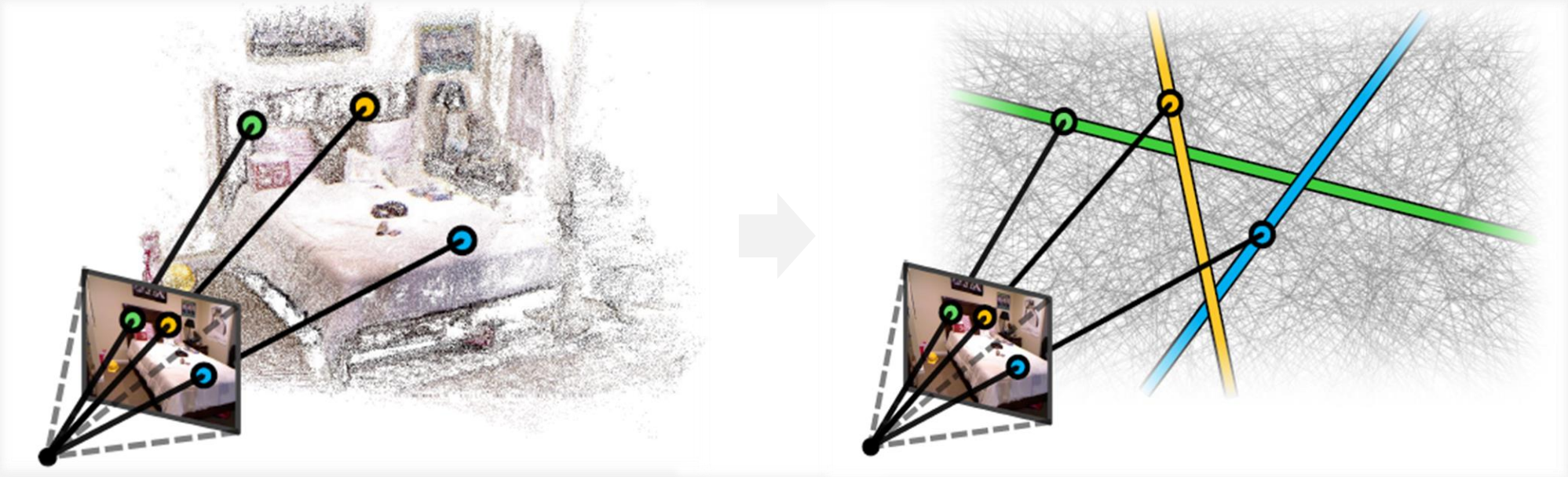


Towards Privacy-Preserving Spatial AI

Sudipta N. Sinha
Microsoft
(joint work with Francesco Pittaluga, Sanjeev Koppal, Sing Bing Kang,
Pablo Speciale, Johannes Schonberger and Marc Pollefeys)

Advancing AI @ WSU seminar, October 14, 2020

Microsoft HoloLens, Mixed Reality & AI

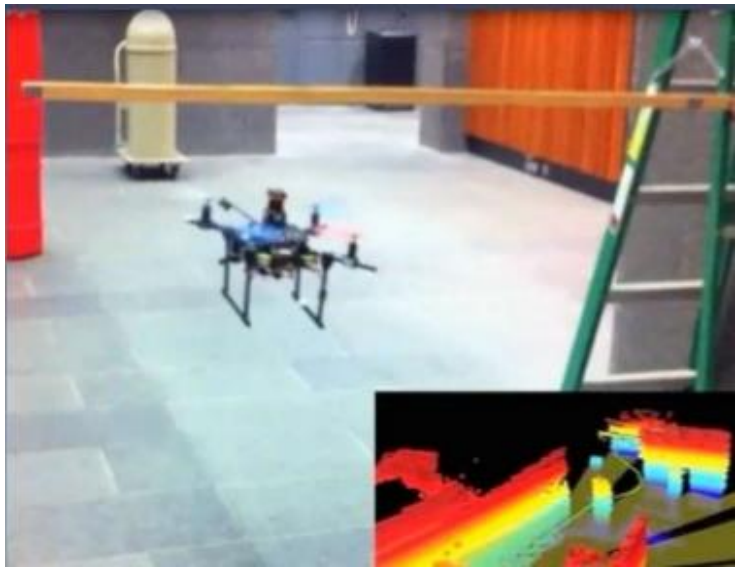
- Milgram and Kishino 1994, *"A Taxonomy of Mixed Reality Visual Displays".*
- Microsoft HoloLens is an untethered mixed reality device that blends world-locked 3D digital content (holograms) into the real world.

Microsoft HoloLens, Mixed Reality & AI

- Enabled by advances in computer vision, AI, graphics and display technologies.
- Human understanding and environment understanding

Environment Understanding (Spatial AI)

Mixed Reality



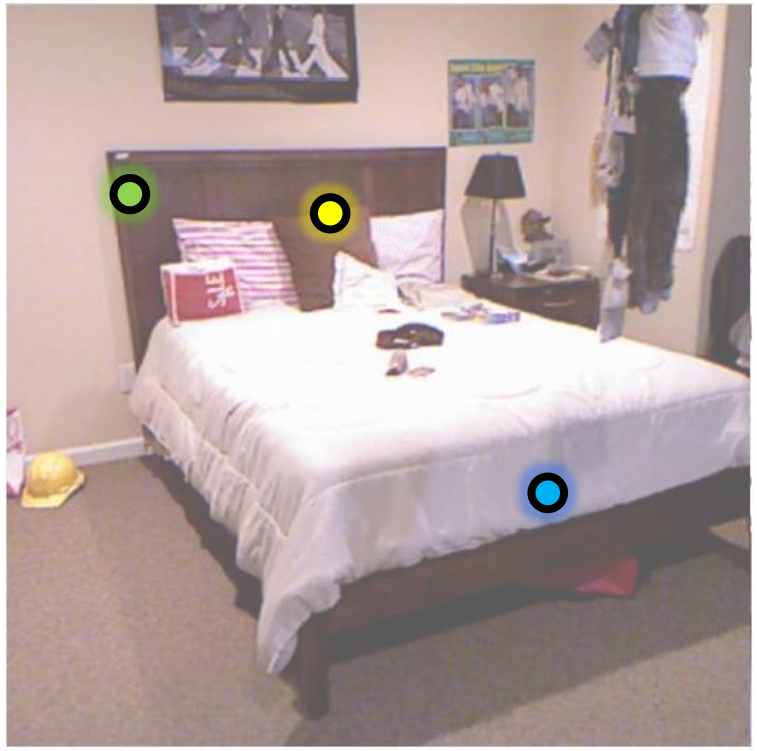
Autonomous Drones, Robots

Augmented Reality
on mobile devices

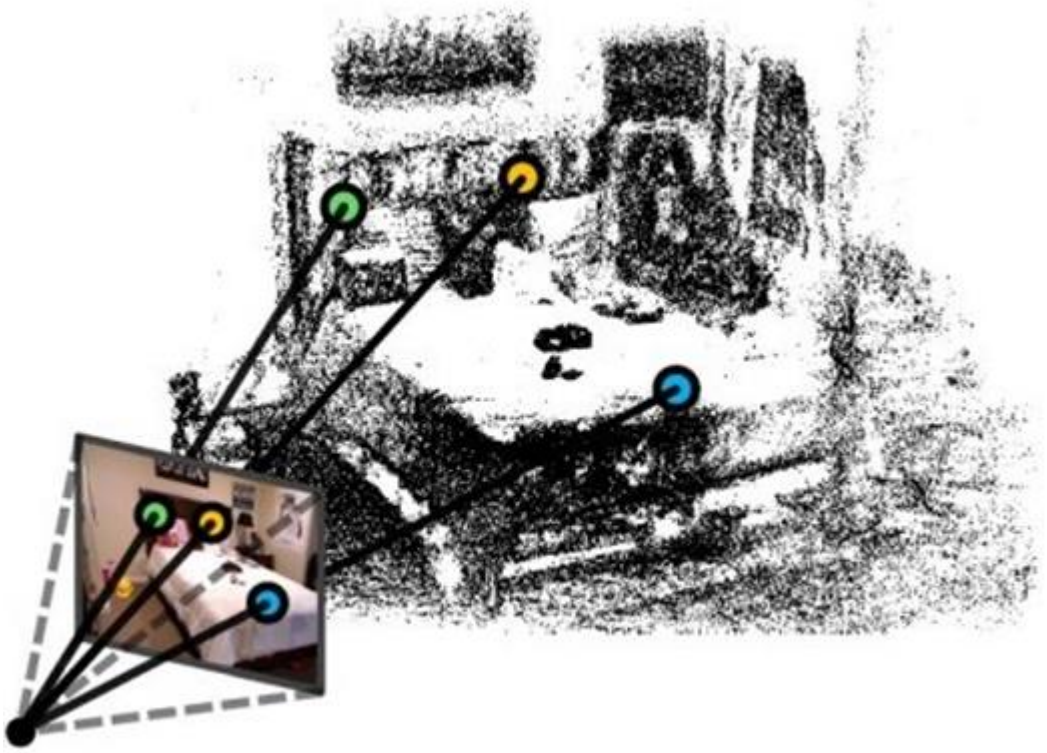
- Vision/IMU-based Mapping and Localization.
- Real-time 6-DOF pose provides situational awareness (cloud back-end for scalability).

Camera Localization

Query Image



Precomputed Map (3D Point Cloud)



