Sync-NeRF: Generalizing Dynamic NeRFs to Unsynchronized Videos

Seoha Kim*[1] Jeongmin Bae*[1] Youngsik Yun[1] Hahyun Lee[2] Gun Bang[2] Youngjung Uh[1] * Equal contribution

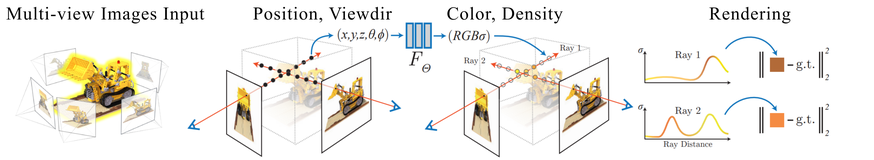
[1] Yonsei University [2] Electronics and Telecommunications Research Institute

Project Page

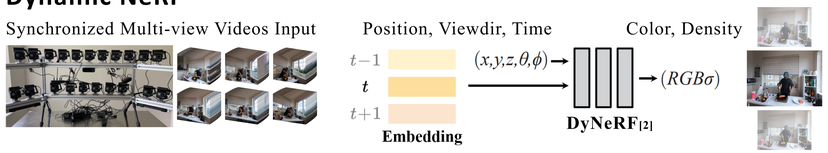
AAAI 2024
Association for the Advancement of Artificial Intelligence

Preliminaries

NeRF[1]



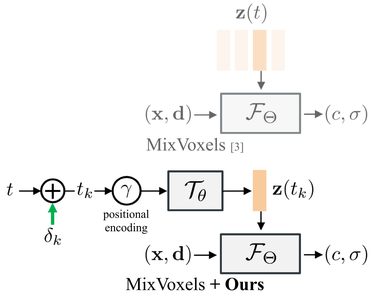
Dynamic NeRF



Method

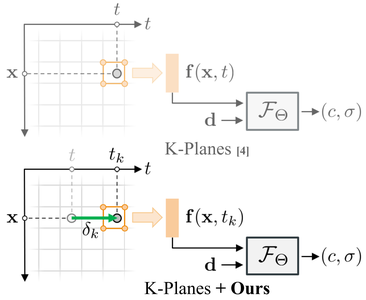
Implicit function-based approach

Use implicit function instead of discrete embeddings



Grid-based approach

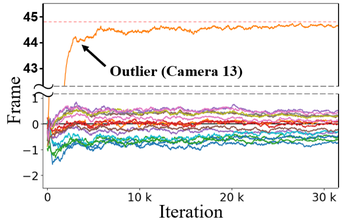
Query the grid at time t_k calibrated by the offset δ_k



Learnable Time Offset

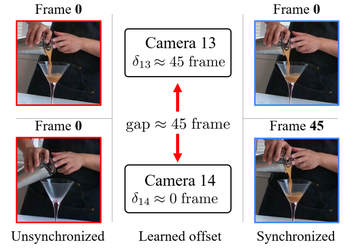
Convergence

Our method efficiently optimizes the offset in the early iteration.

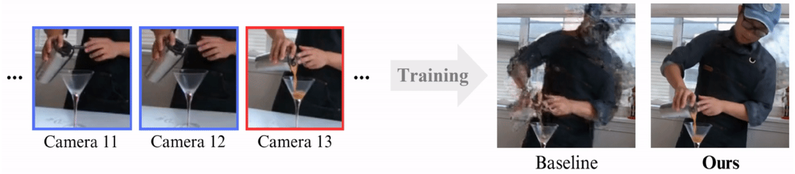


Synchronize by Offset

Learned offset is equivalent to automatically synchronizing the videos by shifting.

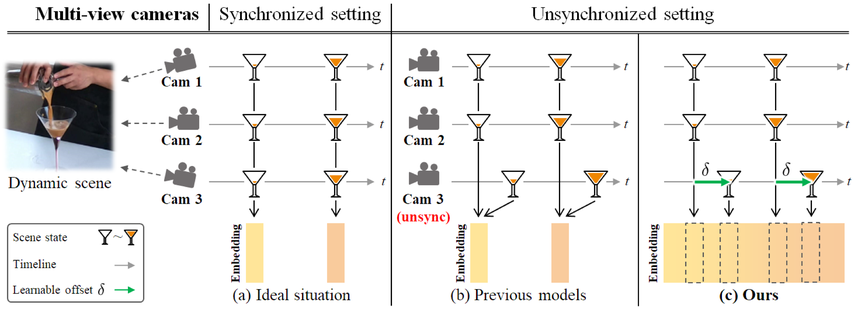


Our Goal



To successfully represent the dynamic scene from **unsynchroznied** multi-view videos.

