Tangent Sampson Error: Fast Approximate Two-view Reprojection Error for General Camera Models Supplementary Material

Mikhail Terekhov
ETH Zurich
mterekhov@ethz.ch

Viktor Larsson
Lund University
viktor.larsson@math.lth.se

Supplementary Material

In the supplementary material we present the following:

- Section A discusses the use of the Essential matrix over the Fundamental matrix in the pinhole errors.
- Section B shows a comparison with the optimal triangulation method from Kukelova et al. [4] which uses the one parameter division model from Fitzgibbon [2].
- Section C provides more qualitative examples comparing the reprojection error with the other error metrics.
- Section D shows the detailed derivations for the covariance-aware Tangent-Sampson error.
- Section E presents the additional details for the experimental setups in the main paper.
- Table 1 shows the runtime comparison without intrinsic optimizations.

A. Essential Matrix and Pinhole Error Metrics

In the main paper we define the pinhole error metrics ($\mathcal{E}_{SED}$, $\mathcal{E}_S$, $\mathcal{E}_{ML}$) in terms of the Essential matrix $\boldsymbol{E}$ instead of the Fundamental matrix $\boldsymbol{F} = \boldsymbol{K}_2^{-\intercal}\boldsymbol{E}\boldsymbol{K}_1^{-1}$, where $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are the respective calibration matrices. If both cameras have the same intrinsic parameters with zero-skew and unit-aspect ratio (a common situation in practice), i.e.

$$\boldsymbol{K}_1 = \boldsymbol{K}_2 = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

then the two variants coincide up to a scaling factor (the focal length f), i.e. $\mathcal{E}_S^2(\boldsymbol{x}, \boldsymbol{F}) = f^2\mathcal{E}_S^2(\boldsymbol{K}^{-1}\boldsymbol{x}, \boldsymbol{E})$, etc.

In the paper we decided to only focus on the $\boldsymbol{E}$ version of the errors for ease of presentation. For all of our quantitative experiments, the two cameras have the same $\boldsymbol{K}$-matrix and aspect ratio very close to one, so we would have gotten similar results using $\boldsymbol{F}$-variant. Thus we believe that all our conclusions should hold for this case as well.

The use of the $\boldsymbol{E}$-variant is also more natural for images with complex distortions (e.g. fisheye or spherical images) that first need to be undistorted to a pinhole image before computing the error. This undistortion can of course map to any image plane (i.e. any focal length), but using the normalized image plane ($\boldsymbol{K} = \boldsymbol{I}$) is a natural choice.

B. Comparison with Kukelova et al.

In [4], Kukelova and Larsson presented a generalization of the optimal triangulation method from Lindstrom [5] that handles radial distortion. Specifically, they assume the that the cameras are modeled with the one parameter division model from Fitzgibbon [2]. The Tangent Sampson error is applicable for this particular model as well, and here we show a small comparison of the two approaches. We take the GoPro7 images of checkerboards (together with intrinsic calibration) from [4] and for all pairs compare the Kendall-Tau correlation between the residuals (as in the main paper). The dataset is split into two categories, medium and wide, corresponding to the FoV-setting on the GoPro camera. The results are shown in Figure 1. For comparison we also include the pinhole triangulations from Lindstrom [5]. The Tangent Sampson error is perfectly correlated with the optimal estimate from Kukelova et al. [4]. Note however that while the results are the same, evaluating the Tangent Sampson error is approximately 18x and 3x times faster than [4] and [5] respectively.

C. Qualitative comparison

In figures 2, 3, and 4 we present qualitative comparisons of all considered error formulations on different camera models. Figure 2 shows the errors in the same setting as Figure 4 from the main paper, where the correspondence is between a fisheye and a panoramic image, but now we also include the pinhole errors and $\mathcal{E}_{\pi SED}$. Figure 3 shows the

