Privacy-Preserving Camera Localization

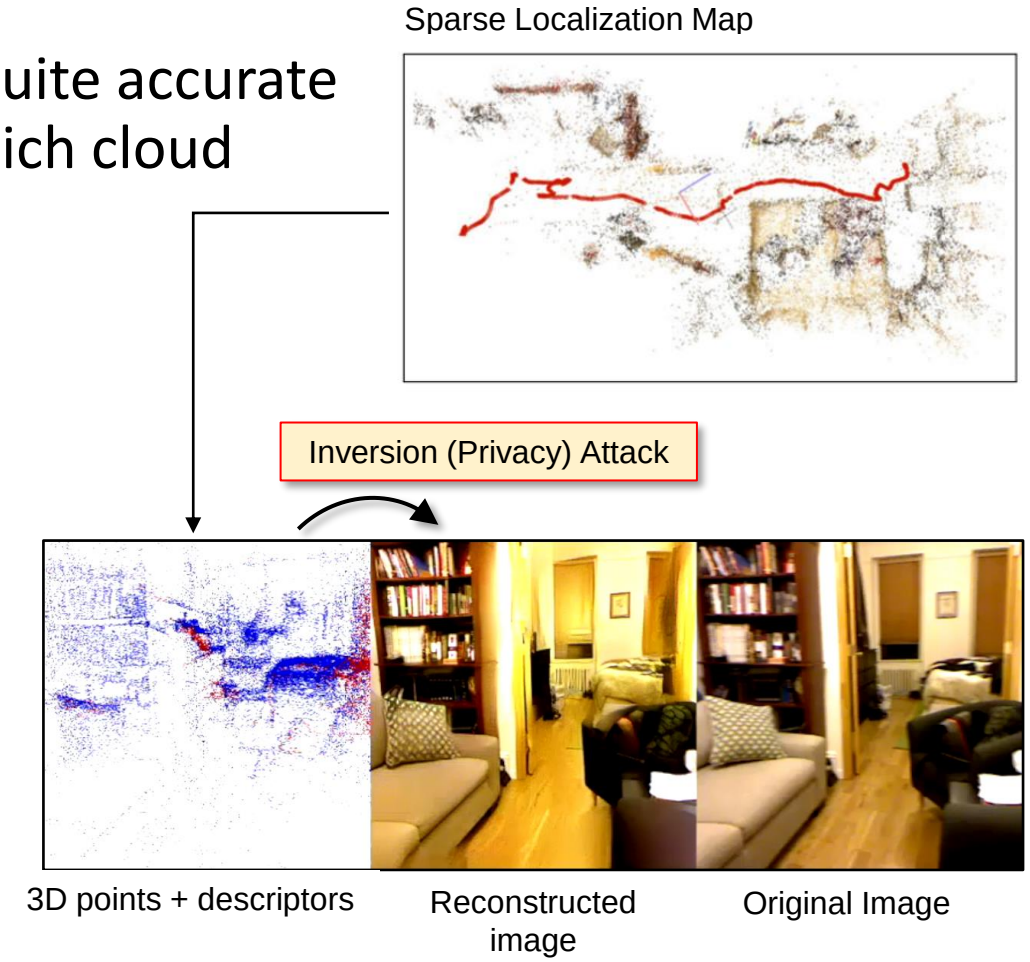
Sudipta N. Sinha

Large-Scale Visual Localization Tutorial

@ CVPR 2023, June 19, 2023

Motivation

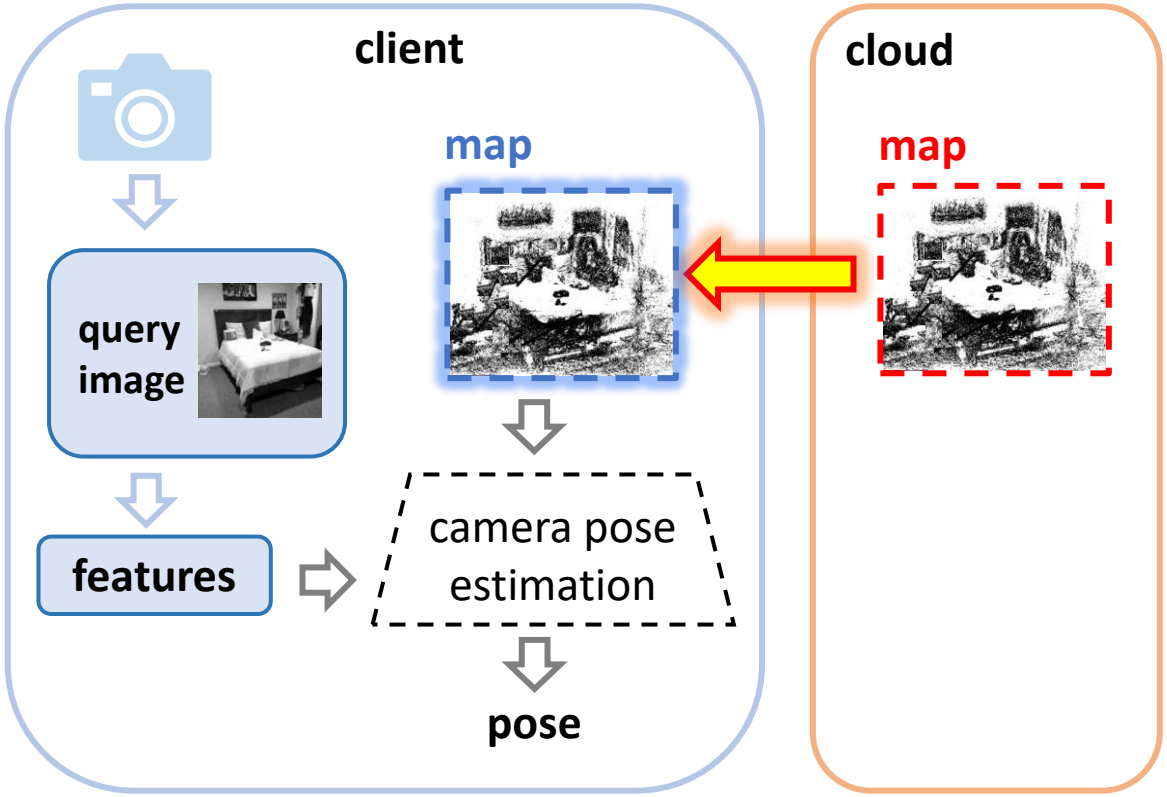
- Feature-based localization methods are often quite accurate but tend to have high storage needs; due to which cloud compute and storage might be needed.
- However, persistent storage of localization maps in the cloud could raise privacy concerns.
- It is possible to invert visual features and localization maps to reconstruct detailed images of the scene [Pittaluga+ 2019].
- The images could reveal sensitive information about the scene or subjects.



Pittaluga et al. 2019, "**Revealing Scenes by Inverting Structure from Motion Reconstructions**", in CVPR.