Overview

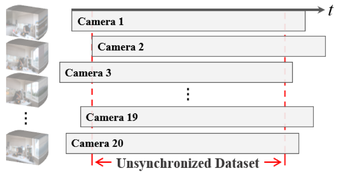


In an unsynchronized setting, previous methods suffer from the temporal mismatch. Ours resolve this problem by introducing **learnable time offsets δ** for each cameras.

Experiments

Unsynchronized Dataset

We unsynchronize the scene by randomly translating the multi-view videos. We use overlapped 270 frames out of a total of 300 frames.



Quantitative Results

model \ metric		PSNR	SSIM	LPIPS$_{alex}$
MixVoxels		29.96	0.9059	0.1669
Sync-MixVoxels		**30.53**	**0.9101**	**0.1570**
K-Planes	hybrid	29.16	0.9120	0.1278
Sync-K-Planes	hybrid	**30.44**	**0.9243**	**0.1064**
K-Planes	explicit	28.51	0.9042	0.1484
Sync-K-Planes	explicit	**29.97**	**0.9223**	**0.1144**

Unsynchronized Plenoptic Video Dataset

model \ scene		2.0× cook spinach	2.0× flame salmon	2.0× fox
MixVoxels		30.48	27.66	29.81
Sync-MixVoxels		**31.50**	**28.74**	**35.58**
K-Planes	hybrid	29.13	26.01	29.71
Sync-K-Planes	hybrid	**31.85**	**28.22**	**40.36**
K-Planes	explicit	28.46	24.83	29.10
Sync-K-Planes	explicit	**31.16**	**27.09**	**40.29**

Unsynchronization lengths x2

model \ metric		PSNR	SSIM	LPIPS$_{alex}$
MixVoxels		31.06	0.9753	0.0237
Sync-MixVoxels		**37.11**	**0.9841**	**0.0120**
K-Planes	hybrid	32.66	0.9771	0.0175
Sync-K-Planes	hybrid	**39.40**	**0.9863**	**0.0066**
K-Planes	explicit	32.15	0.9751	0.0246
Sync-K-Planes	explicit	**39.35**	**0.9865**	**0.0072**

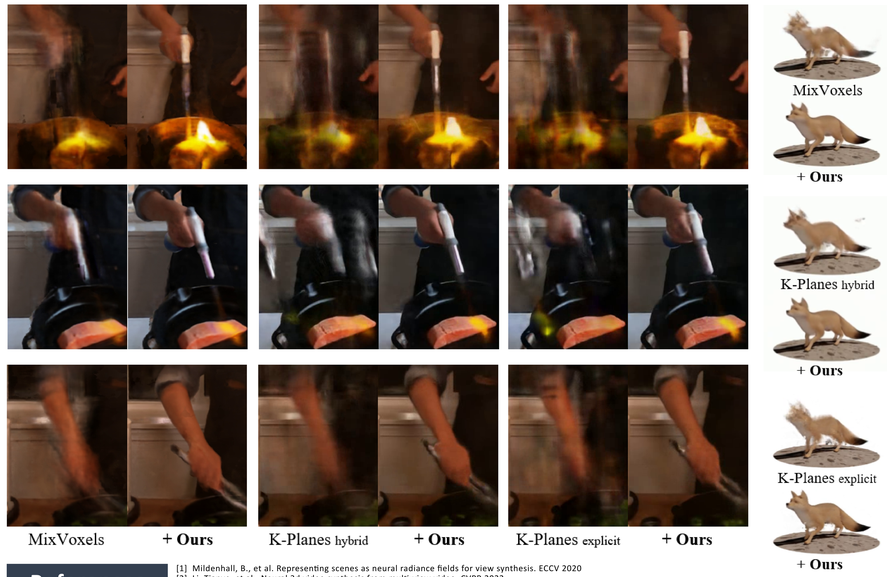
Unsynchronized Blender Dataset

		MAE (seconds)
Sync-MixVoxels		0.0154
Sync-K-Planes	hybrid	0.0156
Sync-K-Planes	explicit	0.0229

Ours achieves about 0.01 sec of MAE with ground truth offsets

Accuracy of time offset

Qualitative Results



References

[1] Mildenhall, B., et al. Representing scenes as neural radiance fields for view synthesis. ECCV 2020
[2] Li, Tianye, et al., Neural 3d video synthesis from multi-view video. CVPR 2022
[3] Wang, Feng, et al. Mixed neural voxels for fast multi-view video synthesis. ICCV 2023
[4] Fridovich-Keil, Sara, et al. K-planes: Explicit radiance fields in space, time, and appearance. CVPR 2023