CHALMERS

FACULTY OF ELECTRICAL ENGINEERING CTU IN PRAGUE

CZECH INSTITUTE OF INFORMATICS ROBOTICS AND CYBERNETICS CTU IN PRAGUE

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Privacy-Preserving Representations are not Enough - Recovering Scene Content from Camera Poses.




Kunal Chelani[1] Torsten Sattler[3] Fredrik Kahl[1] Zuzana Kukelova[2]

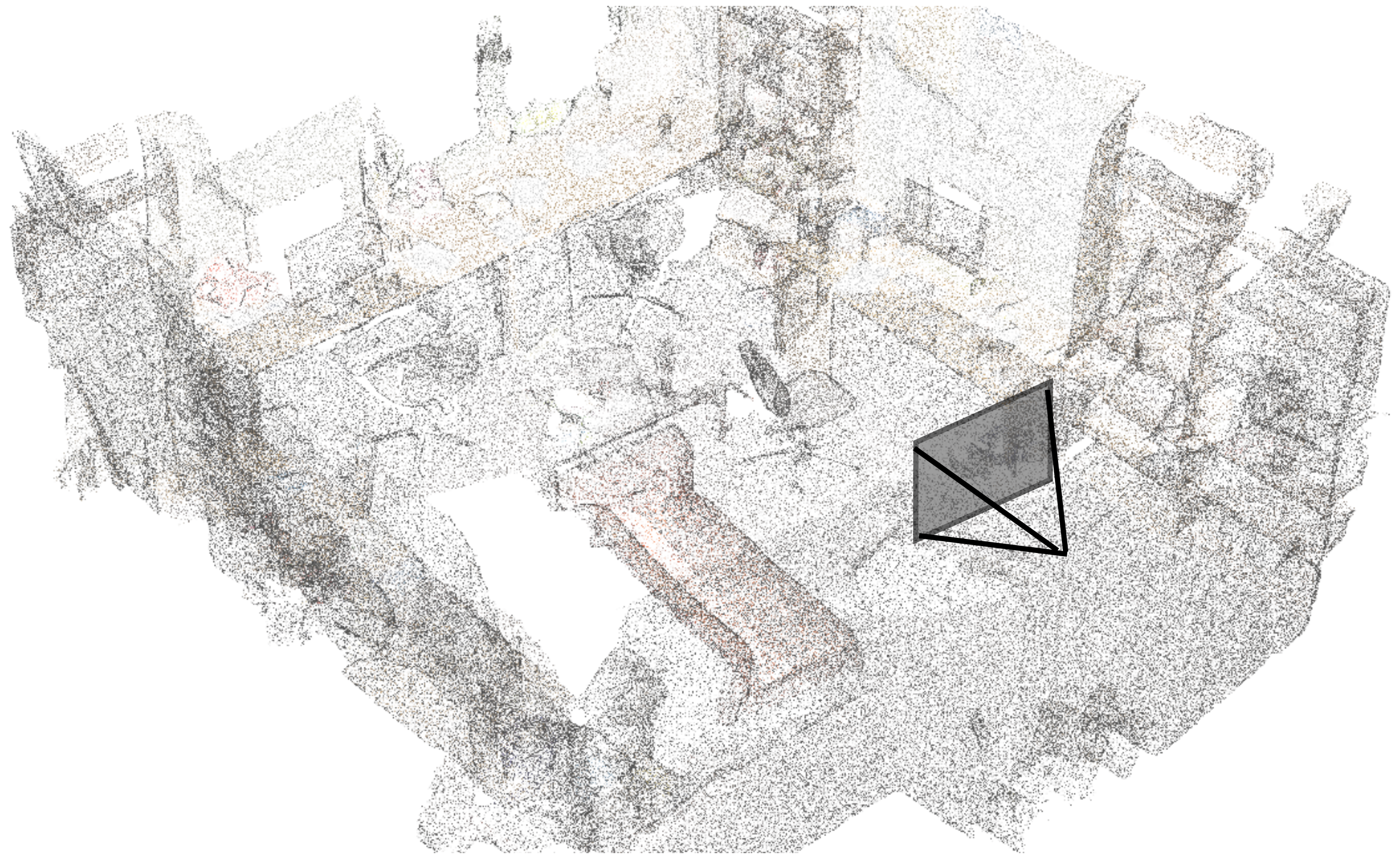
[1]Chalmers University of Technology
[2]Visual Recognition Group, Department of Electrical Engineering, CTU in Prague
[3]Czech Institute of Informatics, Robotics and Cybernetics, CTU in Prague

WED-PM-074

Visual localization

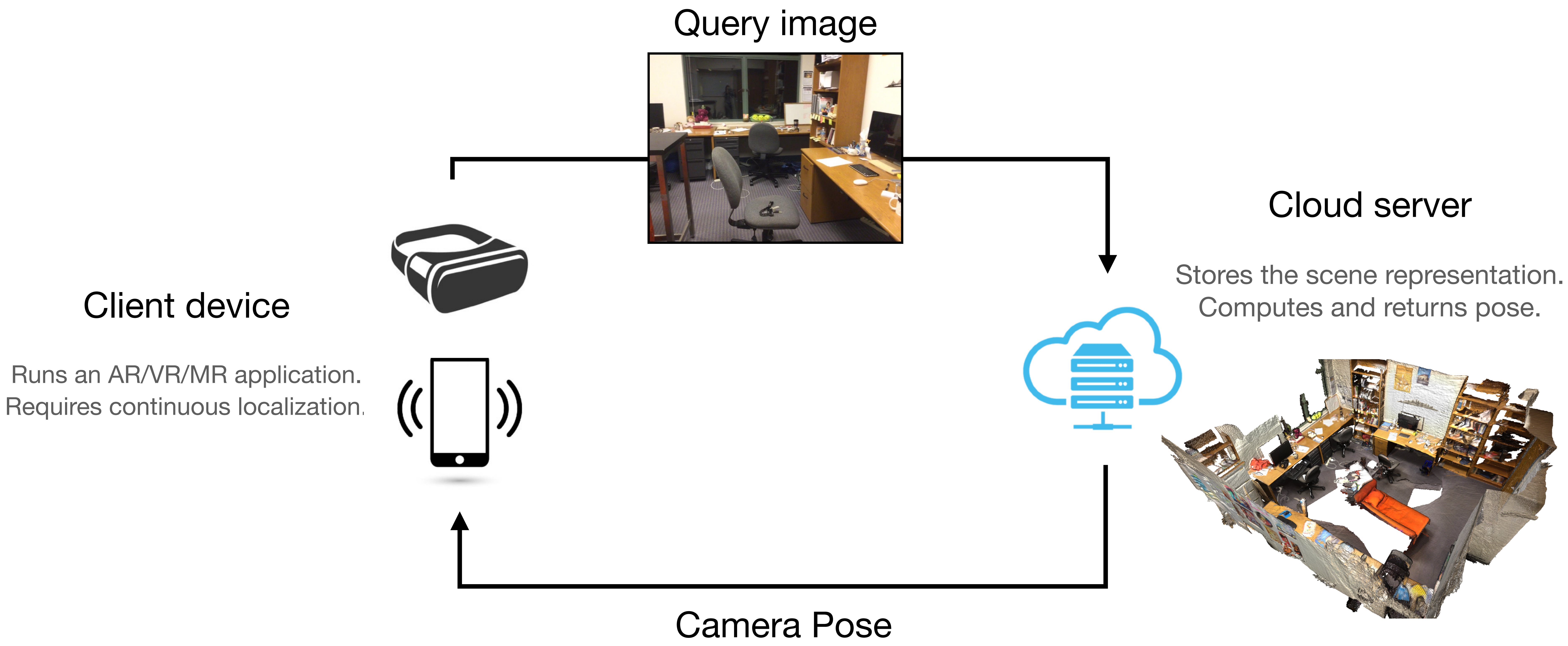
Query image

3D Scene defining coordinate system

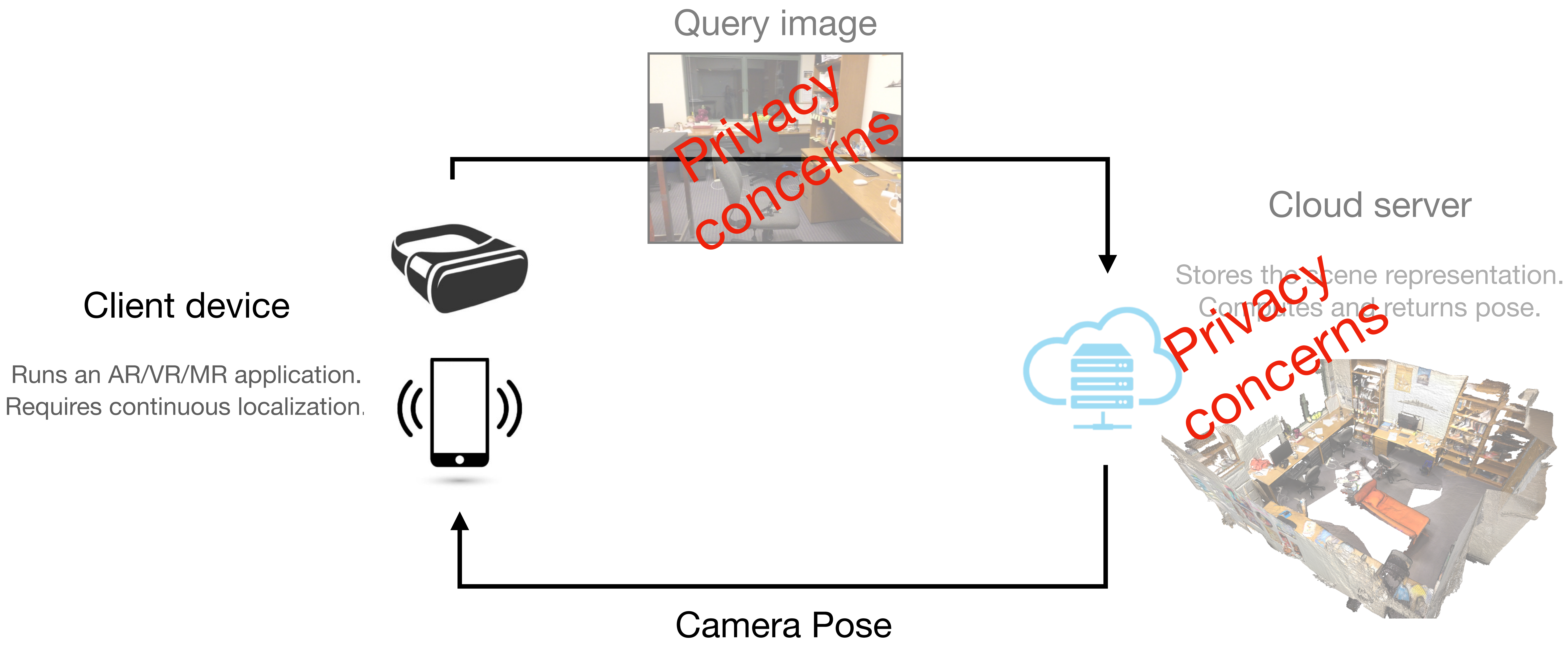


Camera pose

Client-server based visual localization

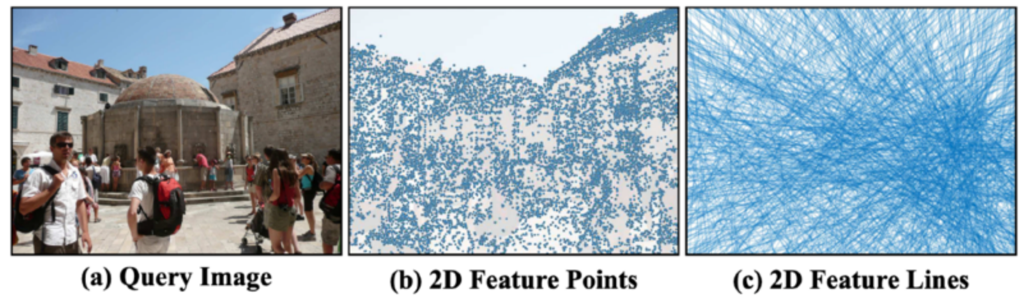


Client-server based visual localization

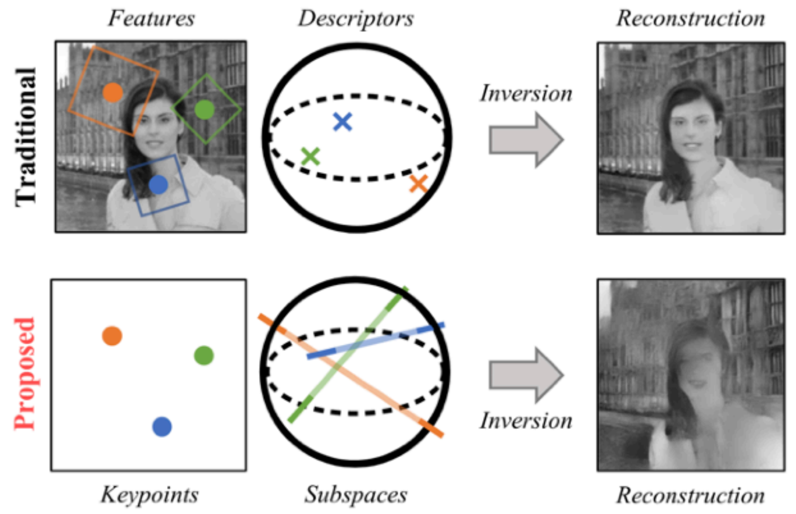


Privacy-preserving representations

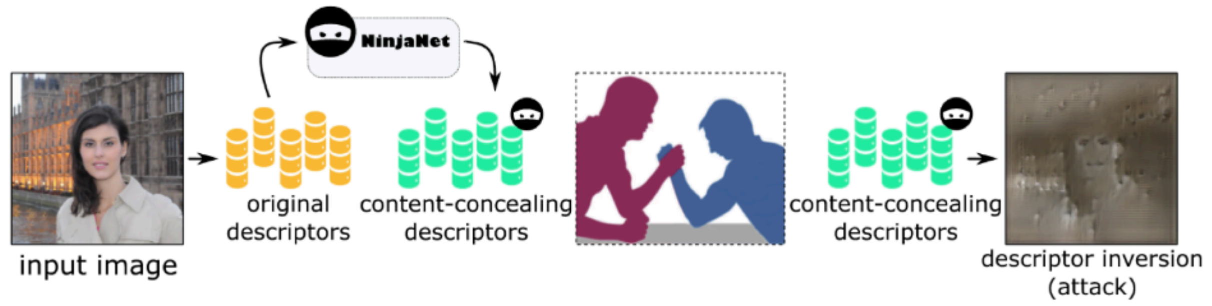
Query image



Speciale et al. Privacy Preserving Image Queries for Camera Localization, CVPR 2019

