[2026 summer seminar]

Physics-Guided Dynamic World Generation through Hybrid Generative Simulation

Pusan National University
Computer Vision & Signal Processing Lab, Dept. of Electronic Engineering

Presented By
김우형

CVSP Lab
Computer Vision & Signal Processing

Index

- Background
- 논문 소개
  - WonderPlay: Dynamic 3D Scene Generation from a Single Image and Actions [ICCV 2025 (Highlight)]
  - PerpetualWonder: Long-Horizon Action-Conditioned 4D Scene Generation [CVPR 2026 (Highlight)]
- Conclusion

Background

- Task
  - Action-conditioned Dynamic 3D Scene Generation from Single Image
    - Single Image로부터 Action의 결과를 나타내는 Dynamic 3D Scene 생성



< Single Image >



< Action >



< Dynamic 3D Scene >

Background

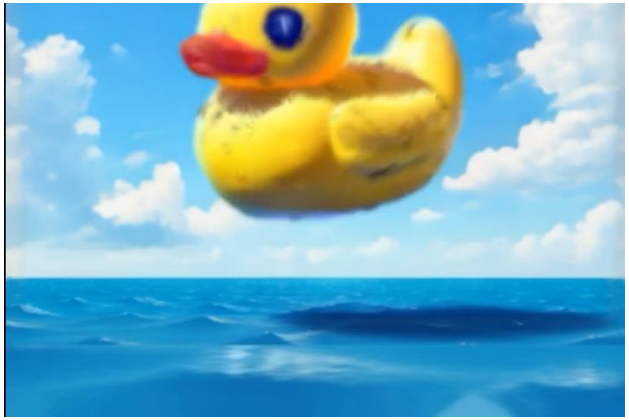
- 기존 방식의 한계
  - Video-Based Method
    - 현실적인 dynamics생성에 강하지만 물리적 action을 잘 제어하지 못함
  - Physics-Based Method
    - 물리적 action을 잘 제어하지만 현실적인 dynamics를 표현하기 어려움



< WonderPlay >



< Video-Based Method >

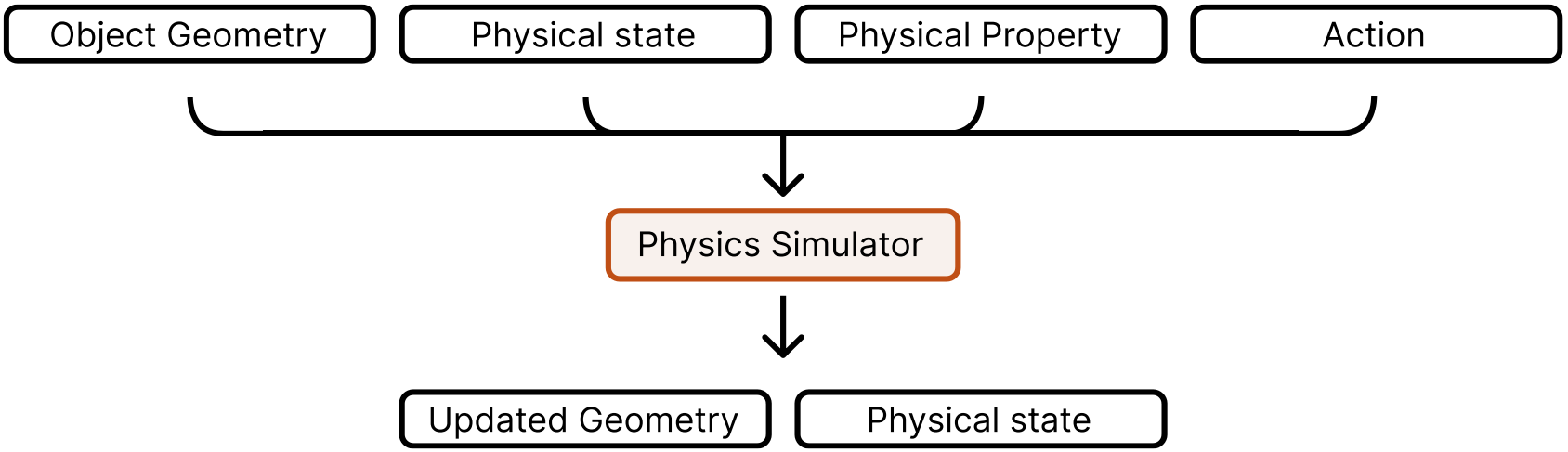


< Physics-Based Method >

CVSP Lab
Computer Vision & Signal Processing

Background

- Physics Simulator
  - 3D object geometry와 현재 physical state, material property, action을 입력으로 받아 다음 physical state를 계산
  - 시간에 따라 반복해 Dynamic 3D Scene 생성



CVSP Lab
Computer Vision & Signal Processing

Background

- Physics solver
  - Simulator 내에 물리 계산을 하는 모듈
  - 물체의 특성에 따라 서로 다른 solver 사용
    - MPM solver
      - 변형되는 물질을 다루기 위한 solver

        Ex) sand, jam
      - Particle : 물체를 이루는 작은 물질 조각

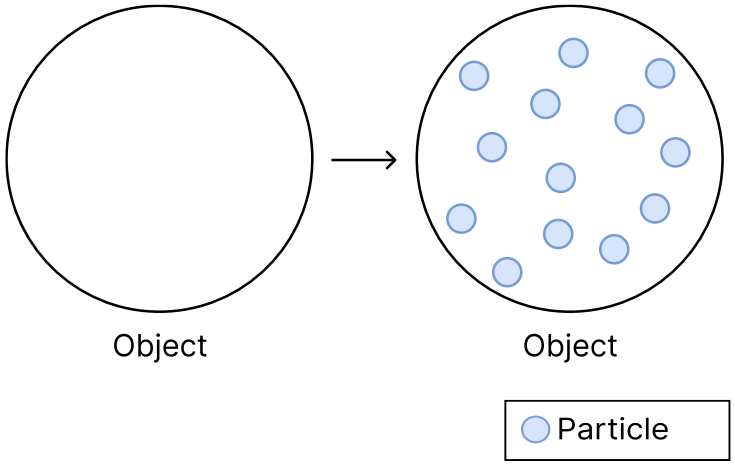
        ✓ 각 particle은 위치, 속도, 변형 등의 물리적 정보를 지님

