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

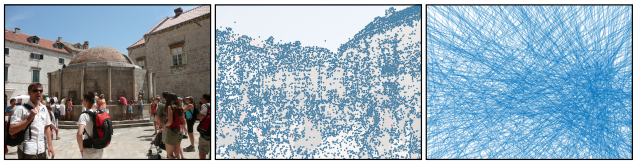
Abstract

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 2D positions. It will be shown that this feature representation hides the image contents, and thereby protects user privacy, yet still provides sufficient geometric constraints to enable robust and accurate 6-DOF camera pose estimation from feature correspondences. Our proposed method can handle single- and multi-image queries as well as exploit additional information about known structure, gravity, and scale. Numerous experiments demonstrate the high practical relevance of our approach.

1. Introduction

Estimating the 6-DOF camera pose from an image is a fundamental problem in computer vision. It is crucial for localization and navigation tasks in augmented/mixed reality (AR/MR) and robotics. Structure-based camera pose estimation methods are now quite mature [28, 38, 40, 54, 79] and deployed in many products (*e.g.*, Microsoft HoloLens, Windows MR Headsets, Magic Leap One, Oculus Quest, Google Maps AR). These methods first match local feature points in a query image to a pre-computed 3D point cloud of the scene. Each 2D–3D point correspondence provides two geometric constraints for estimating camera pose.

Recently, Pittaluga *et al*. [46] showed that sparse 3D point clouds and descriptors can be inverted to synthesize detailed and recognizable images of the scene. Their work emphasizes the inherent privacy risks associated with the persistent storage and sharing of 3D point clouds models. Speciale *et al*. [60] proposed the first solution to address this problem by developing a privacy preserving camera pose estimation technique. They propose to transform 3D point clouds to 3D line clouds in a way that obfuscates the 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] partially avoid the privacy issues associated with sharing confidential 3D point clouds, as they do not explicitly store 3D models. However, model inversion [41] poses privacy risks even for these methods and they are still not efficient and accurate enough [55, 71] for deployment in products.



(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.

Recently launched commercial cloud-based localization services such as Microsoft Azure Spatial Anchors [12], Google Visual Positioning System [2], 6D.AI [1], and Scape Technologies [3] require mobile devices to upload images or features to a server. A prominent example is the Google Maps AR navigation feature [23], which combines GPS- and VPS-based localization and has already been deployed in 93 countries. Protecting the privacy of the map is less critical for such services as the map data remains on the server. However, uploading the query images or features to a server poses serious privacy risks to the user, because the images could reveal confidential information in the scene to an adversary on the server or a man-in-the-middle attacker.

As localization might be running in the background without users being consciously aware, the privacy implications are relevant for all kinds of users, ranging from enterprise users keen to avoid confidential corporate information from accidentally leaking to third parties to home users who want images from their home to remain private. These privacy risks are present even when sending only local image features instead of the full image, as the adversary can easily reconstruct the original image using feature inversion methods [16, 17]. Note that, for server-side localization, Speciale *et al*.'s work [60] and learning-based approaches cannot protect the user's privacy as they are unable to conceal the uploaded images or features on the server.

In this paper, we propose a new privacy preserving visual localization approach that involves transforming the query image features before sending them to the server. The proposed transformation prevents an adversary from recovering the image appearance and recognizing confidential information in the scene. This is the first solution and a crucial step towards mitigating privacy risks in cloud-based localization services. It enables their use without the risk of man-in-the-middle attacks or having to trust the server.

Our solution is inspired by the transformation proposed by Speciale *et al*. [60]. While they transform 3D points in the map to 3D lines, we propose to replace 2D points in the query image with 2D lines. Specifically, each randomly oriented 2D line passes through the original 2D point which is subsequently discarded (see Fig. 1 for an example). Our proposed pose estimation approach requires uploading only the 2D lines and associated feature descriptors to the server. 2D feature locations are unavailable on the server, making it infeasible to invert the features. We show how to robustly and efficiently estimate the 6-DOF camera pose with respect to a 3D point cloud model given 2D line to 3D point correspondences. Our obfuscation method leads to different geometric constraints from the prior work [60]. In our work, we exploit the fact that 2D image lines back-project to 3D planes which must contain their corresponding 3D points.

Furthermore, we consider another case where both the query and the map are confidential. As such, we enable localization to be performed by a third party without allowing it to gain confidential information from either the query image or the pre-computed 3D point cloud of the scene.

Contributions. In summary, this paper: **(1)** considers a new variant of privacy preserving visual localization which is relevant for server-side localization and preserves the privacy of the user's query images; **(2)** it involves transforming 2D feature points to 2D lines in the image, while retaining sufficient geometric constraints for 6-DOF camera pose estimation; **(3)** the transformation prevents inversion of 2D features and thus conceals the image contents; **(4)** we present a method to conceal both the query and the map; **(5)** we implement, evaluate and study several variants of these problems involving a single input image or multiple images, with and without known structure in the query, the knowledge of the vertical direction, or the scale of the scene.

2. Related Work

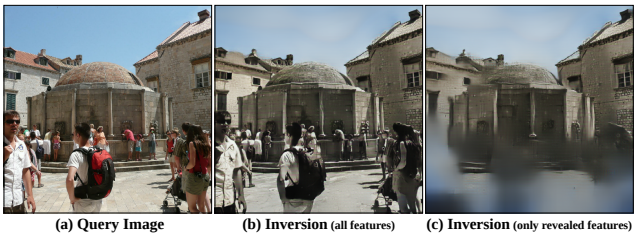
We now discuss relevant work on feature inversion, visual localization, and privacy preserving vision.

Reconstructing Images from Features. Weinzaepfel *et al*. [72] were the first to invert SIFT features. Subsequently, others tried inverting and interpreting HOG features [70], bag of visual words features [29], and CNN features [41, 77, 78]. Dosovitskiy and Brox trained CNN architectures to efficiently invert image features [16, 17]. Pittaluga *et al*. [46] recently used similar CNN models to invert sparse SfM point clouds and descriptors, where they emphasized the privacy risks of storing such data permanently.

