Journal of Transfer and Generalization
Volume 1 · Number 1 · 2026

Transferring a One-Parameter Camera Translation Forecast Across TUM Recording Families

Gregory Harrison[1] · Trevor Carter[2] · Jason Reed[3]

[1] United Kingdom · [2] France · [3] Australia

Original publication: November 1, 2026 | Article 675

Abstract

A simple past-velocity forecast transferred well across the acquired TUM recording families. We fitted one displacement multiplier on 364 anchors from ten fr1 files, then froze it for 5,732 anchors in forty-two fr2 and fr3 files. At a fifty-millisecond horizon, mean file-level squared error was 123.89 mm2 for holding position, 2.61 mm2 for ordinary constant velocity, and 2.59 mm2 for the fitted multiplier. The fitted-versus-fixed gain was only 0.0197 mm2 despite a working interval below zero. Most observed improvement came from using recent motion, rather than estimating its nearly unit multiplier.

Short-horizon reference-trajectory analysis. Future reference positions evaluate the forecast; they are not inputs available to the predictor.

Transferring a One-Parameter Camera Translation Forecast Across TUM Recording Families

> Short-horizon reference-trajectory analysis. Future reference positions evaluate the forecast; they are not inputs available to the predictor.

Abstract

A simple past-velocity forecast transferred well across the acquired TUM recording families. We fitted one displacement multiplier on 364 anchors from ten fr1 files, then froze it for 5,732 anchors in forty-two fr2 and fr3 files. At a fifty-millisecond horizon, mean file-level squared error was 123.89 mm² for holding position, 2.61 mm² for ordinary constant velocity, and 2.59 mm² for the fitted multiplier. The fitted-versus-fixed gain was only 0.0197 mm² despite a working interval below zero. Most observed improvement came from using recent motion, rather than estimating its nearly unit multiplier.

A minimal transport question before a complex motion model

Camera-motion systems often combine a geometric representation with an estimated relationship between recent and future movement. A useful preliminary question is whether even a very small fitted relationship survives a change of recording family. If a complex system is compared only with holding the current camera position, an apparent advantage may mostly reflect the elementary benefit of extrapolating recent motion.

We ask whether a scalar velocity extrapolation fitted on one TUM source family improves short-horizon translation prediction in the other two families, relative to holding position and to fixed constant velocity. This creates two distinct comparisons. The first tests the benefit of using recent displacement. The second tests the additional benefit of estimating one multiplier rather than setting it to one.

Meng, 2026 studies scale-calibrated camera-motion transfer through a gauge-invariant parallax number. Our experiment does not implement that transfer method or estimate its parallax quantity. Its relevance is the assessment of a calibration component across recording conditions. A learned scale-like coefficient should be compared with an explicit simple reference before its contribution is credited to calibration itself.

The geometry perspectives in Barfoot, 2017 and Blanco-Claraco, 2021 motivate keeping the coordinate and prediction meanings separate. We predict the source camera-center translation in the source world coordinate frame. We do not predict image flow, subject movement, camera rotation, or a transformed view. The fitted scalar changes an extrapolated translation vector, not the metric unit of the dataset.

Three source families and one frozen coefficient

The underlying public benchmark is documented in Sturm et al., 2012, with timestamp and pose fields specified in TUM Computer Vision Group, 2026. We use the acquired reference-position texts rather than RGB images, depth frames, or a complete SLAM system. Source filenames divide the numerical recordings into fr1, fr2, and fr3 families. These prefixes define the operational domains used for fitting and evaluation.

The local curation preserves original text files, removes ambiguous conflicting-time rows, and produces fifty-two numerical trajectories with unique ordered timestamps. No physical camera error is inferred from that curation. We retain calibration as well as motion recordings. These source roles make the

collection broader than an ordinary navigation-only cohort, and related recordings are not treated as independent device replications.

Ten fr1 files provide the fitting anchors. The forty-two held files comprise twenty fr2 and twenty-two fr3 recordings. Prefix membership is known before computing prediction errors. It does not isolate a pure hardware, date, or motion-class intervention; differences in those attributes may travel together. The term domain transfer therefore refers specifically to these source-family partitions.

