THE 21ST IEEE INTERNATIONAL SYMPOSIUM ON MIXED AND AUGMENTED REALITY

ISMAR 2022 SINGAPORE

17-21 OCTOBER 2022

INDIAN INSTITUTE OF SCIENCE

भारतीय विज्ञान संस्थान

Temporal View Synthesis of Dynamic Scenes through 3D Object Motion Estimation with Multi-Plane Images

Nagabhushan Somraj

Pranali Sancheti

Rajiv Soundararajan

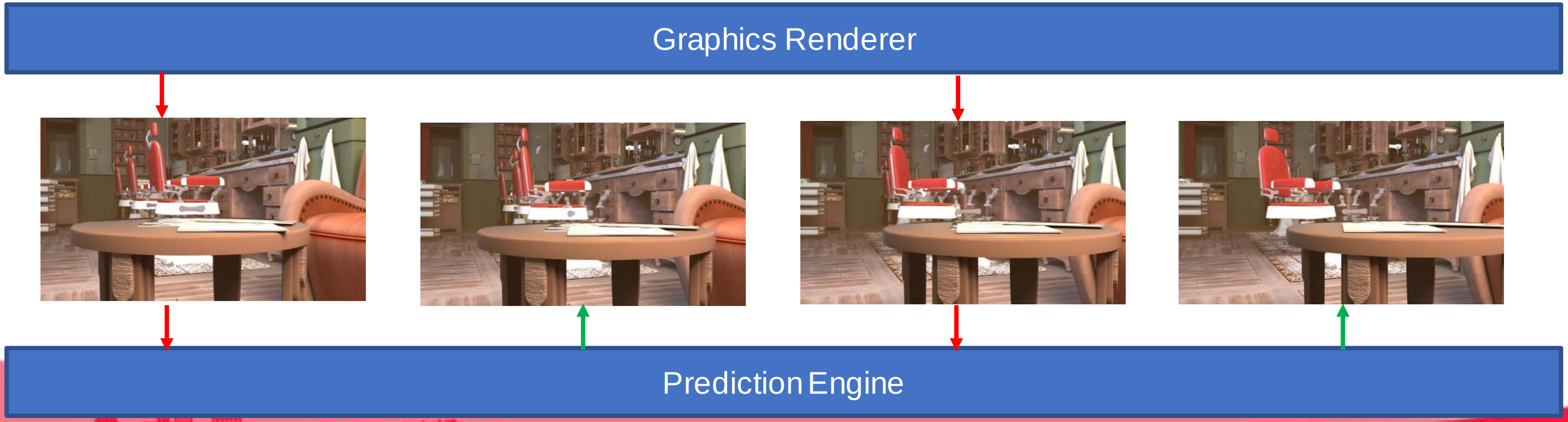
Indian Institute of Science, Bengaluru, India.

Funding Acknowledgement: Qualcomm India Pvt. Ltd.

IEEE COMPUTER SOCIETY IEEE vgtc

Temporal View Synthesis (TVS)

- Consider a user exploring a virtual environment on a head mounted display.
- Can we generate next frame using past frames and next head position?
- Applications: Frame-rate upsampling of graphics videos in low compute devices or natural videos in remote presence applications.



Temporal View Synthesis (TVS)

Two different settings based on motion in the scene:

- Static Scene: Only camera motion
  - Challenge: Synthesizing disoccluded regions. (Kanchana et al. WACV 2022.)
- Dynamic Scene: Both camera and object motion

  Challenges:
  - Predicting object motion.
  - Effective use of camera motion.
  - Infilling disoccluded regions.

Focus of our work

Static Scene

Dynamic Scene

Related Work

View Synthesis

- Synthesizes scene from any novel view-point given the scene from a few view-points.
- Does not predict object motion in dynamic scenes.

Novel View Synthesis

- Use volumetric scene representations when depth is unavailable.
- Multi-Plane Images (Zhou et al. TOG 2018), Neural Radiance Fields (Mildenhall et al. ECCV 2020)

Dynamic View Synthesis

- Synthesizes given frame from novel view-point of a monocular dynamic video (Li et al. CVPR 2021).

Video Prediction

- Predicts future frames of a video given past frames.
- Does not use camera motion and depth.

