

Short Papers

Bound and Conquer: Improving Triangulation by Enforcing Consistency

Adam Scholefield¹, *Member, IEEE*,
Alireza Ghasemi, *Student Member, IEEE*, and
Martin Vetterli¹, *Fellow, IEEE*

Abstract—We study the accuracy of triangulation in multi-camera systems with respect to the number of cameras. We show that, under certain conditions, the optimal achievable reconstruction error decays quadratically as more cameras are added to the system. Furthermore, we analyze the error decay-rate of major state-of-the-art algorithms with respect to the number of cameras. To this end, we introduce the notion of consistency for triangulation, and show that consistent reconstruction algorithms achieve the optimal quadratic decay, which is asymptotically faster than some other methods. Finally, we present simulation results supporting our findings. Our simulations have been implemented in MATLAB and the resulting code is available in the supplementary material, which can be found on the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2019.2939530>.

Index Terms—Multi-camera imaging, multiple-view geometry, triangulation

1 INTRODUCTION

CAMERAS are finite; yet, we commonly assume Gaussian noise models with infinite tails. Clearly, this disparity, which appears across all of science, is offset by the approximation accuracy and mathematical convenience of the Gaussian distribution. However, by replacing cost functions based on the ℓ_2 -norm with their ℓ_∞ -counterparts, the computer vision community has begun to investigate the feasibility of bounded noise models [3], [4], [5], [6].

In this paper, we continue this trend by formalizing some of the benefits of bounded noise models in triangulation problems; i.e., problems which aim to estimate the three-dimensional (3-D) positions of feature points from their two-dimensional (2-D) projections.

For example, consider triangulating from a set of calibrated cameras. In the noise-free, infinite-resolution case, the exact location of the world point can be reconstructed by intersecting rays originating from each camera. However, in practice, various sources of uncertainty mean that the rays do not necessarily intersect. Error minimization techniques are thus employed with the goal of finding the most probable world point [4], [7]. The degree to which we achieve this goal depends on two factors of the cost function: its correct modelling of the uncertainty and how accurately we can find its global minimum.

The simplest approach is to construct an over-complete set of linear equations, each corresponding to one of the rays, and take the pseudo-inverse to find the least-squares solution. Although we find the global minimum, the cost function is not particularly meaningful and it is unlikely that it provides the best model of the underlying uncertainty. However, this technique performs reasonably well,

particularly if the coordinates are correctly normalized [7], [8], and it is a good choice when time complexity is the principal concern.

Alternatively, we can attempt to minimize the ℓ_2 -norm of the reprojection error between the prospective 3-D points and the known image locations, which results in the maximum-likelihood estimator if we assume that the projected points are subjected to i.i.d. zero-mean Gaussian noise in the image plane. Although this assumption is very likely a much better approximation of the underlying uncertainty, the resulting cost function is non-convex and extremely difficult to solve.

Although for a small number of views the global minimum can be found by polynomial root finding [7], [9], [10], [11], [12], [13], [14], the degree of the resulting polynomial grows quadratically with the number of views [11] and thus this is, in general, not practical. Therefore, often, we have to resort to iterative approaches that only converge to a local minimum, or more time-consuming branch and bound techniques [15]. On a positive note, it is possible to verify whether a solution is globally optimal in the ℓ_2 -sense [16].

Inspired by these difficulties, the ℓ_∞ -norm has recently been considered as a measure of the reprojection error [3], [4], [17]. Although this may not correspond to the best model of the underlying uncertainty, the resulting cost is a quasi-convex function and efficient algorithms can find the global optimum [5]. However, since the ℓ_∞ -norm implicitly assumes a bounded noise model, care must be taken to limit the potential catastrophic effect of outliers [18], [19].

In this paper, we analyze the performance of multi-camera systems and triangulation algorithms under the assumption of bounded noise and pixelization. We provide two main contributions. First, we prove that, under certain conditions, the highest achievable point localization accuracy of a multi-camera system is quadratically related to the number of cameras in the system. Second, we introduce the notion of consistency and show that consistent triangulation algorithms achieve the optimal quadratic decay rate.

In the stereo case, there have been a number of excellent studies thoroughly analyzing triangulation accuracy with relation to multiple parameters [20], [21]. Furthermore, the error of depth estimation in linear camera arrays, has been analyzed [22].

However, to the best of our knowledge, there is no work analyzing arbitrary camera setups and certainly no work deriving fundamental scaling laws for the accuracy of triangulation, with respect to the number of cameras. Given the rapid increase in popularity of multi-camera systems, we hope this analysis provides a significant contribution.

Our work is inspired by results derived from frame quantization. In particular, similar error decay rates have been derived for signal reconstruction from over-complete quantized projections [23], [24], [25]. In imaging terms, this corresponds to circular arrays of 2-D pixelised orthographic cameras. Furthermore, these results have recently been generalized to uniform bounded noise and any consistent estimate [26], [27]. We extend these results to more general arrays of 3-D cameras using central projections, which is more typical in the context of imaging.

In what follows, we give a very brief overview of the problem setup and formally define the triangulation problem. After introducing major state-of-the-art techniques, we present our main results on the error decay rate of consistent triangulation algorithms and the best possible performance of multi-camera systems. Finally, simulations are provided to support our findings.

• The authors are with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland. E-mail: {adam.scholefield, alireza.ghasemi, martin.vetterli}@epfl.ch.

Manuscript received 2 Mar. 2018; revised 5 June 2019; accepted 27 Aug. 2019. Date of publication 13 Sept. 2019; date of current version 4 Aug. 2020.

