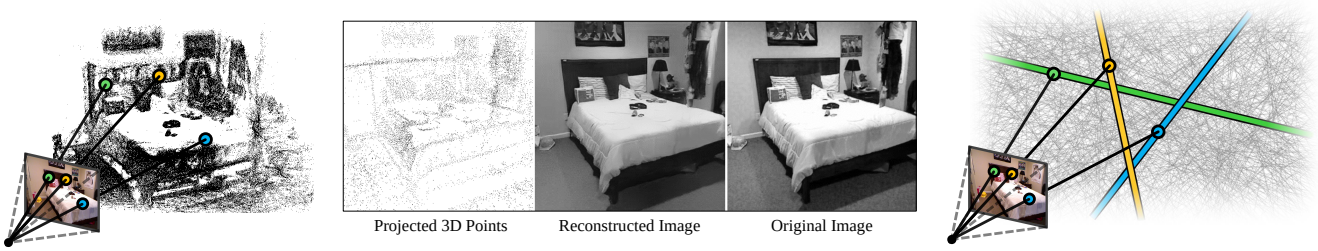
Privacy Preserving Image-Based Localization

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.

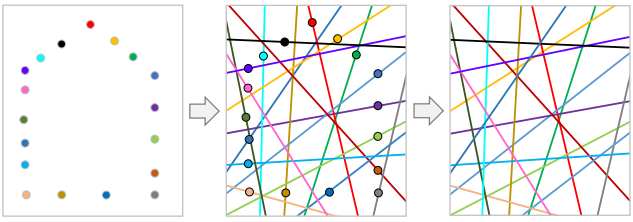
Abstract

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 mapping is a background process that the user might not be fully aware of. We pose the following question: How can we avoid disclosing confidential information about the captured 3D scene, and yet allow reliable camera pose estimation? This paper proposes the first solution to what we call privacy preserving image-based localization. The key idea of our approach is to lift the map representation from a 3D point cloud to a 3D line cloud. This novel representation obfuscates the underlying scene geometry while providing sufficient geometric constraints to enable robust and accurate 6-DOF camera pose estimation. Extensive experiments on several datasets and localization scenarios underline the high practical relevance of our proposed approach.

1. Introduction

Localizing a device within a scene by computing the camera pose from an image is a fundamental problem in computer vision, with high relevance in applications such as robotics [16, 19, 64], augmented/mixed reality (AR/MR) [36, 46], and structure from motion (SfM) [25, 60, 62]. Arguably, the most common approach to image-based localization is structure-based [19, 33, 46, 58] and tackles the problem by first matching the local 2D features of an image to the 3D point cloud model of the scene. Geometric 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 System (VPS) [79] as well as 6D.AI's mapping platform [1] indicate rising demand for image-based localization services that enable spatial persistence in AR/MR and robotics. Even today, devices like HoloLens, MagicLeap1, or iRobot Roomba continuously map their 3D environment to operate. This is a background process that users often are not consciously aware of. As robotics and AR/MR become increasingly relevant in consumer and enterprise applications, more and more 3D maps of our environment will be stored on device or in the cloud and then shared with other clients. Even though the source images are typically discarded after mapping, a person can easily infer the scene layout and presence of potentially confidential objects based on a casual visual inspection of the 3D point cloud (see Fig. 1-(a)). Furthermore, methods that reconstruct images from local features (such as [18, 50]) make it possible to recover remarkably accurate images of the scene from point clouds (see Fig. 1-(b)). From a technical standpoint, these privacy risks have been widely ignored so far. However, these will become increasingly relevant as localization services are adopted by more users as well as when mapping and localization capabilities will be more and more integrated with the cloud. As a consequence, there has recently been a lot of discussion in the AR/MR community around the privacy implications of this development [48, 54, 78].



(a) 3D Point Cloud (b) Lifting to 3D Lines (c) 3D Line Cloud

Figure 2: 3D Line Cloud. The main idea of the proposed 3D line cloud representation that hides the geometry of the map.

In general, we predict three scenarios in which the privacy of users will be compromised. First, if the scene itself is confidential (*e.g.*, a worker in a factory or a person at home), then storing maps with a cloud-based localization service is inherently risky. Running localization on trusted servers with maps stored securely could address privacy concerns, but even then the risk of unauthorized access remains. In the second scenario, the scene itself is not confidential but there is a secret object or information (*e.g.*, a hardware prototype in a workshop or some private details at home). Yet, we still want to enable persistent localization in the same environment later on, without the risk of the secret information leaking via the 3D map of the scene. This is especially relevant since users are typically not aware that mapping and localization services often continuously run in the background. The final scenario involves low-latency and offline applications that need localization to run on client devices, which requires 3D maps to be shared among authorized users. Obviously, distributing 3D maps with other users also compromises privacy.

To address these privacy concerns, we introduce a new research direction which we call *privacy preserving image-based localization* (see Fig. 1). The goal is to encode the 3D map in a confidential manner (thus preventing sensitive information from being extracted), while maintaining the ability to perform robust and accurate camera pose estimation. To the best of our knowledge, we are the first to propose a solution to this novel problem.

The key idea of our solution is to obfuscate the geometry of the scene in a novel map representation, where every 3D point is lifted to a 3D line with a random direction but passing through the original 3D point. Only the 3D lines and the associated feature descriptors of the 3D points are stored, whereas the original 3D point locations are discarded. We refer to such maps as *3D line clouds* (see Fig. 2). The 3D line cloud representation hides the underlying scene geometry and prevents the extraction of sensitive information.

To localize an image within a 3D line clouds, we leverage the traditional approach of feature matching [33, 58] to obtain correspondences between local 2D image features and 3D features in the map. Each correspondence provides the geometric constraint that the 2D image observation must lie on the image projection of its corresponding 3D line. Based on this constraint, the problem of absolute camera pose estimation from 3D line clouds entails the intersection of a set of camera rays and their corresponding 3D lines in the map. Towards leveraging this concept for privacy preserving localization, we show that a 3D line cloud can be interpreted as a generalized camera. As a consequence, absolute camera pose estimation from 3D line clouds boils down to solving a generalized relative or absolute pose problem, which means we can repurpose existing algorithms [30, 39, 42, 67–69] to solve our task.

In our paper, we study several variants of our approach. We first consider the case where the input is a single image and then generalize this concept to the case of jointly localizing multiple images. We also present several specializations of our method for localization scenarios, where the vertical direction or the scale of the scene is known. These specializations are especially valuable in practical applications and underline the high relevance of our approach.

Contributions. We make the following contributions: **(1)** We introduce the *privacy preserving image-based localization* problem and propose a first solution for it. **(2)** We propose a novel 3D map representation based on lifting 3D points to 3D lines, which preserves sufficient geometric constraints for pose estimation without revealing the 3D geometry of the mapped scene. **(3)** We propose minimal solvers for computing the camera pose given correspondences between 2D points in the image and 3D lines in the map. We study eight variants when the input is either a single image or multiple images, with and without the knowledge of the gravity direction or the scale of the scene.

