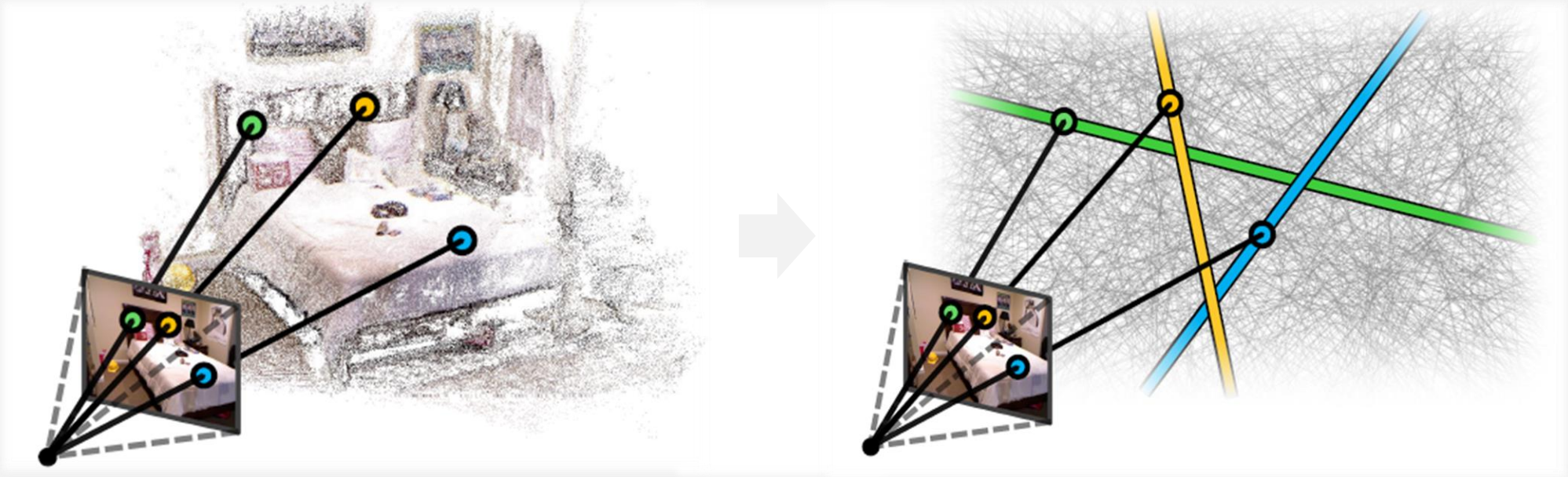


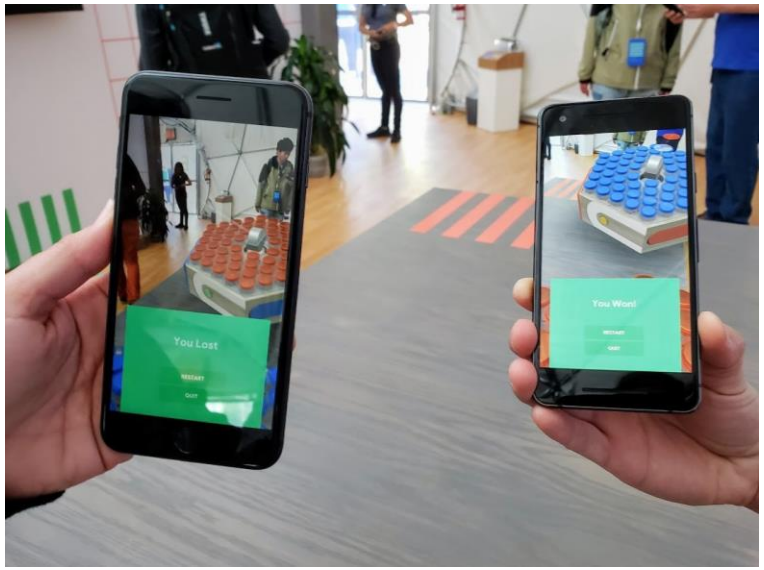
Privacy-Preserving Camera Localization

Sudipta N. Sinha

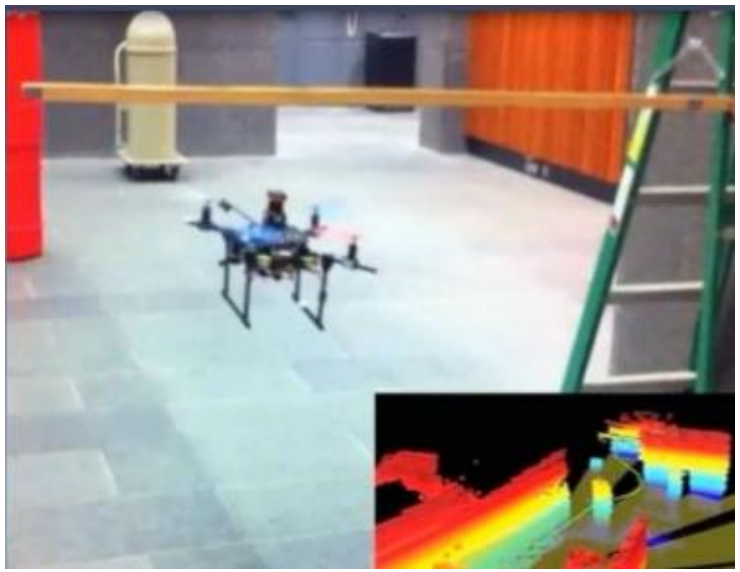
Large-Scale Visual Localization Tutorial

@ ICCV 2021, October 17, 2021

Motivation



Mixed Reality, Augmented Reality



Robots in home, workplaces

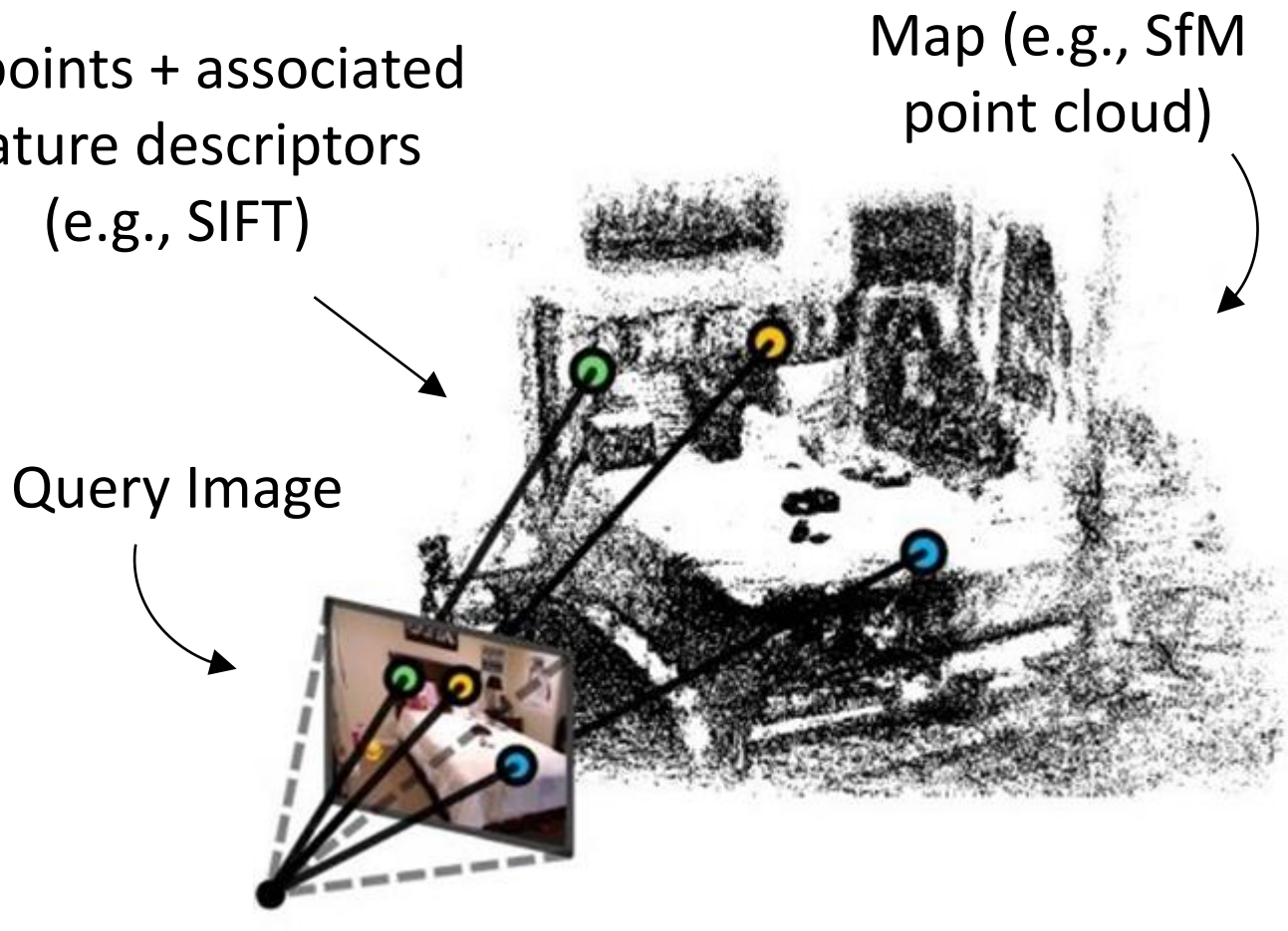
As localization systems get deployed in the real world,
- what are the privacy implications of such systems?
- what solutions have been developed so far and how effective are they?
- where is more work needed?

