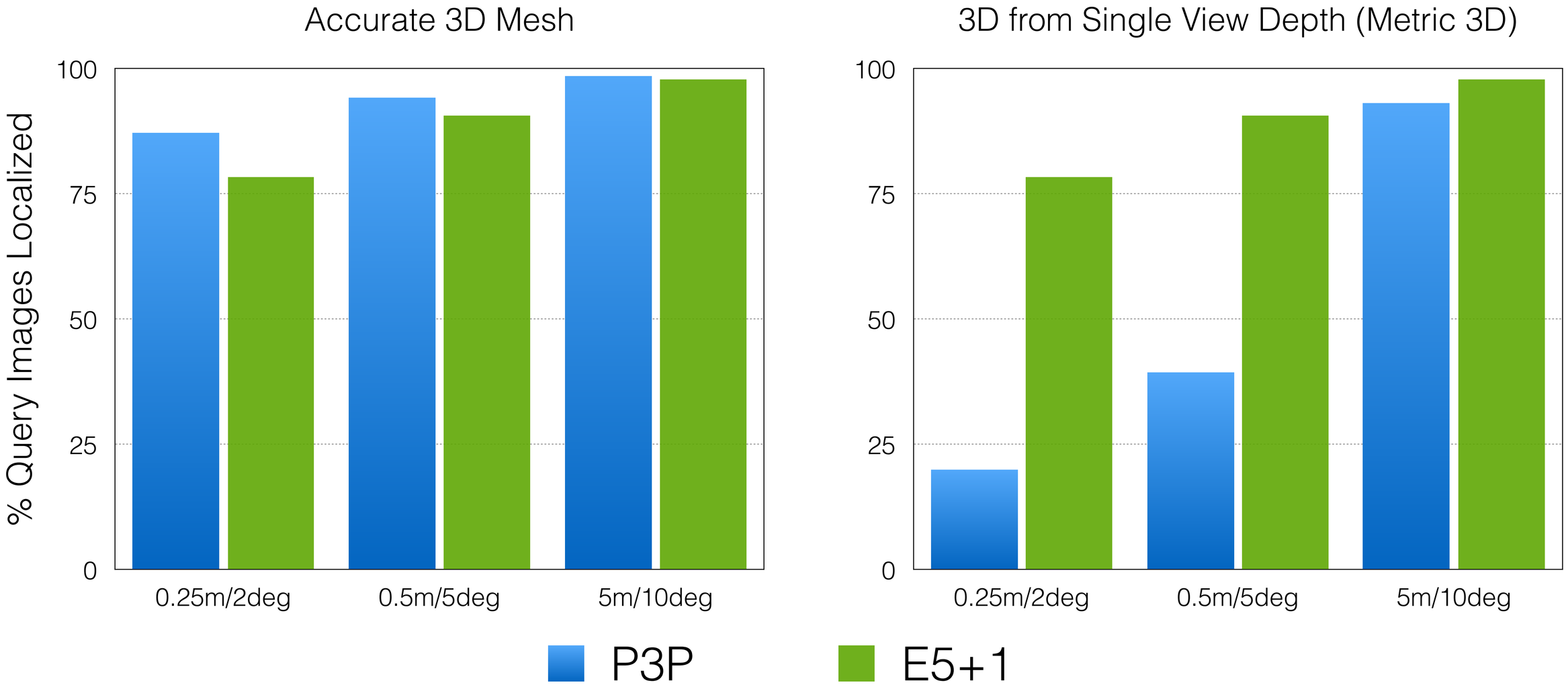
Find N most similar database images

Feature matching between query and top-N database images

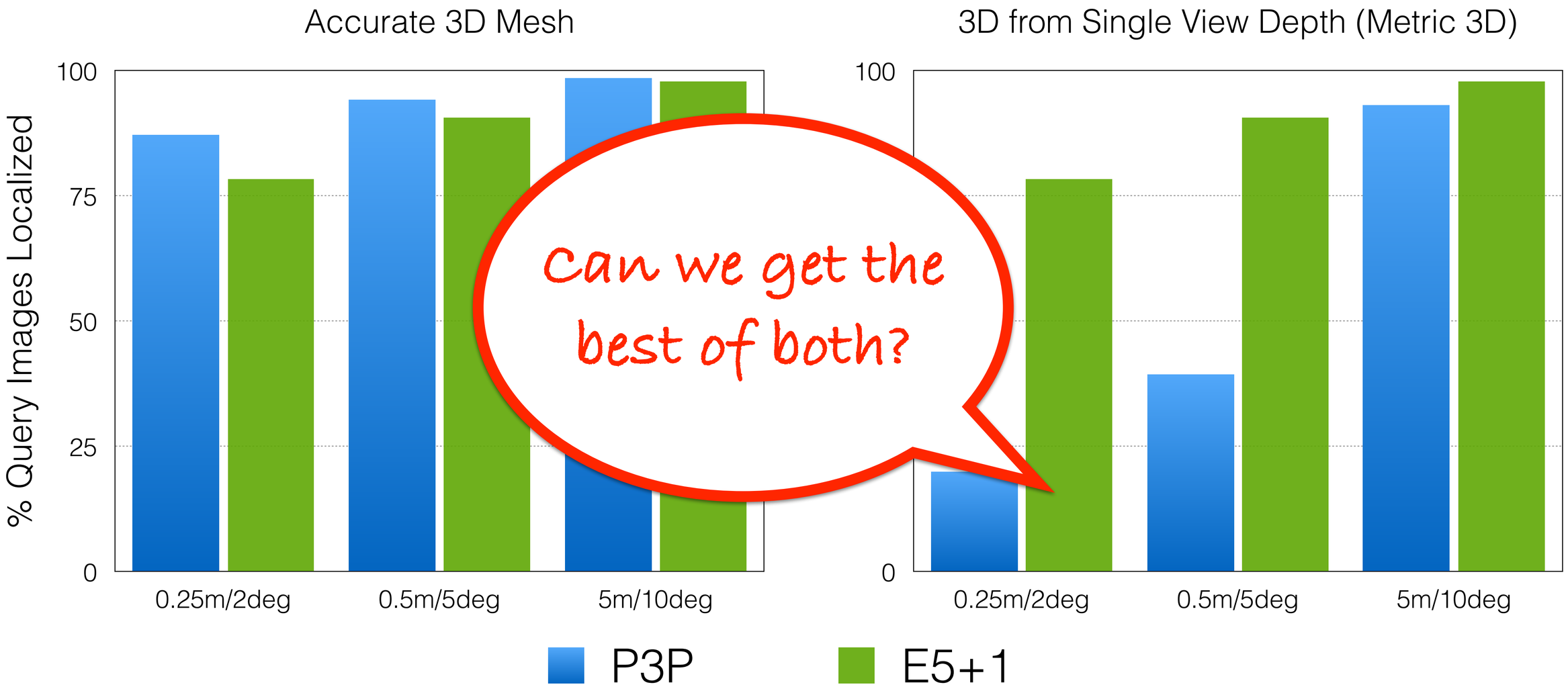
2D-2D matches

Pose estimation from 2D-3D matches (E5+1-LO-RANSAC)

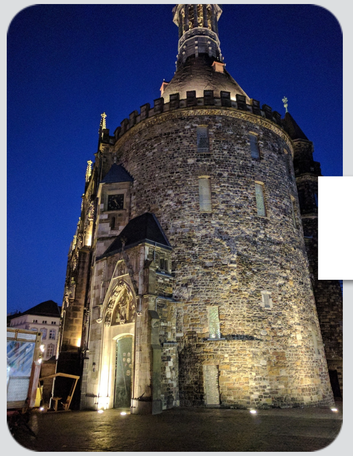
Structure-Based vs. Structure-Less Localization



Structure-Based vs. Structure-Less Localization



An Adaptive Strategy



Find N most similar database images

