Supplementary Material – PREF: Predictability Regularized Neural Motion Fields

Liangchen Song[12], Xuan Gong[12], Benjamin Planche[2], Meng Zheng[2], David Doermann[1], Junsong Yuan[1], Terrence Chen[2], and Ziyan Wu[2]

[1] University at Buffalo, Buffalo NY, USA
[2] United Imaging Intelligence, Cambridge MA, USA
first.last@united-imaging.com

A Scope and Requirement Comparison

We compare the settings in our paper with other similar works in Tab. S1.

Table S1. Comparison of methods for 3D motion estimation. *Topologically varying scenes* mean that new surfaces may appear mid sequence; *physical motion* means that methods estimate the motion of real 3D points, as opposed to canonical motion.

Method	Topologically Varying Scenes	Physical Motion	Precomputed Data-driven Prior
D-NeRF [6]	✗	✓	None
Nerfies [4]	✗	✓	None
NR-NeRF [7]	✗	✓	None
DCT-NeRF [8]	✗	✓	Depth & Optical Flow
NSFF [3]	✓	✓	Depth & Optical Flow
VideoNeRF [9]	✓	✓	Depth
NeRFlow [1]	✓	✓	Optical Flow
HyperNeRF [5]	✓	✗	None
PREF (Ours)	✓	✓	None

B Implementation and Experimental Details

B.1 Architecture

We use the same architecture as in NeRF for the space-time field: The positional encoding of the input location (x) and the timestamp t is passed through 8 fully-connected ReLU layers. Each FC layer is with 256 channels. A skip connection is added to concatenate the input to the 5th layer's activation. Note that different from NeRF, in our experiments, the feature vector from 8th layer is not concatenated with the viewing directions. The predictor consists of 5 fully

connected layers with a width of 128 and ReLU activations. The input of the predictor is the concatenation of three motion weights. The motion field uses the same architecture as the space-time field, except the inputs are now location and motion embedding and the outputs are the motion vector.

B.2 Dataset

We select four representative clips from the Panoptic dataset [2]. The details of the four clips are as follows:

	full name	starting frame	end frame
SPORTS	`161029_sports1`	3600	4000
TOOLS	`161029_tools1`	3700	4100
IAN	`160906_ian5`	600	1000
CELLO	`171026_cello3`	220	620

Full name indicates the name of the full sequence in the dataset. Starting frame and end frame are the frame index acquired with the official video-to-image conversion scripts.

C Additional Experimental Results

C.1 Validation on 2D Toy Example

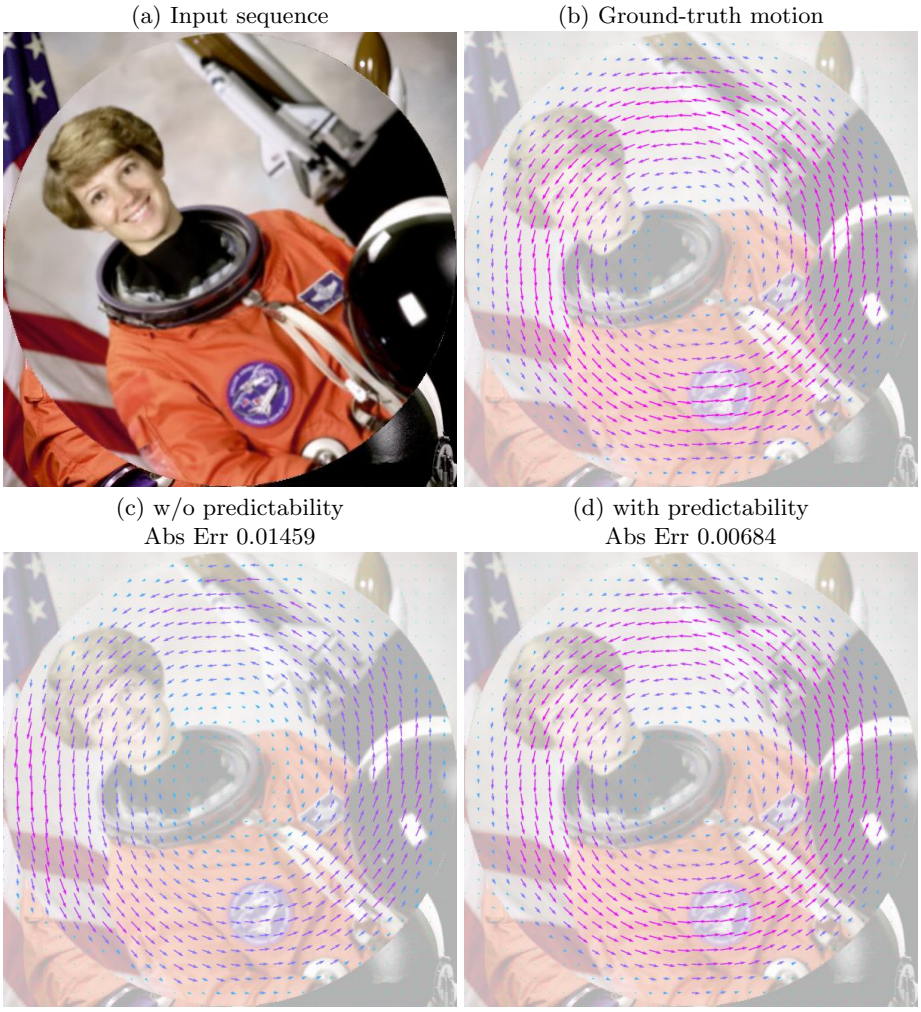
Fig. S1 provides some quantitative results on a 2D toy dataset for validating the proposed predictability regularization.