Camera Localization

Arth+ 2009

Irschara+ 2009

Sattler+ 2011

Li+ 2012

Lim+ 2012

Ventura+ 2014

Zeisl+ 2015

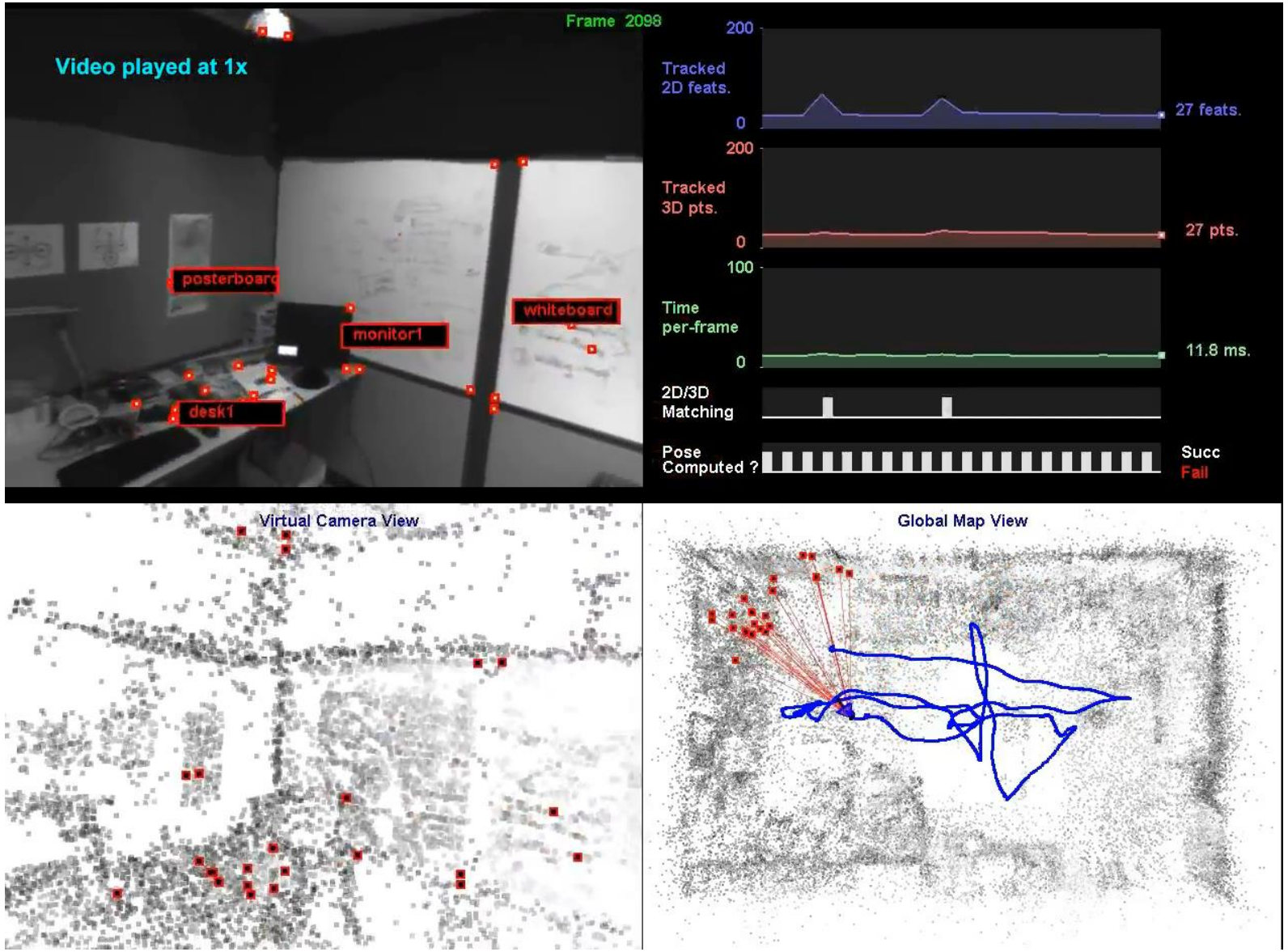
Sattler+ 2015

Lynen+ 2015

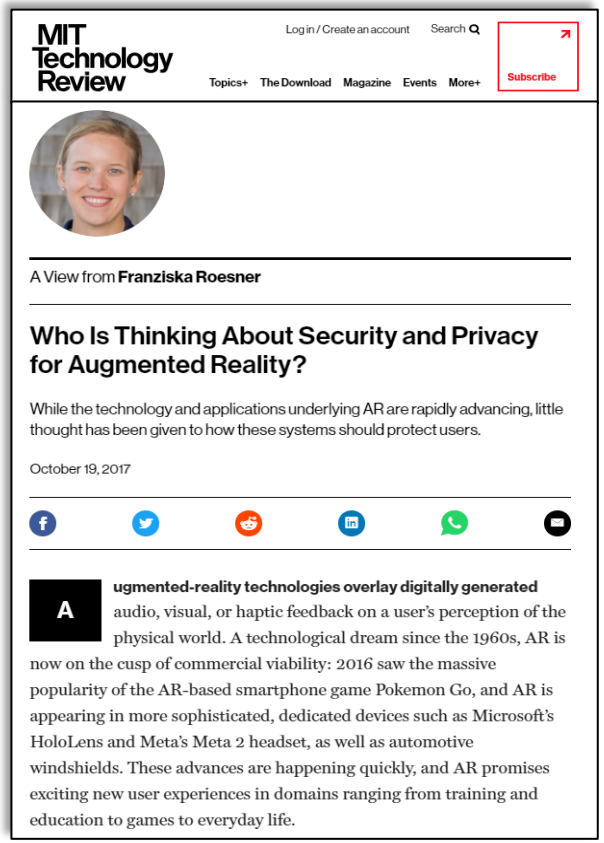
Kendall+ 2015

Weyand+ 2016

…



Privacy Concerns in Augmented Reality, Robotics



MIT Technology Review

Log in / Create an account Search Subscribe

Topics+ The Download Magazine Events More+

A View from **Franziska Roesner**

Who Is Thinking About Security and Privacy for Augmented Reality?

While the technology and applications underlying AR are rapidly advancing, little thought has been given to how these systems should protect users.

October 19, 2017

Augmented-reality technologies overlay digitally generated audio, visual, or haptic feedback on a user's perception of the physical world. A technological dream since the 1960s, AR is now on the cusp of commercial viability: 2016 saw the massive popularity of the AR-based smartphone game Pokemon Go, and AR is appearing in more sophisticated, dedicated devices such as Microsoft's HoloLens and Meta's Meta 2 headset, as well as automotive windshields. These advances are happening quickly, and AR promises exciting new user experiences in domains ranging from training and education to games to everyday life.

Privacy Manifesto for AR Cloud Solutions

DRAFT v 0.1.3, October 18th at AWE EU Münich

Jan-Erik Vinje Follow

Oct 17, 2018 · 5 min read

The next era of computing is upon us. New technology to capture and combine physical spaces with digital content has the potential to profoundly improve the way we see and interact with the world and each other.

As the world's largest companies and organizations race to create the required "AR Cloud" infrastructure to build and fuel these systems, we face unprecedented new challenges and risks to privacy and individual's rights

BeyondStandards

IEEE STANDARDS ASSOCIATION

Home › Augmented Reality › Augmented Reality and its Impact on the Internet, Security, and Privacy

Augmented Reality and its Impact on the Internet, Security, and Privacy

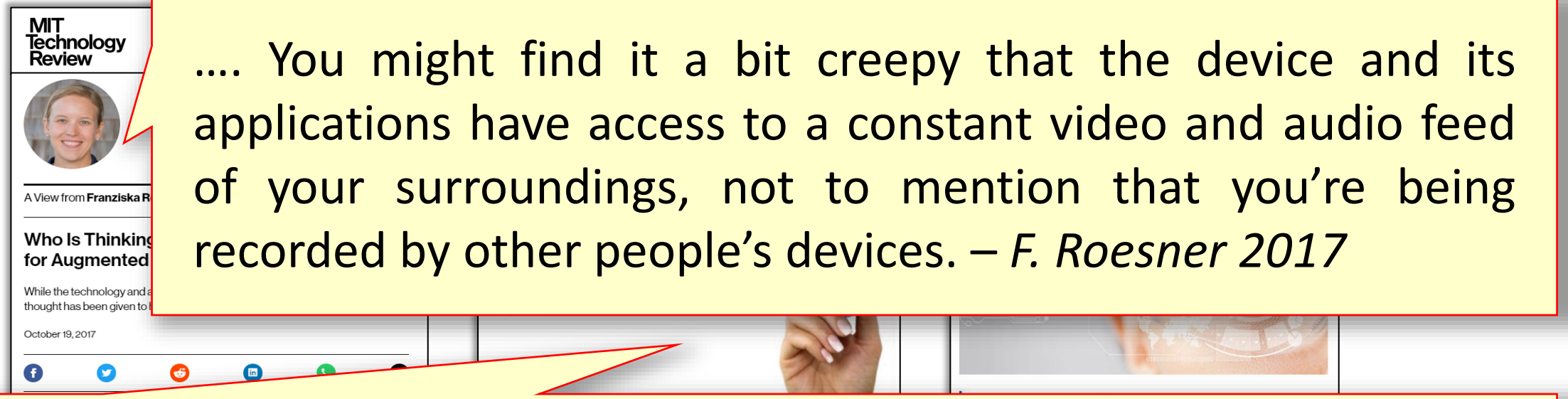
10 July 2015 0 comments Guest Contributor Posted in Augmented Reality

By Mary Lynne Nielsen, Global Operations and Outreach Program Director, IEEE Standards Association

Recently, I had the opportunity to lead networking sessions at Augmented World Expo (AWE) 2015. I interacted with conference attendees to explore the challenges and opportunities facing the evolution of the Internet. Now in its sixth year, AWE is dedicated to exploring technology that turns ordinary experiences into the extraordinary and empowers people to be better at anything they do in work and life. Nearly 3000 people from the augmented and virtual reality, wearable tech, and Internet of Things spaces attended this year's event.

https://beyondstandards.ieee.org/augmented-reality/augmented-reality-and-its-impact-on-the-internet-security-and-privacy/
https://www.technologyreview.com/s/609143/who-is-thinking-about-security-and-privacy-for-augmented-reality/
https://medium.com/openarcloud/privacy-manifesto-for-ar-cloud-solutions-9507543f50b6

Privacy Concerns in Augmented Reality, Robotics



…. You might find it a bit creepy that the device and its applications have access to a constant video and audio feed of your surroundings, not to mention that you're being recorded by other people's devices. – *F. Roesner 2017*

…….. the risk that such *(private)* data could be collected, analyzed, transmitted and stored in databases or distributed and sold to third parties without the explicit consent of users or worse, unsuspecting citizens that happen to be within sensor range of mixed reality enabled devices. – *Jan-Erik Vinje 2018*

Outline

- Revealing Scenes by Inverting Structure from Motion Reconstructions

  Francesco Pittaluga, Sanjeev Koppal, Sing Bing Kang and Sudipta N. Sinha

  CVPR 2019

- Privacy-Preserving Image-based Localization

  Pablo Speciale, Johannes L. Schönberger, Sing Bing Kang, Sudipta N. Sinha and Marc Pollefeys

  CVPR 2019

- Privacy-Preserving Image Queries for Camera Localization

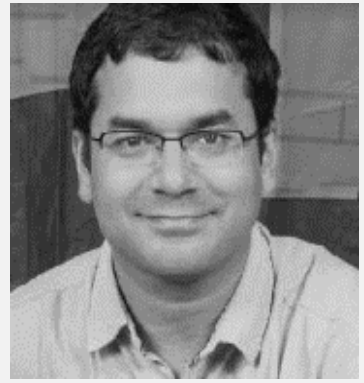
  Pablo Speciale, Johannes L. Schönberger, Sudipta N. Sinha and Marc Pollefeys