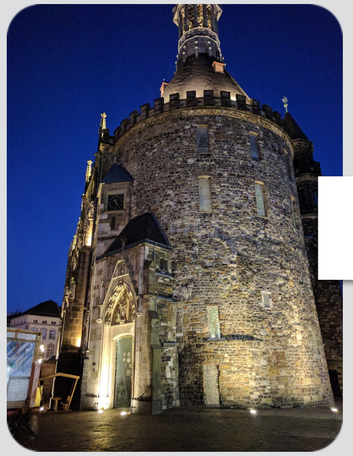
Feature matching between query and top-N database images

2D-2D matches

Pose estimation from 2D-3D matches (E5+1-LO-RANSAC)

[Panek, Sattler, Kukelova, Combining Absolute and Semi-Generalized Relative Poses for Visual Localization, arXiv:2409.14269]

An Adaptive Strategy



Find N most similar database images

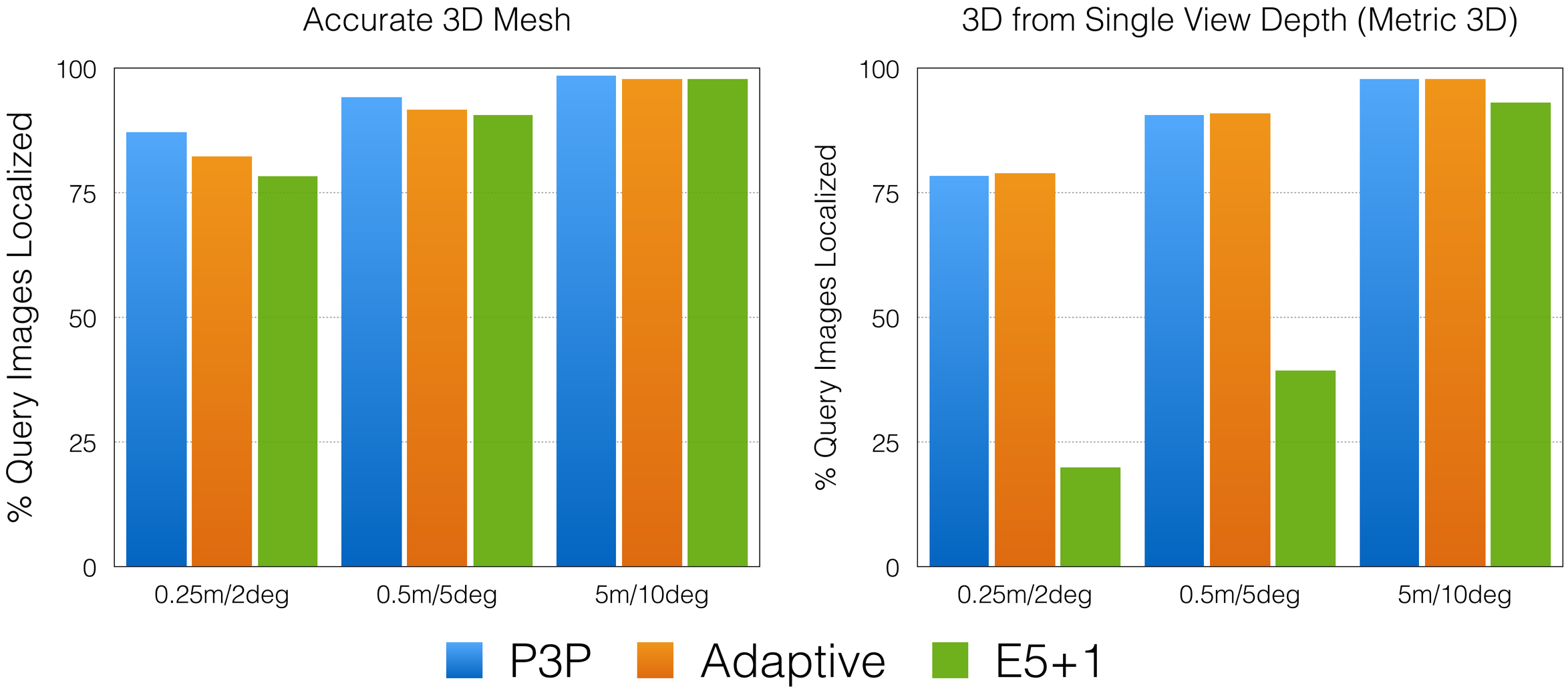
Feature matching between query and top-N database images

