Linear Fundamental Matrix Estimation from 7 or 5 Points

Supplementary Material

8. Additional Details on Real Data Results

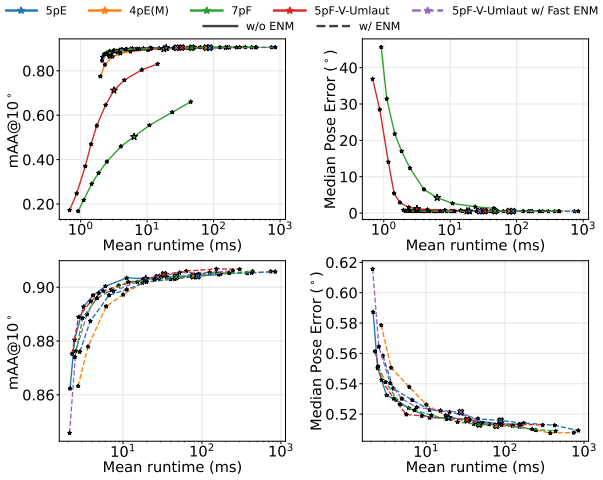

Figure 6. Runtime versus pose accuracy results for the Sacré-Cœur scene in the PhotoTourism [27] validation set. Top row shows all methods, bottom row zooms in on top performers. All methods use the PoseLib LO-RANSAC [30] implementation, with the number of RANSAC iterations in the set $\{10, 20, 50, 100, 200, 500, \mathbf{1000}, 2000, 5000, 10000\}$ and a 1-pixel epipolar threshold. Combined relative pose error is calculated as $\max(\theta_\mathbf{R}, \theta_\mathbf{t})$. Results at 1000 RANSAC iterations are highlighted with larger markers, with stars for methods without ENM and X markers for methods with ENM.

We show the Sacré-Cœur scene results from the PhotoTourism validation set in Figure 6, where most methods perform best. Table 2 additionally reports results for each PhotoTourism scene separately. These results show that `V-Umlaut` can provide comparable performance to top methods without using `ENM` when a scene is generally easier for all methods. Although `V-Umlaut` shows a larger performance drop in more challenging scenes, its runtime advantage is preserved across all scenes, leaving room for improvement through increased RANSAC iterations or by using `ENM`.

9. Details of Derivation

In Figure 7, we provide code for the computer algebra system Macaulay2 [18] that can be used to derive the equations of the rational map $\mathbf{F}_{\text{coords}} = \Phi(s, x, y)$ discussed in Section 3. The formulas for these rational expressions are rather long, but still compact enough to easily cut-and-paste them into our C++ implementation of Solver 3.1.

For the coefficients in equation (18), we find from the

```
FF = frac(QQ[x_1..x_3,y_1..y_3,
        s_(1,1)..s_(2,2)]);
R = FF[f_(2,1),f_(3,1),f_(1,2),
       f_(1,3),f_(2,3)];
F = matrix{
       {0,      f_(1,2), f_(1,3)},
       {f_(2,1),      0, f_(2,3)},
       {f_(3,1),      1, 0      }
       };
I = id_(R^3);
xs = apply(3, i -> I_{i});
xs = xs | {sum xs}
ys = xs;
comb = (pts, i, j, c) ->
       c*pts#i + (1-c) * pts#j;
xs = xs | {
       comb(xs,0,1,s_(1,1)),
       comb(xs,0,2,s_(1,2)),
       matrix{{x_1},{x_2},{x_3}}};
ys = ys | {
       comb(ys,0,1,s_(2,1)),
       comb(ys,0,2,s_(2,2)),
       matrix{{y_1},{y_2},{y_3}}};
Ilin = ideal apply(xs, ys, (x,y) ->
       transpose y * F * x);
netList Ilin_*
eq4 = (Ilin_*)#4;
eq5 = (Ilin_*)#5;
factor((det F)%(ideal(eq4, eq5)))
c1s = s_(1,1)*s_(2,2)*
       (1-s_(1,2))*(1-s_(2,1))
c2s = s_(1,2)*s_(2,1)*
       (1-s_(1,1))*(1-s_(2,2))
I = Ilin + ideal(c1s*f_(2,3)+c2s)
netList apply(gens R, g -> g => g%I)
```

Figure 7. Macaulay2 code.

output of the `factor` command in Figure 7 that

$$c_1(s) = s_{11}s_{22}(1 - s_{12})(1 - s_{21}), \tag{23}$$

$$c_2(s) = s_{12}s_{21}(1 - s_{11})(1 - s_{22}). \tag{24}$$

The operator `%` used in this code is the remainder after polynomial division modulo an ideal, which is computed using Gröbner bases [11]. The final line in this code prints the coordinate functions of the rational map $\Phi(s, x, y)$, giving each entry of $\mathbf{F}_{\text{coords}}$ in terms of input data.

