Bound and Conquer: Improving Triangulation by Enforcing Consistency

Adam Scholefield, *Member, IEEE*, Alireza Ghasemi, *Student Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

SUPPLEMENTARY MATERIAL

Abstract—In this supplementary material, we give the proofs for the two theorems and one proposition of the paper.

1 Proof of Theorem 1

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 pixelisation, the expected reconstruction error of any triangulation algorithm is lower-bounded by a term that is inverse-quadratically dependent on the number of cameras; i.e.,

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

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. A single $N \times N$ pixel camera partitions the world space into N^2 regions. Combined with the partitions of other cameras, this leads to a finite number of partitions. Therefore, when a multi-camera system views the region of interest, it splits it into a finite number of partitions. Let $\mathcal{P}$ be the set containing the resulting partitions of $\mathcal{R}$.

We can now consider the the expected reconstruction error split over these partitions:

$$\begin{aligned}\mathbb{E}\left(\left\|\hat{\mathbf{U}}-\mathbf{U}\right\|^2\right) &= \frac{1}{\mathcal{V}(\mathcal{R})}\iiint_{\mathcal{R}}\left\|\hat{\mathbf{U}}-\mathbf{U}\right\|_2^2 d\mathbf{U}\\ &= \frac{1}{\mathcal{V}(\mathcal{R})}\sum_{\mathcal{C}\in\mathcal{P}}\iiint_{\mathcal{C}}\left\|\hat{\mathbf{U}}-\mathbf{U}\right\|_2^2 d\mathbf{U}. \end{aligned}\tag{2}$$

The localisation error over each partition depends on both its size and shape. Among all partitions with the same volume, the value of this integral would be minimised if the shape was a sphere and the estimate, $\hat{\mathbf{U}}$, was at the centre of that sphere:

$$\iiint_{\mathcal{C}}\left\|\hat{\mathbf{U}}-\mathbf{U}\right\|_2^2 d\mathbf{U} \geq \iiint_{H_r}\|\boldsymbol{c}-\mathbf{U}\|_2^2\, d\mathbf{U}, \tag{3}$$

where H_r is a sphere with centre $\boldsymbol{c}$ and radius $r = \sqrt[3]{3\mathcal{V}(\mathcal{C})/(4\pi)}$. Evaluating this integral, we obtain

$$\iiint_{H_r}\|\boldsymbol{c}-\mathbf{U}\|_2^2\, d\mathbf{U} = \frac{4\pi}{5}r^5 = K\mathcal{V}(\mathcal{C})^{\frac{5}{3}}, \tag{4}$$

where $K = \frac{4\pi}{5}\sqrt[3]{\frac{3}{4\pi}}$.

Combining (2), (3) and (4) yields

$$\mathbb{E}\left(\left\|\hat{\mathbf{U}}-\mathbf{U}\right\|^2\right) > \frac{K}{\mathcal{V}(\mathcal{R})}\sum_{\mathcal{C}\in\mathcal{P}}\mathcal{V}(\mathcal{C})^{\frac{5}{3}}.$$

This lower-bound would be minimised if the available volume, $\mathcal{V}(\mathcal{R})$, was split equally among each of the regions in the sum:

$$\begin{aligned}\mathbb{E}\left(\left\|\hat{\mathbf{U}}-\mathbf{U}\right\|^2\right) &> K\frac{1}{\mathcal{V}(\mathcal{R})}\sum_{\mathcal{C}\in\mathcal{P}}\left(\frac{\mathcal{V}(\mathcal{R})}{\#\mathcal{P}}\right)^{\frac{5}{3}}\\ &= K\left(\frac{\mathcal{V}(\mathcal{R})}{\#\mathcal{P}}\right)^{\frac{2}{3}}. \end{aligned}\tag{5}$$

Here, $\#\mathcal{P}$ is the number of partitions (the cardinality of $\mathcal{P}$).