  ICCV 2019

Revealing Scenes by Inverting Structure from Motion Reconstructions

CVPR 2019

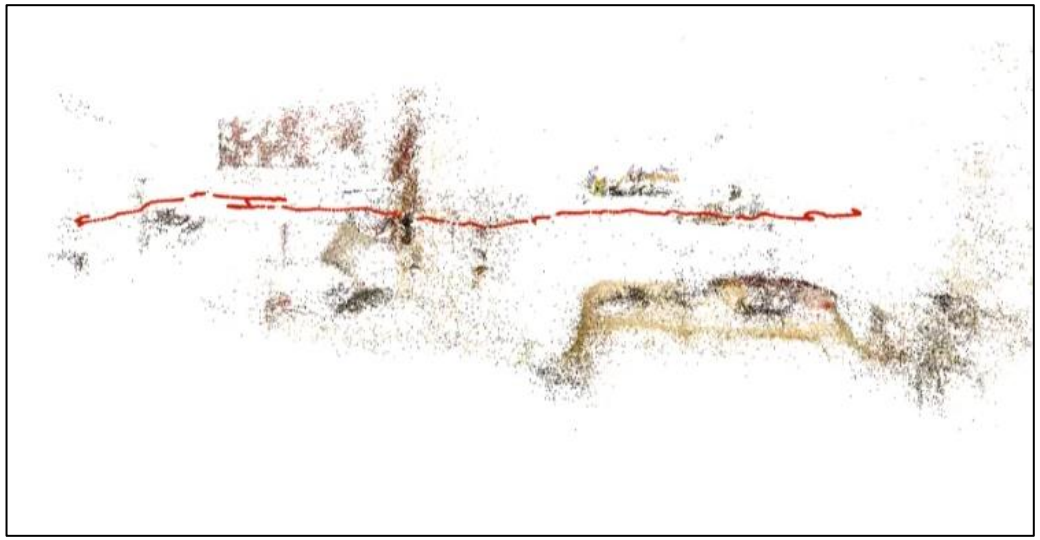


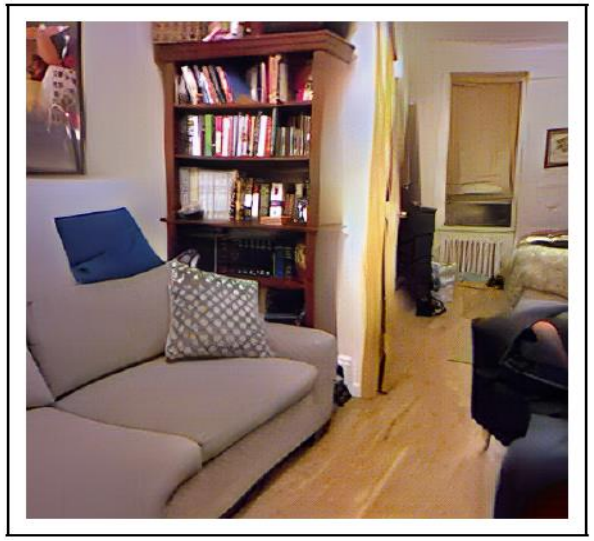
Francesco Pittaluga [1] Sanjeev Koppal [1] Sing Bing Kang[2] Sudipta N. Sinha[2]

[1] University of Florida [2] Microsoft Research

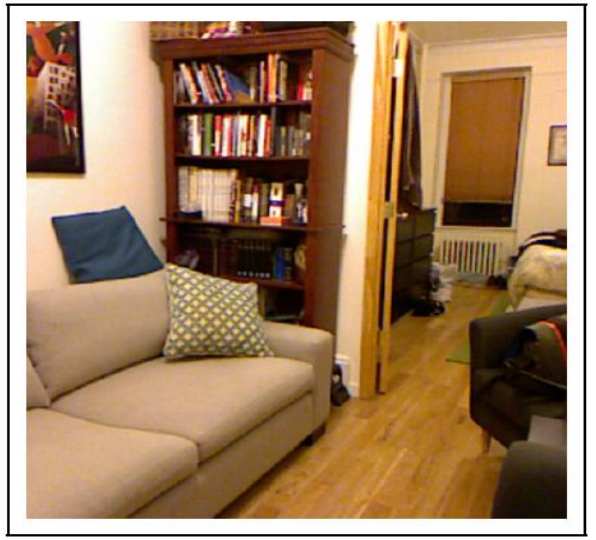
New Privacy Attack on 3D Maps



3D point cloud map



Reconstructed Image



Original Image

Map Inversion

- Input images are discarded after map construction. Sparse 3D points and descriptors are stored.
- The attack aims to reconstruct images of the scene from the stored 3D points and descriptors.
- We implement the attack using a deep neural network.

Reconstruction of Source Video used in Mapping

Projected 3D Points
(network input)

Reconstructed Frames

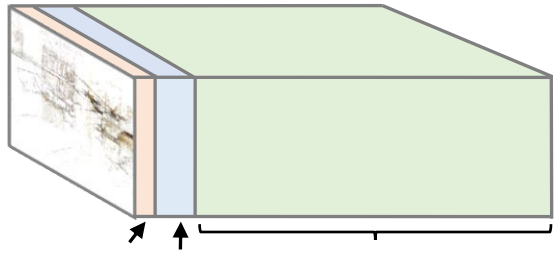
Original Frames

Neural Net Architecture

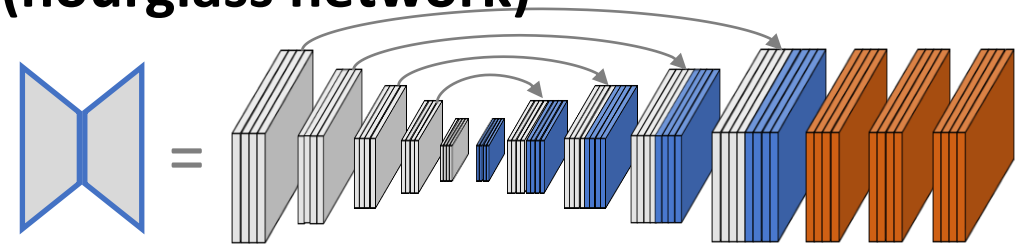
- Input: 2D projection of points + features.
- Training Data: Structure from motion data.

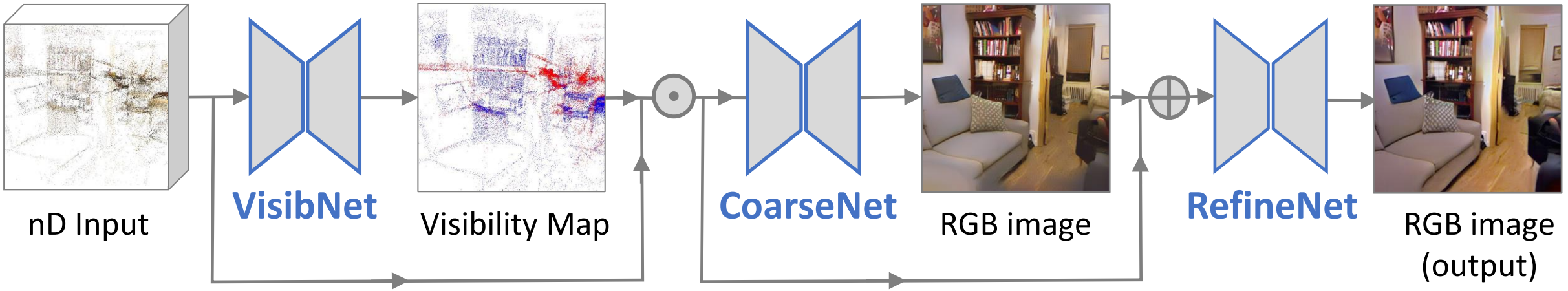
- Model: Three U-Net modules in series.
- Loss: L1 + Perceptual + Adversarial Loss

Input Tensor

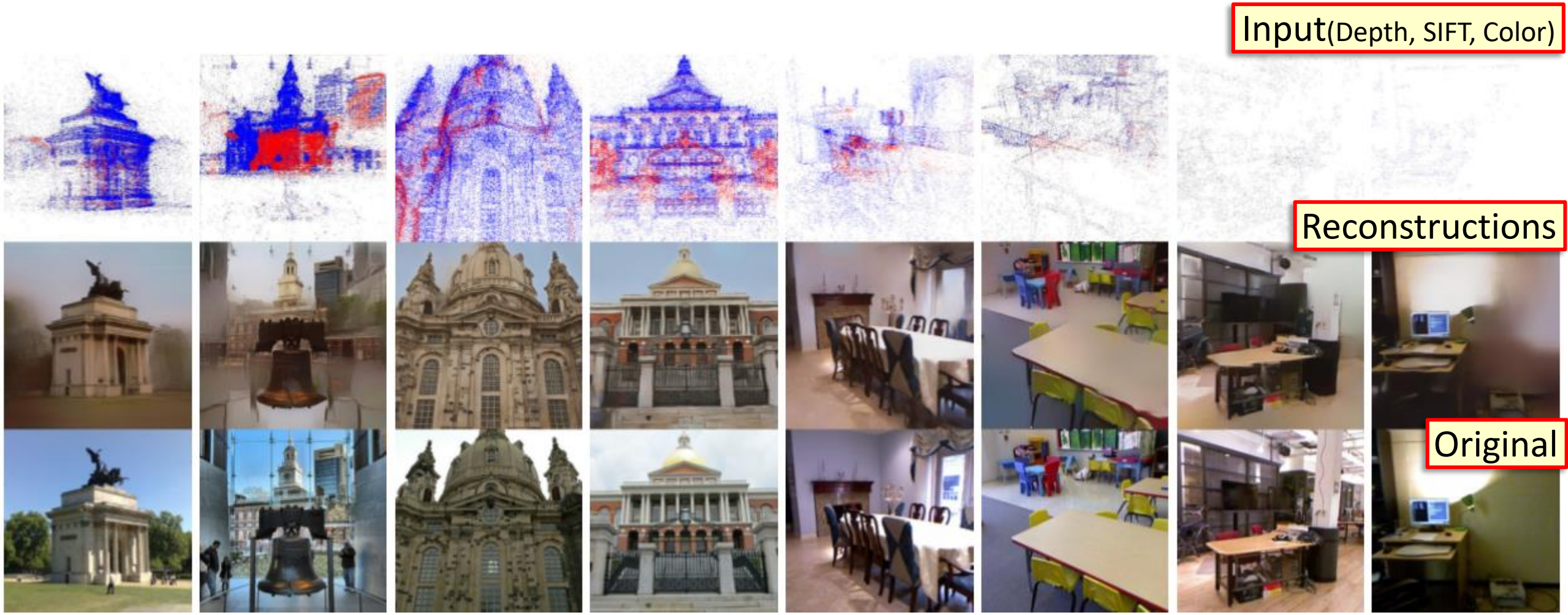


U-Net (hourglass network)