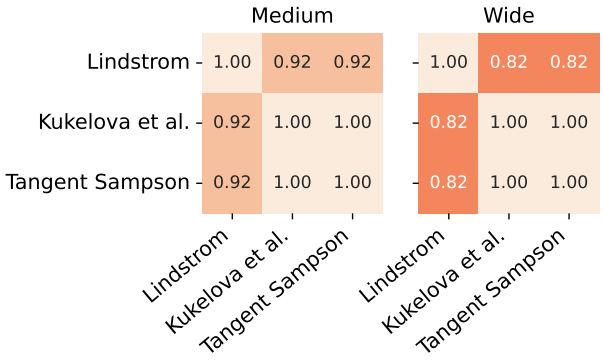

Figure 1. Comparison to Kukelova et al. [4]. The figure shows the Kendall-Tau correlation between the reprojection error obtained from the optimal triangulation from Kukelova et al. [4] and the Tangent Sampson error. For comparison we also show the pinhole *ML* error computed using the method for Lindstrom [5]

	Name		Runtime (Eigen) Relative	Absolute
$\mathcal{E}_{ALG}$	Algebraic error		1.0	2.3 ns
$\mathcal{E}_{CS}$	Cosine error		1.3	3.1 ns
$\mathcal{E}_{ML}$	(Pinhole) Reproj. error		16	37 ns
$\mathcal{E}_{SED}$	(Pinhole) Sym. Epi. Dist.		1.1	2.6 ns
$\mathcal{E}_{S}$	(Pinhole) Sampson error		1.1	2.5 ns
$\mathcal{E}_{\pi ML}$	Reproj. error		495	1160 ns
$\mathcal{E}_{\pi SED}$	Proj. Sym. Epi. Dist.		52	121 ns
$\mathcal{E}_{TS}$	Tangent Sampson		3.5	8.2 ns

Table 1. Runtime comparison for the error metrics compared in the experimental evaluation. The errors are grouped by the domain where the error is computed: geometric errors ( ), the undistorted image ( ) and the original image ( ). Computational cost is shown relative to the algebraic error which is the cheapest. **The table shows the runtime for naive implementations using the Eigen [3] library.** We will make the intrinsic optimized versions (as reported in the main paper) available as open-source.

errors on the pinhole image. Figure 4 shows the errors for a pair of fisheye images. Tangent Sampson is the only error formulation that consistently approximates the true reprojection error faithfully. Only two error formulations, $\mathcal{E}_{ALG}$ and $\mathcal{E}_{TS}$, do not degenerate around the epipole for all image pairs that we consider. However, the algebraic error does not reproduce the spread of the true reprojection error values around the epipolar curve. We believe that stable behavior around the epipole contributes to the better convergence properties we observed when using $\mathcal{E}_{TS}$.

D. Covariance-aware Sampson Error

Here we derive the generalization of the Tangent Sampson error in the case of known covariances in point positions, arriving at the equation (33) from Section 3.2. We begin by linearizing the constraint in the maximum-likelihood error (31):

$$\mathcal{E}_{TS}^2(\boldsymbol{z}, \boldsymbol{E}) = \min_{\hat{\boldsymbol{z}}} \quad (\boldsymbol{z}-\hat{\boldsymbol{z}})^\intercal \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\hat{\boldsymbol{z}}) \qquad (2)$$
$$\text{s.t.} \quad C(\boldsymbol{z}) + \boldsymbol{J}_C(\hat{\boldsymbol{z}}-\boldsymbol{z}) = 0.$$

The Lagrangian is given by

$$\mathcal{L}(\hat{\boldsymbol{z}}, \lambda) = (\boldsymbol{z}-\hat{\boldsymbol{z}})^\intercal \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\hat{\boldsymbol{z}}) + \lambda\left(C(\boldsymbol{z}) + \boldsymbol{J}_C(\hat{\boldsymbol{z}}-\boldsymbol{z})\right), \qquad (3)$$

and the first-order constraints are

$$\hat{\boldsymbol{z}} = \boldsymbol{z} - \frac{1}{2}\lambda \boldsymbol{\Sigma}\boldsymbol{J}_C^\intercal = 0, \quad C(\boldsymbol{z}) + \boldsymbol{J}_C(\hat{\boldsymbol{z}}-\boldsymbol{z}) = 0. \qquad (4)$$

Inserting the first equation into the second yields

$$\lambda = \frac{2C(\boldsymbol{z})}{\boldsymbol{J}_C\boldsymbol{\Sigma}\boldsymbol{J}_C^\intercal} \implies \hat{\boldsymbol{z}} = \boldsymbol{z} - \frac{\boldsymbol{\Sigma}\boldsymbol{J}_C^\intercal C(\boldsymbol{z})}{\boldsymbol{J}_C\boldsymbol{\Sigma}\boldsymbol{J}_C^\intercal}. \qquad (5)$$