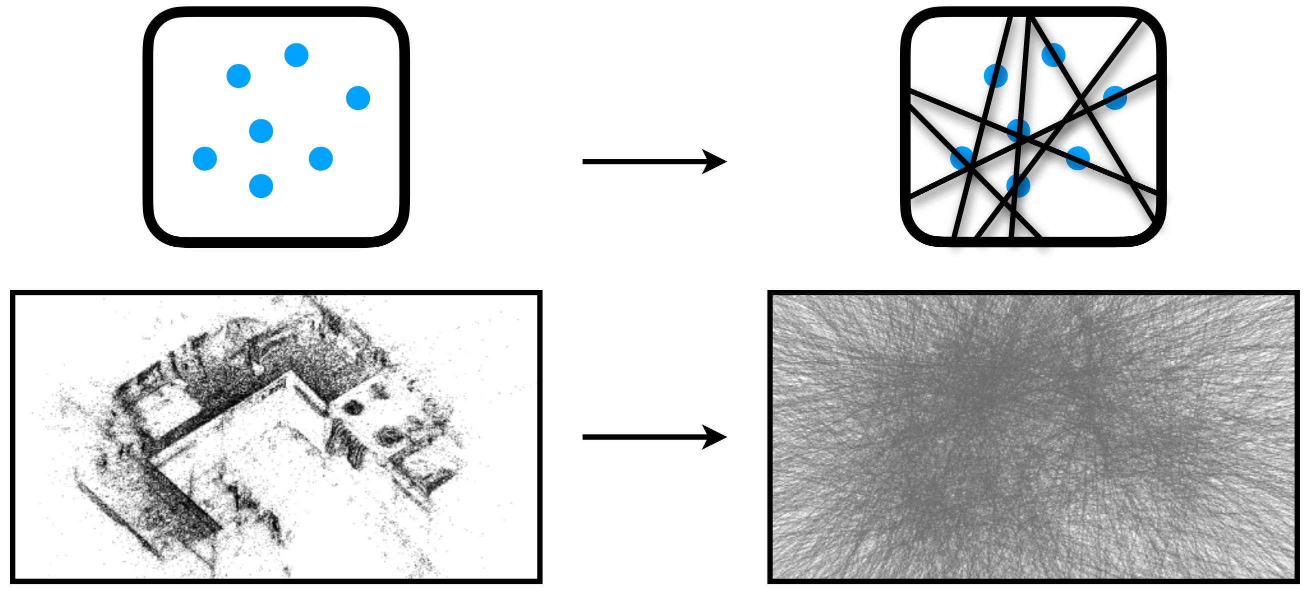


Dusmanu et al. Privacy-Preserving Image Features via Adversarial Affine Subspace Embeddings, CVPR 2021

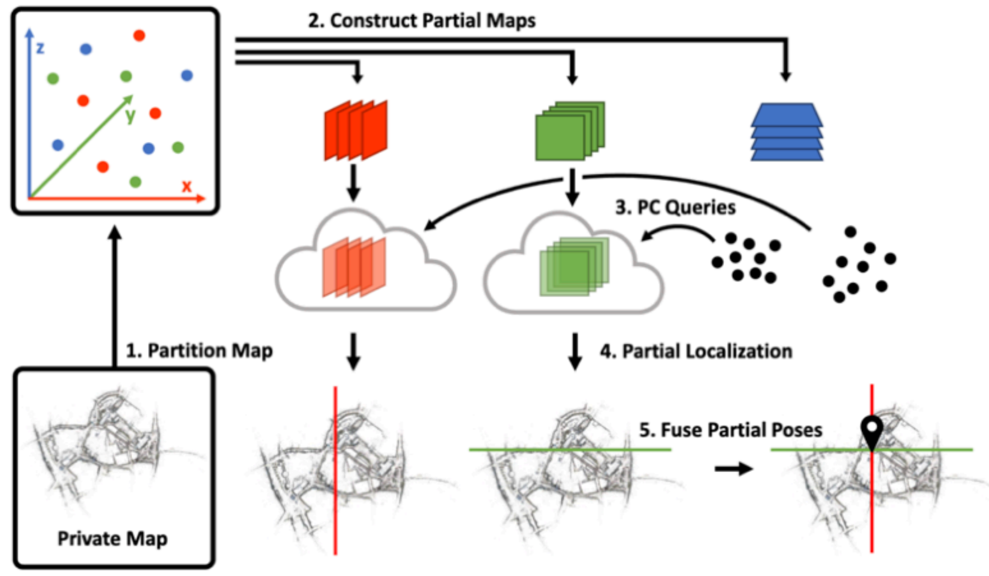


Ng et al. NinjaDesc, CVPR 2022

Map stored on cloud server

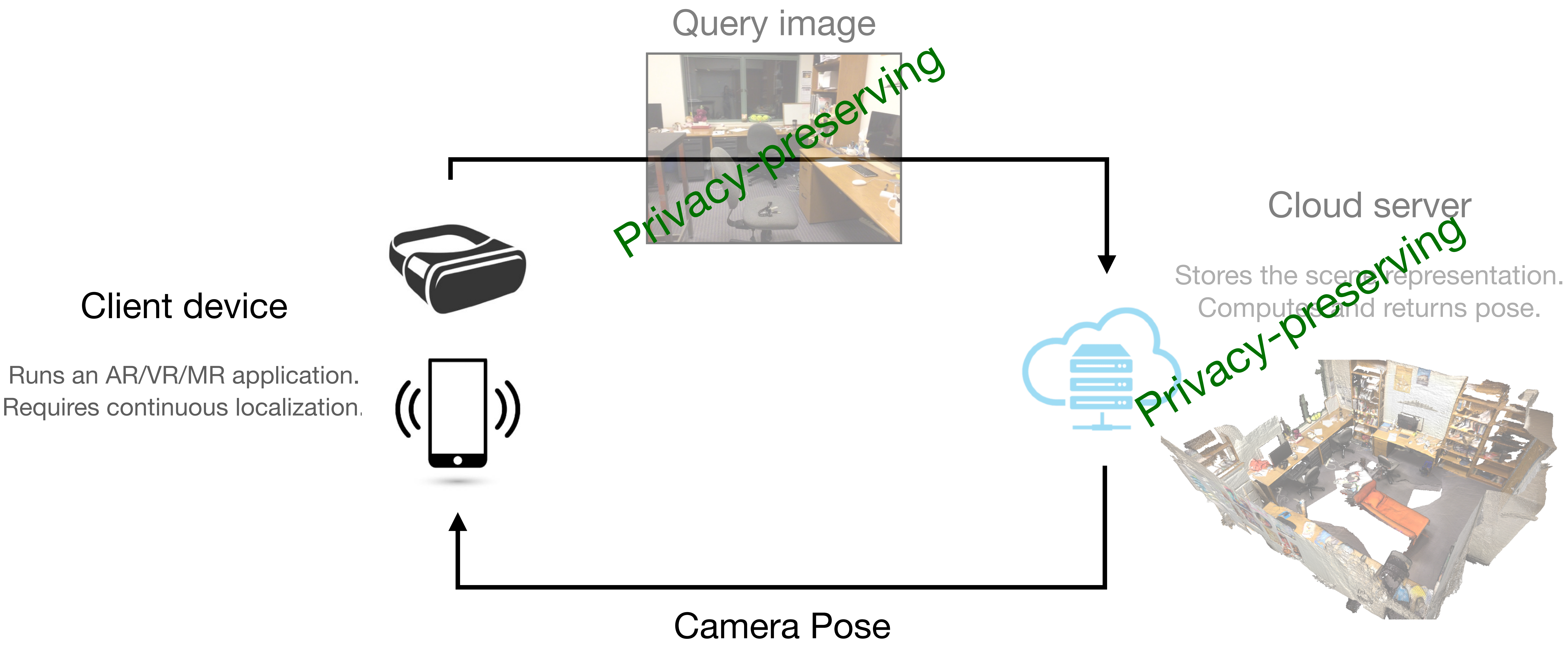


Speciale et al. Privacy Preserving Image-Based Localization, CVPR 2019

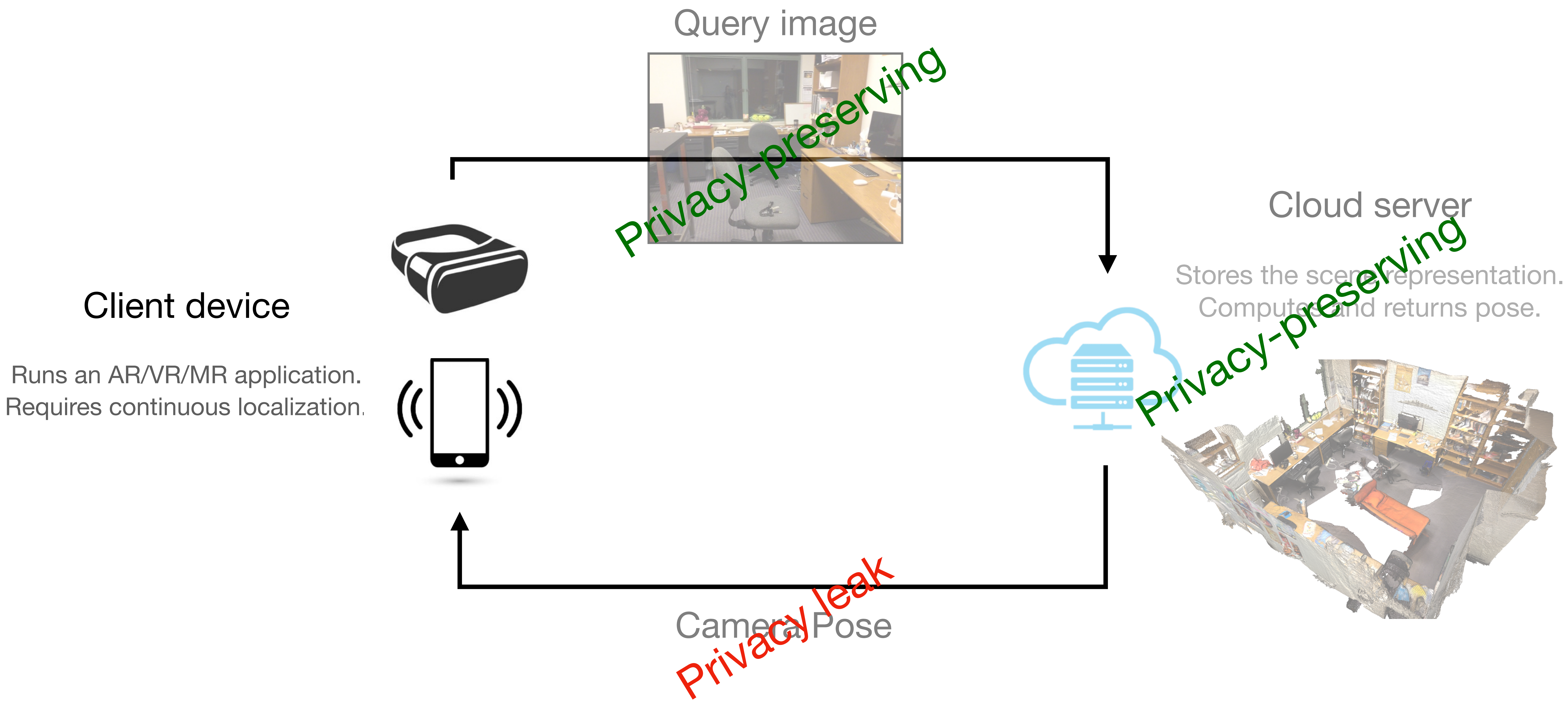


Geppert et al. Privacy Preserving Partial Localization, CVPR 2022

This paper


Query image
Privacy-preserving
Client device
Runs an AR/VR/MR application.
Requires continuous localization.
Cloud server
Stores the scene representation.
Computes and returns pose.
Privacy-preserving
Camera Pose

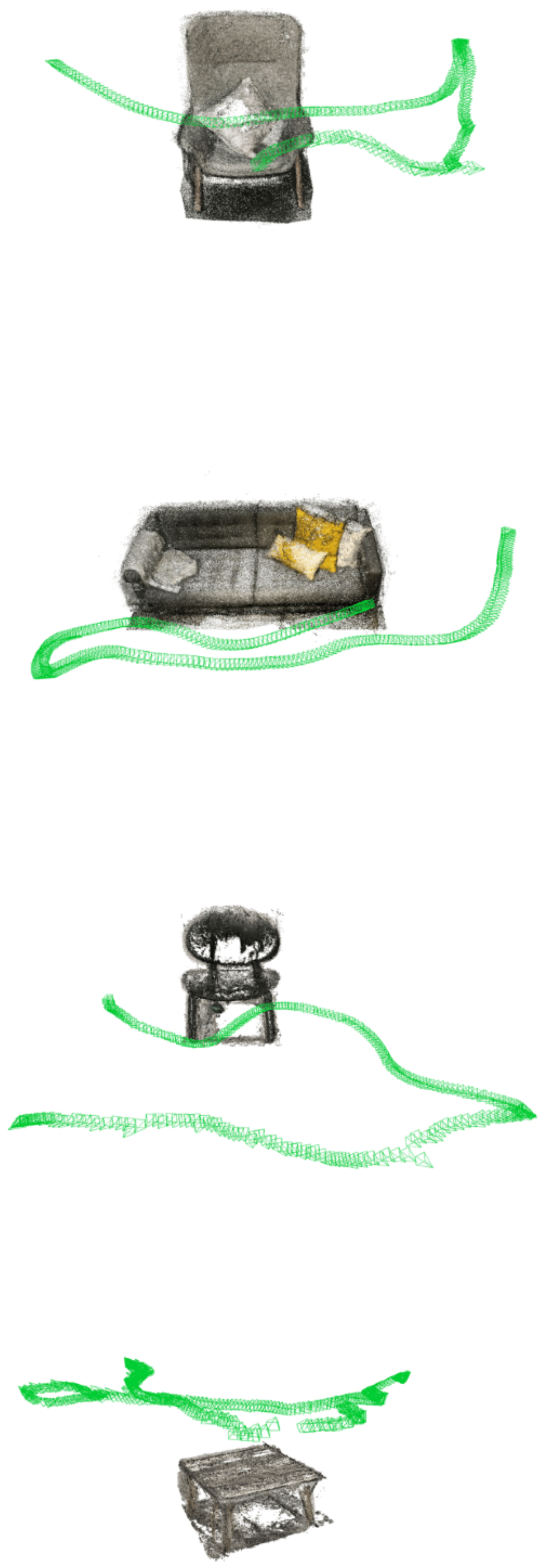
This paper


Query image
Privacy-preserving
Cloud server
Stores the scene representation.
Computes and returns pose.
Privacy-preserving
Client device
Runs an AR/VR/MR application.
Requires continuous localization.
Camera Pose
Privacy leak

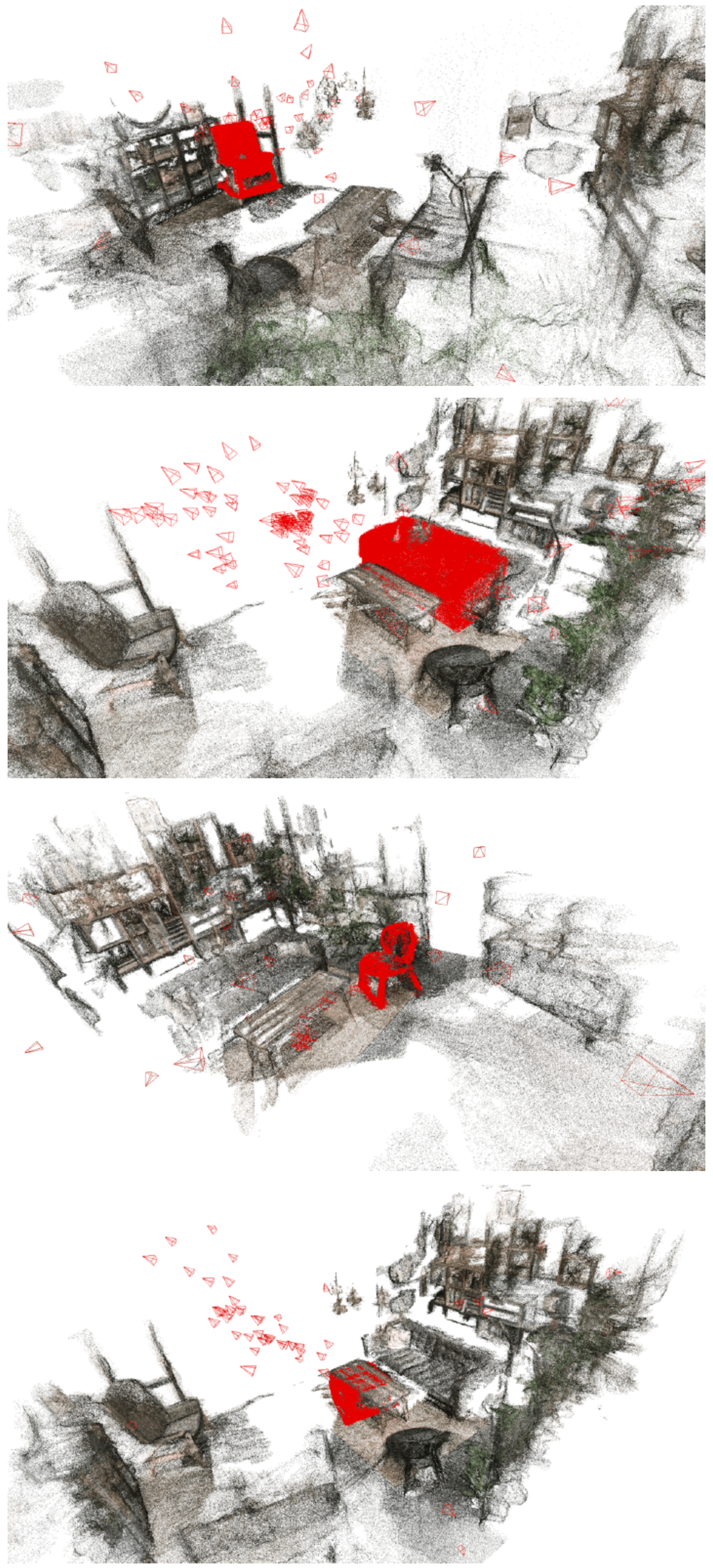
Recovering Scene Content from Camera Poses