Visual Localization and Camera Pose Estimation. Recent progress in image-based localization techniques have led to methods that are robust to changes in scene appearance and illumination [7, 57], scalable [36, 53, 54, 79], and efficient [9, 15, 18, 28, 30, 36–38, 69]. Most localization approaches first recover putative matches between query image features and features associated with 3D structure. Then, the camera pose is typically estimated using minimal solvers within a RANSAC-based robust estimation framework. Often, the camera pose is computed from 2D point to 3D point matches by solving a perspective-n-point (PnP) problem [20]. Various solvers for the minimal case of 3 points (P3P) are known [20, 24] for central cameras and specialized solvers have been proposed for the known vertical direction case [34]. Nister and Stewenius [43] proposed minimal solvers for generalized cameras, whereas Sweeney *et al*. [64] dealt with unknown scale. Meanwhile, structure-less pose estimation methods are based on 2D to 2D point matches between the query image and the mapped images [82], while hybrid methods [14] are based on both 2D–2D and 2D–3D point matches. Generally, camera pose estimation is not limited to point-based features only. For example, the Perspective-N-Line (PNL) problem uses 2D line to 3D line correspondences [48, 75, 80]. In contrast, Speciale *et al*. [60] exploit 2D point to 3D line correspondence for camera pose estimation in a variety of settings. The resulting problems are solved using well known generalized camera pose estimation algorithms [35,43,61,62,64].

3D Point-to-Plane Registration. 3D geometric registration problems can be solved using various techniques ranging from iterative closest point (ICP) to optimal methods for registration of 3D points, lines, and planes [44]. ICP variants for point-to-plane registration have also been proposed [39, 45]. However, more efficient methods are known for the case of known point to plane correspondences [31, 49, 50]. Ramalingam *et al*. [49, 50] proposed minimal solvers for registering points to planes that require six correspondences. These have been used for point-plane SLAM using RGB-D sensors [65]. In most previous work [31,50,65], the planes arise from planar surfaces in the scene or sometimes from other geometric primitives [49]. However, while our method uses a point-to-plane solver, the planes in our method are virtual in the sense that they are obtained by back-projection of randomly oriented 2D image lines.

Privacy Preserving Visual Recognition. Avidan and Butman [10, 11] were one of the first to study privacy-aware techniques for computer vision for the face detection task.



(a) Query Image (b) Inversion (all features) (c) Inversion (only revealed features)

Figure 2: Feature Inversion Attack. Shown from left to right are a) the original query image, the image reconstructed b) using all the SIFT descriptors and c) using only the descriptors whose 2D positions are revealed during our pose estimation method. Notice, how people present in the scene are quite well concealed.

Similar approaches were explored for image retrieval [59], face recognition [19], video surveillance [67], biometric verification [68], and detecting computer screens in first-person video [33]. Recent works on privacy in vision include anonymization for activity recognition [51,52], learning models on private or encrypted data [4, 22, 76] and localizing objects within indoor scenes using active cameras [81]. However, these privacy preserving vision algorithms focus on recognition tasks and cannot be used for geometric vision problems such as camera pose estimation.

Privacy Preserving Localization in Mobile Systems. Efficient privacy preserving approaches for localizing smartphones in GPS-denied indoor scenes applicable to WiFi and cellular fingerprinting as well as cloud-based services have been developed [32]. Other anonymization approaches for location privacy for mobile systems have been explored [5, 6, 8, 21, 42]. However, these approaches are not suitable for precise pose estimation or other geometric vision problems. One crucial difference with our work is that these approaches are geared towards concealing the user's location whereas our approach focuses on concealing the appearance of the query images. Concealing the computed camera pose, *i.e.*, the user location on the server, is an interesting open problem that is out of the scope of this paper.

3. Proposed Method

This section first introduces the key concepts behind our privacy preserving method before discussing the case of localizing a single query image. We then extend this theory to jointly localizing multiple image queries and present an additional solution to the scenario where both the query and the map remain confidential. We also present solutions to several practical special cases, including known gravity direction and the case where we can obtain a local reconstruction of the scene with known or unknown scale.

3.1. Privacy Preserving Localization System

Our privacy preserving localization approach relies on a client-server architecture, where the client first extracts local 2D features from the query image (*e.g.*, SIFT [39]) and then sends them to the server for computing the camera pose w.r.t. a pre-computed 3D point cloud. The server takes the 2D image features from the client and matches them against the associated features of the 3D point cloud. The resulting 2D to 3D correspondences then provide constraints for camera pose estimation. Our approach is based on the same general architecture. However, it relies on a novel privacy preserving representation of the 2D feature points, which lead to different geometric constraints for pose estimation.

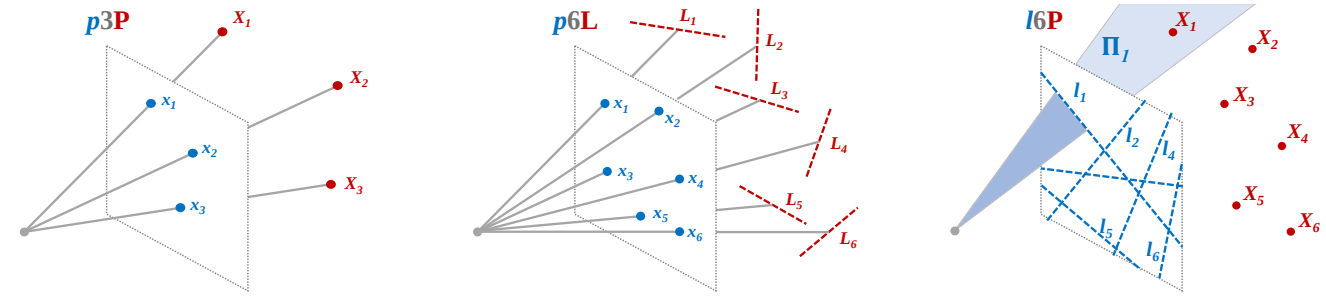
The main underlying idea is to obfuscate the 2D features extracted from the query image before sending them to the server. Opposed to actively detecting and masking potentially confidential objects (which can be error-prone), our method inherently conceals the whole image by transforming all the 2D feature points to randomly oriented 2D lines passing through the original point. The 2D line representation provides a single geometric constraint for accurate and efficient camera pose estimation. During the pose estimation procedure, the original 2D feature point locations of permanent scene structures will be revealed, whereas any confidential transient objects remain concealed. Note that revealing the permanent structures during pose estimation does not compromise privacy, because such structures are already present in the 3D map. An example of what is revealed is shown in Fig. 2. Furthermore, without possession of the 3D point cloud, all features remain concealed and no information about the image can be inferred. The latter is an effective defense against man-in-the-middle attacks that intercept the client-server connection.