(Corresponding author: Adam Scholefield.)

Recommended for acceptance by J.-M. Frahm.

Digital Object Identifier no. 10.1109/TPAMI.2019.2939530

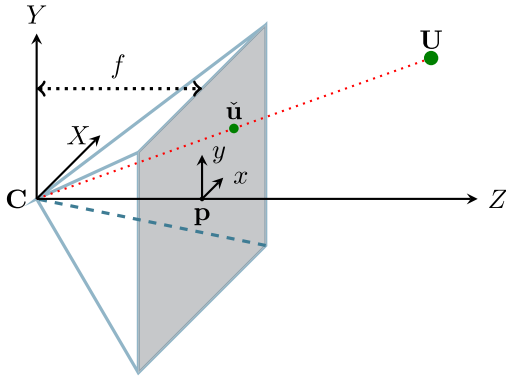


Fig. 1. An example of central projection in a pinhole camera.

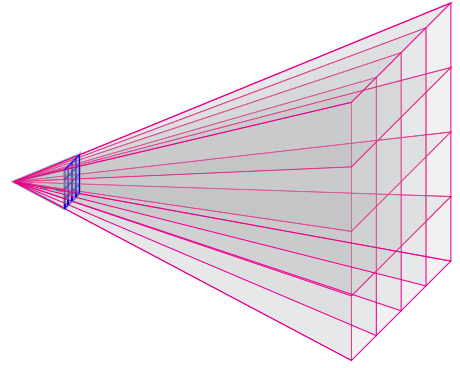


Fig. 2. Pixelization in a digital camera.

2 BACKGROUND

2.1 Pinhole Camera Model

As is typical, we assume the pinhole camera model, which we will now briefly summarize. For a more thorough introduction, we refer the reader to [28].

As depicted in Fig. 1, a pinhole camera projects a point in 3-D space, called the world point, to a point on the camera's 2-D image plane via a central projection. In homogeneous coordinates, pinhole projection can be expressed as a linear matrix multiplication:

$$\tilde{\mathbf{u}}_h = \mathbf{P}\mathbf{U}_h, \quad (1)$$

where $\tilde{\mathbf{u}}_h$ and \mathbf{U}_h are the homogeneous representation of the projected point and world point, respectively. The 3×4 matrix, \mathbf{P} , is the camera matrix, which can be decomposed as

$$\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I} - \mathbf{C}]. \quad (2)$$

Here \mathbf{C} denotes the camera centre, \mathbf{R} the camera orientation, and \mathbf{I} the column-wise concatenation operator. Together, the camera centre and the camera orientation form the extrinsic, or pose parameters. Finally, the matrix \mathbf{K} contains the intrinsic camera parameters: namely, the focal length f and the coordinates of the principal point $p = (c_x, c_y)$.

In this paper, it will generally be more convenient to work with the Cartesian coordinates of euclidean geometry. In this case, we will write the non-linear pinhole projection as

$$\tilde{\mathbf{u}} = \mathcal{P}(\mathbf{U}) = \frac{1}{\tilde{\mathbf{u}}_h[3]} \begin{bmatrix} \tilde{\mathbf{u}}_h[1] \\ \tilde{\mathbf{u}}_h[2] \end{bmatrix}, \quad (3)$$

where $\tilde{\mathbf{u}}$ and \mathbf{U} are the Cartesian coordinate representations of the projected point and world point, respectively, and $\tilde{\mathbf{u}}_h$ is the homogeneous representation of the projected point.

2.2 Sources of Uncertainty

Due to various sources of uncertainty, the true image location $\tilde{\mathbf{u}}$ is perturbed to yield the measurement \mathbf{u} . The error term incorporates both deterministic (e.g., pixelization) and random (e.g., measurement noise) perturbation.

2.2.1 Pixelization

Even when we are in a hypothetical noiseless scenario, we still have to deal with the uncertainty caused by the finite resolution of the camera sensors.

This source of uncertainty, which we call pixelization, is deterministic and, when projected back to the world space, leads to semi-infinite regions in the world space instead of rays. These regions, originating from the boundaries of the pixels, partition the world space into a finite number of regions. Each region consists of all world

points whose projections map to the same pixel. Fig. 2 depicts a simple example of such a partitioning for a camera with sixteen pixels.

When multiple cameras view the same region of interest, these regions intersect producing a finite number of regions, each corresponding to a particular combination of pixels in the cameras. Clearly, smaller regions lead to a smaller uncertainty.

2.2.2 Non-Deterministic Sources of Uncertainty

In reality, pixelization is not the only source of uncertainty, with additionally noises arising from image sensor noise, as well as the error of corner localization algorithms.

These perturbation sources, combined with pixelization, eventually lead to ambiguity in measuring the exact location of an image point. This can be modelled as an additive noise, yielding

$$\mathbf{u} = \tilde{\mathbf{u}} + \boldsymbol{\epsilon} = \mathcal{P}(\mathbf{U}) + \boldsymbol{\epsilon}. \quad (4)$$

2.2.3 Bounded Noise Models

In this paper, we will be interested in bounded noise models:

$$\|\boldsymbol{\epsilon}\|_q \leq \delta, \quad (5)$$

where q specifies the shape of the bounded noise and δ is referred to as the bandwidth of the noise.

The main advantage of bounded noise is that it allows us to completely dismiss regions of the solution space as impossible and, as we will show, the size of the remaining feasible region decays in such a way that the squared reconstruction error decays quadratically with the number of measurements.