Review: Camera Localization and Basic Terminology

Input
- Query Image.
- Map point cloud + features (pre-computed).
- Camera intrinsics (typically known).

Pipeline
- Extract features in query image.
- Match query features to map features.
- Robust camera pose estimation.

Output
- Camera pose in map coordinate system.



CVPR 2019

Revealing Scenes by Inverting Structure from Motion Reconstructions

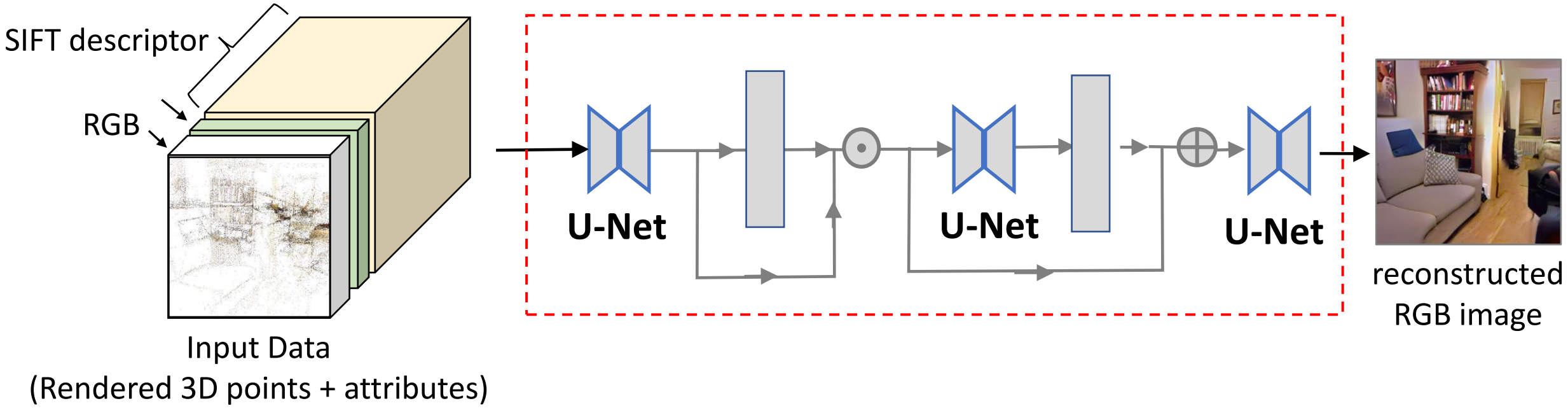
Francesco Pittaluga[1] Sanjeev J. Koppal[1] Sing Bing Kang[2] Sudipta N. Sinha[2]

[1] University of Florida [2] Microsoft Research

(a) SfM point cloud (top view) (b) Projected 3D points (c) Synthesized Image (d) Original Image

Figure 1: SYNTHESIZING IMAGERY FROM A SFM POINT CLOUD: From left to right: (a) Top view of a SfM reconstruction of an indoor scene, (b) 3D points projected into a viewpoint associated with a source image, (c) the image reconstructed using our technique, and (d) the source image. The reconstructed image is very detailed and closely resembles the source image.

Inverting Map Feature Descriptors



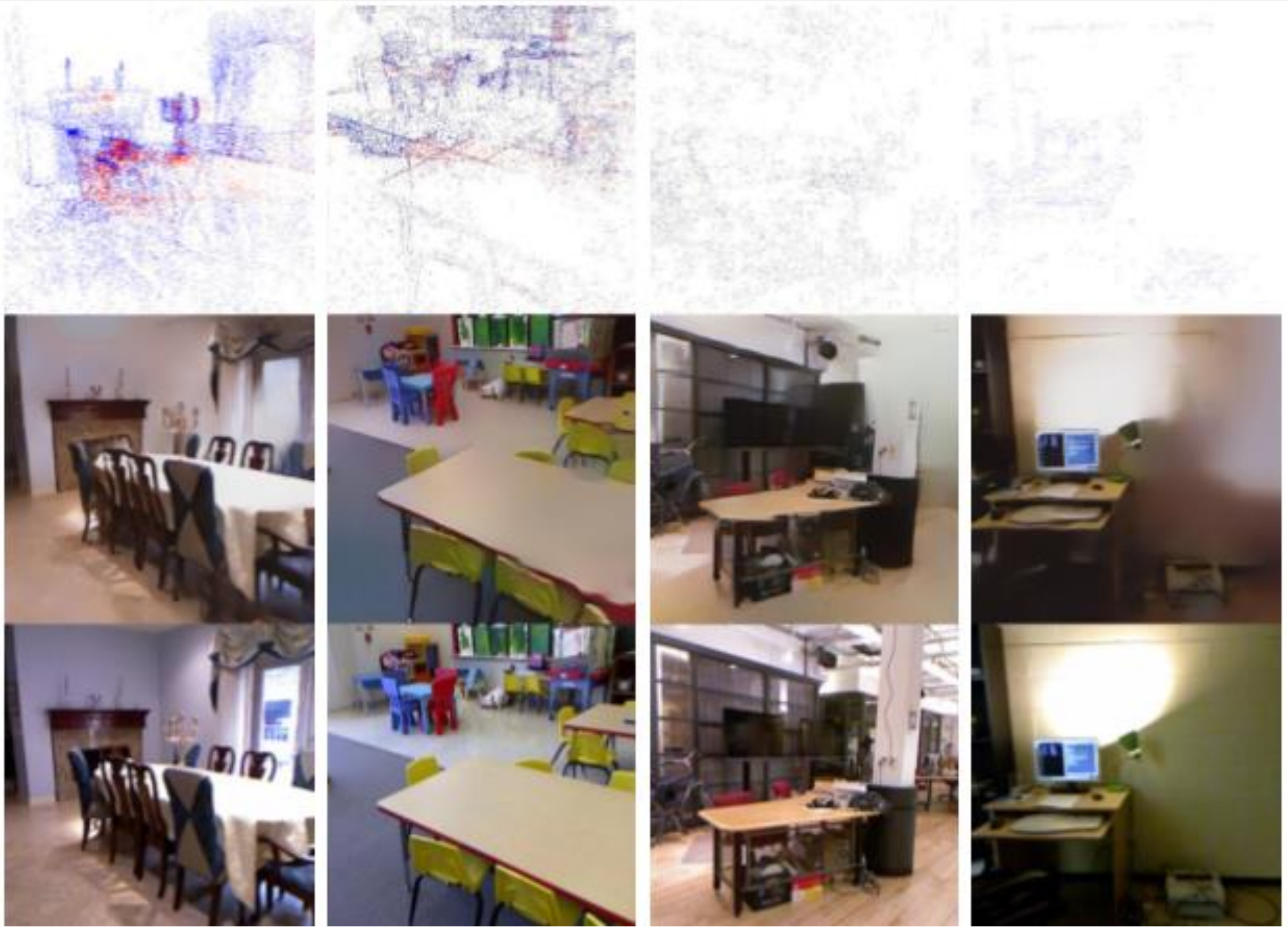
- The inversion method uses a deep neural network trained on Internet Photo datasets.
- An attacker could therefore recover detailed scene appearance using the inversion method on visual features stored in the cloud for camera localization.