Direct Frame Prediction

- Use sequential models (LSTMs) to capture past frames in latent representation, which is used to predict future frames (Villegas et al. 2017).

Motion Prediction as Optical Flow

- DPG (Gao et al. ICCV 2019): Predicts future motion as optical flow and infills disocclusions.
- **Suitable to incorporate camera motion.**

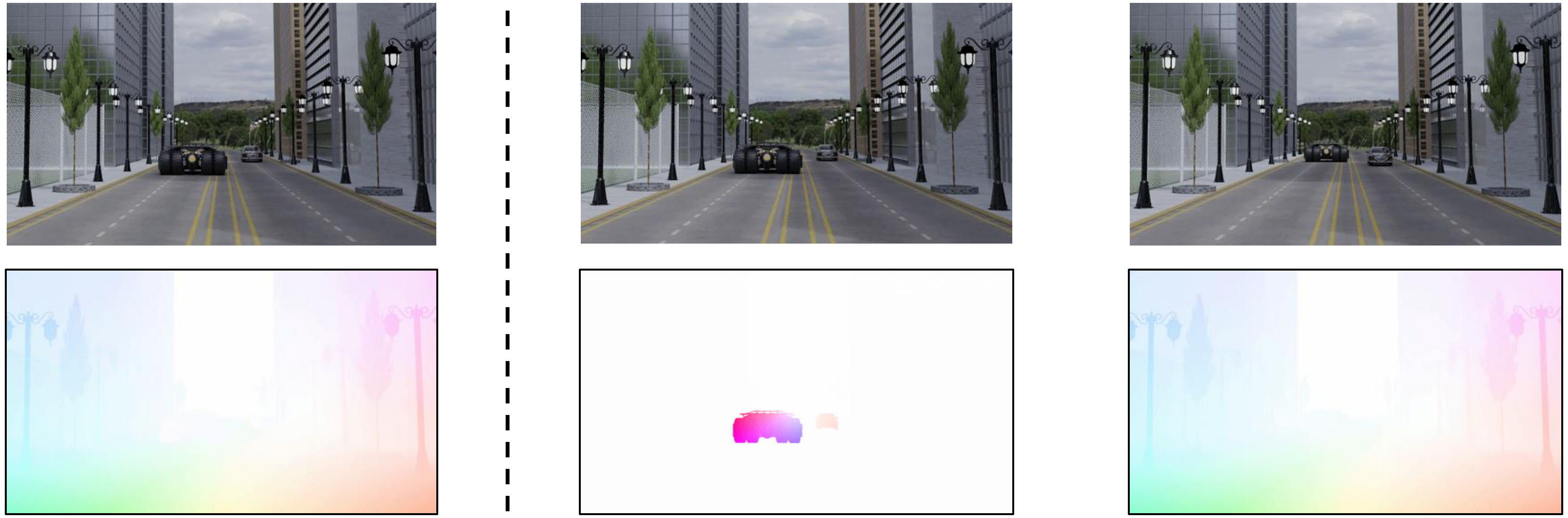
Motion Decomposition

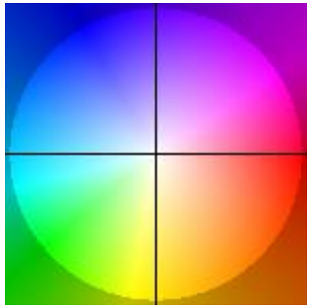
- Decompose overall motion into object and camera motion.
  - Predict object motion and apply camera motion using known camera poses.