3.1.1 Privacy Preserving Single-Image Queries

Our major contribution is inspired by Speciale *et al.* [60], who transform the 3D point cloud to a privacy preserving 3D line cloud. In the following, we adopt their notation and denote the normalized positions of the local 2D features in the query image as $\boldsymbol{x} \in \mathbb{R}^2$ and their corresponding 3D points in the map as $\boldsymbol{X} \in \mathbb{R}^3$. Equivalent to their approach, we assume known intrinsic camera parameters and rely on sparse 3D point clouds. The traditional approach, which is not privacy preserving, leverages 2D–3D point correspondences to derive the 6-DOF camera pose $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{R} & \boldsymbol{T}\end{bmatrix}$ with $\boldsymbol{R} \in \mathrm{SO}(3)$ and $\boldsymbol{T} \in \mathbb{R}^3$ based on the constraint

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

where $\bar{\boldsymbol{x}} \in \mathbb{P}^2$ and $\bar{\boldsymbol{X}} \in \mathbb{P}^3$ are the homogeneous representations of $\boldsymbol{x}$ and $\boldsymbol{X}$ in projective space, respectively. To account for outlier correspondences, this equation system is usually robustly solved for an initial estimate of $\boldsymbol{P}$ using minimal solvers like $p3$P [24] embedded in a variant of RANSAC [20]. A refined solution is then obtained by



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

Figure 3: Absolute Camera Pose Estimation. Left: using traditional 2D point to 3D point matches. Middle: using privacy preserving 3D line cloud representation [60]. Right: using our proposed privacy preserving 2D feature lines.

optimizing the following non-linear least-squares problem

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

formulated over the set of inliers from RANSAC.

As already noticed by Speciale *et al.* [60], such an approach reveals the geometry of the scene stored in the 3D point cloud. In order to conceal the 3D point cloud, they transform the 3D points $\boldsymbol{X}$ of the map to randomly oriented 3D lines $\boldsymbol{L} \in \mathbb{P}^5$. In contrast, we address the issue of leaking confidential image information through the 2D feature points. We propose to lift the 2D points $\boldsymbol{x} \in \mathbb{R}^2$ to randomly oriented 2D lines $\boldsymbol{l} \in \mathbb{P}^2$ in the image that pass through the original points such that $\boldsymbol{l}^T\bar{\boldsymbol{x}} = 0$. Since the original 2D feature positions are completely discarded and can be anywhere along the random lines, this representation obfuscates their layout in the query image. While this completely hides the image information, we will show that 2D line to 3D point correspondences still provide sufficient geometric constraints for camera pose estimation.

In our case, the lifted 2D lines lie in the image plane and do not represent rays in 3D space (see Fig. 3). Instead, their back-projection to 3D space defines a 3D plane $\boldsymbol{\Pi} = \boldsymbol{P}^T\boldsymbol{l} \in \mathbb{P}^3$ passing through the camera projection center and the 2D line in the image. Observe that, for an optimal pose $\boldsymbol{P}$, these back-projected 3D planes should contain their corresponding 3D points in the map. This observation can be formulated in the following geometric constraint

$$0 = \boldsymbol{\Pi}^T\bar{\boldsymbol{X}} = (\boldsymbol{P}^T\boldsymbol{l})^T\bar{\boldsymbol{X}} = \boldsymbol{l}^T\boldsymbol{P}\bar{\boldsymbol{X}} \ , \tag{3}$$

which can be used for camera pose estimation. This geometric problem is equivalent to the 3D point to 3D plane registration problem. Therefore, we can leverage existing minimal solvers [49, 50] for solving the equation system inside RANSAC. The solution to the problem requires a minimum of six 3D plane (back-projected 2D line) to 3D point correspondences and so we denote the problem as *l*6P.

The discussed minimal solution optimizes a 3D point-to-plane distance and produces an initial estimate of the camera pose. In a second step, we aim to refine this initial solution in order to obtain a more accurate result. Towards this goal, we interpret the geometric problem in a different way, such that we can formulate the error in 2D image space. Concretely, the projection of a 3D point must always be close to its corresponding 2D line in the image plane, independent of the 2D line orientation. The geometric constraint in Eq. (3) can be simply reinterpreted as

$$\boldsymbol{0} = \boldsymbol{l}^T\bar{\boldsymbol{x}} = \boldsymbol{l}^T\boldsymbol{P}\bar{\boldsymbol{X}} \ , \tag{4}$$

where we first project the 3D point $\boldsymbol{X}$ to image space. By minimizing the 2D point to 2D line distance in the image using non-linear least squares optimization of

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \frac{\boldsymbol{l}^T \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix}}{\sqrt{l_1^2 + l_2^2}} \quad \text{with } \boldsymbol{l} = \begin{bmatrix} l_1 & l_2 & l_3 \end{bmatrix}^T \ , \tag{5}$$

we obtain the final camera pose estimate. After deriving the theory for single-image localization, we next generalize our approach to the joint localization of multiple images.

3.1.2 Generalization to Multi-Image Queries

The joint localization of multiple images, if their relative pose $\boldsymbol{P}_c$ is known through rigid mounting or local tracking, brings great benefits in terms of recall and accuracy in image-based localization [60]. Instead of determining a separate pose $\boldsymbol{P}$ for each camera, we can reparameterize the pose of multiple images jointly as

$$\boldsymbol{P} = \boldsymbol{P}_c\,\boldsymbol{P}_m \quad \text{with} \quad \boldsymbol{P}_m = s_m \begin{bmatrix} \boldsymbol{R}_m & \boldsymbol{T}_m \\ \boldsymbol{0} & s_m^{-1} \end{bmatrix}. \tag{6}$$