One may question the applicability of bounded noise. Clearly, in the presence of outliers, additional work must be done to prevent these methods breaking down. However, if outlier techniques are applied, bounded approaches can be useful. It would be interesting to fully investigate this comparison, for practical problems; however, this is beyond the scope of this theoretical study.

2.3 The Triangulation Problem

This paper focuses on triangulation—a fundamental problem in multiple-view geometry, which, as well as being interesting in its own right, provides a basic building block for many higher-level computer vision tasks, such as visual metrology, Simultaneous Localization And Mapping (SLAM), and Structure from Motion (SfM).

The aim is to recover the location of an unknown 3-D point \mathbf{U} from its projections in M calibrated cameras; i.e., the camera matrices \mathbf{P}_1 through \mathbf{P}_M are known and we estimate \mathbf{U} from the measured projections $\mathbf{u}_1, \dots, \mathbf{u}_M$. To be able to prove things about triangulation, we formally define it as follows.

Definition 1. A triangulation problem takes as input

$$T = \{(\mathbf{u}_i, \mathbf{P}_i) \mid 1 \leq i \leq M\}, \quad (6)$$

and estimates the underlying unknown 3-D world point \mathbf{U} as follows:

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \sum_{i=1}^M \|\mathbf{u}_i - \mathcal{P}_i(\mathbf{U})\|_{p'}^p. \quad (7)$$

Here, $\mathcal{P}_i(\cdot)$ denotes the projection operator corresponding to the camera matrix \mathbf{P}_i and the $(\mathbf{u}_i, \mathbf{P}_i)$ pairs denote the camera matrices of the M cameras along with the projections of the unknown 3-D world point \mathbf{U} on their image planes.

In the previous definition, p' and p , are known as the image-space and residual-space norms, respectively. When we talk about algorithms that minimize the $(\ell_{p'}, \ell_p)$ -norm of the reprojection error, for some particular p' and p , we are referring to the image-space and residual-space norms in this order.

When $p = \infty$, we assume that (7) becomes

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \max_{i=1..M} \|\mathbf{u}_i - \mathcal{P}_i(\mathbf{U})\|_{p'}. \quad (8)$$

2.4 Equivalence with Camera Localization

As an aside, we briefly mention the connection between triangulation and a restricted version of camera localization. In particular, by replacing the M cameras and single feature point of the triangulation problem with M feature points and a single camera, one can easily show that the triangulation problem, as just defined, is mathematically equivalent to localizing a single camera from the image points of M feature points at known locations. The catch is that the equivalence is only valid if one assumes that the orientation of the camera is known.

Of course, in most practical problems, the camera orientation is not known and heuristics must be applied if one wishes to adapt triangulation techniques to camera localization. Therefore, for clarity, we restrict our analysis to triangulation and present the details of this equivalence in the supplementary material, available online. However, we emphasize that the scaling laws we derive also apply to this restricted version of camera localization. Deriving scaling laws for the full camera localization problem is an interesting open research problem.

3 RECONSTRUCTION ALGORITHMS

In this section, we briefly review the main techniques used to solve triangulation and other geometric reconstruction problems.

3.1 Linear Triangulation

The simplest approach to triangulation is to construct a linear system of equations. Let \mathbf{p}_l^T be the l -th row of a camera matrix \mathbf{P} . Then,

$$\hat{\mathbf{u}}[1] = \frac{\hat{\mathbf{u}}_h[1]}{\hat{\mathbf{u}}_h[3]} = \frac{\mathbf{p}_1^T \mathbf{U}_h}{\mathbf{p}_3^T \mathbf{U}_h}, \quad (9)$$

and similarly for $\hat{\mathbf{u}}[2]$. Therefore, for a single camera, we have

$$\begin{bmatrix} \hat{\mathbf{u}}[1] \mathbf{p}_3^T - \mathbf{p}_1^T \\ \hat{\mathbf{u}}[2] \mathbf{p}_3^T - \mathbf{p}_2^T \end{bmatrix} \mathbf{U}_h = \mathbf{0}. \quad (10)$$

For M cameras, we can stack the $2M$ equations into a matrix $\mathbf{A} \in \mathbb{R}^{2M \times 4}$, with $\mathbf{A} \mathbf{U}_h = \mathbf{0}$.

This equation can be solved efficiently with standard techniques, such as the Singular Value Decomposition (SVD). However, due to the conversion to homogeneous coordinates, this does not minimize the desired cost; i.e., a norm of the residual vector.

The advantages of linear triangulation are its speed and simplicity and it performs particularly well when the cameras are at almost the same depth to the point of interest. Its robustness can also be further improved by normalizing the focal length and other metric distances in the problem instance [7], [8].

3.2 Reprojection Error Minimization

If increased computational resources are available, one can directly attempt to minimize (7) for different values of p' and p .

The most common choice is the (ℓ_2, ℓ_2) -norm, but, as explained in the introduction, the resulting cost function is non-convex and often difficult to solve exactly. Therefore, often, linear triangulation is used to initialize a gradient descent approach thus accepting convergence to a local minimum [16], [28].

Alternatively, more computationally intensive branch and bound techniques can be used, which guarantee convergence to the global minimum. In particular, in [29], the authors present a branch and bound technique for minimizing the (ℓ_2, ℓ_2) , (ℓ_2, ℓ_1) and (ℓ_1, ℓ_1) -norms.

