

Novelty Assessment Report

Paper: What Happens Next? Anticipating Future Motion by Generating Point Trajectories

PDF URL: <https://openreview.net/pdf?id=t1vMY11yhe>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

We consider the problem of forecasting motion from a single image, i.e., predicting how objects in the world are likely to move, without the ability to observe other parameters such as the object velocities or the forces applied to them. We formulate this task as conditional generation of dense trajectory grids with a model that closely follows the architecture of modern video generators but outputs motion trajectories instead of pixels. This approach captures scene-wide dynamics and uncertainty, yielding more accurate and diverse predictions than prior regressors and generators. Although recent state-of-the-art video generators are often regarded as world models, we show that they struggle with forecasting motion from a single image, even in simple physical scenarios such as falling blocks or mechanical object interactions, despite fine-tuning on such data. We show that this limitation arises from the overhead of generating pixels rather than directly modeling motion.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **forecasting motion from a single image using point trajectories**

A total of **50 papers** were analyzed and organized into a taxonomy with **23 categories**.

Taxonomy Overview

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

- **Single-Image Motion Forecasting and Trajectory Generation**
- **Point Tracking in Videos and Temporal Sequences**
- **Multi-Object Tracking with Motion Prediction**
- **Human and Pedestrian Trajectory Prediction**
- **Autonomous Driving and Vehicle Motion Planning**
- **Specialized Tracking and Prediction Applications**
- **Tracking Infrastructure and Algorithmic Techniques**

Complete Taxonomy Tree

- forecasting motion from a single image using point trajectories Survey Taxonomy
- Single-Image Motion Forecasting and Trajectory Generation
 - Dense Trajectory and Pixel-Level Motion Prediction ★ (3 papers)
 - [0] What Happens Next? Anticipating Future Motion by Generating Point Trajectories (Anon et al., 2026) [View paper](#)
 - [5] An uncertain future: Forecasting from static images using variational autoencoders (Jacob Walker, 2016) [View paper](#)
 - [15] MoTDiff: High-resolution Motion Trajectory estimation from a single blurred image using Diffusion models (Choi¹/⁴] Won-Tae, 2025) [View paper](#)
 - Human Pose and Body Dynamics Forecasting (2 papers)
 - [16] Forecasting human dynamics from static images (Yu-Wei Chao, 2017) [View paper](#)
 - [37] Human Trajectory Prediction Based on a Single Frame of Pose and Initial Velocity Information (Yucheng Huang, 2025) [View paper](#)
 - 3D Scene Structure and Camera Motion Estimation (2 papers)
 - [25] Unsupervised video prediction from a single frame by estimating 3d dynamic scene structure (Henderson, 2021) [View paper](#)
 - [26] Image as an IMU: Estimating Camera Motion from a Single Motion-Blurred Image (Chen, 2025) [View paper](#)
 - Image-to-Video Generation with Motion Synthesis (1 papers)
 - [24] Learning to forecast and refine residual motion for image-to-video generation (Long Zhao, 2018) [View paper](#)
 - Point Tracking in Videos and Temporal Sequences
 - General Point and Feature Tracking Frameworks (5 papers)
 - [2] Exploring Temporally-Aware Features for Point Tracking (In[~]'s Hyeonsu Kim, 2025) [View paper](#)
 - [14] Particle video: Long-range motion estimation using point trajectories (Peter Sand, 2008) [View paper](#)
 - [44] TAPNext: Tracking Any Point (TAP) as Next Token Prediction (Zholus, 2025) [View paper](#)
 - [47] TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement (Carl Doersch, 2023) [View paper](#)
 - [48] Point Prompting: Counterfactual Tracking with Video Diffusion Models (Shrivastava, 2025) [View paper](#)
 - Robotics and Manipulation with Point Trajectory Prediction (3 papers)
 - [1] Track2act: Predicting point tracks from internet videos enables generalizable robot manipulation (Homanga Bharadhwaj, 2024) [View paper](#)
 - [8] Pre-training auto-regressive robotic models with 4d representations (Niu Dantong, 2025) [View paper](#)
 - [36] Track2Act: Predicting Point Tracks from Internet Videos enables Diverse Zero-shot Robot Manipulation (Bharadhwaj, 2024) [View paper](#)
 - 3D Reconstruction and Articulated Object Modeling (1 papers)