C.2 Video Results

Attached to this document, readers can find a video illustrating our methods and its results on a variety of dataset. A high-definition video is available at pref.uiius.com. The video is encoded with H.264 codec, which can be downloaded here: `https://www.codecguide.com/download_kl.htm`.

C.3 Pose Tracking Visualization

We provide additional visualizations of the tracked pose in Fig. S2. We can observe that some provided body joints are wrong in Panoptic, such as the top two rows of the SPORTS sequence. This is caused by the inaccurate 3D detector used by Panoptic as the 3D joints are not manually labeled.



To see the animation, please view the document with compatible software, *e.g.*, *Adobe Acrobat* or *KDE Okular*; otherwise, the animation is also provided as a separate file attached to the submission.

Fig. S1. Ablation study on a toy example. A 2D neural motion field (pixel coordinate uv as inputs) is studied. Two deformation patterns are applied alternatively. Abs Err is the mean absolute difference between the estimated and GT motion.

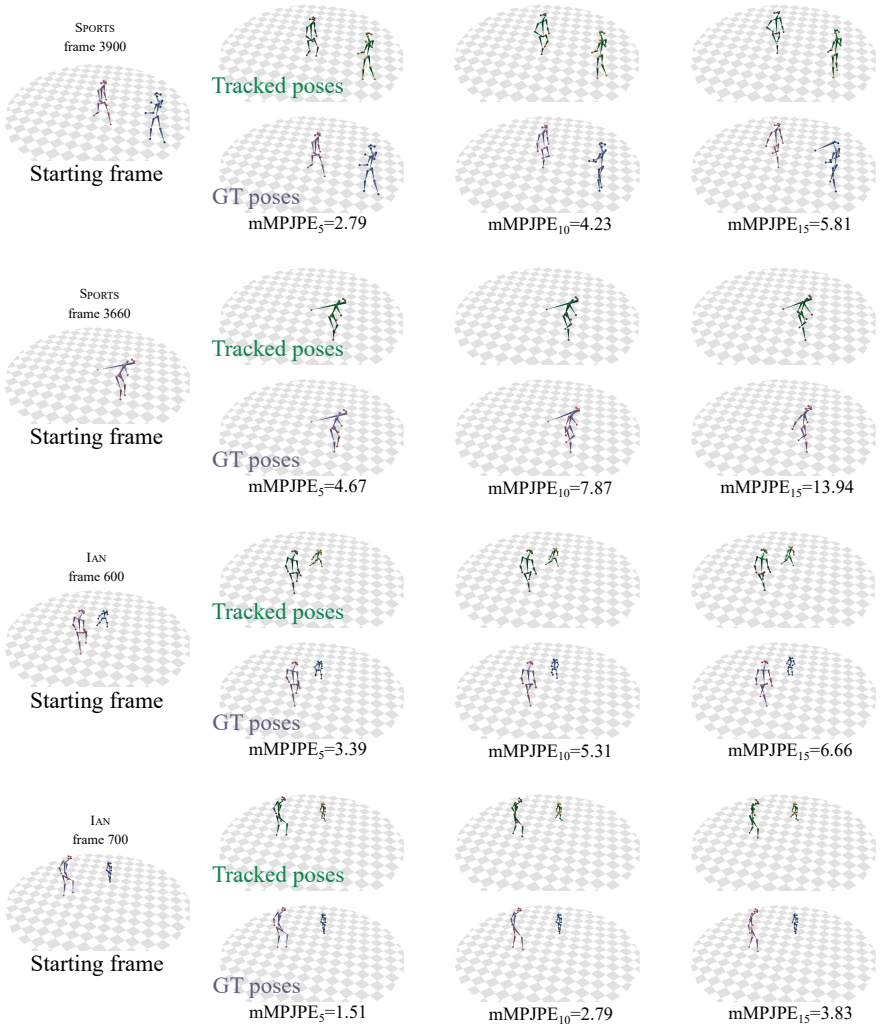


Fig. S2. Visualization of motion tracking results on the Panoptic dataset.

References

1. Du, Y., Zhang, Y., Yu, H.X., Tenenbaum, J.B., Wu, J.: Neural radiance flow for 4d view synthesis and video processing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2021)
2. Joo, H., Liu, H., Tan, L., Gui, L., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., Sheikh, Y.: Panoptic studio: A massively multiview system for social motion capture. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3334–3342 (2015)
3. Li, Z., Niklaus, S., Snavely, N., Wang, O.: Neural scene flow fields for space-time view synthesis of dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021)
4. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5865–5874 (October 2021)
5. Park, K., Sinha, U., Hedman, P., Barron, J.T., Bouaziz, S., Goldman, D.B., Martin-Brualla, R., Seitz, S.M.: Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. ACM Trans. Graph. **40**(6) (dec 2021)
6. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-nerf: Neural radiance fields for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10318–10327 (2021)
7. Tretschk, E., Tewari, A., Golyanik, V., Zollhöfer, M., Lassner, C., Theobalt, C.: Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12959–12970 (October 2021)
8. Wang, C., Eckart, B., Lucey, S., Gallo, O.: Neural trajectory fields for dynamic novel view synthesis. arXiv preprint arXiv:2105.05994 (2021)
9. Xian, W., Huang, J.B., Kopf, J., Kim, C.: Space-time neural irradiance fields for free-viewpoint video. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9421–9431 (2021)