Examples

Input
(Depth, SIFT, …)

Reconstructed
Images

Original
Images



Reconstruction of Source Video used in Mapping

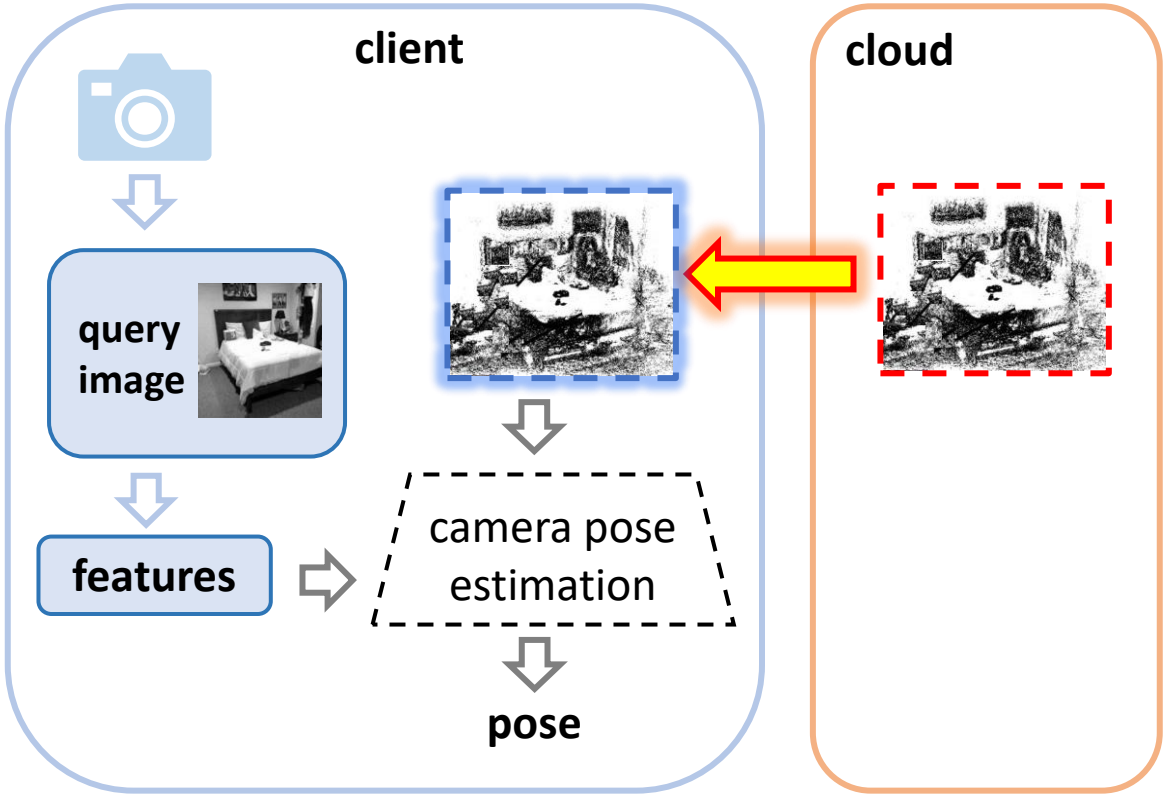
Projected 3D Points
(network input)

Reconstructed Frames

Original Frames

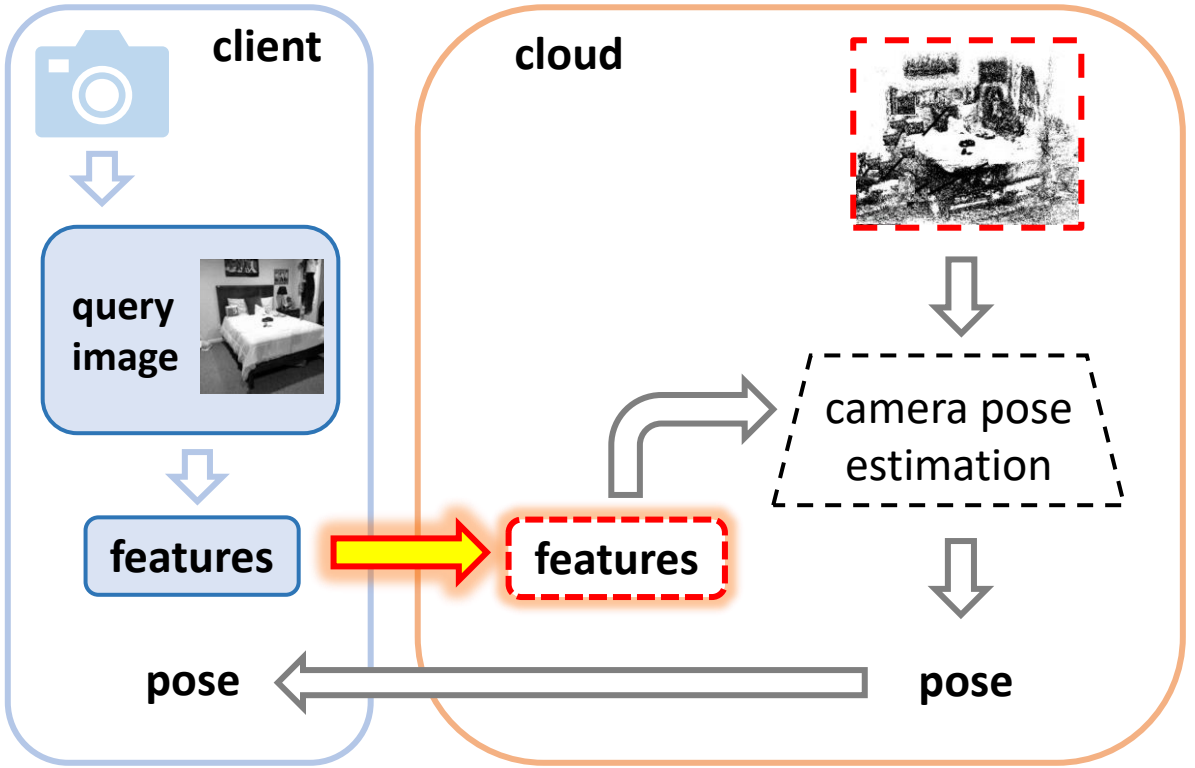
Privacy Implications for Camera Localization

Shared map scenario



- Server shares map data with clients
- Goal: conceal map features

Cloud-based localization scenario



- Client shares image features with server
- Goal: conceal query image features

Privacy-Preserving Camera Localization

Shared map scenario

Goal: An attacker has access to the map features. The task is to compute the camera pose but prevent the attacker from reconstructing images of the scene from the *map features.*

Methods and known results:

- 3D line cloud-based localization

  [Speciale et al. 2019]

- Merely using line clouds alone is not enough!

  [Chelani et al. 2021]

Current Status:

- No satisfactory solution exists yet!

Cloud-based localization scenario

Goal: An attacker has access to the query image features. The task is to compute the camera pose but prevent the attacker from reconstructing the query image from the *query features.*

Existing methods:

- 2D feature line-based localization

  [Speciale et al. 2019b]

- Transform feature descriptors to affine subspaces; new matching strategy.

  [Dusmanu et al. 2021]

Privacy Preserving Image-Based Localization

Pablo Speciale[1,2] Johannes L. Schönberger[2] Sing Bing Kang[2] Sudipta N. Sinha[2] Marc Pollefeys[1,2]

[1] ETH Zürich [2] Microsoft

(a) 3D Point Cloud (b) Inversion Attack on SIFT of 3D Point Cloud (c) 3D Line Cloud

Figure 1: **(a)** Traditional image-based localization using 3D point cloud, which reveals potentially confidential information in the scene. **(b)** Reconstructed image using projected sparse 3D points and their SIFT features [50]. **(c)** Our proposed 3D line cloud protects user privacy by concealing the scene geometry and preventing inversion attacks, while still enabling accurate and efficient localization.