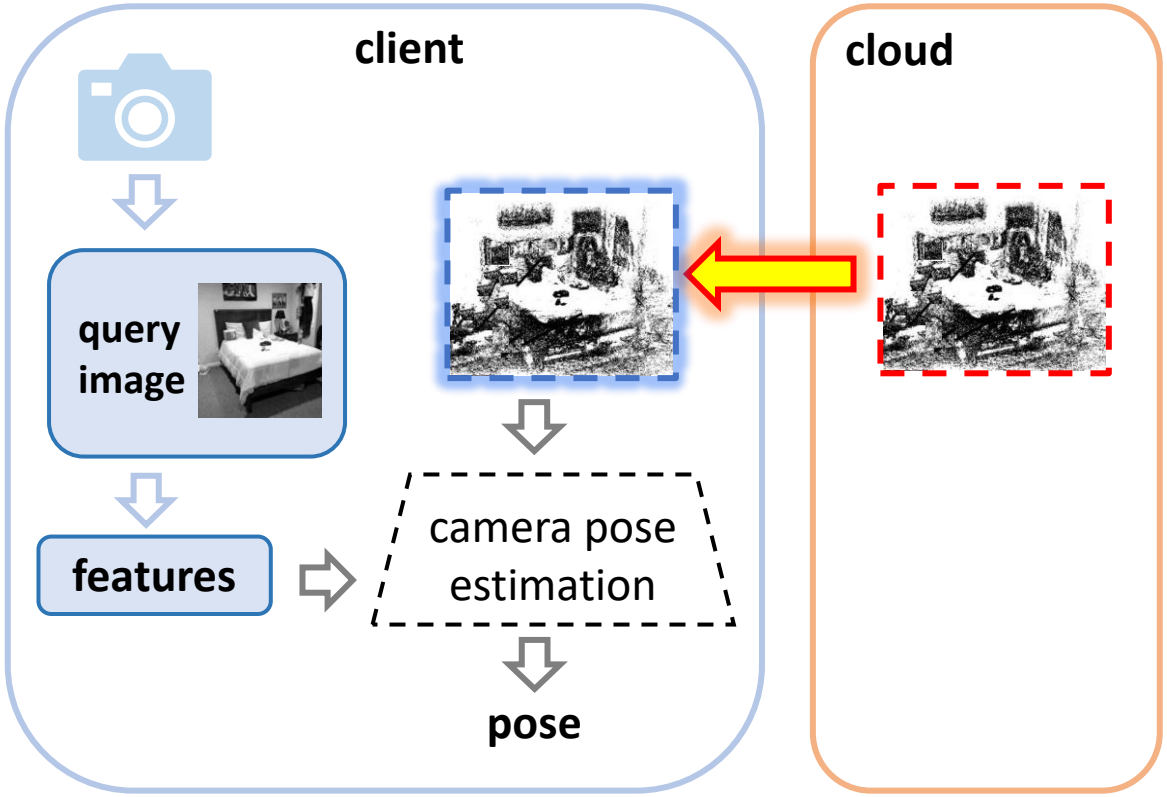


Results



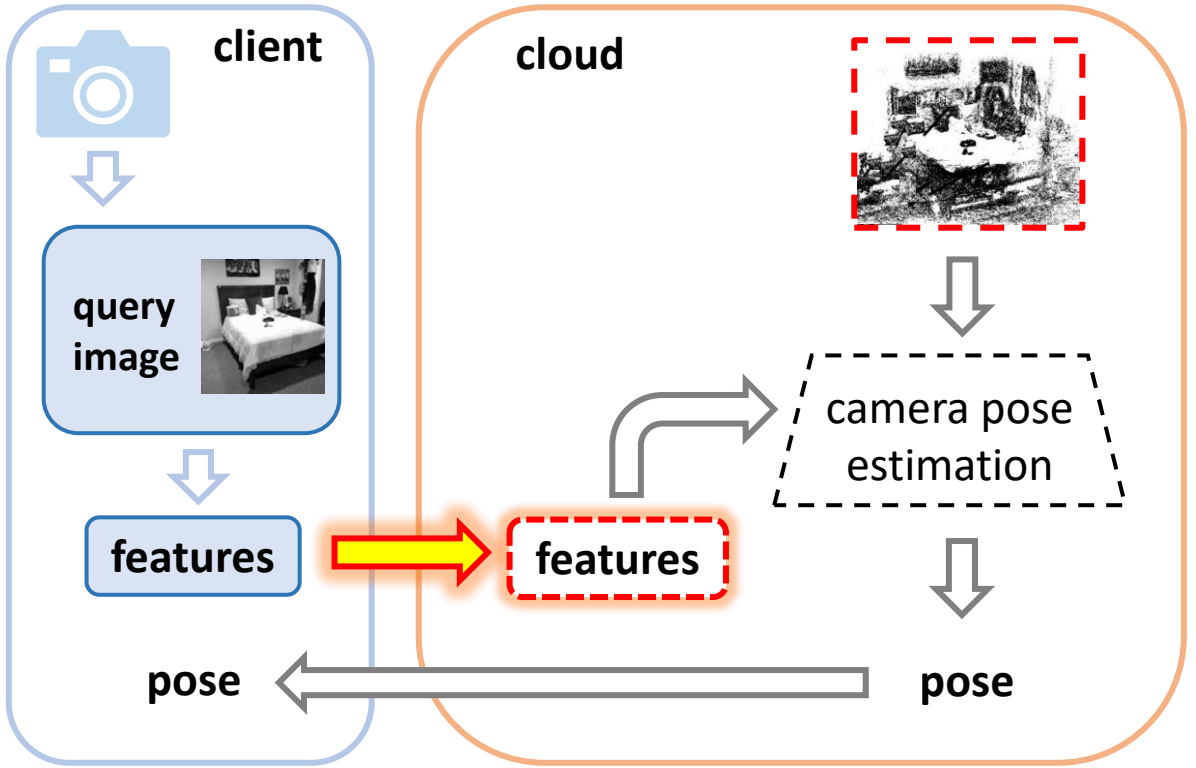
Privacy Implications for Camera Localization

Processing on Client



- Server shares map with clients
- Privacy of map data is a concern

Processing on Server



- Client shares image features with server
- Privacy of query image is a concern

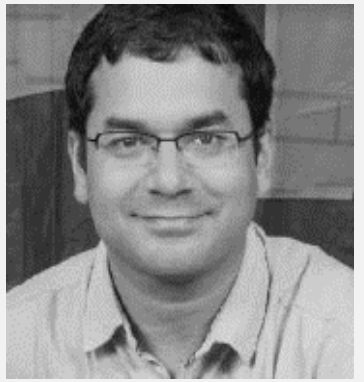
Privacy-Preserving Image-based Localization

CVPR 2019

Pablo Speciale[1]

Johannes L. Schönberger[1]

Sing Bing Kang[2]



Sudipta N. Sinha[2]

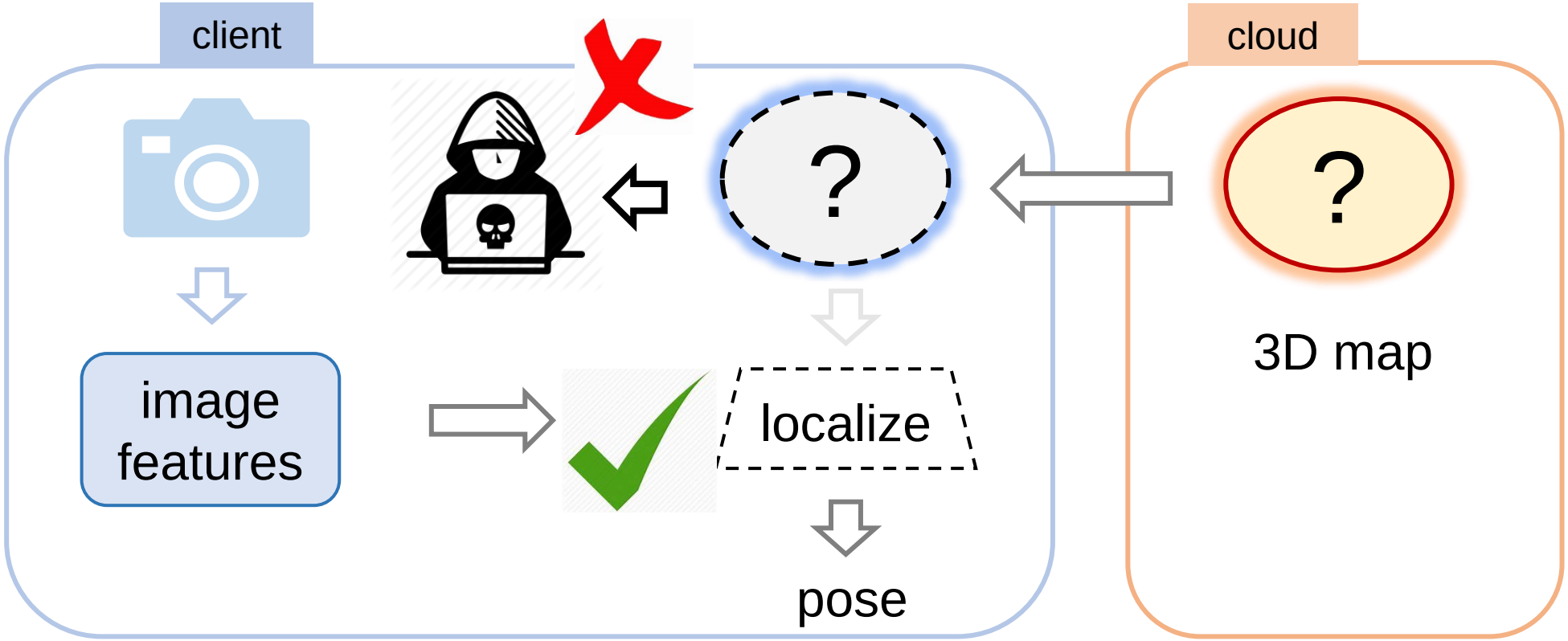
Marc Pollefeys[1,3]