2D-2D **and** 2D-3D matches

Select between P3P-RANSAC and E5+1-RANSAC pose

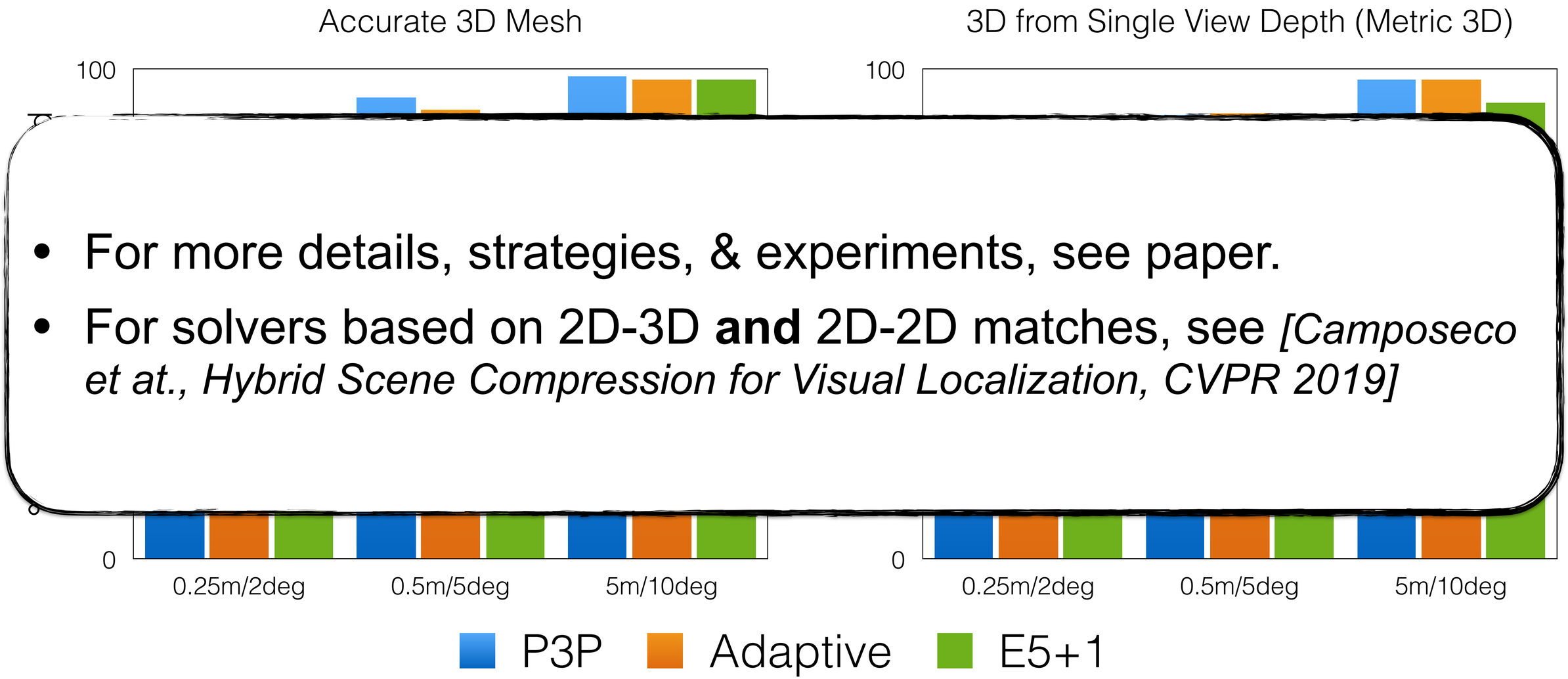
[Panek, Sattler, Kukelova, Combining Absolute and Semi-Generalized Relative Poses for Visual Localization, arXiv:2409.14269]

Structure-Less vs. Structure-Based Localization



[Panek, Sattler, Kukelova, Combining Absolute and Semi-Generalized Relative Poses for Visual Localization, arXiv:2409.14269]