The minimum in (3) is thus given by

$$\mathcal{E}_{TS}^2 = \frac{C(\boldsymbol{z})^2}{\boldsymbol{J}_C\boldsymbol{\Sigma}\boldsymbol{J}_C^\intercal}. \qquad (6)$$

Now, we remember that $C(\boldsymbol{z}) = \pi^{-1}(\boldsymbol{p}_1)^\intercal \boldsymbol{E} \pi^{-1}(\boldsymbol{p}_2)$, so its Jacobian can be evaluated using the procedure from Section 3. Together with the fact that $\boldsymbol{\Sigma} = diag(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, this gives us the desired expression:

$$\mathcal{E}_{TS}^2 = \frac{(\boldsymbol{d}_2^\intercal \boldsymbol{E} \boldsymbol{d}_1)^2}{\|\boldsymbol{d}_2^\intercal \boldsymbol{E}\boldsymbol{J}_1^\dagger\|_{\Sigma_1}^2 + \|\boldsymbol{d}_1^\intercal \boldsymbol{E}^\intercal \boldsymbol{J}_2^\dagger\|_{\Sigma_2}^2}. \qquad (7)$$

E. Additional Experiment Details

E.1. Pose refinement and RANSAC evaluations

Pose refinement for experiments in Sections 4.4 and 4.5 was performed using Ceres [1]. Optimization was run over the $S^2 \times \mathrm{SO}(3)$ manifold to work around the inherent scale ambiguity in single-camera relative pose estimation. We used Levenberg-Marquardt optimization with the default Ceres parameters and ran it for a maximum of 100 iterations.

For a more fair comparison in RANSAC, we selected an optimal outlier threshold for each error metric used with RANSAC. We did this through a two-step procedure. First, we extracted 1000 image pairs with at least 200 shared observed 3D points from each reconstruction. For each pair, we evaluated $\mathcal{E}_{\pi ML}$ and got an inlier ratio with inlier correspondences having $\mathcal{E}_{\pi ML} < 1$ pixel. We then selected

the initial threshold for each metric so that the inlier ratio was the same. Next, we searched for an optimal threshold among $\left\{2^i t,\ i \in \{-3, -2, \ldots, 3\}\right\}$ where t is the initial value we got in the first step. RANSAC was run for each threshold candidate on the same set of image pairs used in the first step. The results' median of the pose error (maximal of angular translation and rotation error) was used to decide which threshold is the optimal one.

To better distinguish between the methods, we ran validation for harder image pairs, that only have 50-200 shared 3D points. We selected all such pairs for each dataset (4342 pairs in the Grossmunster dataset and 2731 pairs in Kirchenge) and ran RANSAC on each one for each error metric with the selected inlier threshold.

E.2. Averaging Kendall's τ

The original Kendall's τ coefficient for two rankings $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$ is defined as $\tau(p, q) =$

$$\frac{|\{\text{concordant pairs in } p, q\}| - |\{\text{discordant pairs in } p, q\}|}{\binom{n}{2}}. \quad (8)$$

This can be generalized to the case where we have multiple rankings per method (e.g. correspondences from different image pairs). Given a set of m image pairs with ground truth relative pose, two error formulations provide two sets $p^1, \ldots, p^m$ and $q^1, \ldots, q^m$ of orderings such that $\left|p^i\right| = \left|q^i\right|$. Then, τ can be naturally generalized as $\tau =$

$$\frac{\sum_{i=1}^{m} \left(|\{\text{concordant in } p^i, q^i\}| - |\{\text{discordant in } p^i, q^i\}|\right)}{\sum_{i=1}^{m} \binom{|p^i|}{2}}. \quad (9)$$

To compute average correlation coefficients that we report in the paper, we use equation (9).