[1] Microsoft Mixed Reality & AI Group, Zurich

[2] Microsoft Research Redmond

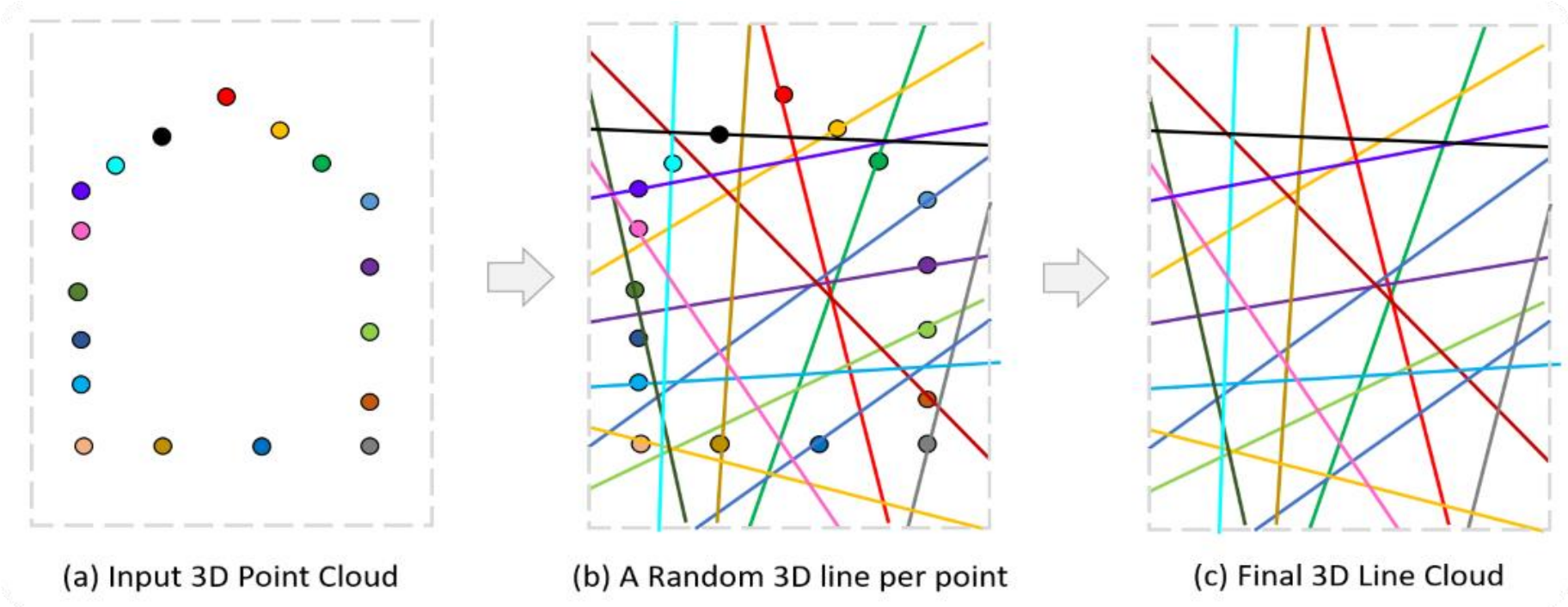
[3] ETH Zurich

Goal: Keep the Map Confidential



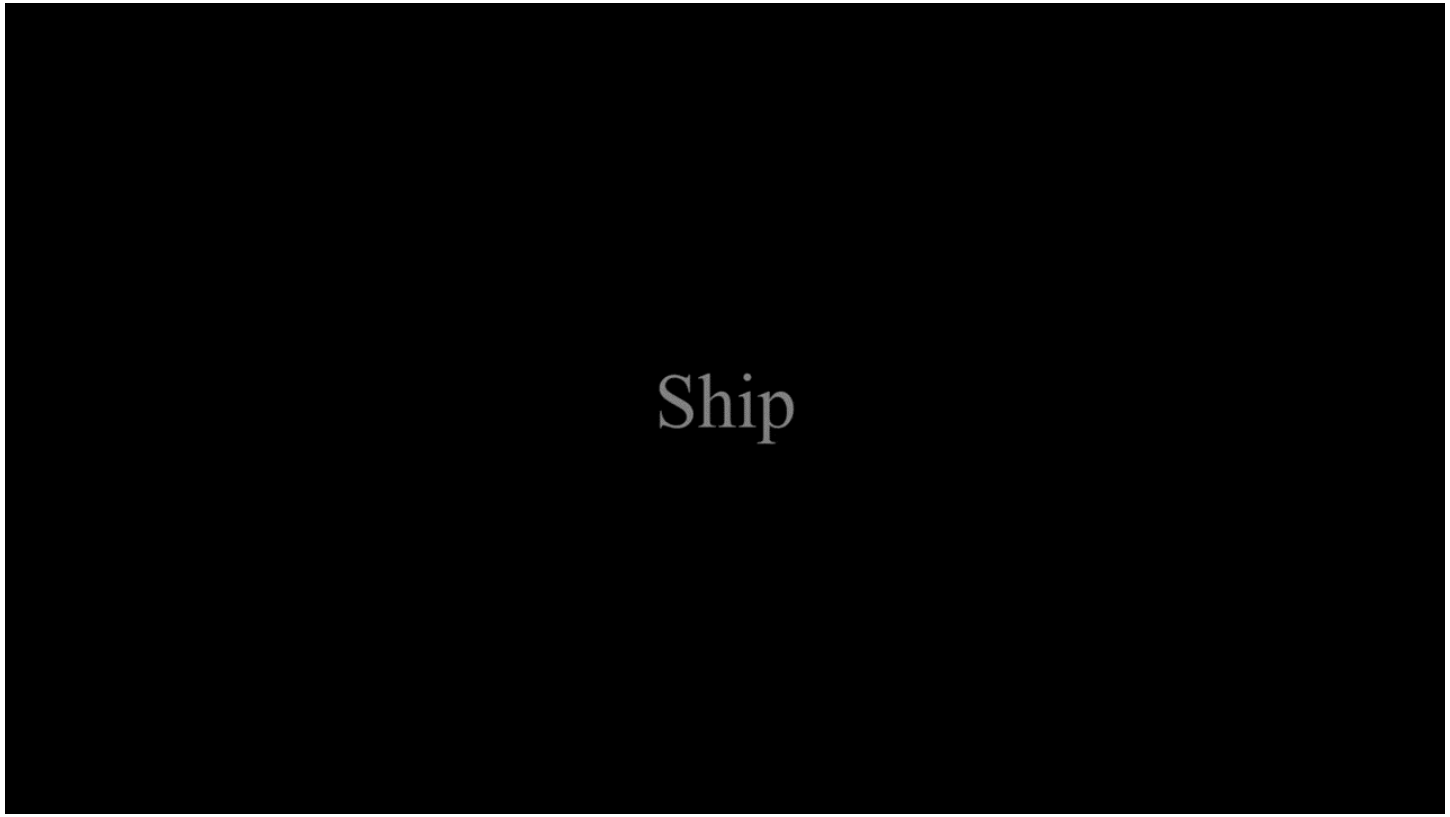
1. Conceal the 3D map; prevent inversion attacks.
2. Yet, somehow allow camera pose estimation!

New Map Representation



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

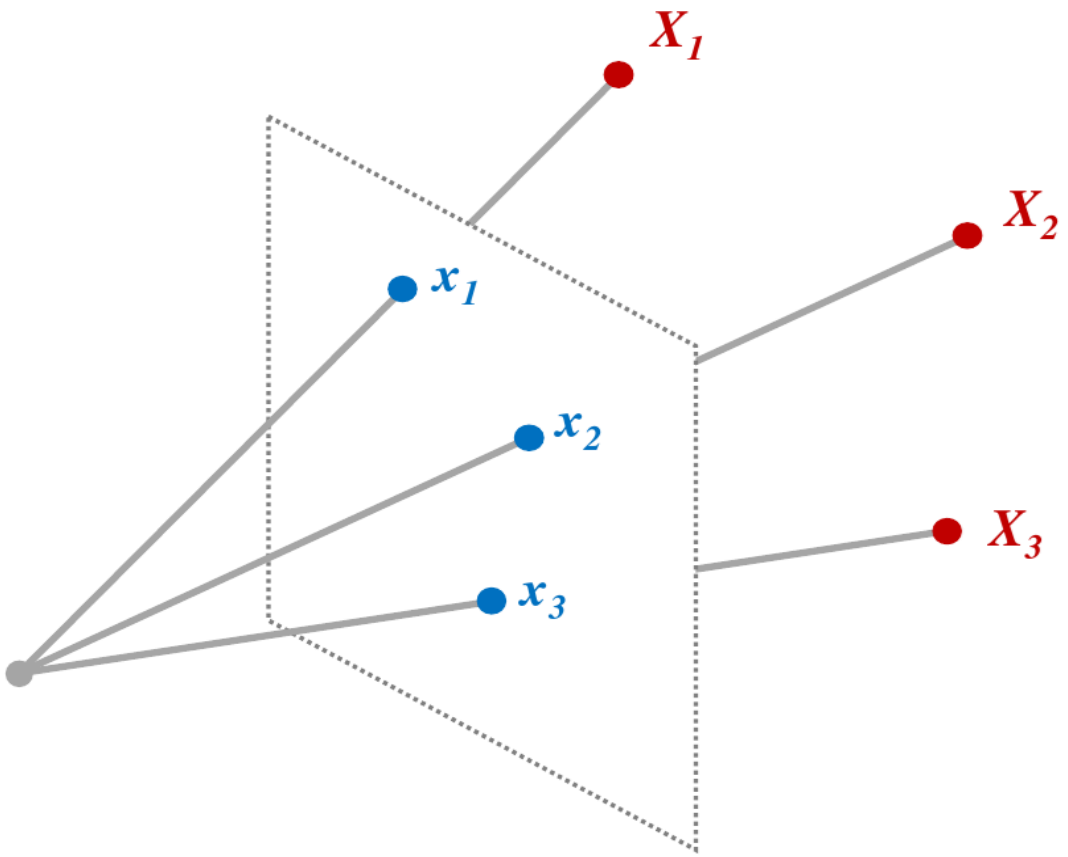
Proposed Idea



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

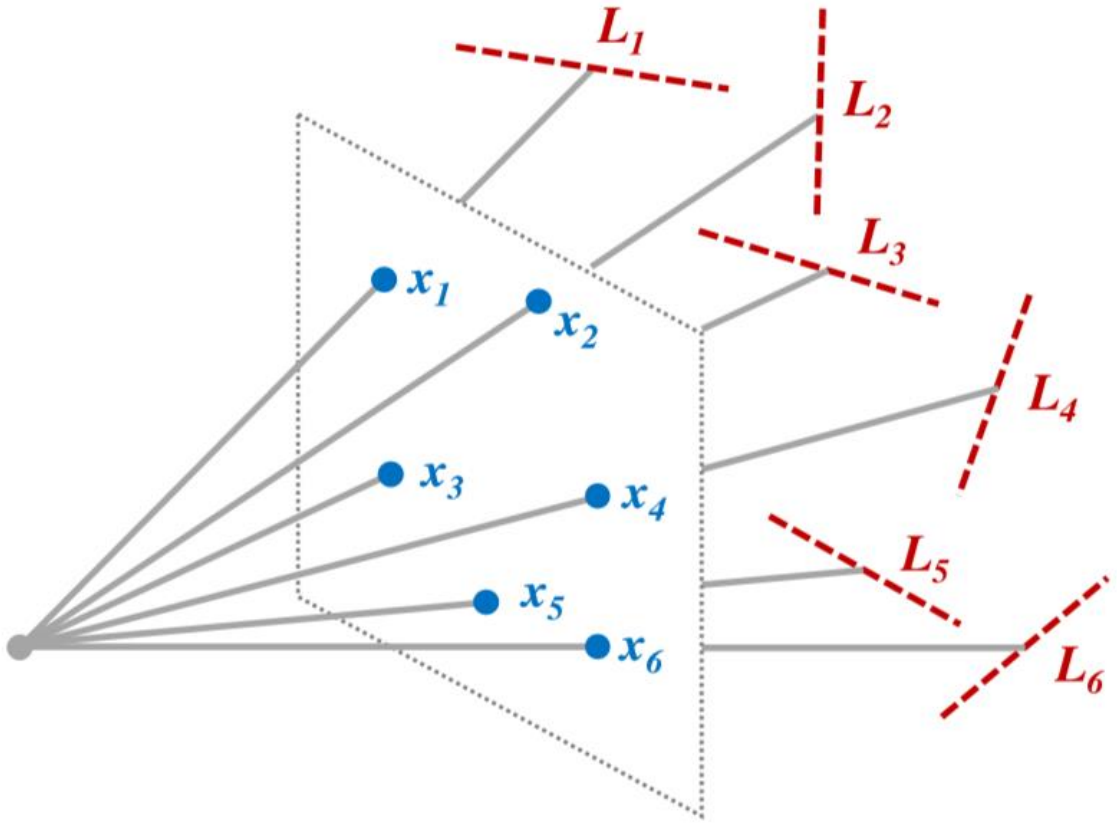
Camera Pose Estimation

Traditional Method ($p3P$)



Three “image point”–3D point correspondences

Proposed Method ($p6L$)



Six “image point”–3D line correspondences

Camera Pose Estimation

1. Our minimal problem can be cast as **generalized relative pose problem** [1].

   [1] Stewenius et al. 2005

2. Proposed **several variants** with:
   - Query 3D point cloud (from multiple images),
   - known vertical direction,
   - known scale.