Structure-Less vs. Structure-Based Localization

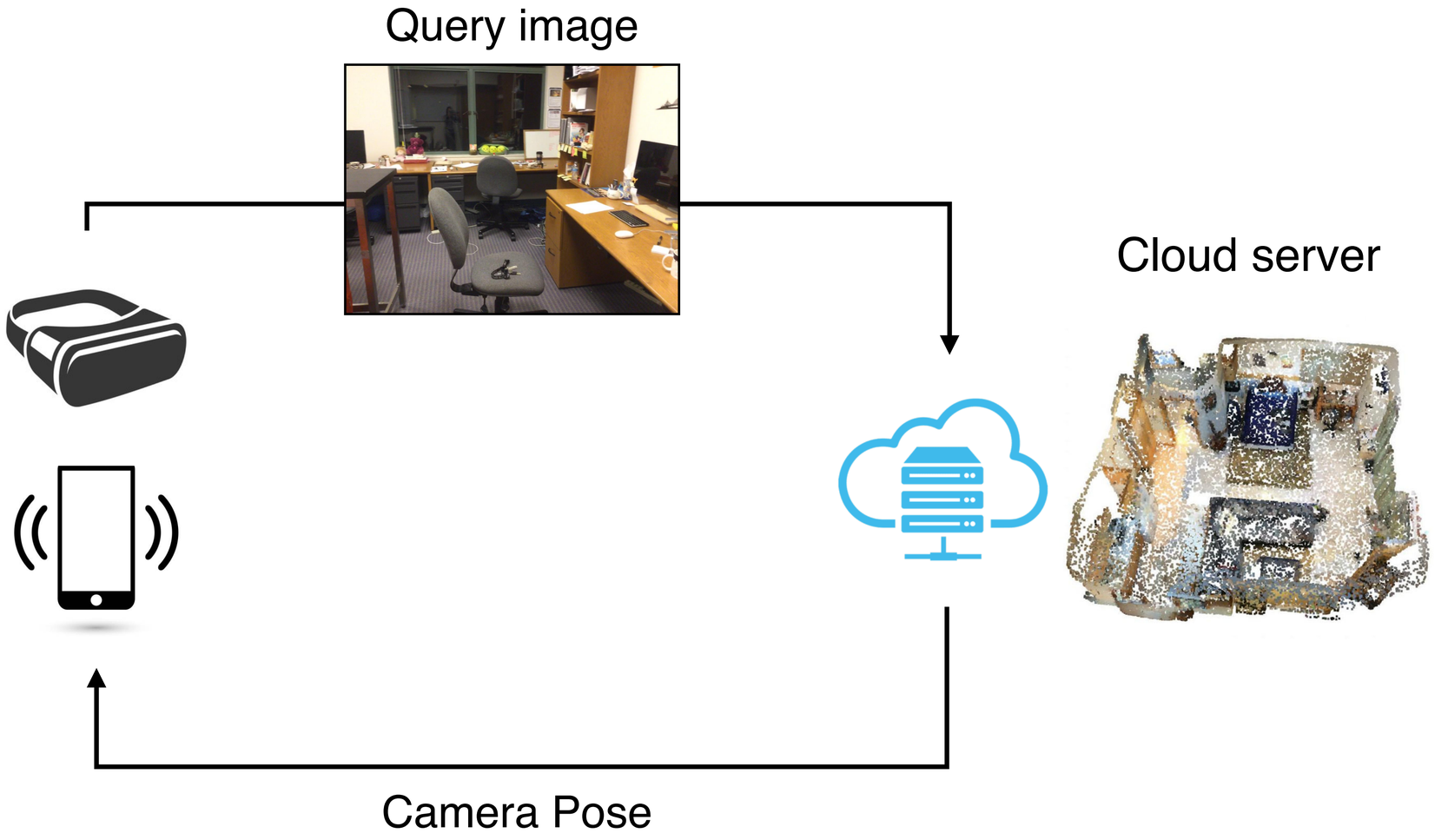


- For more details, strategies, & experiments, see paper.
- For solvers based on 2D-3D **and** 2D-2D matches, see *[Camposeco et at., Hybrid Scene Compression for Visual Localization, CVPR 2019]*

[Panek, Sattler, Kukelova, Combining Absolute and Semi-Generalized Relative Poses for Visual Localization, arXiv:2409.14269]

What About Privacy?

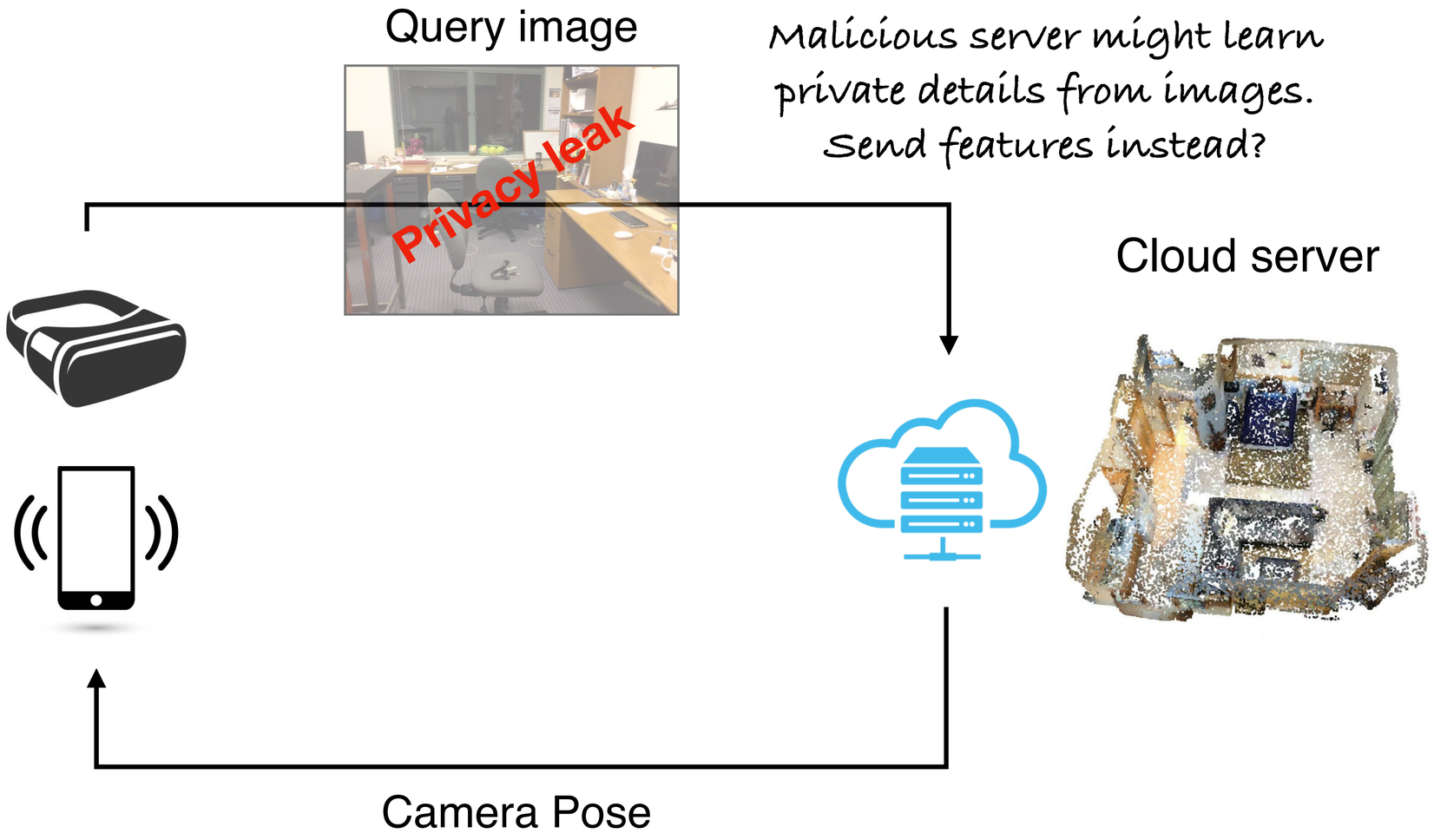
What About Privacy?