Thus, we can estimate only a single transformation $\boldsymbol{P}_m \in \mathrm{Sim}(3)$, while the known relative poses $\boldsymbol{P}_c = \begin{bmatrix} \boldsymbol{R}_c & \boldsymbol{T}_c \end{bmatrix}$ of the individual cameras are fixed. Note that, if we know the relative scale of $\boldsymbol{P}_c$ w.r.t. the 3D points $\boldsymbol{X}$ in the map, we can drop the scale $s_m \in \mathbb{R}^+$ and simplify $\boldsymbol{P}_m$ to $\mathrm{SE}(3)$. The extension of Eq. (3) to multiple cameras is then straightforward by substituting $\boldsymbol{P}$ by $\boldsymbol{P}_c\,\boldsymbol{P}_m$ such that

$$0 = \boldsymbol{l}^T\boldsymbol{P}\bar{\boldsymbol{X}} = \boldsymbol{l}^T\boldsymbol{P}_c\,\boldsymbol{P}_m\bar{\boldsymbol{X}} = \boldsymbol{\Pi}_c^T\bar{\boldsymbol{X}}_m \ , \tag{7}$$

where $\boldsymbol{\Pi}_c = {\boldsymbol{P}_c}^T\boldsymbol{l} \in \mathbb{P}^3$ and $\bar{\boldsymbol{X}}_m = \boldsymbol{P}_m\bar{\boldsymbol{X}} \in \mathbb{P}^3$. While the 3D planes $\boldsymbol{\Pi}$ in Eq. (3) all pass through the projection

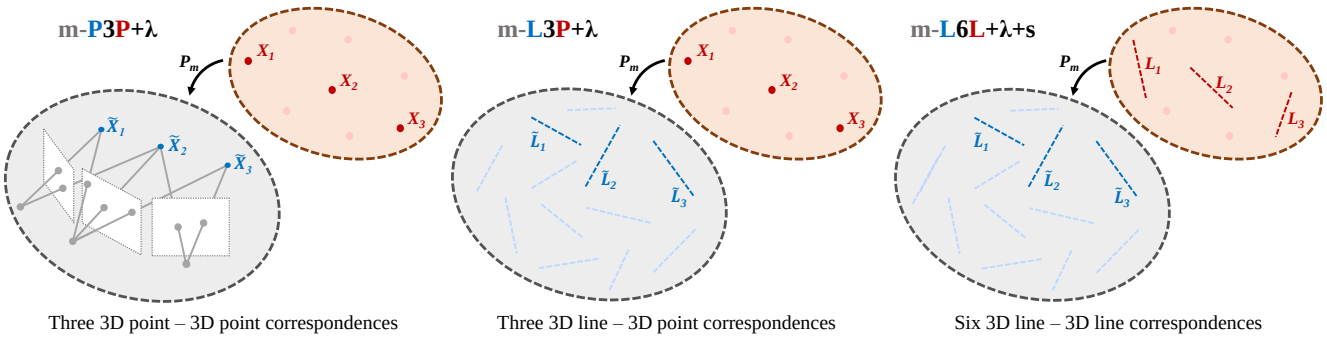


Figure 4: Pose Estimation with Known Structure. Left: traditional setup with 3D point cloud maps. Middle: our proposed approach using 3D line cloud query and the 3D map point cloud. Right: our extended proposed approach using 3D line cloud for both queries and map.

center of a single camera, we now have multiple bundles of planes $\boldsymbol{\Pi}_c$ passing through the respective projection centers of their cameras. The minimal solution to this problem can leverage the same solvers [49,50] as in the single-image query scenario and we denote the problem as m-*l*6P. Similarly, Eq. (5) still serves as our constraint for non-linear refinement of the minimal solution.

3.1.3 Pose Estimation with Known Structure

Equivalent to Speciale *et al*. [60], we also propose a solution to the scenario, where we can determine 3D structure in the query image, *e.g*., through multi-view triangulation or an active depth sensor. In other words, we now know the depth λ of an image observation $\boldsymbol{x}$, *i.e*., the 3D structure in the query image can be computed as $\tilde{\boldsymbol{X}} = \lambda\bar{\boldsymbol{x}}$. In the traditional localization problem, the camera pose can then be estimated as the 3D transformation that best aligns the two corresponding 3D point sets using the constraint

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

This equation system has a direct and efficient solution [27, 66], which we denote as m-P3P+λ in the case of unknown scale and as m-P3P+λ+s in the case of known scale.

In contrast to Speciale *et al*. [60], we want to hide the 3D structure of the query instead of the map. We therefore lift the 3D points $\tilde{\boldsymbol{X}}$ of the query to randomly oriented 3D lines in Plücker coordinates [47] as $\tilde{\boldsymbol{L}} = \begin{bmatrix}\tilde{\boldsymbol{v}} & \tilde{\boldsymbol{w}}\end{bmatrix}^T \in \mathbb{P}^5$ with $\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{X}} \times \tilde{\boldsymbol{v}}$, leading to the new geometric constraints

$$\mathbf{0} = (\tilde{\boldsymbol{v}} \times \tilde{\boldsymbol{w}} + \beta\tilde{\boldsymbol{v}}) - \boldsymbol{P}\bar{\boldsymbol{X}} \ , \tag{9}$$

which we can exploited for camera pose estimation. It turns out that this problem can be solved using existing minimal solutions, and it is geometrically equivalent to the generalized absolute pose problem with known [25] and unknown scale [64]. We denote the minimal problems as m-L3P+λ+s for known scale and m-L4P+λ for unknown scale.

3.1.4 Confidential Query and Confidential Map

Both Speciale *et al*. [60] and this paper so far only described techniques to protect the confidentiality of either the query or the map. In this section, we address this limitation by deriving an approach that obfuscates both the query and the map simultaneously, yet still allows for camera pose estimation. This enables localization where both the query images as well as the pre-computed maps contain confidential objects, enabling localization to be performed by an untrusted third party without leaking information.

In order to solve this problem, we obfuscate both the query and the map representation. Note that this only works when we have known structure on the query side, because solving the problem without known structure relies on lifting 2D points to 2D lines. However, correspondences between 3D planes (back-projected 2D lines) in the query with their corresponding 3D lines in the map do not provide geometric constraints for camera pose estimation, as they generally always intersect, independent of their alignment.

To protect query and map information, we lift the 3D points $\tilde{\boldsymbol{X}}$ of the query and $\boldsymbol{X}$ of the map to 3D lines $\tilde{\boldsymbol{L}}$ and $\boldsymbol{L}$, respectively. The resulting constraints are then

$$\mathbf{0} = (\tilde{\boldsymbol{v}} \times \tilde{\boldsymbol{w}} + \beta\tilde{\boldsymbol{v}}) - \boldsymbol{P}\begin{bmatrix}\boldsymbol{v} \times \boldsymbol{w} + \alpha\boldsymbol{v} \\ 1\end{bmatrix} \ , \tag{10}$$

which can be interpreted as a 3D line to 3D line intersection problem. It turns out that this scenario is yet again geometrically equivalent to the generalized relative pose problem. Each 3D line in our case can be represented with a separate camera ray in the relative pose estimation setting. As such, we can rely on existing minimal solvers [61] for finding an initial estimate of the camera pose inside RANSAC.