- [9] Generalizable Articulated Object Reconstruction from Casually Captured RGBD Videos (W Peng, 2025) [View paper](#)
- Multi-Object Tracking with Motion Prediction
 - Learned Motion Predictors for Object Association (3 papers)
 - [3] Motiontrack: Learning motion predictor for multiple object tracking (Xiao, 2024) [View paper](#)
 - [11] Ettrack: enhanced temporal motion predictor for multi-object tracking (Xudong Han, 2024) [View paper](#)
 - [28] TrajMOT: A Multi-Object Tracking Framework via Trajectory Prediction (Ruining Zhang, 2025) [View paper](#)
 - ID Prediction and Appearance-Based Tracking (1 papers)
 - [4] Multiple object tracking as id prediction (Ruopeng Gao, 2025) [View paper](#)
 - Spatio-Temporal Reasoning and 3D Multi-Camera Tracking (4 papers)
 - [6] Standing between past and future: Spatio-temporal modeling for multi-camera 3d multi-object tracking (Ziqi Pang, 2023) [View paper](#)
 - [12] FASTer: Focal token Acquiring-and-Scaling Transformer for Long-term 3D Object Detection (C Dang, 2025) [View paper](#)
 - [18] Tracking people by predicting 3d appearance, location and pose (Jathushan Rajasegaran, 2022) [View paper](#)
 - [19] Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net (Wenjie Luo, 2018) [View paper](#)
 - Single-Frame Detection with Motion Anticipation (1 papers)
 - [20] Objects do not disappear: Video object detection by single-frame object location anticipation (Xin Liu, 2023) [View paper](#)
- Human and Pedestrian Trajectory Prediction
 - Video-Based Pedestrian Trajectory Forecasting (2 papers)
 - [7] Pedestrian motion trajectory prediction in intelligent driving from far shot first-person perspective video (Yingfeng Cai, 2021) [View paper](#)
 - [29] A predictor of moving objects for first-person vision (Ricardo Sánchez-Matilla, 2019) [View paper](#)
 - Point Cloud and LiDAR-Based Trajectory Prediction (2 papers)
 - [22] Pedestrian Trajectory Prediction Method Based On 3D Point Cloud (Meiqiang Li, 2025) [View paper](#)
 - [34] Behavioral-Aware Early Destination Prediction Using LiDAR Point Clouds and Movement Trajectories (Ohno, 2025) [View paper](#)
 - Social Forces and Multi-Agent Interaction Modeling (2 papers)
 - [10] Learning cooperative trajectory representations for motion forecasting (Hongzhi Ruan, 2024) [View paper](#)
 - [32] People tracking with human motion predictions from social forces (M. Luber, 2010) [View paper](#)
 - Survey and Taxonomy of Human Motion Prediction (1 papers)
 - [17] Human motion trajectory prediction: A survey (Andrey Rudenko, 2020) [View paper](#)
- Autonomous Driving and Vehicle Motion Planning
 - End-to-End Trajectory Planning with Momentum and Stability (1 papers)
 - [49] Don't Shake the Wheel: Momentum-Aware Planning in End-to-End Autonomous Driving (Ziying Song, 2025) [View paper](#)