The benchmark diversity in Burri et al., 2016 and Schubert et al., 2018 shows that trajectory evaluation can involve different sensors, scenes, and collection procedures. We do not acquire their trajectories for this experiment or claim transport to them. Geiger et al., 2012, Yohann et al., 2020, and Dosovitskiy et al., 2017 describe still other driving or simulated environments. They mark useful external boundaries: a fifty-millisecond indoor reference forecast does not establish accuracy on those distinct motion distributions.

Constructing a forecast without future input

Within each curated trajectory, every hundredth source row becomes a candidate anchor. At each anchor, the predictor uses the current position and a reference position twenty milliseconds earlier. The past displacement is multiplied by fifty divided by twenty to create a constant-velocity displacement at a fifty-millisecond horizon. Holding position instead predicts zero displacement over that horizon.

The fitted rule multiplies this constant-velocity displacement by a single scalar alpha. We estimate alpha through least squares without an intercept on the fr1 anchors, jointly using the three displacement components. No held-family position is used to choose alpha, features, the horizon, or the observation stride. The fitted value, 0.991561, remains unchanged throughout evaluation.

Positions at requested times are taken from an exact source timestamp within two microseconds where available. Otherwise they are linearly interpolated only when neighboring source positions bracket the request and their separation is at most fifty milliseconds. There is no extrapolation to construct evaluation labels or past observations. The past bracket is explicitly checked to end no later than the current anchor.

The future interpolated position is used solely as the outcome reference. A predictor that read that future value would have privileged information and answer a different question. Leakage concerns in Kapoor and Narayanan, 2023 motivate documenting this boundary, while validation principles in Varma and Simon, 2006 motivate keeping the source-family fit separate from the evaluation family summaries.

The procedure yields 364 fitting anchors and 5,732 held anchors. Eligibility requires supported past and future requests, so anchors near ends or unsupported gaps can be excluded. The retained sample is not a uniform continuous-time sample: taking every hundredth row follows the source observation sequence. Its interpretation remains tied to that sampling rule rather than a generic camera operating at an asserted fixed frequency.

What the reference error measures

For each eligible anchor, error is the squared Euclidean distance between the predicted and future reference camera position. We average errors within a file and then give each held file equal weight. Values in square meters are multiplied by one million for reporting in squared millimeters. These are squared-position errors, not millimeter RMSE values and not average signed displacement errors.

The trajectory-evaluation framework in Zhang and Scaramuzza, 2018 emphasizes matching an error definition to the comparison. We do not align a newly estimated trajectory, remove scale, or compute a

standard absolute-trajectory-error leaderboard score. The forecast starts from the source current reference pose and predicts one short displacement. Reference labels are themselves recorded and sometimes interpolated observations, rather than proof of the exact continuous physical path.

The measurement-error perspective in Carroll et al., 2006 matters because short-baseline differencing can amplify position noise. The scalar alpha is consequently not identified as a correction for a physical sensor scale defect. It is simply the coefficient that minimizes the fitting displacement error under our source observation and interpolation convention. A value near one does not demonstrate perfect reference calibration.

The numerical concerns in Higham, 2002 motivate retaining coordinate units and avoiding a hidden change of normalization between families. Hartley and Zisserman, 2004 distinguishes the geometry of camera projection from world-frame translation. Since no image points are projected here, the resulting translation error cannot be relabeled as a pixel reprojection error or a perceptual video-quality effect.

Comparing predictions at the file level

We form two paired error contrasts in each held file: fitted velocity minus holding position, and fitted velocity minus fixed velocity. Each contrast is summarized through an intercept-only OLS regression over forty-two files. Both models use HC3 standard errors and t41 intervals. There are no controls or fixed effects, and intercept-only R-squared is zero apart from numerical rounding.

White, 1980 supplies the heteroskedasticity-robust rationale. Cameron and Miller, 2015 discusses why dependence can require care beyond treating every repeated observation as independent, which motivates avoiding thousands of nominal anchor-level replications. Averaging by file does not resolve all relationships among related recordings, so the intervals remain conditional working descriptions rather than population guarantees.