slide credit: Kunal Chelani

[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]

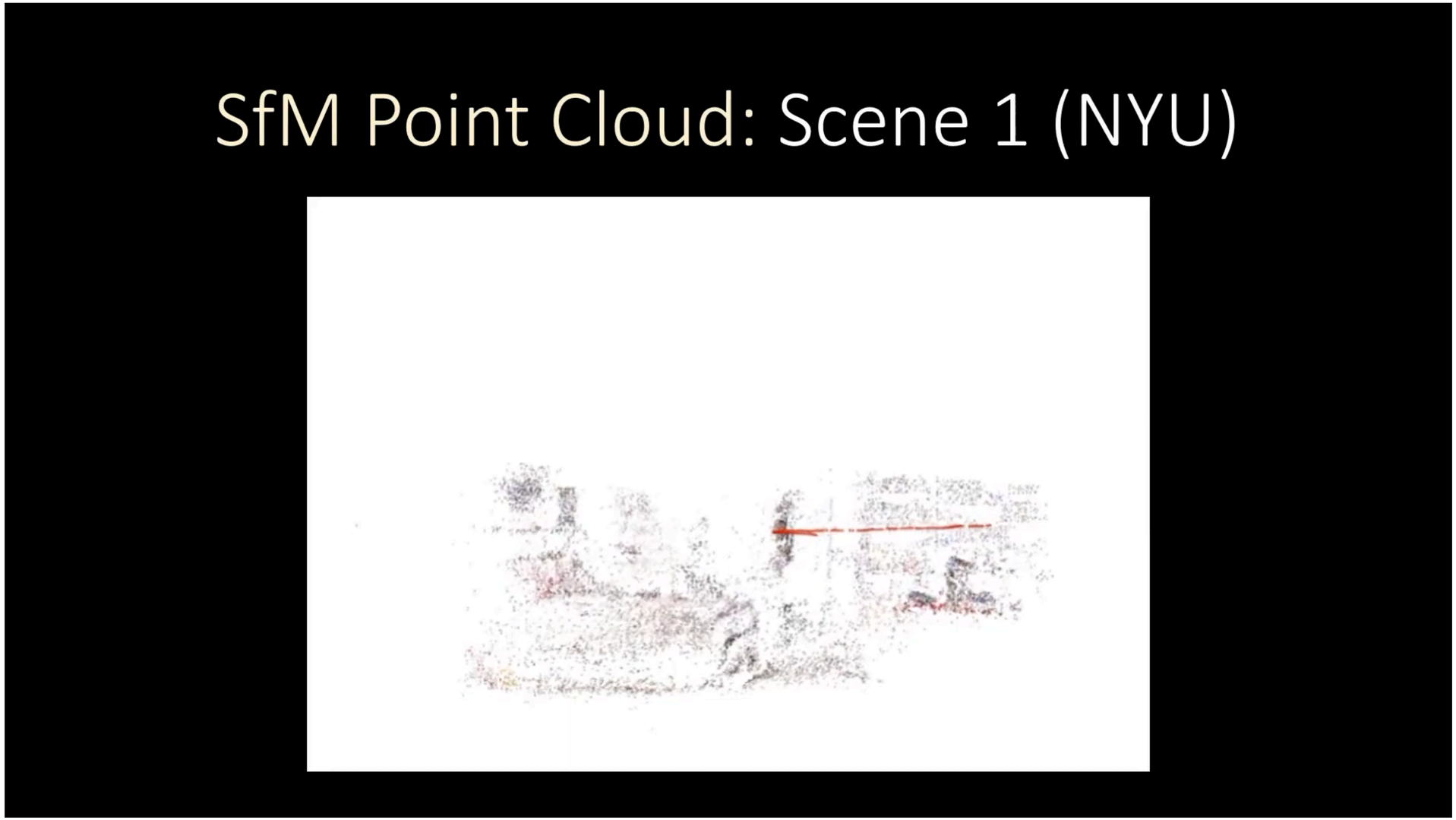
What About Privacy?