The non-linear refinement of the initial pose is more difficult, as we cannot easily compute an error in image space anymore. The reason being that we neither know the original 2D point location in the image nor the corresponding original 3D point location in the map. Minimizing the reprojection error in image space is, however, required in order to find the maximum likelihood estimate under the assumption that the image observations $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{x}})$ have Gaussian distributed errors. We transform the 3D line $\tilde{\boldsymbol{L}}$ of the query to the map coordinate system using the current pose estimate as $[\dot{\boldsymbol{L}}]_\times = \boldsymbol{P}_m^{-1}[\tilde{\boldsymbol{L}}]_\times \boldsymbol{P}_m^{-T}$. Assuming that the 3D map is error-free, we can find the maximum likelihood estimate $\dot{\boldsymbol{X}}$ of the true 3D point location of $\boldsymbol{X}$ by finding

Constraints	*Query Type*	POINT TO POINT (Traditional)				LINE TO POINT (Proposed)			
2D – 3D	*Single-Image*	*p*3P	[24]	*p*2P+u	[62]	*l*6P	[49]	*l*4P+u	[49]
	Multi-Image	m-*p*3P	[25]	m-*p*2P+u	[27]	m-*l*6P	[49]	m-*l*4P+u	[49]
3D – 3D	*Multi-Image*	m-P3P+λ	[66]	m-P2P+λ+u	[66]	m-L4P+λ	[64]	m-L3P+λ+u	[14]
		m-P3P+λ+s	[27]	m-P2P+λ+u+s	[27]	m-L3P+λ+s	[25]	m-L2P+λ+u+s	[62]
						LINE TO LINE (Proposed)			
3D– 3D	*Multi-Image*	m-P3P+λ+s	[27]	m-P2P+λ+u+s	[27]	m-L6L+λ+s	[61]	m-L4L+λ+u+s	[63]

Table 1: Camera Pose Problems. Traditional methods using point to point correspondences are called ***p**P*** (2D to 3D) and **P*P** (3D to 3D), whereas the ones using lines to points are called ***l**P*** (2D to 3D) and **L*P** (3D to 3D). We also refer as **L*L** (3D to 3D) to the case where map and query images are both obfuscated with lines. In the first row, the methods take single query images as input, whereas the rest of the methods jointly localize multiple query images (prefix **m**). These methods include general solvers as well specialized ones for known vertical direction (suffix **+u**). The methods in the last three rows exploit known 3D structure estimated from multiple query images (suffix $+\boldsymbol{\lambda}$ for known structure and suffix **+s** for known scale).

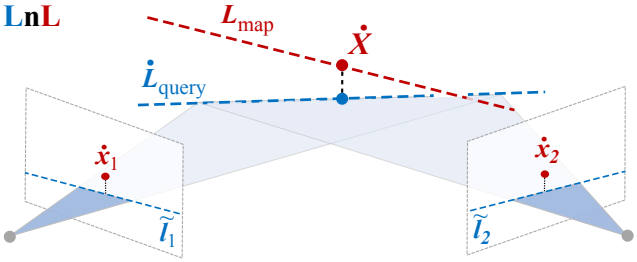


Figure 5: LnL Methods. After computing the camera pose estimation $\boldsymbol{P}$, we need to find the closest point $\dot{\boldsymbol{X}}$ on the 3D line $\boldsymbol{L}$ of the map in order to compute a 2D geometric error.

the closest point of transformed line $\dot{\boldsymbol{L}}$ on the 3D line $\boldsymbol{L}$ in the map coordinate system, as it is illustrated in Fig. 5.

Using the determined 3D point location $\dot{\boldsymbol{X}}$, we can compute the geometric error between the two 3D lines in image space by projecting both $\dot{\boldsymbol{X}}$ and $\dot{\boldsymbol{L}}$ to the respective cameras from which the query lines $\tilde{\boldsymbol{l}}$ are observed. Concretely, we non-linearly optimize the cost function

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \frac{\tilde{\boldsymbol{l}}^T \boldsymbol{P} \dot{\boldsymbol{X}}}{\sqrt{\tilde{l}_1^2 + \tilde{l}_2^2}} \text{ with } [\tilde{\boldsymbol{l}}]_\times = \boldsymbol{P}_c [\tilde{\boldsymbol{L}}]_\times \boldsymbol{P}_c^T, \quad (11)$$

where $[\tilde{\boldsymbol{l}}]_\times$ and $[\tilde{\boldsymbol{L}}]_\times$ are defined as

$$[\tilde{\boldsymbol{l}}]_\times = \begin{bmatrix} 0 & -\tilde{l}_3 & \tilde{l}_2 \\ \tilde{l}_3 & 0 & -\tilde{l}_1 \\ -\tilde{l}_2 & \tilde{l}_1 & 0 \end{bmatrix}, \; [\tilde{\boldsymbol{L}}]_\times = \begin{bmatrix} -[\tilde{\boldsymbol{w}}]_\times & -\tilde{\boldsymbol{v}} \\ \tilde{\boldsymbol{v}}^T & 0 \end{bmatrix}. \quad (12)$$

The geometric error between the projected confidential 3D line $\tilde{\boldsymbol{L}}$ and the projected 3D point $\dot{\boldsymbol{X}}$ can be differentiated analytically w.r.t. $\boldsymbol{P}$ and its minimization thus efficiently yields the final camera pose. We denote the solution to the privacy preserving query and map problem as m-L6L+λ+s, as it requires a minimum of six 3D line to 3D line correspondences with known structure and scale. Note that we do not consider the case with unknown scale, because we are not aware of a minimal solution to the generalized relative pose problem with unknown scale.

3.1.5 Specialization with Known Vertical

Often, an estimate of the gravity direction in both the reference frame of the query and the map is available through means of an inertial measurement unit or vanishing point detection in the images. By pre-aligning the two reference frames to the same vertical direction, one can reduce the number of rotational pose parameters from three to one. The parameterization of the rotation then simplifies to a single quadratic constraint, leading to more efficient and numerically stable solutions. Furthermore, the minimal solutions require fewer correspondences and thus result in a better runtime of RANSAC. We implement the known gravity setting for all described problems, indicated by the suffix *+u*. See Table 1 for an overview of all the methods.