 - Adversarial Robustness and Attack Analysis (1 papers)
 - [40] A Stealthy Steering Angle Manipulation Attack Against End-to-End Autonomous Driving Systems (Sawant, 2025) [View paper](#)
- Specialized Tracking and Prediction Applications
 - Agricultural and Small-Object Tracking (1 papers)
 - [27] Robust Real-Time Blueberry Counting in Greenhouses Using Small-Object Detection and Mamba-Driven Multi-Step Trajectory Completion (Naiqi Zhang, 2025) [View paper](#)
 - VR/AR and Interactive Pointing Prediction (3 papers)
 - [13] Continuous Prediction of Pointing Targets With Motion and Eye-Tracking in Virtual Reality (Choongho Chung, 2024) [View paper](#)
 - [21] Shooting a moving target: Motion-prediction-based transmission for 360-degree videos (Yanan Bao, 2016) [View paper](#)
 - [23] TRACK: A New Method From a Re-Examination of Deep Architectures for Head Motion Prediction in 360 Videos (MFR Rondán, 2021) [View paper](#)
 - Biological and Ecological Movement Analysis (5 papers)
 - [33] Grid cells accurately track movement during path integration-based navigation despite switching reference frames. (Jing-Jie Peng, 2025) [View paper](#)
 - [35] Exploring the interplay between small and large scales movements in a neotropical small mammal (Edgardo Brigatti, 2025) [View paper](#)
 - [38] Movement tracking of psychological processes: A tutorial using mousetrap. (Dirk U. Wulff, 2025) [View paper](#)
 - [46] Quantifying and Analyzing Mosquito Movement from Video Tracking Results. (Florian T. Muijres, 2023) [View paper](#)
 - [50] Classification-based motion analysis of single-molecule trajectories using DiffusionLab. (J. Maris, 2022) [View paper](#)
- Tracking Infrastructure and Algorithmic Techniques
 - Feature-Based and Hybrid Tracking Systems (1 papers)
 - [30] Fusing points and lines for high performance tracking (Edward Rosten, 2005) [View paper](#)
 - Prediction-Based Adaptive Tracking and Coordination (4 papers)
 - [39] Automated cell tracking using motion prediction-based matching and event handling (F. Boukari, 2018) [View paper](#)
 - [41] The object tracking and movement prediction based on the Gaussian mixture model (Dongsheng Lyu, 2023) [View paper](#)
 - [42] Analysis of a prediction-based mobility adaptive tracking algorithm (J. Yick, 2005) [View paper](#)
 - [45] Improved tracking of multiple humans with trajectory prediction and occlusion modeling (R Rosales, 1998) [View paper](#)
 - Graph-Based and Subgraph Learning for Trajectory Prediction (2 papers)
 - [31] Robust Trajectory Prediction Based on Efficient Fusion of Heterogeneous Information (Xuecheng Wang, 2025) [View paper](#)
 - [43] Next track point prediction using a flexible strategy of subgraph learning on road networks (Yifan Zhang, 2024) [View paper](#)