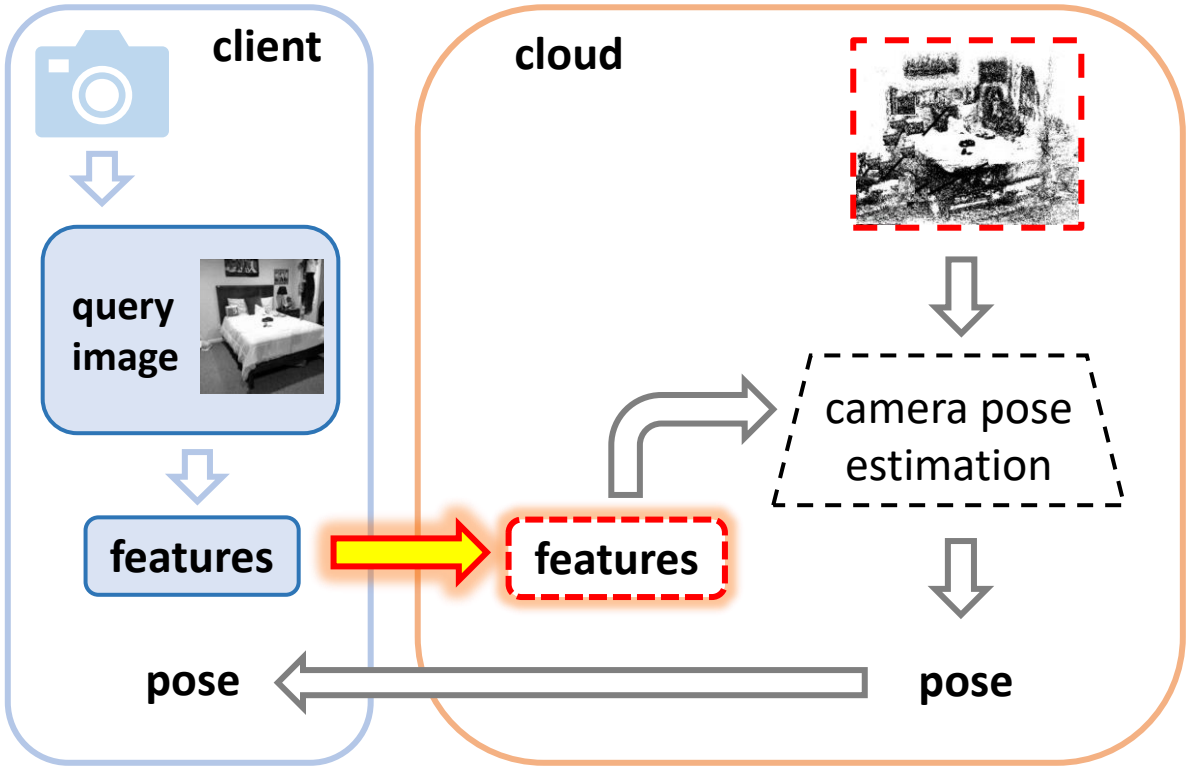
Two Different Settings

Client Setting



- Server shares map data with client devices
- Localization algorithms run locally

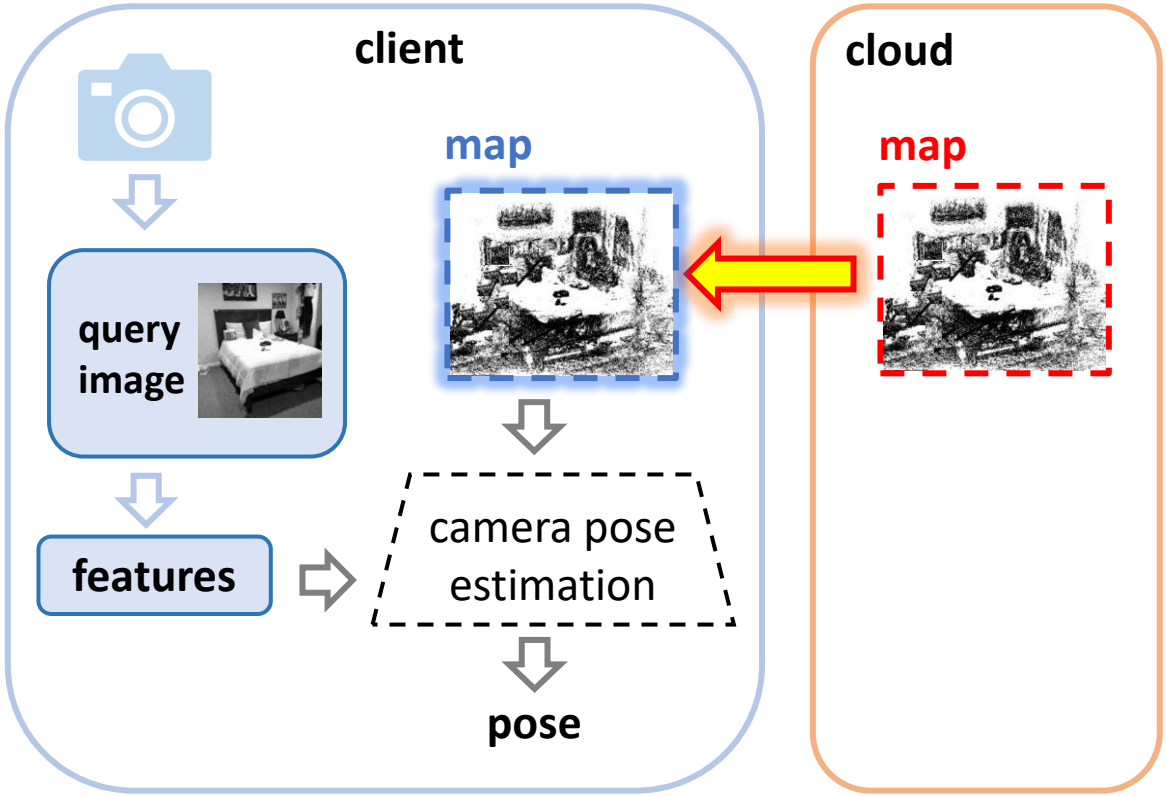
Server Setting



- Client shares image features with server
- Camera localization algorithms run in the cloud

Two Different Settings

Client Setting



- Server shares map data with client devices
- Localization algorithms run locally

Server Setting

- Client shares image features with server
- Camera localization algorithms run in the cloud

Privacy-Preserving Localization on the Client

Setting:

Attacker has access to the *map features*.

Objective:

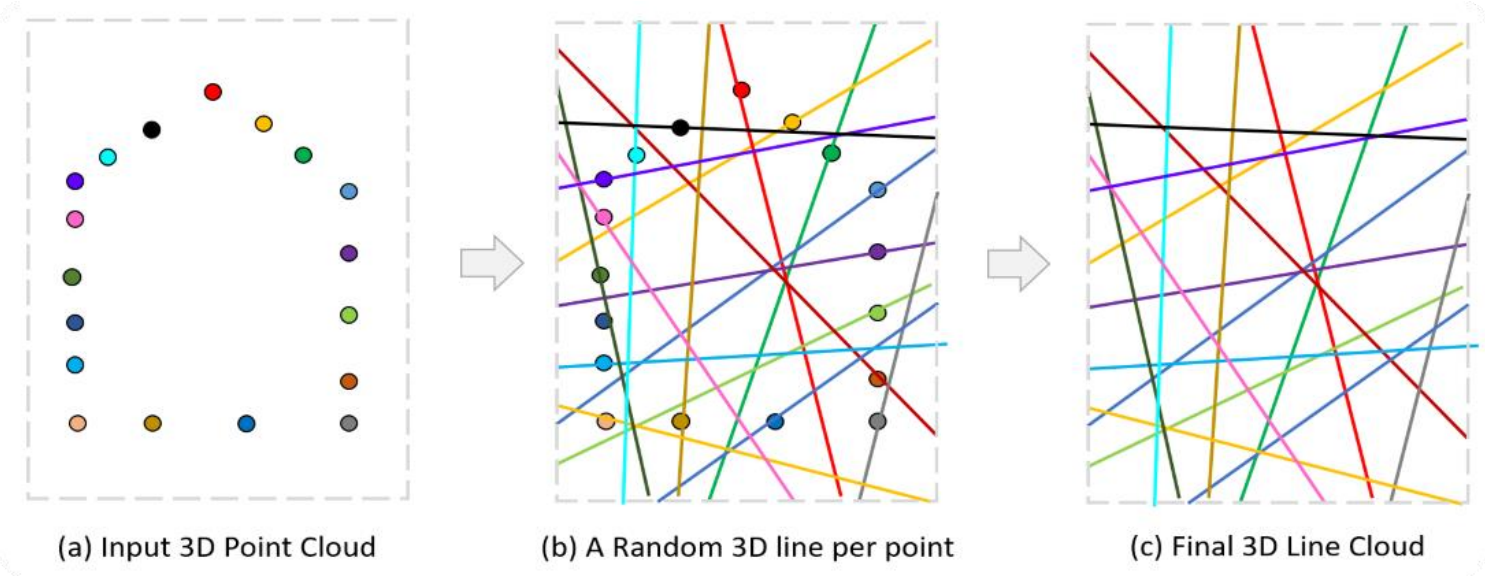
Compute the camera pose but also prevent the attacker from reconstructing images of the scene from the *map features.*

