

1 A Demo Videos

2 To further showcase the effectiveness of OriGS in real-world scenarios, we include additional video
3 demonstrations of in-the-wild 4D reconstruction in the supplementary material. Following the
4 same experimental setup as in the main paper, we present visualizations of both training sequence
5 reconstruction and novel view synthesis, comparing OriGS with the recent state-of-the-art 4D
6 reconstruction system. In addition, we provide video demonstrations of the OriGS variants designed
7 in our ablation study, illustrating how individual components affect reconstruction quality. These
8 results collectively highlight the temporal consistency and structural fidelity of OriGS across diverse
9 scenes and motion patterns.

10 B Implementation Details

11 **Oriented Anchor Initialization.** We follow a similar initialization pipeline provided in recent
12 open-source codebases of dynamic reconstruction frameworks [12, 8]. Specifically, we first extract
13 long-range 2D point trajectories using SpatialTracker [13] and obtain per-frame depth maps using
14 DepthCrafter [5]. These 2D correspondences are lifted into 3D space using estimated camera
15 parameters from bundle adjustment [8]. The resulting 3D trajectories are then converted into oriented
16 anchors, whose initial principal directions are estimated by applying PCA over early-frame motion.
17 In our experiments, we set the temporal window size $W = 5$. This step provides a stable estimate of
18 forward direction, which is then temporally propagated to form our Global Orientation Field.

19 **Hyper-Gaussian Parameterization.** We adopt the 3D Gaussian Splatting (3DGS) [7] as the base rep-
20 resentation and extend each Gaussian primitive with a hyper-Gaussian parameterized by a canonical
21 state μ_{ξ} and a Cholesky-decomposed covariance $\Sigma_{\xi} = \mathbf{L}_{\xi}\mathbf{L}_{\xi}^{\top}$, where \mathbf{L}_{ξ} is a lower-triangular matrix
22 [3, 4, 2]. To improve computational efficiency, we avoid constructing the full covariance matrix Σ_{ξ} .
23 Instead, we further factor it into two learnable subcomponents: (i) a marginal covariance $\Sigma_{(t, \mathbf{O})}$
24 over temporal and orientational variables, which inherits the Cholesky-decomposed parameterization
25 to ensure positive semi-definiteness, and (ii) a cross-covariance term $\Sigma_{(\Delta \mathbf{p}, \Delta \mathbf{g}), (t, \mathbf{O})}$ that captures
26 the correlation between dynamic geometry $(\Delta \mathbf{p}, \Delta \mathbf{g})$ and temporal-orientational context. During
27 training, we jointly optimize these covariance terms along with the canonical mean μ_{ξ} and the
28 original 3DGS parameters (position, scale, rotation, color, and opacity). This factorization strategy
29 circumvents the overhead of full matrix parameterization while retaining the necessary statistical
30 coupling for efficient conditioned slicing.

31 **Optimization.** We optimize the model using the differentiable rasterization-based rendering pipeline
32 from 3DGS [7], adapted to support high-dimensional slicing and dynamic modulation. The loss
33 function combines: (i) photometric loss [7], an RGB reconstruction loss between rendered and
34 ground-truth images, (ii) 2D correspondence loss, alignment to long-range 2D tracks and depth
35 priors from foundation models, as in [12, 8, 9, 10], and (iii) deformation regularization, an as-rigid-
36 as-possible constraint [13, 11, 1, 6] applied to anchor-guided transformations. For scalability and
37 high-fidelity reconstruction, we also design a pruning-and-densification scheme: Gaussian primitives
38 with low opacity are pruned, while spatial regions exhibiting high response gradients with respect
39 to $\mu_{\mathbf{p}}$ and $\mu_{\Delta \mathbf{p}}$ are densified through local duplication. All experiments are conducted on a single
40 NVIDIA RTX A6000 GPU, and the full optimization of a typical scene takes approximately 0.5–2
41 hours, depending on video length and complexity.

42 C Limitations

43 While OriGS demonstrates promising results, it also presents certain limitations. (i) In scenes
44 dominated by simple motion, such as near-rigid translation along a fixed axis, orientation tends
45 to remain approximately constant over time. Consequently, conditioning on these orientational
46 cues may offer limited additional benefit, yet our framework still incurs unnecessary optimization
47 complexity of high-dimensional modeling. Future work could explore adaptive mechanisms that
48 modulate the use or dimensionality of orientation cues based on the complexity of motion in the
49 scene. (ii) OriGS leverages 2D priors such as point trajectories and depth from vision foundation
50 models to initialize and optimize the orientation field. The reliability of these priors can affect the
51 quality of 4D reconstruction. This reflects a deeper challenge tied to the development of reliable
52 visual priors, which has long been a cornerstone of progress in computer vision research.