Object Images

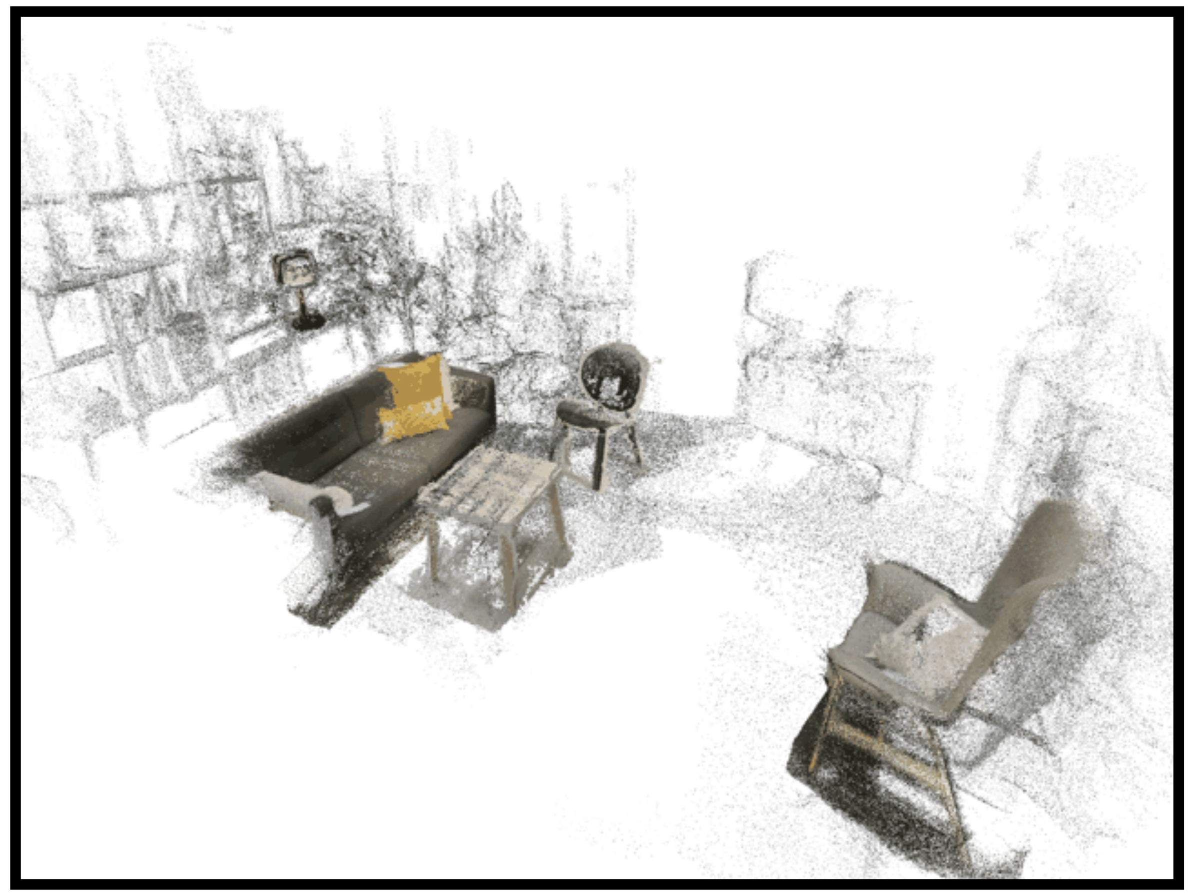
Local poses + SfM models



Poses from localization



Recovered scene layout



Inferred layout in colour against
Underlying scene in grey

CHALMERS

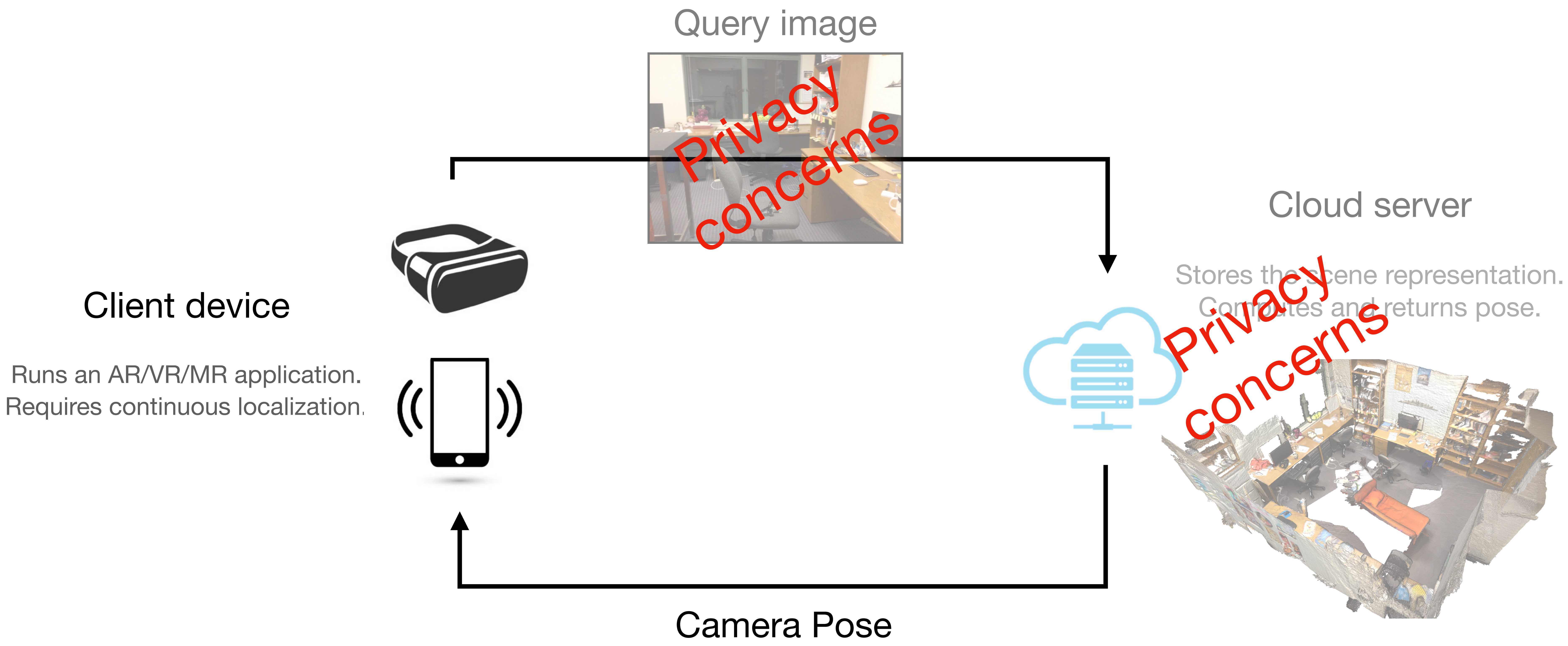
Privacy-Preserving Representations are not Enough - Recovering Scene Content from Camera Poses.




Kunal Chelani[1] Torsten Sattler[3] Fredrik Kahl[1] Zuzana Kukelova[2]

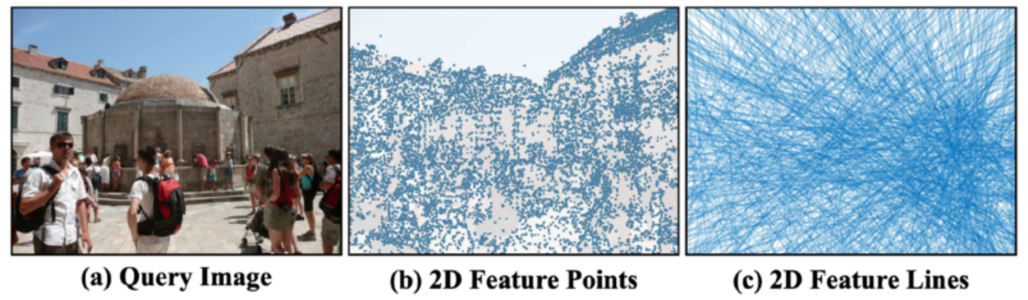
[1]Chalmers University of Technology
[2]Visual Recognition Group, Department of Electrical Engineering, CTU in Prague
[3]Czech Institute of Informatics, Robotics and Cybernetics, CTU in Prague

Client-server based visual localization

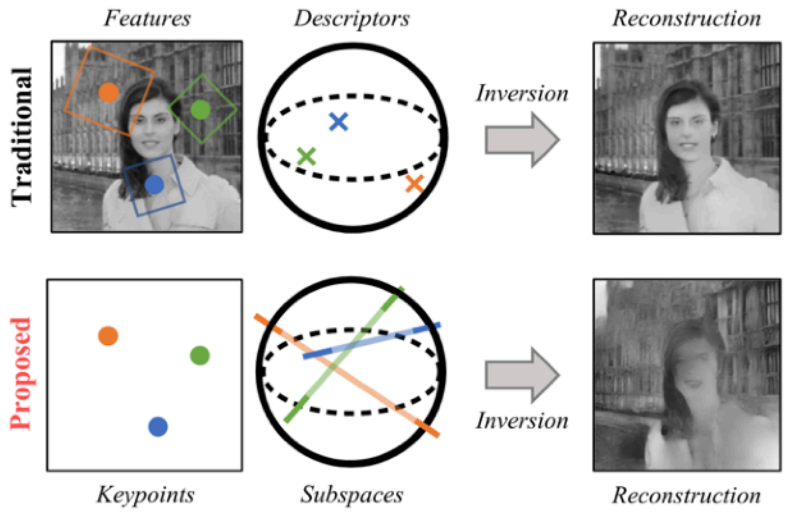


Privacy-preserving representations

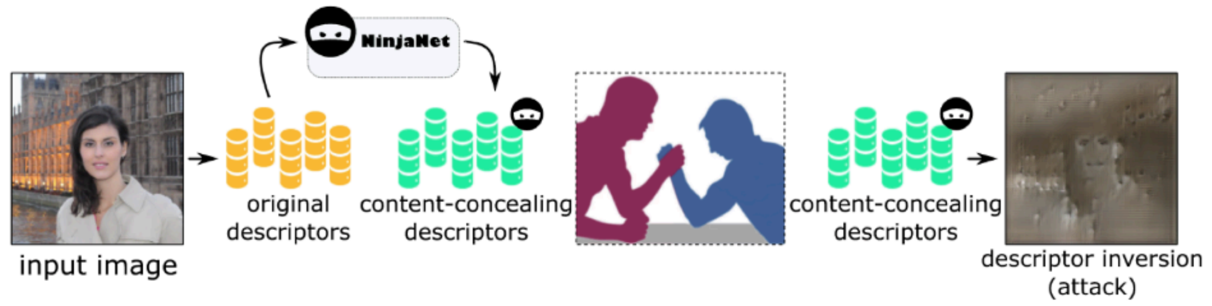
Query image



Speciale et al. Privacy Preserving Image Queries for Camera Localization, CVPR 2019

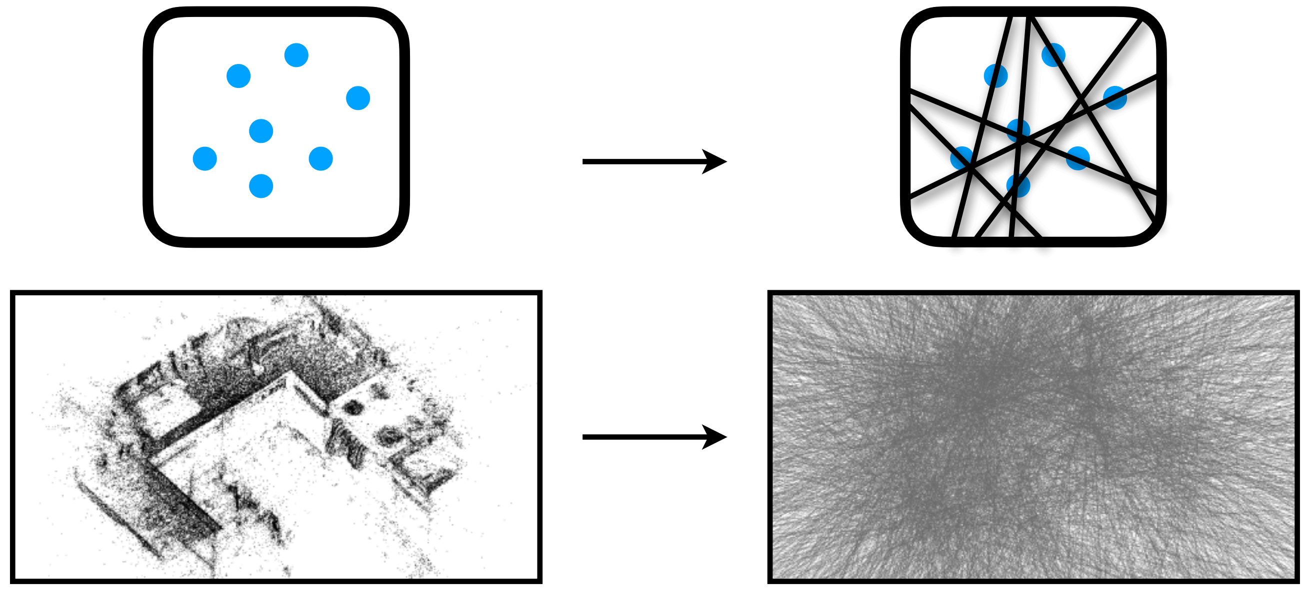


Dusmanu et al. Privacy-Preserving Image Features via Adversarial Affine Subspace Embeddings, CVPR 2021