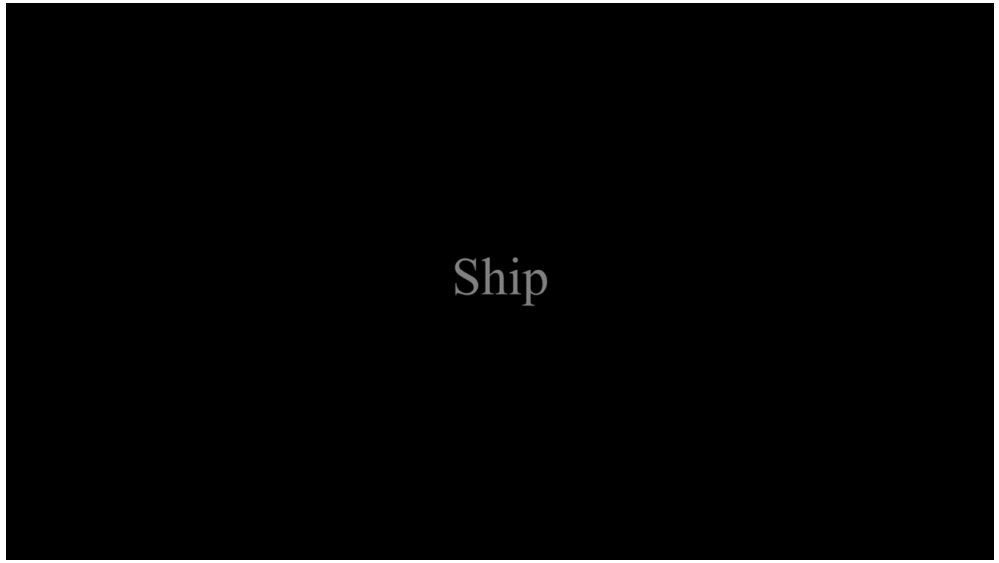
Known results and proposed methods:

- [Speciale *et al.* 2019] 3D line cloud-based localization ☺
- [Shibuya *et al.* 2020] 3D line cloud-based SLAM ☺
- [Chelani *et al.* 2021] Line clouds do not always obfuscate underlying scene structure ☹
- [Lee *et al.* 2023] *(CVPR)* Line clouds built using Paired-Point Lifting are secure! ☺

3D Line Cloud-based Localization

Speciale et al. 2019, "**Revealing Scenes by Inverting Structure from Motion Reconstructions**", in CVPR.





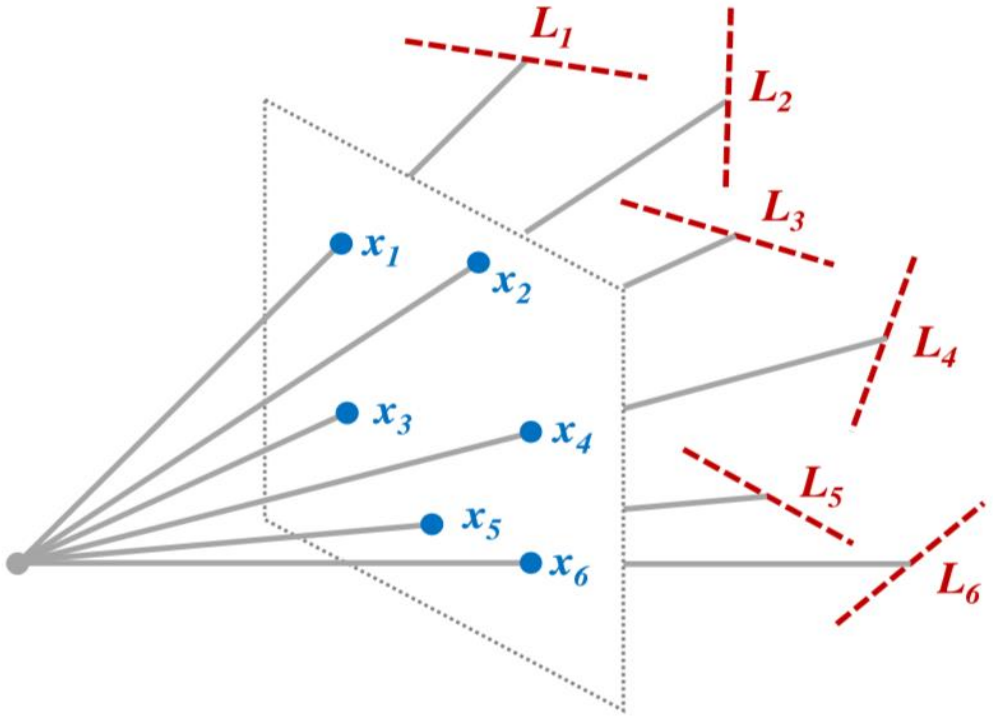
Main Idea:

- For each 3D point, pick a randomly oriented 3D line passing through the point.
- Then **discard** the 3D point.

3D Line Cloud-based Localization

Speciale et al. 2019, "**Revealing Scenes by Inverting Structure from Motion Reconstructions**", in CVPR.

point-to-line constraints ($p6L$)



At least six correspondences needed to compute pose

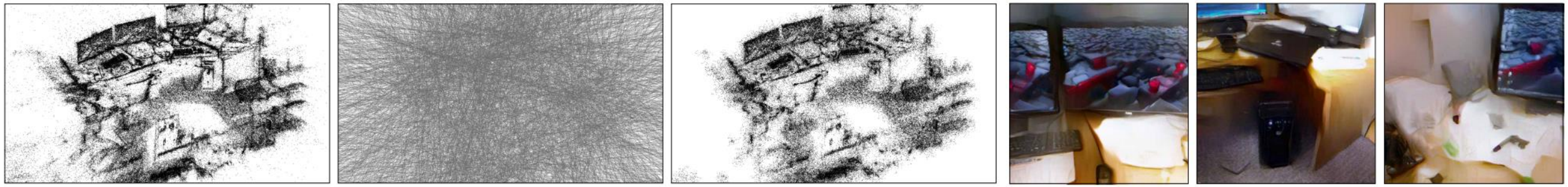
- Minimal problem same as the generalized relative pose problem [Stewenius et al. 2005].
- Explored variants of the method that use multiple images for pose estimation, assume known vertical direction, or known scale.
- Solvers for these variants already exist!

  [Nister+ 2007, Lee+ 2014, Stewenius+ 2005, Sweeney+ 2014, Sweeney+ 2015a, Sweeney+ 2015b]

- Uses minimal solvers along with RANSAC.

Line Clouds do not always obfuscate scene structure

Chelani et al. 2021, “How privacy-preserving are line clouds? recovering scene details from 3d lines”, CVPR.