	Median Pose Error (°)				Mean Runtime (ms)			
	w/o ENM							
Scene	5pE	4pE(M)	7pF	5pF-V-Umlaut	5pE	4pE(M)	7pF	5pF-V-Umlaut
Brandenburg Gate	1.33	1.36	6.86	5.36	19.97	19.38	6.26	2.68
Buckingham Palace	1.43	1.51	4.00	5.31	23.53	22.70	7.45	3.03
Colosseum Exterior	1.13	1.13	4.29	3.64	27.94	26.20	8.74	3.92
Grand-Place Brussels	1.89	1.96	7.81	18.80	29.90	28.37	10.63	3.93
Notre-Dame Front Facade	1.42	1.42	8.89	3.00	27.48	25.67	8.43	3.94
Palace of Westminster	1.12	1.17	20.46	8.05	18.59	17.77	5.90	2.68
Pantheon Exterior	1.43	1.53	6.14	8.28	20.19	19.63	6.94	2.82
Reichstag	1.31	1.30	2.90	3.04	35.70	34.79	11.11	4.51
Sacré-Cœur	0.52	0.52	4.24	1.17	19.35	17.85	6.37	3.17
St. Peter's Square	1.71	1.85	7.28	7.19	24.71	24.05	7.87	3.43
Taj Mahal	0.69	0.72	5.84	2.97	22.49	22.19	7.23	3.49
Temple Nara Japan	1.05	1.09	3.40	2.10	30.17	29.64	10.10	4.88
Trevi Fountain	1.57	1.65	3.82	4.93	20.89	19.40	7.87	3.25
Overall	**1.26**	**1.31**	**5.53**	**4.26**	**24.68**	**23.66**	**8.07**	**3.52**
	w/ ENM							
Scene	5pE	4pE(M)	7pF	5pF-V-Umlaut	5pF-V-Umlaut w/ Fast ENM	5pE	4pE(M)	7pF	5pF-V-Umlaut	5pF-V-Umlaut w/ Fast ENM
Brandenburg Gate	1.33	1.33	1.34	1.32	1.33	96.56	91.58	50.76	34.99	28.96
Buckingham Palace	1.44	1.42	1.42	1.42	1.41	113.20	106.07	61.56	42.09	34.99
Colosseum Exterior	1.11	1.11	1.11	1.11	1.11	131.04	118.17	69.50	50.29	40.73
Grand-Place Brussels	1.88	1.87	1.88	1.87	1.90	145.21	132.74	75.90	53.27	43.20
Notre-Dame Front Facade	1.40	1.42	1.42	1.41	1.41	128.61	116.27	65.91	49.58	39.15
Palace of Westminster	1.12	1.14	1.13	1.14	1.14	85.18	78.80	47.27	31.86	26.94
Pantheon Exterior	1.43	1.44	1.42	1.42	1.46	94.91	89.97	53.23	35.68	30.14
Reichstag	1.31	1.31	1.32	1.30	1.32	172.78	164.01	85.98	66.29	54.45
Sacré-Cœur	0.52	0.51	0.51	0.52	0.52	87.81	77.42	46.96	33.18	27.55
St. Peter's Square	1.69	1.71	1.70	1.70	1.72	119.32	112.80	61.95	44.94	36.98
Taj Mahal	0.70	0.70	0.69	0.70	0.71	108.73	105.41	54.40	40.53	33.62
Temple Nara Japan	1.08	1.08	1.07	1.07	1.07	141.87	135.95	73.38	56.16	47.33
Trevi Fountain	1.58	1.56	1.55	1.58	1.60	85.05	76.31	51.93	34.75	28.80
Overall	**1.26**	**1.26**	**1.26**	**1.25**	**1.26**	**116.17**	**108.11**	**61.44**	**44.12**	**36.37**

Table 2. PhotoTourism [27] all 13 scenes from the training and validation sets. Per-scene results shown here are aggregated in Figure 5 and Table 1. Median combined pose error is computed as $\max(\theta_{\mathbf{R}}, \theta_{\mathbf{t}})$, and runtime in ms is averaged over all pairs. RANSAC iteration count is fixed at 1000.

