Journal of Data Efficiency
Volume 1 · Number 1 · 2026

How Many Driving Positions Can Be Omitted? A Trajectory-Anchor Audit on KITTI

Owen Mason[1] · Sean Douglas[2] · James Harrison[3]

[1] Australia · [2] Sweden · [3] Ireland

Original publication: October 24, 2026 | Article 563

Abstract

On eleven KITTI driving sequences, retaining roughly half of the recorded positions yielded an equal-sequence omitted-frame interpolation RMSE of 0.011 metres. Retaining approximately one position in sixteen raised it to 0.317 metres, and every sequence showed increasing error across the tested spacings. The paired mean increase was 0.307 metres, with a sequence-level 95% interval of [0.238, 0.375]. This simple audit identifies a geometric cost of sparse trajectory anchors. It does not measure video compression efficiency: positions, pixels, camera orientation, and generated scene content require separate evidence.

Research draft based on deterministic transformations of public training trajectories. No video generator, learned compressor, or driving controller was executed.

How Many Driving Positions Can Be Omitted? A Trajectory-Anchor Audit on KITTI

Research draft based on deterministic transformations of public training trajectories. No video generator, learned compressor, or driving controller was executed.

Abstract

On eleven KITTI driving sequences, retaining roughly half of the recorded positions yielded an equal-sequence omitted-frame interpolation RMSE of 0.011 metres. Retaining approximately one position in sixteen raised it to 0.317 metres, and every sequence showed increasing error across the tested spacings. The paired mean increase was 0.307 metres, with a sequence-level 95% interval of [0.238, 0.375]. This simple audit identifies a geometric cost of sparse trajectory anchors. It does not measure video compression efficiency: positions, pixels, camera orientation, and generated scene content require separate evidence.

1. A resource decision before video synthesis

Driving-generation systems can be conditioned on a trajectory that contains many closely spaced positions. Fewer supplied positions appear cheaper to store, transmit, or expose to an operator, but a sparse sequence also leaves more of the path to be reconstructed. A smooth curve through a few points can look adequate without following the recorded path between them. Before attaching a claim about resource efficiency to this reduction, it is useful to ask how much positional information is lost by a fully specified, inexpensive reconstruction rule.

OmniDrive brings language, geometry, and multi-view video into a shared generative framework and discusses co-compression of its conditioning and video representations (Meng et al., 2026). That research setting motivates examining the geometry supplied to a model. Our experiment is a separate baseline on public position trajectories. It measures neither that framework's latent representation nor its reported performance. The research question is how omitted-frame positional error changes when the interval between retained trajectory anchors doubles, holding the original observations and interpolation procedure fixed.

We distinguish two propositions. P1 anticipates a positive within-sequence relationship between anchor spacing and translation reconstruction error. P2 concerns the directly interpretable difference between the sparsest and densest tested choices. These propositions refer to an information-reduction operation that we execute ourselves. They do not concern the causal effect of allocating tokens to a generative model, the perceptual quality of a rendered road scene, or the safety of a driving policy that uses the reconstructed path.

2. Trajectories are one component of a driving representation

Driving datasets contain more than interchangeable position lists. KITTI supplies a benchmark setting with ground-truth trajectories, while nuScenes and the Waymo Open Dataset organize richer multimodal perception resources (Geiger et al., 2012; Caesar et al., 2019; Sun et al., 2019). BDD100K and Argoverse 2 represent other collections with their own task and sampling choices (Yu et al., 2018; Wilson et al., 2023). We use only KITTI's eleven public training pose sequences. Mentioning other resources identifies a boundary for transportability rather than implying that their data were downloaded or included in our sample.

Simulation and generation introduce additional representational choices. CARLA provides an urban driving simulation platform; MagicDrive and DriveDreamer address controllable driving generation, and DriveDiffusion develops multi-view driving-scene synthesis (Dosovitskiy et al., 2017; Gao et al., 2023; Wang et al., 2023; Li et al., 2023). These systems motivate maintaining geometric controls when creating synthetic observations. A reconstruction error calculated from a position file alone cannot determine how any of them will render an omitted maneuver. A generator may smooth, correct, or amplify a control discrepancy depending on information absent from our experiment.