slide credit: Kunal Chelani

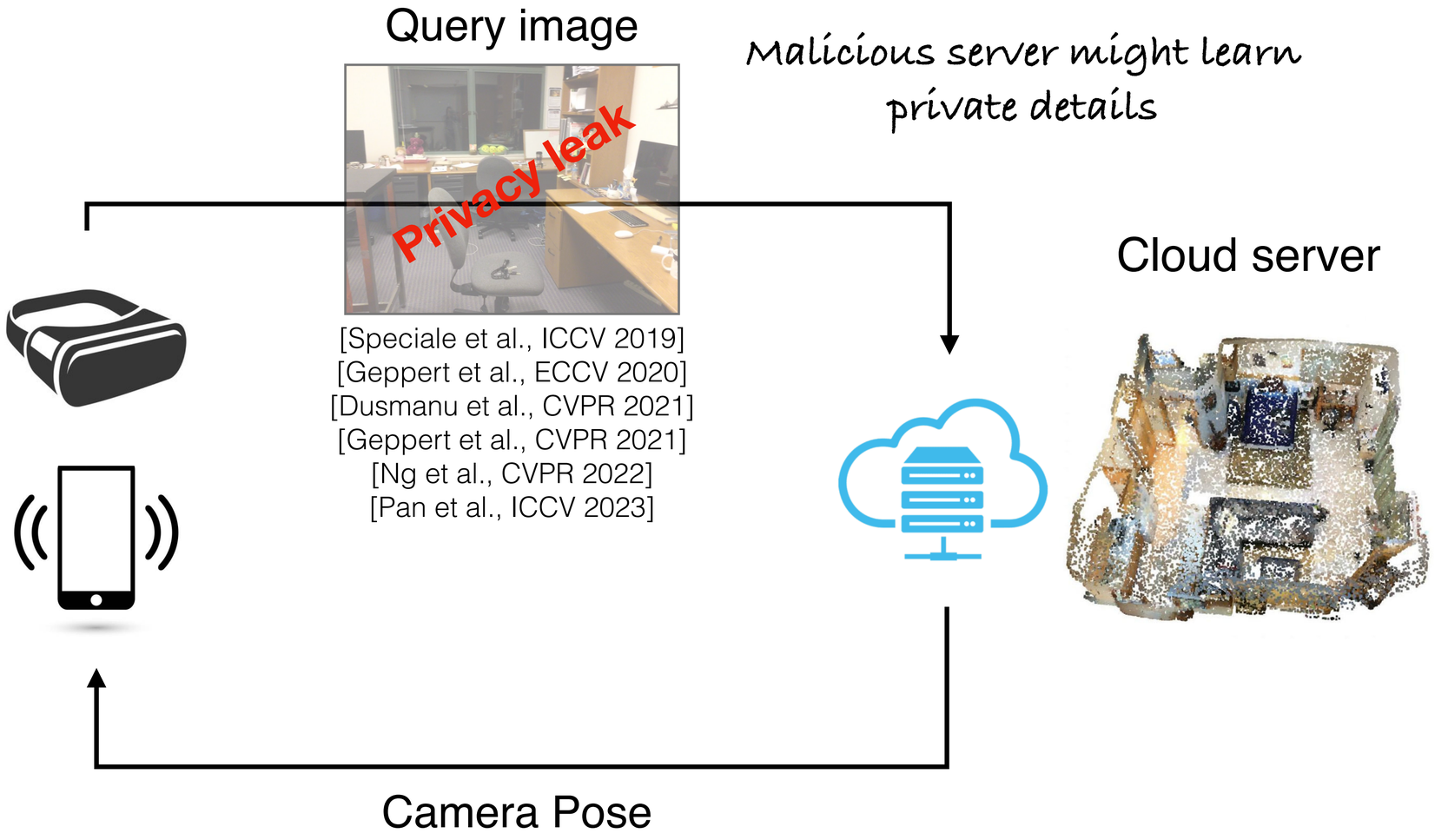
[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]

Privacy Issues in Visual Localization



[Pittaluga, Koppal, Kang, Sinha, Revealing Scenes by Inverting Structure From Motion Reconstructions, CVPR 2019]

What About Privacy?



slide credit: Kunal Chelani

[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]
[Chelani, Benbihi, Kahl, Sattler, Kukelova, Obfuscation Based Privacy Preserving Representations are Recoverable Using Neighborhood Information, arXiv:2409.11536]

What About Privacy?

Query image

[Speciale et al., ICCV 2019]
[Geppert et al., ECCV 2020]
[Dusmanu et al., CVPR 2021]
[Geppert et al., CVPR 2021]
[Ng et al., CVPR 2022]
[Pan et al., ICCV 2023]

Malicious server might learn private details

Cloud server

[Speciale et al., CVPR 2019]
[Lee et al., CVPR 2023]
[Pan et al., ICCV 2023]

Attacker might learn private details from maps shared by users

Camera Pose

slide credit: Kunal Chelani

[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]
[Chelani, Benbihi, Kahl, Sattler, Kukelova, Obfuscation Based Privacy Preserving Representations are Recoverable Using Neighborhood Information, arXiv:2409.11536]

What About Privacy?

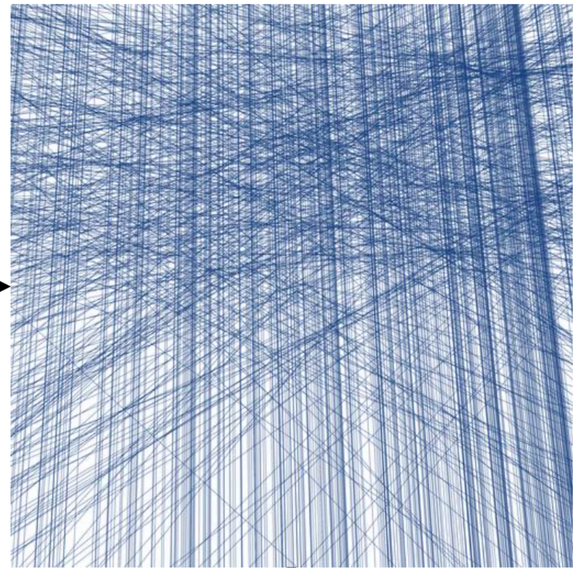


image credit: Marcel Geppert

- Methods based on obfuscating scene / image geometry
- Original geometry can be recovered quite easily if neighborhood information is available
- Neighborhoods can be approximately recovered from descriptors

  See [Chelani, Benbihi et al., arXiv:2409.11536] for details

al., CVPR 2019]
., CVPR 2023]
., ICCV 2023]