Combined motion | Object motion | Camera motion

Frames

Optical Flow





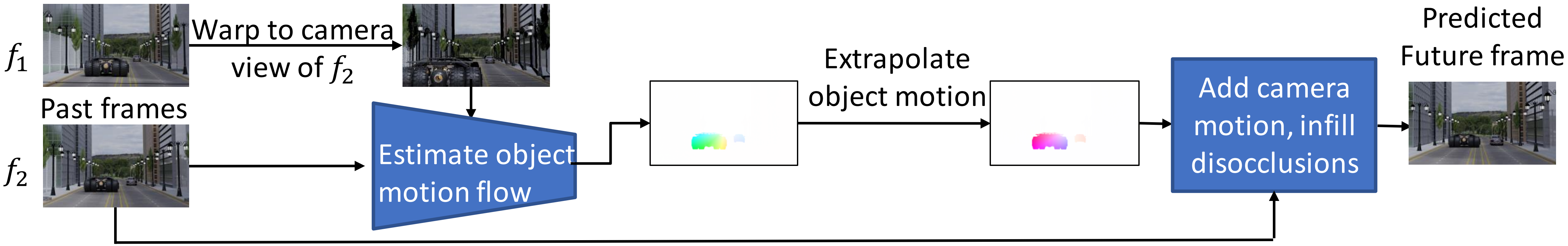
Optical flow visualization color wheel

Difficult to accurately predict object motion

Object motion prediction is easier

Object Motion Isolation and Prediction

- Estimate object motion between past frames and extrapolate it.
- Challenge: Camera and object motion are mixed in past frames.
- Solution: Warp all past frames to same camera view.
  - Isolates object motion.



Contribution 1: Decomposing Motion into camera and object motion and the isolation of object motion between the past frames.

Object Motion Estimation

- Problem: Flow estimation in occluded regions is incorrect.

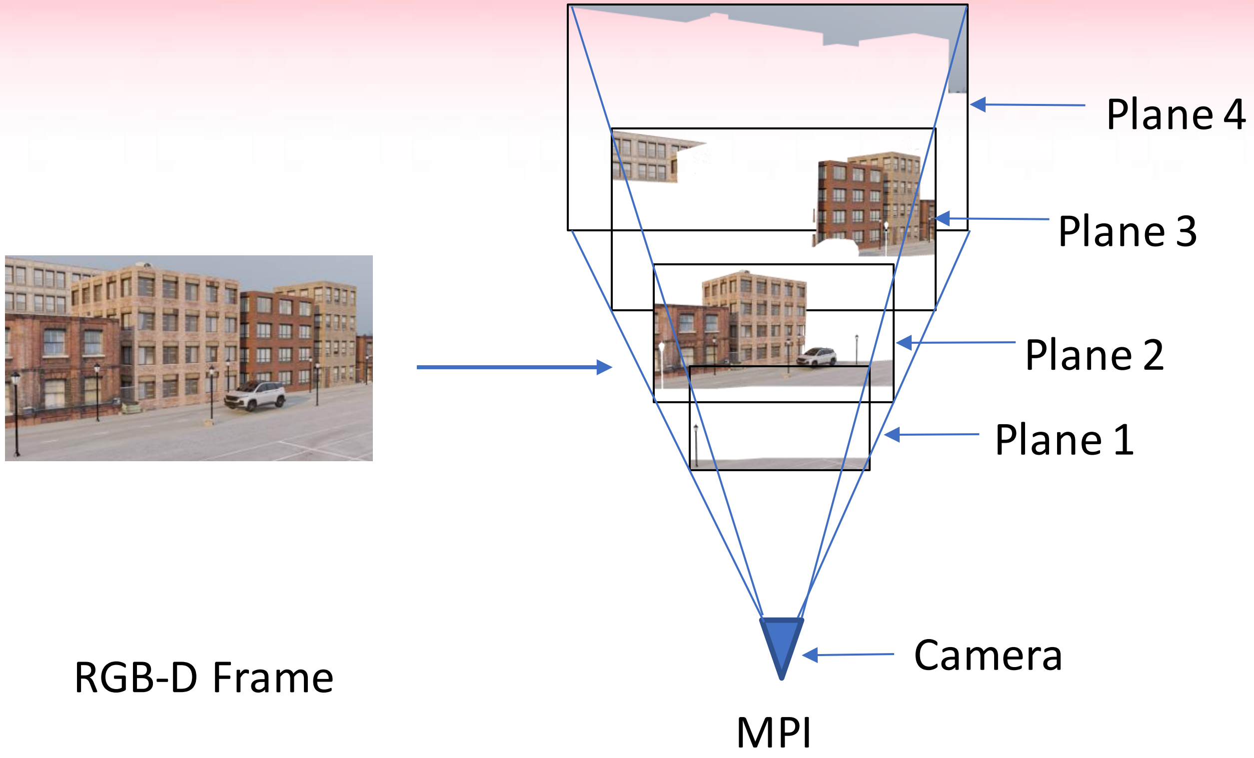
- Reason: Occluded regions do not have matching points.
  - Flow estimation is guided by spatial smoothness
  - Occluded flow depends on flow in both foreground and background.
- <u>Our solution</u>: Use a 3D scene representation – Multi-Plane Images.
  - Pushes foreground and background objects apart.
  - Occluded flow depends on flow in background only.