2. Related Work

Image-Based Localization. Recent progress in image-based localization has led to methods that are now quite robust to changes in scene appearance and illumination [4,61], scale to large scenes [43,56,58,83], and are suitable for real-time computation and mobile devices [8, 33, 35, 43–45, 57, 76] with compressed map representations [15, 21]. Traditional localization methods based on image retrieval [34,66] and based on learning [12, 35, 80, 81] have the advantage of not requiring the explicit storage of 3D maps. However, model inversion techniques [47] pose privacy risks even for these methods. Besides, they are generally not accurate enough [59, 80] to enable persistent AR and robotics applications. Overall, to the best of our knowledge, there is no prior work on privacy preserving image-based localization or on privacy-aware methods in other 3D vision tasks.

Privacy-Aware Recognition. Privacy-aware object recognition and biometrics have been studied in vision since Avidan and Butman [9, 10], who devised a secure face detection system. Other applications include image retrieval [65], face recognition [22], video surveillance [74],

biometric verification [75], activity recognition in videos to anonymize faces [53, 55], and detecting computer screens in first-person video [40]. A recent line of work explores learning data-driven models from private or encrypted datasets [2, 27, 82]. All related works on privacy in computer vision focus on recognition problems, whereas ours is the first to focus on geometric vision. While our work aims at keeping the scene geometry confidential, it is worth exploring confidential features as well. However, this is beyond the scope of this paper.

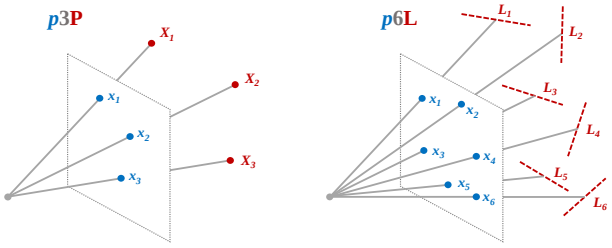
Privacy Preserving Databases. Privacy preserving techniques have been studied for querying data without leaking side information [17]. Differential privacy [20] and k-anonymity [71] have been applied to the problem of location privacy [3, 6, 26]. Learning data-driven models from private datasets has also received attention [2, 27, 82]. However, existing techniques are inapplicable for geometric vision problems such as image-based localization.

Generalized Camera Pose Estimation. An important insight we present in the paper is that privacy preserving camera pose estimation from 3D line clouds has a close relation to generalized cameras. After Grossberg and Nayar [28] formulated the theory of generalized cameras, Pless [51] derived generalized epipolar constraints from the the Plücker representation of 3D lines. Stewenius *et al*. [67] proposed the first minimal solver for the generalized relative pose problem, whereas numerous other solvers have been proposed for various generalized pose problems [13, 14, 23, 37, 38, 41, 42, 49, 68–70, 72, 77].

Generalized cameras are mostly used to model rigid multi-camera rigs or for dealing with multiple groups of calibrated cameras with known extrinsics [68–70]. In those settings, generalized cameras typically have a small number of pinhole cameras with several observations per image. In contrast, our 3D line clouds can be viewed as generalized cameras with one pinhole camera (and one observation) per 3D line. While existing generalized pose solvers can be prone to degeneracies, we avoid this problem by choosing lines with random directions. This not only enhances privacy but also leads to better conditioning of the problem.

3. Proposed Method

In this section, we describe our proposed solution to privacy preserving image-based localization. To give context, we first introduce the traditional approach to this problem for a single camera and then present the key concepts behind our privacy preserving method. We then describe the extension of these concepts for jointly localizing multiple cameras. Finally, we discuss practical solutions for several special cases, where the gravity direction is known or where we can obtain a local reconstruction of the scene with known or unknown scale. In our description, we focus on the high level intuitions behind our approach and refer the reader to related literature for details on the underlying algorithms needed to solve the various cases.



Three 2D point – 3D point correspondences Six 2D point – 3D line correspondences

Figure 3: Camera Pose Estimation. Left: using traditional 3D point cloud. Right: using our privacy preserving 3D line cloud.

3.1. Traditional Camera Pose Estimation

We follow the traditional approach of structure-based visual localization [33, 58], where the map of a scene is represented by a 3D point cloud, which is typically reconstructed from images using SfM [60]. To localize a pinhole camera with known intrinsics in the reconstructed scene, one estimates its absolute pose $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{R} & \boldsymbol{T}\end{bmatrix}$ with $\boldsymbol{R} \in \mathrm{SO}(3)$ and $\boldsymbol{T} \in \mathbb{R}^3$ from correspondences between normalized 2D observations $\boldsymbol{x} \in \mathbb{R}^2$ in the image and 3D points $\boldsymbol{X} \in \mathbb{R}^3$ in the map. To establish 2D–3D correspondences, the classical approach is to either use direct or indirect matching from 2D image features to 3D point features [33, 58]. Each 2D–3D point correspondence provides two geometric constraints for absolute camera pose estimation in the form of

$$\mathbf{0} = \bar{\boldsymbol{x}} - \boldsymbol{P}\bar{\boldsymbol{X}} = \lambda \begin{bmatrix}\boldsymbol{x} \\ 1\end{bmatrix} - \boldsymbol{P}\bar{\boldsymbol{X}} \ , \tag{1}$$

where λ is the depth of the image observation $\boldsymbol{x}$ while $\bar{\boldsymbol{x}} \in \mathbb{P}^2$ and $\bar{\boldsymbol{X}} \in \mathbb{P}^3$ are the lifted representations of $\boldsymbol{x}$ and $\boldsymbol{X}$ in projective space, respectively. Naturally, we need a minimum of three 2D–3D correspondences to estimate the six unknowns in $\boldsymbol{P}$. In the general case, this problem is typically referred to as the *p*nP problem and in the minimal case as the *p*3P problem. Since the matching process is imperfect and leads to outliers in the set of 2D–3D correspondences, the standard procedure is to use robust algorithms such as RANSAC [24, 52] in combination with efficient minimal solvers to optimize Eq. (1) for computing an initial pose estimate. Subsequently, that estimate is then refined by solving the non-linear least-squares problem

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \, \|\bar{\boldsymbol{x}} - \boldsymbol{P}\bar{\boldsymbol{X}}\|_2 \ , \tag{2}$$

which gives the maximum likelihood estimate based on a Gaussian error model $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{x}})$ for the image observations. This approach has been widely used [33, 45, 46, 58, 83] and enables efficient and accurate image-based localization in large scenes. However, it requires knowledge about

the scene geometry in the form of the 3D point cloud $\boldsymbol{X}$ and thereby this approach inherently reveals the geometry of the scene. In the next sections, we present our novel localization approach that overcomes this privacy limitation.