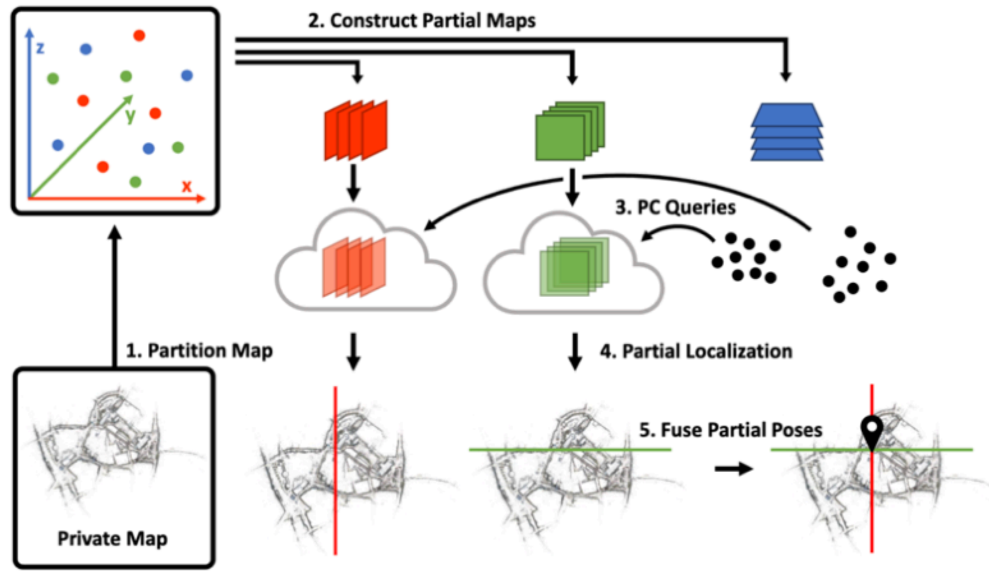


Ng et al. NinjaDesc, CVPR 2022

Map stored on cloud server

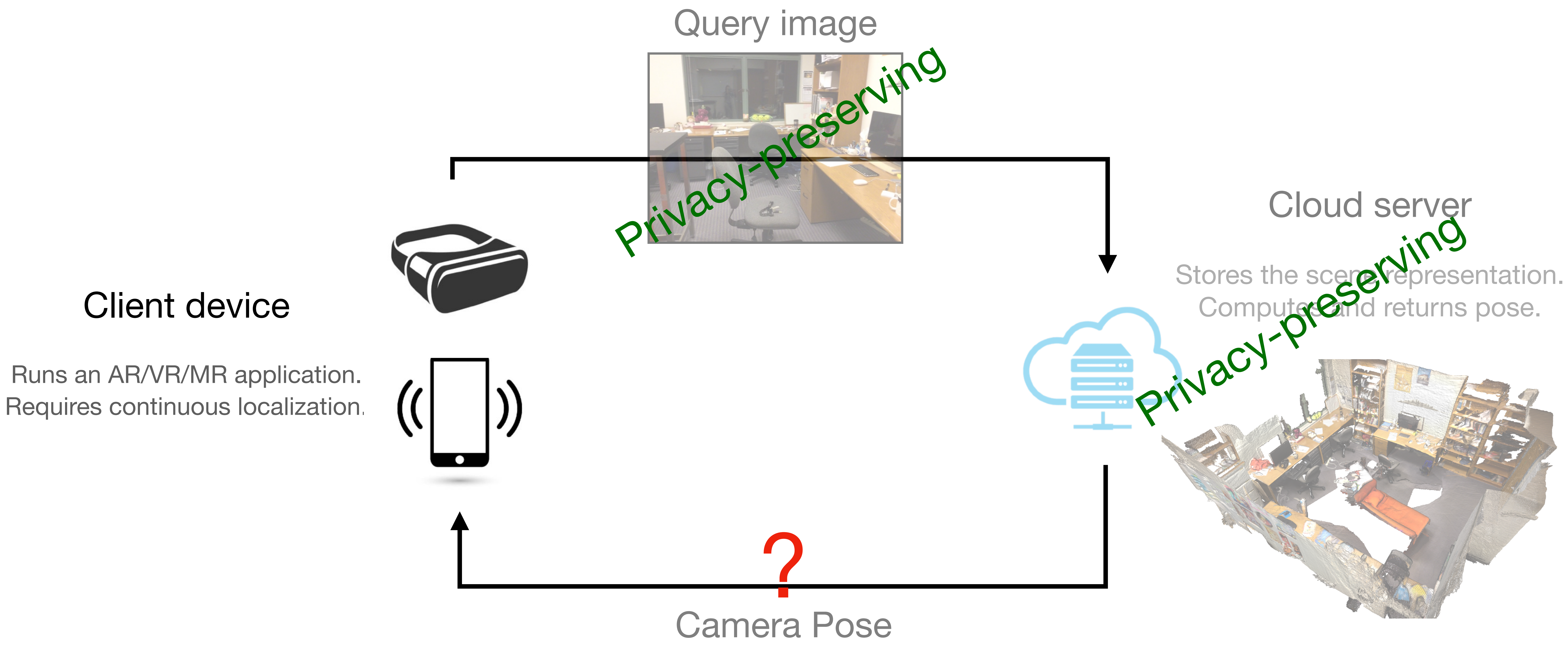


Speciale et al. Privacy Preserving Image-Based Localization, CVPR 2019



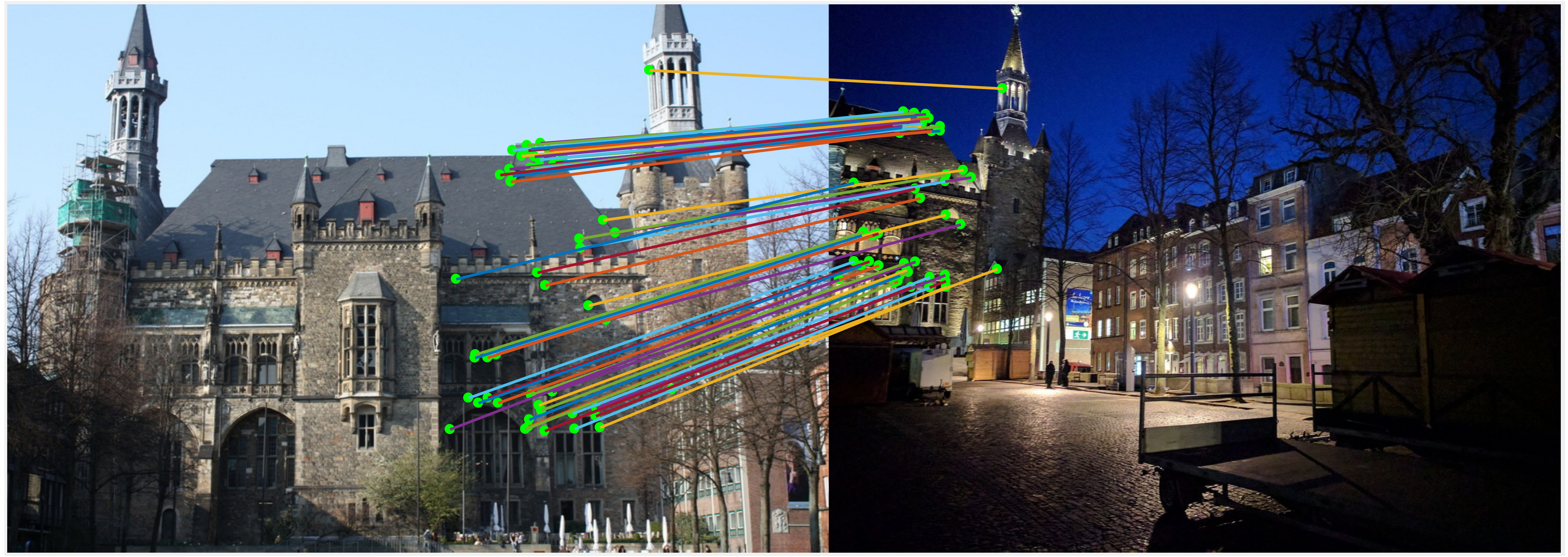
Geppert et al. Privacy Preserving Partial Localization, CVPR 2022

Can camera poses leak private information?



Motivation

Modern localization pipelines designed to maximise robustness!



Example from “D2-Net-A Trainable CNN for Joint Detection and Description of Local Features” Dusmanu et al. CVPR 2019

Enough matches to localize images of different object instances across different scenes!

Matches between 2 very different bicycles in different scenes.

Matches between different bookshelves in two different scenes.

Outline

Query image

Client device
Adversary

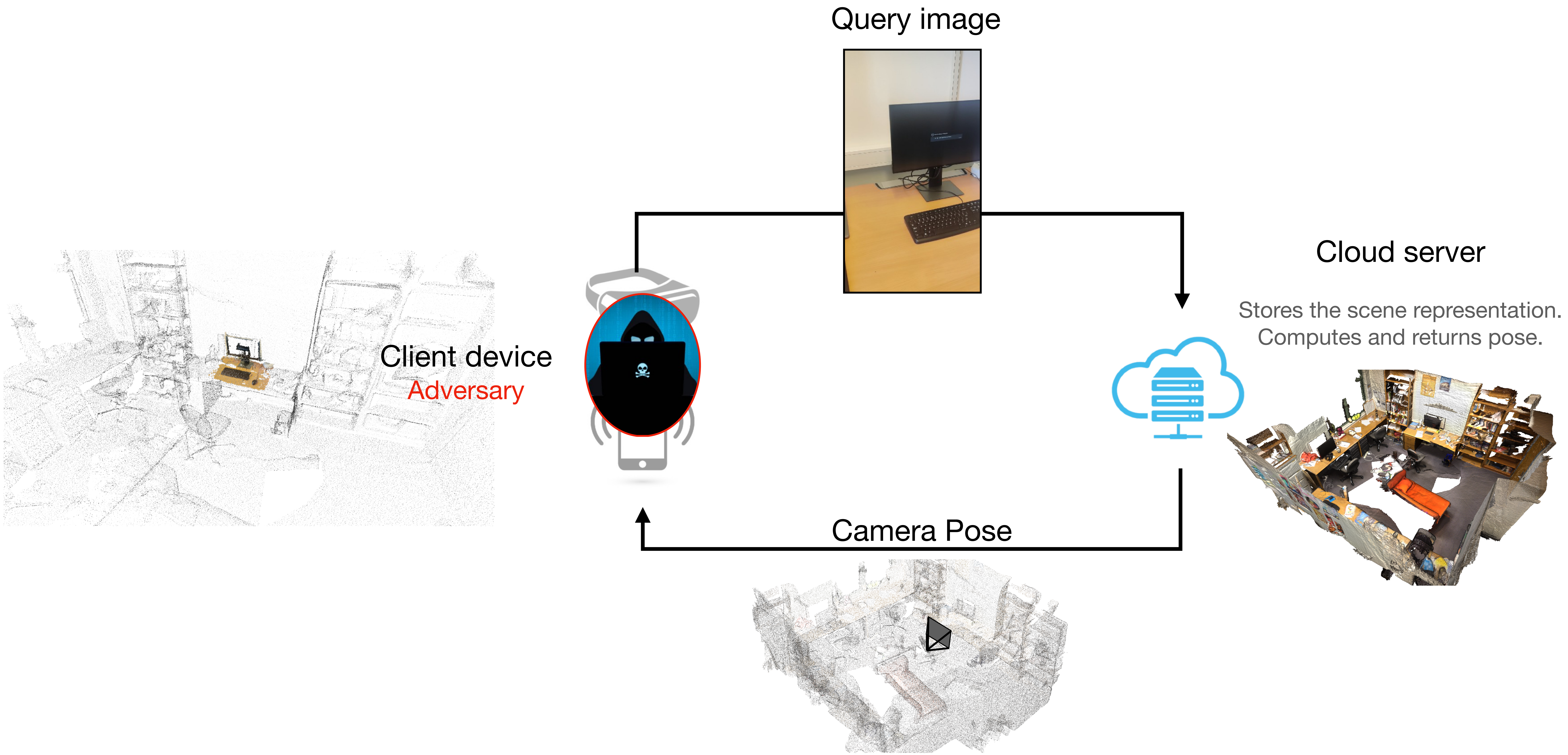
Cloud server

Stores the scene representation.
Computes and returns pose.

Camera Pose

Just by using these, the attacker can infer approximate scene layout!