The ℓ_∞ -norm corresponds to a bounded noise assumption. In the case of the (ℓ_2, ℓ_∞) -norm, the image points are bounded to circles on the image plane, which back project as cones in the 3-D world space. Consequently, Second Order Cone Programming (SOCP) can be applied [4].

As with all approaches based on bounded noise, ℓ_∞ -based methods suffer from being extremely sensitive to even a single outlier. Therefore, it is critical that these techniques are either combined with standard outlier removal techniques or, as has been recently proposed, relaxations applied [18], [19], [30].

4 CONSISTENT RECONSTRUCTION AND THE ACCURACY OF MULTI-CAMERA SYSTEMS

In this section, we present two results concerning the triangulation accuracy of multi-camera systems as more cameras are added to the system.

In the first, we prove a lower bound for the average reconstruction error of any triangulation algorithm using any multi-camera system, over a region of interest. This lower bound decreases quadratically as more cameras are added to the system.

Next, we introduce the concepts of consistency and consistent triangulation and then prove that the reconstruction error of a consistent triangulation algorithm is upper bounded by a term that decreases quadratically as more cameras are added to the system. Therefore, consistent triangulation algorithms achieve the optimal error decay rate.

This is not necessarily the case for other algorithms: linear triangulation, for example, can be shown to yield a linear error decay rate with respect to the number of cameras.

4.1 Lower Bound for the Accuracy of a Multi-Camera System

We would like to lower-bound the best possible reconstruction error achievable by a multi-camera system, using any possible triangulation algorithm. To do this, we first formally define what we mean by a triangulation algorithm.

Definition 2. A triangulation algorithm is any mapping from $T = \{(\mathbf{u}_i, \mathbf{P}_i) : 1 \leq i \leq M\}$ to $\hat{\mathbf{U}} \in \mathbb{R}^3$.

Since we are seeking a lower-bound, it makes sense to limit the uncertainties in the system. Therefore, in the following theorem, we assume that pixelization is the only source or uncertainty; however, note that the size of pixels is arbitrary and thus this is a mild assumption that is interesting even if an image point can be localized with subpixel precision.

Theorem 1. Consider a multi-camera system of M cameras, each with an $N \times N$ pixel image sensor and define a fixed region of interest, \mathcal{R} , with a finite non-zero volume.

If we assume that the only source of uncertainty is pixelization, the expected ℓ_2 reconstruction error of any triangulation algorithm is lower-bounded by a term that is inverse-quadratically dependent on

the number of cameras; i.e.,

$$\mathbb{E}\left(\|\hat{\mathbf{U}} - \mathbf{U}\|_2^2\right) = \Omega\left(\frac{1}{M^2}\right), \quad (11)$$

where $\mathbf{U} \in \mathcal{R}$ is any point in the region of interest, and $\hat{\mathbf{U}}$ is the result of reconstructing \mathbf{U} , from its images in the multi-camera system, using any triangulation algorithm. Here, the expectation is taken over the location of the point \mathbf{U} in the region of interest.

Proof. For the proof, please refer to the supplementary material, available online. \square

The above theorem states that, under certain assumptions, no triangulation algorithm can do better than a quadratic decay with respect to the number of cameras, regardless of the camera setup. In what follows, we show that, if the camera array is properly constructed, the expected reconstruction error of certain triangulation algorithms can be upper-bound by a term that decays quadratically with the number of cameras. Therefore, in doing so, we show that these triangulation algorithms reach the best possible decay rate. So which triangulation algorithms achieve this optimal decay? It turns out that the key property is *consistency*.

4.2 Consistency and Consistent Reconstruction

A key advantage of bounded noise models is that, given a noisy image point, we can restrict the true image point to a finite 2-D region, which we call an *image-space q -consistency region*:

$$I_{\mathbf{u},\delta,q} = \{\mathbf{u} \in \mathbb{R}^2 : \|\mathbf{u} - \mathbf{u}\|_q \leq \delta\}. \quad (12)$$

The shape of these regions depends on the type of bounded noise: a circle when $q = 2$, a diamond when $q = 1$ and a square when $q = \infty$. Note that, we assume that $q \geq 1$ so the norm is properly defined. In this case, the image-space q -consistency region is convex.

If we back-project an image-space q -consistency region into the world space, we obtain a convex 3-D region, which we call a *world-space q -consistency region*:

$$W_{\mathbf{u},\delta,\mathbf{P},q} = \{\mathbf{U} \in \mathbb{R}^3 : \|\mathbf{u} - \mathcal{P}(\mathbf{U})\|_q \leq \delta\}. \quad (13)$$

Again, depending on the type of bounded noise, these 3-D regions can be either a cone ($q = 2$), a diamond-based pyramid ($q = 1$), or a square-based pyramid ($q = \infty$).

We know that the true 3-D point must lie in the intersection of all world-space q -consistency regions:

$$V_{\mathbf{u},\delta,\mathfrak{P},q} = \{\mathbf{U} \in \mathbb{R}^3 : \bigwedge_{i=1}^M \|\mathbf{u}_i - \mathcal{P}_i(\mathbf{U})\|_q \leq \delta\}, \quad (14)$$

where $\mathfrak{P} = \{\mathbf{P}_i : i \in [1, M]\}$. For a particular triangulation problem, we call this region the *q -consistent region* and any estimate that lies within it a *q -consistent estimate*. In addition, if a triangulation algorithm always returns a *q -consistent estimate*, we call it a *q -consistent triangulation algorithm*. This is stated more formally in the following definition.