Narrative

Core task: forecasting motion from a single image using point trajectories. This field addresses the challenge of predicting how points in a static image will move over time, a problem that spans computer vision, robotics, and autonomous systems. The taxonomy reveals a diverse landscape organized into seven main branches. Single-Image Motion Forecasting and Trajectory Generation focuses on methods that extract future motion directly from still frames, often producing dense pixel-level or sparse point trajectories. Point Tracking in Videos and Temporal Sequences emphasizes techniques like TAPIR[47] and TAPNext[44] that follow points across video frames, building temporal correspondences. Multi-Object Tracking with Motion Prediction and Human and Pedestrian Trajectory Prediction branches address entity-level forecasting, where works such as Pedestrian Trajectory Prediction[7] model social interactions and scene context. Autonomous Driving and Vehicle Motion Planning applies these ideas to navigation and safety-critical scenarios, while Specialized

Tracking and Prediction Applications explores domain-specific problems ranging from robotic manipulation to biological movement analysis. Finally, Tracking Infrastructure and Algorithmic Techniques provides foundational methods for association, occlusion handling, and efficient computation.

Several active lines of work reveal key trade-offs between dense versus sparse representations, single-frame versus temporal modeling, and deterministic versus probabilistic forecasting. Early efforts like Uncertain Future[5] and Particle Video[14] explored stochastic motion from limited observations, while recent approaches increasingly leverage learned features and diffusion models, as seen in MoTDiff[15]. Anticipating Future Motion[0] sits within the Dense Trajectory and Pixel-Level Motion Prediction cluster, emphasizing the generation of detailed point trajectories from a single image without relying on video sequences. This contrasts with video-based trackers like Motiontrack[3], which exploit temporal continuity, and with probabilistic frameworks that model multiple plausible futures. The work shares thematic connections with Single Frame Prediction[37] in its reliance on static input, yet distinguishes itself by focusing on trajectory-level rather than frame-level synthesis, positioning it at the intersection of generative modeling and motion understanding.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. An uncertain future: Forecasting from static images using variational autoencoders

Authors: Jacob Walker, Carl Doersch, A. Gupta, M. Hebert, ABHINAV GUPTA, et al. (6 authors total) | **Year/Venue:** 2016 | **URL:** [View paper](#)

Abstract

In a given scene, humans can easily predict a set of immediate future events that might happen. However, pixel-level anticipation in computer vision is difficult because machine learning struggles with the ambiguity in predicting the future. In this paper, we focus on predicting the dense trajectory of pixels in a scene—what will move in the scene, where it will travel, and how it will deform over the course of one second. We propose a conditional variational autoencoder as a solution to this ...

Relationship Analysis

Both papers belong to the Dense Trajectory and Pixel-Level Motion Prediction category, focusing on generating dense pixel trajectories from single images using generative models. They overlap in their core approach of predicting point trajectories to forecast motion and both employ variational autoencoders (VAEs) for handling uncertainty in future predictions. The key difference is that the original paper uses modern flow matching techniques with a transformer-based architecture and evaluates on physical simulation datasets with rigidity metrics, while the candidate paper uses a classical VAE with CNN-based architecture, applies DCT compression for trajectory encoding, and evaluates primarily on action recognition datasets (UCF101) using log-likelihood metrics.

2. MoTDiff: High-resolution Motion Trajectory estimation from a single blurred image using Diffusion models

Authors: Choi^{1/4} Won-Tae, Wontae Choi, Jaelin Lee, Jeon^{1/4} Byeungwoo, Hyung Sup Yun, et al. (9 authors total) | **Year/Venue:** 2025 | **URL:** [View paper](#)

Abstract

Accurate estimation of motion information is crucial in diverse computational imaging and computer vision applications. Researchers have investigated various methods to extract motion information from a single blurred image, including blur kernels and optical flow. However, existing motion representations are often of low quality, i.e., coarse-grained and inaccurate. In this paper, we propose the first high-resolution (HR) Motion Trajectory estimation framework using Diffusion models (MoTDiff). ...

Relationship Analysis

Both papers belong to the Dense Trajectory and Pixel-Level Motion Prediction category, focusing on generating motion representations from single images. While the original paper forecasts future motion by generating dense point trajectories across entire scenes using flow matching, this candidate paper estimates past motion trajectories from motion-blurred images using diffusion models for deblurring applications. The key difference lies in their temporal direction (future forecasting vs. past motion recovery) and application domains (general motion prediction vs. computational photography).

Contributions Analysis

Overall novelty summary. The paper formulates motion forecasting from a single image as conditional generation of dense trajectory grids, positioning itself within the Dense Trajectory and Pixel-Level Motion Prediction leaf of the taxonomy. This leaf contains only three papers total, including the original work, indicating a relatively sparse research direction. The sibling papers explore related themes: one addresses uncertain future prediction with stochastic motion models, while another examines particle-based video representations. This small cluster suggests the specific combination of single-image input, dense trajectory output, and generative modeling remains underexplored compared to video-based tracking branches, which contain significantly more papers.

The taxonomy reveals that neighboring research directions are substantially more populated. The Point Tracking in Videos branch contains multiple well-developed subcategories with works like TAPIR and TAPNext that leverage temporal sequences for correspondence. The Multi-Object Tracking subtree addresses entity-level forecasting with learned motion predictors and appearance-based methods. The paper's approach diverges from these by eliminating temporal input entirely, instead inferring motion from static visual cues alone. This boundary is reinforced by the taxonomy's exclude notes, which explicitly separate single-image methods from video-based tracking and multi-frame prediction categories, highlighting the distinct challenge of forecasting without observing actual motion.