Diebold and Mariano, 1995 concerns predictive-accuracy comparison, but we do not claim to implement its time-series test here. Newey and West, 1987 provides a serial-covariance approach that would be relevant to an explicitly timed residual-series design. Our fitted models instead use file summaries with a declared source-family split. Naming those alternatives prevents a generic forecast-comparison label from obscuring the estimator actually executed.

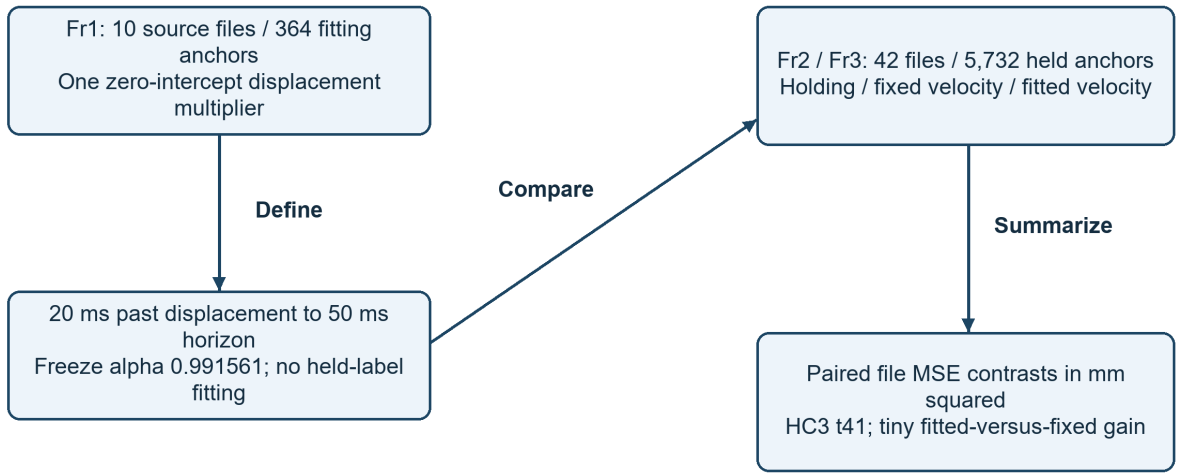


Figure 1. *Twenty-millisecond past motion supports a fifty-millisecond forecast. Held future positions supply labels only. The main comparison distinguishes motion extrapolation from the smaller benefit of coefficient fitting.*

Strong extrapolation, a small calibration increment

The equal-held-file mean squared error was 123.892240 mm² when position was held fixed. Ordinary constant velocity reduced it to 2.610790 mm². The fitted multiplier gave 2.591102 mm². Recent displacement therefore accounted for nearly all of the measured improvement relative to the stationary reference.

Model 1 estimates fitted-minus-holding error of -121.301138 mm², with standard error 15.978801 and interval -153.570981 to -89.031295. Model 2 estimates fitted-minus-fixed-velocity error of -0.019689 mm², with standard error 0.003530 and interval -0.026817 to -0.012560. The second difference is detectable within the working file-level analysis but small on the scale of the already low constant-velocity error.

Table 1. *Paired forecast MSE contrasts in squared millimeters. N equals forty-two held file means. Intercept-only OLS, HC3 covariance, t41 intervals; no controls or fixed effects. The coefficient was fitted only on fr1 and is held fixed for both models.*

Model	Term	Estimate (SE)	95% CI	N
M1	Mean outcome/contrast	-121.301138 (15.978801)	[-153.570981, -89.031295]	42
M1: fit	R-squared = 0.0000	-	-	42
M2	Mean outcome/contrast	-0.019689 (0.003530)	[-0.026817, -0.012560]	42
M2: fit	R-squared = 0.0000	-	-	42

Model labels: M1 = Fitted velocity minus holding-position MSE, mm squared; M2 = Fitted velocity minus fixed-velocity MSE, mm squared.

The two held families show the same qualitative pattern descriptively. Fr2 file means were about 109.38 mm² for holding position and 2.54 mm² for fitted velocity. Fr3 means were about 137.08 and 2.64 mm², respectively. Those family summaries do not constitute a fitted test of between-family differences, and the observed transfer cannot be attributed to one known physical source of domain change.