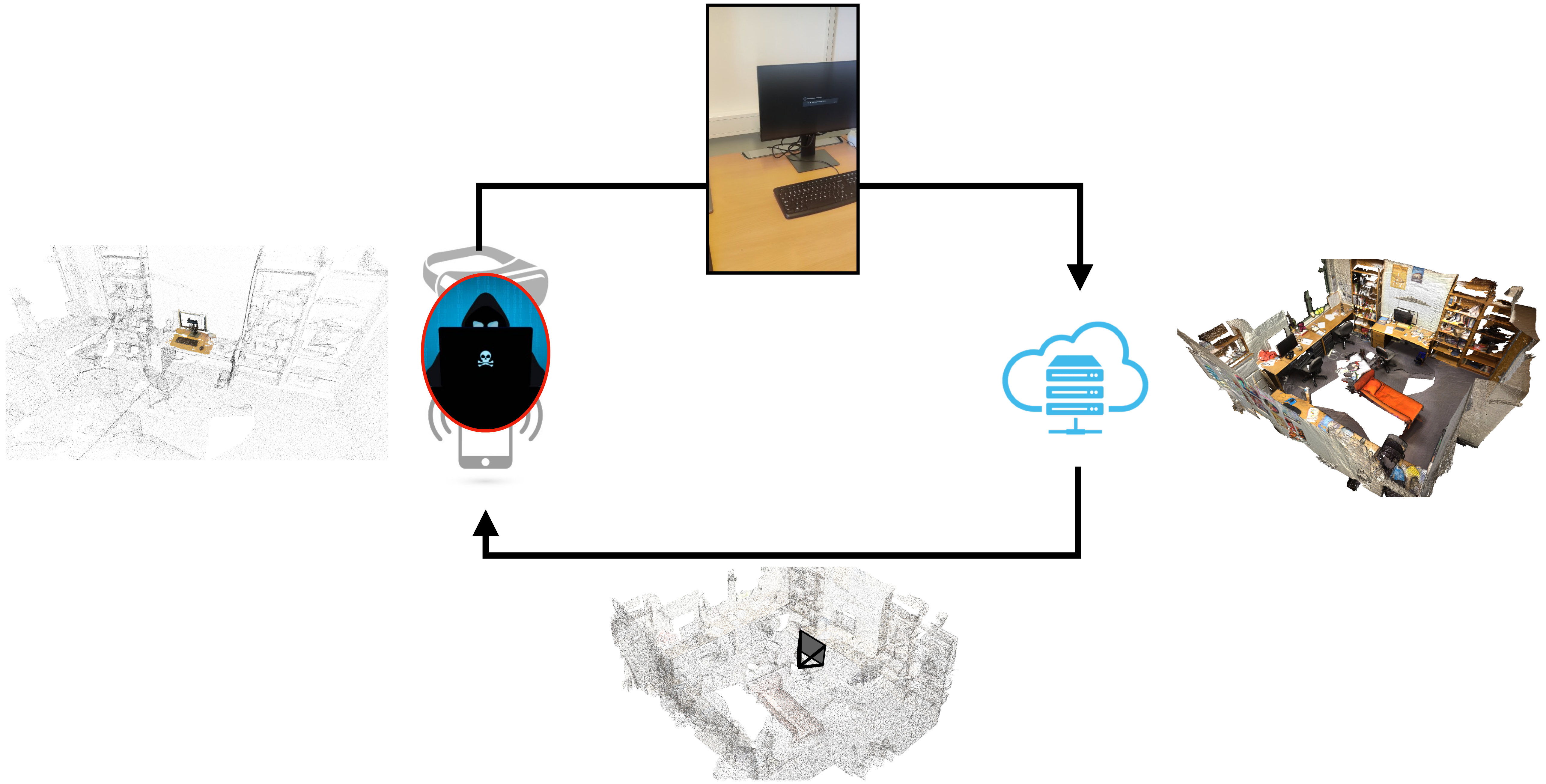
Simplest attack



Simplest attack - Challenges

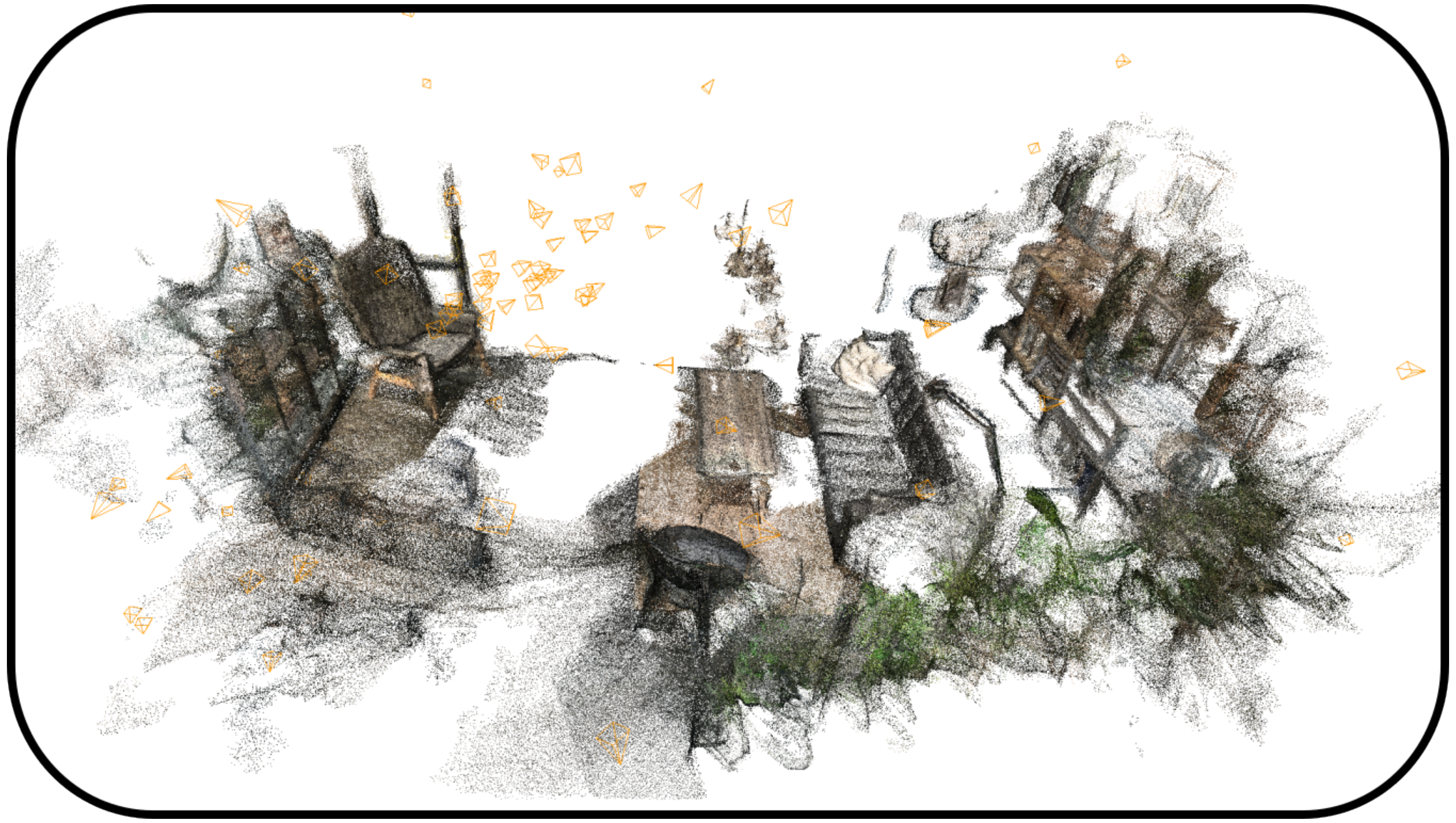
1. Every image gets a pose - cannot decide which object is present and which isn't.

2. Returned pose can be quite noisy (far from object) - incorrect positioning.



Using multi-view images

Suggestion : Use information from multiple images of each object taken from different view points

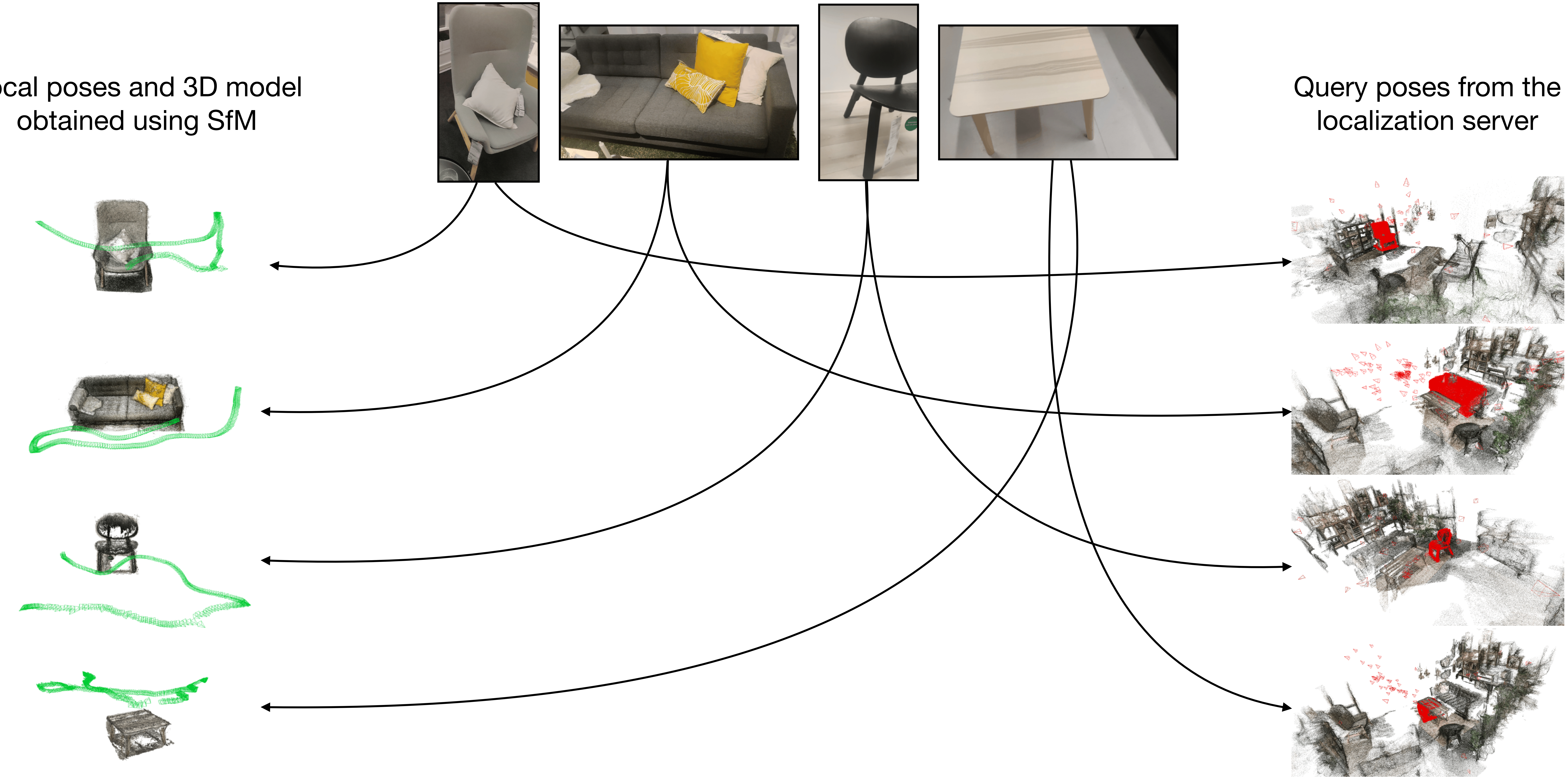


1. Some of the viewpoints would align well with the scene - allow correctly positioning - Challenge 2.
2. Distribution of the obtained poses can allow to decide if the object is present or not - Challenge 1.

Attack pipeline

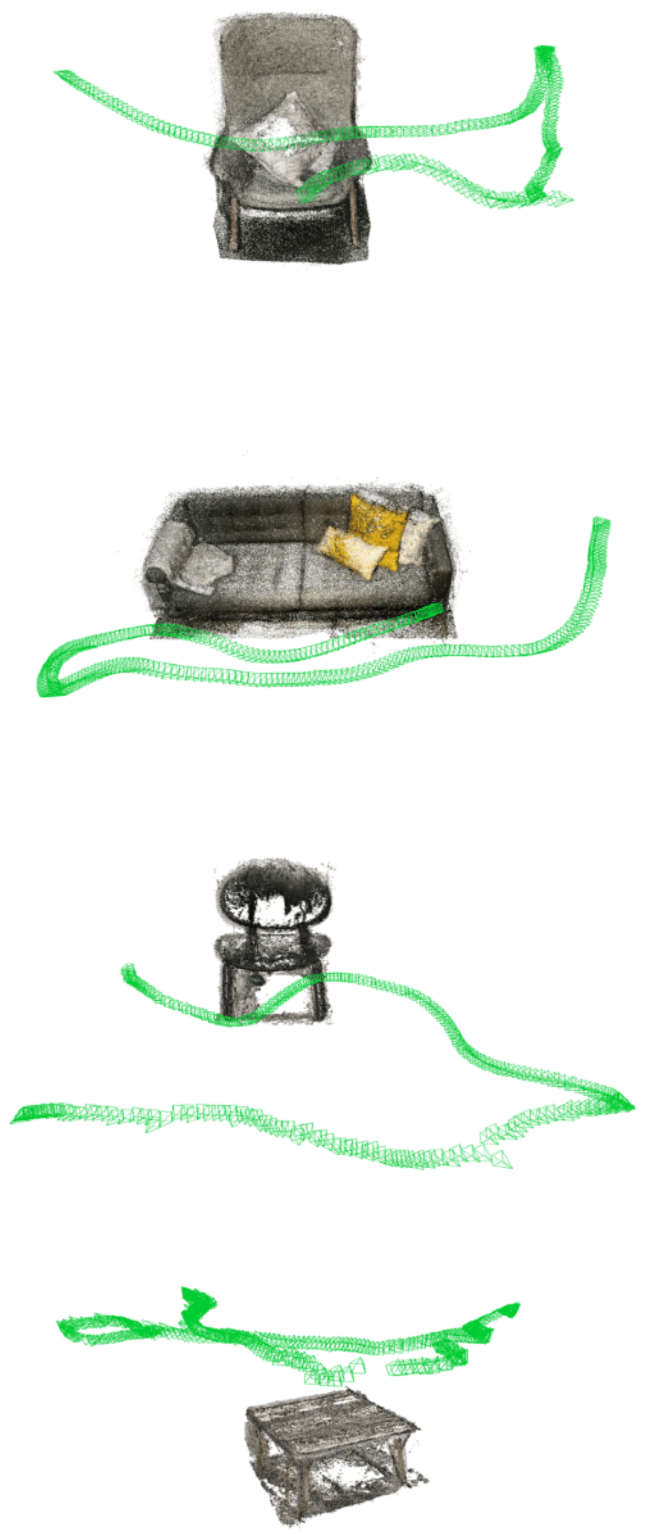
Local poses and 3D model obtained using SfM