Among the 19 candidates examined through semantic search, none clearly refute the three main contributions. The first contribution—formulating motion forecasting as dense trajectory generation—examined 10 candidates with no refutable overlaps. The second contribution comparing trajectory generation to regressors and video generators examined 4 candidates, again with no clear prior work. The third contribution analyzing pixel generation overhead in video models examined 5 candidates without finding substantial precedent. This limited search scope suggests the specific framing and comparative analysis may be novel within the examined literature, though the small candidate pool (19 papers) means potentially relevant work outside top semantic matches remains unassessed.

The analysis indicates the work occupies a sparsely populated research direction, with its taxonomy leaf containing minimal prior work and neighboring branches focusing on fundamentally different input modalities. The absence of refutable candidates across 19 examined papers suggests novelty in the specific formulation and comparative insights, though this conclusion is constrained by the limited search scope. A more exhaustive literature review covering broader semantic neighborhoods or citation networks might reveal additional relevant precedents, particularly in adjacent areas like image-to-video generation or probabilistic motion modeling that were less thoroughly explored in this analysis.

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

Contribution 1: Formulation of motion forecasting as conditional generation of dense trajectory grids

Description: The authors propose a new formulation for predicting future motion from a single image by generating dense trajectory grids rather than regressing trajectories or generating RGB pixels. This approach models scene-wide dynamics and uncertainty using flow matching in a latent space learned by a trajectory variational autoencoder.

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. TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement

URL: [View paper](#)

Brief Assessment

TAPIR[47] focuses on point tracking across video frames, not motion forecasting from a single image. The candidate extracts trajectories from existing videos rather than generating future motion predictions conditioned on static images.

2. IKMo: Image-Keyframed Motion Generation with Trajectory-Pose Conditioned Motion Diffusion Model

URL: [View paper](#)

Brief Assessment

IKMo[62] focuses on human motion generation from keyframe images and poses, not on forecasting future motion from single images via dense trajectory grids. The candidate addresses a different problem domain (human animation control) with different inputs (keyframe images/poses) and outputs (human skeletal motion).

3. Training-free Dense-Aligned Diffusion Guidance for Modular Conditional Image Synthesis

URL: [View paper](#)

Brief Assessment

Dense-Aligned Diffusion[65] focuses on conditional image synthesis using text, layout, and drag conditions, not motion forecasting from images. The candidate addresses a fundamentally different problem domain (image synthesis vs. motion prediction).

4. COTA-motion: Controllable image-to-video synthesis with dense semantic trajectories

URL: [View paper](#)

Brief Assessment

COTA-motion[66] focuses on controllable image-to-video synthesis with dense semantic trajectories for video generation tasks, not motion forecasting from single images as a conditional generation problem for predicting future dynamics.

5. Magicmotion: Controllable video generation with dense-to-sparse trajectory guidance

URL: [View paper](#)

Brief Assessment

Magicmotion[60] focuses on controllable video generation from images using trajectory guidance (masks, boxes, sparse boxes), not on forecasting future motion from a single image as a conditional generation problem with dense trajectory grids in latent space.

6. Meta Motion Sense and Motion Trajectory Prediction

URL: [View paper](#)

Brief Assessment

Meta Motion Sense[63] focuses on optical flow estimation from wide-angle metalens imaging systems and trajectory prediction for motion sensing applications, not on conditional generation of dense trajectory grids as a formulation for motion forecasting from single images.

7. SIGHT: Synthesizing Image-Text Conditioned and Geometry-Guided 3D Hand-Object Trajectories

URL: [View paper](#)

Brief Assessment

SIGHT[64] focuses on generating 3D hand-object interaction trajectories from images and text, not general scene-wide motion forecasting from single images using dense trajectory grids as in the original paper.