    - PBD solver
      - 변형되는 물질 중에서도 particle 사이에 거리를 유지하는 물체

        EX) cloth, rope

    - Rigid-body solver
      - 형태가 변하지 않는 물체를 다루기 위한 solver

        Ex) box, ball
      - 물체 전체를 하나의 rigid body로 간주하여 위치, 회전, 각속도 등 변수를 업데이트



CVSP Lab
Computer Vision & Signal Processing

Wonderplay: Dynamic 3D Scene Generation From a Single Image and Actions
ICCV 2025 (Highlight)

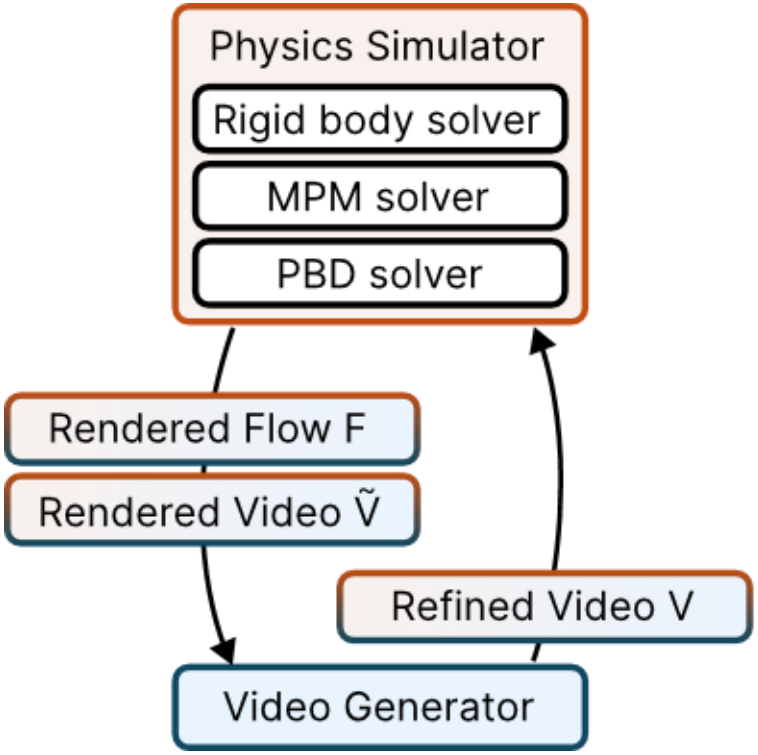
WonderPlay

- Task
  - Action-conditioned Dynamic Scene Generation from Single Image
    - Single Image로부터 Action의 결과를 나타내는 Dynamic 3D Scene 생성

- Motivation
  - 기존 방식의 한계
    - Video-Based Method
      - 현실적인 dynamics생성에 강하지만 물리적 action을 잘 제어하지 못함
    - Physics-Based Method
      - 물리적 action을 잘 제어하지만 현실적인 dynamics를 표현하기 어려움

WonderPlay

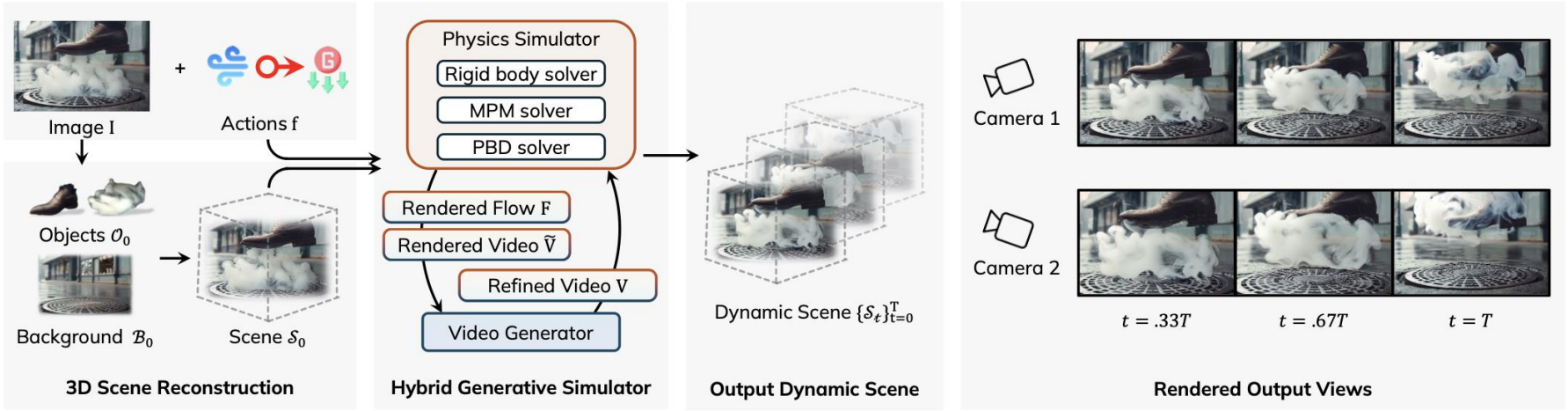
- Motivation
  - Hybrid Generative Simulator
    - Physics Simulator와 Video Generator의 장점을 합친 loop Framework 제안



CVSP Lab
Computer Vision & Signal Processing

WonderPlay

- Method
  - Overview
    - Image (**I**)로부터 정적인 3D Scene ($\mathbf{S_0}$) Reconstruction
    - 초기 3D Scene ($\mathbf{S_0}$) 과 Action (**f**) 을 입력으로 받아 Hybrid Generative Simulator를 통과해 Dynamic 3D Scene 생성 $\{\mathbf{S_t}\}_{\mathbf{t=0}}^{\mathbf{T}}$