Panacea and Street Gaussians illustrate further approaches to producing or representing driving scenes (Wen et al., 2023; Yan et al., 2024). Their outputs have spatial, visual, and temporal properties that our scalar distance does not summarize. Likewise, BEVFormer aggregates multi-camera and temporal information, while planning-oriented autonomous-driving work connects perception with downstream tasks (Li et al., 2022; Hu et al., 2022). A geometry audit can identify an input discrepancy relevant to such pipelines without establishing its downstream consequence. That consequence requires an actual perception or planning experiment with a stated loss.

KITTI-360 also combines scene understanding with geometric information in a distinct dataset (Liao et al., 2021). Its existence reinforces the need to state which trajectory source is being used. Our eleven sequences are from the earlier odometry training release, and we do not join them to KITTI-360 imagery. Route characteristics, localization accuracy, and traffic conditions can differ across these collections. Treating all resources carrying a related benchmark name as one homogeneous sample would obscure the unit on which the current calculations depend.

3. Source poses and the omission operation

The downloaded pose archive contains eleven text files corresponding to sequences 00 through 10, with 23,201 native frames in total. Each row has twelve entries that reshape into a three-by-four pose matrix. We extract its translation column as a three-dimensional position in metres. All parsed values are finite. The ZIP integrity check passes, and the source file hash is recorded. We retain every available position in each of these sequences before constructing any reduced representation.

For each sequence we apply anchor strides of two, four, eight, and sixteen native frames. The first frame is retained, followed by every stride-th frame, and the final frame is always appended if it is not already selected. This rule prevents the reconstruction from extrapolating beyond its final anchor. It also creates a shorter terminal interval for some sequence lengths. We preserve that feature in the processing record rather than forcing every trajectory to have an artificial length divisible by sixteen.

Between anchors, each of the three position coordinates is reconstructed by linear interpolation in frame index. Retained coordinates are copied exactly. The Euclidean difference between a reconstructed and original position is then calculated at every omitted frame. The primary error is the square root of the mean squared Euclidean discrepancy over omitted frames only. Including exact anchors in the error denominator would mechanically reduce the reported average, particularly for the densest representation, so their exclusion is part of the metric definition.

This design produces four errors per sequence, giving 44 sequence-stride observations. The errors use different sets of omitted frames across strides. We therefore describe them as the cost of each complete omission policy, not as repeated measurements on an identical held-out frame set. Each sequence receives equal weight in the final cross-sequence summaries, irrespective of its native length. The analysis file records native frames, retained anchors, omitted-frame counts, mean-square error summaries, and upper-tail positional discrepancies for each policy.

4. Estimands and the limits of uncertainty

For P1 we regress the natural logarithm of sequence-level RMSE on the base-two logarithm of anchor stride and sequence indicators. The slope is the fitted change in log error when spacing doubles within the same recorded route. An intercept and ten sequence indicators absorb route-level average differences. This is a compact descriptive approximation across four omission levels, not a physical law governing every road maneuver. The outcome is strictly positive in all observed cells, so no arbitrary constant is required before taking its logarithm.

We cluster the covariance by sequence, use a finite-sample correction, and evaluate the slope with a Student reference distribution based on eleven clusters. Cluster-robust methodology clarifies why 44 cells should not be treated as 44 independent routes (Cameron and Miller, 2015). Heteroskedasticity-consistent covariance alone would not address the dependence between different omissions of the same trajectory (White, 1980). Even the chosen calculation remains uncertain with only eleven clusters. Its interval describes a hypothetical route-sampling interpretation, rather than randomness in our deterministic interpolation code.

For P2 we compute, within each sequence, the RMSE at stride sixteen minus the RMSE at stride two. An intercept-only regression on these eleven differences yields their equal-sequence mean and an ordinary Student interval with ten residual degrees of freedom. This model's R-squared is zero up to numerical rounding because there is no predictor beyond the intercept. It is included as a descriptive inferential summary, not presented as a fitted prediction model with explanatory power.

The distinction between observation count and independent information is familiar in structured validation and dependent processes (Roberts et al., 2017; Haslett and Raftery, 1989). We do not apply a frame-wise bootstrap that would ignore trajectory organization. Time-series bootstrap methods and heteroskedasticity-autocorrelation covariance offer alternative tools for appropriate time-indexed estimands (Bühlmann, 2002; Newey and West, 1987), but they are not silently substituted for the sequence-level analysis performed here. Figure 1 separates the transformation itself from the two summaries of its consequences.