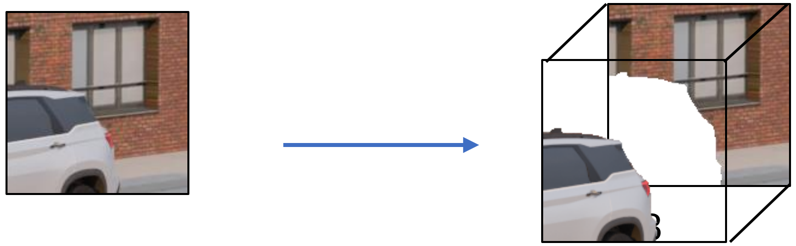
Contribution 2

Multi-Plane Images (MPIs)

- Splits a single RGB frame into multiple planes at different depths.

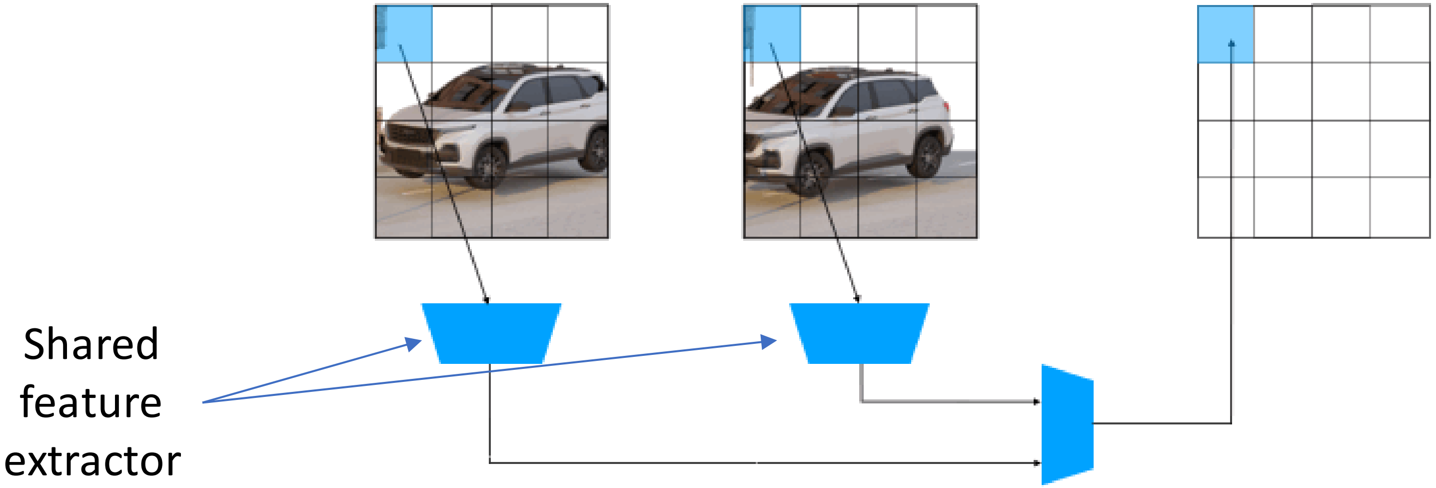
- Moving car and static building are separated onto different planes.





Flow Estimation with MPIs

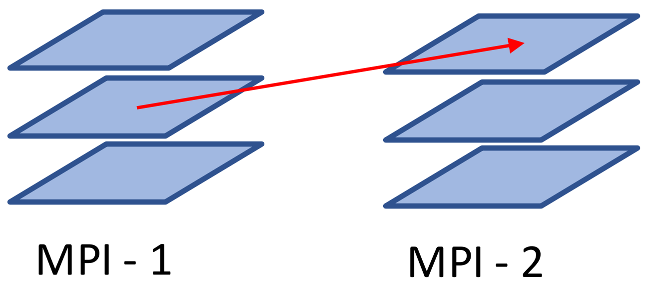
- Flow Estimation includes finding correspondences through correlation between different regions of input MPIs.
- Problem: Empty regions can cause incorrect correspondences.