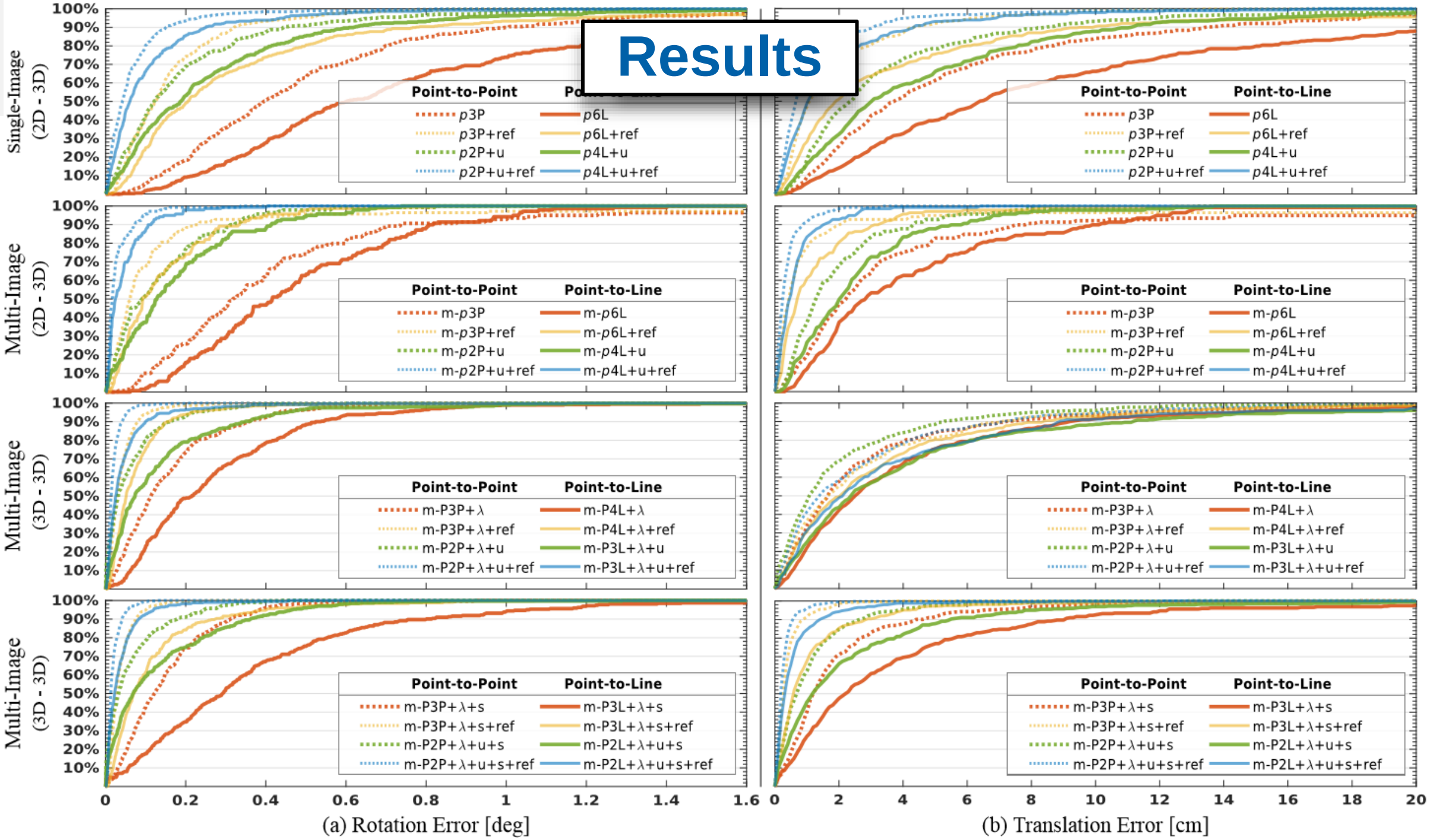
   We leverage existing minimal solvers [2-7].

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

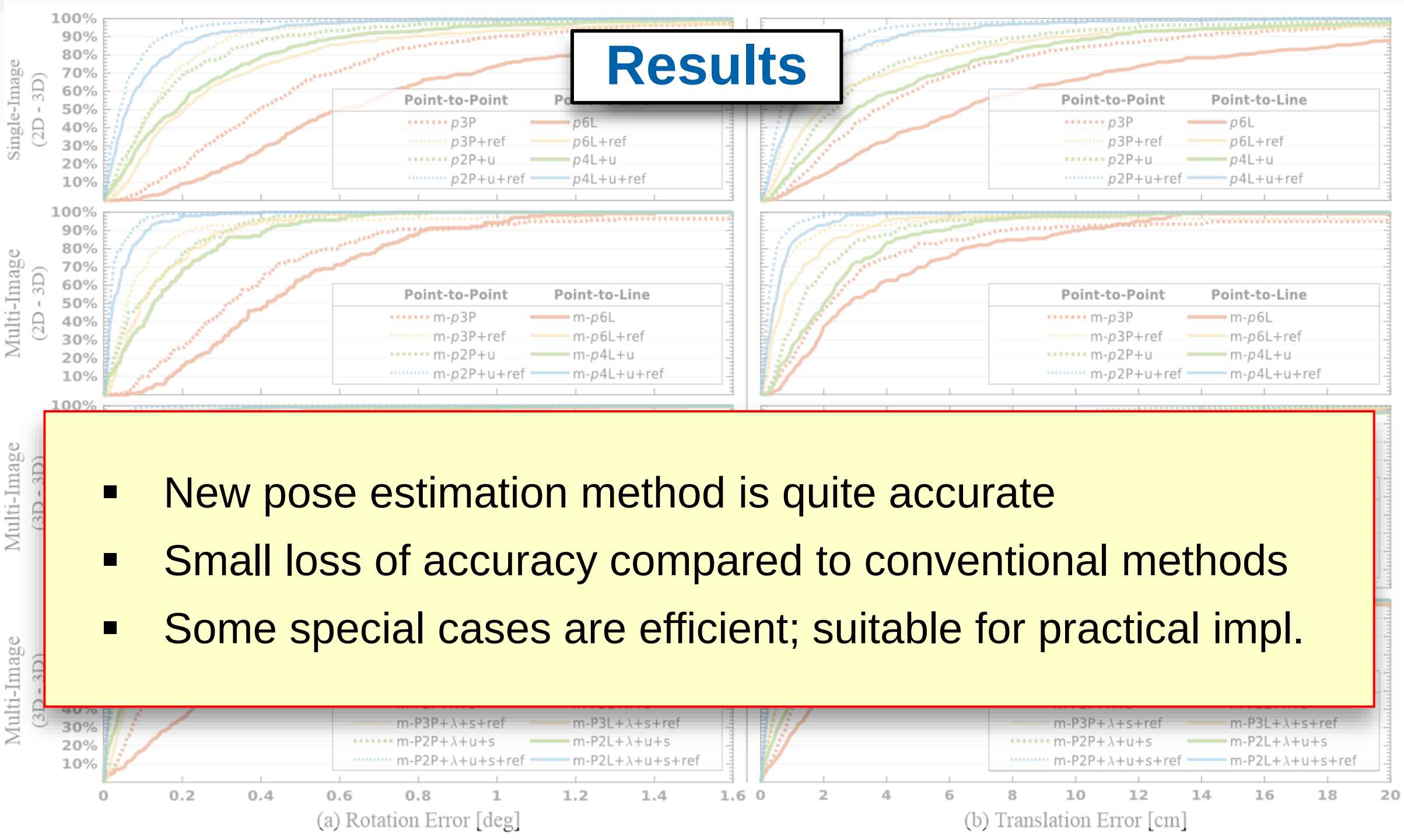
   [5] Sweeney et al. 2015a [6] Sweeney et al. 2015b [7] Sweeney et al. 2014

3. Most variants are **computationally efficient** and can be used with RANSAC.

Results


Single-Image (2D - 3D)
Multi-Image (2D - 3D)
Multi-Image (3D - 3D)
Multi-Image (3D - 3D)
Point-to-Point
Point-to-Line
p3P
p3P+ref
p2P+u
p2P+u+ref
p6L
p6L+ref
p4L+u
p4L+u+ref
m-p3P
m-p3P+ref
m-p2P+u
m-p2P+u+ref
m-p6L
m-p6L+ref
m-p4L+u
m-p4L+u+ref
m-P3P+λ
m-P3P+λ+ref
m-P2P+λ+u
m-P2P+λ+u+ref
m-P4L+λ
m-P4L+λ+ref
m-P3L+λ+u
m-P3L+λ+u+ref
m-P3P+λ+s
m-P3P+λ+s+ref
m-P2P+λ+u+s
m-P2P+λ+u+s+ref
m-P3L+λ+s
m-P3L+λ+s+ref
m-P2L+λ+u+s
m-P2L+λ+u+s+ref
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
(a) Rotation Error [deg]
0
2
4
6
8
10
12
14
16
18
20
(b) Translation Error [cm]

Results



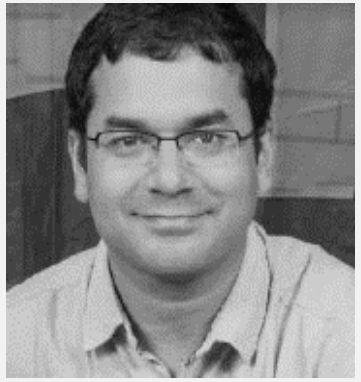
- New pose estimation method is quite accurate
- Small loss of accuracy compared to conventional methods
- Some special cases are efficient; suitable for practical impl.

Privacy-Preserving Image Queries for Camera Localization

ICCV 2019

Pablo Speciale[1]

Johannes L. Schönberger[1]



Sudipta N. Sinha[2]

Marc Pollefeys[1,3]

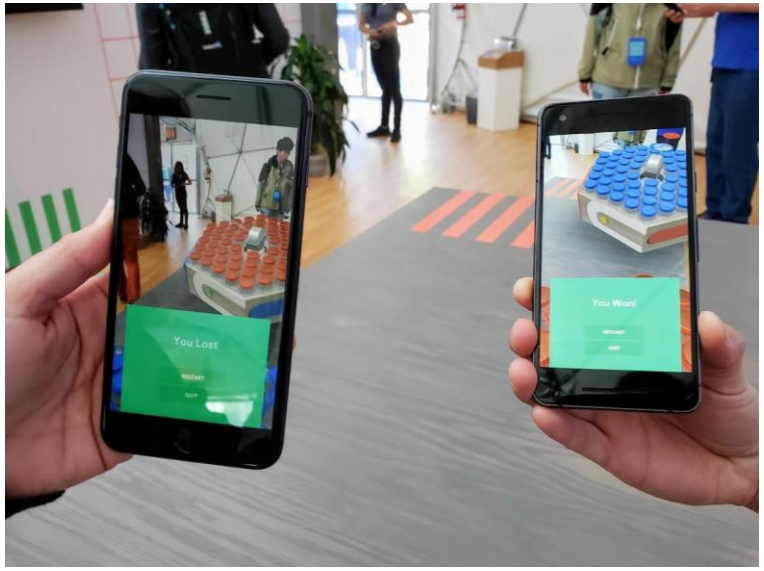
[1] Microsoft Mixed Reality & AI Group, Zurich

