UC Davis
IDAV Publications

Title
Uncertainty, Baseline, and Noise Analysis for L1 Error-Based Multi-View Triangulation

Permalink
https://escholarship.org/uc/item/6nk233jn

Authors
Hess-Flores, Mauricio
Recker, Shawn
Joy, Kenneth I.

Publication Date
2014

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Uncertainty, Baseline, and Noise Analysis for L_1 Error-Based Multi-View Triangulation

Mauricio Hess-Flores*, Shawn Recker‡, and Kenneth I. Joy§
Institute for Data Analysis and Visualization
University of California, Davis
Email: *mhessf@ucdavis.edu, ‡strecker@ucdavis.edu, §kenneth.i.joy@gmail.com

***Abstract*—A comprehensive uncertainty, baseline, and noise analysis in computing 3D points using a recent L_1-based triangulation algorithm is presented. This method is shown to be not only faster and more accurate than its main competitor, linear triangulation, but also more stable under noise and baseline changes. A Monte Carlo analysis of covariance and a confidence ellipsoid analysis were performed over a large range of baselines and noise levels for different camera configurations, to compare performance between angular error-based and linear triangulation. Furthermore, the effect of baseline and noise was analyzed for true multi-view triangulation versus pairwise stereo fusion. Results on real and synthetic data show that L_1 angular error-based triangulation has a positive effect on confidence ellipsoids, lowers covariance values and results in more-accurate pairwise and multi-view triangulation, for varying numbers of cameras and configurations.**

***Keywords*—triangulation; noise analysis; uncertainty; multi-view reconstruction; structure-from-motion;**

I. Introduction

During the past years, there has been a surge in the amount of work dealing with multi-view reconstruction of scenes, for applications such as robotics, surveillance and virtual reality. In the reconstruction process, three-dimensional (3D) objects and scenes can be computed from a collection of images taken from different camera viewpoints. Most common reconstruction algorithms produce a point cloud representing the scene's structure. State-of-the-art algorithms [1], [2] can obtain very accurate reconstructions from large numbers of images, even from the Internet. Most commonly, such reconstruction algorithms are based on sparse feature detectors such as SIFT [3]. Despite notable recent advances, increasing the accuracy, efficiency, and reconstruction density are still areas of ongoing research.

Triangulation is an important step in the process of scene reconstruction, typically following the initial steps of feature tracking and camera calibration [4]. Triangulation determines the 3D location of a scene point X from its imaged pixel location x_i in two or more images. When X reprojects exactly onto its x_i coordinates, triangulation is trivial through simple, linear methods. However, in the presence of image noise, the reprojected coordinates of X will not coincide with each respective x_i. In the general setting with an arbitrary number of cameras, possibly noisy camera parameters, and inaccurate image measurements (*feature tracks*), the goal becomes finding the point X that best fits a given track. There are a number of triangulation methods in the literature. The midpoint method [4], though inaccurate in general, is by far the fastest method given two views. Another common, fast method that solves for 3D points based on linear least squares is *linear triangulation* [4]. Such methods are not very accurate under noise, and the obtained solution is not necessarily the best from all feasible solutions.

A number of optimal algorithms exist in the literature, but most solve for either two views (such as Lindstrom [5]) or three (such as Byrod et al. [6]). For more than three views, the traditional approach has been a two-phase method, where an initial linear method such as N-view linear triangulation [4] is applied to obtain an initial point, followed by non-linear *bundle adjustment* optimization to reduce the L_2-norm of reprojection error [7]. Bundle adjustment is expensive and prone to miring in local minima, hence it requires an accurate initialization. A few N-view optimal solvers have been proposed, with very limited results. Most recently, a triangulator by Recker et al. [8] uses a novel L_1 angular error cost function that is optimized with adaptive gradient descent given an initial midpoint estimate. It shows a significant speed increase and better reprojection errors than other triangulators, such as N-view linear. However, its performance under noise and baseline changes has not been well-studied, and that is the main contribution of this paper.

Triangulation uncertainty is a function of feature tracking accuracy, camera calibration, and reconstruction angle. Beder and Steffen [9] provide an algorithm that computes the confidence ellipsoid for a 3D point. Based upon a set of points' average 'roundness', the algorithm determines the best pair to start a 3D reconstruction, by essentially fixing its scale. The roundness of the expected covariance matrices, which is related to the condition number of the normal equation system for reconstruction, is maximized, in order to obtain frames that enable a stable reconstruction. Rumpler et al. [10] provide an algorithm to derive a similar measure of covariance, but based on Monte Carlo simulations, and also analyze the effect of true multi-view triangulation as opposed to pairwise fusion on accuracy. Both algorithms are based on linear triangulation [4].