3.2. Privacy Preserving Camera Pose Estimation

The core idea behind our approach to enable privacy preserving localization is to obfuscate the geometry of the map in a way that conceals information about the underlying scene without losing the ability to localize the camera within the scene. In order to obfuscate the 3D point cloud geometry, we lift each 3D point $\boldsymbol{X}$ in the map to a 3D line $\boldsymbol{L}$ with a random direction $\boldsymbol{v} \in \mathbb{R}^3$ passing through $\boldsymbol{X}$. The lifted 3D line $\boldsymbol{L}$ in Plücker coordinates [51] is defined as

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} \in \mathbb{P}^5 \qquad \text{with} \qquad \boldsymbol{w} = \boldsymbol{X} \times \boldsymbol{v} \ . \tag{3}$$

Importantly, since direction $\boldsymbol{v}$ is chosen at random and due to the cross product being a rank-deficient operation, the original 3D point location $\boldsymbol{X}$ cannot be recovered from its lifted 3D line $\boldsymbol{L}$. We only know that $\boldsymbol{L}$ passes through $\boldsymbol{X}$ somewhere and that this also holds for their respective 2D projections $\boldsymbol{l}$ and $\boldsymbol{x}$ in the image. Formally, a 2D image observation $\boldsymbol{x}$ passes through the projected 2D line $\boldsymbol{l}$, if it satisfies the geometric constraint

$$0 = \bar{\boldsymbol{x}}^T \boldsymbol{l} \quad \text{with} \quad [\boldsymbol{l}]_\times = \begin{bmatrix} 0 & -l_3 & l_2 \\ l_3 & 0 & -l_1 \\ -l_2 & l_1 & 0 \end{bmatrix} = \boldsymbol{P}[\boldsymbol{L}]_\times \boldsymbol{P}^T, \tag{4}$$

where $[\boldsymbol{L}]_\times$ is the Plücker matrix defined as

$$[\boldsymbol{L}]_\times = \begin{bmatrix} -[\boldsymbol{w}]_\times & -\boldsymbol{v} \\ \boldsymbol{v}^T & 0 \end{bmatrix} \ . \tag{5}$$

Using this constraint for absolute camera pose estimation requires a minimum of six 2D point to 3D line correspondences to solve for the six unknowns in $\boldsymbol{P}$. This is in contrast to the traditional approach, where each correspondence provides two constraints and thus only three correspondences are needed. Similar to the traditional pnP and p3P problems, we denote the general problem as pnL and the minimal problem as p6L. Geometrically, solving the pnL problem is equivalent to rotating and translating the bundle of rays defined by $\boldsymbol{x}$ and passing through the pinhole of the camera, such that the bundle of camera rays intersect with their corresponding 3D lines in the map (see Fig 3). Note, this is a specialization of the generalized relative pose problem [67], where the rays in the first generalized camera represent the known 3D lines of the map, and the rays of the second generalized camera represent the 2D image observations of the pinhole camera that we want to localize.

We embed this concept into the traditional localization pipeline by robustly estimating an initial pose estimate using RANSAC with the minimal solver by Stewenius *et al*. [67] to solve Eq. (4). We then non-linearly refine the initial pose by minimizing the geometric distance between the observed 2D point and the projected 3D line as

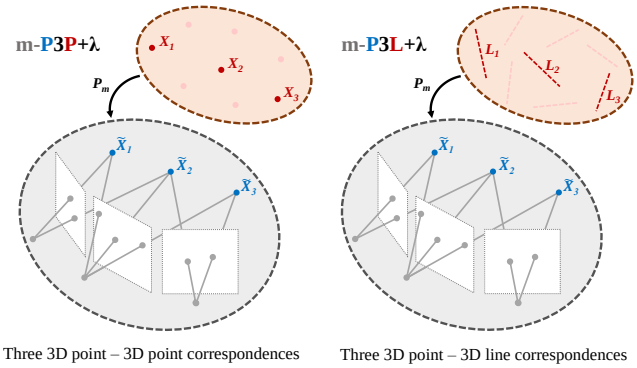


Figure 4: Camera Pose Estimation with Known Structure. Left: traditional setup with 3D point cloud maps. Right: our proposed approach using 3D line cloud maps.

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \frac{\bar{\boldsymbol{x}}^T \boldsymbol{l}}{\sqrt{l_1^2 + l_2^2}} \ . \tag{6}$$

After deriving the theory for a single camera in this section, we next generalize our approach to the joint localization of multiple images and the special case with known vertical.

3.2.1 Generalization to Multiple Cameras

While existing localization approaches typically only consider a single image [33, 46, 58, 83], many devices such as head mounted displays, robots, or vehicles are equipped with multiple rigidly mounted cameras, that have been calibrated a priori. Jointly localizing multiple cameras together brings great benefits for localization by leveraging the combined field of view to retrieve more 2D–3D correspondences and by reducing the number of unknown pose parameters for an increased redundancy in the estimation problem. In addition, many mobile devices nowadays have built-in SLAM capabilities (*e.g.*, ARKit, ARCore), which can be leveraged to take advantage of the same simplifications as with multi-camera systems by treating a local camera trajectory as an extrinsic calibration of multiple images.

The joint localization of multiple cameras differs from the case of a single camera primarily in how the problem is parameterized. Instead of determining a separate pose $\boldsymbol{P} \in \mathrm{SE}(3)$ for each camera, we reparameterize the pose as

$$\boldsymbol{P} = \boldsymbol{P}_c \boldsymbol{P}_m \quad \text{with} \quad \boldsymbol{P}_m = s_m \begin{bmatrix} \boldsymbol{R}_m & \boldsymbol{T}_m \\ \boldsymbol{0} & s_m^{-1} \end{bmatrix}. \tag{7}$$

We now estimate only a single 3D similarity transformation $\boldsymbol{P}_m \in \mathrm{Sim}(3)$, while the known relative extrinsic calibrations $\boldsymbol{P}_c$ of the individual cameras stay fixed. Note that if we know the relative scale of $\boldsymbol{P}_c$ with respect to the 3D points $\boldsymbol{X}$ in the map, we can eliminate the scale factor $s_m \in \mathbb{R}^+$ and reduce $\boldsymbol{P}_m$ to a 3D rigid transformation.

Constraints	*Query Type*	POINT TO POINT (Traditional)		POINT TO LINE (Privacy Preserving)	
2D – 3D	*Single-Image*	*p*3P [29]	*p*2P+u [68]	*p*6L [67]	*p*4L+u [69]
	Multi-Image	m-*p*3P [30]	m-*p*2P+u [32]	m-*p*6L [67]	m-*p*4L+u [69]
3D – 3D	*Multi-Image*	m-P3P+λ [73]	m-P2P+λ+u [73]	m-P4L+λ [70]	m-P3L+λ+u [13]
		m-P3P+λ+s [32]	m-P2P+λ+u+s [32]	m-P3L+λ+s [30]	m-P2L+λ+u+s [68]

Table 1: Camera Pose Problems. Traditional methods are ***p****P*** (2D point to 3D point) and **P**P** (3D point to 3D point), whereas privacy preserving methods are ***p****L*** (2D point to 3D line) and **P**L** (3D point to 3D line). The methods in the first row localize single images, whereas the rest jointly localizes multiple images (prefix *m*). We have general solvers as well specialized ones for known vertical direction (suffix +*u*). The bottom two rows exploit known 3D structure (suffix $+\lambda$ and suffix +*s* for known scale) local to the camera to be localized.