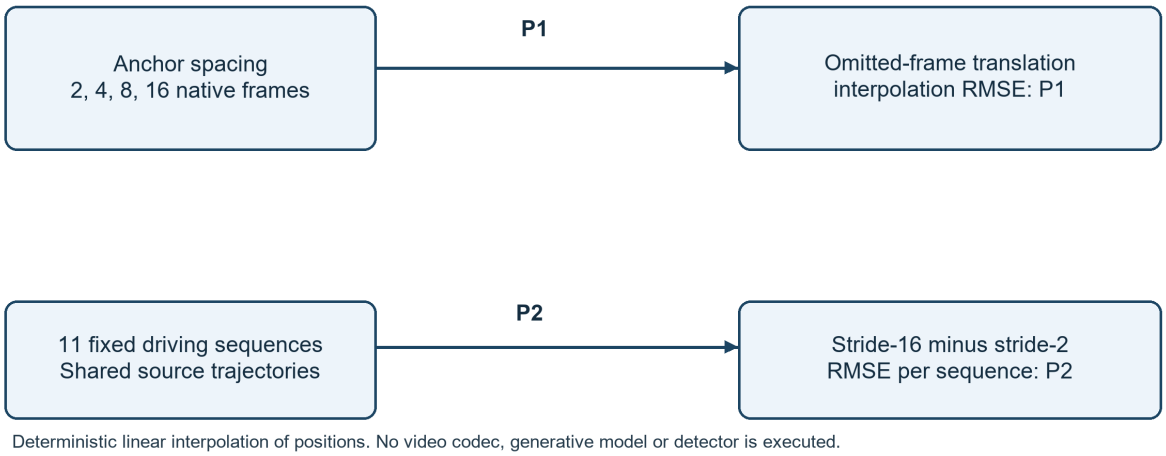


Figure 1. Positional information-reduction experiment. P1 summarizes the spacing-error relationship; P2 compares the two endpoint policies within each route. Arrows describe executed transformations and descriptive contrasts.

5. The observed spacing-error profile

The equal-sequence retained-anchor fraction is approximately 50.05% at stride two, 25.09% at stride four, 12.61% at stride eight, and 6.36% at stride sixteen. Small departures from exact fractions follow from the endpoint rule and sequence lengths. The corresponding mean omitted-frame RMSE values are 0.0106, 0.0253, 0.0864, and 0.3172 metres. Every sequence has larger RMSE at each successive tested stride. This consistency of direction is a property of the eleven observed trajectories, without an assumed probability of monotonicity in other routes.

The upper tail also grows. Averaging each sequence's 95th-percentile omitted-frame discrepancy gives 0.0205 metres for stride two and 0.7069 metres for stride sixteen. These are averages of within-sequence percentiles, not the percentile of all pooled frame errors. The distinction matters because a long route would otherwise dominate the summary. We report the construction explicitly so that a reader does not interpret the sparse-policy mean RMSE as a maximum error bound for every part of every route.

In Table 1, the fitted log-error slope for one doubling of spacing is 1.1283, with a sequence-clustered standard error of 0.0362 and a 95% interval of [1.0477, 1.2089]. Exponentiating the coefficient gives an approximate factor of 3.09 in the fitted RMSE relationship per doubling. That factor summarizes a common slope across four settings; it is not a guarantee that every adjacent pair of settings or every individual sequence exhibits precisely the same multiplier. The high in-sample R-squared includes route indicators and reflects a deliberately regular transformation experiment.

Table 1. Executed sequence-level estimates. The first model uses 44 sequence-stride cells, route indicators, and eleven sequence clusters with a t(10) reference. The second model uses eleven paired RMSE differences and an intercept only; its conventional Student uncertainty describes variation across sequences.