The main goal in this paper is to provide a comprehensive analysis of noise, baseline, and true multi-view triangulation versus two-view fusion for angular error-based triangulation, based mainly on the procedures presented by Beder and Steffen [9] and Rumpler et al. [10]. Such an analysis has not yet been performed for this algorithm, and this is of great importance given it's the state-of-the-art triangulation algorithm and with the potential to completely replace linear triangulation in future reconstruction pipelines. This is specially relevant with large data, where both great speed and accuracy may be required. Furthermore, it is studied whether this algorithm makes pair-wise fusion become more accurate with respect

to true multi-view stereo even with very small baselines, which is essential for a number of applications in which accurate fusion may be required. Performance comparisons are performed mainly with respect to linear triangulation, given the overall lack of other truly N-view triangulation methods for which to compare with. Related work will be discussed in Section II. Full experimental results will be described in detail in Section III, followed by conclusions in Section IV.

II. Related Work

There are many general scene reconstruction algorithms in the literature, several of which have been successful for certain applications. A comprehensive overview and comparison of different methods is given in Strecha et al. [11]. In this paper, our focus will be on triangulation analysis and not on the other components of reconstruction. Therefore, the following sections will present some background on Recker et al.'s L_1 cost function, confidence ellipsoids for triangulated points, Monte Carlo covariance analysis, and merging of stereo pairs.

A. Linear Versus Angular Triangulation

Triangulation is one of the final steps in reconstruction and its accuracy is a direct function of previously-computed feature tracking, camera intrinsic calibration, and pose estimation [4]. Typically, 3×4 projection matrices are used to encapsulate all camera intrinsic and pose information. The most widely-used method in the literature is *linear triangulation* [4]. A system of the form $AX = 0$ is solved by eigen-analysis or singular value decomposition (SVD), where the A matrix is a function of feature track and camera projection matrix values. The obtained solution is a direct, best-fit solve. Numerical stability issues arise with near-parallel cameras.

Recker et al. [8] proposed an L_1 cost function based on an angular error measure for a candidate 3D position, p, with respect to its feature track t. This method will be referred to from now on as *angular triangulation*. After initializing through the simple midpoint method, adaptive gradient descent is applied on the cost function. Finally, a statistical sampling component, based upon confidence levels, reduces the number of rays used for triangulation of a given feature track. It is mentioned in Recker et al. [8] that an initialization based on linear triangulation is also possible, but in practice did not lead to better final results after optimization, and is more expensive to compute than the midpoint estimate. As for the choice of L_1 norm, it measures the median of noise and is more robust to outliers than the L_2 or L_∞ norms because it does not amplify their effect. Furthermore, the L_2 least-squares solution is the maximum likelihood (ML) estimate under Gaussian image noise, but typically contains many local minima, while the L_∞ model assumes uniform bounded noise, which is not always realistic. The inputs to the algorithm are a set of feature tracks across N images and their respective 3×4 camera projection matrices P_i. The error for p is computed as follows. A unit direction vector v_i is first computed between each camera center C_i and p. A second unit vector, w_{ti}, is computed as the ray from each C_i through its 2D feature track t in each image plane. Since t generally does not coincide with the projection of p in each image plane, there is frequently a non-zero angle between each possible v_i and w_{ti}. Finally, the average of the dot products $v_i \cdot w_{ti}$ across all cameras is obtained. Each dot product can vary from $[-1, 1]$. Given C_i cameras, T the set of all feature tracks, and $p = (X, Y, Z)$ a 3D evaluation position, the cost function for p with respect to a track $t \in T$ is displayed in (1).

$$f_{t \in T}(p) = \frac{\sum_{i \in I}(1 - \hat{v_i} \cdot \hat{w_{ti}})}{||I||} \tag{1}$$

Here, $I = \{C_i | t$ "appears in" $C_i\}$, $\vec{v_i} = (p - C_i)$, and $\vec{w_{ti}} = P_i^+ t_i$. The right pseudo-inverse of P_i is given by P_i^+, and t_i is the homogeneous coordinate of track t in camera i. The equation can be expanded with $v_i = (v_{i,X}, v_{i,Y}, v_{i,Z}) = (X - C_{i,X}, Y - C_{i,Y}, Z - C_{i,Z})$, with normalized $\hat{v_i} = \frac{v_i}{||v_i||}$ and $\hat{w_{ti}} = \frac{w_{ti}}{||w_{ti}||}$. Gradient values are defined in equations $5 - 7$ of Recker et al. [8]. Variation in space is very smooth for this function, as explained in Recker et al. [8], due to the continuous nature of sums of dot products. As for the choice of angular error versus reprojection error, it is important to mention that while they are intrinsically similar, angular error requires a lower operation count to compute. Furthermore, Recker et al. [8] were able to prove that it is very unlikely to converge to local minima while optimizing the L_1-norm of this function.