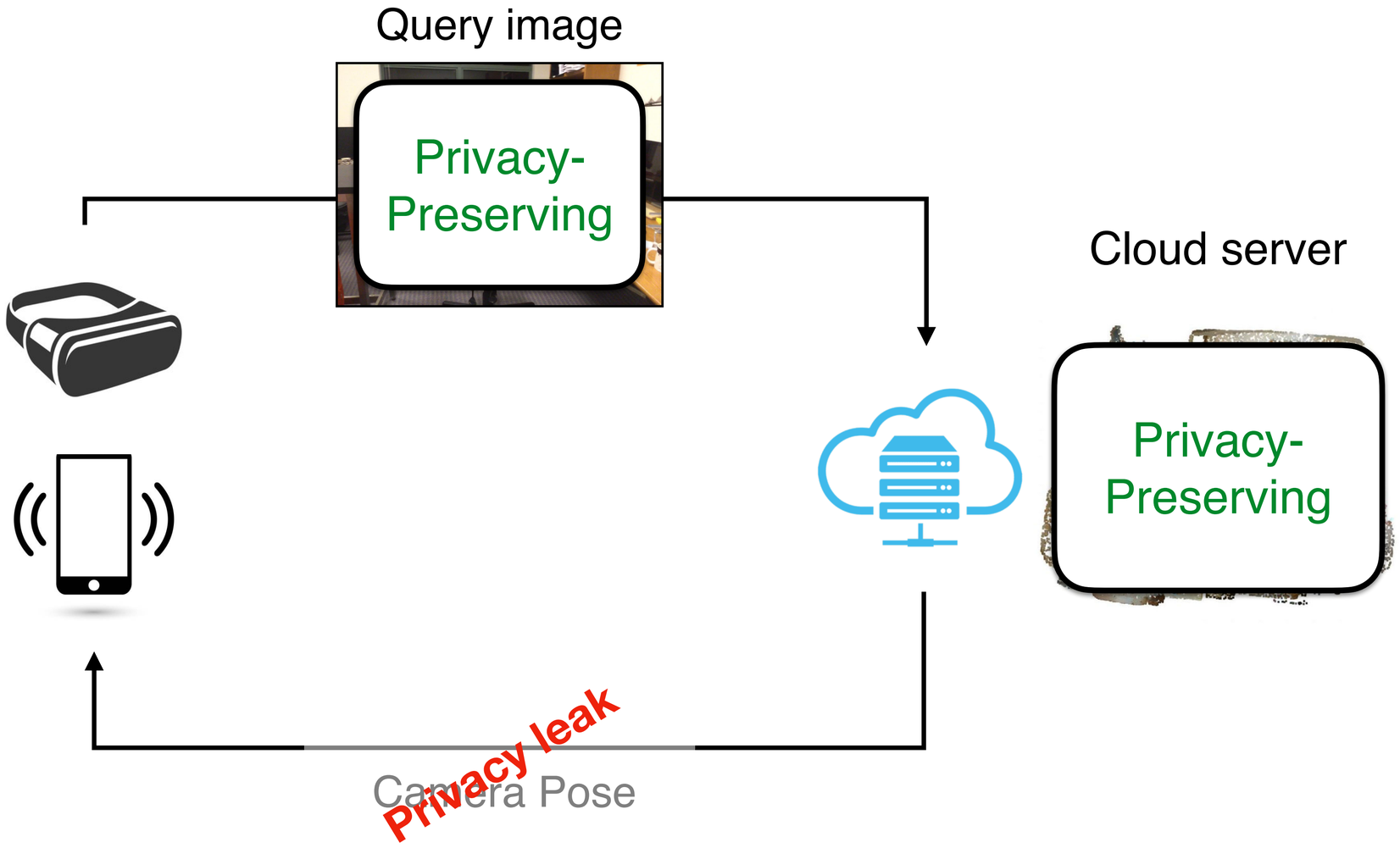
vate

j users

. Kunal Chelani

[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]
[Chelani, Benbihi, Kahl, Sattler, Kukelova, Obfuscation Based Privacy Preserving Representations are Recoverable Using Neighborhood Information, arXiv:2409.11536]

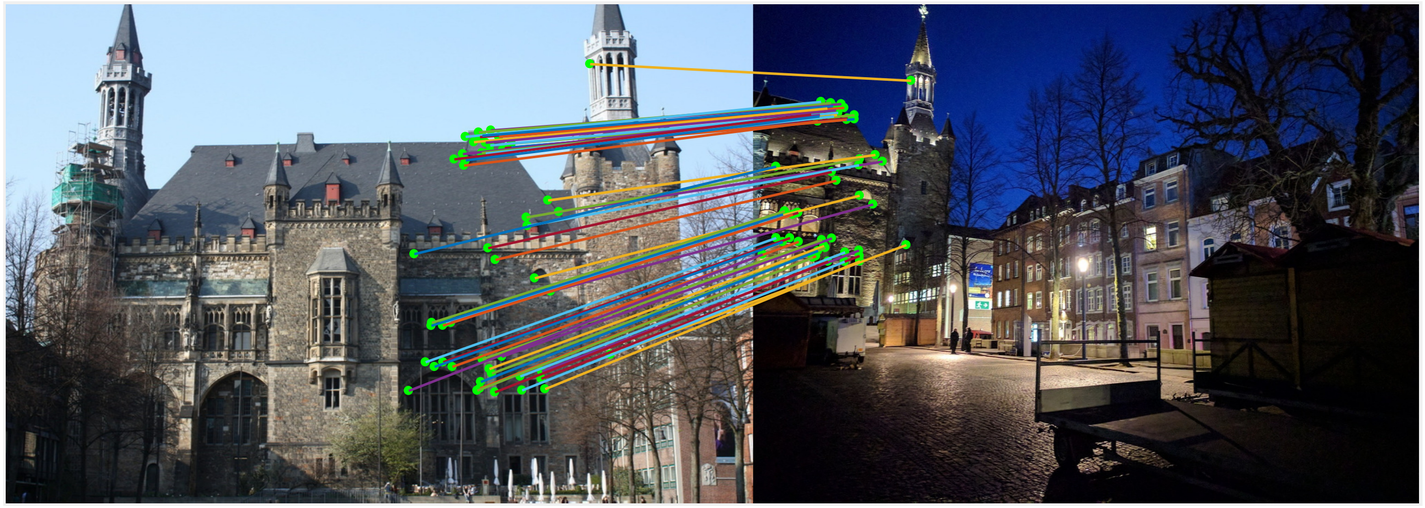
Are Privacy-Preserving Representations Enough?



slide credit: Kunal Chelani

[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]