- Our Solution: Masked correlation layers and partial convolution layers.

Flow Estimation with MPIs

- Objects can move across different depth planes.
- Problem: MPI has discrete depth planes.
- Resolution along depth is much lower compared to resolution along height and width.
- Solution: Estimate flow as probability distribution of motion across planes.



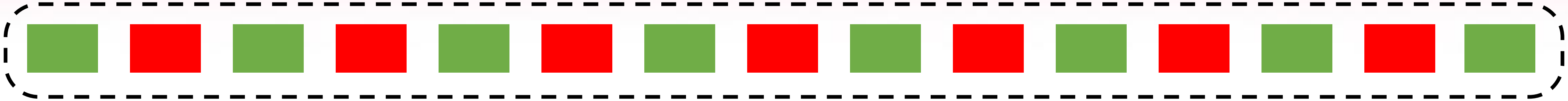
IISc VEED-Dynamic Database*

- No existing large-scale database of dynamic videos with necessary ground truth – frames, depth and camera pose.
- Generate videos using blender at high spatial and temporal resolutions: 1920x1080 at 30fps.
- 200 unique scenes, 800 videos in total.

*Indian Institute of Science Virtual Environment Exploration Database for Dynamic scenes

Single Frame Prediction

Sequence of frames



Input video (15 fps)

DPG (30 fps)

Ours (30 fps)

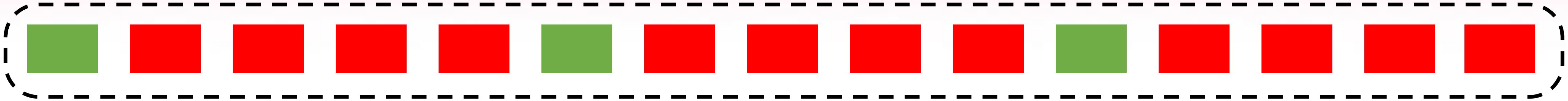
Ground Truth (30 fps)

Graphically Rendered frames

Predicted frames

Multi-Frame Prediction

Sequence of frames



Input video (6 fps)

DPG (30 fps)

Graphically Rendered frames

Predicted frames

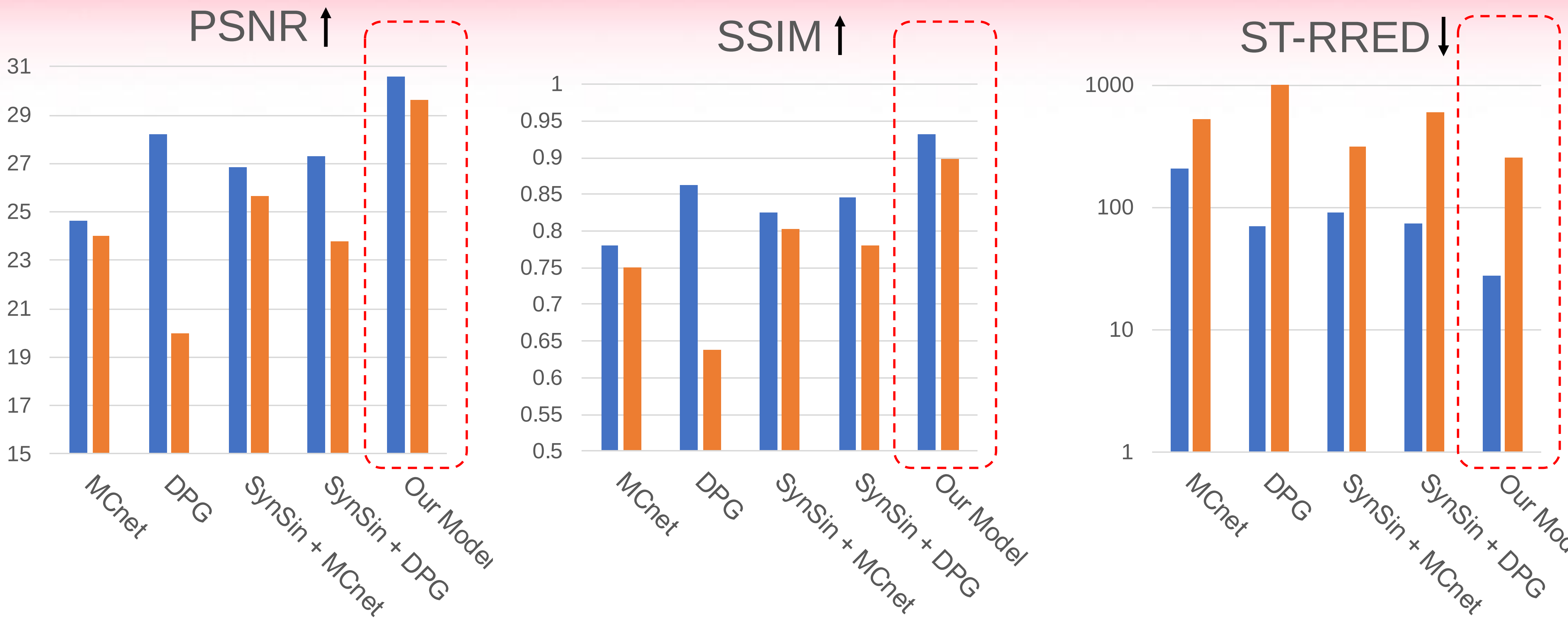
Ours (30 fps)