B. Confidence Ellipsoids for Triangulated Points

Beder and Steffen [9] present an algorithm to determine the best initial pair for a multi-view reconstruction. Their analysis is based on computing a confidence ellipsoid for each computed 3D scene point X, such that its roundness measures the quality of each obtained point. This measure is related to the condition number of the normal equation system for reconstruction, such that the frame pair with the highest average roundness provides the most stable points for reconstruction. For two views, the covariance matrices of image feature matches x' and $x"$ are given by C' and $C"$, respectively. The covariance matrix C_{XX} of the distribution of the scene point coordinates X is proportional to the upper left 4×4 submatrix $N^{-1}_{1:4,1:4}$ for the inverse of the 5×5 matrix N given by (2). The A and B matrices encode information related to the projection matrices for the two cameras, the image coordinates of the feature match yielding the scene point, and the 3D point coordinates.

$$N = \begin{pmatrix} A^T \left(B \begin{pmatrix} \text{C'} & 0 \\ 0 & \text{C"} \end{pmatrix} B^T \right)^{-1} A & X \\ X^T & 0 \end{pmatrix} \tag{2}$$

If the homogeneous vector $X = [X_0^T, X_h]^T$ is normalized to Euclidean coordinates, the covariance matrix of the Euclidean coordinates' distribution is given by (3), where J_e corresponds to the Jacobian of a division of X_0 by X_h.

$$C^{(e)} = J_e C_{XX} J_e^T \tag{3}$$

Finally, with the SVD of the matrix $C^{(e)}$, the roundness R of the confidence ellipsoid is obtained as the square root of the quotient of the smallest singular value λ_3 and the largest singular value λ_1, per $R = \sqrt{\frac{\lambda_3}{\lambda_1}}$. The value of R lies between 0 and 1, and only depends on the relative geometry of the two poses, the feature positions and the 3D point; radial distortion is not explicitly taken into account. The eigenvectors of $C^{(e)}$ define the orientation of the principal axes of the ellipse, while its eigenvalues define their lengths.

1) A Further Analysis of Confidence Ellipsoids: In (2), it is possible to include image covariances C' and C'', but in this case we are interested in analyzing only the geometric properties of triangulation. To this end, both image covariance matrices are set to identity, I_{3x3}. In this setting, let the interior matrix to be inverted in N be denoted by G, and defined per (4). The expression is further simplified with $B_1 = -S(P'X)$ and $B_2 = -S(P''X)$ respective 2×3 matrices. Here, $P'X = \hat{x}'$ and $P''X = \hat{x}''$ can be interpreted as image measurements after reprojecting the 3D point X back into each respective image plane. Applying properties of block matrices, the inverse of G is defined in (5). Now, let $W = A^T G^{-1} A = f(P', P'', x', x'', X)$, such that N can be re-written as in (6).

$$G = B \begin{pmatrix} \text{C'} & 0 \\ 0 & \text{C"} \end{pmatrix} B^T = \begin{pmatrix} B_1 B_1^T & 0 \\ 0 & B_2 B_2^T \end{pmatrix} \quad (4)$$

$$G^{-1} = \begin{pmatrix} (B_1 B_1^T)^{-1} & 0 \\ 0 & (B_2 B_2^T)^{-1} \end{pmatrix} \quad (5)$$

$$N = \begin{pmatrix} W & X \\ X^T & 0 \end{pmatrix} \quad (6)$$

For a block matrix such as N, of size $(m+1) \times (m+1)$ with the upper-left submatrix of size $m \times m$, its inverse is shown in (7), where $K = -X^T W^{-1} X$ is a scalar.

$$N^{-1} = \begin{pmatrix} W^{-1} + \frac{W^{-1} X X^T W^{-1}}{K} & -\frac{W^{-1} X}{K} \\ -\frac{X^T W^{-1}}{K} & \frac{1}{K} \end{pmatrix} \quad (7)$$