Abstract

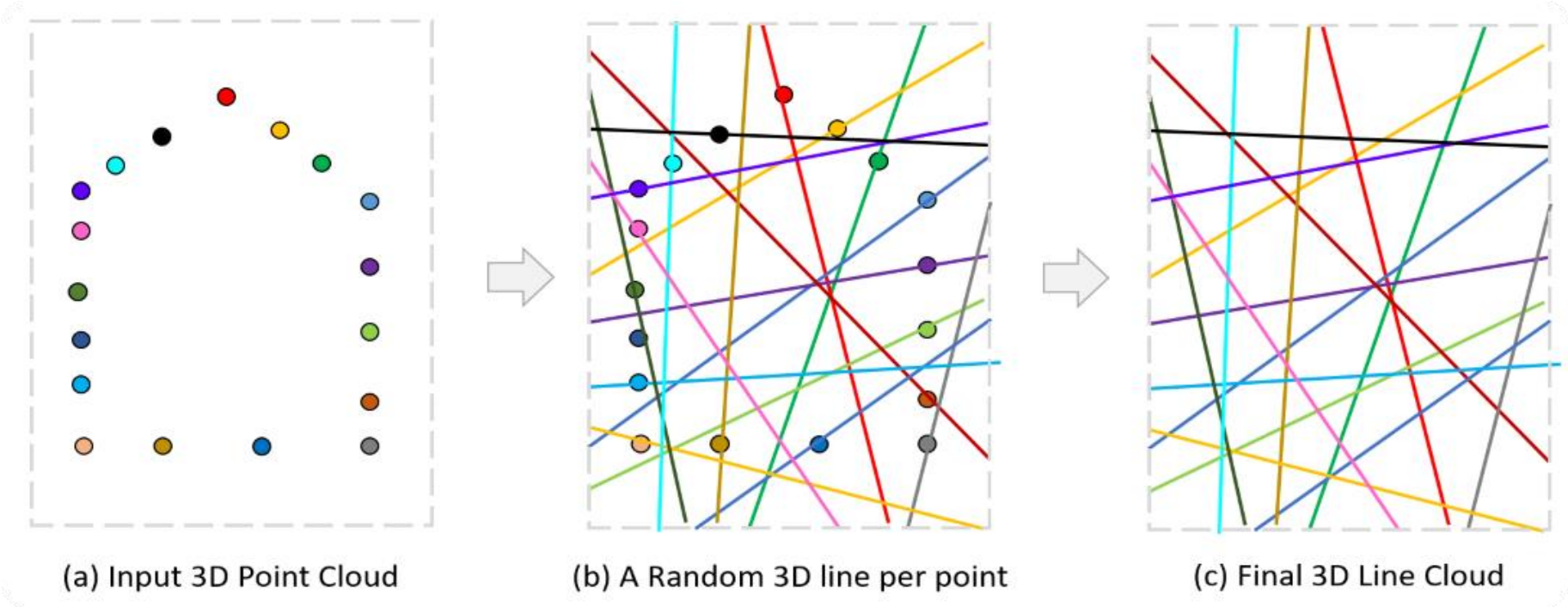
Image-based localization is a core component of many augmented/mixed reality (AR/MR) and autonomous robotic systems. Current localization systems rely on the persistent storage of 3D point clouds of the scene to enable camera pose estimation, but such data reveals potentially sensitive scene information. This gives rise to significant privacy risks, especially as for many applications 3D map-

constraints derived from the matched 2D–3D point correspondences are then used to estimate the camera pose. Inherently, the traditional approach to image-based localization thus requires the persistent storage of 3D point clouds.

The popularity of AR platforms such as ARCore [5] and ARKit [7], wearable AR devices such as Microsoft HoloLens [31], and announcements of Microsoft's Azure Spatial Anchors (ASA) [11], Google's Visual Positioning

Main Idea: Transform 3D points to 3D lines

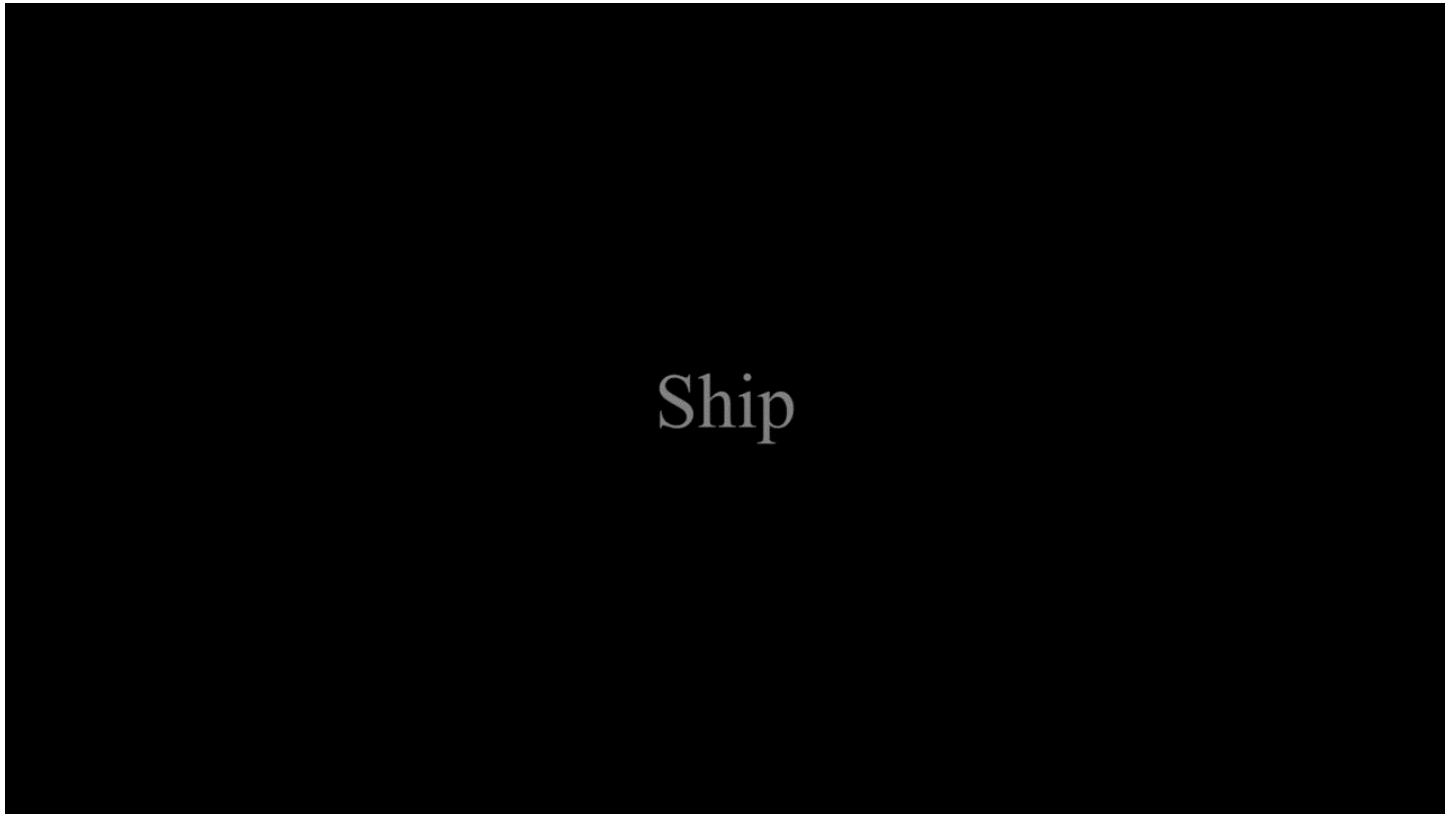
[Speciale et al. 2019]



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

Main Idea: Transform 3D points to 3D lines

[Speciale et al. 2019]