8. Scene compliant trajectory forecast with agent-centric spatio-temporal grids

URL: [View paper](#)

Brief Assessment

Scene Compliant Forecast[61] focuses on forecasting human trajectories in traffic scenarios using grid representations of past motion and BEV scene images, not on generating dense trajectory grids from a single image for general physical motion prediction.

9. Motion-Conditioned Diffusion Model for Controllable Video Synthesis

URL: [View paper](#)

Brief Assessment

Motion-Conditioned Diffusion[67] focuses on controllable video synthesis from user-specified strokes, not motion forecasting from a single image. The candidate generates RGB video frames conditioned on strokes and start frames, whereas the original generates dense trajectory grids for motion prediction without stroke inputs.

10. Vehicular Multimodal Motion Forecasting via Conditional Score-based Modeling

URL: [View paper](#)

Brief Assessment

Vehicular Motion Forecasting[68] focuses on vehicle trajectory prediction in autonomous driving scenarios using score-based models, not dense trajectory grids from single images for general scene-wide motion forecasting.

Contribution 2: Demonstration that trajectory generation outperforms prior regressors and video generators

Description: The authors demonstrate through experiments that their generative approach to trajectory prediction surpasses both regression-based trajectory forecasters and large-scale pretrained video generators, even when the latter are fine-tuned on the target domain. They attribute this to modeling uncertainty and reasoning about the entire scene jointly.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Trajectory grid diffusion for multimodal trajectory prediction in autonomous vehicles

URL: [View paper](#)

Brief Assessment

Trajectory Grid Diffusion[51] focuses on multimodal trajectory prediction in autonomous vehicles using grid-based diffusion, not on comparing generative approaches against regression-based forecasters and video generators in general motion forecasting scenarios.

2. Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction

URL: [View paper](#)

Brief Assessment

Language-Based Trajectory Prediction[52] focuses on converting trajectory prediction into a language modeling task using text prompts and tokenization, not on comparing generative versus regression-based approaches or evaluating against video generators for motion forecasting.

3. Short-Term Probabilistic Wind Speed Predictions Integrating Multivariate Linear Regression and Generative Adversarial Network Methods

URL: [View paper](#)

Brief Assessment

Wind Speed Predictions[54] focuses on wind speed interval forecasting using time-series decomposition and GANs, not trajectory generation or motion forecasting from images. The domains are entirely different (meteorological prediction vs. visual motion forecasting).

4. VFR trajectory forecasting using deep generative model for autonomous airspace operations

URL: [View paper](#)

Brief Assessment

VFR Trajectory Forecasting[53] focuses on forecasting visual flight rules aircraft trajectories in airspace operations, not general motion forecasting from images. The candidate addresses a completely different domain (aviation/airspace) with different input modalities and objectives than the original paper's image-based motion forecasting task.

Contribution 3: Analysis showing pixel generation overhead limits motion forecasting in video generators

Description: The authors provide experimental evidence that state-of-the-art video generators struggle with motion forecasting not due to lack of world knowledge, but because generating RGB pixels introduces overhead that reduces focus on motion accuracy and physical plausibility. They demonstrate this by ablating output modality while keeping architecture fixed.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Pixel-Level Tracking and Future Prediction in Video Streams

URL: [View paper](#)

Brief Assessment

Pixel-Level Tracking[58] is a PhD thesis focused on dense pixel-level tracking methods and video synthesis architectures, not on analyzing why video generators struggle with motion forecasting or demonstrating that pixel generation overhead is the limiting factor.

2. A good image generator is what you need for high-resolution video synthesis

URL: [View paper](#)

Brief Assessment

Good Image Generator[55] focuses on using pre-trained image generators to synthesize high-resolution videos by discovering trajectories in latent space, not on analyzing pixel generation overhead as a limitation for motion forecasting. The candidate's motivation is computational efficiency and resolution scaling, not demonstrating that RGB pixel generation reduces focus on motion accuracy.