Since the volume of the region of interest, $\mathcal{V}(\mathcal{R})$, is fixed, we just need to consider how the number of regions, $\#\mathcal{P}$, grows as we add more cameras to the system. To do so, we first consider how many regions can be created from L planes in $\mathbb{R}^3$. In computational geometry, this quantity is known as the number of cells in an arrangement of hyperplanes (see for example [?]). It can be shown that, with L planes, the 3-D space $\mathbb{R}^3$ is partitioned into at most k regions and k grows cubically with L, i.e. $k = \mathcal{O}(L^3)$.

In our case, partitions are created by the boundaries of the pixels. We can see that each camera in a multi-camera

Authors are with with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: firstname.surname@epfl.ch).
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—SPARSAM Nr: 247006.
A. Ghasemi was additionally supported by a Qualcomm Innovation Fellowship.

system partitions the space with at most $2(N+1)$ planes intersected by rays starting from the camera centre and passing through pixel boundaries[1] (we have an upper bound since some or all of these planes may not pass through the region of interest). Therefore, for M cameras, we have at most $2M(N+1)$ such planes passing through the region of interest and thus we can conclude that the number of regions ($\#\mathcal{P}$) satisfies

$$\#\mathcal{P} = \mathcal{O}\left(M^3 N^3\right). \tag{6}$$

Substituting (6) into (5) gives

$$\mathbb{E}\left(\left\|\hat{\mathbf{U}} - \mathbf{U}\right\|^2\right) = \Omega\left(\frac{\mathcal{V}(\mathcal{R})}{M^2 N^2}\right),$$

which proves that $\mathbb{E}\left(\left\|\hat{\mathbf{U}} - \mathbf{U}\right\|^2\right) = \Omega\left(1/M^2\right)$ for fixed N and $\mathcal{R}$, hence the fact that best possible decay rate for a geometric reconstruction algorithm is quadratic. □

2 Proof of Proposition 1

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

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

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

Proof. The proof will be by contradiction. Let $\hat{\mathbf{U}}$ be the minimum (ℓ_q, ℓ_∞)-norm solution:

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{X}} \max_{i=1..M} \|\mathbf{u}_i - \mathcal{P}_i(\mathbf{X})\|_q. \tag{7}$$

Assume that $\hat{\mathbf{U}}$ is not consistent. Then, there exists an i such that

$$\left\|\mathbf{u}_i - \mathcal{P}_i(\hat{\mathbf{U}})\right\|_q > \delta. \tag{8}$$

Alternatively, let $\boldsymbol{X}_c$ be a consistent estimate. By definition,

$$\|\mathbf{u}_i - \mathcal{P}_i(\boldsymbol{X}_c)\|_q \leq \delta \quad \text{for all } i = 1...M. \tag{9}$$

Therefore,

$$\max_{i=1..M} \left\|\mathbf{u}_i - \mathcal{P}_i(\hat{\mathbf{U}})\right\|_q > \max_{i=1..M} \|\mathbf{u}_i - \mathcal{P}_i(\boldsymbol{X}_c)\|_q. \tag{10}$$

But, this contradicts (7) and thus $\hat{\mathbf{U}}$ must be consistent. □

1. In the case of orthogonal projection, rays do not originate from the centre of the camera, but their cardinality and hence the rest of the proof remain unchanged.

3 Proof of Theorem 2

The proof makes use of the following corollary.

Corollary 1 (Powell and Whitehouse 2016)**.** *Assume random vectors $\{\phi_i\}_{i=1}^M \subset \mathbb{R}^d$ are i.i.d. and uniformly distributed on the unit d-dimensional sphere. Suppose a point in $\mathbb{R}^d$ is orthogonal projected onto the random vectors and subjected to zero-mean uniform bounded noise with bandwidth δ. Then, constants $c_1, c_2 > 0$ exist such that*

$$\mathbb{E}\{(W_M)^2\} \leq \frac{c_2 d^3 \delta^3}{M^2}, \qquad \forall M \geq c_1 d \ln d. \tag{11}$$

Here, W_M is the radius of the smallest d-dimensional sphere containing the consistency region formed from the M samples.