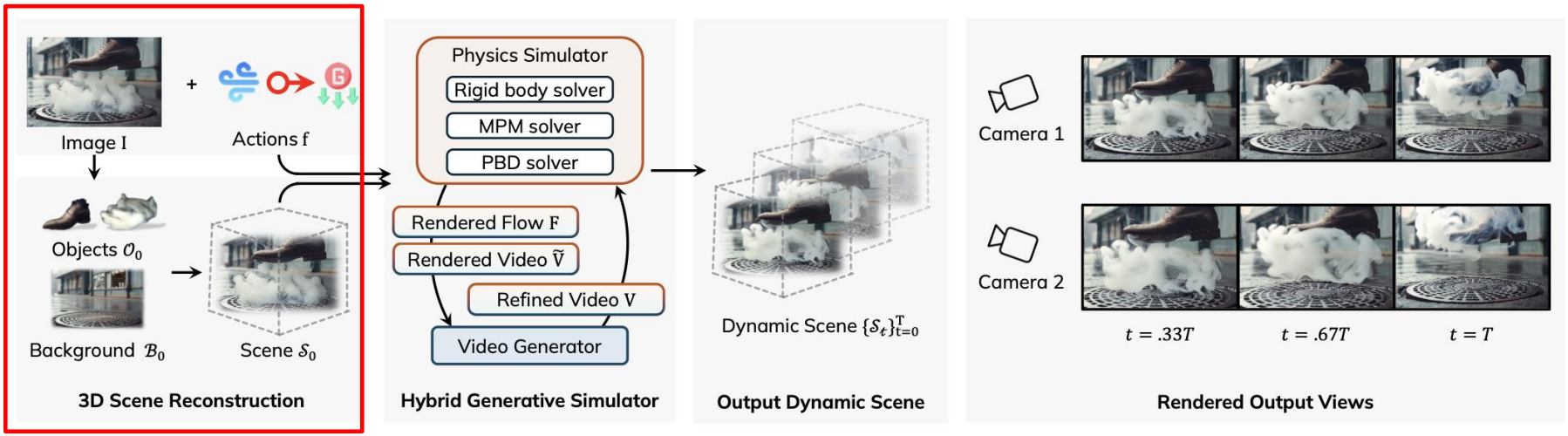


< WonderPlay의 전체 Framework >

CVSP Lab
Computer Vision & Signal Processing

WonderPlay

- Method
  - 3D Scene Reconstruction
    - Single Image (I)로부터 배경 ($\mathbf{B_0}$) 과 객체 ($\mathbf{O_0}$) 를 각각 복원한 뒤 합쳐서 초기 3D Scene ($\mathbf{S_0}$) 형성
      - $S_0 = B_0 \cup O_0$
        - ✓ 정적인 배경 ($\mathbf{B_0}$)은 Gaussian으로 표현
        - ✓ 동적인 객체 ($\mathbf{O_0}$)는 Mesh를 생성한 뒤, 각 vertex에 Gaussian을 결합하여 표현

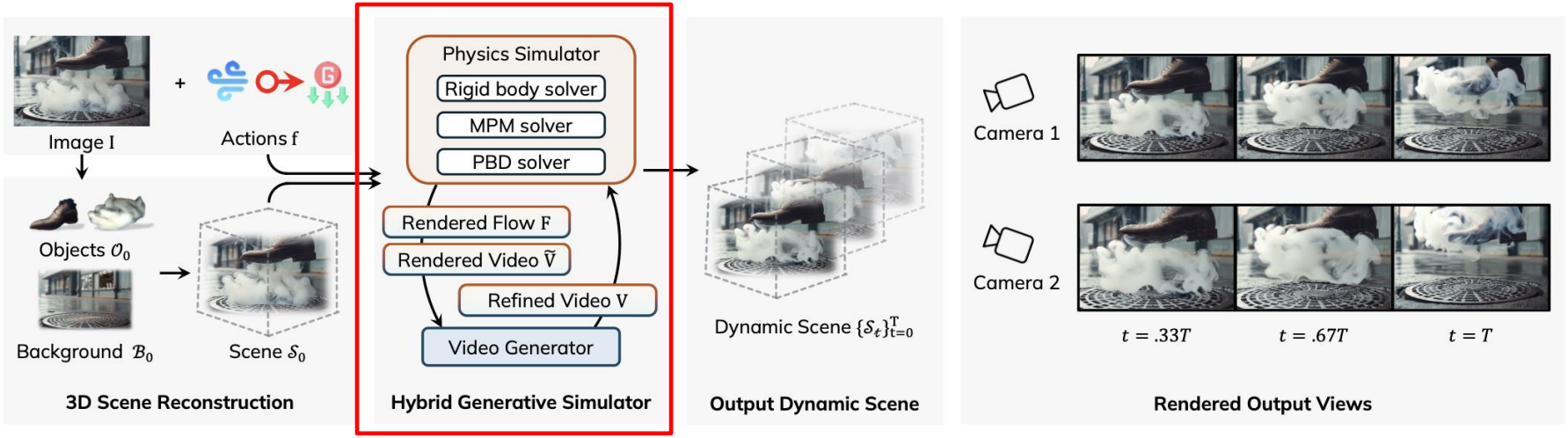


< WonderPlay의 전체 Framework >

CVSP Lab
Computer Vision & Signal Processing

WonderPlay

- Method
  - Hybrid Generative Simulator
    - Physics Simulator
      - 초기 3D Scene ($\mathbf{S_0}$)과 Actions ($\mathbf{f}$)로부터 Coarse Dynamic 3D Scene $\{\tilde{\mathcal{S}}_t\}_{t=0}^{T}$ 형성
      - Video Generator가 읽을 수 있도록 Rendering
        - ✓ Rendered Flow $\mathbf{F}$ – Motion information
        - ✓ Rendered Video $\tilde{\mathbf{V}}$ – Appearance information