3. View synthesis prediction for multiview video coding

URL: [View paper](#)

Brief Assessment

View Synthesis Prediction[57] addresses multiview video coding using depth information and view synthesis for compression efficiency, not motion forecasting from single images. The paper focuses on encoding auxiliary depth data for prediction in multiview scenarios, which is fundamentally different from analyzing pixel generation overhead in video generators for motion forecasting tasks.

4. Future frame synthesis for fast Monte Carlo rendering

URL: [View paper](#)

Brief Assessment

Future Frame Synthesis[56] focuses on accelerating Monte Carlo rendering by predicting future frames to reduce ray sampling, not on analyzing how pixel generation overhead affects motion forecasting accuracy in video generators.

5. Intermediate frames prediction for integral imaging video based on Triple I-3D Net.

URL: [View paper](#)

Brief Assessment

Intermediate Frames Prediction[59] focuses on integral imaging 3D display technology and frame interpolation for element image arrays, not on analyzing pixel generation overhead in motion forecasting for video generators.

Appendix: Text Similarity Detection

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

References

- [0] What Happens Next? Anticipating Future Motion by Generating Point Trajectories [View paper](#)
-

- [1] Track2act: Predicting point tracks from internet videos enables generalizable robot manipulation [View paper](#)
- [2] Exploring Temporally-Aware Features for Point Tracking [View paper](#)
- [3] Motiontrack: Learning motion predictor for multiple object tracking [View paper](#)
- [4] Multiple object tracking as id prediction [View paper](#)
- [5] An uncertain future: Forecasting from static images using variational autoencoders [View paper](#)
- [6] Standing between past and future: Spatio-temporal modeling for multi-camera 3d multi-object tracking [View paper](#)
- [7] Pedestrian motion trajectory prediction in intelligent driving from far shot first-person perspective video [View paper](#)
- [8] Pre-training auto-regressive robotic models with 4d representations [View paper](#)
- [9] Generalizable Articulated Object Reconstruction from Casually Captured RGBD Videos [View paper](#)
- [10] Learning cooperative trajectory representations for motion forecasting [View paper](#)
- [11] Ettrack: enhanced temporal motion predictor for multi-object tracking [View paper](#)
- [12] FASTER: Focal token Acquiring-and-Scaling Transformer for Long-term 3D Objection Detection [View paper](#)
- [13] Continuous Prediction of Pointing Targets With Motion and Eye-Tracking in Virtual Reality [View paper](#)
- [14] Particle video: Long-range motion estimation using point trajectories [View paper](#)
- [15] MoTDiff: High-resolution Motion Trajectory estimation from a single blurred image using Diffusion models [View paper](#)
- [16] Forecasting human dynamics from static images [View paper](#)
- [17] Human motion trajectory prediction: A survey [View paper](#)
- [18] Tracking people by predicting 3d appearance, location and pose [View paper](#)
- [19] Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net [View paper](#)
- [20] Objects do not disappear: Video object detection by single-frame object location anticipation [View paper](#)
- [21] Shooting a moving target: Motion-prediction-based transmission for 360-degree videos [View paper](#)
- [22] Pedestrian Trajectory Prediction Method Based On 3D Point Cloud [View paper](#)
- [23] TRACK: A New Method From a Re-Examination of Deep Architectures for Head Motion Prediction in 360 Videos [View paper](#)
- [24] Learning to forecast and refine residual motion for image-to-video generation [View paper](#)
- [25] Unsupervised video prediction from a single frame by estimating 3d dynamic scene structure [View paper](#)
- [26] Image as an IMU: Estimating Camera Motion from a Single Motion-Blurred Image [View paper](#)
- [27] Robust Real-Time Blueberry Counting in Greenhouses Using Small-Object Detection and Mamba-Driven Multi-Step Trajectory Completion [View paper](#)
