GimbalDiffusion: Gravity-Aware Camera Control for Video Generation

UNIVERSITÉ LAVAL [1] Adobe [2]

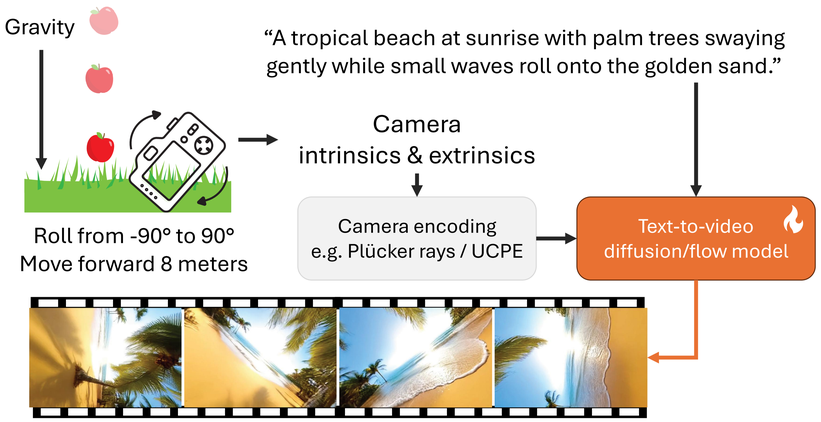
Frédéric Fortier-Chouinard[1], Yannick Hold-Geoffroy[2], Valentin Deschaintre[2], Matheus Gadelha[2], Jean-François Lalonde[1]

ECCV 2026 | MALMÖ | SEPT 8–12

Videos & code available

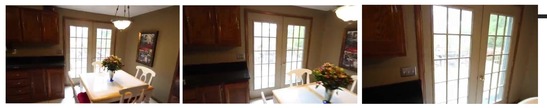
Goal

Control video generation models using a textual description and a precise camera path.



Typical training pipeline

Problem #1: Training on regular perspective videos



× Limited camera path diversity × Relative to first frame

It does not generalize to extreme camera paths.

Example result from UCPE — Correct — Incorrect

Problem #2: Annotating only the visible scene elements

× Doesn't include the general context, e.g. "clear sky"

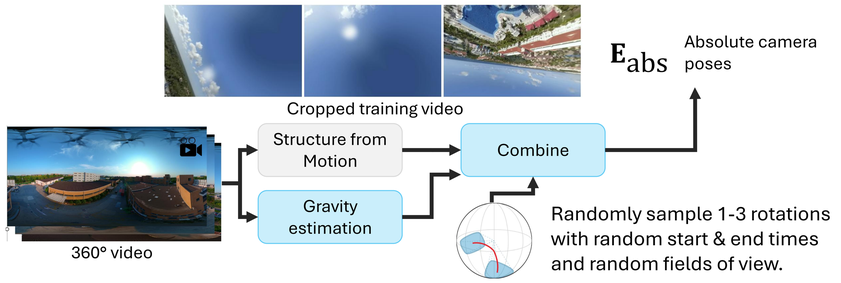
It forces every element in the prompt to be generated.

Static pitch angle +90° 0° -90°

Prompt :
"A dirt path through a lush, green forest [...] The sky is clear and blue, with a few scattered clouds [...]"

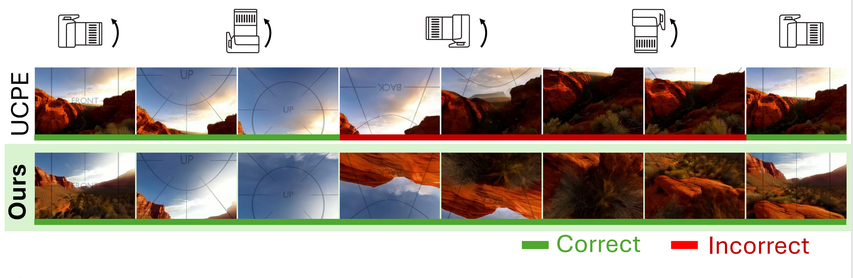
Method

Solution #1: Simulating diverse camera paths in 360° videos



✓ Diverse camera rotations ✓ Trajectories anchored to the gravity vector

Our method allows for generating more extreme rotations.



Correct Incorrect

Solution #2: Annotating the general scene context (null-pitch conditioning)

✓ General scene context is included in the caption

Qualitative impact of null-pitch conditioning

Prompt : "A dirt path through a lush, green forest [...] The sky is clear and blue, with a few scattered clouds [...]"

Static pitch angle +90° 0° -90°

Evaluation

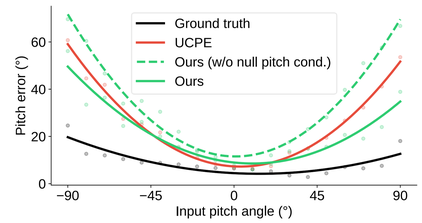
Video camera control benchmark **(new)**

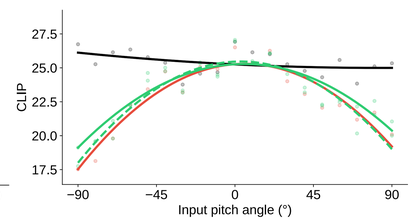
- Rebalanced SpatialVID videos to have a uniform pitch distribution.
- Randomly sampled rotations and kept the original translations.
- Use VGGT & Perspective Fields to estimate the generated camera path.

Method	PitchErr ↓	GravityErr ↓	RotErr ↓	TransErr ↓	FID ↓
CogVideoX-2B + Plücker camera encoder + ControlNet					
AC3D	39.09	44.13	29.27	**0.62**	113.0
Ours	**19.67**	**23.31**	**18.55**	0.67	**110.8**
WAN 2.1 1.3B + UCPE encoding + UCPE module training					
UCPE	16.25	18.88	18.38	**0.44**	**110.7**
Ours w/o null-pitch	19.13	22.07	**11.68**	0.48	114.3
Ours w/o absolute	51.16	56.92	15.15	0.53	115.7
Ours (default)	**13.95**	**16.95**	12.26	0.46	114.0

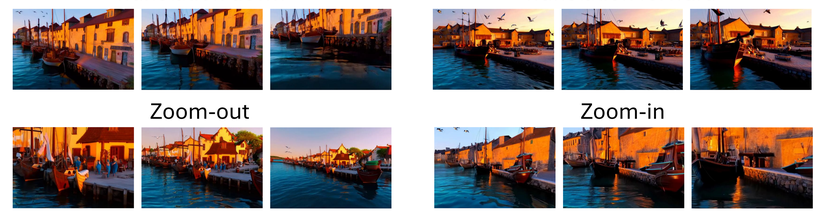
Prompt-camera entanglement benchmark **(new)**

- 20 panoramas from PolyHaven along with text annotation for forward-looking crop.
- Methods are evaluated to generate videos with a static camera, with pitch angles going from -90° to +90°, in 10° increments.





Other camera motions



Takeaways

- Regular perspective video datasets are limited in terms of camera path diversity; data augmentation is needed to improve camera control.
- Current captioning pipelines annotate only the visible content in the scene, limiting the ability of models to reason about the general scene context. Simply modifying the captioning improves performance.