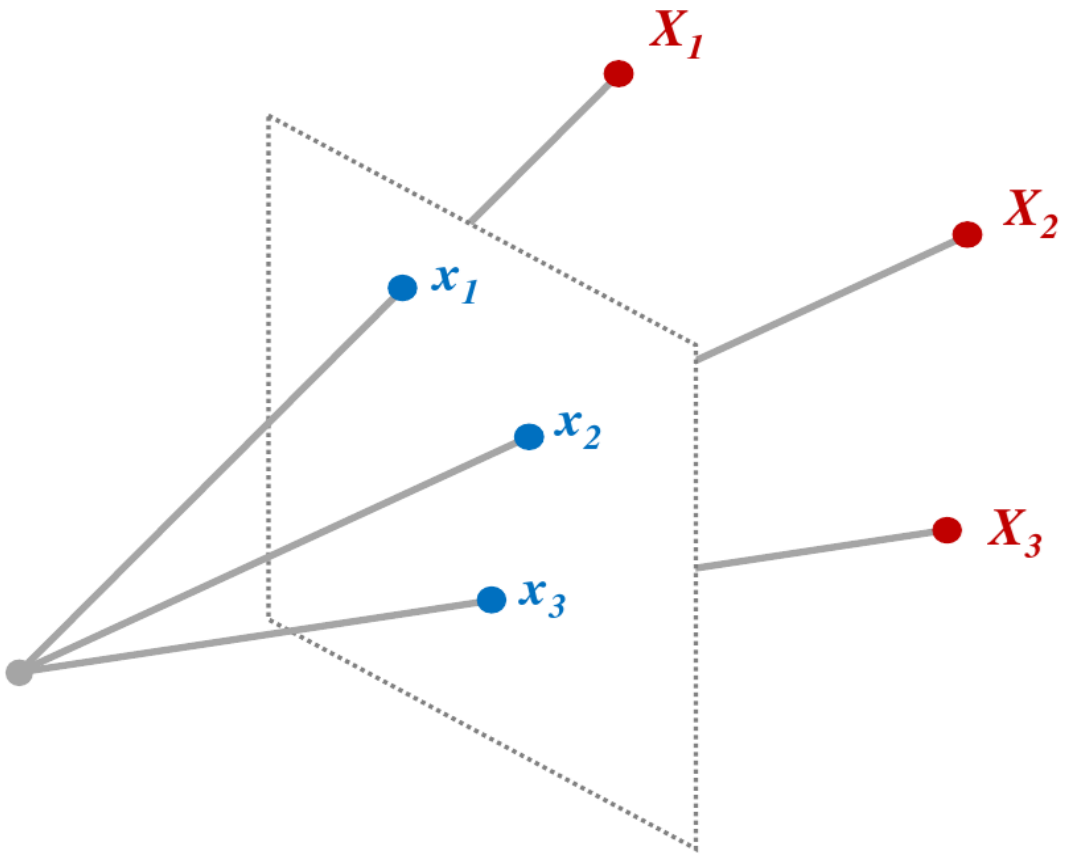


- For each 3D point, pick a randomly oriented 3D line passing through the point.
- Then **discard** the 3D point.

Geometric Constraints

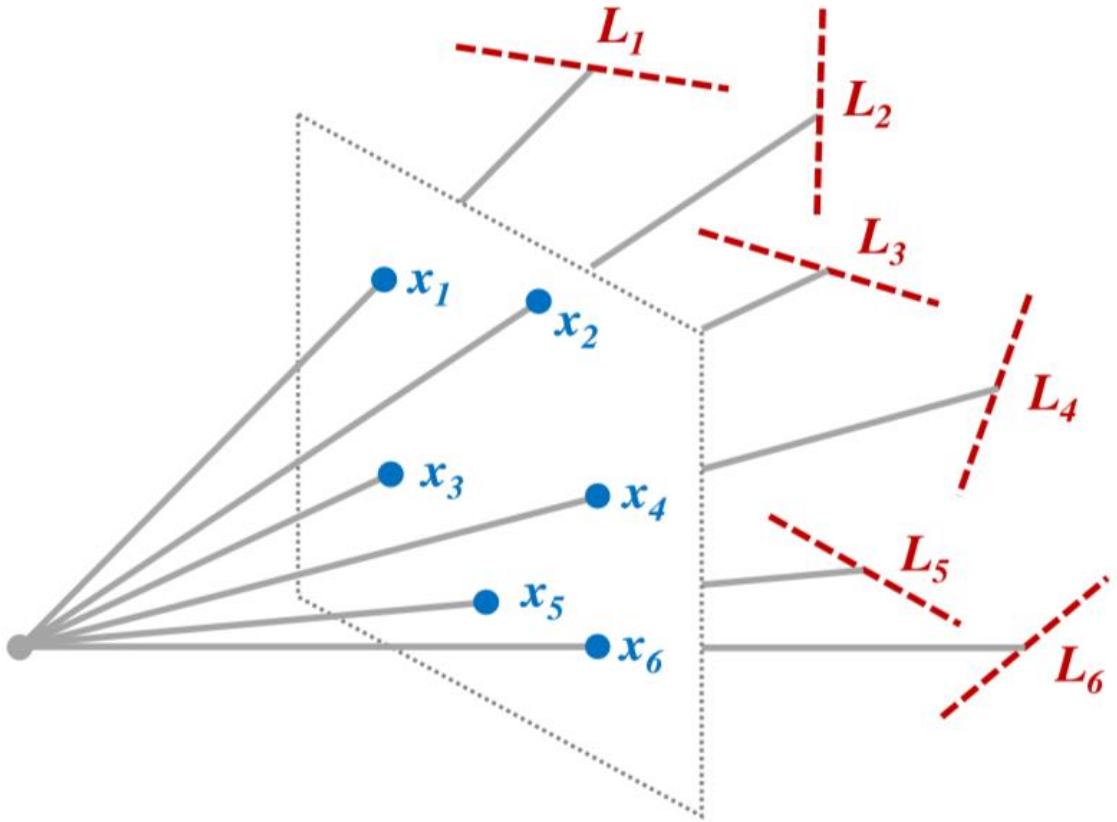
[Speciale et al. 2019]

point-to-point ($p3P$)



Three “image point”–3D point correspondences

point-to-line ($p6L$)



Six “image point”–3D line correspondences

3D Line-based Pose Estimation

1. Minimal problem same as the **generalized relative pose problem**.
   [Stewenius et al. 2005]

2. Variants:
   - Query 3D point cloud (from multiple images),
   - known vertical direction,
   - known scale.

   Leverage existing minimal solvers [2-7].

   [2] Nister et al. 2007 [3] Lee et al. 2014 [4] Stewenius et al. 2005

   [5] Sweeney et al. 2015a [6] Sweeney et al. 2015b [7] Sweeney et al. 2014

3. Computationally efficient and can be used with RANSAC.

How Privacy-Preserving are Line Clouds? Recovering Scene Details from 3D Lines

Kunal Chelani[1] Fredrik Kahl[1] Torsten Sattler[1,2]
[1]Chalmers University of Technology [2]Czech Technical University in Prague

Abstract

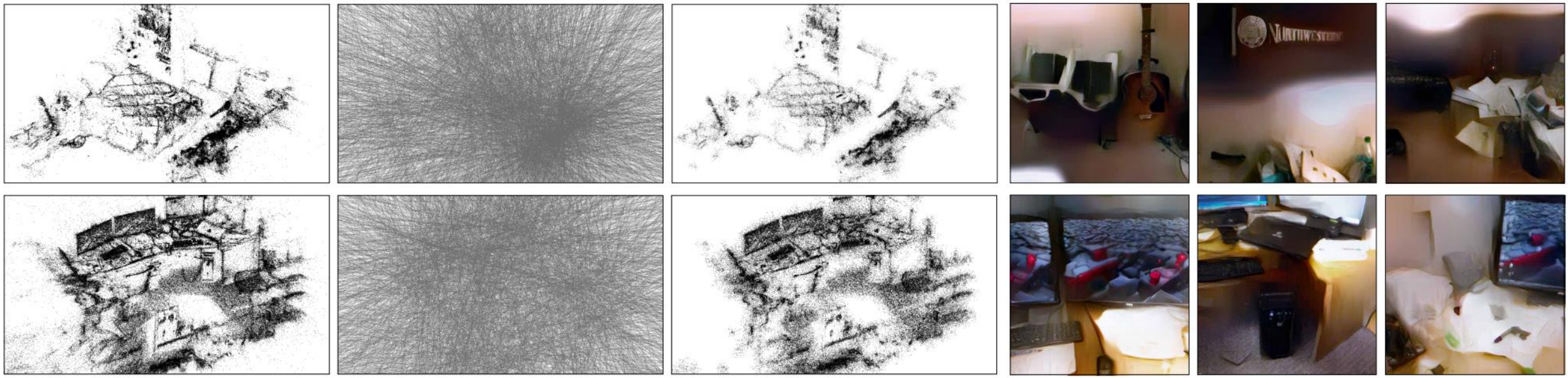
Visual localization is the problem of estimating the camera pose of a given image with respect to a known scene. Visual localization algorithms are a fundamental building block in advanced computer vision applications, including Mixed and Virtual Reality systems. Many algorithms used in practice represent the scene through a Structure-from-

image and the 3D points then yields a set of 2D-3D matches that can be used for RANSAC-based camera pose estimation [9, 17, 23, 36–39].

Traditionally, work on visual localization has focused on accurate and scalable algorithms able to cover large areas [15, 30, 40, 58, 74, 76, 86] or to run in real-time on mobile devices with limited memory and compute capabili-

Recovering the Original Point Cloud from the Line Cloud

[Chelani et al. 2019]



Original Point Cloud → 3D Line Cloud → Approximate recovery of 3D points → Recovered Images!