Definition 3. A triangulation algorithm is q -consistent, over the region of interest \mathcal{R} , if

$$\mathbf{u}_i \in I_{\mathcal{P}_i(\mathbf{U}),\delta,q} \quad \forall i \in [1, M] \quad \Rightarrow \quad \hat{\mathbf{U}} \in V_{\mathbf{u},\delta,\mathfrak{P},q}, \quad (15)$$

for any $\mathbf{U} \in \mathcal{R}$ and valid projection matrices $\mathfrak{P} = \{\mathbf{P}_i : 1 \leq i \leq M\}$. Note that a q -consistent triangulation algorithm returns no estimate when $V_{\mathbf{u},\delta,\mathfrak{P}} = \emptyset$.

4.3 Finding Consistent Estimates

As stated in the following proposition, (ℓ_q, ℓ_∞) -based triangulation is q -consistent.

Proposition 1. Consider a multi-camera system viewing a point and assume that the image points are subjected to ℓ_q -norm bounded noise:

$$\|\mathbf{u}_i - \mathcal{P}_i(\mathbf{U})\|_q \leq \delta \quad \text{for } i = 1 \dots M.$$

Then, any algorithm that minimizes the (ℓ_q, ℓ_∞) -norm of the reprojection error is a q -consistent triangulation algorithm.

Proof. For the proof, please refer to the supplementary material, available online. \square

For ℓ_2 -bounded noise, the 2-consistent regions are cones and the SOCP technique outlined by Kahl et.al. [4] returns a 2-consistent estimate.

For ℓ_∞ -bounded noise, the following simple linear program (LP) can be used to find a ∞ -consistent estimate. Recall that, for the linear triangulation algorithm, we used \mathbf{p}_l^T to denote the l -th row of a camera matrix \mathbf{P} . To avoid homogeneous coordinates, let's separate the first three elements from the last: $\mathbf{p}_l^T = [\bar{\mathbf{p}}_l^T \quad p_{l4}]$. Then,

$$\check{\mathbf{u}}[1] = \frac{\check{\mathbf{u}}_h[1]}{\check{\mathbf{u}}_h[3]} = \frac{\bar{\mathbf{p}}_1^T \mathbf{U} + p_{14}}{\bar{\mathbf{p}}_3^T \mathbf{U} + p_{34}}, \quad (16)$$

and similarly for $\check{\mathbf{u}}[2]$. For ℓ_∞ -bounded noise, with bandwidth δ ,

$$\mathbf{u}[i] - \delta \leq \check{\mathbf{u}}[i] \leq \mathbf{u}[i] + \delta, \quad i = 1, 2.$$

Therefore, for a single camera, we have

$$\begin{bmatrix} (\mathbf{u}[1] - \delta)\bar{\mathbf{p}}_3^T - \bar{\mathbf{p}}_1^T \\ (\mathbf{u}[2] - \delta)\bar{\mathbf{p}}_3^T - \bar{\mathbf{p}}_2^T \\ \bar{\mathbf{p}}_1^T - (\mathbf{u}[1] + \delta)\bar{\mathbf{p}}_3^T \\ \bar{\mathbf{p}}_2^T - (\mathbf{u}[2] + \delta)\bar{\mathbf{p}}_3^T \end{bmatrix} \mathbf{U} \leq \begin{bmatrix} p_{14} - (\mathbf{u}[1] - \delta)p_{34} \\ p_{24} - (\mathbf{u}[2] - \delta)p_{34} \\ (\mathbf{u}[1] + \delta)p_{34} - p_{14} \\ (\mathbf{u}[2] + \delta)p_{34} - p_{24} \end{bmatrix}.$$

Note that, here, we have assumed that the point is in front of the camera so that $\bar{\mathbf{p}}_3^T \mathbf{U} + p_{34} > 0$.

For M cameras, we can stack the $4M$ inequalities producing a matrix $\mathbf{A} \in \mathbb{R}^{4M \times 3}$ and vector $\mathbf{b} \in \mathbb{R}^{4M}$, such that $\mathbf{A}\mathbf{U} \leq \mathbf{b}$. Any point satisfying these constraints is ∞ -consistent and thus the following LP will return a ∞ -consistent estimate:

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \mathbf{c}^T \mathbf{U}, \quad \text{s.t. } \mathbf{A}\mathbf{U} \leq \mathbf{b}. \quad (17)$$

Here, the vector $\mathbf{c} \in \mathbb{R}^3$ dictates which point in the ∞ -consistent region is the optimum. As we show in the next section, for many cameras, any ∞ -consistent estimate is good and thus in our simulations we use a standard LP solver with $\mathbf{c} = \mathbf{0}$.

As an aside, we note that more complex LPs can be used to select a more desirable point in the ∞ -consistent region. As just stated, this will only be beneficial for small M , since asymptotically they will achieve the same performance.

For example, one can take the previous linear program and remove all redundant constraints. Assume this produces \bar{M} equivalent non-redundant constraints: $\bar{\mathbf{A}}\mathbf{U} \leq \bar{\mathbf{b}}$. Then, a LP that finds the point in the ∞ -consistent region that minimizes the average distance to the \bar{M} planes at the limit of the constraints can be designed as follows. Let $\bar{\mathbf{a}}_l^T$ be the l -th row of $\bar{\mathbf{A}}$. Then, the minimum distance from any point $\mathbf{V} \in \mathbb{R}^3$ to the l -th plane, is

$$d_l(\mathbf{V}) := \frac{\bar{\mathbf{a}}_l^T \mathbf{V} - \bar{b}_l}{\|\bar{\mathbf{a}}_l\|_2}. \quad (18)$$

We thus wish to solve

$$\arg \min_{\mathbf{U}} \sum_l |d_l(\mathbf{U})|, \quad \text{s.t. } \bar{\mathbf{A}}\mathbf{U} \leq \bar{\mathbf{b}}. \quad (19)$$

It is a standard exercise in linear programming to convert this problem into standard form using \bar{M} auxiliary variables, one for each distance.