Query poses from the localization server

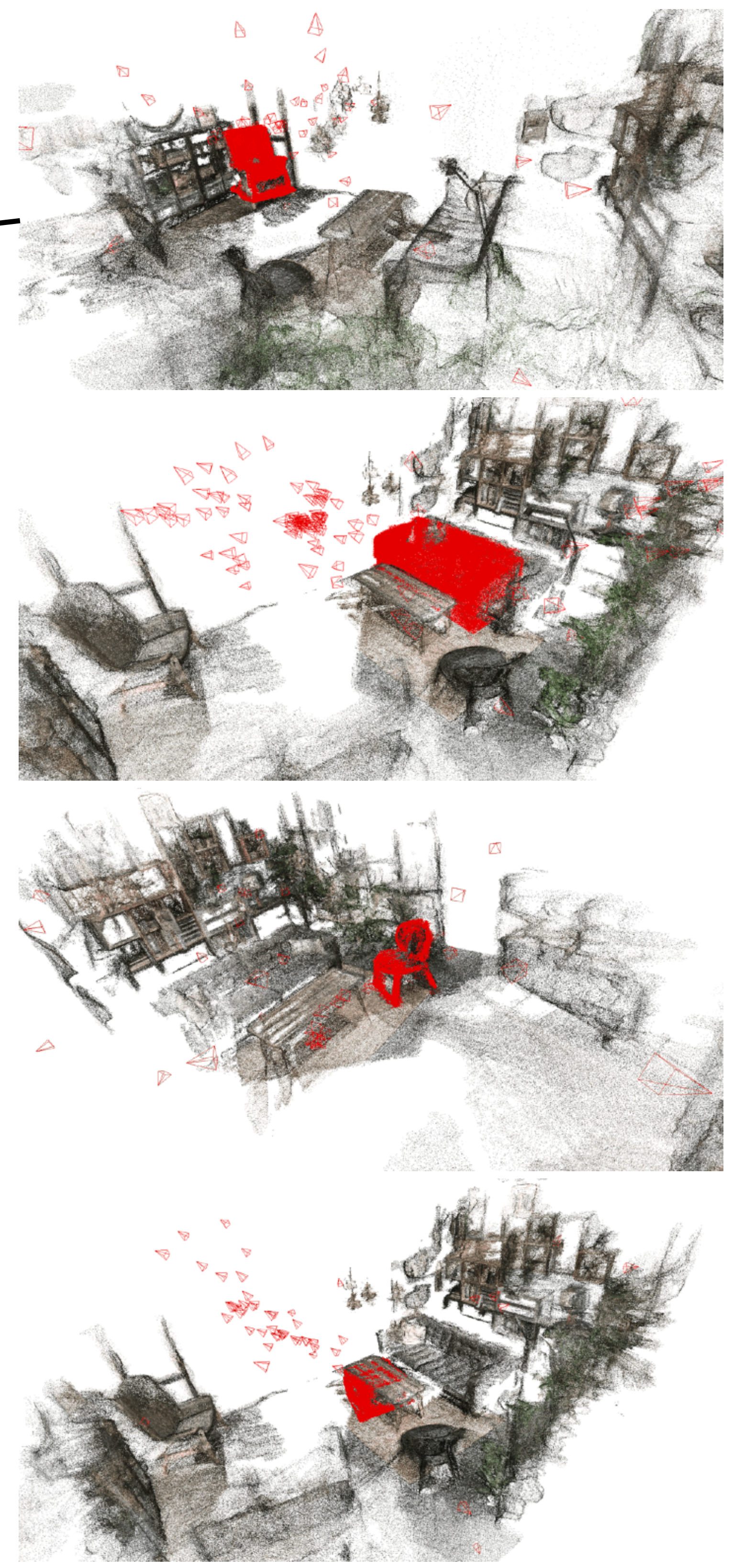


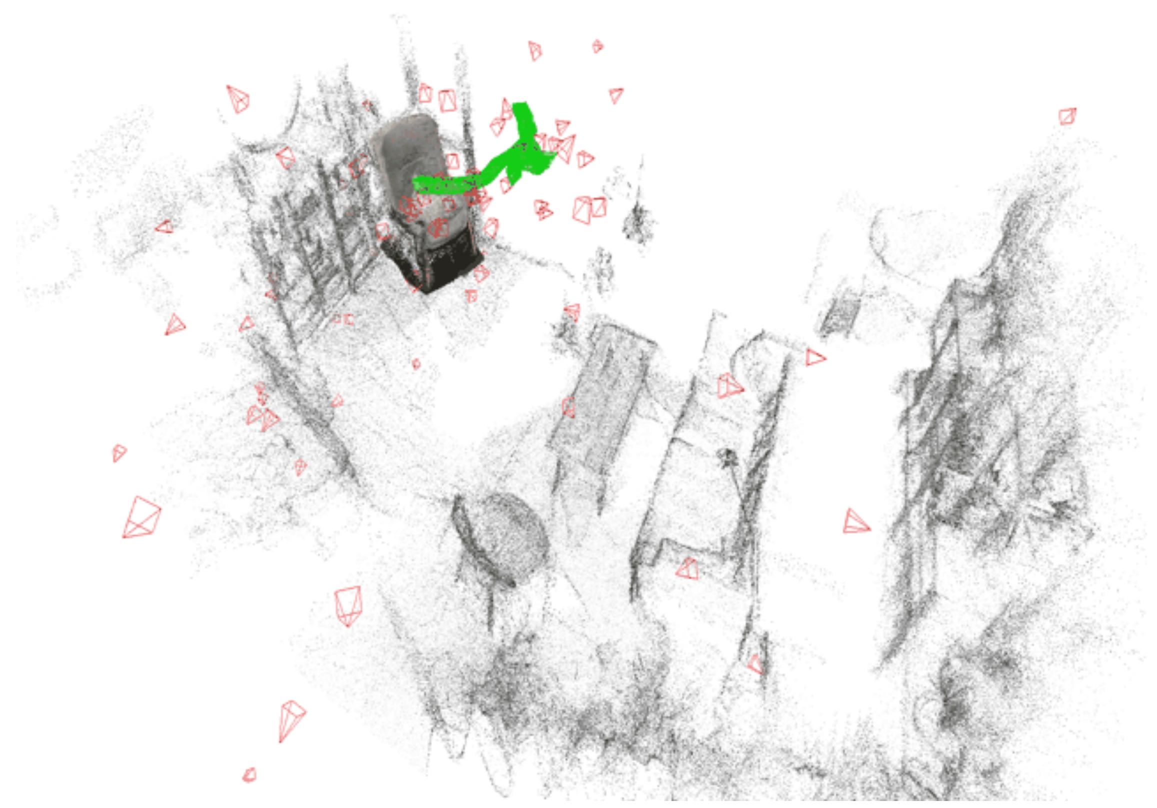
Attack pipeline

Local poses and 3D model obtained using SfM



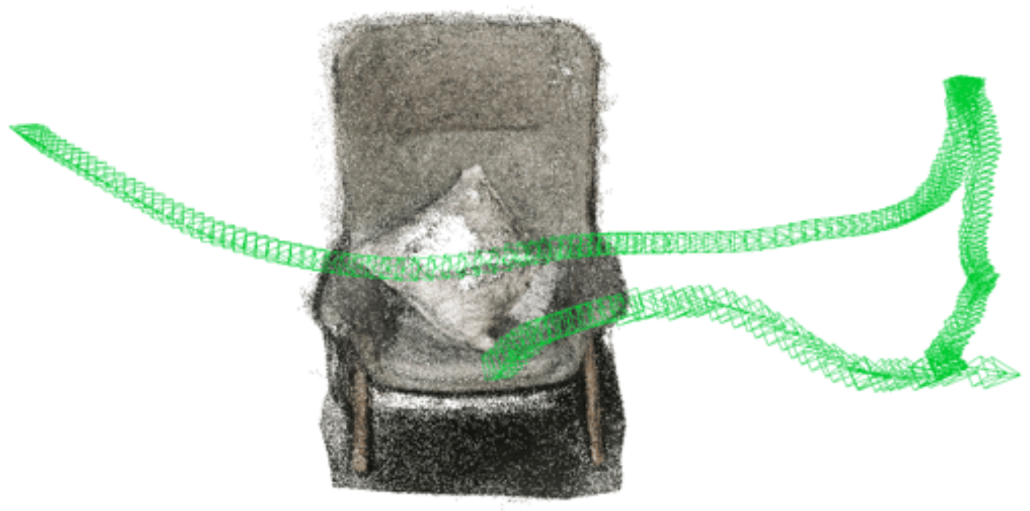
Query poses from the localization server



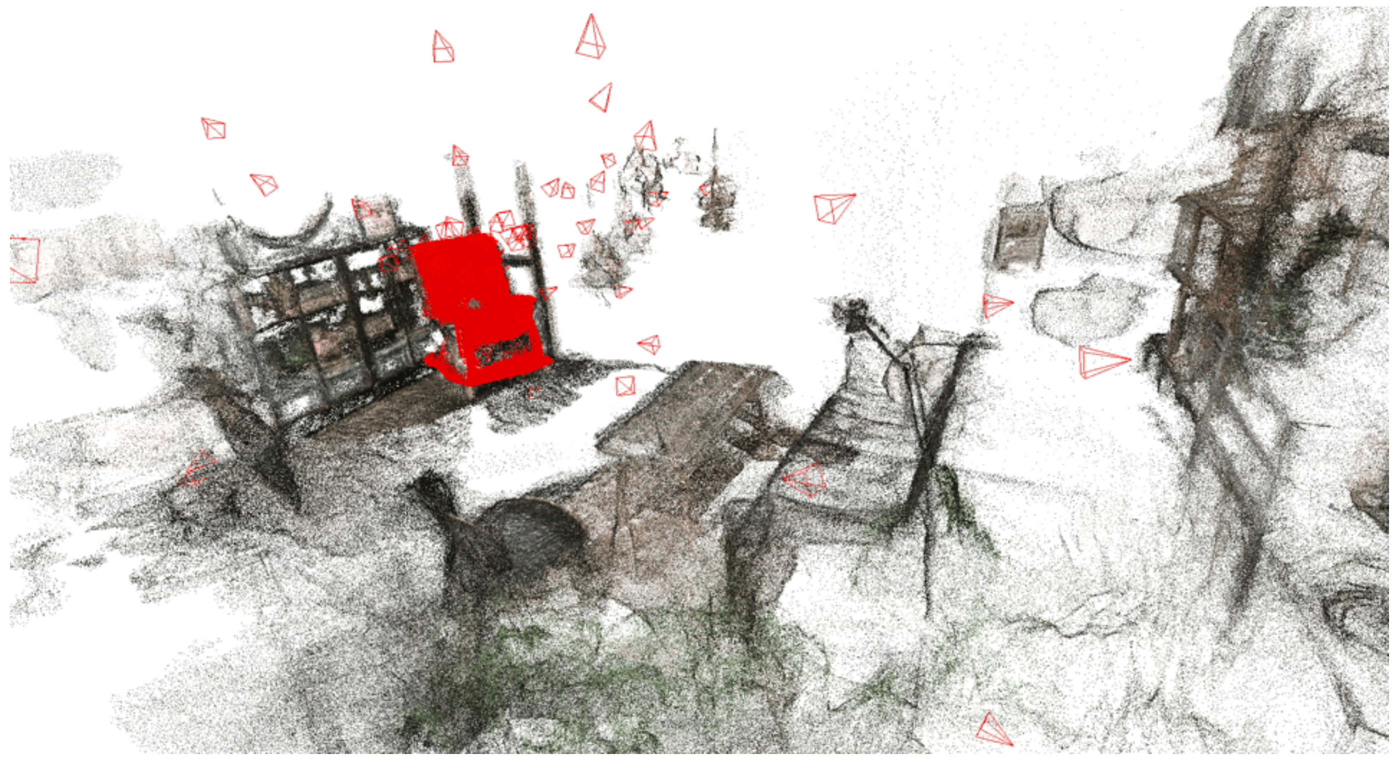


Position object by aligning poses

Robust pose set alignment



Local poses and 3D model obtained using SfM



Poses from querying the localization server

Algorithm 1 Best single camera based alignment between sets of poses

Input $\mathbf{P}_o = \{[\mathbf{R}_i|\mathbf{t}_i]\}, \hat{\mathbf{P}}_o = \{[\hat{\mathbf{R}}_i|\hat{\mathbf{t}}_i]\}, \delta_r, \delta_t$
Output $\mathbf{R}_{best}, \mathbf{t}_{best}, \epsilon$

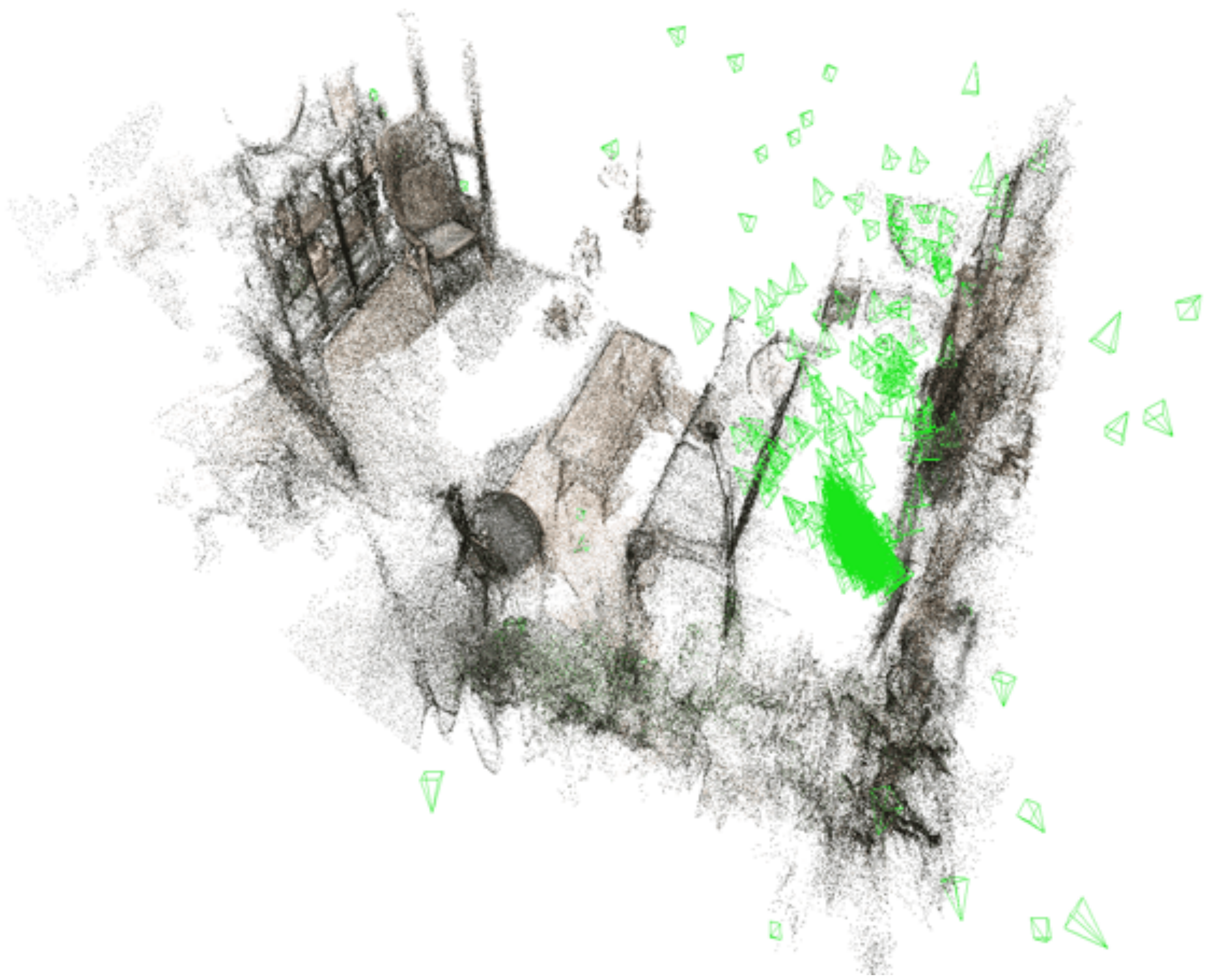
procedure GET-BEST-ALIGNMENT
  N $\leftarrow |\mathbf{P}_o|$
  `Inliers_best` $\leftarrow \phi$
  for `i = 1 to N` **do**
    $\mathbf{R}_{est} \leftarrow \hat{\mathbf{R}}_i^\top \mathbf{R}_i$
    $\mathbf{t}_{est} \leftarrow \hat{\mathbf{R}}_i^\top (\mathbf{t}_i - \hat{\mathbf{t}}_i)$
    `Inliers` $\leftarrow \phi$
    for `j = 1 to N` **do**
      $\Delta_r \leftarrow \angle(\mathbf{R}_j \mathbf{R}_{est}^\top \hat{\mathbf{R}}_j^\top)$
      $\Delta_t \leftarrow ||\hat{\mathbf{R}}_j^\top \hat{\mathbf{t}}_j - \mathbf{R}_{est} \mathbf{R}_j^\top \mathbf{t}_j + \mathbf{t}_{est})||$
      if $\Delta_r < \delta_r$ **and** $\Delta_t < \delta_t$ **then**
        `Inliers` $\leftarrow$ `Inliers` $\cup$ `{j}`
    if `|Inliers|` $>$ `|Inliers_best|` **then**
      `Inliers_best` $\leftarrow$ `Inliers`
  $\epsilon \leftarrow$ `|Inliers_best|/N`
  $\mathbf{R}_{best}, \mathbf{t}_{best} \leftarrow$ **Average(`Inliers_best`)**

For each corresponding camera, compute the relative motion and use that to transform all other cameras.

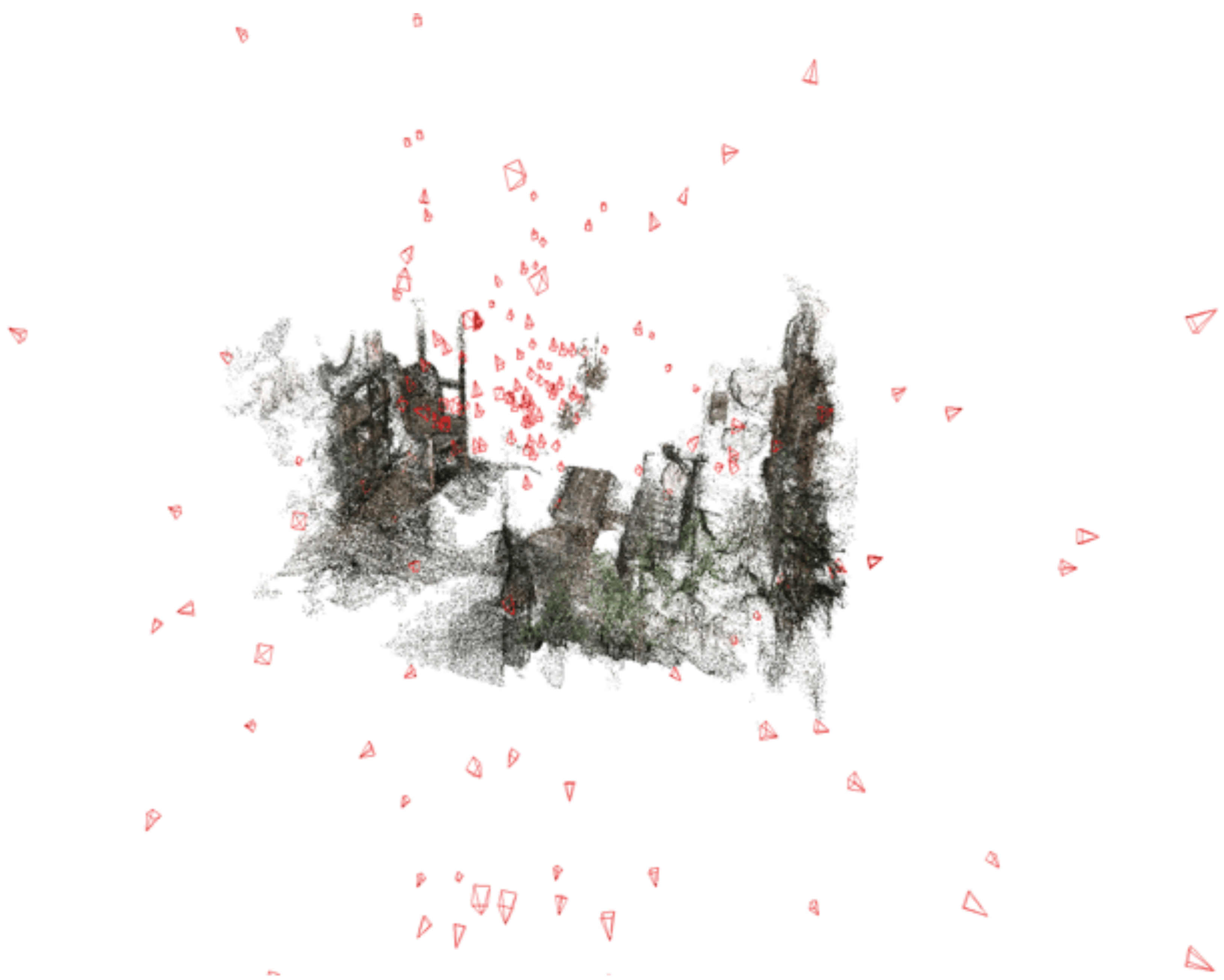
Check how well other cameras agree with this by counting inliers within some thresholds.