In the literature, this problem is referred as the generalized absolute pose problem [30, 49], which is analogous to the traditional problem and does not conceal the 3D point cloud. In most practical applications, we can assume that the scale $s_m = 1$, because multi-camera setups are typically calibrated to metric scale, and due to the fact that most SLAM systems recover scale from integrated inertial measurements. In the following, we therefore initially restrict our work to the scenario where $\boldsymbol{P}_m \in \mathrm{SE}(3)$. We refer to the solution of this problem as *m-pnP* in the general case and *m-p3P* in the minimal case. However, efficient solutions also exist for the more general case $\boldsymbol{P}_m \in \mathrm{Sim}(3)$ [70, 77].

In the privacy preserving setting, the generalization to multiple images again boils down to solving a generalized relative pose problem [67]. However, the rays of the second generalized camera arise from 2D image observations of multiple instead of a single pinhole camera. We refer to the generalized solutions in the privacy preserving setting as *m-p*nL in the general and *m-p*6L in the minimal case.

3.2.2 Pose Estimation with Known Structure

So far, we have discussed a way to estimate the camera pose from the rays of 2D image observations directly. In many situations though, it is possible to obtain the depth λ of an image observation $\boldsymbol{x}$, after which, its 3D location relative to the camera is computed as $\tilde{\boldsymbol{X}} = \lambda\bar{\boldsymbol{x}}$. Such 3D data can be extracted through an active depth camera that yields RGB-D images or through multi-view triangulation. In the traditional localization problem, we can therefore directly estimate the camera pose as the transformation that best aligns the two corresponding 3D point sets using the constraint

$$\mathbf{0} = \tilde{\boldsymbol{X}} - \boldsymbol{P}\bar{\boldsymbol{X}} \ . \tag{8}$$

To solve this equation in the minimal case, we need only three correspondences for the 6-DOF of the 3D rigid transformation $\boldsymbol{P}$. Eq. (8) is typically solved in a least-squares fashion, and in this form has a direct and computationally efficient solution [32, 73]; we refer to this as m-PnP+λ and m-P3P+λ in the general and minimal cases, respectively.

Similarly, we can also take advantage of the local 3D points $\tilde{\boldsymbol{X}}$ in our privacy preserving approach. Instead of solving a generalized relative pose problem to find the intersection between the 3D lines of the map and the camera rays, we now try to find a pose such that the 3D lines $\boldsymbol{L}$ of the map pass through the 3D points $\tilde{\boldsymbol{X}}$. We can formalize this in the following geometric constraint

$$\mathbf{0} = \tilde{\boldsymbol{X}} - \boldsymbol{P}\begin{bmatrix}\boldsymbol{v}\times\boldsymbol{w} + \alpha\boldsymbol{v} \\ 1\end{bmatrix} \ , \tag{9}$$

where α is the unknown distance from the random origin $\boldsymbol{v}\times\boldsymbol{w}$ of the 3D line $\boldsymbol{L}$ to the secret 3D point $\boldsymbol{X}$. By inverting the role of the camera and the map, this problem is geometrically equivalent to the generalized absolute pose problem, *i.e.*, we can repurpose m-*p*nP to solve for the unknown pose $\boldsymbol{P}$. As such, we now only need a minimum of three 3D point to 3D line correspondences compared to the six correspondences needed to solve m-*p*6L (see Fig.4). Note that requiring fewer points to solve the minimal problem is advantageous in RANSAC, which has an exponential runtime complexity in the number of sampled points. The solution to Eq. (9) is also more efficient to compute [30] as compared to m-*p*6L. We refer to this problem as m-*p*nL+λ in the general and m-P3L+λ in the minimal case.

3.2.3 Extension to Unknown Scale

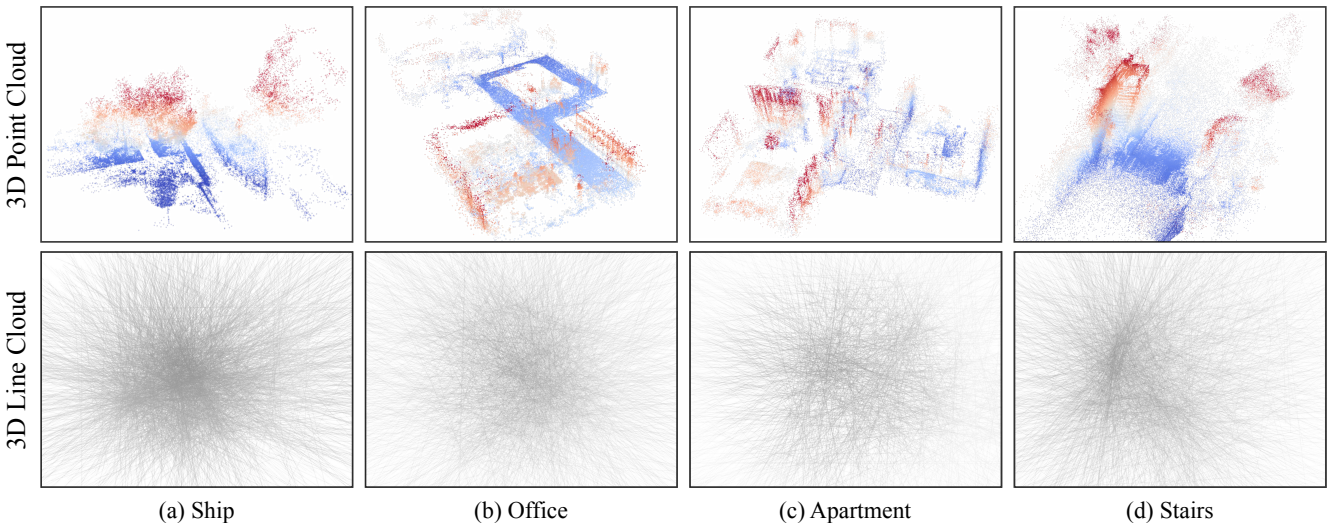
The approach described in the previous section can be sensitive to inaccurate 3D point locations $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$. This is problematic, even if the two 3D point clouds have only slightly different scale, *e.g.*, due to drift in SLAM or slight miscalibrations of the multi-camera system. In comparison, the constraints used by *p*nP and *p*nL are less susceptible to this issue. This is because the viewpoints used to triangulate $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ are inherently similar in image-based localization and the uncertainties σ_λ in the depths λ are usually larger than the uncertainties $\boldsymbol{\sigma}_{\boldsymbol{x}}$ in image space.

To overcome this issue, it is typically better to estimate a 3D similarity transformation $s\boldsymbol{P}$ with $s \in \mathbb{R}^+$ instead of a 3D rigid transformation, when performing structure-based alignment. The constraint in Eq. (8) then becomes

$$\mathbf{0} = \tilde{\boldsymbol{X}} - s\boldsymbol{P}\bar{\boldsymbol{X}} \ , \tag{10}$$

while the privacy preserving constraint in Eq. (9) becomes

$$\mathbf{0} = \tilde{\boldsymbol{X}} - s\boldsymbol{P}\begin{bmatrix}\boldsymbol{v}\times\boldsymbol{w} + \alpha\boldsymbol{v} \\ 1\end{bmatrix} \ . \tag{11}$$



(a) Ship (b) Office (c) Apartment (d) Stairs

Figure 5: Dataset Visualization. The original 3D point cloud with the corresponding 3D line cloud is shown.

