# Novelty Assessment Report

**Paper**: Light-X: Generative 4D Video Rendering with Camera and Illumination Control

**PDF URL**: https://openreview.net/pdf?id=VBew6vESGL

**Venue**: ICLR 2026 Conference Submission

**Year**: 2026

**Report Generated**: 2026-01-01

## Abstract

Recent advances in illumination control extend image-based methods to video, yet still facing a trade-off between lighting fidelity and temporal consistency. Moving beyond relighting, a key step toward generative modeling of real-world scenes is the joint control of camera trajectory and illumination, since visual dynamics are inherently shaped by both geometry and lighting. To this end, we present Light-X, a video generation framework that enables controllable rendering from monocular videos with both viewpoint and illumination control. 1) We propose a disentangled design that decouples geometry and lighting signals: geometry and motion are captured via dynamic point clouds projected along user-defined camera trajectories, while illumination cues are provided by a relit frame consistently projected into the same geometry. These explicit, fine-grained cues enable effective disentanglement and guide high-quality illumination. 2) To address the lack of paired multi-view and multi-illumination videos, we introduce Light-Syn, a degradation-based pipeline with inverse-mapping that synthesizes training pairs from in-the-wild monocular footage. This strategy yields a dataset covering static, dynamic, and AI-generated scenes, ensuring robust training. Extensive experiments show that Light-X outperforms baseline methods in joint camera–illumination control. Besides, our model surpasses prior video relighting methods in text- and background-conditioned settings. Ablation studies further validate the effectiveness of the disentangled formulation and degradation pipeline. Code, data and models will be made public.



## Core Task Landscape

This paper addresses: **Joint Control of Camera Trajectory and Illumination in Video Generation**

A total of **35 papers** were analyzed and organized into a taxonomy with **17 categories**.

### Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Unified Multi-Modal Control Frameworks**
- **Camera Trajectory Control**
- **Illumination and Relighting Control**
- **Motion Control and Trajectory Specification**
- **3D Scene Representation and Rendering**
- **Cinematic and Optical Control**
- **Benchmarks, Taxonomies, and Evaluation Frameworks**
- **Survey and Review Literature**

### Complete Taxonomy Tree

- Joint Control of Camera Trajectory and Illumination in Video Generation Survey Taxonomy
- Unified Multi-Modal Control Frameworks
  - Disentangled Geometry-Lighting Control ★ (2 papers)
  - [0] Light-X: Generative 4D Video Rendering with Camera and Illumination Control (Anon et al., 2026) View paper
  - [1] Vidcraft3: Camera, object, and lighting control for image-to-video generation (Zheng Sixiao, 2025) View paper
  - Implicit Joint Control via Unified Conditioning (2 papers)
  - [2] Uni3C: Unifying Precisely 3D-Enhanced Camera and Human Motion Controls for Video Generation (Cao Chenjie, 2025) View paper
  - [5] Direct-a-Video: Customized Video Generation with User-Directed Camera Movement and Object Motion (Shiyuan Yang, 2024) View paper
- Camera Trajectory Control
  - 3D-Consistent Camera Control (3 papers)
  - [3] CamCo: Camera-Controllable 3D-Consistent Image-to-Video Generation (Xu, 2024) View paper
  - [4] Gen3c: 3d-informed world-consistent video generation with precise camera control (Xuanchi Ren, 2025) View paper
  - [21] FloVD: Optical Flow Meets Video Diffusion Model for Enhanced Camera-Controlled Video Synthesis (Wonjoon Jin, 2025) View paper
  - Decoupled Camera-Scene Dynamics (2 papers)
  - [22] Enhancing Human-Computer Interaction Through Decoupling Motion and Camera Control in Human-Centric Video Generation (G.T.Y. Chen, 2025) View paper
  - [23] BulletTime: Decoupled Control of Time and Camera Pose for Video Generation (Yiming Wang, 2025) View paper
  - Camera Control for Specialized Domains (2 papers)
  - [11] Beyond Static Scenes: Camera-controllable Background Generation for Human Motion (Yao Mingshuai, 2025) View paper
  - [32] MyGo: Consistent and Controllable Multi-View Driving Video Generation with Camera Control (Yao Yining, 2024) View paper
  - Post-Capture Camera Trajectory Editing (2 papers)