References

[1] Dhruv Agarwal, Taci Kucukpinar, Joshua Fraser, Jeffrey Kerley, Andrew R. Buck, Derek T. Anderson, and Kannappan Palaniappan. Simulating city-scale aerial data collection using Unreal Engine. In *IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2023. 6

[2] Sameer Agarwal, Erin Connelly, Annalisa Crannell, Timothy Duff, and Rekha R Thomas. A computer vision problem in flatland. *SIAM Journal on Applied Algebra and Geometry*, 10(1):14–45, 2026. 2

[3] Hadi Aliakbarpour, Kannappan Palaniappan, and Guna Seetharaman. Robust camera pose refinement and rapid SfM for multiview aerial imagery – Without RANSAC. *IEEE Geoscience and Remote Sensing Letters*, 12(11): 2203–2207, 2015. 6

[4] Hadi Aliakbarpour, Kannappan Palaniappan, and Guna Seetharaman. Fast structure from motion for sequential and wide area motion imagery. In *IEEE Int. Conf. on Computer Vision, Workshop on Video Summarization for Large-scale Analytics*, pages 1086–1093, 2015. 6

[5] Daniel Barath. Five-point fundamental matrix estimation for uncalibrated cameras. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 235–243. Computer Vision Foundation / IEEE Computer Society, 2018. 2

[6] Robert C. Bolles and Martin A. Fischler. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI '81, Vancouver, BC, Canada, August 24-28, 1981*, pages 637–643. William Kaufmann, 1981. 2

[7] Martin Bråtelund. A classification of critical configurations for any number of projective views. *arXiv preprint arXiv:2401.03450*, 2024. 2

[8] Martin Bråtelund. Critical configurations for two projective views, a new approach. *Journal of Symbolic Computation*, 120:102226, 2024. 2

[9] Jaired R. Collins, Taci Kucukpinar, Timothy Duff, Joshua Fraser, and Kannappan Palaniappan Guna Seetharaman. Fundamental constraints on camera centers arising from 3D-2D matches of points and lines. *Int. Journal of Computer Vision*, 134:94,19pp, 2026. 2

[10] Erin Connelly, Sameer Agarwal, Alperen Ergur, and Rekha R. Thomas. The geometry of rank drop in a class of face-splitting matrix products: Part I. *Advances in Geometry*, 24(3):369–394, 2024. 2

[11] David A. Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer Nature Switzerland, fifth edition, 2025. 1

[12] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: self-supervised interest point detection and description. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 337–33712, 2018. 7

[13] Timothy Duff, Kathlén Kohn, Anton Leykin, and Tomas Pajdla. PLMP–Point-Line Minimal Problems in Complete Multi-view Visibility. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1):421–435, 2023. 2

[14] Timothy Duff, Kathlén Kohn, Anton Leykin, and Tomás Pajdla. $\mathrm{Pl}_1\mathrm{p}$: Point-line minimal problems under partial visibility in three views. *Int. J. Comput. Vis.*, 132(8): 3302–3323, 2024. 2

[15] Ahmed Elliethy and Gaurav Sharma. Vehicle tracking in wide area motion imagery via stochastic progressive association across multiple frames. *IEEE Transactions on Image Processing*, 27(7):3644–3656, 2018. 6

[16] Olivier D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Computer Vision - ECCV'92, Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19-22, 1992, Proceedings*, pages 563–578. Springer, 1992. 2

[17] Ke Gao, Hadi Aliakbarpour, Gunasekaran Seetharaman, and Kannappan Palaniappan. DCT-based local descriptor for robust matching and feature tracking in wide area motion imagery. *IEEE Geoscience and Remote Sensing Letters*, 18(8):1441–1445, 2021. 6

[18] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at `http://www2.macaulay2.com`. 1

[19] Banglei Guan, Ji Zhao, Saibal Mitra, and Laurent Kneip. Six-point method for multi-camera systems with reduced solution space. *Int. J. Comput. Vis.*, 133(10):7270–7292, 2025. 2

[20] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. 2, 3, 4, 5, 6

[21] Richard I. Hartley and Fredrik Kahl. Critical configurations for projective reconstruction from multiple views. *Int. J. Comput. Vis.*, 71(1):5–47, 2007. 2

[22] Richard I. Hartley, Rajiv Gupta, and Tom Chang. Stereo from uncalibrated cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1992, Proceedings, 15-18 June, 1992, Champaign, Illinois, USA*, pages 761–764. IEEE, 1992. 2

[23] Robert J. Holt and Arun N. Netravali. Uniqueness of solutions to three perspective views of four points. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(3):303–307, 1995. 2

[24] Petr Hruby, Viktor Korotynskiy, Timothy Duff, Luke Oeding, Marc Pollefeys, Tomas Pajdla, and Viktor Larsson. Four-view geometry with unknown radial distortion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8990–9000, 2023. 2