Now we need at least four correspondences to estimate the 7-DOF 3D similarity. Note that Eq. (10) has a comparatively simple and efficient solution [73] that we refer to as m-PnP+λ+s. In the privacy preserving setting, the problem of computing a 3D rigid transformation is exactly minimal, *i.e.*, we now need a fourth correspondence to estimate the additional scale parameter using the constraints in Eq. (11). This is equivalent to the generalized absolute pose and scale problem [70], where the role of cameras and map is again reversed. We refer to the reversed problem as m-PnL+λ+s in the general and as m-P4L+λ+s in the minimal case.

3.2.4 Specialization with Known Vertical

Oftentimes, an estimate of the gravity direction in both the reference frame of the camera and the 3D map may be available, *e.g.*, from inertial measurements or vanishing point detection. By pre-aligning the two reference frames to the vertical direction, we can reduce the number of rotational pose parameters from three to one such that $\boldsymbol{R} \in \mathrm{SO}(2)$. This parameterization of the rotation simplifies the geometric constraints, and leads to more efficient and numerically stable solutions for these problems. In addition, the minimal cases require fewer points, leading to a better runtime of RANSAC. We implement the known gravity setting for all described problems and indicate this with the suffix $+u$. An overview of all the problems is given in Table 1.

4. Experimental Evaluation

To demonstrate the high practical relevance of our approach, we conduct an extensive list of experiments on real-world data with ground-truth. We evaluate the pose estimation performance in terms of accuracy/recall and robustness to the input by comparing our privacy preserving approach using 3D line clouds to the traditional approach of using 3D point clouds. In the following, we first describe the experimental setup before presenting the results.

4.1. Setup

Datasets. We collect 15 real-world datasets of complex indoor and outdoor scenes (see Fig. 5) using a mix of mobile phones and the research mode of the Microsoft HoloLens [31]. To realistically simulate an image-based localization scenario, we captured *map images* used to reconstruct a 3D point cloud of the scene and *query images* from significantly different viewpoints used for evaluating localization. For sparse scene reconstruction and camera calibration, we feed all the recorded (*map* and *query*) images into the COLMAP SfM pipeline [60, 63] to obtain high-quality camera calibrations. The obtained camera poses of the query images serve as ground-truth $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{T}}$ for our evaluations. All query images alongside their corresponding 3D points are then carefully removed from the obtained reconstructions to prepare the 3D map for localization. Afterwards, we perform another bundle adjustment with fixed camera poses to optimize the remaining 3D points given only the map images. These steps are to reconstruct accurate ground-truth poses for the query images, and to also ensure a realistic 3D map for localization, in which we are only given the map images. Across the datasets, we captured 375 single-image and 402 multi-image queries.

Protocol. To establish 2D–3D correspondences, we use indirect matching of SIFT features at the default settings of the SfM pipeline [60, 63]. In the single-image scenario, we treat each query image separately, while for the multi-image scenario, we group several consecutive images in the camera stream as one generalized camera. When evaluating the multi-image case and pose estimation with known structure, we reconstruct the 3D points $\tilde{\boldsymbol{X}}$ and camera poses $\boldsymbol{P}_c$ using SfM [60,63] from only the query images. For a fair compar-

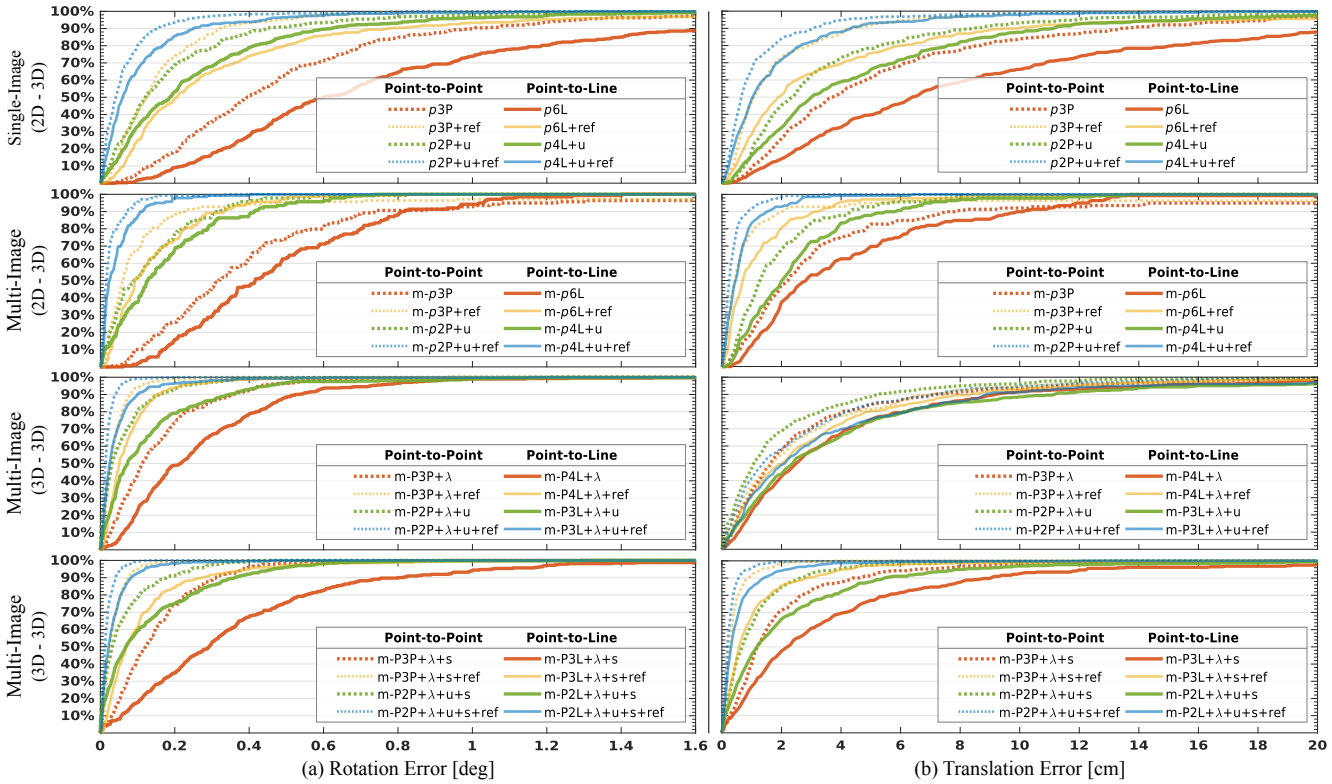


Figure 6: Camera Pose Estimation Errors Plots. Cumulative rotation and translation error histograms for all 16 evaluated methods.

ison, all methods use exactly the same 2D–3D correspondences, thresholds, and RANSAC implementation [24]. See supplementary material for more details.