4. Experimental Evaluation

4.1. Setup

Datasets. For evaluating the different specializations (known structure, scale, and gravity) of our proposed solvers, we use 15 real-world datasets [60] of complex indoor and outdoor scenes captured using a mix of mobile phones and the Microsoft HoloLens [26]. In the 15 datasets, there are a total of 375 single-image and 402 multi-image queries for evaluation, evenly spread across the scenes. In addition, we also evaluate our unconstrained single-image approach (*p*3P and *l*3P) on four large-scale Internet photo collection datasets [37, 74], which are well-established benchmarks in the community. For sparse 3D scene reconstruction, we rely on the COLMAP SfM pipeline [56, 58]. For a fair comparison, all methods use exactly the same 2D–3D correspondences, thresholds, and RANSAC implementation [20]. See supplementary material for more details and visualizations.

Metrics. We compute the rotational and translational errors as $\Delta R = \arccos \frac{\text{Tr}(\boldsymbol{R}^T \hat{\boldsymbol{R}}) - 1}{2}$ and $\Delta T = \|\boldsymbol{R}^T \boldsymbol{T} - \hat{\boldsymbol{R}}^T \hat{\boldsymbol{T}}\|_2$. In the supplementary material, we also report the average point-to-point and line-to-point reprojection errors (*c.f.*

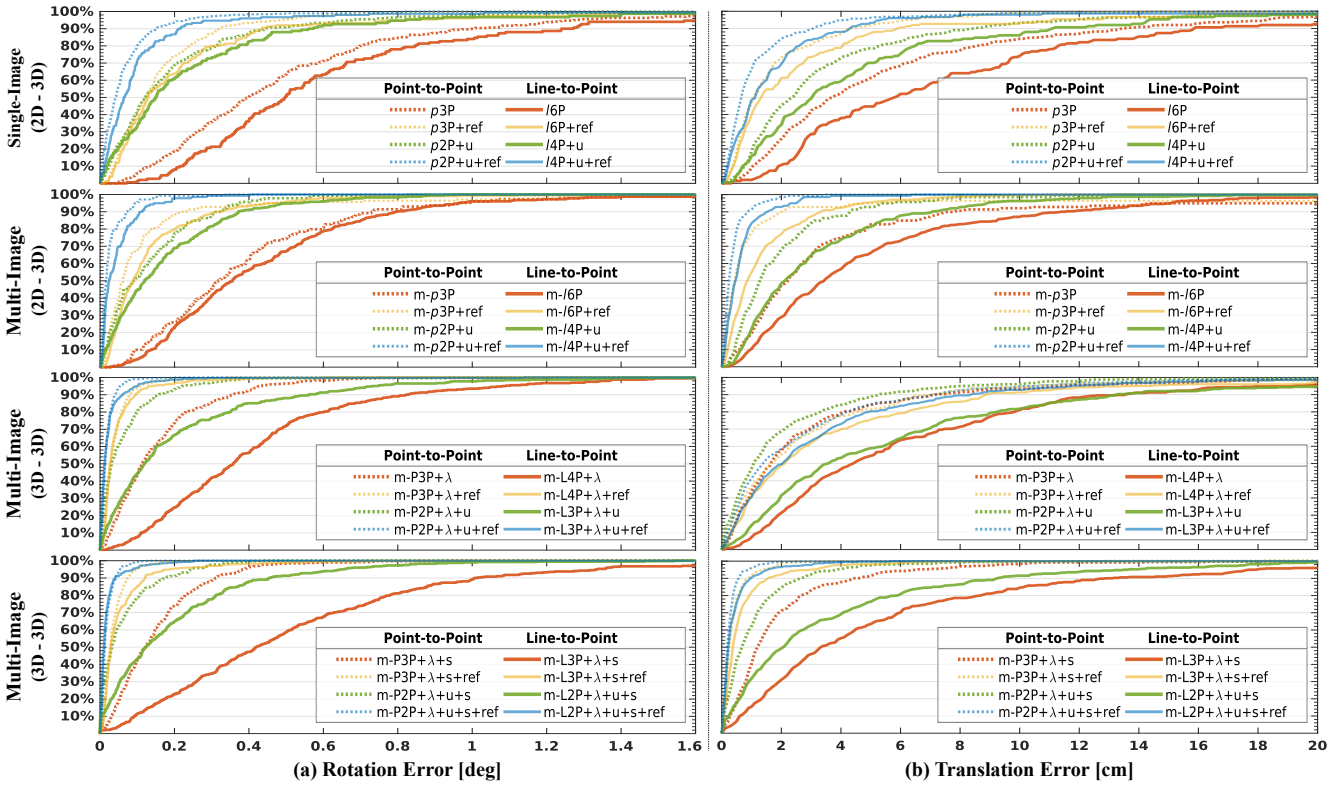


Figure 6: Results for Privacy Preserving Image Query. Cumulative rotation and translation error histograms for all 16 evaluated methods. Privacy preserving method yields achieve almost as accurate results as the traditional approach, especially after non-linear refinement.

Eq. (2) and (5)) in addition to other details on runtime, *etc*.

Methods. We compare all our proposed privacy preserving to their corresponding 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. The initial results are then further refined (suffix *+ref*) using a Levenberg-Marquardt optimization of the cost functions in Eqs. (5) and (11) based on the inliers from RANSAC.

4.2. Results

Mobile and HoloLens Datasets. Fig. 6 shows detailed accuracy and recall statistics, where our proposed privacy preserving method achieves almost as accurate results as the traditional approach, despite leveraging a single instead of two geometric constraints per correspondence. In absolute terms, both the traditional as well as our approach achieve state-of-the-art results, sufficient to enable precise localization in AR and robotics scenarios. Similar to Speciale *et al*. [60], the localization results improve by incorporating known information about structure, gravity, and scale.

Internet Photo Collections. We also report results on well-known large-scale localization benchmarks crowd-sourced from the Internet (see Fig. 7). Even though good performance on these datasets is mainly determined by the performance of correspondence search, this experiment demonstrates the feasibility of privacy preserving single-image localization in large-scale outdoor environments. Our approach consistently yields a median localization accuracy below 1° and 2cm, comparable to state-of-the-art approaches. Furthermore, our approach achieves the same recall as the traditional localization method.