[2] Microsoft Research Redmond

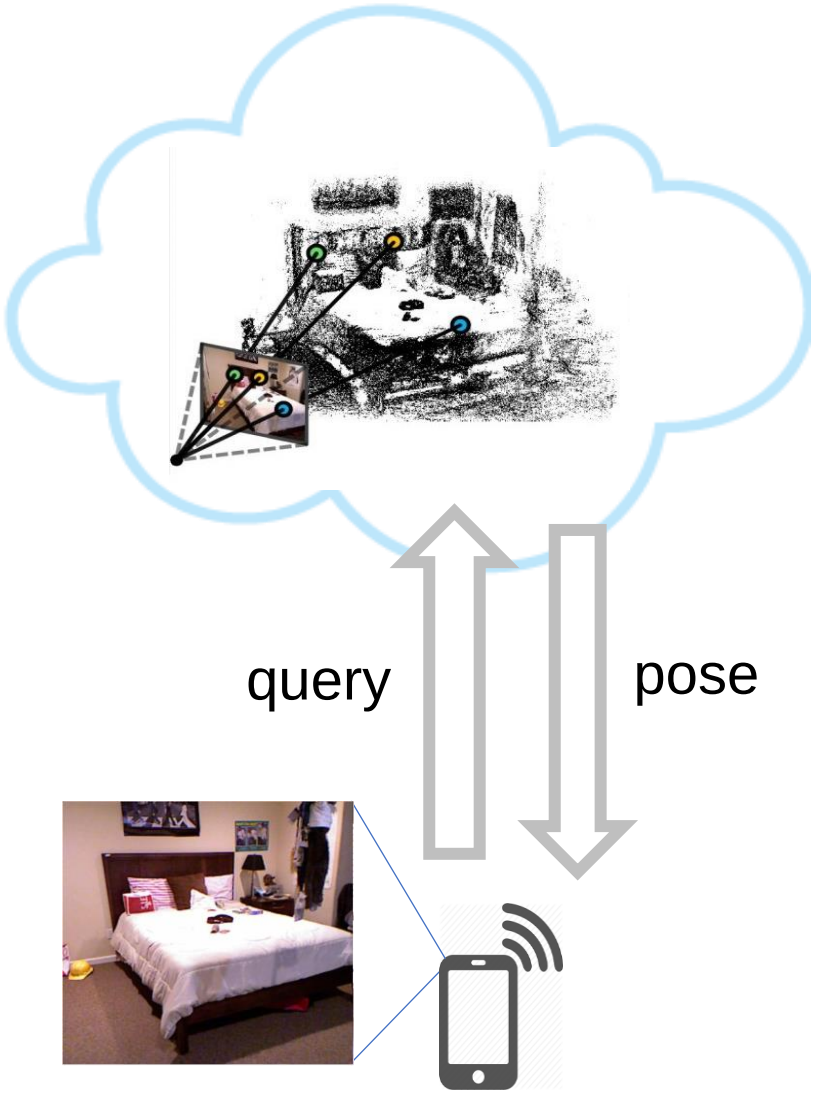
[3] ETH Zurich

Localization in the Cloud

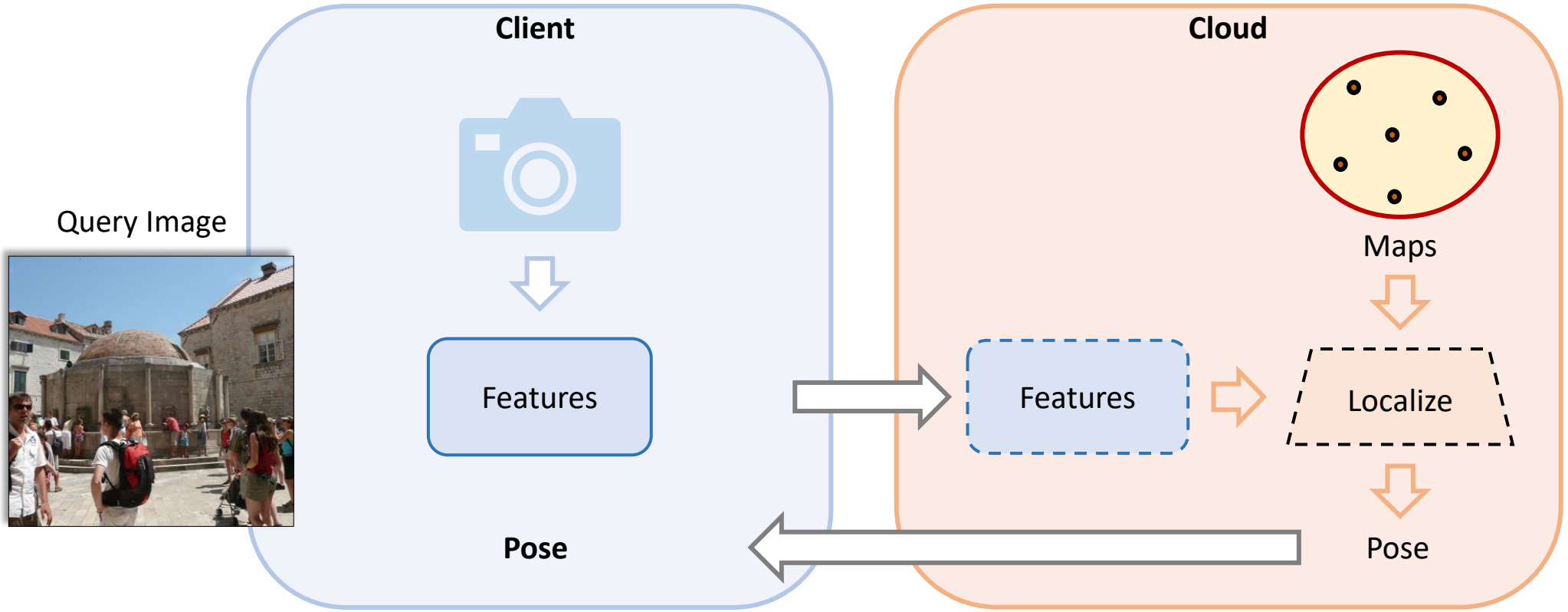
Microsoft ASA
(Azure Spatial Anchors)



Google AR Core
(Cloud Anchors)

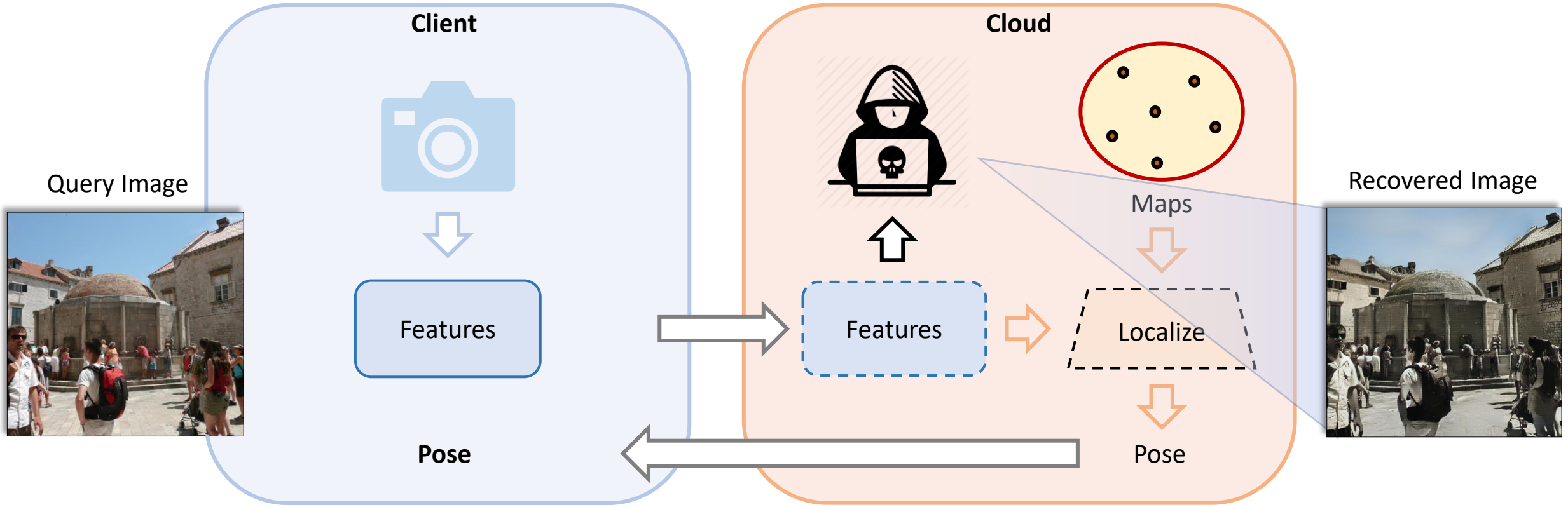


Privacy Risk in Cloud-based Localization



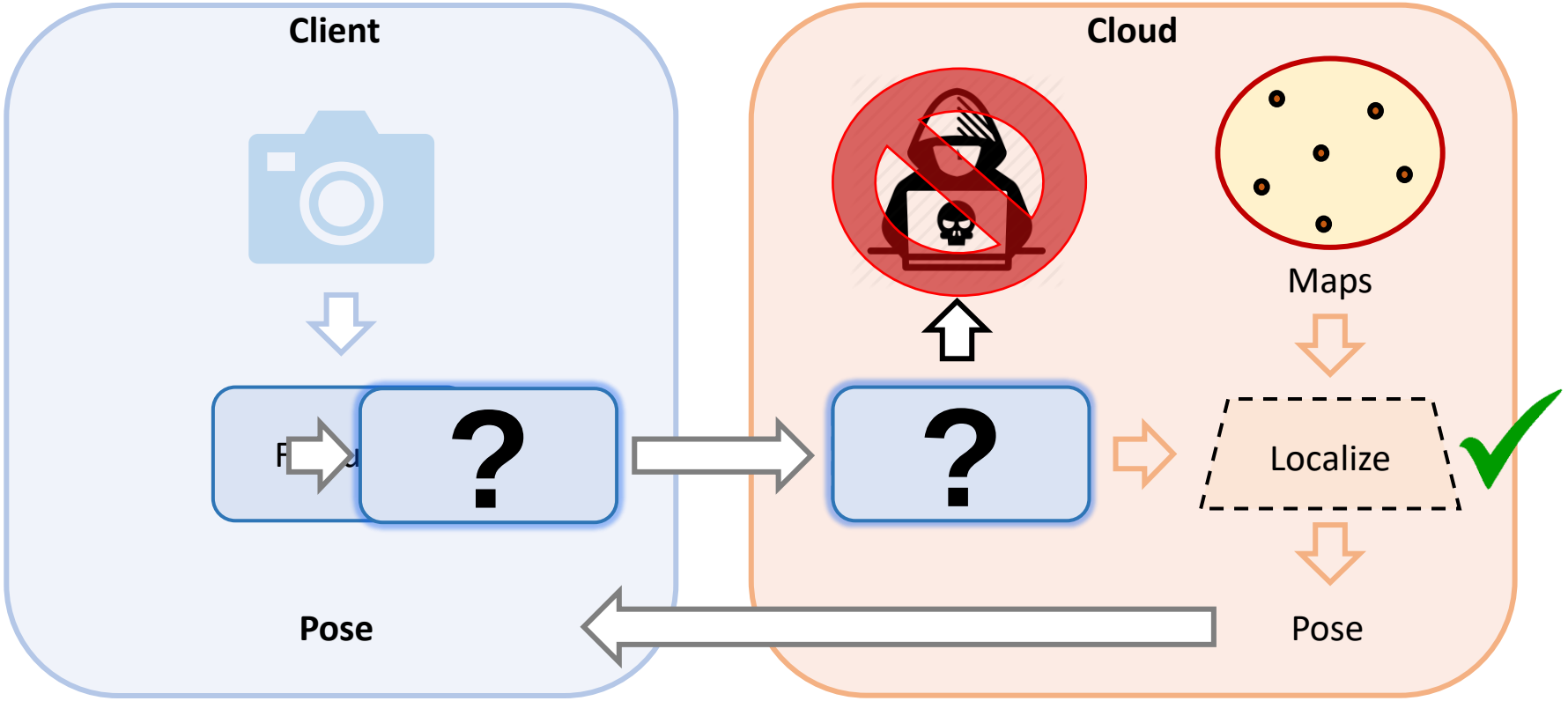
- Client sends image features to cloud
- Localization runs on cloud server
- Pose is sent back to Client