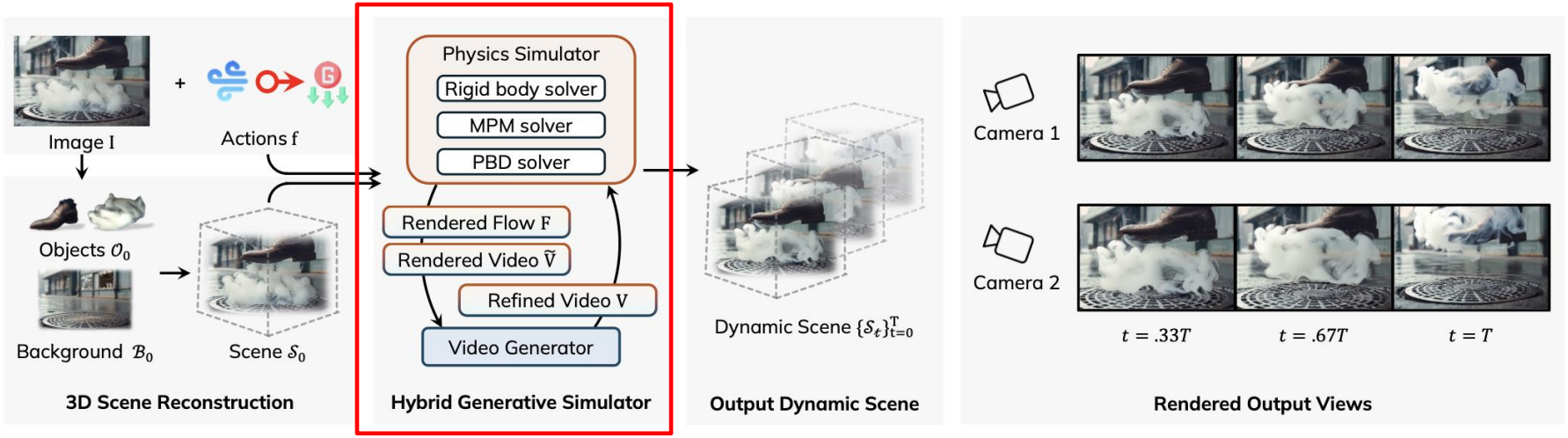


< WonderPlay의 전체 Framework >

CVSP Lab Computer Vision & Signal Processing

WonderPlay

- Method
  - Hybrid Generative Simulator
    - Physics Simulator
      - Physical simulation 결과를 Gaussian에 전달
        - ✓ Rigid body
          - Mesh의 위치와 회전이 update되면 mesh와 연결된 gaussian도 함께 update
        - ✓ MPM / PBD
          - 인접한 10개의 particle의 평균 위치 update로 Gaussian의 위치 update

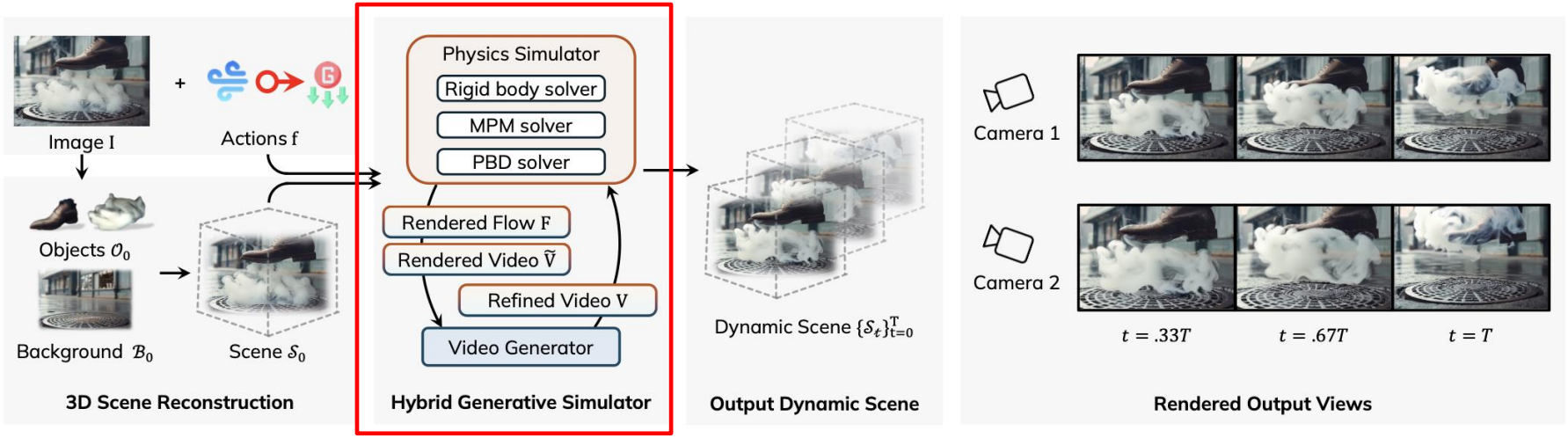


< WonderPlay의 전체 Framework >

CVSP Lab
Computer Vision & Signal Processing

WonderPlay

- Method
  - Hybrid Generative Simulator
    - Video Generator
      - Image ($\mathbf{I}$), Rendered Flow ($\mathbf{F}$), Rendered Video ($\tilde{\mathbf{V}}$)를 condition으로 Refined Video ($\mathbf{V}$) 생성
      - $\mathbf{V} = \mathbf{g}(\mathbf{F}, \tilde{\mathbf{V}}, \mathbf{I})$
        - ✓ $\mathbf{F}$: Rendered Flow $\mathbf{F}$
        - ✓ $\tilde{\mathbf{V}}$: Rendered Video $\tilde{\mathbf{V}}$
        - ✓ $\mathbf{I}$: Image