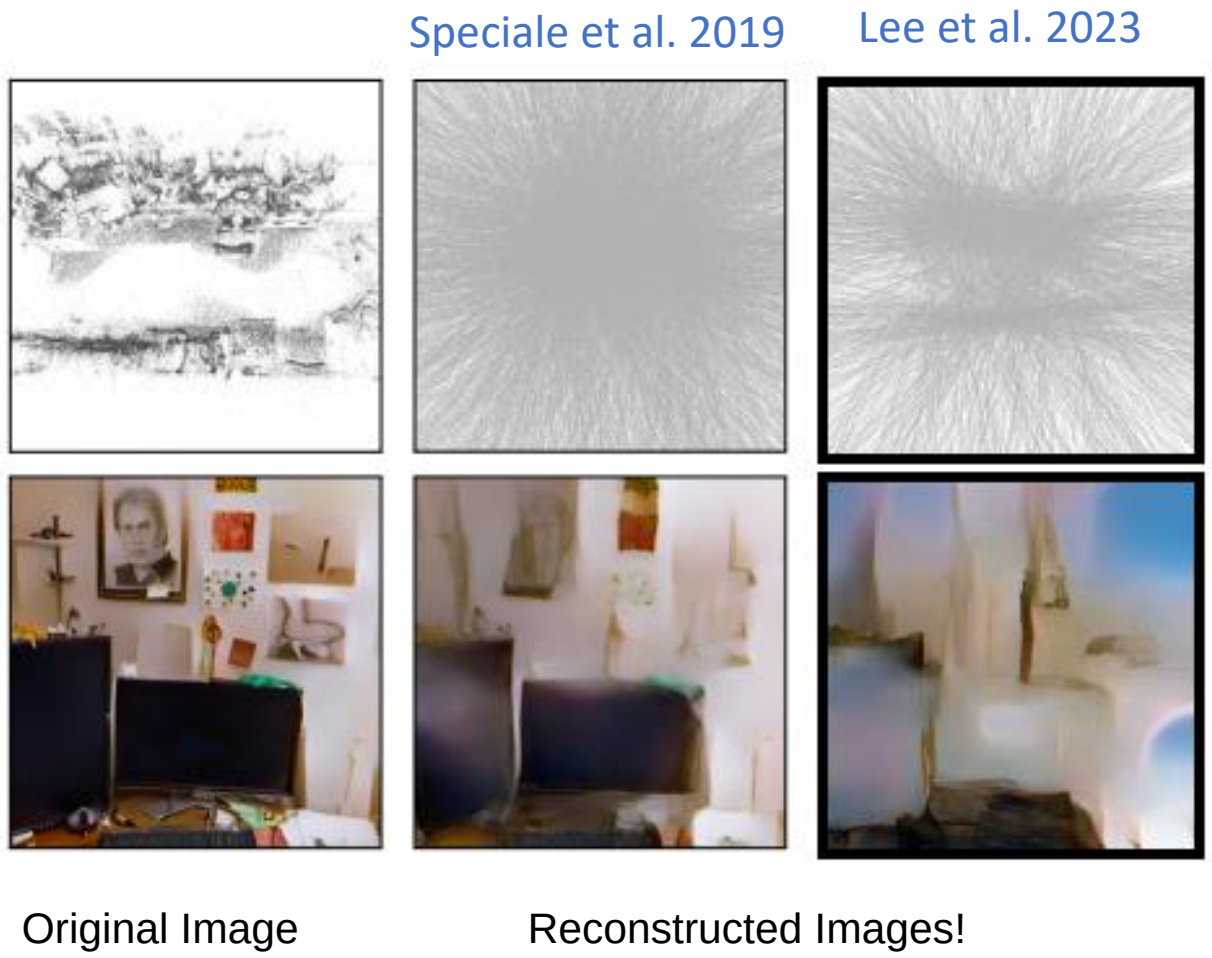


- When 3D line directions are chosen uniformly at random, then information about local neighborhoods is retained in the representation. Within local neighborhoods, the position of the closest points between line pairs provide a good approximation to the original 3D points.
- An iterative approach for approximate recovery of the original 3D points; subsequently the inversion attack can be carried out to recover scene appearance.
- **Key insight:** Alternative strategies for selecting line directions may prevent the proposed attack.

Better 3D Line Clouds via Paired-Point Lifting

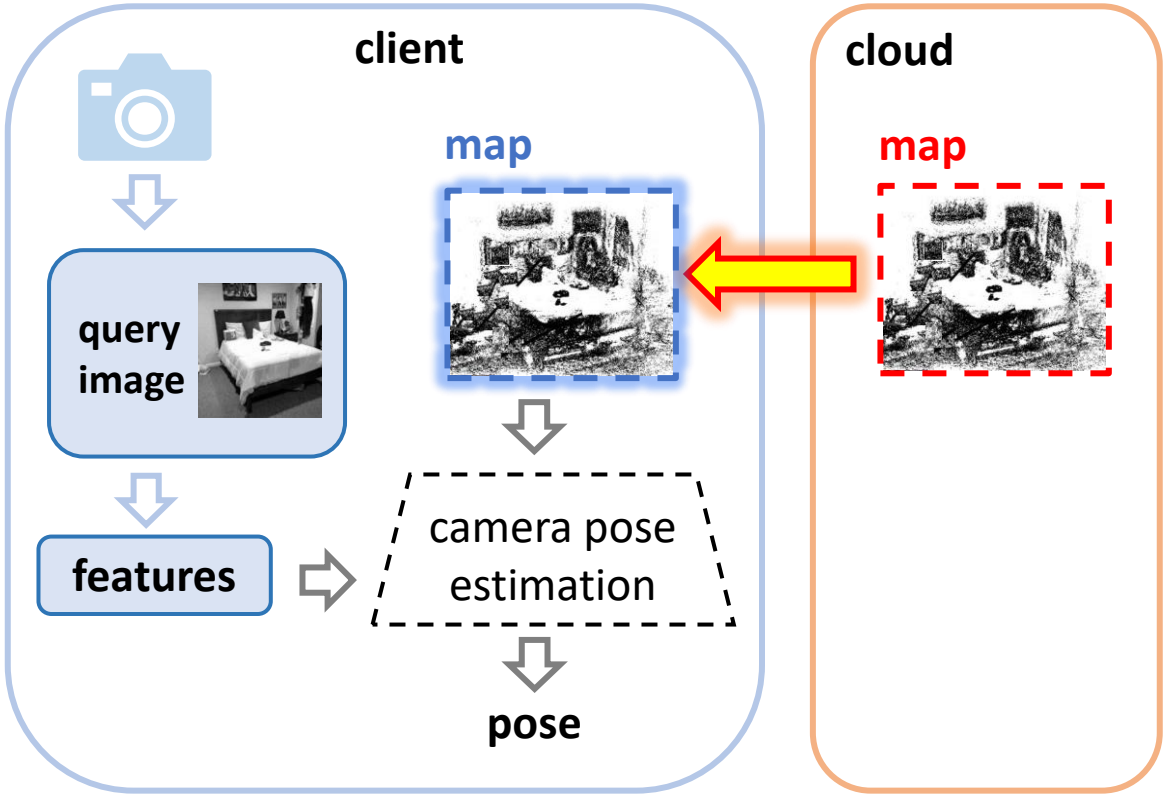
Lee et al. 2023, "**Paired-Point Lifting for Enhanced Privacy-Preserving Visual Localization**", in CVPR.

- A new method to construct 3D line clouds.
  - Selects 3D point pairs and joins them to form 3D lines.
- Non-trivial distribution of line directions
  - prevents recovery of underlying 3D points as shown by [Chelani et al. 2021].
- The new method does not compromise camera localization accuracy.