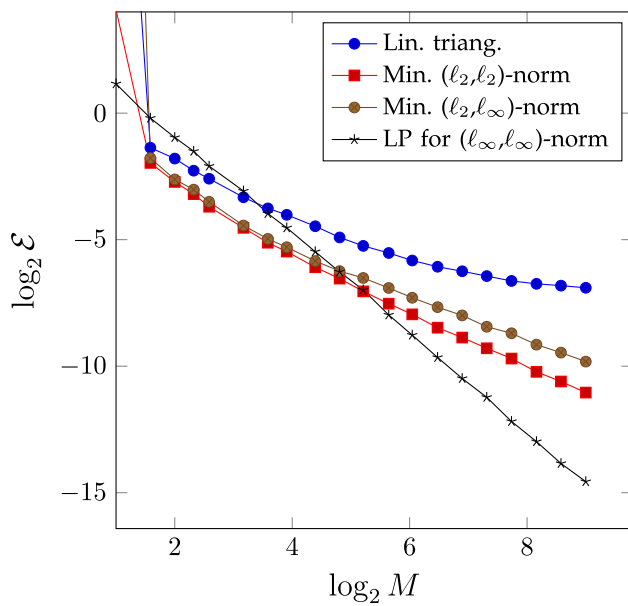
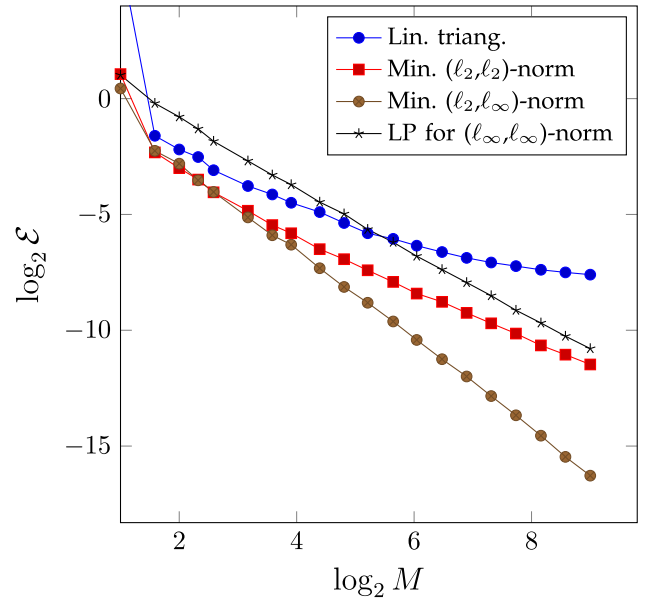

 (a) Additive ℓ_∞ -norm bounded noise.

 (b) Additive ℓ_2 -norm bounded noise.

 Fig. 3. Verification of Conjecture 1. Expected squared error (\mathcal{E}) as a function of the number of cameras (M).

4.4 Error Decay in Consistent Reconstruction

We will shortly present a theorem stating that, for circular camera arrays, the expected reconstruction error of q -consistent triangulation algorithms is upper bounded by a term that decays quadratically with the number of cameras. However, since simulations suggest that the result holds in many more cases, including linear camera arrays and even quite general random setups, we present the following more general conjecture, which is numerically tested in the following section.

Conjecture 1. Define the region of interest, \mathcal{R} , to be a sphere of finite radius r and place a point anywhere in this region. Place M cameras inside a larger finite radius sphere, with the same centre as \mathcal{R} , i.i.d. uniformly at random such that they all see the whole region of interest.

Furthermore, assume that the images of the world point in the cameras are perturbed with uniform bounded noise; i.e., for the world point \mathbf{U} , the image \mathbf{u}_i in the i -th camera is computed as

$$\mathbf{u}_i = \mathcal{P}_i(\mathbf{U}) + \epsilon_i, \quad (20)$$

where ϵ_i is zero-mean uniform bounded random satisfying $\|\epsilon_i\|_q \leq \delta$.

In this situation, the expected ℓ_2 reconstruction error of any q -consistent triangulation algorithm is upper-bounded by a term which decreases quadratically with the number of cameras; i.e.,

$$\mathbb{E}(\|\hat{\mathbf{U}} - \mathbf{U}\|_2^2) = \mathcal{O}\left(\frac{1}{M^2}\right), \quad (21)$$

where $\mathbf{U} \in \mathcal{R}$ is any point in the region of interest, and $\hat{\mathbf{U}}$ is the result of reconstructing \mathbf{U} , from its images in the multi-camera system, using a q -consistent triangulation algorithm. Here, the expectation is taken over both the noise vector ϵ and the camera locations.

Now, the theorem for circular camera arrays. Once again, simulation results are presented in the following section to support this theorem.

Theorem 2. Place M cameras in a plane, i.i.d. uniformly at random on a finite radius circle oriented towards the centre of the circle. Define the region of interest, \mathcal{R} , to be the intersection of the field of view of all cameras as $M \rightarrow \infty$ and place a point anywhere in this region.