Metrics. In our evaluation, we compute the rotational error as $\Delta R = \arccos \frac{\mathrm{Tr}(\boldsymbol{R}^T \hat{\boldsymbol{R}})-1}{2}$ and the translational error as $\Delta T = \|\boldsymbol{R}^T \boldsymbol{T} - \hat{\boldsymbol{R}}^T \hat{\boldsymbol{T}}\|_2$. We also report the average point-to-point and point-to-line (*c.f.* Eq. (2) and Eq. (6)) reprojection errors w.r.t. to the obtained pose estimate.

Methods. We compare the results of our proposed 8 privacy preserving to the corresponding 8 variants of traditional pose estimators, see Table 1. The initial pose estimates of all methods are computed using standard RANSAC and a minimal solver for the geometric constraints. We also compare the results of a non-linear refinement (suffix *+ref*) of the initial pose using a Levenberg-Marquardt optimization of Eqs. (2) and (6) based on the inliers from RANSAC.

4.2. Results

Accuracy and Recall. The accuracy/recall curves are presented in Fig. 6, and reprojection errors in Table 2. As expected, the traditional approaches achieve better accuracy/recall, because their solutions leverage two constraints for pose estimation. Surprisingly, even though ours uses only a single geometric constraint, it comes very close to the results achieved by the traditional approach. Moreover, incorporating known information about structure, gravity, and scale leads to an additional improvement of the results for all methods.

Runtime. Table 2 reports the mean number of required RANSAC iterations, inlier ratio, generated solutions in the minimal solver, and time required to estimate a single minimal solution. The results show that, while our method is slower than the conventional approach, it provides runtimes that are suitable for practical real-time applications. Especially, the specialized solvers with known structure and gravity achieve competitive runtime. We used the same RANSAC threshold for all the methods, but in practice this threshold could be chosen smaller for the privacy preserving methods, since the point-to-line is always smaller than point-to-point reprojection errors. This, together with the possibility to more easily include some additional outliers along the line due to mistakes in feature matching, leads to slightly higher inlier ratios for our method, see Table 2.

Robustness. We study robustness with respect to point cloud density and image noise. In Fig. 7, we demonstrate reliable pose estimation even when only retaining every 20^{th} point of an already sparse SfM point cloud. Fig. 8 shows similar behavior for ours and the conventional approach under varying noise $\boldsymbol{\sigma_x}$ on the image observations.

5. Discussion

Let us now discuss to what extent the privacy risks have been addressed, and highlight directions for future work.

POINT TO POINT (Traditional)				POINT TO LINE (Privacy Preserving)			
***p*3P**	**i:** 31 **r:** 64 **s:** 2.05 **t:** 3.54	***p*2P+u**	**i:** 15 **r:** 64 **s:** 2 **t:** 3.21	***p*6L**	**i:** 158 **r:** 69 **s:** 64 **t:** 1002	***p*4L+u**	**i:** 40 **r:** 70 **s:** 4 **t:** 5.27
m-*p*3P	**i:** 38 **r:** 64 **s:** 1.70 **t:** 3.81	**m-*p*2P+u**	**i:** 11 **r:** 65 **s:** 2 **t:** 3.16	**m-*p*6L**	**i:** 64 **r:** 72 **s:** 64 **t:** 691	**m-*p*4L+u**	**i:** 23 **r:** 72 **s:** 4 **t:** 3.61
m-P3P+λ	**i:** 23 **r:** 62 **s:** 1 **t:** 1.82	**m-P2P+λ+u**	**i:** 12 **r:** 62 **s:** 1 **t:** 0.97	**m-P4L+λ**	**i:** 27 **r:** 68 **s:** 1.5 **t:** 26.3	**m-P3L+λ+u**	**i:** 24 **r:** 68 **s:** 1 **t:** 4.07
m-P3P+λ+s	**i:** 24 **r:** 62 **s:** 1 **t:** 4.19	**m-P2P+λ+u+s**	**i:** 12 **r:** 62 **s:** 1 **t:** 2.37	**m-P3L+λ+s**	**i:** 18 **r:** 68 **s:** 2.1 **t:** 2.3	**m-P2L+λ+u+s**	**i:** 9 **r:** 68 **s:** 2 **t:** 2.10

Notation: **i:** mean number of iterations, **r:** inlier ratio [%], **s:** mean number of solutions, **t:** minimal solver time [ms].

***p*3P**	1.84 / 1.42	***p*2P+u**	1.88 / 1.43	***p*6L**	1.55 (4.20) / 1.10 (3.24)	***p*4L+u**	1.45 (3.62) / 1.08 (3.11)
m-*p*3P	2.06 / 1.51	**m-*p*2P+u**	1.88 / 1.51	**m-*p*6L**	1.56 (4.23) / 1.13 (3.29)	**m-*p*4L+u**	1.46 (3.90) / 1.12 (3.17)
m-P3P+λ	1.71 / 1.42	**m-P2P+λ+u**	1.62 / 1.42	**m-P4L+λ**	1.39 (4.20) / 1.08 (3.52)	**m-P3L+λ+u**	1.62 (4.92) / 1.17 (3.73)
m-P3P+λ+s	1.72 / 1.43	**m-P2P+λ+u+s**	1.63 / 1.41	**m-P3L+λ+s**	1.47 (4.29) / 1.13 (3.48)	**m-P2L+λ+u+s**	1.31 (4.06) / 1.07 (3.52)

Table 2: Quantitative Results. RANSAC statistics (top) and reprojection errors in pixels (bottom) for *initial / refined* results in traditional (Point to Point, Eq. (2)) and privacy preserving setting (Point to Line, Eq. (6); in brackets also Point to Point, if we had the secret points).

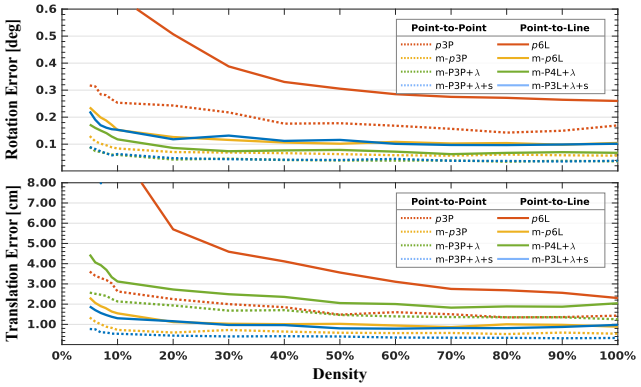


Figure 7: Point Cloud Density. Rotation and translation error with varying point cloud density by uniformly dropping a percentage of points/lines at random from the map.

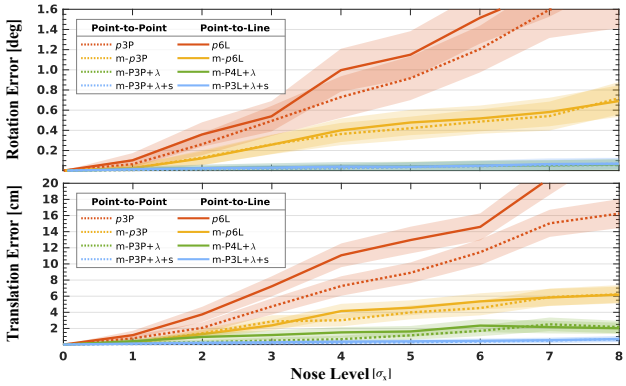


Figure 8: Measurement Noise Sensitivity. Rotation and translation errors for varying Gaussian noise level on the measurements.