[Speciale et al. 2019] This paper [Pittaluga et al. 2019]

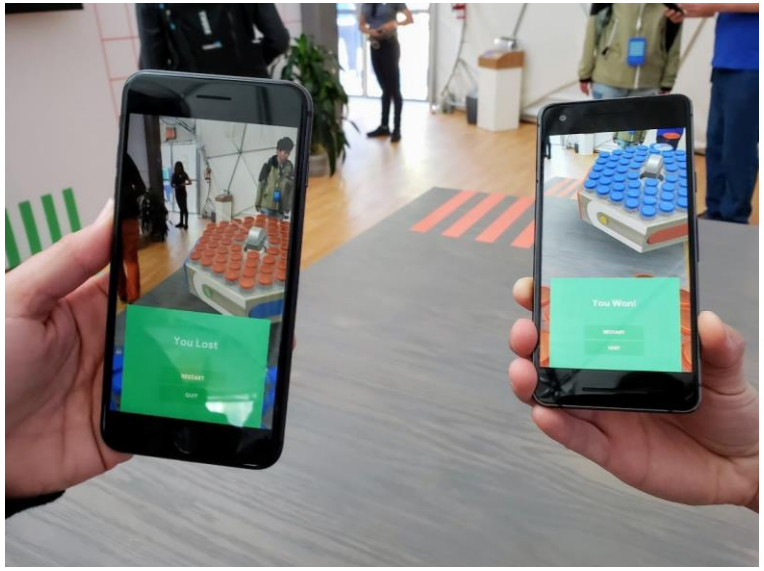
Main Insight

[Chelani et al. 2019]

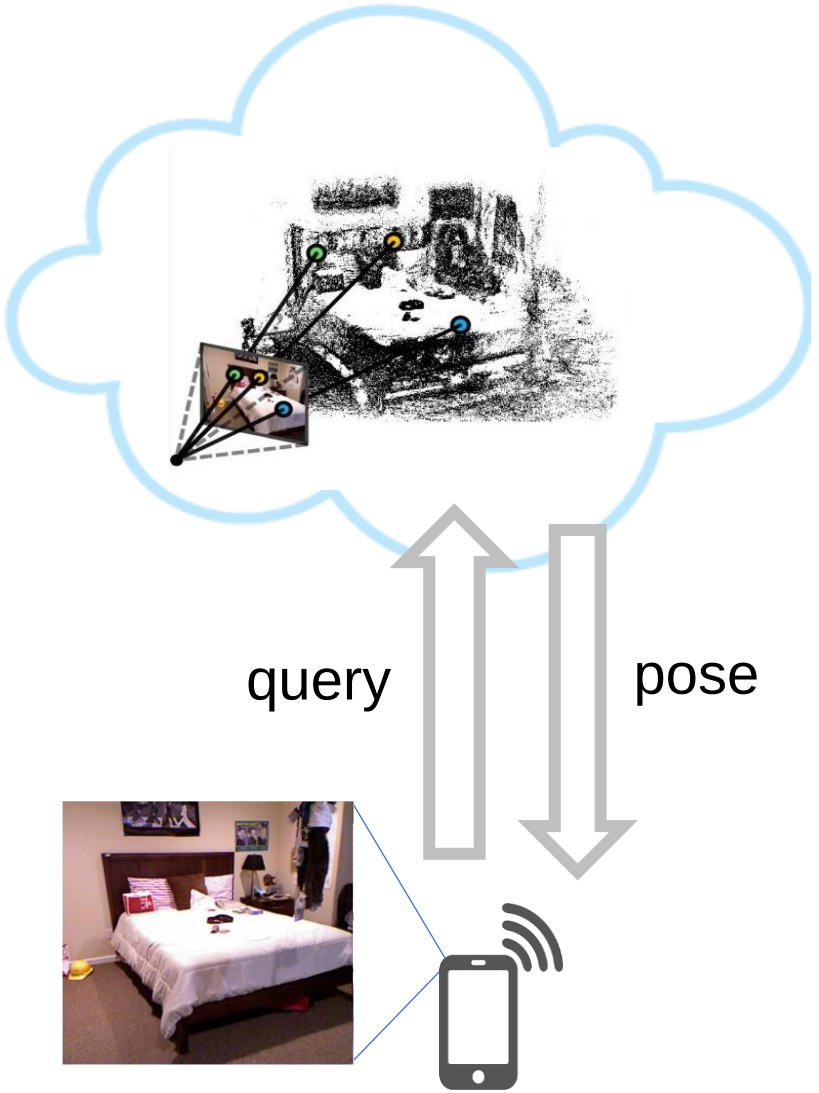
- When 3D line directions are chosen uniformly at random, then the position of the two closest points between pairs of lines is often near the original 3d points; especially true when the two points are near each other.
- Repeat for each line
  - Find approximate point neighborhoods for each point/line.
  - For those points, find closest points on the 3D line.
  - Select a single candidate per line via robust peak finding.
- Run multiple iterations

Localization in the Cloud

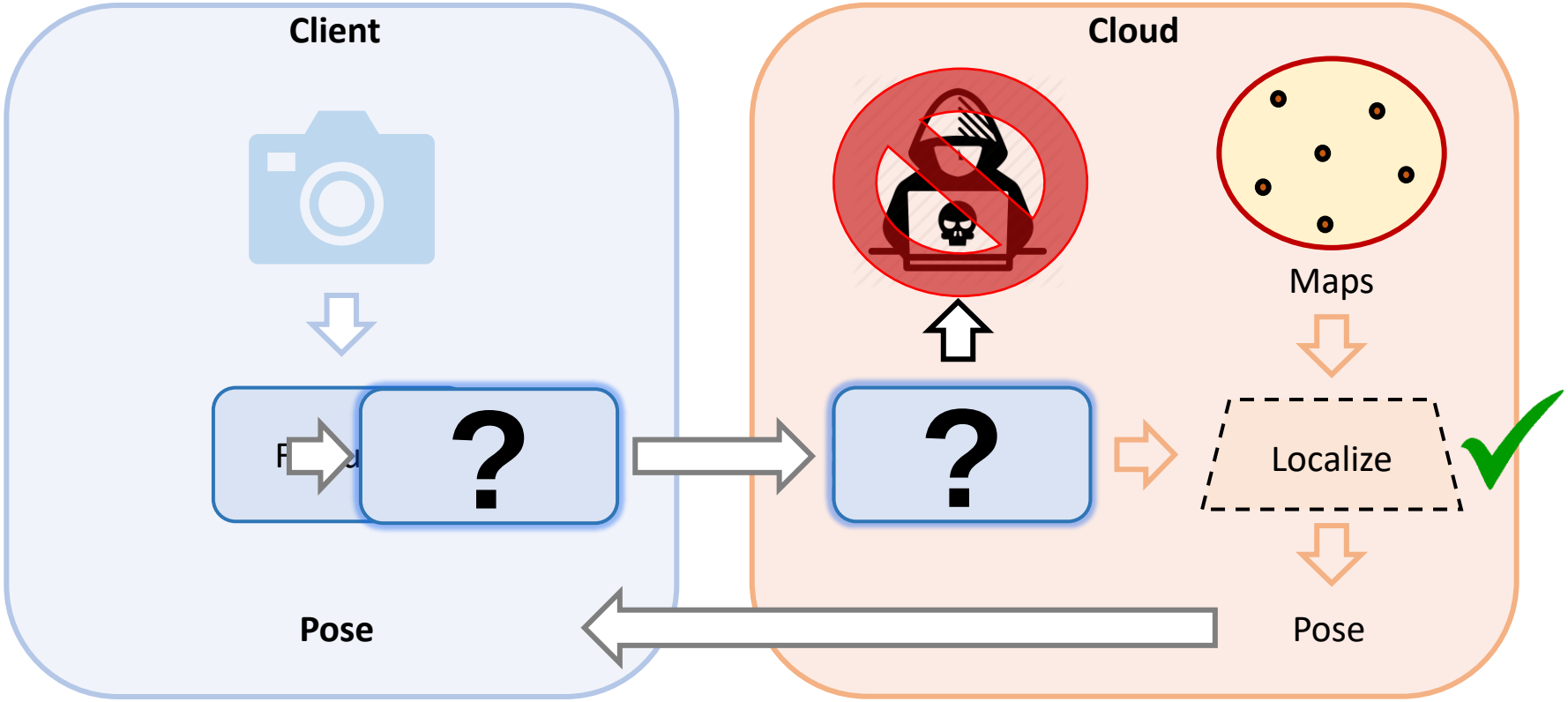
Microsoft ASA
(Azure Spatial Anchors)