Furthermore, assume that the images of the world point in the cameras are perturbed with ℓ_∞ uniform bounded noise; i.e., for the world point \mathbf{U} , the image \mathbf{u}_i in the i -th camera is

$$\mathbf{u}_i = \mathcal{P}_i(\mathbf{U}) + \epsilon_i, \quad (22)$$

where ϵ_i is zero-mean uniform bounded random satisfying $\|\epsilon_i\|_\infty \leq \delta$.

In this situation, the expected ℓ_2 reconstruction error of any ∞ -consistent triangulation algorithm is upper-bounded by a term which decreases quadratically with the number of cameras; i.e.,

$$\mathbb{E}(\|\hat{\mathbf{U}} - \mathbf{U}\|_2^2) = \mathcal{O}\left(\frac{1}{M^2}\right), \quad (23)$$

where $\mathbf{U} \in \mathcal{R}$ is any point in the region of interest, and $\hat{\mathbf{U}}$ is the result of reconstructing \mathbf{U} , from its images in the multi-camera system, using a ∞ -consistent triangulation algorithm. Here, the expectation is taken over both the noise and the camera locations.

Proof. The proof makes use of [27, Corollary 6.2] and appears in the supplementary material, available online. \square

5 SIMULATIONS

We now present simulations to verify our theoretical results. To approximate the expected reconstruction error, we take the average of many realizations. This means that any algorithm we use is run many times and thus has to be extremely robust. Unfortunately, this prevented us from using the branch and bound technique proposed in [29], since we were unable to prevent their implementation from crashing for certain realizations. Therefore, for the (ℓ_2, ℓ_2) -norm we used the non-linear approach from Hartley and Zisserman [28]. The technique is based on Newton iterations and we initialized it with the solution of linear triangulation. For the linear triangulation, we used the normalized implementation by the same authors [28]. For the (ℓ_2, ℓ_∞) -norm, we used the SOCP implementation provided by Kahl [4] and, finally, for the $(\ell_\infty, \ell_\infty)$ -norm minimization, we used the linear program defined in (17) with $c = 0$. Of course, this just returns a ∞ -consistent estimate and does not fully minimize the norm.

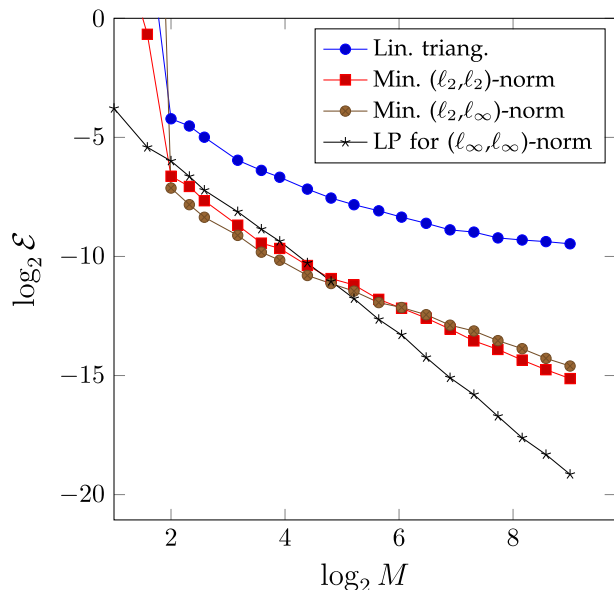


Fig. 4. Verification of Theorem 2. Expected squared error (\mathcal{E}) as a function of the number of cameras (M).

5.1 Numerical Simulation of Conjecture 1

We simulated the setup explained in Conjecture 1. In order to generate cameras uniformly at random that see the whole region of interest, we use rejection sampling; i.e., we repeatedly generate a camera centre and rotation matrix uniformly at random and reject ones that do not see the whole region of interest. To generate rotation matrices uniformly at random we use the technique outlined in [32].

The results are shown in Fig. 3, for ℓ_∞ -norm and ℓ_2 -norm bounded noise. As expected, q -consistent algorithms, where q matches the norm of the noise, have a quadratic decay; i.e., the LP has a quadratic decay for ℓ_∞ -norm bounded noise and the minimum (ℓ_2, ℓ_∞) -norm has a quadratic decay for ℓ_2 -norm bounded noise.

Non-consistent techniques, such as the (ℓ_2, ℓ_2) -norm minimization, can perform well for small M , but, as M increases, fail to reach the quadratic decay.

5.2 Numerical Simulation of Theorem 2

Finally, we present a simulation to experimentally test Theorem 2. The algorithms are the same as for the simulation of Conjecture 1 and the results are shown in Fig. 4. As can be seen in the figure, all algorithms behave as expected.

6 CONCLUSION

We presented an analysis of the accuracy of multi-camera systems and their error decay rate with respect to the number of cameras. In doing so, we derived fundamental scaling laws stating that, under certain conditions, the accuracy of a multi-camera imaging system in triangulating 3-D points increases quadratically with respect to the number of cameras.

We also analyzed the performance of state-of-the-art algorithms with respect to their error decay rate. To do this, we introduced the notion of q -consistency and showed that q -consistent reconstruction algorithms achieve the optimal quadratic error decay. Furthermore, we showed that (ℓ_q, ℓ_∞) -norm based minimization is q -consistent and presented two simple linear programs that are ∞ -consistent.