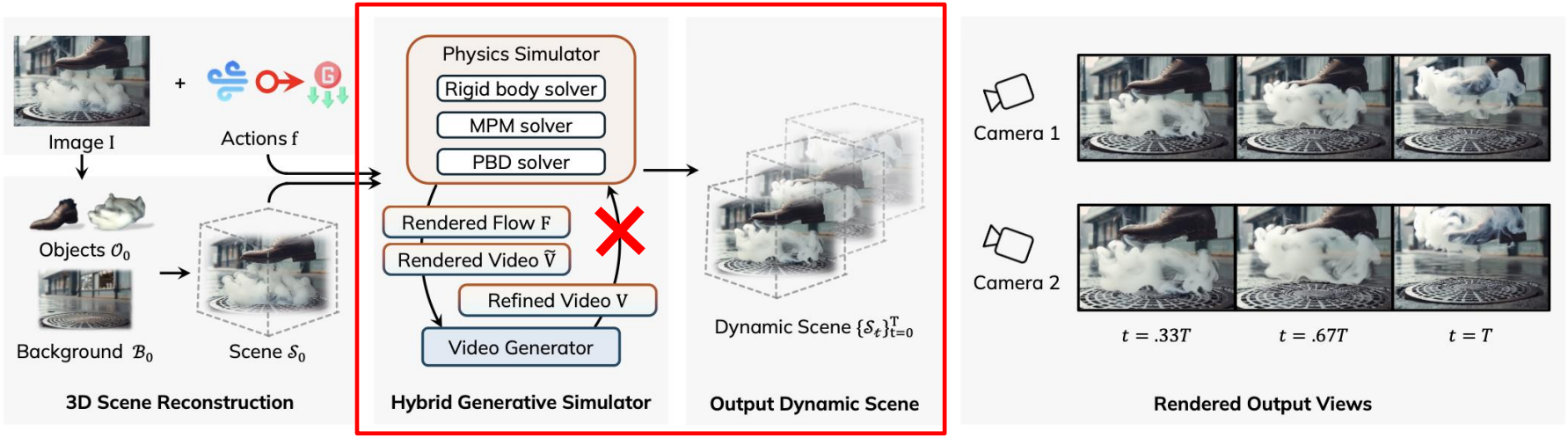


< WonderPlay의 전체 Framework >

CVSP Lab
Computer Vision & Signal Processing

WonderPlay

- Method
  - Output Dynamic Scene
    - Update coarse dynamic scene $\{\tilde{S}_t\}_{t=0}^{T}$
      - Refined video ($\mathbf{V}$)를 supervision으로 사용하여 coarse dynamic scene $\{\tilde{S}_t\}_{t=0}^{T}$ 의 Gaussian 최적화
      - $\mathbf{L_1} = \min_{\{c_t^B, O_t\}_{t=0}^{T}} \|\mathbf{V} - \tilde{\mathbf{V}}\|_1$
        - ✓ $\tilde{\mathbf{V}}$: Rendered Video
        - ✓ $\mathbf{V}$: Refined Video



< WonderPlay의 전체 Framework >

CVSP Lab
Computer Vision & Signal Processing

WonderPlay

- Experiment
  - Physics-Based Method / Video-Based Method와 Metric 비교
    - **Imaging** : 영상의 전반적인 품질 (blur, noise, artifact …)
    - **Aesthetic** : 영상의 전반적인 매력도(색감, 구도, 전체적인 조화 …)
    - **Motion** : 움직임의 자연스러움
    - **Consistency** : Frame간 시간적 일관성
    - **PhysReal** : Dynamics의 물리적 현실성 (GPT-4o 기반)

Methods	Imaging (↑)	Aesthetic (↑)	Motion (↑)	Consistency (↑)	PhysReal (↑)
PhysGen	<u>0.692</u>	0.593	0.992	0.212	0.545
PhysGaussian	0.492	0.564	<u>0.994</u>	0.206	0.350
CogVideoX	0.686	0.574	0.993	**0.219**	0.670
Tora	0.644	**0.620**	0.992	0.210	0.530
Ours	**0.695**	<u>0.610</u>	**0.995**	<u>0.217</u>	**0.700**

CVSP Lab
Computer Vision & Signal Processing

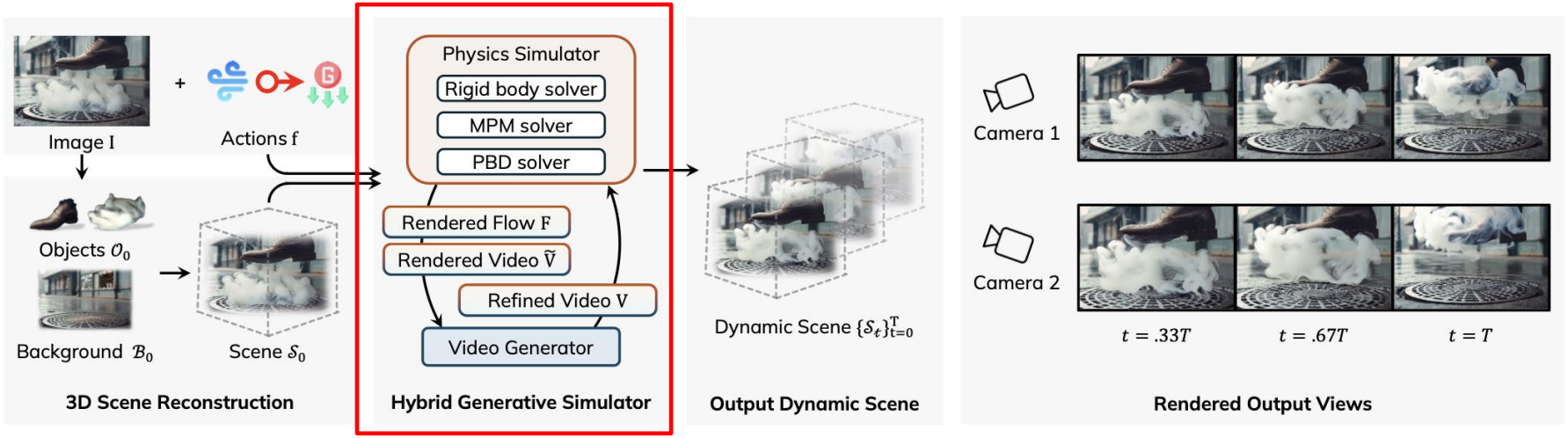
1) Jiang, Zhangqi, et al. "Devils in middle layers of large vision-language models: Interpreting, detecting and mitigating object hallucinations via attention lens." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.