- [28] ReCapture: Generative Video Camera Controls for User-Provided Videos using Masked Video Fine-Tuning (David Junhao Zhang, 2024) View paper
  - [35] Simultaneous Camera Path Optimization and Distraction Removal for Improving Amateur Video. (Fang-Lue Zhang, 2016) View paper
- Illumination and Relighting Control
  - Portrait and Human-Centric Relighting (3 papers)
  - [7] High-Fidelity Relightable Monocular Portrait Animation with Lighting-Controllable Video Diffusion Model (Mingtao Guo, 2025) View paper
  - [13] Real-time 3d-aware portrait video relighting (Ziqi Cai, 2024) View paper
  - [20] Uniavatar: Taming lifelike audio-driven talking head generation with comprehensive motion and lighting control (Sun Wenzhang, 2024) View paper
  - Scene-Level Relighting and View Synthesis (2 papers)
  - [8] Neural light transport for relighting and view synthesis (Xiu-Ming Zhang, 2021) View paper
  - [19] Free-viewpoint indoor neural relighting from multi-view stereo (Philip, 2021) View paper
  - Volumetric Capture with Lighting Variation (2 papers)
  - [26] Virtually Being: Customizing Camera-Controllable Video Diffusion Models with Multi-View Performance Captures (Xu, 2025) View paper
  - [27] Virtually Being: Customizing Camera-Controllable Video Diffusion Models with Volumetric Performance Captures (Yuancheng Xu, 2025) View paper
- Motion Control and Trajectory Specification
  - Sparse and Dense Motion Trajectory Control (2 papers)
  - [16] Motion Prompting: Controlling Video Generation with Motion Trajectories (Daniel Geng, 2024) View paper
  - [17] Motionstream: Real-time video generation with interactive motion controls (Shin, 2025) View paper
  - Optical Flow-Based Motion Representation (1 papers)
  - [9] Free-form motion control: A synthetic video generation dataset with controllable camera and object motions (Ding, 2025) View paper
- 3D Scene Representation and Rendering
  - Procedural and Parametric 3D Scene Generation (2 papers)
  - [12] VideoFrom3D: 3D Scene Video Generation via Complementary Image and Video Diffusion Models (Kim, 2025) View paper
  - [33] Scene Co-pilot: Procedural Text to Video Generation with Human in the Loop (Zhaofang Qian, 2024) View paper
  - NeRF-Based Video Reconstruction and Enhancement (2 papers)
  - [24] Enhancing NeRFs for High-Quality Indoor Video Generation: A Study on Parameterization and Recording Methods (Tafnes Silva Barbosa, 2025) View paper
  - [34] Augmenting Visual Memories (Holynski, 2023) View paper
  - Multi-Layer Segmentation and Reconstruction (1 papers)
  - [25] Joint multi-layer segmentation and reconstruction for free-viewpoint video applications (Jean-Yves Guillemaut, 2011) View paper
- Cinematic and Optical Control (2 papers)
  - [10] Motionbooth: Motion-aware customized text-to-video generation (Kai Chen, 2024) View paper
  - [30] AKiRa: Augmentation Kit on Rays for optical video generation (Xi Wang, 2024) View paper
- Benchmarks, Taxonomies, and Evaluation Frameworks (2 papers)
  - [6] Stable Cinemetrics: Structured Taxonomy and Evaluation for Professional Video Generation (Entezari, 2025) View paper
  - [18] CineTechBench: A Benchmark for Cinematographic Technique Understanding and Generation (Wang Xinran, 2025) View paper
- Survey and Review Literature (4 papers)
  - [14] Reviewing Intelligent Cinematography: AI research for camera-based video production (Azzarelli, 2024) View paper
  - [15] Controllable video generation: A survey (Ma, 2025) View paper
  - [29] Virtual Cinematography and Camera Control for Film and Video Production (Mukesh Joshi, 2023) View paper
  - [31] Video Production with Generative AI (Brent Rabowsky, 2025) View paper