From (7), notice that $C_{XX} = (N^{-1})_{1:4,1:4} = W^{-1} + \frac{1}{K} W^{-1} X X^T W^{-1}$. This demonstrates that C_{XX} can actually be represented by two additive terms. Next, focus will return to (3), where $C^{(e)} = J_e C_{XX} J_e^T$. It was mentioned that the 3×4 matrix J_e corresponds to the Jacobian of a division of X_0 by X_h, and can be written as in (8) with the homogeneous vector $X = [X_0^T, X_h]^T$, where $X_0^T = (X_X, X_Y, X_Z)^T$.

$$J_e = \frac{1}{X_h} \left(I_{3\times 3} | - \frac{X_0}{X_h} \right) \quad (8)$$

By multiplying per (3), an expression is obtained for the 3×3 matrix $C^{(e)}$, where all terms have a similar form. For example, element $(1,1)$ of $C^{(e)}$ is shown in (9). If the homogeneous coordinate X_h is very large, this value tends to be small, and vice-versa, and this is related to the condition number in the reconstruction of a 3D point.

$$C_{11}^{(e)} = \frac{C_{XX,11}}{X_h^2} - \frac{(C_{XX,14} + C_{XX,41}) X_X}{X_h^3} + \frac{(C_{XX,41}) X_X^2}{X_h^4} \quad (9)$$

One question that immediately arises is what the space of solutions for the roundness of confidence ellipsoids actually looks like in the neighborhood of a 3D position. For example, Recker et al. [8] showed how their angular cost function has a very smooth variation, such that slightly different 3D positions show function values that do not differ much.

To illustrate this, a scalar field analysis was performed. At each location of a densely-populated uniform scalar field, centered at a chosen scene structure position, both Recker et al.'s [8] cost function and Beder and Steffen's roundness [9] were computed, as displayed in Fig. 1. To compute the roundness scalar field, the homogeneous coordinate at the current position was computed from the SVD of the A matrix, and applied to the non-homogeneous current 3D coordinates (by multiplication). The scalar field visualization was implemented and stored in *VTK* format [12]. Samples are color-coded such that red indicates low values, whereas purple indicates high values within the scalar field. A *VisIt* [13] marching cubes implementation was used to generate isosurfaces, which enclose regions of different ranges of values, in order to create a *contour plot* [13]. The transition between isosurfaces shows that the roundness measure in Fig. 1(b) is smooth, but very different from the values for the angular cost function in Fig. 1(a). The highest scalar field value for roundness does not generally coincide with the convergence position for the angular cost function. Furthermore, notice that roundness values seem to be biased towards points that lie closest to the cameras, such that values for those points are higher. Meanwhile, the angular cost function is depth invariant by definition, since it is a simple angular measure. Analyzing (9), it can be seen that low values for the homogeneous coordinate X_h, typically indicative of bad condition numbers, lead to higher values of $C^(e)$, which may be a possible cause of closer positions having higher values for roundness R.

C. Monte Carlo Estimation of Covariance

Rumpler et al. [10] performed an analysis based on Monte Carlo estimation of covariances, to show that multi-view reconstruction produces much less depth uncertainty than merging stereo-pair reconstructions. Formally, the depth error in stereo is shown in (10), where z is the depth, b the baseline, f the focal length in pixels, and ϵ_d the matching error in pixels (disparity values). Depth precision is mainly a function of the angle formed between the cameras' principal rays, such that with a small angle these rays are nearly parallel.

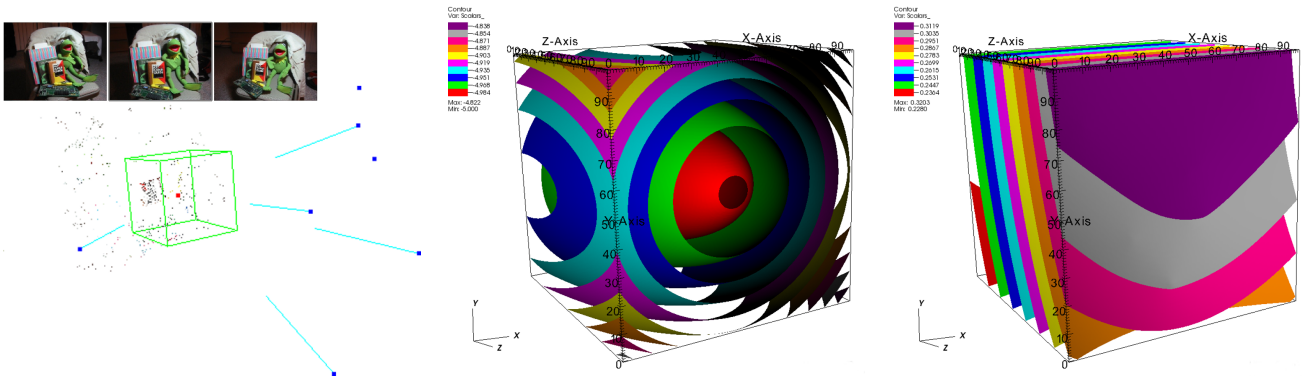
$$\epsilon_z = \frac{z^2 \epsilon_d}{bf} \quad (10)$$