Proof. See [Powell and Whitehouse 2016, Corollary 6.2]. □

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 $M \to \infty$ and place a point anywhere in this region.*

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

$$\mathbf{u}_i = \mathcal{P}_i(\mathbf{U}) + \boldsymbol{\epsilon}_i, \tag{12}$$

where $\boldsymbol{\epsilon}_i = [\epsilon_{i,x}, \epsilon_{i,y}]^T$ and $\epsilon_{i,x}$, $\epsilon_{i,y}$ are zero-mean uniform bounded random variables with bandwidth δ.

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

$$\mathbb{E}\left(\left\|\hat{\mathbf{U}} - \mathbf{U}\right\|^2\right) = \mathcal{O}\left(\frac{1}{M^2}\right), \tag{13}$$

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. Let $\mathbf{U} = (U_X, U_Y, U_Z)$ and assume, without loss of generality, that the circle lies in the X-Z plane.

Before considering the central projection case, we assume the cameras are orthographic. In this case, the vertical coordinate of the image points are given by

$$u_{i,y} = U_Y + \epsilon_{i,y}, \quad i \in [1, M]. \tag{14}$$

The consistent region for U_Y, which we will denote by $\mathcal{C}_y$, is simply a 1-D interval:

$$\mathcal{C}_y = \left\{\hat{U}_Y : \max_i \epsilon_{i,y} - \delta/2 \leq \hat{U}_Y - U_Y \leq \min_i \epsilon_{i,y} + \delta/2\right\}.$$

Therefore, the maximum reconstruction error is

$$\begin{aligned}\mathcal{E} &:= \max_{\hat{U}_Y \in \mathcal{C}_y} \left|\hat{U}_Y - U_Y\right| \\ &= \max\left\{\left|\max_i \epsilon_{i,y} - \frac{\delta}{2}\right|, \left|\min_i \epsilon_{i,y} + \frac{\delta}{2}\right|\right\} \\ &= \max\left\{\mathcal{E}_l, \mathcal{E}_u\right\}, \end{aligned} \tag{15}$$

where $\mathcal{E}_l := \left|\max_i \epsilon_{i,y} - \frac{\delta}{2}\right| = \frac{\delta}{2} - \max_i \epsilon_{i,y}$ and $\mathcal{E}_u := \left|\min_i \epsilon_{i,y} + \frac{\delta}{2}\right| = \min_i \epsilon_{i,y} + \frac{\delta}{2}$ are the absolute values of the lower and upper bounds, respectively.

The expected maximum squared error can be computed as

$$\mathbb{E}\left(\mathcal{E}^2\right) = \int_0^\infty \lambda^2 \frac{d\mathbb{P}\left(\mathcal{E} \le \lambda\right)}{d\lambda} d\lambda = 2\int_0^\infty \lambda \mathbb{P}\left(\mathcal{E} \ge \lambda\right) d\lambda.$$

Furthermore, from (15), we have

$$\begin{aligned}\mathbb{P}\left(\mathcal{E} \ge \lambda\right) &= \mathbb{P}\left(\mathcal{E}_l \ge \lambda \cup \mathcal{E}_u \ge \lambda\right) \\ &= \mathbb{P}\left(\mathcal{E}_l \ge \lambda\right) + \mathbb{P}\left(\mathcal{E}_u \ge \lambda\right) - \mathbb{P}\left(\mathcal{E}_l \ge \lambda \cup \mathcal{E}_u \ge \lambda\right).\end{aligned}$$

Each term can be calculated as

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_l \ge \lambda\right) &= \mathbb{P}\left(\epsilon_{i,y} \le \frac{\delta}{2} - \lambda,\, i \in [1, M]\right) \\ &= \left(1 - \frac{\lambda}{\delta}\right)^M \quad \text{for } 0 \le \lambda \le \delta,\end{aligned}$$