53 D Broader Impacts

54 OriGS provides a unified framework for reconstructing dynamic scenes from casual monocular
55 videos, which can benefit various real-world applications. For instance, in virtual and augmented
56 reality, OriGS can reconstruct dynamic environments or actors from consumer-grade video input
57 alone, lowering the barrier to immersive content creation. Our work can also support robotics and
58 behavioral analysis, especially where low-cost monocular capture is the only viable option. However,
59 as with other scene reconstruction techniques, OriGS could potentially be misused for unauthorized
60 replication of environments or individuals. While our framework is not designed for such misconduct,
61 responsible usage should consider privacy and ethical concerns.

62 References

- 63 [1] Marc Alexa, Daniel Cohen-Or, and David Levin. As-rigid-as-possible shape interpolation. In
64 *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 165–172. 2023.
- 65 [2] Stavros Diolatzis, Tobias Zirr, Alexander Kuznetsov, Georgios Kopanas, and Anton Kaplanyan.
66 N-dimensional gaussians for fitting of high dimensional functions. In *ACM SIGGRAPH 2024*
67 *Conference Papers*, pages 1–11, 2024.
- 68 [3] Zhongpai Gao, Benjamin Planche, Meng Zheng, Anwesa Choudhuri, Terrence Chen, and Ziyang
69 Wu. 6dgs: Enhanced direction-aware gaussian splatting for volumetric rendering. In *ICLR*,
70 2025.
- 71 [4] Zhongpai Gao, Benjamin Planche, Meng Zheng, Anwesa Choudhuri, Terrence Chen, and Ziyang
72 Wu. 7dgs: Unified spatial-temporal-angular gaussian splatting. *arXiv preprint arXiv:2503.07946*,
73 2025.
- 74 [5] Wenbo Hu, Xiangjun Gao, Xiaoyu Li, Sijie Zhao, Xiaodong Cun, Yong Zhang, Long Quan, and
75 Ying Shan. Depthcrafter: Generating consistent long depth sequences for open-world videos.
76 *arXiv preprint arXiv:2409.02095*, 2024.
- 77 [6] Takeo Igarashi, Tomer Moscovich, and John F Hughes. As-rigid-as-possible shape manipulation.
78 *ACM transactions on Graphics (TOG)*, 24(3):1134–1141, 2005.
- 79 [7] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian
80 splatting for real-time radiance field rendering. In *ACM Trans. Graph.*, 2023.
- 81 [8] Jiahui Lei, Yijia Weng, Adam Harley, Leonidas Guibas, and Kostas Daniilidis. Mosca: Dynamic
82 gaussian fusion from casual videos via 4d motion scaffolds. In *CVPR*, 2025.
- 83 [9] Yiming Liang, Tianhan Xu, and Yuta Kikuchi. Himor: Monocular deformable gaussian
84 reconstruction with hierarchical motion representation. In *CVPR*, 2025.
- 85 [10] Jongmin Park, Minh-Quan Viet Bui, Juan Luis Gonzalez Bello, Jaeho Moon, Jihyong Oh, and
86 Munchurl Kim. Splinesg: Robust motion-adaptive spline for real-time dynamic 3d gaussians
87 from monocular video. In *CVPR*, 2025.
- 88 [11] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on*
89 *Geometry processing*, volume 4, pages 109–116. Citeseer, 2007.
- 90 [12] Colton Stearns, Adam Harley, Mikaela Uy, Florian Dubost, Federico Tombari, Gordon Wet-
91 zstein, and Leonidas Guibas. Dynamic gaussian marbles for novel view synthesis of casual
92 monocular videos. In *SIGGRAPH Asia*, 2024.
- 93 [13] Yuxi Xiao, Qianqian Wang, Shangzhan Zhang, Nan Xue, Sida Peng, Yujun Shen, and Xiaowei
94 Zhou. Spatialtracker: Tracking any 2d pixels in 3d space. In *Proceedings of the IEEE/CVF*
95 *Conference on Computer Vision and Pattern Recognition*, pages 20406–20417, 2024.