Model	Term	Estimate (SE)	95% CI	N
Log RMSE versus anchor stride	Log2(anchor stride): P1	1.1283 (0.0362)	[1.0477, 1.2089]	44
Log RMSE versus anchor stride: fit	R-squared = 0.9858	-	-	44
Stride-16 minus stride-2 RMSE	Mean RMSE difference: P2	0.3066 (0.0306)	[0.2384, 0.3748]	11
Stride-16 minus stride-2 RMSE: fit	R-squared = 0.0000	-	-	11

The endpoint contrast is 0.3066 metres, with a standard error of 0.0306 and a 95% interval of [0.2384, 0.3748]. It answers a more concrete question than the fitted slope: how much greater is the average omitted-frame error when retaining about one-sixteenth rather than one-half of native positions under this interpolation rule? It does not attach monetary cost or execution time to the retained fraction. Our calculation stores coordinates and counts anchors, rather than measuring network traffic or memory allocation inside a trained model.

6. What can be inferred about data requirements?

The empirical result establishes a tradeoff in this chosen representation: larger spacing saves anchors and loses positional detail. It does not identify a universally sufficient sampling rate. A sufficient rate would require an acceptable error criterion, a destination task, and a deployment distribution. Work on forecast accuracy emphasizes that metrics serve particular comparison purposes, while proper scoring rules concern evaluated predictive distributions (Hyndman and Koehler, 2006; Gneiting and Raftery, 2007). Our geometric RMSE is neither a perceptual score nor a probability forecast, and those different quantities should not be substituted for it.

The source positions are themselves measurements. Small interpolation discrepancies may be affected by localization uncertainty, and the official odometry documentation cautions about evaluating very short subsequences. Measurement-error analysis provides a reason to avoid treating every centimetre-scale difference as exact physical truth (Carroll et al., 2006). No independent GPS precision estimate was introduced. We therefore interpret the fine-scale results as reconstruction of the supplied benchmark coordinates, with physical trajectory truth and sensor uncertainty outside the measured estimand.

Nor does high agreement of route shapes imply interchangeable coordinates for a downstream application. Agreement analysis distinguishes association from acceptable differences (Bland and Altman, 1986). A road-scene generator might tolerate some positional deviations while a collision calculation cannot. Without specified tolerances, our study cannot label any stride safe, high quality, or operationally optimal. It provides reproducible quantities with which a later application can state such a tolerance and test it, rather than supplying the missing criterion by assertion.

Endpoint handling also illustrates a general missing-information issue. Our interior omissions are intentionally constructed and reconstructible from known anchors; naturally missing sensor packets could have irregular gaps and different selection mechanisms. Missing-data methodology warns that one mechanism need not represent another (Little and Rubin, 2002). The current results should not be presented as evidence of robustness to dropped packets, adverse weather, or failure of a localization sensor. Such questions would require additional transformations and appropriate evidence about the source of the missingness.

7. Reuse and follow-up boundaries

A useful next experiment would compare interpolation rules on a declared held-out route set, using a development set to choose any tuning parameters. Model-selection bias is relevant if a researcher chooses the best rule after repeatedly examining its test errors (Varma and Simon, 2006). We did not tune splines, smoothing parameters, or adaptive anchor schedules here. Their potential benefits remain proposals. Influence diagnostics could also identify routes that dominate an average, but deleting an inconvenient route after seeing its error would change the target comparison (Cook, 1977).

Multiple alternative routes, errors, and interpolation choices could create a large family of comparisons. False-discovery-rate control offers one approach when that family is defined (Benjamini and Hochberg, 1995). Our two propositions and four fixed spacing policies are retained together, rather than using favorable significance to select a headline rate. The source files, retained counts, transformation code, and outcome definitions are preserved in a provenance record consistent with datasheet principles (Gebru et al., 2021). This makes later changes to the representation explicit and auditable.

8. Conclusion

The eleven recorded trajectories show that inexpensive linear interpolation preserves fine positional detail with dense anchors and loses progressively more detail as anchors are thinned. The clearest summary is the paired 0.307-metre increase between the endpoint policies, alongside the retained fractions that produce it. This is evidence about public trajectory coordinates under an executed omission rule. Connecting it to latent video compression, multi-view consistency, or planning performance remains a separate empirical task requiring those outputs to be measured.