- [28] TrajMOT: A Multi-Object Tracking Framework via Trajectory Prediction [View paper](#)
- [29] A predictor of moving objects for first-person vision [View paper](#)
- [30] Fusing points and lines for high performance tracking [View paper](#)
- [31] Robust Trajectory Prediction Based on Efficient Fusion of Heterogeneous Information [View paper](#)
- [32] People tracking with human motion predictions from social forces [View paper](#)
- [33] Grid cells accurately track movement during path integration-based navigation despite switching reference frames. [View paper](#)
- [34] Behavioral-Aware Early Destination Prediction Using LiDAR Point Clouds and Movement Trajectories [View paper](#)
- [35] Exploring the interplay between small and large scales movements in a neotropical small mammal [View paper](#)
- [36] Track2Act: Predicting Point Tracks from Internet Videos enables Diverse Zero-shot Robot Manipulation [View paper](#)
- [37] Human Trajectory Prediction Based on a Single Frame of Pose and Initial Velocity Information [View paper](#)
- [38] Movement tracking of psychological processes: A tutorial using mousetrap. [View paper](#)
- [39] Automated cell tracking using motion prediction-based matching and event handling [View paper](#)
- [40] A Stealthy Steering Angle Manipulation Attack Against End-to-End Autonomous Driving Systems [View paper](#)
- [41] The object tracking and movement prediction based on the Gaussian mixture model [View paper](#)
- [42] Analysis of a prediction-based mobility adaptive tracking algorithm [View paper](#)
- [43] Next track point prediction using a flexible strategy of subgraph learning on road networks [View paper](#)
- [44] TAPNext: Tracking Any Point (TAP) as Next Token Prediction [View paper](#)
- [45] Improved tracking of multiple humans with trajectory prediction and occlusion modeling [View paper](#)
- [46] Quantifying and Analyzing Mosquito Movement from Video Tracking Results. [View paper](#)
- [47] TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement [View paper](#)
- [48] Point Prompting: Counterfactual Tracking with Video Diffusion Models [View paper](#)
- [49] Don't Shake the Wheel: Momentum-Aware Planning in End-to-End Autonomous Driving [View paper](#)
- [50] Classification-based motion analysis of single-molecule trajectories using DiffusionLab. [View paper](#)
- [51] Trajectory grid diffusion for multimodal trajectory prediction in autonomous vehicles [View paper](#)
- [52] Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction [View paper](#)
- [53] VFR trajectory forecasting using deep generative model for autonomous airspace operations [View paper](#)
- [54] Short-Term Probabilistic Wind Speed Predictions Integrating Multivariate Linear Regression and Generative Adversarial Network Methods [View paper](#)
- [55] A good image generator is what you need for high-resolution video synthesis [View paper](#)
- [56] Future frame synthesis for fast Monte Carlo rendering [View paper](#)
- [57] View synthesis prediction for multiview video coding [View paper](#)
- [58] Pixel-Level Tracking and Future Prediction in Video Streams [View paper](#)
- [59] Intermediate frames prediction for integral imaging video based on Triple I-3D Net. [View paper](#)
- [60] Magicmotion: Controllable video generation with dense-to-sparse trajectory guidance [View paper](#)
- [61] Scene compliant trajectory forecast with agent-centric spatio-temporal grids [View paper](#)
- [62] IKMo: Image-Keyframed Motion Generation with Trajectory-Pose Conditioned Motion Diffusion Model [View paper](#)
- [63] Meta Motion Sense and Motion Trajectory Prediction [View paper](#)
- [64] SIGHT: Synthesizing Image-Text Conditioned and Geometry-Guided 3D Hand-Object Trajectories [View paper](#)
- [65] Training-free Dense-Aligned Diffusion Guidance for Modular Conditional Image Synthesis [View paper](#)
- [66] COTA-motion: Controllable image-to-video synthesis with dense semantic trajectories [View paper](#)
- [67] Motion-Conditioned Diffusion Model for Controllable Video Synthesis [View paper](#)
- [68] Vehicular Multimodal Motion Forecasting via Conditional Score-based Modeling [View paper](#)