Google AR Core
(Cloud Anchors)



Privacy Preserving Cloud-based Localization



Our Goal:

- Hide query features
- Prevent feature inversion on server
- Allow camera pose estimation

ICCV 2019

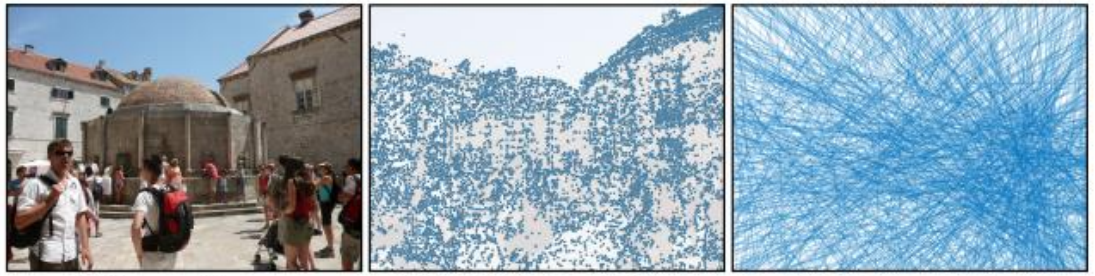
Privacy Preserving Image Queries for Camera Localization

Pablo Speciale[1,2] Johannes L. Schönberger[2] Sudipta N. Sinha[2] Marc Pollefeys[1,2]

[1] ETH Zürich [2] Microsoft

Abstract

Augmented/mixed reality and robotic applications are increasingly relying on cloud-based localization services, which require users to upload query images to perform camera pose estimation on a server. This raises significant privacy concerns when consumers use such services in their homes or in confidential industrial settings. Even if only image features are uploaded, the privacy concerns remain as the images can be reconstructed fairly well from feature locations and descriptors. We propose to conceal the content of the query images from an adversary on the server or a man-in-the-middle intruder. The key insight is to replace the 2D image feature points in the query image with randomly oriented 2D lines passing through their original



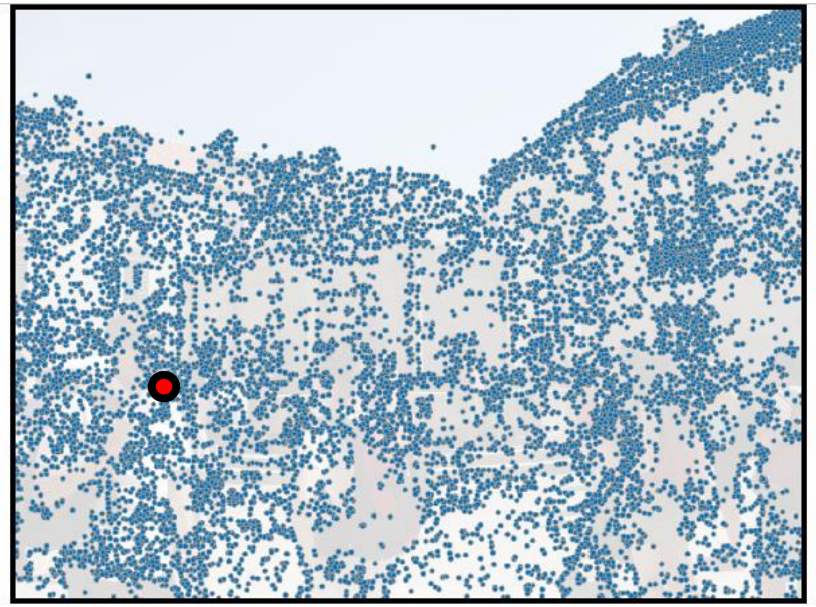
(a) Query Image (b) 2D Feature Points (c) 2D Feature Lines

Figure 1: Main Idea. Replace each 2D feature point in the query image with a randomly oriented 2D feature line passing through it.

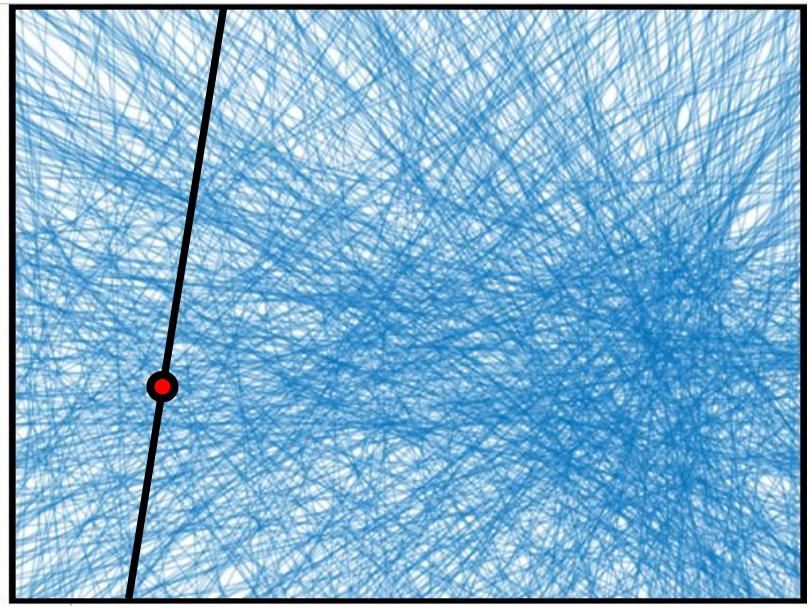
scene geometry while retaining sufficient constraints for robust and accurate camera pose estimation in many settings. Their representation [60] thus makes it possible to share maps with client devices without compromising privacy and enables privacy preserving localization on a local device. Alternatively, learning-based methods [13, 30, 71, 73] par-