PerpetualWonder: Long-Horizon Action-Conditioned 4D Scene Generation

CVPR 2026 (Highlight)

PerpetualWonder

- Motivation
  - WonderPlay의 한계
    - Particle의 motion은 Gaussian에 전달되지만, video refinement로 update된 Gaussian의 상태는 Particle에 전달되지 않음
      - Visual - Physical State의 Mismatch가 누적되어, action에 대한 장기적인 simulation이 어려움
    - Single-view optimization
      - 3D Dynamics Scene의 novel view에서 geometry inconsistency및 artifact 발생



PerpetualWonder

- Motivation
  - WonderPlay의 한계
    - Particle의 motion은 Gaussian에 전달되지만, video refinement로 update된 Gaussian의 상태는 Particle에 전달되지 않음
      - Visual-Physical State의 Mismatch가 누적되어, 연속적인 action에 대한 장기 simulation이 어려움
    - Single-view optimization
      - 3D Dynamics Scene의 novel view에서 geometry inconsistency및 artifact 발생

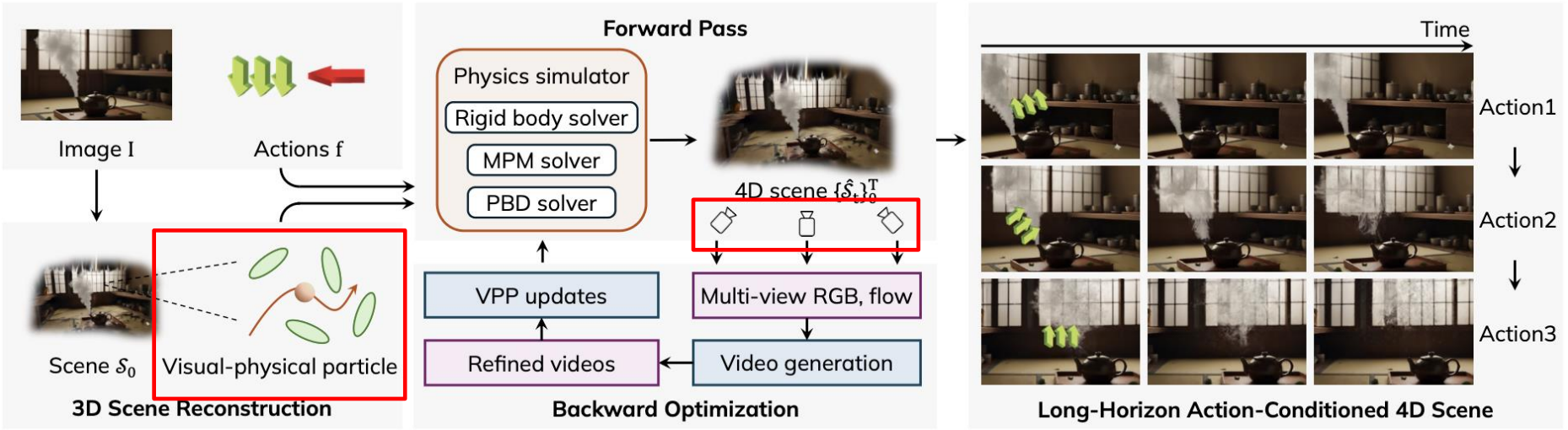


< PerpetualWonder > < WonderPlay >

CVSP Lab
Computer Vision & Signal Processing

PerpetualWonder

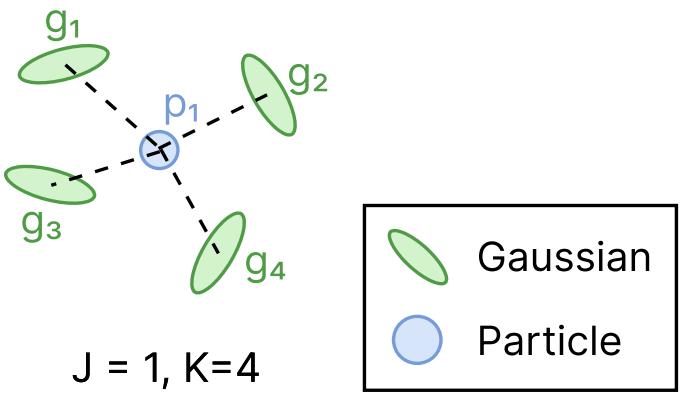
- Method
  - Overview
    - Image (I)로부터 정적인 3D Scene (S_0) Reconstruction
    - 3D Scene (S_0) 과 Actions (f) 을 입력으로 받아 Hybrid Generatvie Simulator를 통과해 Dynamic 3D scene 생성
    - 개선된 점
      - VPP를 통한 Gaussian과 Particle 사이의 bidirectional connection 형성
      - Multi-view optimization을 통해 geometry inconsistency와 artifact 완화