However, Rumpler et al. [10] demonstrated that depth error can alternatively be computed by first estimating a population of points generated by perturbing input feature tracks with zero-mean Gaussian noise and computing the covariance matrix of the resulting set, in a *Monte Carlo* estimation procedure. The largest singular value of this matrix corresponds to the uncertainty in depth estimates for the set of points. For a column vector $\mathbf{X} = [X_1 \dots X_n]$ of 3D positions, the covariance matrix is $\Sigma = \mathrm{E}\left[(X_i - \mu)(X_i - \mu)^T\right]$, where μ is the average 3D position of the set.

Based on this Monte Carlo estimation of depth uncertainty, it was shown that true multi-view linear triangulation is up to ten times more accurate than mean or median fusion of stereo pairs. In the next section, it will be analyzed if the same result holds for angular triangulation.

III. Results

Comprehensive results were obtained for Recker et al.'s algorithm on both real and synthetic datasets. All tests were conducted on an 8-core Intel(R) Core(TM) $i7 - 2600$ CPU at 3.40GHz, on one thread, running Linux Mint 13 (Maya).



(a) Evaluated structure position (b) L_1 cost function [8] (c) Roundness of confidence ellipsoid [9]

Fig. 1. Reconstruction of the *Kermit* dataset [2], where an evaluation 3D position, shown in red, is encased inside a green bounding box (a). Sample images are shown along the top. Inside this bounding box, scalar fields were rendered with Recker et al.'s angular cost function [8](b), and with Beder and Steffen's roundness of confidence ellipsoid metric [9] (c).

In the following tests, these triangulation types will be used: pairwise angular *AP*, multi-view angular *AM*, pairwise linear *LTP*, multi-view linear *LTM*, pairwise midpoint *MP*, and multi-view midpoint triangulation *MM*.

A. Synthetic Setup

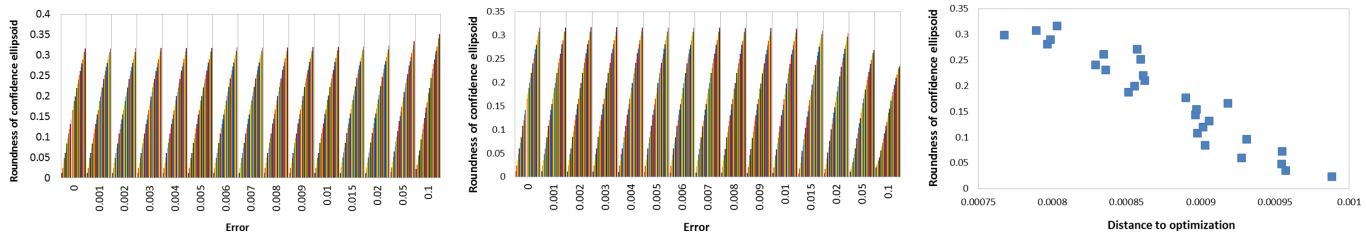
To analyze performance, the following synthetic datasets were generated. The same synthetic camera was used per dataset, with a fixed focal length and principal point. First, a set of cameras were placed in a quarter-circle camera configuration, with an increasing baseline between $0°$-$90°$. A total of 30 cameras are viewing a point located at $(0, 0, 0)$, where 500 trials with random Gaussian noise were performed for each baseline and noise level and then averaged. Then, a similar setup was created for cameras ranging from $0° - 180°$ and from $0° - 360°$. Finally, a set of 30 cameras were placed in a linear configuration, also looking at $(0, 0, 0)$. The following 14 Gaussian noise levels were used: $\sigma = 0.1\%$, 0.2% ..... 1.0%, 1.5%, 2.0%, 5.0% and 10.0%, where the percentages correspond to the image diagonal size and correspond to the standard deviation used for the Gaussian noise.