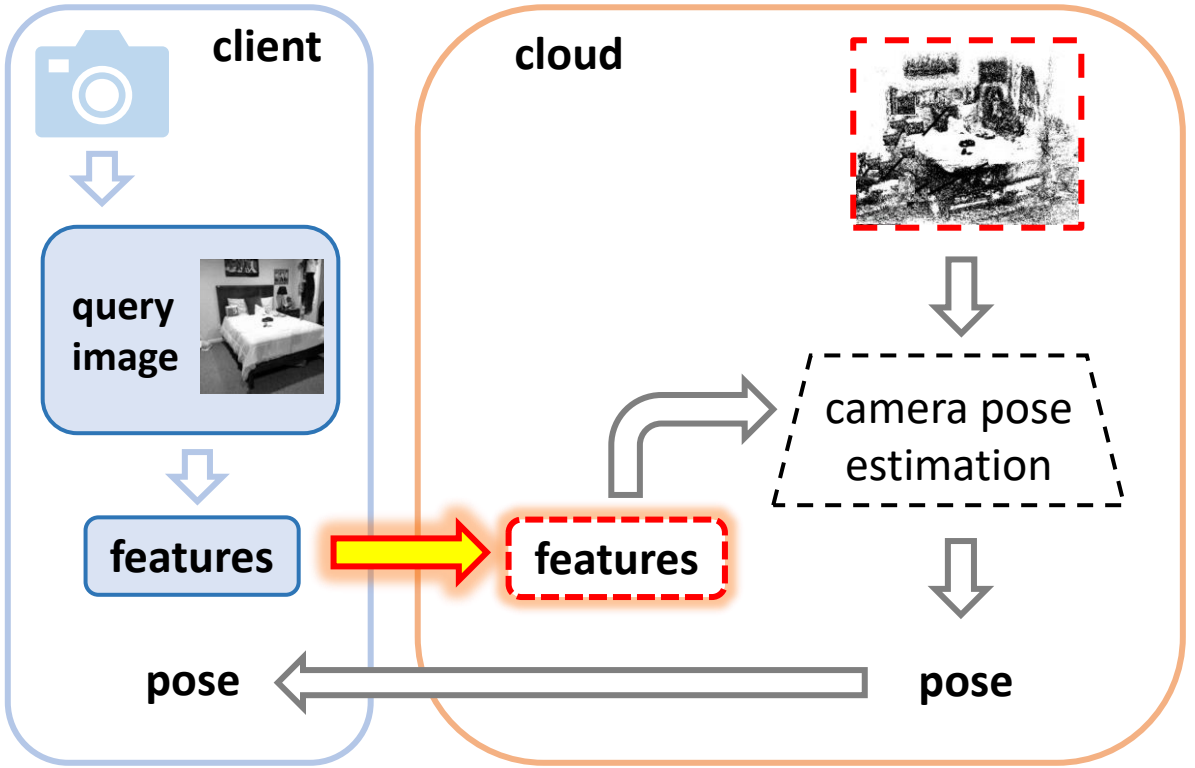
Two Different Settings

Client Setting



- Server shares map data with client devices
- Localization algorithms run locally

Server Setting



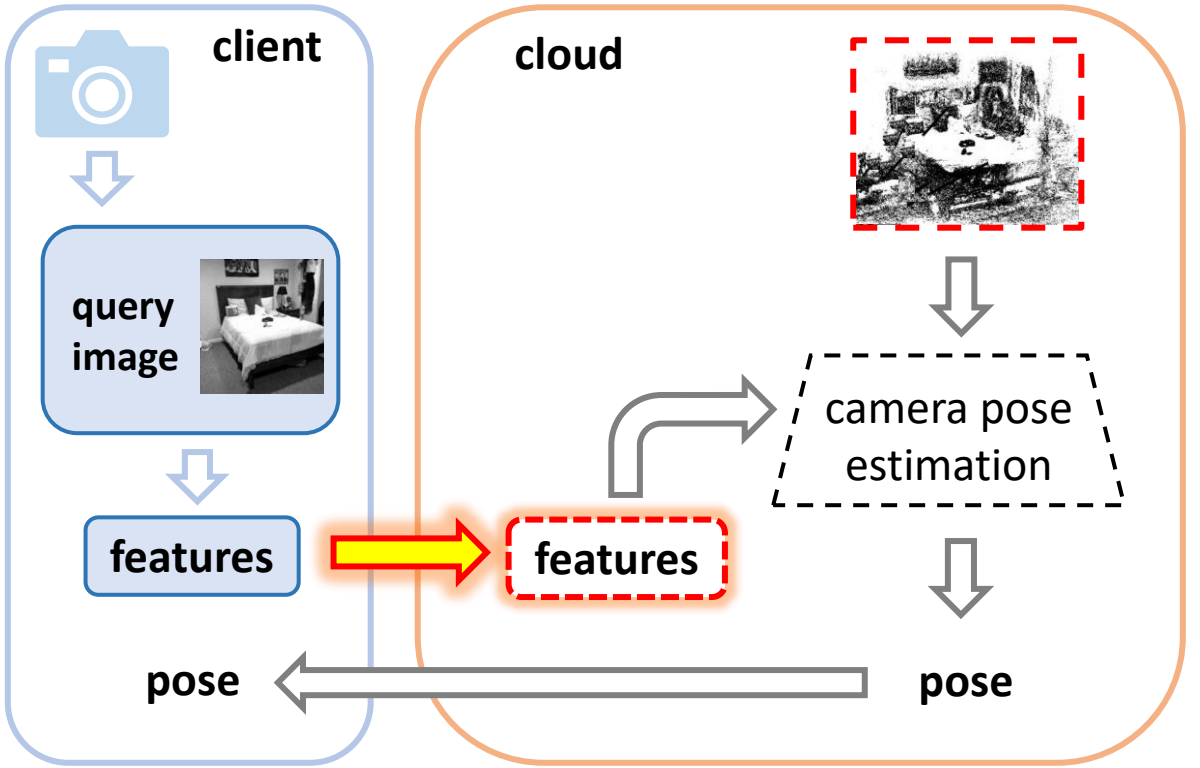
- Client shares image features with server
- Camera localization algorithms run in the cloud

Two Different Settings

Client Setting

- Server shares map data with client devices
- Localization algorithms run locally

Server Setting



- Client shares image features with server
- Camera localization algorithms run in the cloud

Privacy-Preserving Localization in the Cloud

Setting:

Attacker has access to the *query image features*.

Objective:

Compute the camera pose in the cloud but prevent the attacker from reconstructing the query images from the *query image features.*

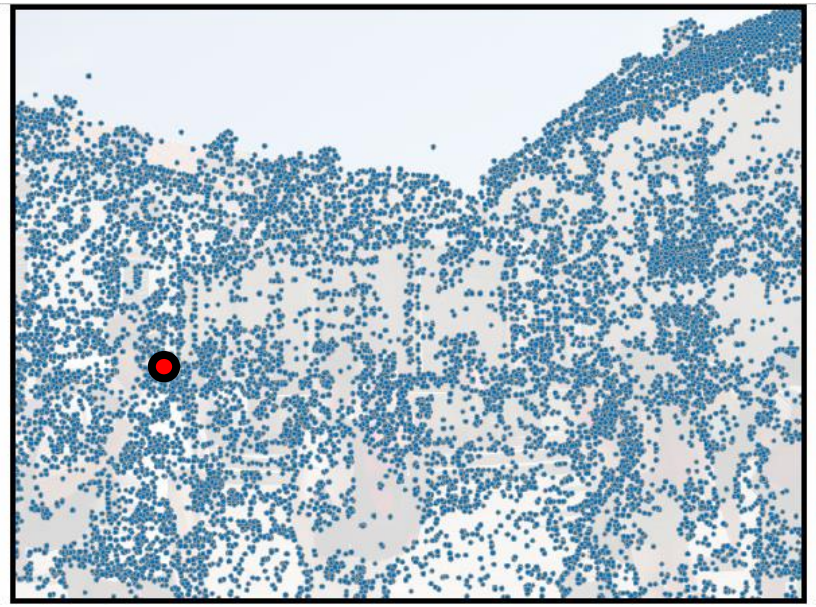
Existing Approaches:

- **New Feature Representation**
  - [Dusmanu *et al.* 2021] Affine feature subspace embeddings
  - [Ng *et al.* 2022] Adversarial Learning to learn features that cannot be inverted.
- **New Map Representation via Geometric Lifting**
  - [Speciale *et al.* 2019b] 2D feature line-based localization
  - [Geppert *et al.* 2022] Privacy preserving partial localization

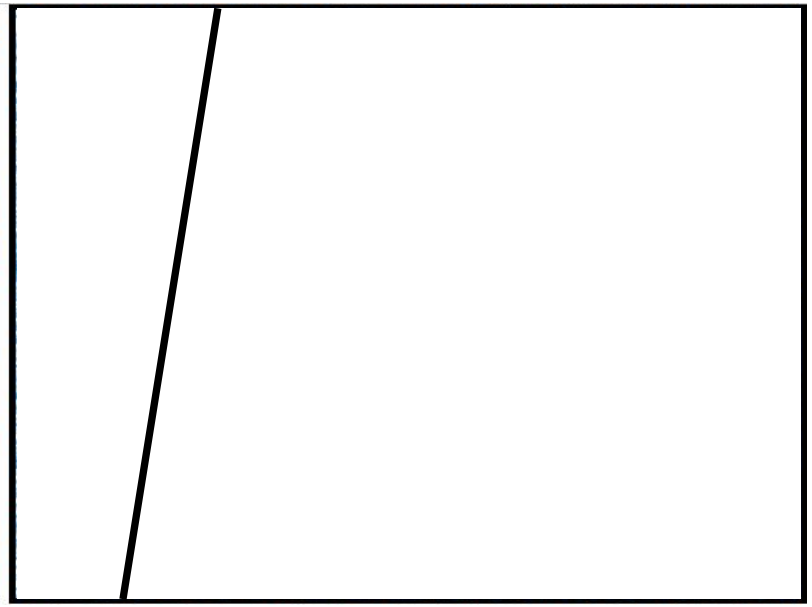
2D Feature Line-based Localization

Speciale et al. 2019, "**Revealing Scenes by Inverting Structure from Motion Reconstructions**", in ICCV.

Query Image



2D Feature Points



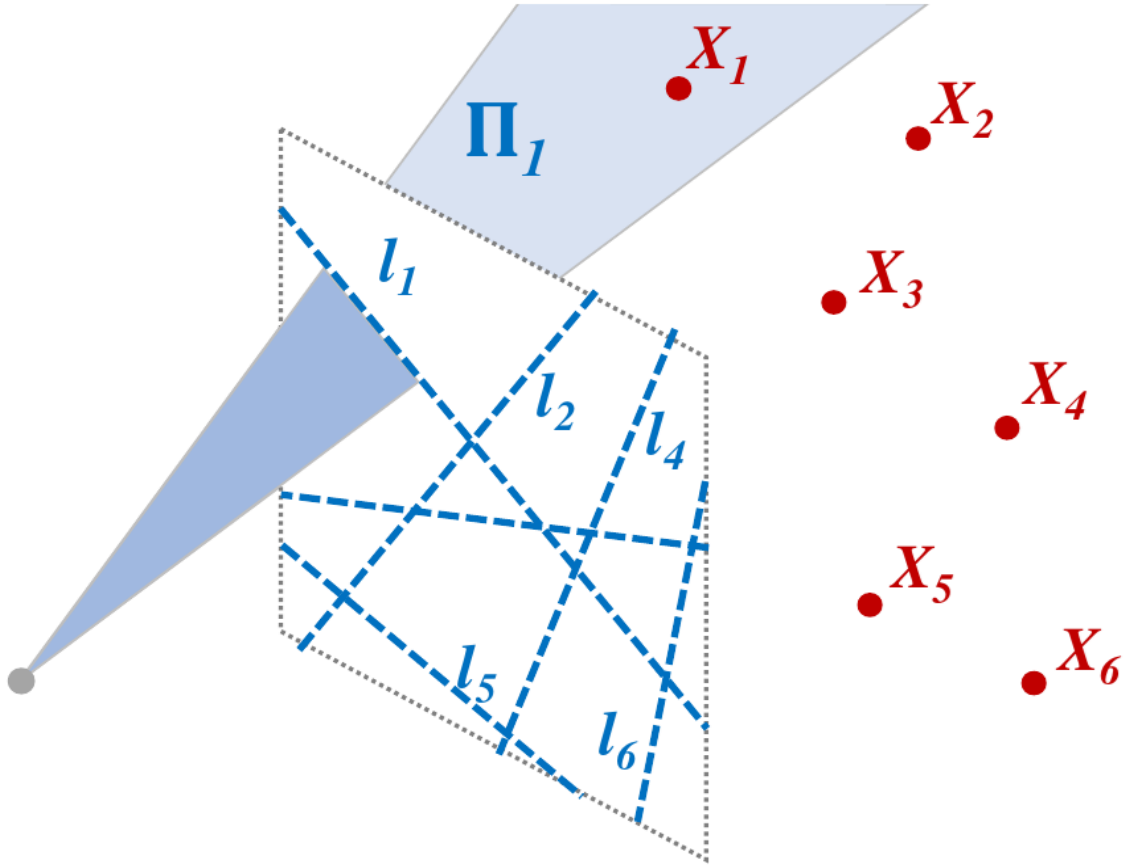
2D Feature Lines

- Select a randomly oriented 2D line through each 2D feature point
- Discard the 2D feature points
- Upload 2D features lines + descriptors to the cloud

2D Feature Line-based Localization

Speciale et al. 2019, "**Revealing Scenes by Inverting Structure from Motion Reconstructions**", in ICCV.

Line-to-point constraint (*l*6P)



At least six correspondences needed to compute pose

- Can be cast as **Point-to-Plane problem** [Ramalingam et al. 2013]
- Efficient solvers exist for multi-image queries, known vertical direction, known scale variants of the problem.
- Computationally efficient; used with RANSAC.

2D Feature Line-based Localization

Speciale et al. 2019, "**Revealing Scenes by Inverting Structure from Motion Reconstructions**", in ICCV.

- The appearance of transient objects and subjects in the query images is not revealed as the true 2D feature point locations in the query image are not revealed after pose estimation.

Query Image

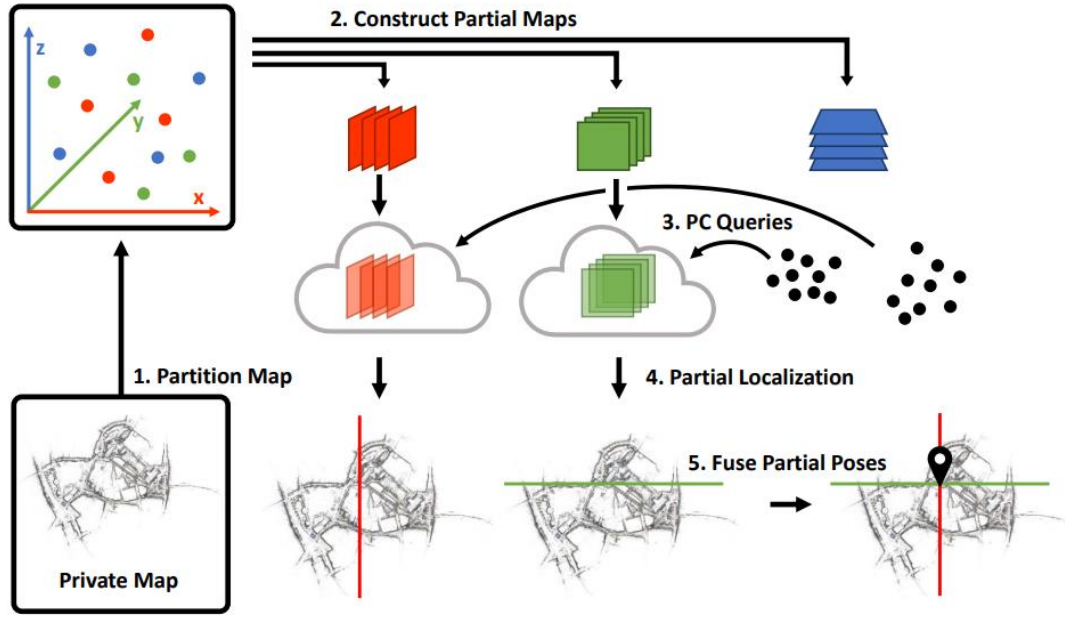
Reconstructed Image
(using all features)

Reconstructed Image
(using only *revealed* features)

Privacy-Preserving *Partial* Localization

Geppert et al. 2022, "**Privacy-Preserving Partial Localization**", in ECCV.