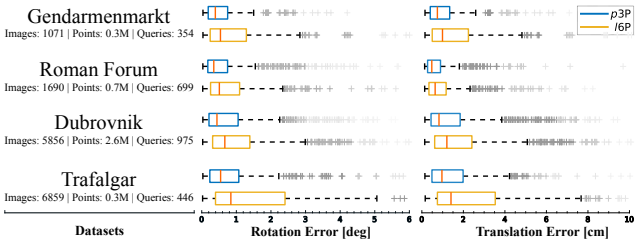


Figure 7: Internet Photo Collection Dataset Results. Dataset names [37, 74] and statistics on the left with corresponding localization errors for single-image scenario as box plots to the right.

Confidential Query and Map. The previous two paragraphs discussed results for privacy preserving image queries with a traditional 3D point cloud based map representation. Fig. 8 also evaluates the scenario where both the query and the map are confidential. Compared to either concealing the query (see m-LnP in Fig. 6) or the map (see m-PnL in [60]), this approach has slightly lower accuracy. Nevertheless, it still produces accurate localization results in absolute numbers, especially the refined solutions and the more constrained scenario with known gravity.

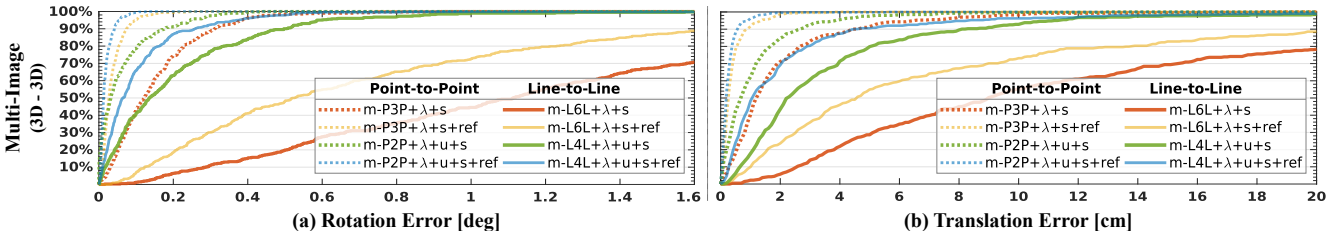


Figure 8: Privacy Preserving Image Query and Map Results. Cumulative rotation and translation error histograms with both concealed query and the map show slightly lower compared to m-LnP or m-PnL, yet still competitive in absolute terms, especially for m-L4L+λ+u+s.

5. Discussion

We now discuss other insights and potential future work.

Feature Line Triangulation Attack. When sending multiple queries from the same confidential scene to the server, corresponding 2D feature lines in different views can potentially be triangulated to a 3D point. Since 2D image lines back-project to 3D planes, at least three non-coplanar planes must intersect to yield a 3D point. This 3D point can then be back-projected to the views to find the approximate 2D position of the original features in the images. Such an attack could be exploited by an attacker on the server to apply feature inversion techniques in 2D or 3D [16, 17, 46] to recover the appearance of confidential content. A common expected scenario for our method is where the confidential content is transient and moving (*e.g.*, people in the scene). This inherently ensures the above attack would not work because the triangulation scheme is only valid for objects which are stationary across different views. A more principled defense requires the user to never upload image feature descriptors to the server. Instead, the server must send the map descriptors to the client. The client then matches the image features to the map features and only sends successfully matched 2D lines to the server. Using such an approach, the client cannot recover any information from the map, as it does not have the 3D point coordinates. Also, the server cannot infer anything from the query either, because it does not have the image descriptors. Although this may require a larger exchange of data (at least for the first query). Alternately, the converse approach is also valid. The client uploads only the descriptors to the server (without 2D lines features), the server returns only successfully matched 3D map features to the client (perhaps in 3D line format [60]); and then, clients can perform localization locally. These **two-step protocols** can be more secure than sending all the descriptors and features at once.

Towards Concealing the User's Location. So far, we focused on ensuring that the appearance of the query image remains confidential. However, after successful pose estimation, the precise location of the user w.r.t. the 3D point cloud is revealed. If precise location information is a privacy concern for certain applications, it can be easily hidden from the server by only using multi-image queries with known structure. In this case, the client only sends 3D lines to the server without the corresponding relative camera poses $\boldsymbol{P}_c$ in the query. Note that the server can still perform non-linear refinement but by generating its own virtual cameras. By not having the actual cameras, the server cannot recover the precise location of the user in the scene anymore. However, an approximate user location is still known based on knowing which 3D points were observed in the image. Hiding the location completely is not the focus of this paper and an interesting research direction.

Feature Line Intersection. Similar to the 3D line cloud attack discussed in [60], our 2D feature lines could potentially be intersected with each other to find the approximate positions of the original 2D feature points. This can be especially effective, if nearby feature descriptors were extracted from overlapping local patches. In this case, an attacker can try to find similar descriptors with partial overlap and only intersect their lines to reduce ambiguity. However, such an attack can be easily mitigated by applying sparsification [60] of the 2D features and ensuring that their image patches do not overlap.

Compactness of Representation. The traditional approach based on 2D feature points requires 2 floats to represent the location of an image feature. By discretizing the line directions, we can represent random 2D lines using 1 float for the distance to the origin and 1 byte for the orientation, if using a finite set of 256 directions. Similarly, 3D lines can be compactly represented by 2 floats and 1 byte [60].

6. Conclusion

This paper address privacy concerns associated with current cloud-based visual localization services. In our scenario, we protect user privacy by concealing image information from the server. This is a fundamental step towards enabling image-based localization services to be deployed in a wide range of scenarios. For example, privacy preserving methods are absolutely mandatory, if we aim to not leak confidential corporate information or private details in everyone's home. Our approach is based on novel geometric constraints leading to efficient solutions to the camera pose estimation problem. Experiments on a large variety of datasets demonstrate robust and accurate results for our method. Last but not least, with this work, we not only intend to propose a practical solution to a novel problem, but we also hope to increase privacy awareness in the community and thereby encourage future work in this direction.