B. Confidence Ellipsoid Roundness Comparison

Fig. 2 displays a comparison made between computing the roundness metric [9] with linearly-triangulated points versus angular-triangulated points, for the 30-camera quarter-circle setup described above. In this setup, within each X-axis bin, 29 measurements are displayed, each corresponding to a pairwise triangulation between the first image and a second with a uniformly increasing baseline between $3°$-$90°$. Each of the bins represents the 14 error levels, increasing from left to right within the chart. A total of 500 trials with random Gaussian noise were performed for each baseline and noise level combination and then averaged. For angular-triangulated points in (b), notice that roundness values increase with baseline but decrease with increasing error. In contrast, roundness actually increases with large noise in (a). Also, there is a very strong inverse correlation between roundness and angular distance to optimization [8], shown in (c).

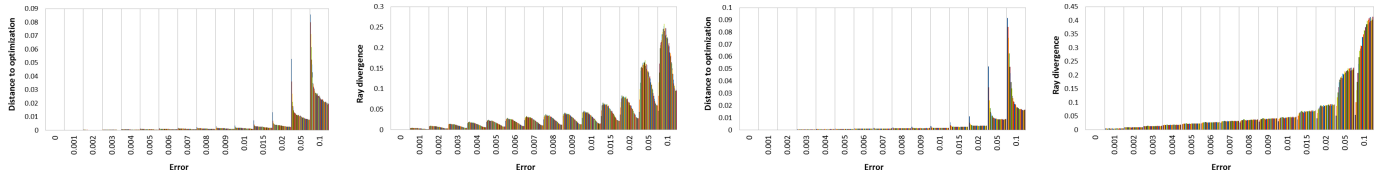
C. Other Baseline and Noise Tests

Other tests were performed to analyze the effect of baseline increase in two-view triangulation under increasing noise. Fig. 3(a) shows the trends seen in the distance from initial midpoint to optimization in *AP*, while (b) displays the corresponding *ray divergences*. Ray divergence is defined as the Euclidean distance between closest points on two rays. It was used as an evaluation metric because of its simple computation and because it correlates well with Beder and Steffen's roundness metric [9], as proven in Hess-Flores et al. [14]. Again, just as in Section III-B, within each X-axis bin measurements are performed with an increasing baseline. Each bin represents the 14 error levels, increasing from left to right, with 500 averaged trials. Notice the clear trends indicating that distances to optimization increase with error but decrease with baseline, while divergences increase with error but show an interesting parabolic shape with respect to increasing baseline. For any of the quarter-circle, semi-circle or circle camera configurations, at $90°$, distances to optimization decrease with respect to lower angles because the triangulation angle is better. However, ray divergences are actually greater. This makes sense if we imagine two parallel rays right next to each other at small baselines, where uncertainty is large but not ray divergence, and spread apart as angle increases. Surprisingly, trends in distances to optimization, roundness and ray divergences continue and stable out near $180°$, perhaps indicating that this is a better triangulation angle than $90°$ [9], [10]. For a linear camera configuration, distances to optimization grow with increasing error, as expected. However, as the baseline grows for a given noise level, the distances to optimization decrease because the angle is wider. For roundness R, results remain practically identical regardless of noise. In general across all configurations, the average distance to optimization increases regardless of the baseline as noise increases. However, for all noise levels, the average distance to convergence decreases as the baseline increases, which is a very good sign. For small baselines, values tend to be highest since there is a lot of uncertainty in the measurement, and this coincides with the literature [8]–[10]. In all tests, timing data for angular triangulation is remarkably stable despite increasing noise, which is a very important property.



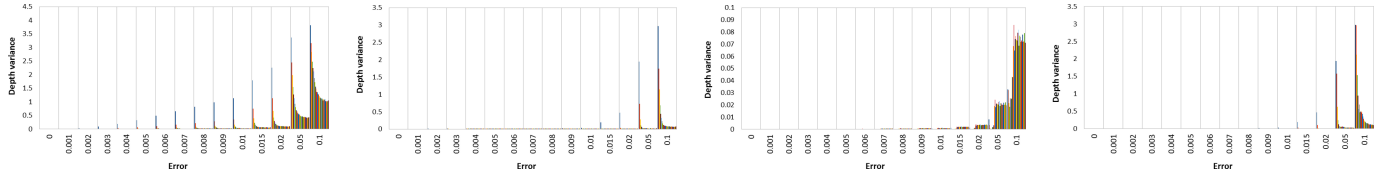
(a) (b) (c)

Fig. 2. Roundness of confidence ellipsoid R [9] computed using *LTP* (a), and using *AP* [8](b), for a quarter-circle camera configuration. Notice that there is an inverse correlation between distances to optimization [8] and the values for R (c).