References

1. Yan, Yunzhi; Lin, Haotong; Zhou, Chenxu; Wang, Weijie; Sun, Haiyang; Zhan, Kun; Lang, Xianpeng; Zhou, Xiaowei; Peng, Sida. (2024). Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting. arXiv:2401.01339. https://arxiv.org/abs/2401.01339.
2. Liao, Yiyi; Xie, Jun; Geiger, Andreas. (2021). KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D. arXiv:2109.13410. https://arxiv.org/abs/2109.13410.
3. 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.
4. Wang, Xiaofeng; Zhu, Zheng; Huang, Guan; Chen, Xinze; Zhu, Jiagang; Lu, Jiwen. (2023). DriveDreamer: Towards Real-world-driven World Models for Autonomous Driving. arXiv:2309.09777. https://arxiv.org/abs/2309.09777.
5. 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.
6. Yu, Fisher, et al. "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning" *arXiv preprint arXiv:1805.04687*, 2018. https://arxiv.org/abs/1805.04687.
7. Meng, Zijie; Liu, Yufei; Ma, Chengqian; Li, Zhiyu; Liu, Jiyuan; Nie, Wenhua; Wei, Bingcai; Chen, Shuqin; Xu, Weichen; Yuan, Jiquan; Zhang, Miao. (2026). OmniDrive: An LLM-Choreographed Multi-Agent World Model with Unified Latent Co-Compression for Multi-View Driving Video Generation. arXiv:2606.17536. https://arxiv.org/abs/2606.17536.
8. Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
9. Sun, Pei, et al. "Scalability in Perception for Autonomous Driving: Waymo Open Dataset" *arXiv preprint arXiv:1912.04838*, 2019. https://arxiv.org/abs/1912.04838.
10. Li, Xiaofan; Zhang, Yifu; Ye, Xiaoqing. (2023). DrivingDiffusion: Layout-Guided multi-view driving scene video generation with latent diffusion model. arXiv:2310.07771. https://arxiv.org/abs/2310.07771.
11. Wilson, Benjamin; Qi, William; Agarwal, Tanmay; Lambert, John; Singh, Jagjeet; Khandelwal, Siddhesh; Pan, Bowen; Kumar, Ratnesh; Hartnett, Andrew; Pontes, Jhony Kaesemodel; Ramanan, Deva; Carr, Peter; Hays, James. (2023). Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting. arXiv:2301.00493. https://arxiv.org/abs/2301.00493.
12. Bühlmann, Peter. (2002). Bootstraps for Time Series. Statistical Science. https://doi.org/10.1214/ss/1023798998.
13. Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.
14. Gao, Ruiyuan; Chen, Kai; Xie, Enze; Hong, Lanqing; Li, Zhenguo; Yeung, Dit-Yan; Xu, Qiang. (2023). MagicDrive: Street View Generation with Diverse 3D Geometry Control. arXiv:2310.02601. https://arxiv.org/abs/2310.02601.
15. 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.
16. 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.
17. Caesar, Holger, et al. "nuScenes: A multimodal dataset for autonomous driving" *arXiv preprint arXiv:1903.11027*, 2019. https://arxiv.org/abs/1903.11027.

18. Hu, Yihan; Yang, Jiazhi; Chen, Li; Li, Keyu; Sima, Chonghao; Zhu, Xizhou; Chai, Siqi; Du, Senyao; Lin, Tianwei; Wang, Wenhai; Lu, Lewei; Jia, Xiaosong; Liu, Qiang; Dai, Jifeng; Qiao, Yu; Li, Hongyang. (2022). Planning-oriented Autonomous Driving. arXiv:2212.10156. https://arxiv.org/abs/2212.10156.

19. Li, Zhiqi; Wang, Wenhai; Li, Hongyang; Xie, Enze; Sima, Chonghao; Lu, Tong; Yu, Qiao; Dai, Jifeng. (2022). BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers. arXiv:2203.17270. https://arxiv.org/abs/2203.17270.

20. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

21. 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.

22. 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.

23. Wen, Yuqing; Zhao, Yucheng; Liu, Yingfei; Jia, Fan; Wang, Yanhui; Luo, Chong; Zhang, Chi; Wang, Tiancai; Sun, Xiaoyan; Zhang, Xiangyu. (2023). Panacea: Panoramic and Controllable Video Generation for Autonomous Driving. arXiv:2311.16813. https://arxiv.org/abs/2311.16813.

24. 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.

25. 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.

26. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.

27. 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.

28. Haslett, John; Raftery, Adrian E.. (1989). Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. Applied Statistics. https://doi.org/10.2307/2347679.

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

30. Bland, J. Martin, and Douglas G. Altman. "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement." *The Lancet*, vol. 327, no. 8476, 1986, pp. 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8.