- First work to consider ***location privacy***, where the location (pose) of the user (client device) is treated as sensitive information and must not be revealed to the cloud.
- **Main idea:**
  - Split the 3D point cloud into three disjoint sets, each associated with one of the three orthogonal directions.
  - For each set, lift 3D points to planes (all share the same plane normal direction).
  - Store the three set of planes on three different servers.



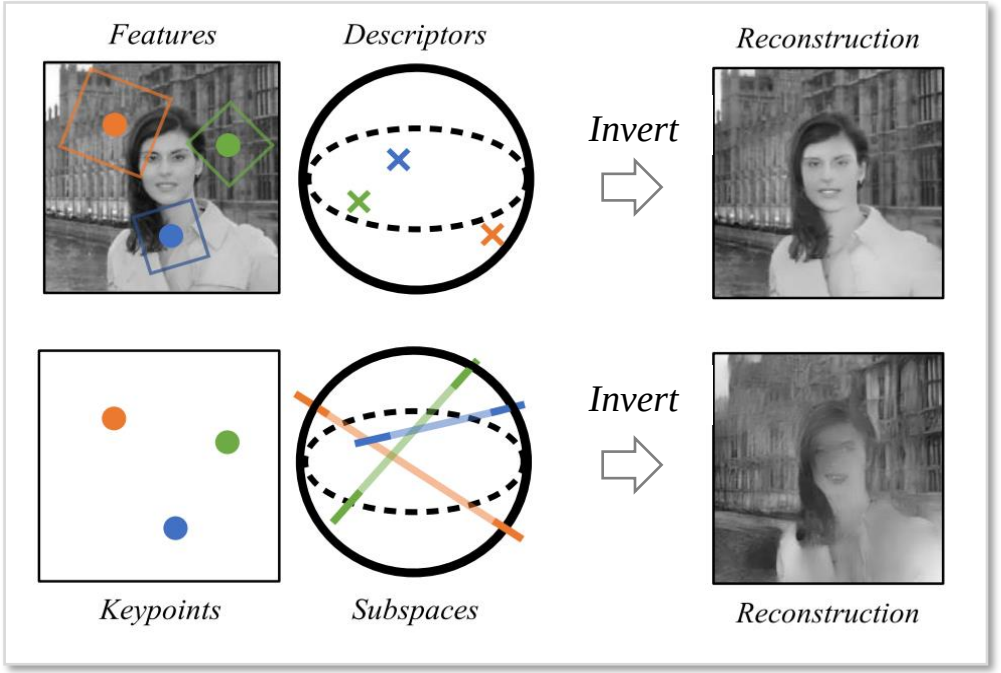
- Each server solves a *partial* localization (pose), *i.e.,* a single row of the matrix $[\boldsymbol{R}\ \boldsymbol{t}]$, where $\boldsymbol{R}$ and $\boldsymbol{t}$ are camera rotation and translation. Client can compose multiple partial poses to obtain the full 6-dof pose.

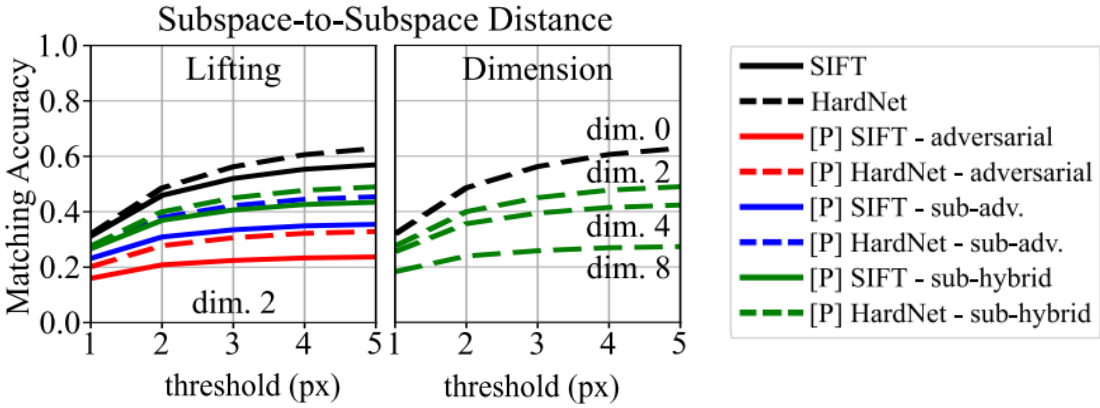
- Caveat: To guarantee privacy, servers must not communicate maps or partial poses with each other!

Privacy-Preserving Feature Descriptors

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

- Main Idea: Transform local descriptors into *private* features, by embedding each descriptor within an affine subspace that contains the original point.
- Feature matching via direct comparison of private descriptors (affine subspaces), based on the notion of subspace-to-subspace distance.
- Accuracy/privacy tradeoff of three subspace construction methods analyzed:
  - Random, Adversarial and Hybrid lifting
- Evaluated with SIFT, HardNet [1] descriptors on localization, structure-from-motion tasks.

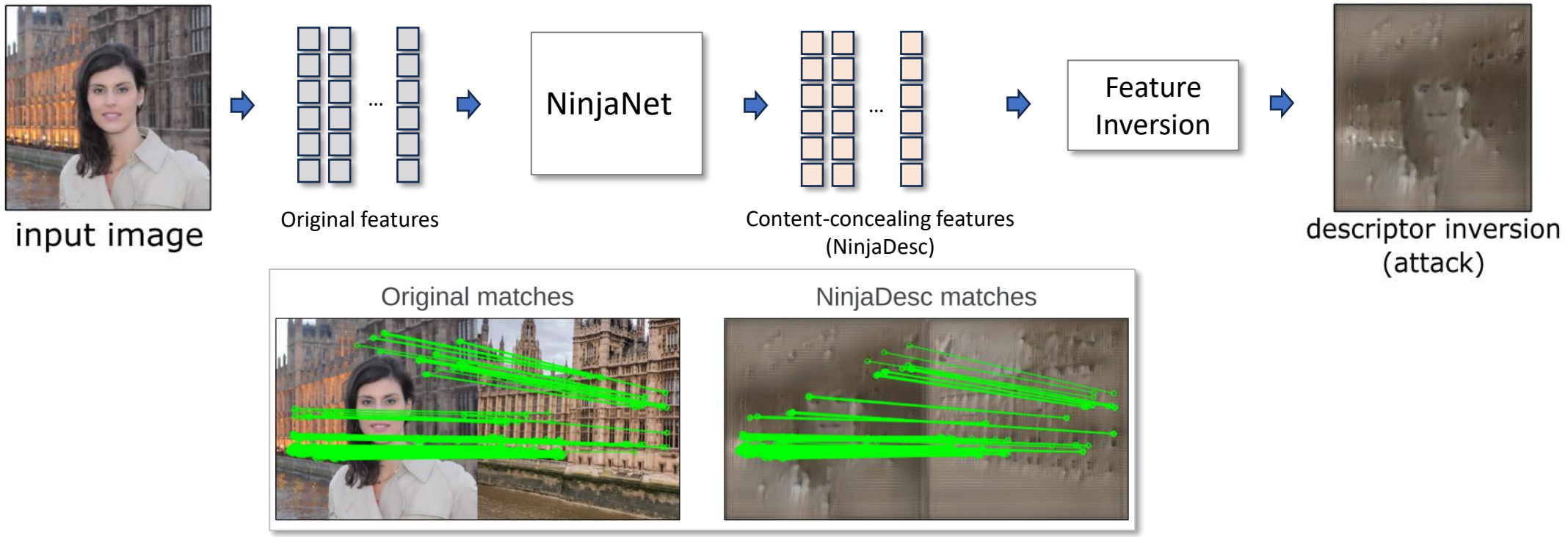




[1] Mischuk *et al.* 2017: **Working hard to know your neighbor's margins: Local descriptor learning loss.**