ACKNOWLEDGMENTS

The authors would like to thank Richard Hartley, Fredrik Kahl, and Pascal Frossard for their advice and fruitful discussions, which have greatly improved the manuscript. This work was in part presented in [1] and [2]. This work was supported by the Commission for Technology and Innovation (CTI) project no. 14842.1 PFES-ES

and ERC Advanced Grant—Support for Frontier Research—SPAR-SAM Nr: 247006. A. Ghasemi was additionally supported by a Qualcomm Innovation Fellowship.

REFERENCES

- [1] A. Ghasemi, A. Scholefield, and M. Vetterli, "On the accuracy of point localisation in a circular camera-array," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 981–985.
- [2] A. Ghasemi, A. Scholefield, and M. Vetterli, "SHAPE: Linear-time camera pose estimation with quadratic error-decay," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2016, pp. 1776–1780.
- [3] R. Hartley and F. Kahl, "Optimal algorithms in multiview geometry," in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 13–34.
- [4] F. Kahl and R. Hartley, "Multiple-view geometry under the ℓ_∞ -norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1603–1617, Sep. 2008.
- [5] S. Donné, B. Goossens, and W. Philips, "Point triangulation through polyhedron collapse using the ℓ_∞ norm," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 792–800, 2015.
- [6] C. Freundlich, M. Zavlanos, and P. Mordohai, "Exact bias correction and covariance estimation for stereo vision," in *Proc. IEEE Comput. Vis. Pattern Recognit. Conf.*, 2015, vol. 1, pp. 3296–3304.
- [7] R. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.
- [8] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, Jun. 1997.
- [9] K. Kanatani and H. Niitsuma, "Optimal two-view planar scene triangulation," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 242–253.
- [10] K. Kanatani, Y. Sugaya, and H. Niitsuma, "Triangulation from two views revisited: Hartley-Sturm versus optimal correction," in *Proc. British Mach. Vis. Conf.*, 2008, pp. 173–182.
- [11] H. Stewenius, F. Schaffalitzky, and D. Nister, "How hard is 3-view triangulation really?," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 686–693.
- [12] J. Hedberg, A. Robinson, and M. Felsberg, "Robust three-view triangulation done fast," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 152–157.
- [13] M. Byröd, K. Josephson, and K. Åström, "Fast optimal three view triangulation," in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 549–559.
- [14] K. Nordberg, "Efficient three-view triangulation based on 3-D optimization," in *Proc. British Mach. Vis. Conf.*, 2008, pp. 1–10.
- [15] S. Agarwal, M. K. Chandraker, F. Kahl, D. Kriegman, and S. Belongie, "Practical global optimization for multiview geometry," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 592–605.
- [16] R. Hartley, F. Kahl, C. Olsson, and Y. Seo, "Verifying global minima for ℓ_2 minimization problems in multiple view geometry," *Int. J. Comput. Vis.*, vol. 101, no. 2, pp. 288–304, 2013.
- [17] H. D. Mittelmann, "An independent benchmarking of SDP and SOCP solvers," *Math. Program.*, vol. 95, no. 2, pp. 407–430, 2003.
- [18] H. Li, "A practical algorithm for ℓ_∞ triangulation with outliers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
- [19] B. Micsik and R. Pflugfelder, "Localizing non-overlapping surveillance cameras under the ℓ_∞ norm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2895–2901.
- [20] S. Blostein and T. S. Huang, "Error analysis in stereo determination of 3-D point positions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 6, pp. 752–765, Jun. 1987.
- [21] J. Rodriguez and J. Aggarwal, "Stochastic analysis of stereo quantization error," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 467–470, May 1990.
- [22] R. Raynor and K. Walli, "Plenoptic camera range finding," in *Proc. AIPR Workshops*, 2013, vol. 1, pp. 1–11.
- [23] V. K. Goyal, M. Vetterli, and N. T. Nguyen, "Quantized overcomplete expansions in \mathbb{R}^n : analysis, synthesis, and algorithms," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 16–31, Jan. 1998.
- [24] Z. Cvetkovic, "Source coding with quantized redundant expansions: Accuracy and reconstruction," in *Proc. Data Compression Conf.*, 1999, vol. 1, pp. 344–353.
- [25] B. Beferull-Lozano and A. Ortega, "Efficient quantization for overcomplete expansions in \mathbb{R}^n ," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 129–150, Jan. 2003.
- [26] S. Rangan and V. K. Goyal, "Recursive consistent estimation with bounded noise," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 457–464, Jan. 2001.
- [27] A. M. Powell and J. T. Whitehouse, "Error bounds for consistent reconstruction: Random polytopes and coverage processes," *Found. Comput. Math.*, vol. 16, no. 2, pp. 395–423, 2016.
- [28] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [29] F. Kahl, S. Agarwal, M. Chandraker, D. Kriegman, and S. Belongie, "Practical global optimization for multiview geometry," *Int. J. Comput. Vis.*, vol. 79, no. 3, pp. 271–284, 2008.
- [30] Y. Seo, H. Lee, and S. Lee, "Outlier removal by convex optimization for ℓ_∞ approaches," in *Proc. Pacific-Rim Symp. Image Video Technol.*, 2009, pp. 203–214.
- [31] J. E. Goodman and R. Pollack, "Upper bounds for configurations and polytopes in \mathbb{R}^d ," *Discrete & Comput. Geometry*, vol. 1, no. 3, pp. 219–227, 1986.
- [32] J. Arvo, "Fast random rotation matrices," in *Graphics Gems III*, D. Kirk, Ed. San Diego, CA, USA: Academic Press Professional, Inc., 1992, p. 117–120.