References

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

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

[3] Scape Technologies. https://scape.io/, 2019. 1

[4] 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

[5] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. 2008. 3

[6] 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

[7] 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

[8] 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

[9] 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

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

[11] S. Avidan and M. Butman. Efficient methods for privacy preserving face detection. In *Conference on Neural Information Processing Systems (NIPS)*, 2007. 2

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

[13] 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. 1

[14] F. Camposeco, A. Cohen, M. Pollefeys, and T. Sattler. Hybrid camera pose estimation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 6

[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] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In *Advances in Neural Information Processing Systems*, 2016. 1, 2, 8

[17] A. Dosovitskiy and T. Brox. Inverting visual representations with convolutional networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1, 2, 8

[18] 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

[19] 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. 3

[20] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. 1981. 2, 3, 6

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

[22] 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. 2016. 3

[23] Google AR Navigation. https://ai.googleblog.com/2019/02/using-global-localization-to-improve.html, 2019. 1

[24] 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. 2, 3, 6

[25] 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. 5, 6

[26] HoloLens. https://www.microsoft.com/en-us/hololens, 2016. 6

[27] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 1987. 5, 6

[28] 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

[29] H. Kato and T. Harada. Image reconstruction from bag-of-visual-words. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 2

[30] 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. 1, 2

[31] K. Khoshelham. Direct 6-dof pose estimation from point-plane correspondences. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2015. 2

[32] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis. Privacy-preserving indoor localization on smartphones. *IEEE Transactions on Knowledge and Data Engineering*, 2015. 3

[33] 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

[34] Z. Kukelova, M. Bujnak, and T. Pajdla. Closed-form solutions to minimal absolute pose problems with known vertical direction. In *Asian Conference on Computer Vision (ACCV)*, 2010. 2

[35] 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. 2

[36] 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

[37] Y. Li, N. Snavely, and D. P. Huttenlocher. Location recognition using prioritized feature matching. In *European Conference on Computer Vision (ECCV)*, 2010. 2, 6, 7

[38] 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. 1, 2

[39] K.-L. Low. Linear least-squares optimization for point-to-plane icp surface registration. *Chapel Hill, University of North Carolina*, 2004. 2, 3

[40] 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

[41] A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. In *CVPR*, 2015. 1, 2

[42] M. E. Nergiz, M. Atzori, and Y. Saygin. Towards trajectory anonymization: a generalization-based approach. In *Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS*, pages 52–61. ACM, 2008. 3

[43] D. Nistér and H. Stewénius. A minimal solution to the generalised 3-point pose problem. *Journal of Mathematical Imaging and Vision*, 2007. 2

[44] C. Olsson, F. Kahl, and M. Oskarsson. The registration problem revisited: Optimal solutions from points, lines and planes. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 2

[45] S.-Y. Park and M. Subbarao. An accurate and fast point-to-plane registration technique. *Pattern Recognition Letters*, 2003. 2

[46] F. Pittaluga, S. Koppal, S. B. Kang, and S. N. Sinha. Revealing scenes by inverting structure from motion reconstructions. In *CVPR*, 2019. 1, 2, 8

[47] R. Pless. Using many cameras as one. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003. 5

[48] B. Pibyl, P. Zemk, and M. adik. Camera pose estimation from lines using plcker coordinates. In *British Machine Vision Conference (BMVC)*, 2015. 2

[49] S. Ramalingam and Y. Taguchi. A theory of minimal 3d point to 3d plane registration and its generalization. *International Journal of Computer Vision (IJCV)*, 2013. 2, 4, 5, 6

[50] S. Ramalingam, Y. Taguchi, T. K. Marks, and O. Tuzel. P2π: A minimal solution for registration of 3d points to 3d planes. In *European Conference on Computer Vision (ECCV)*, 2010. 2, 4, 5

[51] 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

[52] 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

[53] 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

[54] 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

[55] 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. 1

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

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

[58] 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

[59] J. Shashank, P. Kowshik, K. Srinathan, and C. Jawahar. Private content based image retrieval. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 3

[60] P. Speciale, J. Schönberger, S. B. Kang, S. N. Sinha, and M. Pollefeys. Privacy Preserving Image-based Localization. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 3, 4, 5, 6, 7, 8

[61] H. Stewénius, D. Nistér, M. Oskarsson, and K. Aström. Solutions to minimal generalized relative pose problems. *OMNIVIS workshop*, 2005. 2, 5, 6

[62] C. Sweeney, J. Flynn, B. Nuernberger, M. Turk, and T. Hollerer. Efficient computation of absolute pose for gravity-aware augmented reality. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, 2015. 2, 6

[63] 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. 6

[64] 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. 2, 5, 6

[65] Y. Taguchi, Y.-D. Jian, S. Ramalingam, and C. Feng. Point-plane slam for hand-held 3d sensors. In *International Conference on Robotics and Automation (ICRA)*, 2013. 2

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

[67] M. Upmanyu, A. M. Namboodiri, K. Srinathan, and C. Jawahar. Efficient privacy preserving video surveillance. In *International Conference on Computer Vision (ICCV)*, 2009. 3

[68] 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

[69] 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

[70] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba. Hoggles: Visualizing object detection features. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 2

[71] 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. 1

[72] P. Weinzaepfel, H. Jégou, and P. Pérez. Reconstructing an image from its local descriptors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 2

[73] T. Weyand, I. Kostrikov, and J. Philbin. Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision (ECCV)*, 2016. 1

[74] K. Wilson and N. Snavely. Robust global translations with 1dsfm. In *European Conference on Computer Vision (ECCV)*, 2014. 6, 7

[75] C. Xu, L. Zhang, L. Cheng, and R. Koch. Pose estimation from line correspondences: A complete analysis and a series of solutions. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2017. 2

[76] 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

[77] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson. Understanding neural networks through deep visualization. In *ICML Workshop on Deep Learning*, 2015. 2

[78] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision (ECCV)*, 2014. 2

[79] B. Zeisl, T. Sattler, and M. Pollefeys. Camera pose voting for large-scale image-based localization. In *International Conference on Computer Vision (ICCV)*, 2015. 1, 2

[80] L. Zhang, C. Xu, K.-M. Lee, and R. Koch. Robust and efficient pose estimation from line correspondences. In *Asian Conference on Computer Vision (ACCV)*, 2012. 2

[81] J. Zhao, N. Frumkin, J. Konrad, and P. Ishwar. Privacy-preserving indoor localization via active scene illumination. In *Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018. 3

[82] E. Zheng and C. Wu. Structure from motion using structure-less resection. In *International Conference on Computer Vision (ICCV)*, 2015. 2