PerpetualWonder

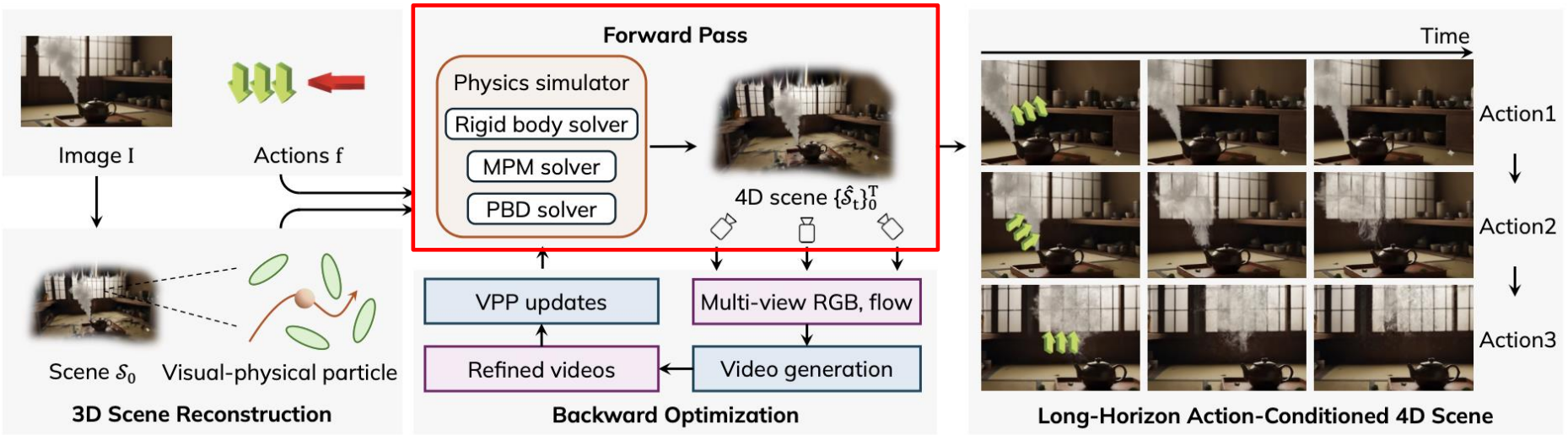
- Method
  - VPP (Visual-Physical Aligned Particle)
    - 각 Particle들을 anchor로 K개의 Gaussian들을 결합해 Physical-Visual representation 통합

- $\mu_{j,k} = p_j + \tanh(\tilde{p}_{j,k}) \cdot \delta$
  - $\mu_{j,k}$: gaussian들의 위치
  - p_j : particle들의 위치
  - $\tilde{p}_{j,k}$: learnable position offset
    - ✓ j : particle index
    - ✓ k : j 번째 particle에 결합된 k번째 gaussian index
  - δ : particle의 크기
    - ✓ Gaussian이 particle로부터 너무 멀리 떨어지지 않게 하기 위함



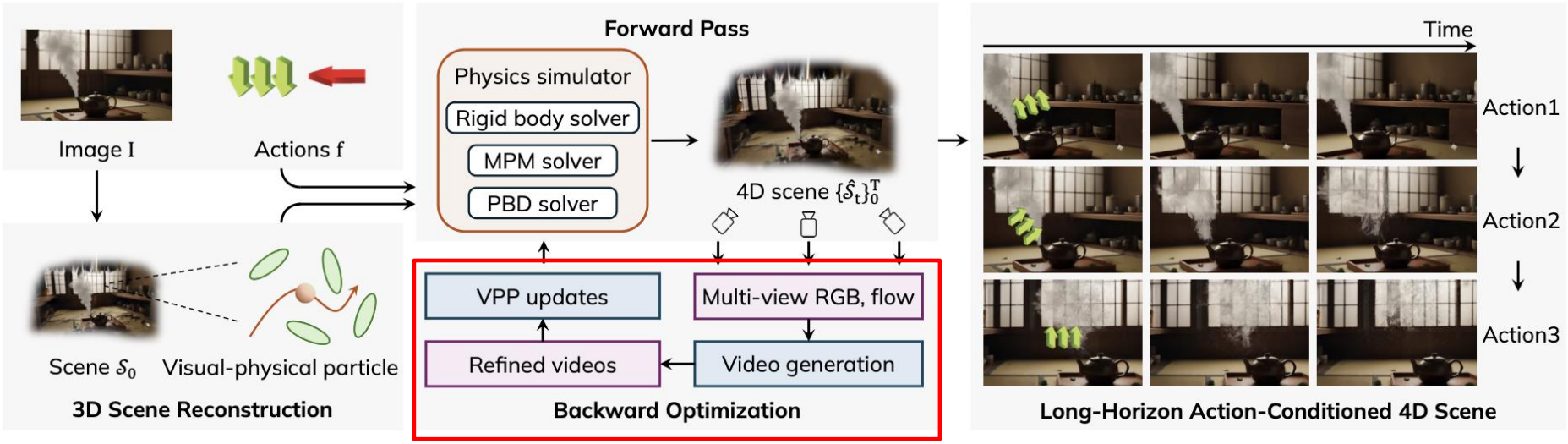
PerpetualWonder

- Method
  - Forward Pass
    - Physics Simulator
      - 초기 3D Scene ($\mathbf{S_0}$)과 Actions ($\mathbf{f}$)로부터 Coarse Dynamic Scene $\{\tilde{\mathcal{S}}_t\}_{t=0}^{T}$ 형성
      - Video Generator가 읽을 수 있도록 Rendering
        - ✓ Multi-view Rendered Flow $\mathbf{F}$
        - ✓ Multi-view Rendered Video $\tilde{\mathbf{V}}$