Proposed Idea

Query Image



2D Feature Points

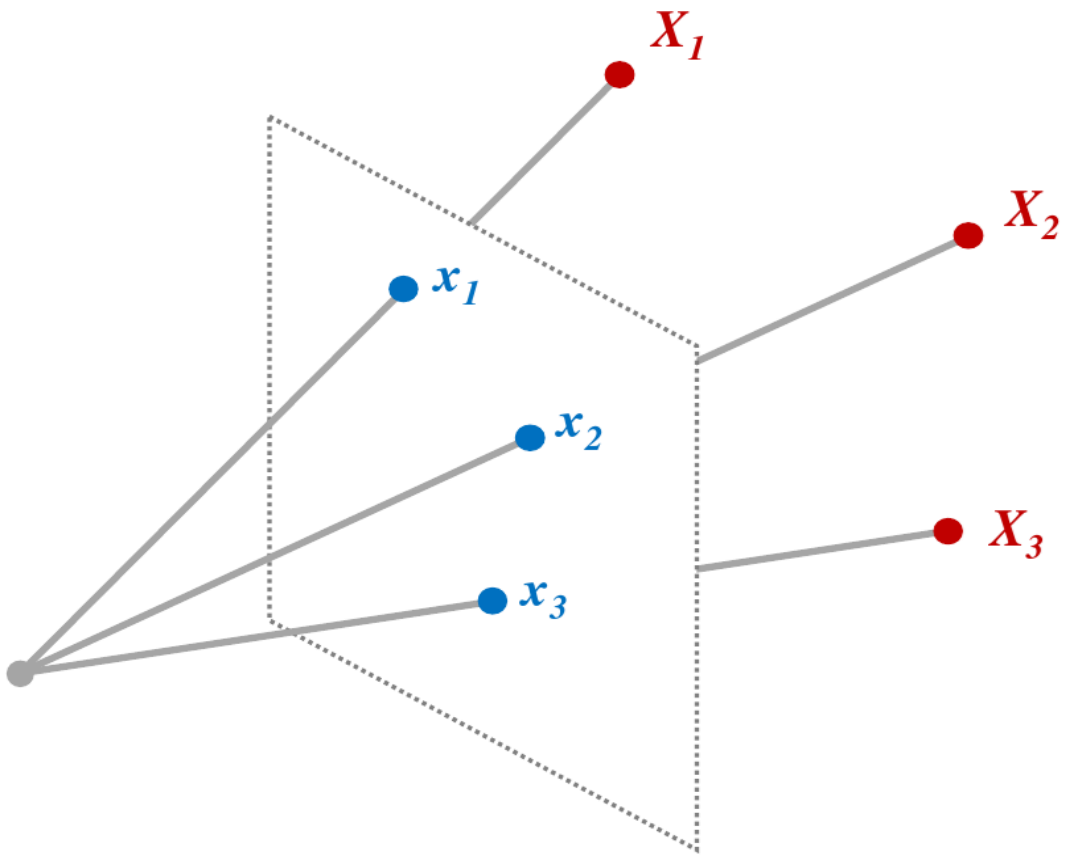


2D Feature Lines

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

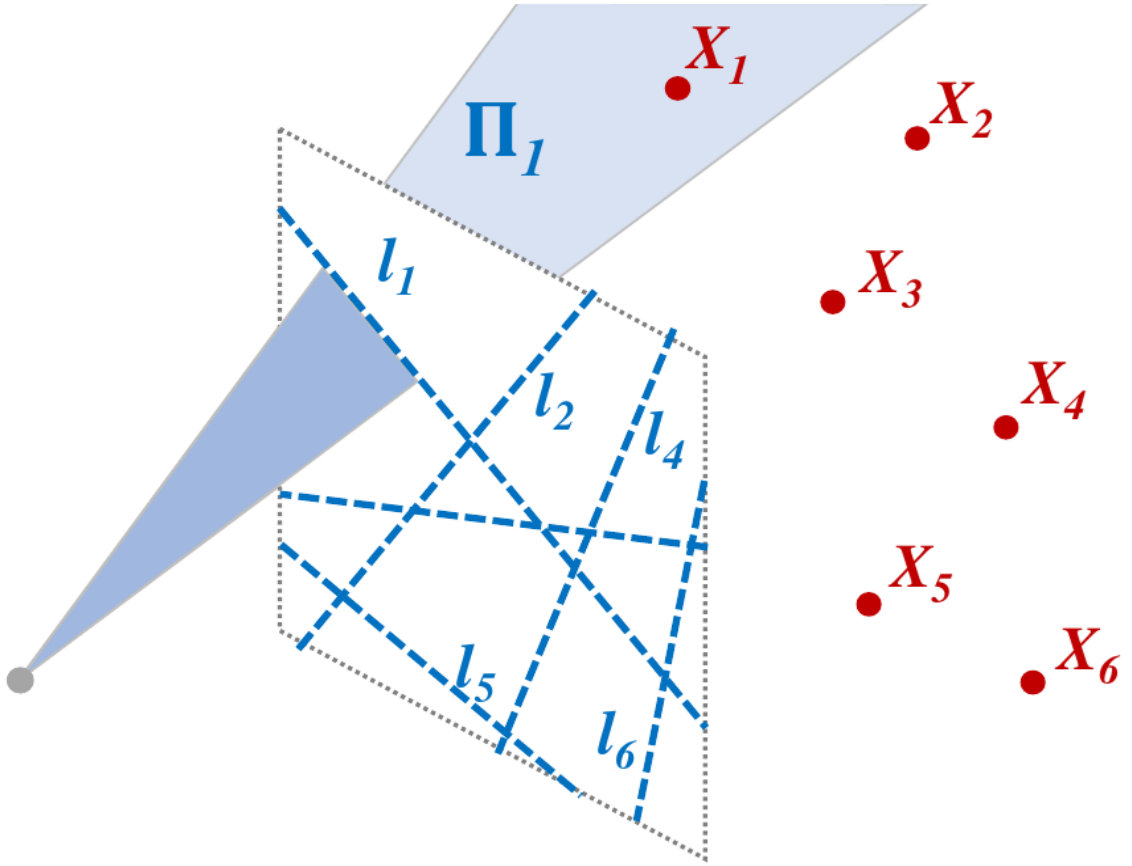
Geometric Constraints

Traditional Method ($p3P$)



Three 2D *image point* – 3D point correspondences

Proposed Method ($l6P$)



Six 2D *image line* – 3D point correspondences

2D Feature Line-based Pose Estimation

1. Can be cast as **Point-to-Plane problem** [Ramalingam et al. 2013]
2. Paper explored several variants
   - multi-image queries,
   - known vertical direction,
   - Known scale.
3. Also leverages existing minimal solvers from the literature.
4. Computationally efficient and can be used with RANSAC.

What gets revealed after pose estimation?

Query Image

Reconstructed Image
(using all features)

Reconstructed Image
(using only *revealed* features)

Privacy-Preserving Image Features via Adversarial Affine Subspace Embeddings

CVPR 2021

Mihai Dusmanu[1] Johannes L. Schönberger[2] Sudipta N. Sinha[2] Marc Pollefeys[1,2]
[1] Department of Computer Science, ETH Zürich [2] Microsoft

Abstract

Many computer vision systems require users to upload image features to the cloud for processing and storage. These features can be exploited to recover sensitive information about the scene or subjects, e.g., by reconstructing the appearance of the original image. To address this privacy concern, we propose a new privacy-preserving feature representation. The core idea of our work is to drop constraints from each feature descriptor by embedding it within an affine subspace containing the original feature as well as adversarial feature samples. Feature matching on the privacy-preserving representation is enabled based on the notion of subspace-to-subspace distance. We experimentally demonstrate the effectiveness of our method and its high practical relevance for the applications of visual localization and mapping as well as face authentication. Compared to the original features, our approach makes it significantly more difficult for an adversary to recover private information.

1. Introduction

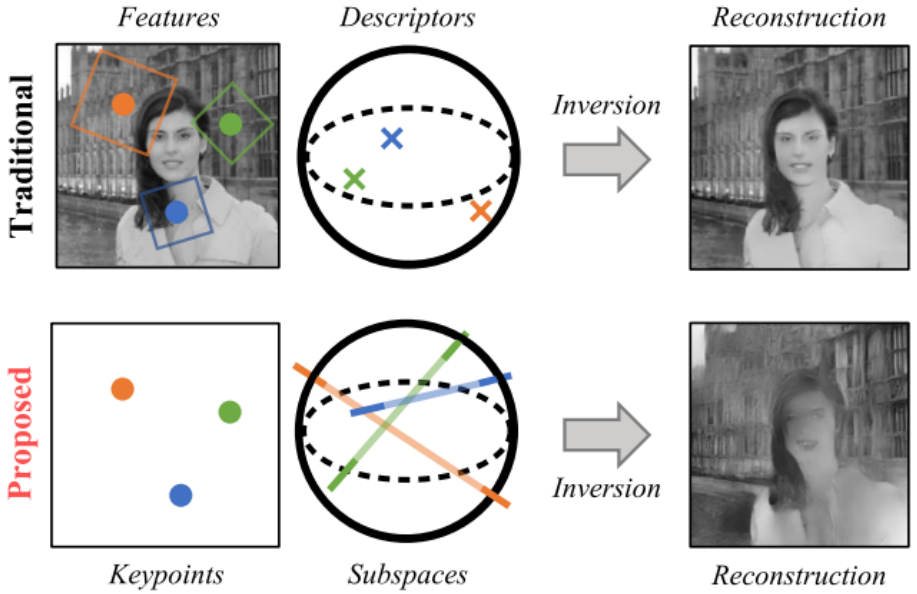


Figure 1: **Privacy-Preserving Image Features.** Inversion of traditional local image features is a privacy concern in many applications. Our proposed approach obfuscates the appearance of the original image by lifting the descriptors to affine subspaces. Distance between the privacy-preserving subspaces enables efficient matching of features. The same concept can be applied to other domains such as face features for biometric authentication. Image credit: *laylam-oran4battersea* (Layla Moran).

Conclusions

- New task and goals: Privacy-preserving camera localization:
  - motivated by Mixed reality, Robotics and spatial AI applications.
- Approaches explored so far:
  - Privacy-preservation of maps
    - Conceal 3D point cloud by transforming points into 3D lines.
    - But shown to be not very effective; no good solutions exist yet!
  - Privacy-preservation of query images for cloud-based localization
    - Conceal 2D feature positions; or conceal the feature descriptor vectors.

Related Work and Future Directions

- Privacy-preserving Structure from Motion, SLAM
  - extends idea of 2D feature lines [Geppert et al. 2020]
  - 3D line clouds in SLAM [Shibuya et. al. 2020]
- Are there better ways to construct 3D line clouds maps?
  - Build upon insights shared by [Chelani et. al. 2021]
- Are there better representations?
- Can learned localization approaches address privacy concerns?
- Can we develop a theoretic framework for more rigorous analysis of the accuracy/privacy tradeoffs?

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.