### Narrative

Core task: joint control of camera trajectory and illumination in video generation. The field has evolved into several distinct branches that reflect different emphases in controllable video synthesis. Unified Multi-Modal Control Frameworks aim to integrate multiple control signals—such as camera motion, lighting conditions, and scene geometry—into a single generation pipeline, often disentangling these factors to enable independent manipulation. Camera Trajectory Control focuses specifically on specifying and executing camera paths, whether through explicit pose sequences or higher-level cinematic directives. Illumination and Relighting Control addresses the challenge of manipulating lighting after capture or during synthesis, drawing on neural rendering and light transport modeling. Motion Control and Trajectory Specification explores how to represent and guide object or scene motion, while 3D Scene Representation and Rendering leverages volumetric or neural scene models to support view synthesis. Cinematic and Optical Control targets film-like aesthetics, and Benchmarks, Taxonomies, and Evaluation Frameworks provide the infrastructure for systematic comparison. Representative works such as CamCo[3] and Direct-a-Video[5] illustrate camera-centric approaches, while Relightable Portrait Animation[7] and Neural Light Transport[8] exemplify lighting-focused methods.

Recent efforts have increasingly sought to unify these control dimensions rather than treating them in isolation. A handful of works, including Uni3C[2] and Gen3c[4], demonstrate multi-modal frameworks that coordinate camera, lighting, and motion signals within a single diffusion or neural rendering backbone. Light-X[0] sits within this emerging cluster of disentangled geometry-lighting control methods, emphasizing the separation of camera trajectory from illumination changes so that each can be adjusted independently during generation. This contrasts with earlier camera-only approaches like CamCo[3], which prioritize pose control but do not explicitly model lighting variation, and with relighting-focused methods such as Relightable Portrait Animation[7], which handle illumination but typically assume fixed viewpoints. By disentangling these two axes, Light-X[0] and its close neighbor Vidcraft3[1] address a key challenge in achieving flexible, cinematic video synthesis where both viewpoint and lighting can be freely manipulated.

## Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

### 1. Vidcraft3: Camera, object, and lighting control for image-to-video generation

**Authors**: Zheng Sixiao, Sixiao Zheng, Z. Y. Peng, Zhou Yanpeng, Yanpeng Zhou, et al. (16 authors total) | **Year/Venue**: 2025 | **URL**: View paper

### Abstract

Controllable image-to-video (I2V) generation transforms a reference image into a coherent video guided by user-specified control signals. In content creation workflows, precise and simultaneous control over camera motion, object motion, and lighting direction enhances both accuracy and flexibility. However, existing approaches typically treat these control signals separately, largely due to the scarcity of datasets with high-quality joint annotations and mismatched control spaces across modaliti...


### Relationship Analysis

Both papers belong to the Disentangled Geometry-Lighting Control category, employing explicit separation of geometric and photometric signals through distinct encoding pathways. They overlap in addressing joint camera trajectory and illumination control for video generation using point cloud-based geometric representations and separate lighting conditioning mechanisms. However, Light-X focuses on monocular video relighting with a degradation-based training pipeline (Light-Syn) and uses IC-Light for single-frame relighting projected into 3D space, while VidCraft3 emphasizes image-to-video generation with simultaneous camera, object motion, and lighting direction control, introducing a VLD dataset and a three-stage training strategy without requiring fully joint annotations.

## Contributions Analysis

**Overall novelty summary.** The paper introduces Light-X, a framework for joint camera trajectory and illumination control in video generation, using a disentangled design that separates geometry (via dynamic point clouds) from lighting (via relit frames). Within the taxonomy, it resides in the 'Disentangled Geometry-Lighting Control' leaf under 'Unified Multi-Modal Control Frameworks'. This leaf contains only two papers, including the original work and one sibling (Vidcraft3), indicating a sparse and emerging research direction. The taxonomy shows that most prior work addresses camera or lighting control separately, making this joint disentangled approach relatively uncommon.

The taxonomy reveals that neighboring leaves include 'Implicit Joint Control via Unified Conditioning' (two papers) and broader branches like 'Camera Trajectory Control' (nine papers) and 'Illumination and Relighting Control' (seven papers). The scope notes clarify that methods controlling only camera (e.g., CamCo) or only lighting (e.g., Relightable Portrait Animation) belong to specialized single-modality branches. Light-X diverges from these by explicitly separating geometric and photometric signals, whereas implicit joint control methods condition on multiple signals without architectural disentanglement. This positioning suggests the work bridges previously isolated research directions.