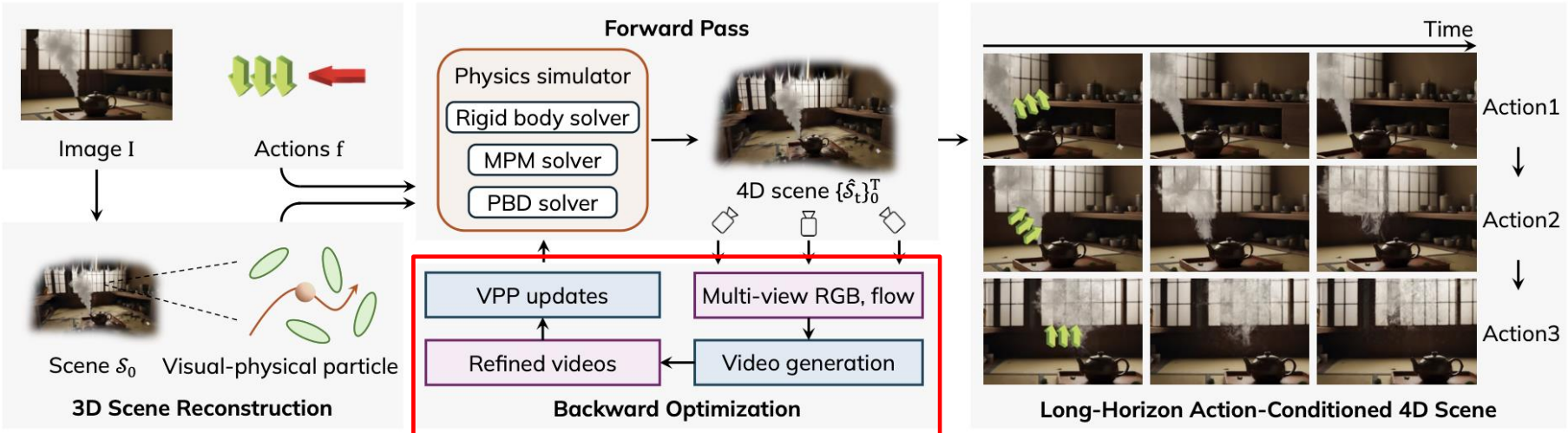
PerpetualWonder

- Method
  - Backward Optimization
    - Video Generation
      - Input Image (**I**), Multi-view Rendered Flow (**F**)와 Rendered Video ($\tilde{\boldsymbol{V}}$)를 condition으로 Refined Videos ($\mathbf{V_t}$) 형성
        - ✓ **F** : Motion information
        - ✓ $\tilde{\mathbf{V}}$: Appearance information



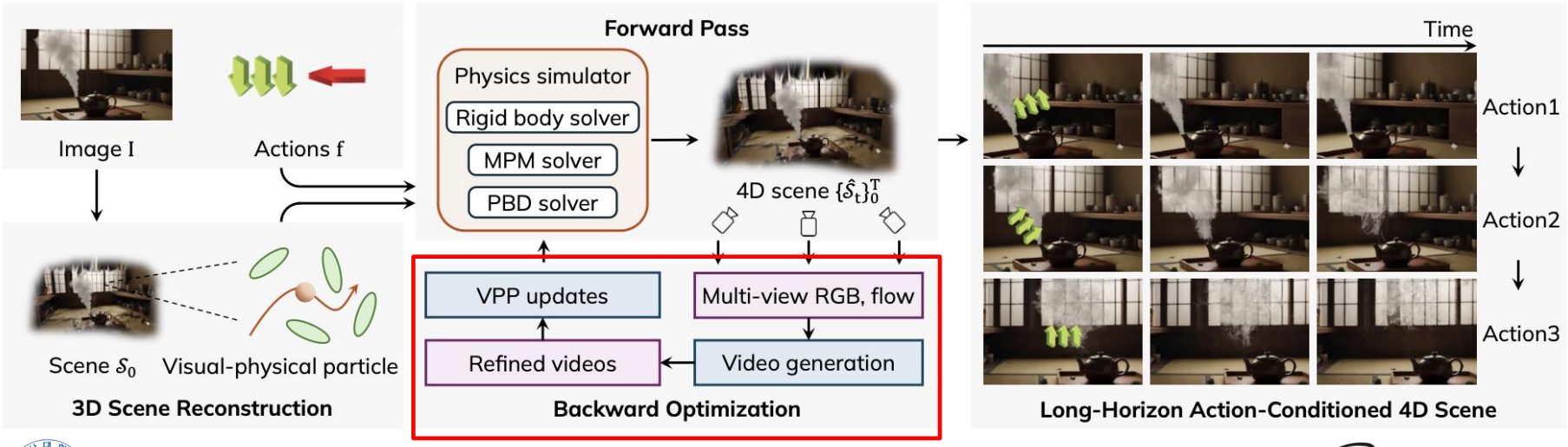
PerpetualWonder

- Method
  - Backward Optimization
    - VPP updates
      - Refined video (**V**)를 supervision으로 사용하여 coarse dynamic scene $\{\tilde{\boldsymbol{S}}_t\}_{t=0}^{T}$ 의 Gaussian 최적화
      - 각 Gaussian의 위치를 평균화 하여 anchor인 Particle의 위치 (p_j)를 update



PerpetualWonder

- Method
  - Backward Optimization
    - Progressive Optimization
      - Image (I)와 동일한 view에서만 Dynamic 3D Scene 최적화
      - Dynamic 3D Scene을 left, right view 에서 Dynamic 3D Scene을 최적화
        - ✓ consistent하면서 artifact가 없는 Dynamic 3D Scene 생성



PerpetualWonder

- Experiment
  - Conditional Video Generation / Hybrid Generative Simulator와 Metric 비교
    - **Camera Ctrl:** 지정된 camera 경로대로 viewpoint가 변하는 정도
    - **3D Consistency:** viewpoint가 변해도 scene geometry와 appearance가 일관되게 유지되는지
    - **Imaging:** 영상의 전반적인 품질 (blur, noise, artifact …)

Method	Camera Ctrl	3D Consist	Imaging
Wan2.2 [40]	59.73	65.35	<u>67.03</u>
GEN3C [34]	<u>80.29</u>	61.69	66.25
WonderPlay [21]	75.95	63.93	36.80
Tora [59]	51.80	60.77	54.37
Wan2.6 [40]	64.75	70.49	66.09
DaS [12]	78.96	62.18	60.23
Veo3.1 [44]	60.61	<u>73.93</u>	**67.82**
PerpetualWonder (ours)	**93.26**	**80.41**	66.98

CVSP Lab
Computer Vision & Signal Processing

Conclusion

- WonderPlay
  - 기존의 Video-Based Method, Physics-Based Method의 한계를 극복하는 Hybrid Generative Simulator 라는 Framework를 제안
  - 한계
    - Dynamic 3D Scene update 과정에서 Particle들의 위치가 update되지 않는다는 한계
    - Novel view에서 geometry inconsistency및 artifact 발생

- PerpetualWonder
  - VPP와 Multi-view optimization을 통해 WonderPlay의 한계 해결

CVSP Lab
Computer Vision & Signal Processing