Learning features that cannot be inverted

Ng et al. 2022, “NinjaDesc: Content-Concealing Visual Descriptors via Adversarial Learning”, in CVPR.



- Learn local features that prevent image reconstruction (inversion) but aids feature matching.
- Adversarial learning framework to train NinjaNet
  - Training loss balances utility (matchability) and reconstructibility (privacy).

Learning-based Localization Provides Privacy

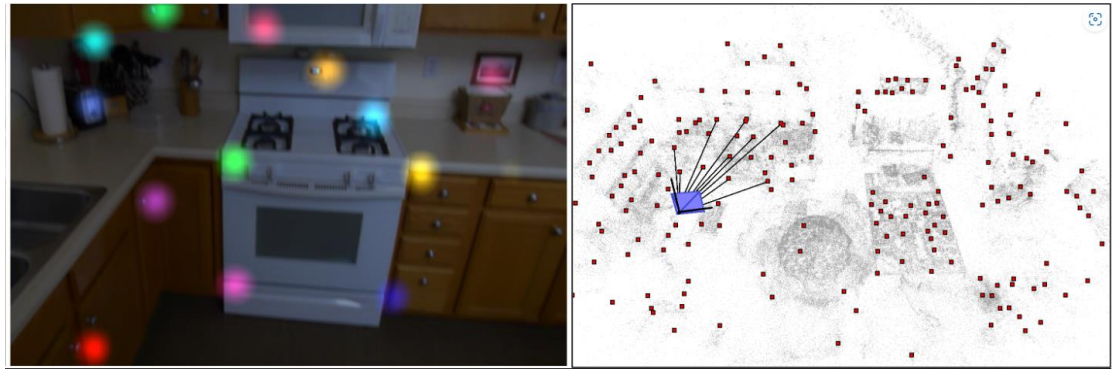
- Learning-based localization methods do not explicitly store maps.
  - Absolute Pose Regression (*e.g.,* PoseNet etc.)
    [Kendall+ 2015]
  - Scene Coordinate Regression (*e.g.,* DSAC, DSAC*, etc.)
    [Brachman+ 2017, Brachman and Rother 2021]

- Instead, they implicitly encode the map as a scene-specific neural net.
- These methods do not store the map after the model is trained.
- Thus, they avoid the privacy issue by design.

- However, model training is not privacy-preserving.

Learning-based Localization Provides Privacy

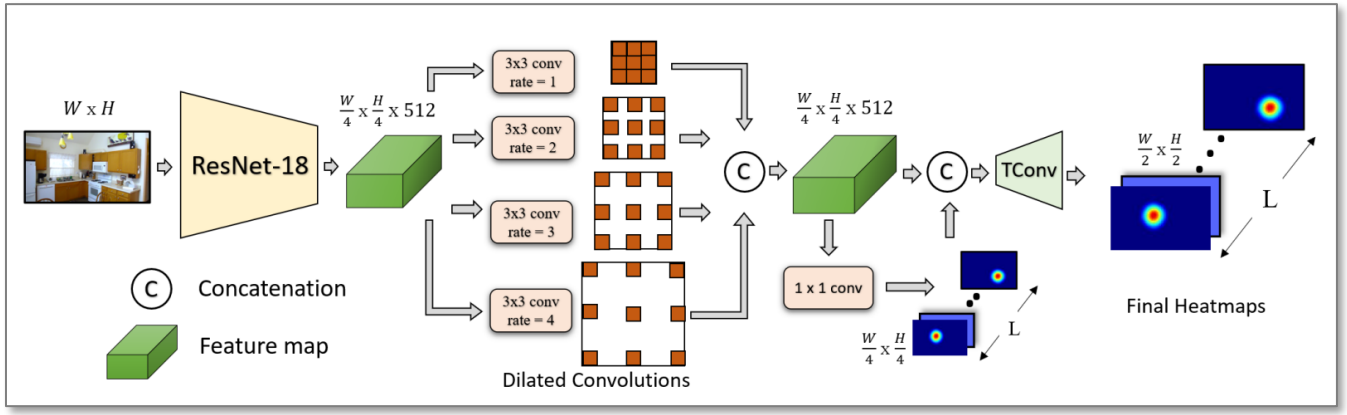
Do et al. 2022, “**Learning to Detect Scene Landmarks for Camera Localization**”, in CVPR.



Scene landmarks detected in query image.

PnP pose estimation

Scene Landmark Detector (SLD) architecture



- Ultra sparse map representation (a set of designated 3D points per scene *aka* scene landmarks.
- Scene-specific CNN model trained to detect the landmarks in query images.
- As few as 300 landmarks enough to represent medium sized indoor scenes.
- Outperforms existing learned localization approaches; but not as accurate as feature-based methods such as **hloc** [Sarlin et al. 2019, Sarlin et al. 2020].

Conclusions

- Privacy Preserving Camera Localization
  - The objective is to robustly compute camera pose but prevent image reconstruction attacks that reveal sensitive information about the scene.
  - The problem is becoming increasingly relevant as AR, robotics and spatial AI technologies are deployed in the real world.
- Reviewed three promising solution strategies explored so far.
  - Geometric Lifting (3D Line clouds, 2D feature lines, 3D planes)
  - Transformations for local feature descriptors.
  - Learning-based localization where map is discarded after training.

Thanks

References:

- Pittaluga et. al. 2019, Revealing Scenes by Inverting Structure from Motion Reconstructions, in CVPR.
- Speciale et. al. 2019, Privacy Preserving Image-Based Localization, in CVPR.
- Speciale et. al. 2019b, Privacy Preserving Image Queries for Camera Localization, in ICCV.
- Shariati et al. 2019, Towards Privacy-Preserving Ego-Motion Estimation using an Extremely Low-Resolution Camera, in RA-L.
- Shibuya et al. 2020, Privacy Preserving Visual SLAM, in ECCV.
- Geppert et. al. 2020, Privacy Preserving Structure-from-Motion, in ECCV.
- Chelani et. al. 2021, How Privacy-Preserving are Line Clouds? Recovering Scene Details from 3D Lines, in CVPR.
- Dusmanu et. al. 2021, Privacy-Preserving Image Features via Adversarial Affine Subspace Embeddings, in CVPR.
- Geppert et. al. 2020, Privacy Preserving Structure-from-Motion, in ECCV.
- Geppert et. al. 2021, Privacy Preserving Localization and Mapping from Uncalibrated Cameras, in CVPR.
- Geppert et. al. 2022, Privacy Preserving Partial Localization, in CVPR.
- Lee et. al. 2023, Paired-Point Lifting for Enhanced Privacy-Preserving Visual Localization, in CVPR.