Average over the best set of inliers.

Decide Object Presence



Object present = Server Poses relatively consistent



Object Absent = Server poses distributed randomly.

Use inlier ratio from the pose-alignment algorithm as a proxy for how random the poses are.
Low inlier ratio = high randomness.

CHALMERS
FACULTY OF ELECTRICAL ENGINEERING CTU IN PRAGUE
CZECH INSTITUTE OF INFORMATICS ROBOTICS AND CYBERNETICS CTU IN PRAGUE
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Results - Datasets

Server maps

IKEA-Scenes

Sequences form 7 inspiration rooms taken at an IKEA store

Attack queries

IKEA-Objects

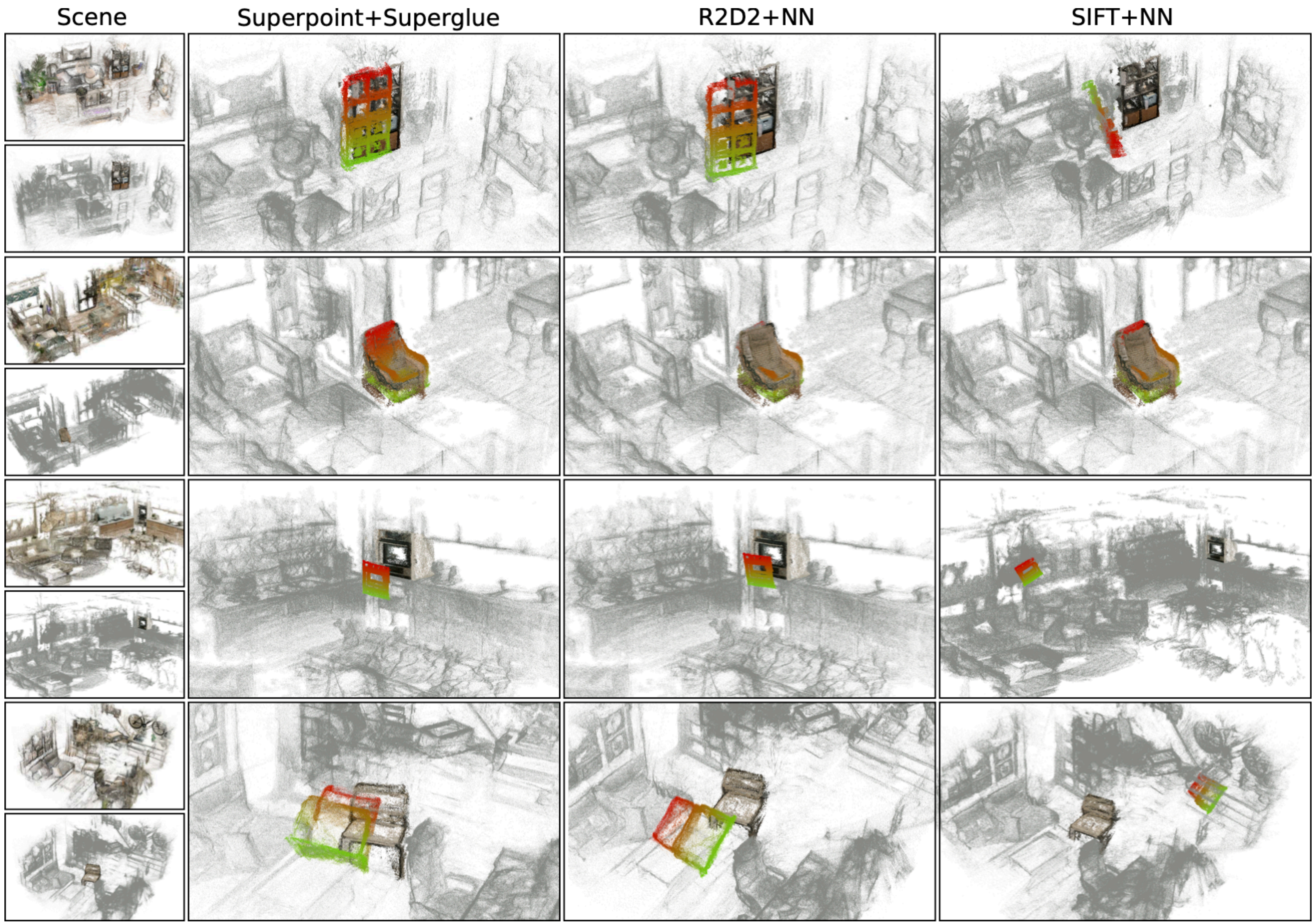
Sequences of similar objects as in IKEA Scenes in a different part of the store

Results - Different local features

- Localization server - Hloc[1]

- Comparison over following features:

  1. Superpoint[2] + Superglue[3]
  2. R2D2[4] + Nearest Neighbor
  3. SIFT[5] + Nearest Neighbor

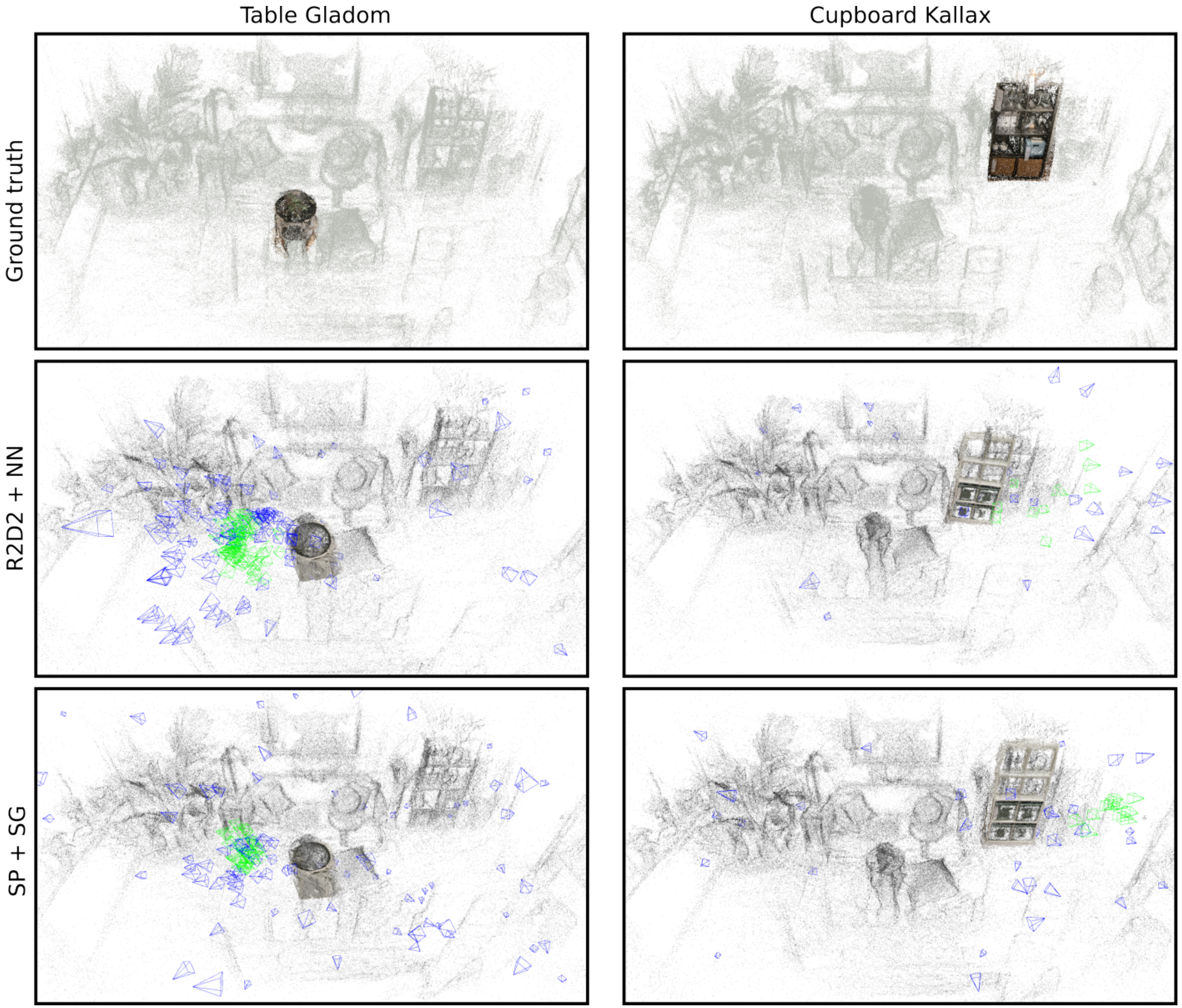


1. “From Coarse to Fine: Robust Hierarchical Localization at Large Scale” Sarlin et al. CVPR 2019
2. “SuperPoint: Self-Supervised Interest Point Detection and Description” DeTone et al. DLV4SLAM 2018 (CVPR workshop)
3. “SuperGlue:Learning Feature Matching with Graph Neural Networks” Sarlin et al. CVPR 2020
4. “R2D2:Repeatable and Reliable Detector and Descriptor” Revaud et al. NeurIPS 2019
5. “Distinctive Image Features from Scale-Invariant Keypoints” Lowe et al. IJCV 2004

Results - Qualitative alignment

Actual object in scene

Attack object



Actual object in scene

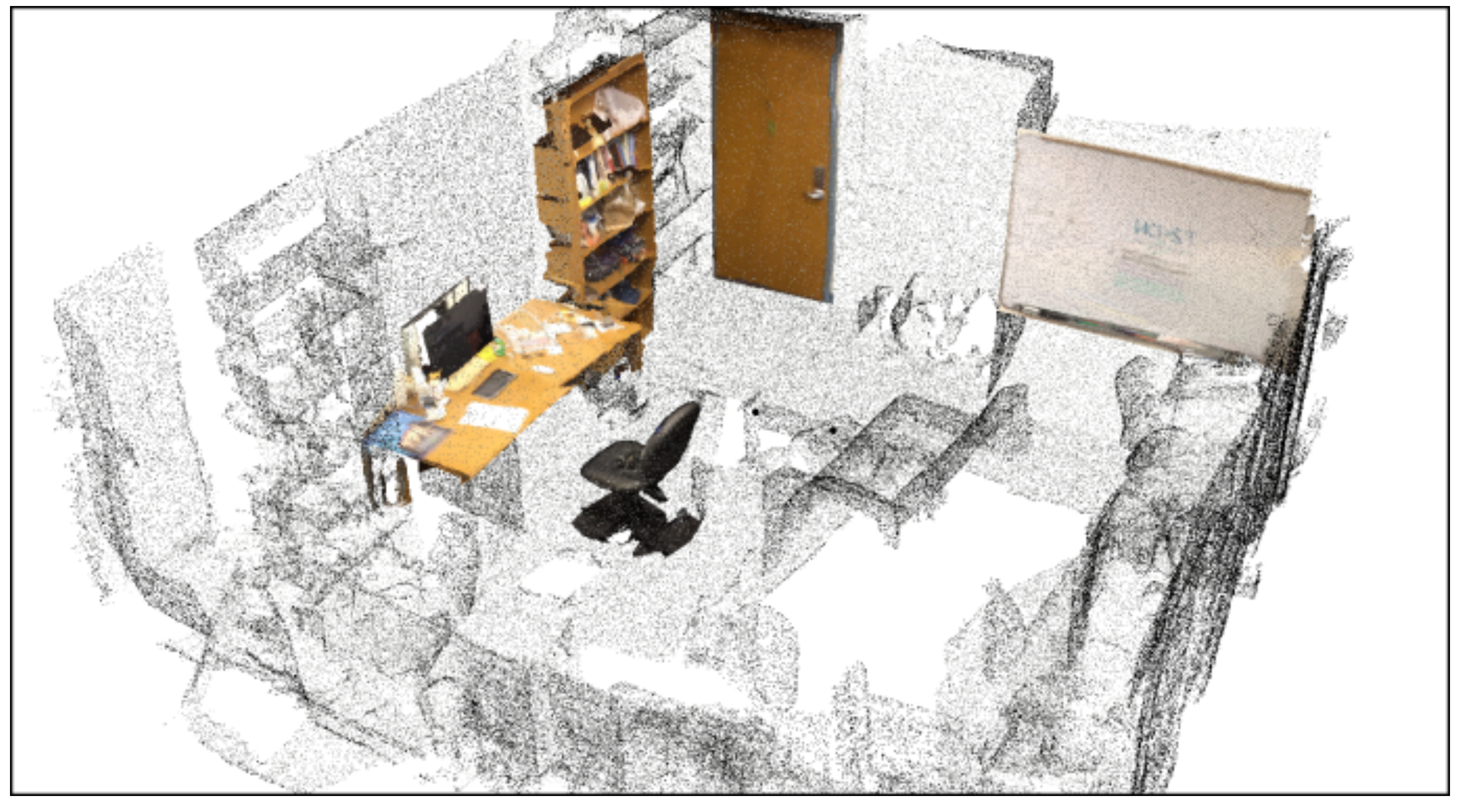
Attack object

Results - Datasets

Server maps

ScanNet[1]-Office

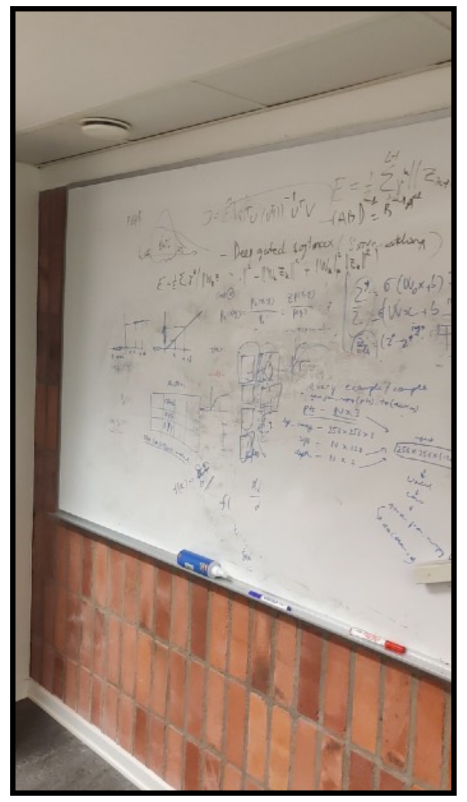
An office scene from the ScanNet dataset