Among 22 candidates examined, the disentangled conditioning scheme (Contribution 2) shows overlap with prior work: 10 candidates were reviewed, and 2 appear to provide refutable evidence. The Light-X framework itself (Contribution 1) and the Light-Syn data pipeline (Contribution 3) examined 10 and 2 candidates respectively, with no clear refutations found. The limited search scope means these statistics reflect top-K semantic matches and citation expansion, not exhaustive coverage. The disentangled conditioning appears less novel given the identified overlaps, while the overall framework and data synthesis strategy show fewer direct precedents within the examined set.

Based on the limited literature search (22 candidates), the work appears to occupy a sparsely populated research direction, with only one sibling paper in its taxonomy leaf. The disentangled conditioning scheme has some precedent among the examined candidates, but the integrated framework and data synthesis approach show fewer overlaps. The analysis covers top-K semantic matches and does not claim exhaustive field coverage, so additional related work may exist beyond the examined scope.

---

This paper presents **3 main contributions**, each analyzed against relevant prior work:

### Contribution 1: Light-X framework for joint camera and illumination control

**Description**: The authors introduce Light-X, the first framework that jointly controls camera trajectory and illumination for video generation from monocular input videos. This enables novel-view synthesis with simultaneous lighting manipulation, addressing a gap left by prior methods that handle only one control dimension.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

#### 1. Neural video portrait relighting in real-time via consistency modeling

**URL**: View paper

**Brief Assessment**

Portrait Relighting Real-time[49] focuses exclusively on video portrait relighting with temporal consistency modeling, without any camera trajectory control capabilities. The candidate addresses a different problem scope (relighting only) compared to the original's joint camera-illumination control framework.

---

#### 2. Neural reconstruction of relightable human model from monocular video

**URL**: View paper

**Brief Assessment**

Neural Face Relighting[48] focuses on reconstructing relightable human models from monocular video for novel illumination rendering, but does not address joint camera trajectory control. The candidate paper's scope is limited to relighting under fixed viewpoints, not simultaneous camera and illumination manipulation for video generation.

---

#### 3. Monocular reconstruction of neural face reflectance fields

**URL**: View paper

**Brief Assessment**

Neural Face Reflectance[50] focuses on monocular face reflectance field reconstruction for relighting faces under different viewpoints and lighting, but does not address joint camera trajectory and illumination control for general video generation from monocular videos as Light-X does.

---

#### 4. Reality's Canvas, Language's Brush: Crafting 3D Avatars from Monocular Video

**URL**: View paper

**Brief Assessment**

Reality's Canvas[52] focuses on 3D avatar creation from monocular video with text-based manipulation of shape, texture, and lighting, but does not address joint camera trajectory and illumination control for video generation. The candidate targets avatar reconstruction and editing, not novel-view video synthesis with simultaneous lighting manipulation.

---

#### 5. Joint self-supervised learning of interest point, descriptor, depth, and ego-motion from monocular video

**URL**: View paper

**Brief Assessment**

Joint Self-supervised Learning[51] focuses on self-supervised learning of interest points, descriptors, depth, and ego-motion from monocular video for visual odometry tasks, not on controllable video generation with joint camera trajectory and illumination control for novel-view synthesis and relighting.

### 6. Real-time 3d-aware portrait video relighting

**URL**: View paper

**Brief Assessment**

Real-time Portrait Relighting[13] focuses on portrait video relighting with novel view synthesis but does not address joint camera trajectory and illumination control from monocular videos in the same manner as the original paper's Light-X framework.

### 7. Vidcraft3: Camera, object, and lighting control for image-to-video generation

**URL**: View paper

**Brief Assessment**

Vidcraft3[1] focuses on image-to-video generation with joint control over camera motion, object motion, and lighting direction, whereas the original paper addresses video-to-video generation with joint camera trajectory and illumination control from monocular videos. The technical approaches differ fundamentally in their input modalities and control mechanisms.

### 8. Surfel-based Gaussian inverse rendering for fast and relightable dynamic human reconstruction from monocular videos

**URL**: View paper

**Brief Assessment**

Surfel Gaussian Relighting[47] focuses on physically-based rendering (PBR) properties reconstruction for dynamic human avatars from monocular videos, not on joint camera trajectory and illumination control for general video generation. The candidate addresses relighting of human avatars with novel poses, while the original paper presents a framework for controllable video generation with both viewpoint and illumination manipulation across diverse scene types.