The multiplier's improvement over fixed velocity is about three quarters of one percent of fixed velocity's mean error. We report this scale alongside its interval because statistical distinguishability is not the same as engineering importance. A more complex calibration mechanism would need to justify its own incremental value, implementation cost, and behavior under longer or less regular motion than the present comparison covers.

Why the reference choice changes the apparent achievement

Holding position is a useful reference when no reliable motion history is available. In this experiment, however, the predictor is given a supported twenty-millisecond history by construction. A comparison only with holding position would credit the fitted model with the elementary information in that history as well as with its estimated coefficient. The fixed-velocity reference removes that ambiguity: it uses the same history and forecast horizon while declining to estimate the scalar.

The two comparisons therefore answer different engineering questions. The large first contrast shows that recent motion carries predictive information over the short horizon in these recorded trajectories. The much smaller second contrast isolates the incremental contribution of this particular coefficient fitting step. Neither result establishes that a larger learned model is necessary, nor that a fitted coefficient would remain useful when history is noisy, motion changes abruptly, or the destination horizon is extended.

The fitting objective also differs from the reporting weights. Alpha minimizes total squared displacement error across the fr1 fitting anchors and their three coordinates. A long or energetic fitting file can contribute more to that objective than a short or nearly stationary one. Held evaluation instead averages within file and then weights files equally. This is a declared design choice rather than an assertion that the fitting and evaluation distributions have identical importance weights.

Error is based on the squared norm of a three-dimensional displacement residual, which sums the three squared coordinate errors. It is not the average of three separately normalized component errors. The conversion to squared millimeters occurs after that vector error is formed and multiplies every method equally. This keeps the very small fitted-versus-fixed increment readable without changing its ranking, interval, or underlying predictive task.

The source labels can be linearly interpolated around a requested future time, but the predictor never receives the future bracket. This makes the comparison implementable with respect to its input timing while leaving the outcome definition dependent on a reference interpolation convention. If actual motion between source observations curved or accelerated differently from that convention, the experiment would not detect the difference. A physical ground-truth claim stronger than the recorded reference would require additional measurement.

Transport is conditional on the forecasting task

Forecasting guidance in Hyndman and Koehler, 2006 and rolling-evaluation principles in Tashman, 2000 emphasize horizon and information availability. Our favorable result is at fifty milliseconds with a twenty-millisecond motion history. It cannot establish the same ordering at a longer horizon, during acceleration, after large observation gaps, or when current pose estimates are themselves produced by a noisy visual estimator.

Ovadia et al., 2019 and Roberts et al., 2017 help distinguish a frozen evaluation across recorded domains from a claim about all future environments. The coefficient was not refitted on fr2 or fr3, but the held files belong to the same benchmark family and use the same pose representation. Their related collection context is a more limited transfer test than deployment in another sensing or navigation system.

The loss-function perspective in Gneiting and Raftery, 2007 matters as well. Mean squared translation error emphasizes larger errors and differs from a thresholded success rate, an absolute-error criterion, or image-plane safety. Robust estimation discussed in Rousseeuw, 1984 would produce another fitting objective; no such alternative was selected after inspecting the held outcomes. Cook, 1977 motivates retaining file-level results rather than removing influential recordings to enlarge the apparent gain.

Evidence retained and evidence still needed

The measurement-scale discipline in Stevens, 1946 keeps squared millimeters distinct from distance, speed, and dimensionless parallax. Documentation in Gebru et al., 2021 and reuse principles in Wilkinson et al., 2016 motivate preserving anchor rules, source-file membership, the fitted coefficient, predictions, units, and result tables. Those records make the calculation reproducible without claiming that the original source motion is error free.

Uncertainty from fitting alpha is not propagated into the displayed intervals; they condition on the fitted value. Source-family dependence and reference-position noise are also unresolved. A stronger transfer experiment would repeat fitting across independently acquired source cohorts and evaluate against new motion with matched measurement conditions. Here the bounded conclusion is simpler: recent-motion extrapolation transferred successfully within the acquired TUM families, while estimating its nearly unit multiplier added only a small measured benefit.