(a) (b) (c) (d)

Fig. 3. Distance to optimization [8] from the initial midpoint estimate in *AP* (a), and corresponding ray divergences (b), for the half-circle camera configuration. Distance (c) and ray divergence (d) are also shown for the linear camera configuration.



(a) *AP* (b) *LTP* (c) *AM* (d) *LTM*

Fig. 4. Average depth uncertainty between pairwise consecutive images obtained using *AP* (a) versus *LTP* (b), for the quarter-circle camera configuration. Noise levels increase from left to right, and baseline increases left-right within each bin. Similarly, average depth uncertainty results for *AM*, obtained by adding one image at a time as baseline increases, are shown in (c), versus *LTM* in (d).

D. Depth Uncertainty Analysis

In order to analyze how stereo pair fusion fares with respect to true multi-view triangulation, we performed an analysis inspired by Rumpler et al. [10]. First, triangulation estimates between images *1-2*, *2-3*, *3-4*, and so on were computed, using *LTP* and *AP*. At each of the previously-mentioned error levels, a total of 500 depth uncertainty samples were computed and averaged for each pair. Figs. 4(a,b) shows the corresponding results for the quarter-circle configuration. It can be seen that depth uncertainty is generally low for both *AP* and *LTP*, but can be very high in both cases under large errors and small baselines. Next, triangulation estimates between images *1-2*, *1-2-3*, *1-2-3-4*, and so on were computed, adding one image at a time to the initial reconstruction of pair *1-2*, now with *LTM* and *AM*. Figs. 4(c,d) shows the corresponding depth uncertainty results for the quarter-circle configuration. Notice that uncertainty for *LTM* can be *1-2* orders of magnitude higher than for *AM* with small baselines and large errors. This is one of the most telling results of our comparisons, and confirms the stability of *AM* versus *LTM*, which is critical for multi-view reconstruction performance.

E. Results on Real Data

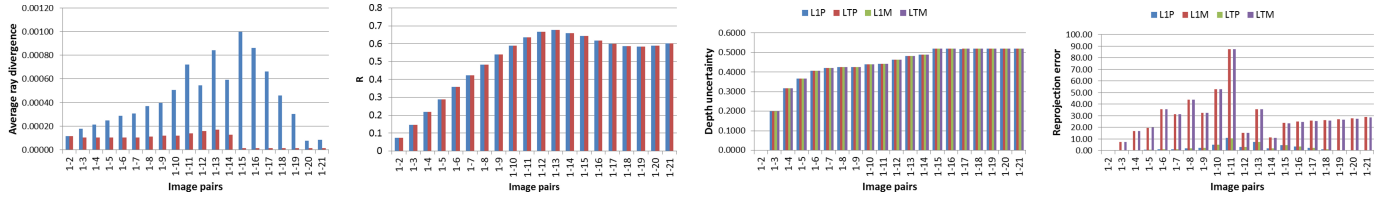
For real datasets, it is assumed that feature tracks and camera projection matrices are available as input. Radial distortion present in the images was not removed prior to testing, and images were acquired sequentially. Table I shows average depth uncertainty results with all triangulators for the well-known *Dinosaur* dataset [15], consisting of 4983 feature tracks, computed similarly to Section III-D. This dataset in general has very accurate feature tracks and cameras. Results are very similar for *AP* and *LTP*, while *AM* slightly outperforms *LTM*. Even though standard deviation values are relatively high, all μ and σ values are very small overall. However, the efficiency and speed benefits of angular triangulation in real datasets are noticed with very large data [8]. Fig. 5 shows baseline analysis data obtained for the same dataset. Ray divergence values in (a) grow and decrease slightly with very long baselines, but values are much smaller when using multiple views. In (b), roundness values with *AP* and *AM* similarly rise, stabilize and settle at lower values as the baseline is increased. In (c), depth uncertainty values are very similar for *AP*, *AM*, *LTP* and *LTM*, with a tendency for covariances to grow with baseline. In (d), reprojection errors are very similar between *AP* versus *LTP* as well as *AM* versus LTM, while midpoint reprojection errors are not displayed since they are at least an order of magnitude higher. Finally, scenes reconstructed with *AM* are shown in Fig. 6.