Privacy Risk in Cloud-based Localization



Adversary on cloud can invert features (recover the image)

Privacy Risk in Cloud-based Localization

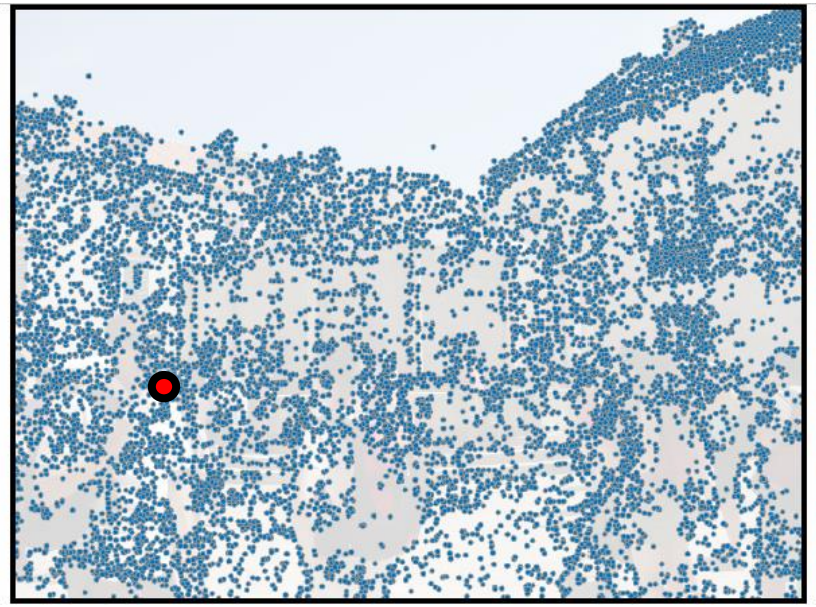


Our Goal:

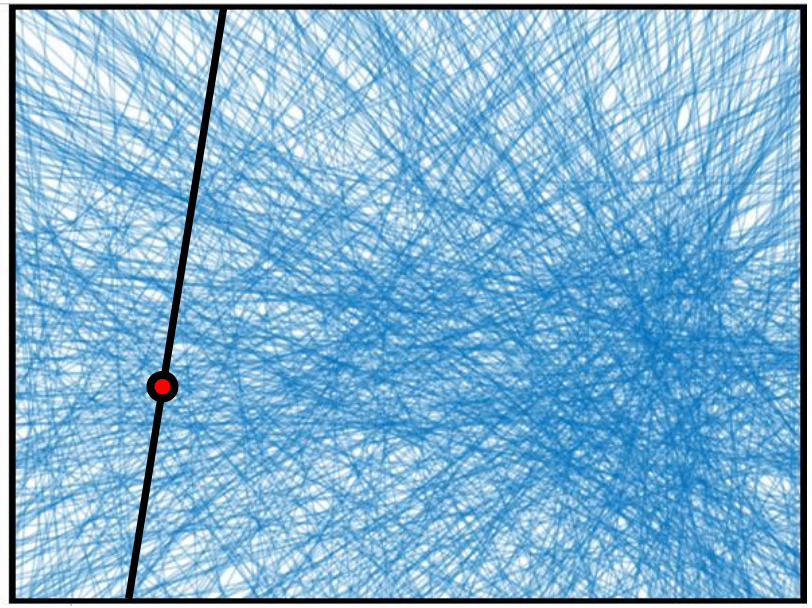
- Hide query features
- Prevent feature inversion on server
- Allow camera pose estimation

Proposed Idea

Query Image



2D Feature Points

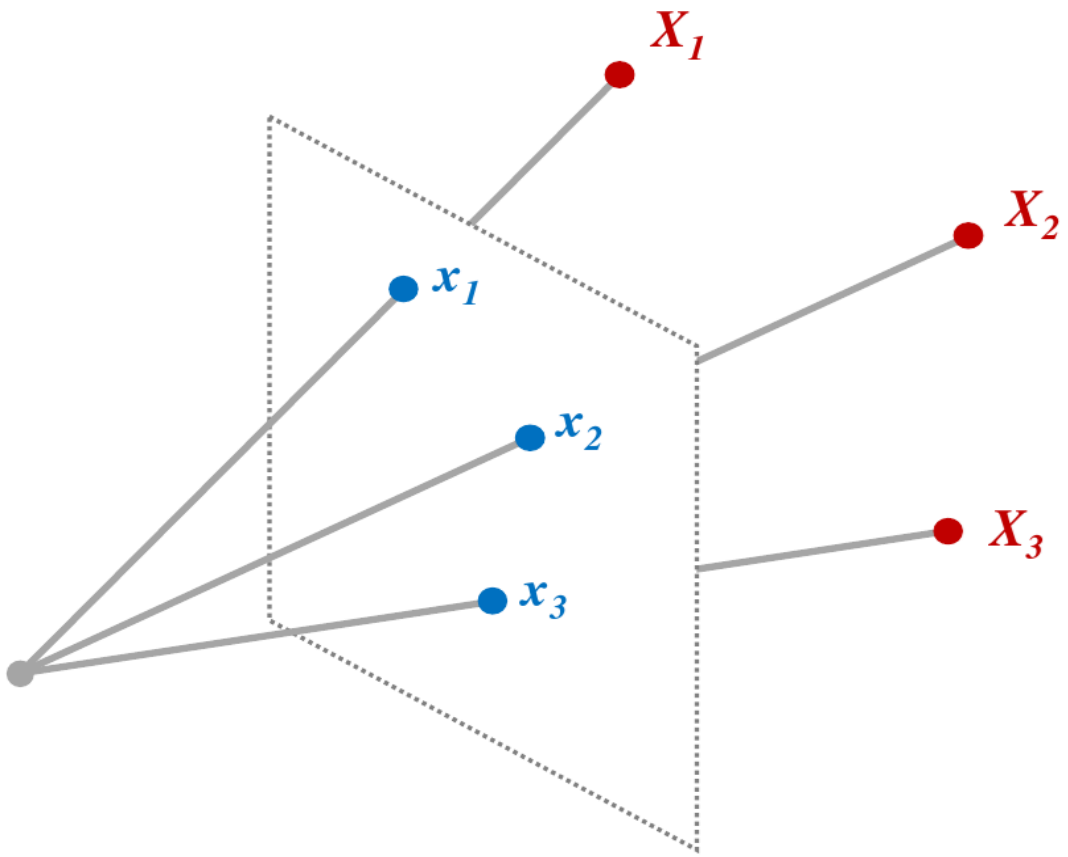


2D Feature Lines

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

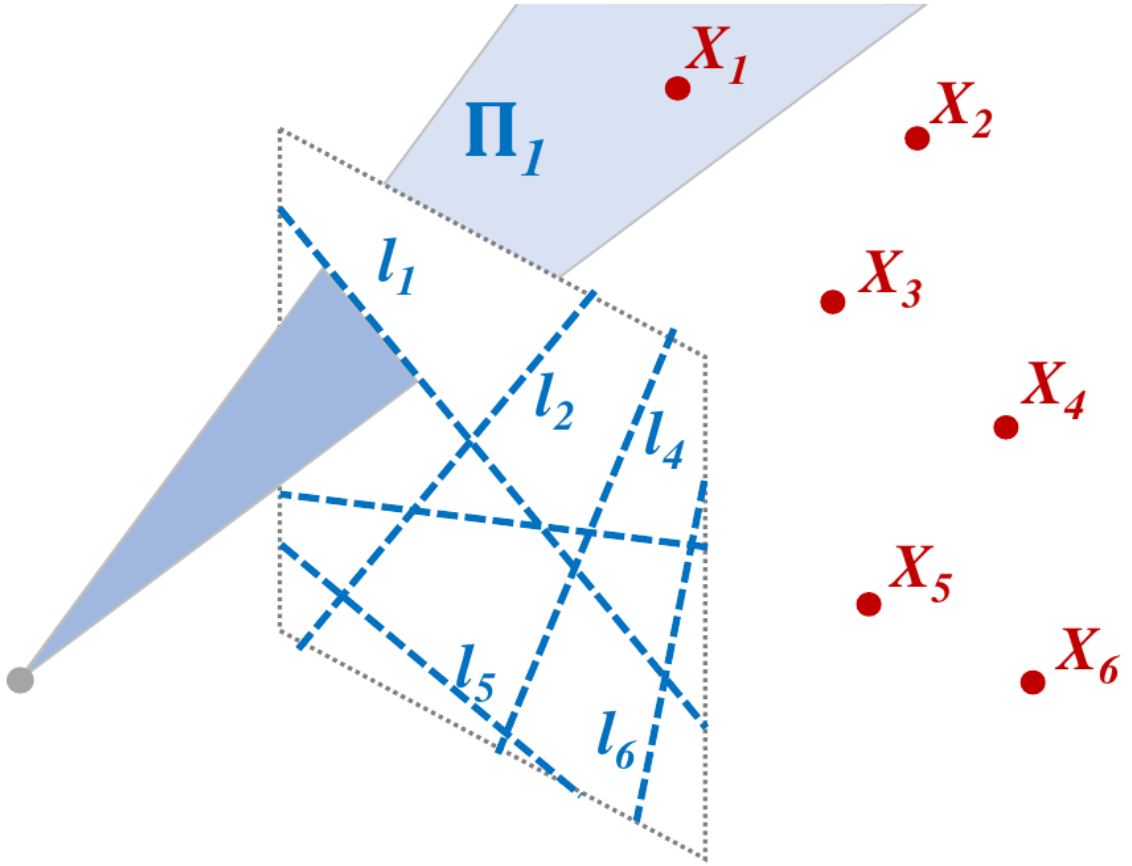
Camera Pose Estimation

Traditional Method ($p3P$)



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

Proposed Method ($l6P$)



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

Camera Pose Estimation

1. Our minimal problem can be cast as **Point-to-Plane problem** [1].

   [1] Ramalingam et al. 2013

2. Proposed **several variants** with:
   - known structure (multiple images),
   - known vertical direction,
   - known scale.

   We leverage existing minimal solvers [2-7].

   [2] Camposeco et al. 2018 [3] Lee et al. 2016 [4] Stewenius et al. 2005

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

3. Most variants are **computationally efficient** and can be used with RANSAC.

What gets revealed after localization?

Query Image

Reconstructed Image
(using all features)

Reconstructed Image
(using only *revealed* features)

Conclusions

- Highlighted new type of privacy issues in Mixed Reality, Robotics and Spatial AI platforms.

- Proposed privacy-preserving camera localization techniques
  - where the **map and query image** remains concealed.
  - built on top of **known minimal solvers**; many of which are accurate and **computationally efficient**.
- Many open problems and avenue for future work.