### 9. IllumiCraft: Unified Geometry and Illumination Diffusion for Controllable Video Generation

**URL**: View paper

**Brief Assessment**

IllumiCraft[36] focuses on integrating HDR maps, relit frames, and 3D point tracks for video relighting, but does not explicitly address joint camera trajectory control alongside illumination manipulation from monocular videos as Light-X does.

### 10. High-Fidelity Relightable Monocular Portrait Animation with Lighting-Controllable Video Diffusion Model

**URL**: View paper

**Brief Assessment**

Relightable Portrait Animation[7] focuses on portrait animation with lighting control but does not address camera trajectory control or novel-view synthesis from monocular videos, which are central to the original paper's contribution.

## Contribution 2: Disentangled conditioning scheme separating geometry and lighting

**Description**: The authors develop a conditioning formulation that explicitly separates geometric and motion information (via dynamic point clouds) from illumination cues (via projected relit frames). This disentanglement enables fine-grained control and effective learning of both camera trajectory and lighting.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

### 1. BEAM: Bridging Physically-based Rendering and Gaussian Modeling for Relightable Volumetric Video

**URL**: View paper

**Brief Assessment**

BEAM[42] focuses on recovering PBR properties (roughness, AO, base color) for relightable volumetric video from multi-view RGB footage using Gaussian representations, not on disentangling geometry and lighting signals for controllable video synthesis from monocular input.

### 2. Lumen: Consistent video relighting and harmonious background replacement with video generative models

**URL**: View paper

**Brief Assessment**

Lumen[40] focuses on video relighting with background replacement using a multi-domain training curriculum and style adapter, rather than explicitly disentangling geometry and lighting signals through dynamic point clouds and projected relit frames as in the original paper.

### 3. X2Video: Adapting Diffusion Models for Multimodal Controllable Neural Video Rendering

**URL**: View paper

**Brief Assessment**

X2Video[41] focuses on intrinsic channel guidance (albedo, normal, roughness, metallicity, irradiance) for video rendering, not on disentangling geometry from lighting signals. The original paper's approach uses dynamic point clouds for geometry/motion and projected relit frames for illumination, which is architecturally distinct from X2Video's intrinsic channel-based control.

### 4. Controllable generation with disentangled representative learning of multiple perspectives in autonomous driving

**URL**: View paper

**Brief Assessment**

Disentangled Autonomous Driving[39] focuses on disentangling weather and speed factors in autonomous driving scenes, not geometry and lighting for video relighting. The technical domains and applications are fundamentally different.

### 5. UniRelight: Learning Joint Decomposition and Synthesis for Video Relighting

**URL**: View paper

**Brief Assessment**

UniRelight[37] uses a joint denoising approach where albedo and relit video latents are concatenated and processed together, rather than explicitly separating geometry (via point clouds) and lighting (via projected relit frames) as conditioning signals. The candidate's joint formulation differs fundamentally from the original's disentangled geometric projection strategy.

### 6. Shape-for-motion: Precise and consistent video editing with 3d proxy

**URL**: View paper

**Brief Assessment**

Shape-for-motion[44] focuses on 3D mesh-based video editing with geometry and texture decoupling for spatial manipulation, not on separating illumination from geometry for lighting control as in the original paper.

### 7. Idol: Instant photorealistic 3d human creation from a single image

**URL**: View paper

**Brief Assessment**

Idol[38] focuses on 3D human reconstruction from single images using Gaussian splatting and SMPL-X models. It does not address video generation, camera trajectory control, or illumination manipulation, which are central to the original paper's contribution.

### 8. Generative multiview relighting for 3d reconstruction under extreme illumination variation

**URL**: View paper

**Brief Assessment**

Multiview Relighting[43] addresses 3D reconstruction from images with varying illumination by relighting images first, then reconstructing. The original paper's disentangled conditioning for video generation (separating dynamic point clouds for geometry/motion from projected relit frames for lighting) differs fundamentally from this candidate's two-stage pipeline (multiview relighting diffusion followed by NeRF-based reconstruction).

### 9. IllumiCraft: Unified Geometry and Illumination Diffusion for Controllable Video Generation

**URL**: View paper

**Prior Art Analysis**

IllumiCraft[36] demonstrates prior work on disentangling geometry and lighting signals for controllable video synthesis. The candidate explicitly describes accepting three complementary inputs that separate geometric information (3D point tracks), lighting control (HDR video maps), and appearance cues (synthetically relit frames). This architecture integrates lighting, appearance, and geometry cues within a unified diffusion framework, showing that the concept of explicitly separating geometric and illumination information for video generation existed before the original paper's submission.