References

[1] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. http://ceres-solver.org. 2

[2] Andrew W Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Computer Vision and Pattern Recognition (CVPR)*, 2001. 1

[3] Gaël Guennebaud, Benoît Jacob, et al. Eigen v3. http://eigen.tuxfamily.org, 2010. 2

[4] Zuzana Kukelova and Viktor Larsson. Radial distortion triangulation. In *Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2

[5] Peter Lindstrom. Triangulation made easy. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1554–1561. IEEE, 2010. 1, 2

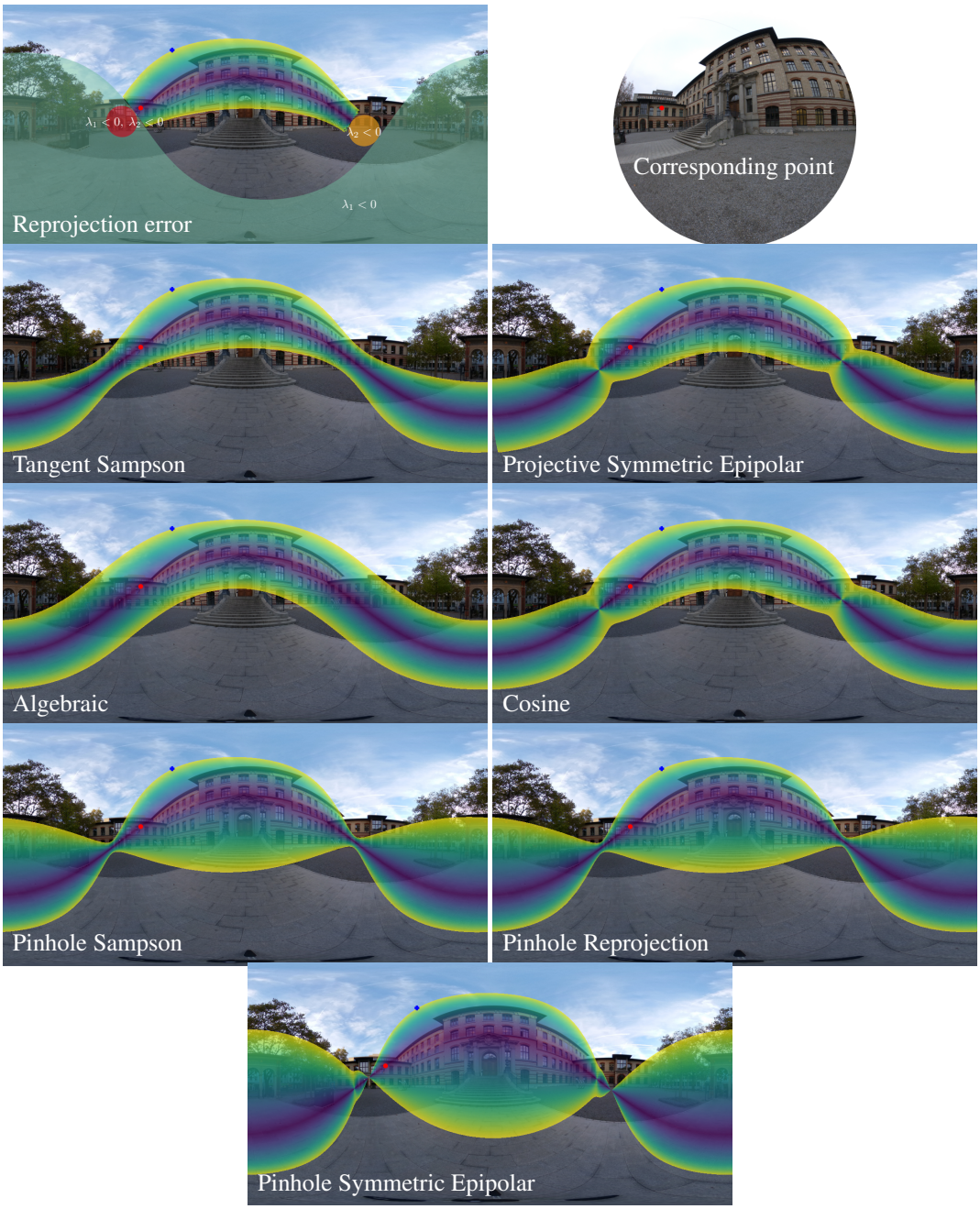

Figure 2. **Qualitative comparison of different error functions for a spherical camera.** The thresholds for colormaps were selected to be the error at the point marked by the blue plus. The geometric error is only shown in the region where the 3D point after midpoint triangulation does not correspond to negative depths of bearing vectors (areas where this does not hold for one or both depths are highlighted).