Ground Truth (30 fps)

Quantitative Evaluation

- Datasets:
  - Ours: 135 train scenes, 65 test scenes.
  - MPI-Sintel (Butler *et al. ECCV* 2012): 13 train scenes, 10 test scenes.
- Quality assessment measures:
  - Frame-level:
    - Peak Signal to Noise Ratio (PSNR)
    - Structural Similarity (SSIM)
  - Video-level:
    - ST-RRED (Soundararajan *et al.* CSVT 2013)

Quantitative Evaluation – Single Frame Prediction



IISc VEED-Dynamic Database

MPI-Sintel Database

Flow Estimation w/ and w/o MPI

Complete frame

Predicted w/o MPI

Predicted w/ MPI

Ground Truth

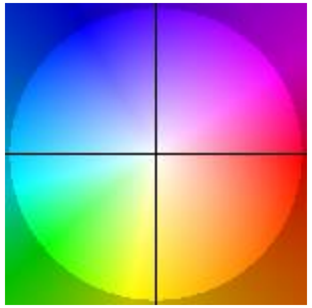
Past frames

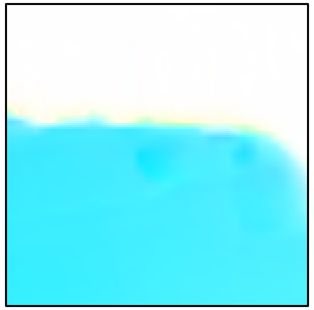
Future predicted frame

Disoccluded regions

Distorted background

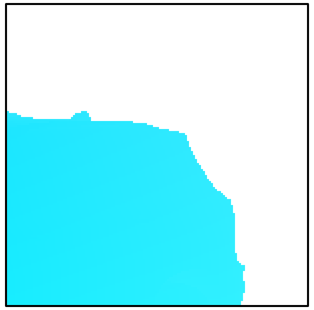
Optical flow visualization color wheel





Predicted flow is not sharp

Sharp edge in predicted flow



Average end point error in estimated flow decreases by **38%** on using MPIs for flow estimation.

Conclusion and Future Work

- Developed a framework for frame-rate upsampling of synthetic dynamic videos by decoupling global and local motion.
- Designed model to predict local object motion by estimating object motion in 3D using multi plane images.
- Designed a challenging database and achieved state-of-the-art performance.
- Future Work:
  - Extend the framework to natural videos – depth may not be available.

ISMAR 2022
SINGAPORE
17-21 OCTOBER 2022

IEEE

IEEE COMPUTER SOCIETY

IEEE vgtc