[25] Petr Hruby, Timothy Duff, and Marc Pollefeys. Efficient solution of point-line absolute pose. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21316–21325, 2024. 2

[26] Petr Hruby, Timothy Duff, Anton Leykin, and Tomás Pajdla. Learning to solve hard minimal problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(11):9386–9401, 2025. 2

[27] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image Matching across Wide Baselines: From Paper to Practice. *International Journal of Computer Vision*, 2020. 7, 8, 1, 2

[28] Elham Soltani Kazemi, Imad Eddine Toubal, Gani Rahmon, Juan Mogollon, and Kannappan Palaniappan. Domain generalization for multiple video object segmentation and tracking using transformers and smart memory. *Int. Journal of Computer Vision*, 2026, Accepted. 7

[29] Kim Kiehn, Albin Ahlbäck, and Kathlén Kohn. PLMP - point-line minimal problems for projective SfM. *CoRR*, abs/2503.04351, 2025. 1, 2

[30] Viktor Larsson and contributors. PoseLib - Minimal Solvers for Camera Pose Estimation, 2020. 6, 8, 1

[31] Viktor Larsson, Marc Pollefeys, and Magnus Oskarsson. Orthographic-perspective epipolar geometry. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5550–5558, 2021. 2

[32] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):756–770, 2004. 2, 6

[33] David Nistér and Frederik Schaffalitzky. Four points in two or three calibrated views: Theory and practice. *Int. J. Comput. Vis.*, 67(2):211–231, 2006. 2

[34] Magnus Oskarsson. Two-view orthographic epipolar geometry: Minimal and optimal solvers. *Journal of Mathematical Imaging and Vision*, 60(2):163–173, 2018. 2

[35] Kannappan Palaniappan, Raghuveer M Rao, and Guna Seetharaman. Wide-area persistent airborne video: Architecture and challenges. *Distributed Video Sensor Networks*, pages 349–371, 2011. 6

[36] Linfei Pan, Dániel Baráth, Marc Pollefeys, and Johannes Lutz Schönberger. Global structure-from-motion revisited. In *European Conference on Computer Vision (ECCV)*, 2024. 1

[37] Rengarajan Pelapur, Sema Candemir, Filiz Bunyak, Mahdieh Poostchi, Guna Seetharaman, and Kannappan Palaniappan. Persistent target tracking using likelihood fusion in wide-area and full motion video sequences. In *15th International Conference on Information Fusion*, pages 2420–2427, 2012. 6

[38] Zhaoshuai Qi, Yifeng Hao, Rui Hu, Wenyou Chang, Jiaqi Yang, and Yanning Zhang. Indoor 3d reconstruction with an unknown camera-projector pair. *arXiv preprint arXiv:2407.01945*, 2024. 2

[39] Long Quan, Bill Triggs, and Bernard Mourrain. Some results on minimal euclidean reconstruction from four points. *J. Math. Imaging Vis.*, 24(3):341–348, 2006. 2

[40] Srikumar Ramalingam, Sofien Bouaziz, and Peter Sturm. Pose estimation using both points and lines for geo-localization. In *2011 IEEE International Conference on Robotics and Automation*, pages 4716–4723. IEEE, 2011. 2

[41] René Ranftl and Vladlen Koltun. Deep fundamental matrix estimation. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part I*, pages 292–309. Springer, 2018. 1

[42] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 7

[43] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[44] Lars Sommer, Wolfgang Krüger, and Michael Teutsch. Appearance and motion based persistent multiple object tracking in wide area motion imagery. In *IEEE Int. Conf. Computer Vision Workshops*, pages 3871–3881, 2021. 6

[45] Henrik Stewénius. *Gröbner Basis Methods for Minimal Problems in Computer Vision*. PhD thesis, Lund University, Sweden, 2005. 2

[46] Peter Sturm. A historical survey of geometric computer vision. In *14th Int. Conf. Computer Analysis of Images and Patterns (CAIP)*, pages 1–8. Springer Berlin Heidelberg, 2011. 2

[47] Charalambos Tzamos, Viktor Kocur, Yaqing Ding, Daniel Barath, Zuzana Berger Haladová, Torsten Sattler, and Zuzana Kukelova. Practical solutions to the relative pose of three calibrated cameras. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pages 21913–21923. Computer Vision Foundation / IEEE, 2025. 2, 7, 8

[48] Tomás Werner. Constraint on five points in two images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 203–208, 2003. 2

[49] Zichao Zhang, Torsten Sattler, and Davide Scaramuzza. Reference pose generation for long-term visual localization via learned features and view synthesis. *International Journal of Computer Vision*, 129:821 – 844, 2020. 7, 8