Figure 3. **Qualitative comparison of different error functions for a pinhole camera.** The thresholds for colormaps were selected to be the error at the point marked by the blue plus. The geometric error is only shown in the region where the 3D point after midpoint triangulation does not correspond to negative depths of bearing vectors (areas where this does not hold for one or both depths are highlighted). Note that level sets for the projection error are slightly curved on the pinhole image because the other camera model is non-linear.

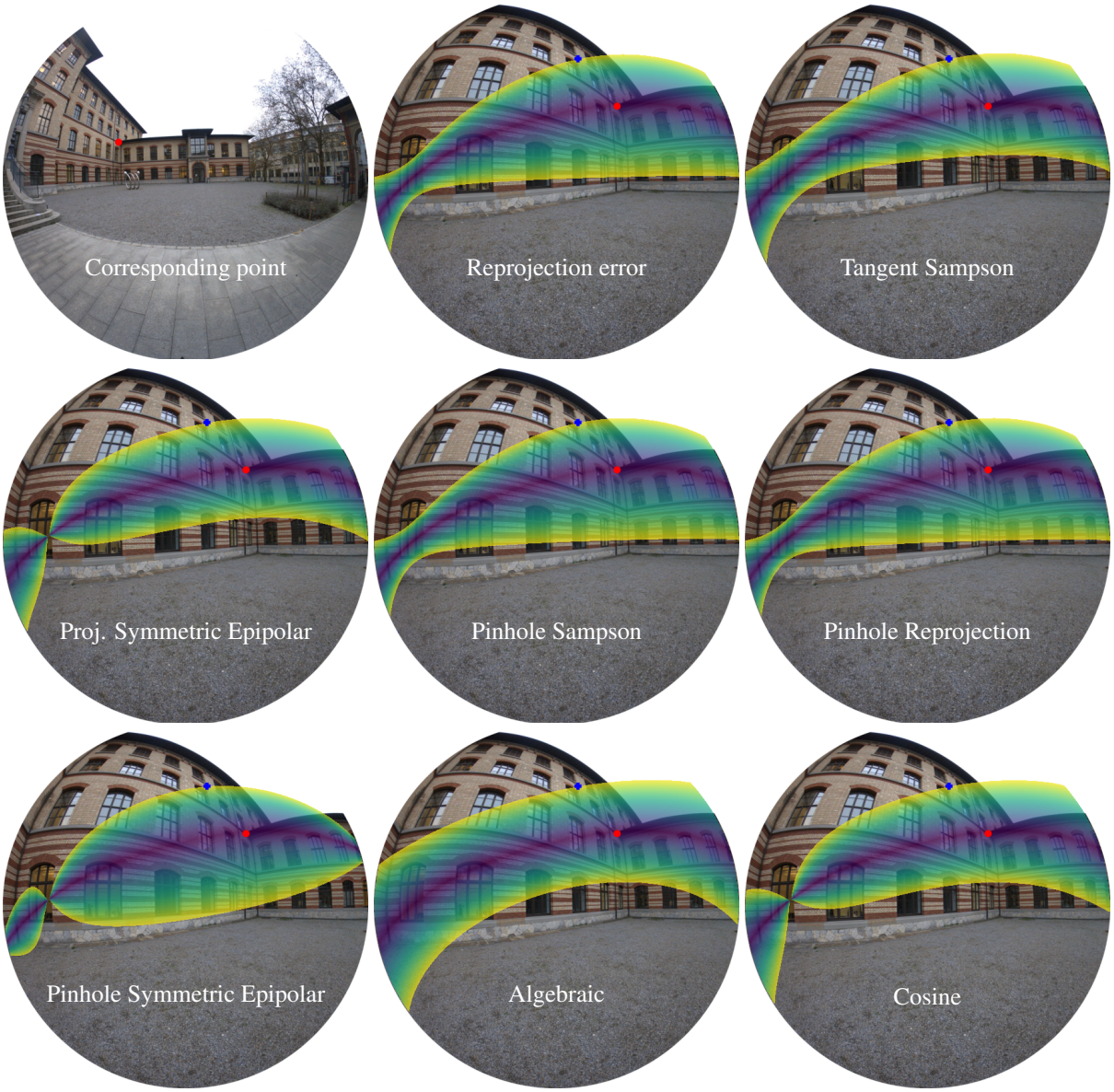

Figure 4. **Qualitative comparison of different error functions for a fisheye camera.** The thresholds for colormaps were selected to be the error at the point marked by the blue plus.