What is revealed during localization? When images are successfully localized within a scene, the inliers of the pose estimate reveal the secret 3D points through intersection of the camera rays with the corresponding 3D lines. On first sight, this might seem like a privacy issue, but in practice only objects visible in the image are revealed, while the rest of the map or any confidential objects remain secret.

Permanent Line Cloud Transformation. The lifting transformation must be performed only once and becomes permanent for a scene; otherwise, an adversary that retains multiple copies of line clouds generated by different lifting transformations can easily recover the secret 3D points by intersecting corresponding 3D lines.

Compactness of Representation. A more compact representation than Plücker lines in Eq. (3) would be to chose a finite set of line directions; *e.g.*, 256 to fit within a byte, and encode the position of the line as the intersection with the plane through the origin and orthogonal to the direction, this reduces memory usage to 2 floats and 1 byte, *i.e.*, even less than the 3 floats to encode a 3D point.

Privacy Attack on Line Clouds. Recovering the location of a single 3D point from its lifted 3D line representation is an ill-posed inversion problem, see Eq. (3). However, by analyzing the density of the 3D line cloud, one could potentially recover information about the scene structure. While 3D line clouds appear effective at making the underlying scene geometry incomprehensible, it really depends on the sampling density of the 3D points in the scene (see suppl. material). In practice, we believe that our method is generally quite robust to such attacks, since image-based localization typically uses sparse SfM point clouds. Besides, a sparsification in Fig 7 of the 3D line cloud is an effective defense mechanism. Nevertheless, a more thorough theoretical analysis is an interesting avenue of future research.

6. Conclusion

This paper introduced a new research direction called *privacy preserving image-based localization*. With this work, we are the first to address potential privacy concerns associated with the persistent storage of 3D point cloud models, as required by a wide range of applications in AR and robotics. Our proposed idea of using confidential 3D line cloud maps conceals the geometry of the scene, while maintaining the ability to perform robust image-based localization based on the standard feature matching paradigm. There are numerous directions for future work and we encourage the community to investigate this problem.

References

[1] 6D.AI. http://6d.ai/, 2018. 1

[2] M. Abadi, A. Chu, I. Goodfellow, B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Conference on Computer and Communications Security (ACM CCS)*, 2016. 3

[3] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Conference on Computer & communications security (SIGSAC)*, 2013. 3

[4] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[5] ARCore. https://developers.google.com/ar/, 2018. 1

[6] C. A. Ardagna, M. Cremonini, E. Damiani, S. D. C. Di Vimercati, and P. Samarati. Location privacy protection through obfuscation-based techniques. In *IFIP Annual Conference on Data and Applications Security and Privacy*, 2007. 3

[7] ARKit. https://developer.apple.com/arkit/, 2018. 1

[8] C. Arth, D. Wagner, M. Klopschitz, A. Irschara, and D. Schmalstieg. Wide area localization on mobile phones. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, 2009. 2

[9] S. Avidan and M. Butman. Blind vision. In *European Conference on Computer Vision (ECCV)*, 2006. 2

[10] S. Avidan and M. Butman. Efficient methods for privacy preserving face detection. In *Advances in neural information processing systems*, 2007. 2

[11] Azure Spatial Anchors. https://azure.microsoft.com/en-us/services/spatial-anchors/, 2019. 1

[12] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother. DSAC-differentiable RANSAC for camera localization. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[13] F. Camposeco, A. Cohen, M. Pollefeys, and T. Sattler. Hybrid camera pose estimation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3, 5

[14] F. Camposeco, T. Sattler, and M. Pollefeys. Minimal solvers for generalized pose and scale estimation from two rays and one point. *European Conference on Computer Vision (ECCV)*, 2016. 3

[15] S. Cao and N. Snavely. Minimal scene descriptions from structure from motion models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 2

[16] M. Cummins and P. Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *International Journal of Robotics Research (IJRR)*, 2008. 1

[17] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *Symposium on Principles of Database Systems*, 2003. 3

[18] A. Dosovitskiy and T. Brox. Inverting visual representations with convolutional networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[19] R. Dubé, D. Dugas, E. Stumm, J. I. Nieto, R. Siegwart, and C. Cadena. Segmatch: Segment based loop-closure for 3D point clouds. In *International Conference on Robotics and Automation (ICRA)*, 2017. 1

[20] C. Dwork. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*, 2008. 3

[21] M. Dymczyk, S. Lynen, M. Bosse, and R. Siegwart. Keep it brief: Scalable creation of compressed localization maps. In *International Conference on Intelligent Robots and Systems (IROS)*, 2015. 2

[22] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft. Privacy-preserving face recognition. In *International Symposium on Privacy Enhancing Technologies Symposium*, 2009. 2

[23] A. Ess, A. Neubeck, and L. J. V. Gool. Generalised linear pose estimation. In *BMVC*, 2007. 3

[24] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981. 3, 7

[25] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, et al. Building rome on a cloudless day. In *European Conference on Computer Vision (ECCV)*, 2010. 1

[26] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 2008. 3

[27] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International Conference on Machine Learning*, 2016. 3

[28] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. In *International Conference on Computer Vision (ICCV)*, 2001. 3

[29] B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal of Computer Vision (IJCV)*, 1994. 5

[30] G. Hee Lee, B. Li, M. Pollefeys, and F. Fraundorfer. Minimal solutions for pose estimation of a multi-camera system. *Springer Tracts in Advanced Robotics*, 2016. 2, 5

[31] Hololens. https://www.microsoft.com/en-us/hololens, 2016. 1, 6

[32] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 1987. 5

[33] A. Irschara, C. Zach, J.-M. Frahm, and H. Bischof. From structure-from-motion point clouds to fast location recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. 1, 2, 3, 4

[34] H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, and C. Schmid. Aggregating local image descriptors into compact codes. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2012. 2

[35] A. Kendall, M. Grimes, and R. Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *International Conference on Computer Vision (ICCV)*, 2015. 2

[36] G. Klein and D. Murray. Parallel Tracking and Mapping on a Camera Phone. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, 2009. 1

[37] L. Kneip and P. Furgale. OpenGV: A unified and generalized approach to real-time calibrated geometric vision. *International Conference on Robotics and Automation (ICRA)*, 2014. 3

[38] L. Kneip and H. Li. Efficient computation of relative pose for multi-camera systems. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[39] L. Kneip and S. Lynen. Direct optimization of frame-to-frame rotation. *International Conference on Computer Vision (ICCV)*, 2013. 2

[40] M. Korayem, R. Templeman, D. Chen, D. Crandall, and A. Kapadia. Enhancing lifelogging privacy by detecting screens. In *Conference on Human Factors in Computing Systems (CHI)*, 2016. 3