The Downside of Robust Localization



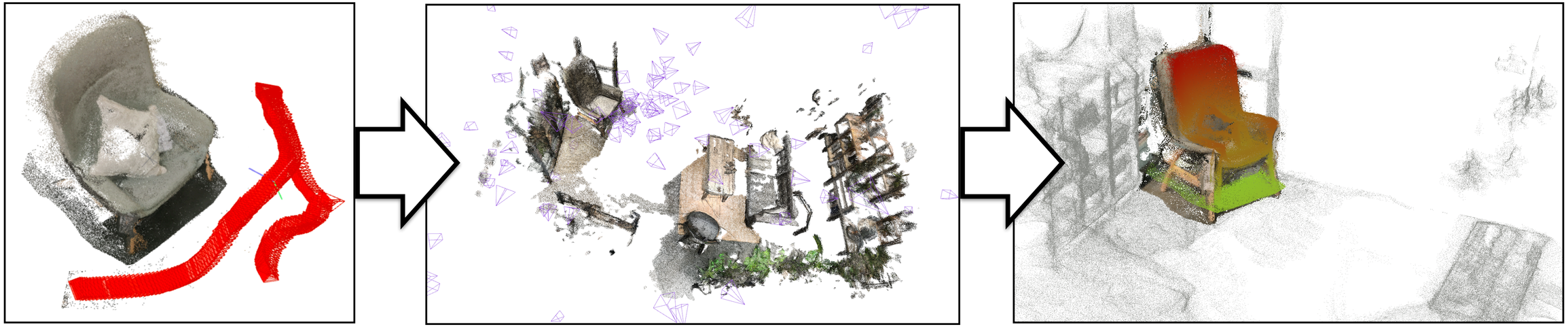
Robustness to shape and appearance variations means we can match images of different object instances

slide credit: Mihai Dusmanu, Kunal Chelani

[Dusmanu, Rocco, Pajdla, Pollefeys, Sivic, Torii, Sattler, D2-Net: A Trainable CNN for Joint Detection and Description of Local Features, CVPR 2019]

Recovering Scene Content from Camera Poses

input:
image sequence of object
(e.g., from Internet)



Attacker runs SfM to get camera poses

Camera poses returned by the server

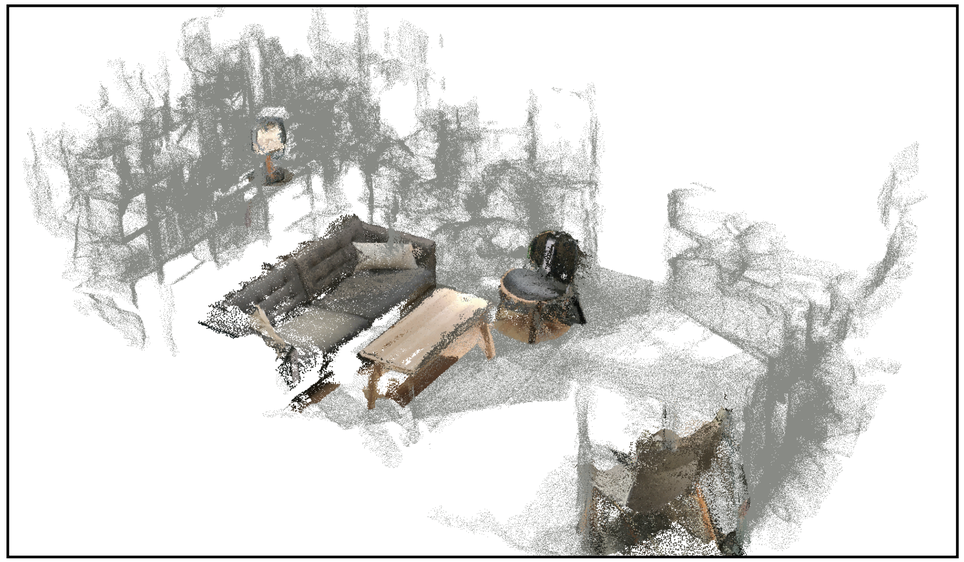
Object position from pose alignment

slide credit: Kunal Chelani

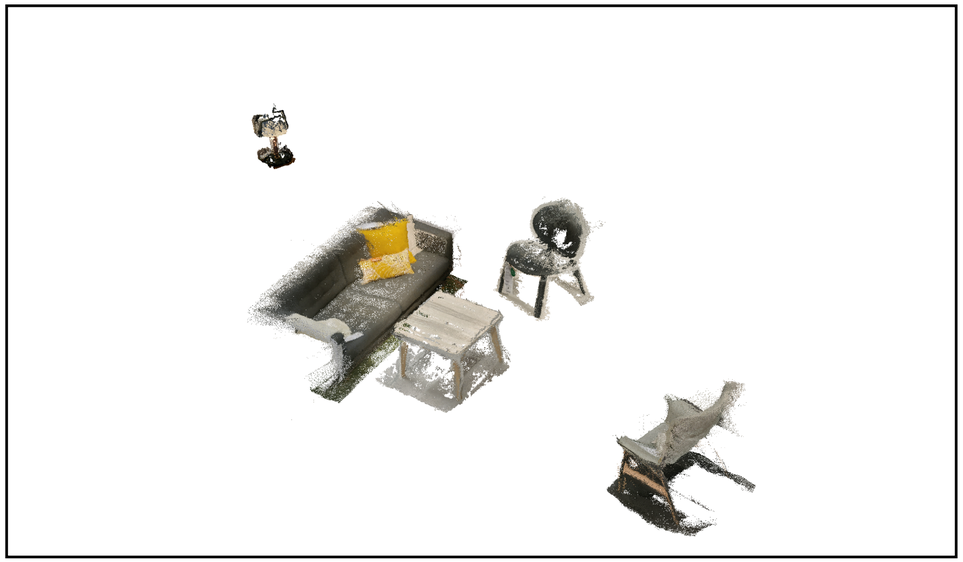
[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]

Qualitative Results

slide credit: Kunal Chelani



Actual scene with highlighted objects



Roughly reconstructed scene

[Chelani, Sattler, Kahl, Kukelova. Privacy-Preserving Representations are not Enough: Recovering Scene Content from Camera Poses, CVPR 2023]

Open Positions

- Open PhD & PostDoc position on camera geometry estimation, starting in 2025
- Contact: kukelova@gmail.com

- One open postdoc positions on privacy-preserving / temporal 3D mapping
- Open PhD position on visual localization
- Contact: torsten.sattler@cvut.cz