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_u \ge \lambda\right) &= \mathbb{P}\left(\epsilon_{i,y} \ge \lambda - \frac{\delta}{2},\, i \in [1, M]\right) \\ &= \left(1 - \frac{\lambda}{\delta}\right)^M \quad \text{for } 0 \le \lambda \le \delta,\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_l \ge \lambda \cup \mathcal{E}_u \ge \lambda\right) &= \mathbb{P}\left(\lambda - \frac{\delta}{2} \le \epsilon_{i,y} \le \frac{\delta}{2} - \lambda, i \in [1, M]\right) \\ &= \left(1 - \frac{2\lambda}{\delta}\right)^M \quad \text{for } 0 \le \lambda \le \frac{\delta}{2}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(\mathcal{E}^2\right) =& 4\int_0^\delta \lambda\left(1 - \frac{\lambda}{\delta}\right)^M d\lambda - 2\int_0^{\delta/2} \lambda\left(1 - \frac{2\lambda}{\delta}\right)^M d\lambda \\ =& \frac{14\delta^2}{4(M+1)(M+2)}\end{aligned}$$

and so

$$\mathbb{E}\left(\left|\hat{U}_Y - U_Y\right|^2\right) \le \frac{14\delta^2}{4(M+1)(M+2)} < \frac{14\delta^2}{4M^2}, \tag{16}$$

for any consistent estimate $\hat{U}_Y$ of U_Y.

Let's now consider the horizontal coordinate of the image points. If we continue to assume orthographic projection, we have

$$\begin{bmatrix} u_{1,x} \\ u_{2,x} \\ \vdots \\ u_{M,x} \end{bmatrix} = \begin{bmatrix} -\sin\theta_1 & \cos\theta_1 \\ -\sin\theta_2 & \cos\theta_2 \\ \vdots & \vdots \\ -\sin\theta_M & \cos\theta_M \end{bmatrix} \begin{bmatrix} U_X \\ U_Z \end{bmatrix} + \begin{bmatrix} \epsilon_{1,x} \\ \epsilon_{2,x} \\ \vdots \\ \epsilon_{M,x} \end{bmatrix}.$$

This is a linear inverse problem in two dimensions, seeking unknowns U_X and U_Z, leading to a 2-D consistent region. The geometry of this consistent region is more complicated than the 1-D case; however, the assumption that the cameras are uniformly distributed on the circle simplifies this geometrical dependence. This is exploited in [Powell and Whitehouse 2016] to prove various bounds including Corollary 1. Directly applying this corollary yields

$$\mathbb{E}\left(\left\|\begin{bmatrix}\hat{U}_X \\ \hat{U}_Z\end{bmatrix} - \begin{bmatrix}U_X \\ U_Z\end{bmatrix}\right\|^2\right) \le \frac{K_1\delta^2}{M^2}, \tag{17}$$

for any consistent estimate $[\hat{U}_X, \hat{U}_Z]^T$ of $[U_X, U_Z]^T$. Here K_1 is a constant independent of the number of cameras and the support of the bounded noise.

Combining (16) and (17) yields

$$\mathbb{E}\left(\left\|\hat{\mathbf{U}} - \mathbf{U}\right\|^2\right) \le \frac{K_2\delta^2}{M^2},$$

for the orthographic case. Here K_2 is a constant independent of the number of cameras and the support of the bounded noise.

Now, to extend this result to the pinhole camera case, let r be the radius of the circle and f be the focal length of all cameras. Then, the pinhole projection consistency region corresponding to an image point measurement with a noise bandwidth of δ has a smaller volume than the consistency region of an orthogonal projection, with larger bandwidth and a circle of interest of radius $r - f$. The bandwidth δ_{equiv} of this corresponding parallel projection camera is computed as

$$\delta_{equiv} = \delta\left(1 + \frac{r-f}{f}\right) = \delta\left(\frac{r}{f}\right). \tag{18}$$

This means that we can upper-bound the reconstruction error of a circular array of M pinhole cameras with a measurement error bandwidth of δ, with the reconstruction error of a circular array of *parallel* cameras, with the bandwidth δ_{equiv} as defined above. Using this fact, we have the following bound:

$$\mathbb{E}\left(\left\|\hat{\mathbf{U}} - \mathbf{U}\right\|^2\right) \le \frac{K_2\delta^2 r^2}{M^2 f^2}. \tag{19}$$

□