References

1. Tashman, Leonard J.. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting. https://doi.org/10.1016/s0169-2070(00)00065-0.
2. Kapoor, Sayash; Narayanan, Arvind. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns. https://doi.org/10.1016/j.patter.2023.100804.
3. Sturm, Jürgen; Engelhard, Nikolas; Endres, Felix; Burgard, Wolfram; Cremers, Daniel. (2012). A benchmark for the evaluation of RGB-D SLAM systems. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. https://doi.org/10.1109/IROS.2012.6385773.
4. TUM Computer Vision Group. (2026). RGB-D SLAM Dataset: File Formats [Official documentation; accessed8November2026, captureyearnotbenchmarkyear]. https://cvg.cit.tum.de/data/datasets/rgbd-dataset/file_formats.
5. Ovadia, Yaniv, et al. “Can You Trust Your Model’s Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift.” *Advances in Neural Information Processing Systems*, vol. 32, 2019. https://arxiv.org/abs/1906.02530.
6. Hyndman, Rob J. and Anne B. Koehler. “Another look at measures of forecast accuracy.” *International Journal of Forecasting*, vol. 22, no. 4, 2006, pp. 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001.
7. Gneiting, Tilmann; Raftery, Adrian E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. https://doi.org/10.1198/016214506000001437.
8. Schubert, David; Goll, Thore; Demmel, Nikolaus; Usenko, Vladyslav; Stückler, Jörg; Cremers, Daniel. (2018). The TUM VI Benchmark for Evaluating Visual-Inertial Odometry. arXiv:1804.06120. https://arxiv.org/abs/1804.06120.
9. Yohann, Cabon,; Naila, Murray,; Martin, Humenberger,. (2020). Virtual KITTI 2. https://arxiv.org/abs/2001.10773.
10. Burri, Michael; Nikolic, Janosch; Gohl, Pascal; Schneider, Thomas; Rehder, Joern; Omari, Sammy; Achtelik, Markus W; Siegwart, Roland. (2016). The EuRoC micro aerial vehicle datasets. The International Journal of Robotics Research. https://doi.org/10.1177/0278364915620033.
11. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.
12. Hartley, Richard; Zisserman, Andrew. (2004). Multiple View Geometry in Computer Vision. Cambridge University Press. https://doi.org/10.1017/cbo9780511811685.
13. Roberts, David R., et al. “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.” *Ecography*, vol. 40, no. 8, 2017, pp. 913-929. https://doi.org/10.1111/ecog.02881.
14. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

15. Zhang, Zichao; Scaramuzza, Davide. (2018). A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). https://doi.org/10.1109/IROS.2018.8593941.

16. Stevens, S. S.. (1946). On the Theory of Scales of Measurement. Science. https://doi.org/10.1126/science.103.2684.677.

17. Rousseeuw, Peter J.. (1984). Least Median of Squares Regression. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1984.10477105.

18. Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

19. Dosovitskiy, Alexey; Ros, German; Codevilla, Felipe; Lopez, Antonio; Koltun, Vladlen. (2017). CARLA: An Open Urban Driving Simulator. arXiv:1711.03938. https://arxiv.org/abs/1711.03938.

20. White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

21. Barfoot, Timothy D.. (2017). State Estimation for Robotics. Cambridge University Press, first edition. https://doi.org/10.1017/9781316671528.

22. Cameron, A. Colin, and Douglas L. Miller. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.

23. Meng, Zijie. (2026). ParaScale: Scale-Calibrated Camera-Motion Transfer via a Gauge-Invariant Parallax Number. https://arxiv.org/abs/2606.19805.

24. Diebold, Francis X.; Mariano, Roberto S.. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics. https://doi.org/10.1080/07350015.1995.10524599.

25. Blanco-Claraco, José Luis. (2021). A tutorial on SE(3) transformation parameterizations and on-manifold optimization. arXiv:2103.15980; arXivversionyear, originaltechnicalreport2010. https://arxiv.org/abs/2103.15980.

26. Geiger, A.; Lenz, P.; Urtasun, R.. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. 2012 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2012.6248074.

27. Higham, Nicholas J.. (2002). Accuracy and Stability of Numerical Algorithms: Second Edition. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718027.

28. Newey, Whitney K., and Kenneth D. West. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, vol. 55, no. 3, 1987, pp. 703-708. https://doi.org/10.2307/1913610.

29. Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; ’t Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18.

30. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.