IV. Conclusions

This paper provided a detailed uncertainty, noise and baseline analysis for a state-of-the-art L_1 angular error-based

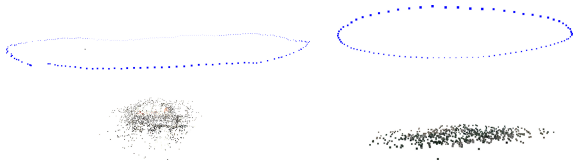
TABLE I. DEPTH UNCERTAINTY DATA FOR THE *Dinosaur* DATASET [15], WITH MEAN μ AND STANDARD DEVIATION σ. VALUES CORRESPOND TO THE AVERAGE LARGEST SINGULAR VALUE OF MONTE CARLO COVARIANCE MATRICES FOR POSITIONAL VARIATION IN EACH CASE.

Triangulation type	μ	σ
AP	1.1144E-05	2.4388E-04
LTP	1.1145E-05	2.4387E-04
AM	2.6643E-06	6.2394E-05
LTM	2.6682E-06	6.2824E-05



(a) d (b) R (c) Depth uncertainty (d) Reprojection error

Fig. 5. Baseline analysis for the *Dinosaur* dataset [15]. Ray divergence d (a) and roundness R [9] (b) are plotted for *AP* (blue) versus *AM* (red). Depth uncertainty (c) and total reprojection error in pixels (d) are plotted for *AP*, *LTP*, *AM*, and *LTM*.



(a) *Brown06* [16] (220 views) (b) *Canyon* [8] (90 views)

Fig. 6. Scenes reconstructed with multi-view angular triangulation. Cameras are rendered in blue.

triangulation method which has been shown to provide greater speed and accuracy than other algorithms in the literature [8]. For varying baselines and noise levels, confidence ellipsoids, Monte Carlo covariance and stereo pair fusion were analyzed versus linear triangulation. On real and synthetic data, it was verified that angular triangulation is not just faster and more accurate than linear triangulation, but is also more stable under noise and baseline changes, and with a lower depth uncertainty. This result is very promising towards applications requiring stability, accuracy and speed in triangulation.

ACKNOWLEDGMENT

This work was supported in part by the Department of Energy, National Nuclear Security Agency through Contract No. DE-FG52-09NA29355 (UCDavis: K. Joy).

REFERENCES

[1] Changchang Wu, "VisualSfM: A visual structure from motion system," http://homes.cs.washington.edu/~ccwu/vsfm/, 2011.

[2] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3D," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*. New York, NY, USA: ACM, 2006, pp. 835–846.

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal On Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[5] P. Lindstrom, "Triangulation Made Easy," in *CVPR*, 2010, pp. 1554–1561.

[6] M. Byröd, K. Josephson, and K. Åström, "Fast optimal three view triangulation," in *Proceedings of the 8th Asian Conference on Computer Vision*, ser. ACCV'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 549–559.

[7] M. I. A. Lourakis and A. A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm," Institute of Computer Science – FORTH, Heraklion, Crete, Greece, Tech. Rep. 340, Aug. 2000.

[8] S. Recker, M. Hess-Flores, and K. I. Joy, "Statistical angular error-based triangulation for efficient and accurate multi-view scene reconstruction," in *Workshop on the Applications of Computer Vision (WACV)*, 2013.

[9] C. Beder and R. Steffen, "Determining an initial image pair for fixing the scale of a 3D reconstruction from an image sequence," in *DAGM-Symposium'06*, 2006, pp. 657–666.

[10] M. Rumpler, A. Irschara, and H. Bischof, "Multi-view stereo: Redundancy benefits for 3D reconstruction," in *Proceedings of the 35th Workshop of the Austrian Association for Pattern Recognition, AAPR/OAGM*, 2011.

[11] C. Strecha, W. von Hansen, L. J. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *CVPR'08*, 2008.

[12] K. Inc., "VTK: Visualization toolkit," 2012. [Online]. Available: http://http://www.vtk.org/

[13] T. Eickermann, W. Frings, and A. Haming, "VISIT - a visualization interface toolkit," 2002. [Online]. Available: https://wci.llnl.gov/codes/visit/home.html

[14] M. Hess-Flores, D. Knoblauch, M. A. Duchaineau, K. I. Joy, and F. Kuester, "Ray divergence-based bundle adjustment conditioning for multi-view stereo," in *Proceedings of the 5th Pacific Rim conference on Advances in Image and Video Technology - Volume Part I*, ser. PSIVT'11. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 153–164. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-25367-6_14

[15] Oxford Visual Geometry Group, "Multi-view and Oxford Colleges building reconstruction," http://www.robots.ox.ac.uk/~vgg/, Aug. 2009.

[16] M. I. Restrepo, B. A. Mayer, A. O. Ulusoy, and J. L. Mundy, "Characterization of 3-d volumetric probabilistic scenes for object recognition." *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, pp. 522–537, Sep. 2012.