**Evidence**

Evidence 1 - **Rationale**: Both papers describe integrating separated geometry and lighting signals within a diffusion framework. The original explicitly decouples geometry/motion from illumination using point clouds and relit frames, while IllumiCraft[36] integrates lighting, appearance, and geometry cues, showing prior work on this disentanglement approach. - **Original**: we introduce a conditioning scheme that explicitly decouples geometry/motion from illumination. camera trajectories are modeled through dynamic point cloud rendering like (yu et al., 2025), while illumination cues are provided by projecting a relit frame (obtained via (zhang et al., 2025b)) into the... - **Candidate**: By integrating the lighting, appearance, and geometry cues within a unified diffusion architecture, illumicraft generates temporally coherent videos aligned with user-defined prompts.

### 10. High-Fidelity Relightable Monocular Portrait Animation with Lighting-Controllable Video Diffusion Model

**URL**: View paper

**Prior Art Analysis**

Relightable Portrait Animation[7] demonstrates prior work on disentangling intrinsic features (identity and appearance) from extrinsic features (pose and lighting) through dedicated feature subspaces. The candidate explicitly addresses the separation of these attributes using adapters that map features into distinct subspaces, similar to the original paper's approach of separating geometry/motion from illumination cues. Both papers employ conditioning schemes that explicitly decouple these signals to enable fine-grained control.

**Evidence**

Evidence 1 - **Rationale**: Both approaches use explicit representations to separate lighting from other attributes—the original projects relit frames for illumination cues, while the candidate renders shading hints to encode extrinsic attributes including lighting. - **Original**: we introduce a conditioning scheme that explicitly decouples geometry/motion from illumination. camera trajectories are modeled through dynamic point cloud rendering like (yu et al., 2025), while illumination cues are provided by projecting a relit frame - **Candidate**: we represent the portrait's extrinsic attributes using shading hints rendered with the reference image's 3d mesh, target lighting, and driving frame's pose, while the intrinsic attributes are represented by the

## Contribution 3: Light-Syn degradation-based data synthesis pipeline

**Description**: The authors propose Light-Syn, a data curation pipeline that addresses the scarcity of paired multi-view and multi-illumination videos by synthesizing training pairs from monocular footage through degradation and inverse geometric mapping. This enables robust training across static, dynamic, and AI-generated scenes.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

### 1. Structure-centric robust monocular depth estimation via knowledge distillation

**URL**: View paper

**Brief Assessment**

Structure-centric Depth[45] focuses on monocular depth estimation robustness through illumination-reflectance decoupling and semantic knowledge distillation, not on multi-view multi-illumination video synthesis from monocular footage. The candidate addresses a fundamentally different problem domain (depth estimation vs. video generation with camera/illumination control).

### 2. DBMovi-GS: Dynamic View Synthesis from Blurry Monocular Video via Sparse-Controlled Gaussian Splatting

**URL**: View paper

**Brief Assessment**

DBMovi-GS[46] focuses on dynamic view synthesis from blurry monocular videos using sparse-controlled Gaussian splatting, not on degradation-based synthesis of multi-view multi-illumination training pairs. The candidate addresses motion blur and camera pose estimation in 3D reconstruction, while the original paper's Light-Syn pipeline synthesizes paired training data for illumination control through degradation and inverse geometric mapping from monocular footage.

## Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

## References


- [0] Light-X: Generative 4D Video Rendering with Camera and Illumination Control View paper
- [1] Vidcraft3: Camera, object, and lighting control for image-to-video generation View paper
- [2] Uni3C: Unifying Precisely 3D-Enhanced Camera and Human Motion Controls for Video Generation View paper
- [3] CamCo: Camera-Controllable 3D-Consistent Image-to-Video Generation View paper
- [4] Gen3c: 3d-informed world-consistent video generation with precise camera control View paper
- [5] Direct-a-Video: Customized Video Generation with User-Directed Camera Movement and Object Motion View paper
- [6] Stable Cinemetrics: Structured Taxonomy and Evaluation for Professional Video Generation View paper
- [7] High-Fidelity Relightable Monocular Portrait Animation with Lighting-Controllable Video Diffusion Model View paper
- [8] Neural light transport for relighting and view synthesis View paper
- [9] Free-form motion control: A synthetic video generation dataset with controllable camera and object motions View paper
- [10] Motionbooth: Motion-aware customized text-to-video generation View paper
- [11] Beyond Static Scenes: Camera-controllable Background Generation for Human Motion View paper
- [12] VideoFrom3D: 3D Scene Video Generation via Complementary Image and Video Diffusion Models View paper
- [13] Real-time 3d-aware portrait video relighting View paper
- [14] Reviewing Intelligent Cinematography: AI research for camera-based video production View paper
- [15] Controllable video generation: A survey View paper
- [16] Motion Prompting: Controlling Video Generation with Motion Trajectories View paper
- [17] Motionstream: Real-time video generation with interactive motion controls View paper
- [18] CineTechBench: A Benchmark for Cinematographic Technique Understanding and Generation View paper
- [19] Free-viewpoint indoor neural relighting from multi-view stereo View paper
- [20] Uniavatar: Taming lifelike audio-driven talking head generation with comprehensive motion and lighting control View paper
- [21] FloVD: Optical Flow Meets Video Diffusion Model for Enhanced Camera-Controlled Video Synthesis View paper
- [22] Enhancing Human-Computer Interaction Through Decoupling Motion and Camera Control in Human-Centric Video Generation View paper
- [23] BulletTime: Decoupled Control of Time and Camera Pose for Video Generation View paper
- [24] Enhancing NeRFs for High-Quality Indoor Video Generation: A Study on Parameterization and Recording Methods View paper
- [25] Joint multi-layer segmentation and reconstruction for free-viewpoint video applications View paper
- [26] Virtually Being: Customizing Camera-Controllable Video Diffusion Models with Multi-View Performance Captures View paper
- [27] Virtually Being: Customizing Camera-Controllable Video Diffusion Models with Volumetric Performance Captures View paper
- [28] ReCapture: Generative Video Camera Controls for User-Provided Videos using Masked Video Fine-Tuning View paper
- [29] Virtual Cinematography and Camera Control for Film and Video Production View paper
- [30] AKiRa: Augmentation Kit on Rays for optical video generation View paper
- [31] Video Production with Generative AI View paper
- [32] MyGo: Consistent and Controllable Multi-View Driving Video Generation with Camera Control View paper
- [33] Scene Co-pilot: Procedural Text to Video Generation with Human in the Loop View paper
- [34] Augmenting Visual Memories View paper
- [35] Simultaneous Camera Path Optimization and Distraction Removal for Improving Amateur Video. View paper
- [36] IllumiCraft: Unified Geometry and Illumination Diffusion for Controllable Video Generation View paper
- [37] UniRelight: Learning Joint Decomposition and Synthesis for Video Relighting View paper
- [38] Idol: Instant photorealistic 3d human creation from a single image View paper
- [39] Controllable generation with disentangled representative learning of multiple perspectives in autonomous driving View paper
- [40] Lumen: Consistent video relighting and harmonious background replacement with video generative models View paper
- [41] X2Video: Adapting Diffusion Models for Multimodal Controllable Neural Video Rendering View paper
- [42] BEAM: Bridging Physically-based Rendering and Gaussian Modeling for Relightable Volumetric Video View paper
- [43] Generative multiview relighting for 3d reconstruction under extreme illumination variation View paper
- [44] Shape-for-motion: Precise and consistent video editing with 3d proxy View paper
- [45] Structure-centric robust monocular depth estimation via knowledge distillation View paper
- [46] DBMovi-GS: Dynamic View Synthesis from Blurry Monocular Video via Sparse-Controlled Gaussian Splatting View paper
- [47] Surfel-based Gaussian inverse rendering for fast and relightable dynamic human reconstruction from monocular videos View paper
- [48] Neural reconstruction of relightable human model from monocular video View paper
- [49] Neural video portrait relighting in real-time via consistency modeling View paper
- [50] Monocular reconstruction of neural face reflectance fields View paper
- [51] Joint self-supervised learning of interest point, descriptor, depth, and ego-motion from monocular video View paper
- [52] Reality's Canvas, Language's Brush: Crafting 3D Avatars from Monocular Video View paper