[41] G. H. Lee, M. Pollefeys, and F. Fraundorfer. Relative pose estimation for a multi-camera system with known vertical direction. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[42] H. Li, R. Hartley, and J.-H. Kim. A linear approach to motion estimation using generalized camera models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 2, 3

[43] Y. Li, N. Snavely, D. Huttenlocher, and P. Fua. Worldwide pose estimation using 3d point clouds. In *European Conference on Computer Vision (ECCV)*, 2012. 2

[44] Y. Li, N. Snavely, and D. P. Huttenlocher. Location recognition using prioritized feature matching. In *European Conference on Computer Vision (ECCV)*, 2010. 2

[45] H. Lim, S. N. Sinha, M. F. Cohen, M. Uyttendaele, and H. J. Kim. Real-time monocular image-based 6-doF localization. *International Journal of Robotics Research (IJRR)*, 2015. 2, 3

[46] S. Lynen, T. Sattler, M. Bosse, J. A. Hesch, M. Pollefeys, and R. Siegwart. Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization. In *Robotics: Science and Systems (RSS)*, 2015. 1, 3, 4

[47] A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2

[48] M. L. Nielsen. Augmented Reality and its Impact on the Internet, Security, and Privacy. https://beyondstandards.ieee.org/augmented-reality/augmented-reality-and-its-impact-on-the-internet-security-and-privacy/, 2015. 1

[49] D. Nistér and H. Stewénius. A minimal solution to the generalised 3-point pose problem. *Journal of Mathematical Imaging and Vision*, 2007. 3, 5

[50] F. Pittaluga, S. Koppal, S. B. Kang, and S. N. Sinha. Revealing scenes by inverting structure from motion reconstructions. In *CVPR*, 2019. 1

[51] R. Pless. Using many cameras as one. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003. 3, 4

[52] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm. USAC: a universal framework for random sample consensus. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2013. 3

[53] Z. Ren, Y. J. Lee, and M. S. Ryoo. Learning to anonymize faces for privacy preserving action detection. *European Conference on Computer Vision (ECCV)*, 2018. 3

[54] F. Roesner. Who Is Thinking About Security and Privacy for Augmented Reality? https://www.technologyreview.com/s/609143/who-is-thinking-about-security-and-privacy-for-augmented-reality/, 2017. 1

[55] M. S. Ryoo, B. Rothrock, C. Fleming, and H. J. Yang. Privacy-preserving human activity recognition from extreme low resolution. In *Proceedings of the Thirty-First Conference on Artificial Intelligence (AAAI)*, 2017. 3

[56] T. Sattler, M. Havlena, F. Radenovic, K. Schindler, and M. Pollefeys. Hyperpoints and fine vocabularies for large-scale location recognition. In *International Conference on Computer Vision (ICCV)*, 2015. 2

[57] T. Sattler, B. Leibe, and L. Kobbelt. Fast image-based localization using direct 2d-to-3d matching. In *International Conference on Computer Vision (ICCV)*, 2011. 2

[58] T. Sattler, B. Leibe, and L. Kobbelt. Efficient & effective prioritized matching for large-scale image-based localization. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2017. 1, 2, 3, 4

[59] T. Sattler, A. Torii, J. Sivic, M. Pollefeys, H. Taira, M. Okutomi, and T. Pajdla. Are large-scale 3D models really necessary for accurate visual localization? In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[60] J. L. Schönberger and J.-M. Frahm. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1, 3, 6

[61] J. L. Schönberger, M. Pollefeys, A. Geiger, and T. Sattler. Semantic visual localization. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[62] J. L. Schönberger, F. Radenović, O. Chum, and J.-M. Frahm. From Single Image Query to Detailed 3D Reconstruction. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1

[63] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016. 6

[64] M. Schreiber, C. Knöppel, and U. Franke. Laneloc: Lane marking based localization using highly accurate maps. In *IEEE Intelligent Vehicles Symposium (IV)*, 2013. 1

[65] J. Shashank, P. Kowshik, K. Srinathan, and C. Jawahar. Private content based image retrieval. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 2

[66] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *International Conference on Computer Vision (ICCV)*, 2003. 2

[67] H. Stewénius, D. Nistér, M. Oskarsson, and K. Aström. Solutions to minimal generalized relative pose problems. *OMNIVIS workshop*, 2005. 2, 3, 4, 5

[68] C. Sweeney, J. Flynn, B. Nuernberger, M. Turk, and T. Hollerer. Efficient computation of absolute pose for gravity-aware augmented reality. *International Symposium on Mixed and Augmented Reality (ISMAR)*, 2015. 2, 3, 5

[69] C. Sweeney, J. Flynn, and M. Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. *International Conference on 3D Vision (3DV)*, 2015. 2, 3, 5

[70] C. Sweeney, V. Fragoso, T. Höllerer, and M. Turk. gDLS: A scalable solution to the generalized pose and scale problem. In *European Conference on Computer Vision (ECCV)*, 2014. 3, 5, 6

[71] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2002. 3

[72] S. Thirthala and M. Pollefeys. Radial multi-focal tensors: Applications to omnidirectional camera calibration. *International Journal of Computer Vision (IJCV)*, 2012. 3

[73] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1991. 5, 6

[74] M. Upmanyu, A. M. Namboodiri, K. Srinathan, and C. Jawahar. Efficient privacy preserving video surveillance. In *International Conference on Computer Vision (ICCV)*, 2009. 2

[75] M. Upmanyu, A. M. Namboodiri, K. Srinathan, and C. Jawahar. Blind authentication: a secure crypto-biometric verification protocol. *IEEE Transactions on Information Forensics and Security*, 2010. 3

[76] J. Ventura, C. Arth, G. Reitmayr, and D. Schmalstieg. Global localization from monocular slam on a mobile phone. *Transactions on Visualization and Computer Graphics (TVCG)*, 2014. 2

[77] J. Ventura, C. Arth, G. Reitmayr, and D. Schmalstieg. A minimal solution to the generalized pose-and-scale problem. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3, 5

[78] J.-E. Vinje. Privacy Manifesto for AR Cloud Solutions. https://medium.com/openarcloud/privacy-manifesto-for-ar-cloud-solutions-9507543f50b6, 2018. 1

[79] VPS. Google Visual Positioning System. https://www.engadget.com/2018/05/08/g/, 2018. 1

[80] F. Walch, C. Hazirbas, L. Leal-Taixé, T. Sattler, S. Hilsenbeck, and D. Cremers. Image-based localization with spatial LSTMs. In *International Conference on Computer Vision (ICCV)*, 2017. 2

[81] T. Weyand, I. Kostrikov, and J. Philbin. Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision (ECCV)*, 2016. 2

[82] R. Yonetani, V. N. Boddeti, K. M. Kitani, and Y. Sato. Privacy-preserving visual learning using doubly permuted homomorphic encryption. In *International Conference on Computer Vision (ICCV)*, 2017. 3

[83] B. Zeisl, T. Sattler, and M. Pollefeys. Camera pose voting for large-scale image-based localization. In *International Conference on Computer Vision (ICCV)*, 2015. 2, 3, 4