Attack queries

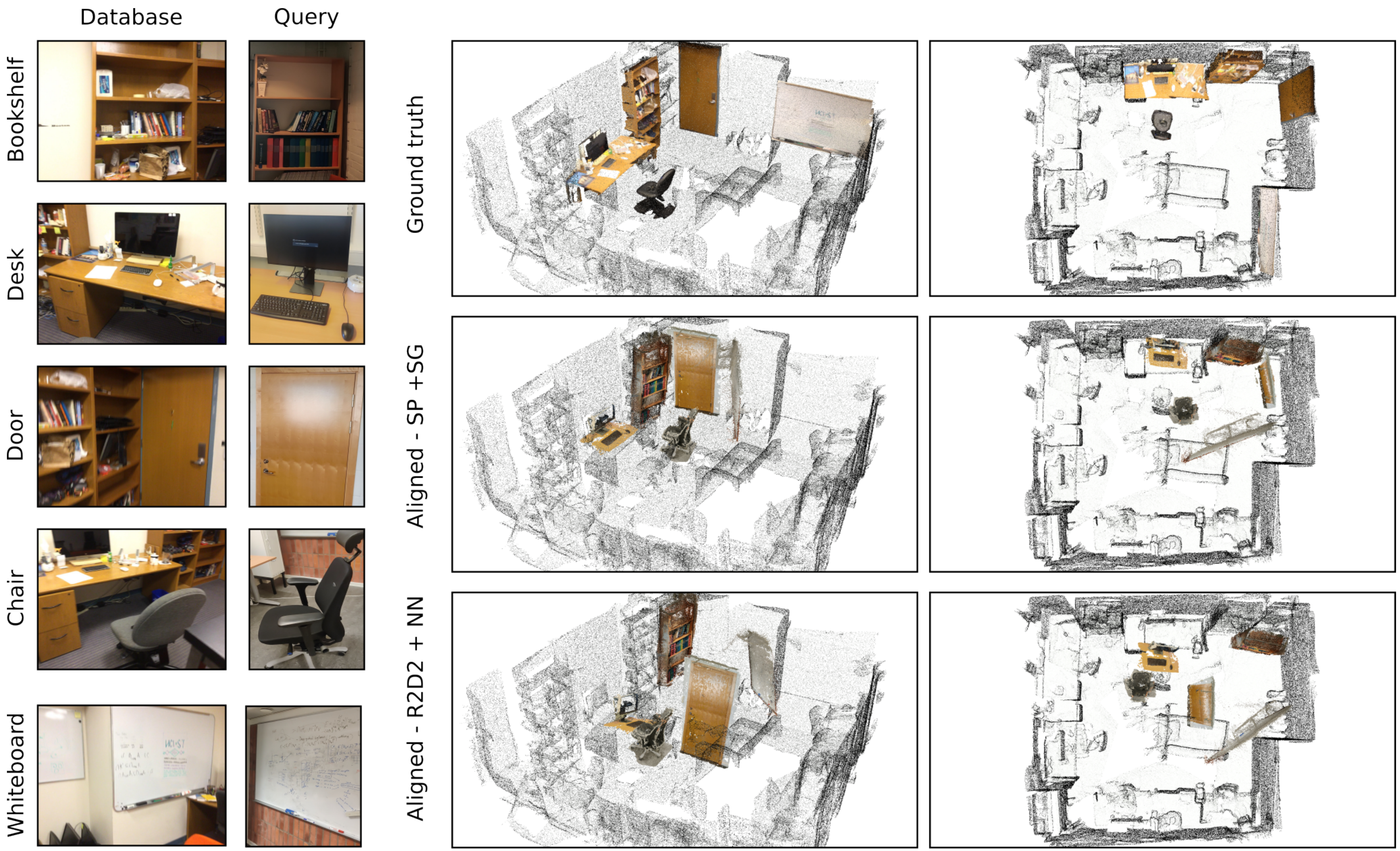
Office-Objects

Image sequences of office objects at our office



1. “ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes” Dai et al. CVPR 2017

Results - Qualitative alignment



Results - Deciding object presence

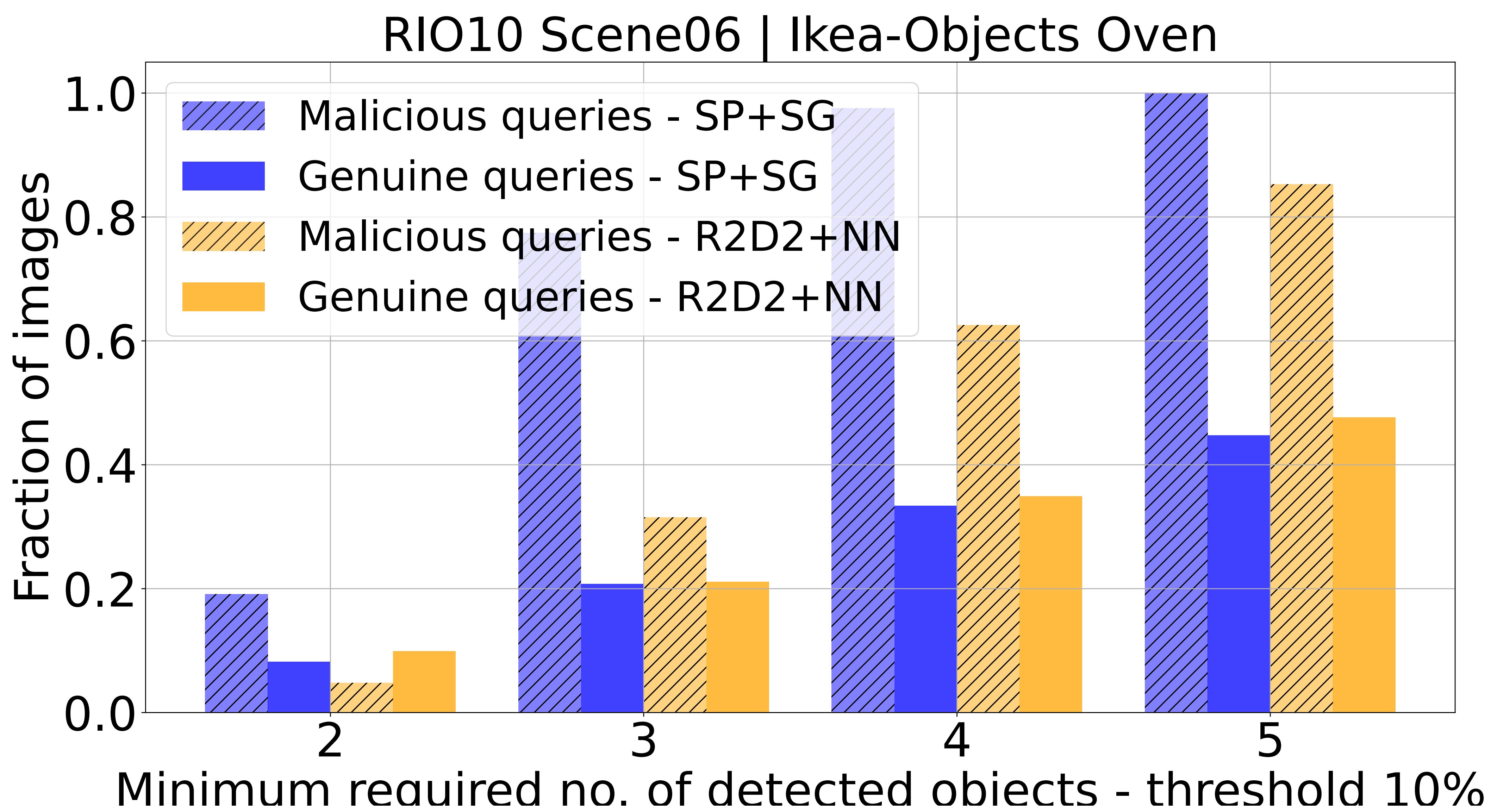
The decision method is not perfect - the task is difficult, however the underlying motivation is definitely holds.

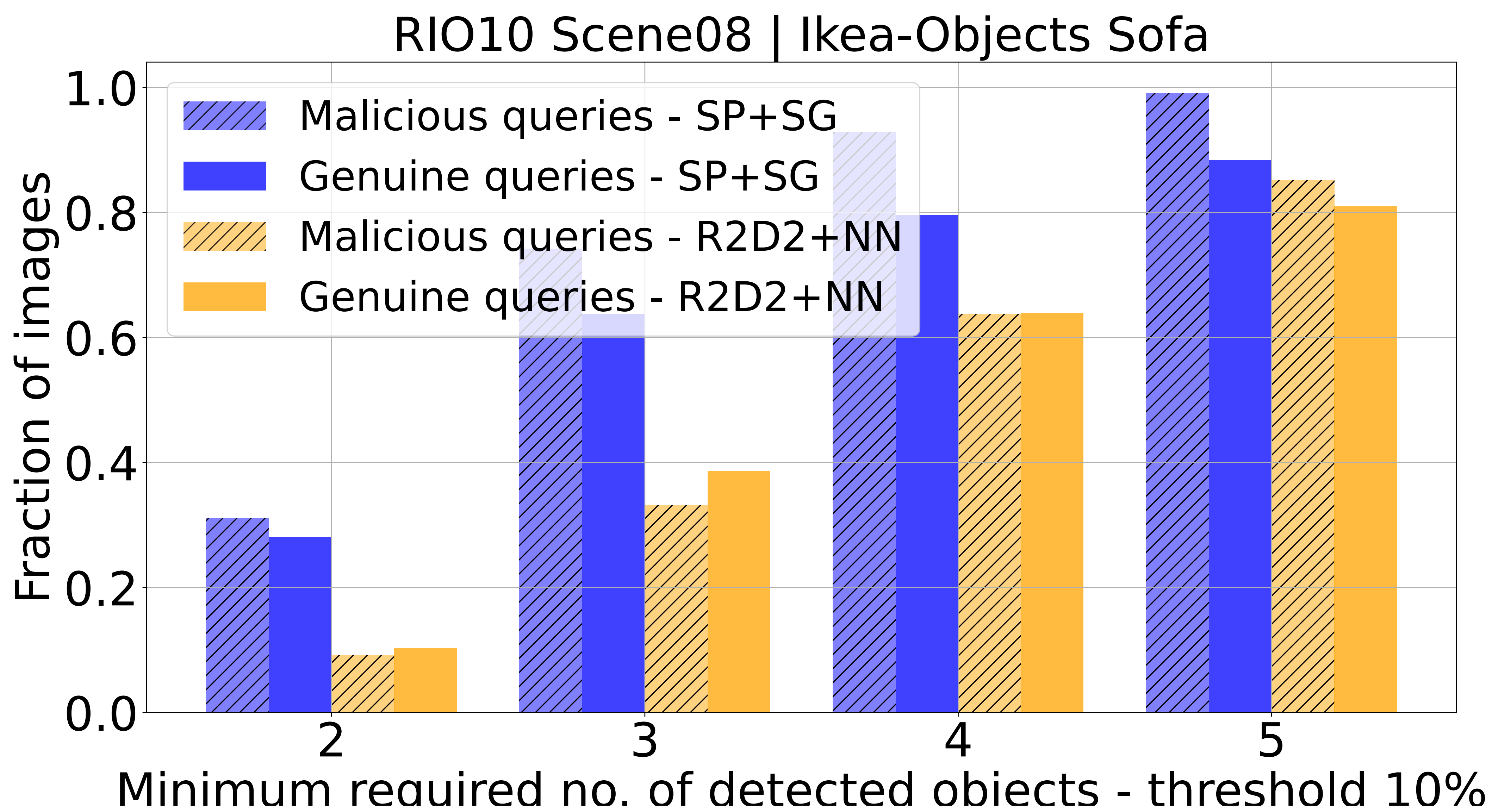
Scene	Objects present (recall)	Objects absent
IKEA Scene01	4/7	28/31
IKEA Scene02	4/10	21/28
IKEA Scene03	5/7	23/31
IKEA Scene04	3/5	28/33
IKEA Scene05	3/5	29/33
IKEA Scene06	2/5	27/33
IKEA Scene07	3/6	30/32

Discussion

Possible defence - Deny localization if 3D point inliers are predominantly from the same object.

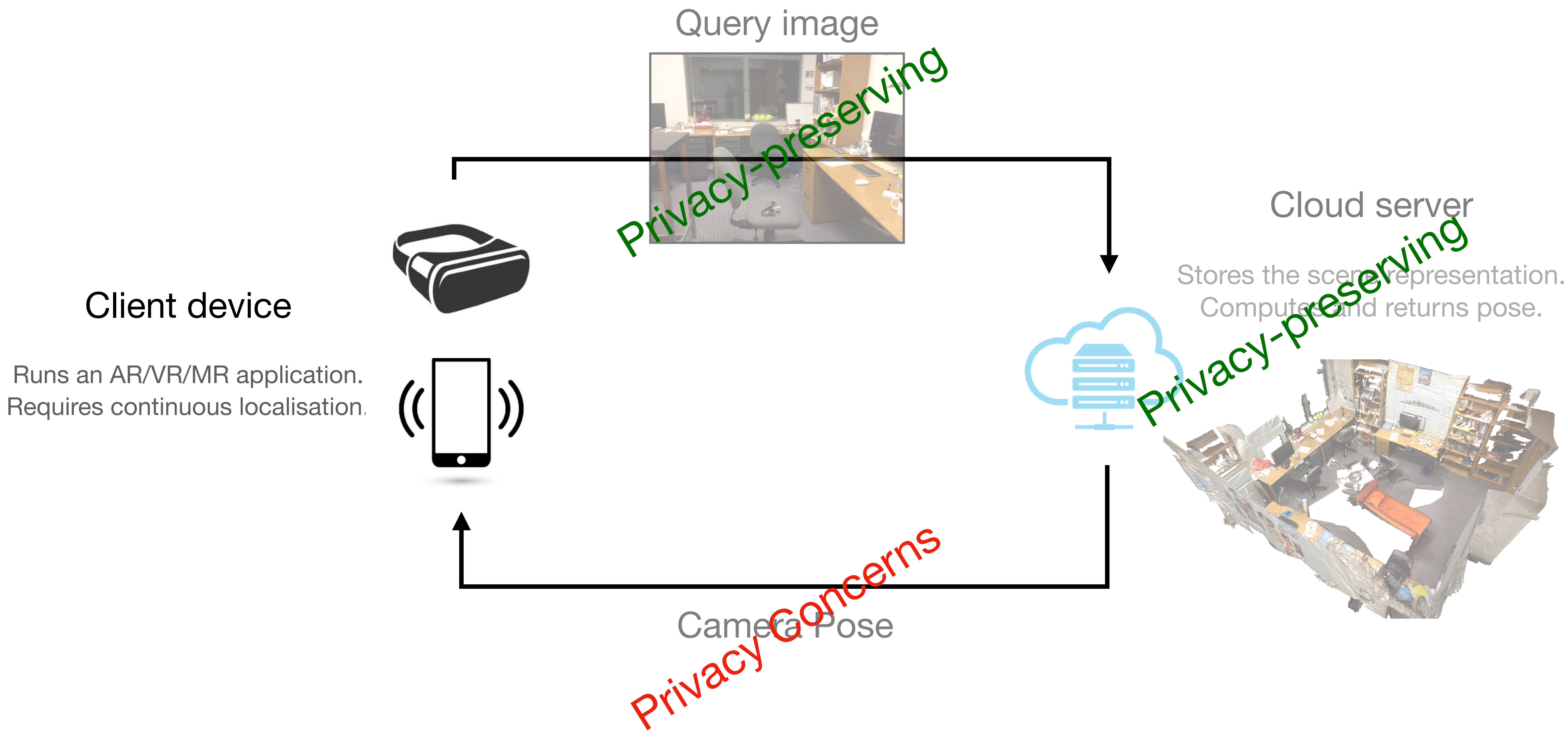
Results in denying several genuine queries as well.





Conclusion

1. A novel privacy-attack via camera poses in a client-server based localization-setup is presented.



Conclusion

1. A novel privacy-attack via camera poses in a client-server based localization-setup is presented.

2. A proof-of-concept attack pipeline is implemented to show the feasibility of the attack and 3 different local features are weighed on the scale of susceptibility to such an attack.

3. It is shown that it might not be trivial to develop a defence without affecting the robustness and reliability of the localization service.

4. More research in the direction of privacy-preserving localization is definitely needed.

WED-PM-074

Project Github page : https://github.com/kunalchelani/ObjectPositioningFromPoses