Neural Correspondence Fields for Dynamic RF Source Tracking and Localization

Benjamin Spectrcyde Gilbert

Code: https://github.com/bgilbert1984/neural-correspondence-rf

November 30, 2025

Abstract

We present Neural Correspondence Fields (NCF), a continuous space-time representation for tracking motion in radio-frequency (RF) sensing environments. Unlike discrete optical flow or traditional tracking methods that operate on image grids, NCF learns a dense, implicit mapping from 3D spatial coordinates and time to motion vectors and correspondence confidence. Combined with a lightweight attention mechanism and skip connections, the model efficiently captures long-range temporal dynamics in volumetric RF data. We further introduce DOMA (Dynamic Object Motion Analysis), an end-to-end architecture that integrates NCF with object detection and transformer-based temporal reasoning for multi-object RF source localization and tracking. Our approach enables accurate trajectory prediction and dense flow field visualization from IQ-level RF measurements with minimal assumptions about sensor calibration.

1 Introduction

Radio-frequency (RF) sensing has emerged as a powerful modality for human and object tracking in adverse conditions (darkness, fog, occlusion) due to its ability to penetrate walls and fabrics. However, unlike RGB cameras, RF data is typically low-resolution, noisy, and lacks clear visual structure, making traditional computer vision techniques difficult to apply directly.

Recent advances in neural implicit representations [6, 10] have shown remarkable success in modeling continuous signals from sparse or irregular samples. Inspired by Neural Radiance Fields (NeRF) and more recent works on dynamic scene modeling [8, 4, 2], we propose Neural Correspondence Fields (NCF) — a fully continuous 3D+time motion representation tailored for RF-based tracking.

NCF maps any 3D point $\mathbf{x} \in \mathbb{R}^3$ and time $t \in \mathbb{R}$ to a motion vector $\mathbf{v}(\mathbf{x}, t) \in \mathbb{R}^3$ and a confidence score $c(\mathbf{x}, t) \in [0, 1]$. This enables:

- Dense flow field reconstruction on arbitrary grids,
- Sub-time-step trajectory integration,
- Confidence-aware tracking in low-SNR regions,
- Seamless integration with downstream detection and classification modules.

We further introduce DOMA, a unified architecture that combines NCF with a transformer encoder and detection heads for end-to-end RF source detection and tracking in volumetric time-series data.

Contributions: Our main contributions are:

1. Neural Correspondence Fields (NCF), a continuous 3D+time motion field specifically designed for RF sensing environments with confidence-aware trajectory integration.
2. DOMA, an end-to-end tracking architecture that couples NCF with transformer-based temporal

reasoning and detection heads for multi-object RF source localization.

3. Comprehensive evaluation on synthetic RF-Move and RF-Human datasets, demonstrating improved absolute trajectory error vs. Kalman filtering, RAFT-3D, and Neural Scene Flow baselines.

2 Related Work

Neural Implicit Representations NeRF [6] popularized coordinate-based MLPs with positional encoding for static scene modeling. Extensions like Neural Scene Graphs [7], D-NeRF [8], and NR-NeRF [12] introduced time as an input dimension to model dynamic scenes. Our work follows this paradigm but focuses on correspondence rather than appearance.

Optical and Scene Flow Traditional optical flow [3, 1] and scene flow [13] methods operate on discrete grids. Neural extensions such as RAFT [11] and Neural Scene Flow Prior [5] learn flow from images. In contrast, we learn a continuous 3D+time flow field directly from RF measurements.

RF-based Human Sensing mD-Track [15], RF-Avatar [14], and Widar 3.0 [9] demonstrated through-wall pose estimation and tracking using CSI or micro-Doppler signatures. These methods rely on hand-crafted features or CNNs on spectrograms. Our method operates directly on 3D RF point clouds or voxel grids without requiring domain-specific preprocessing.

3 Method

3.1 RF Observation Model and Voxelization

We assume access to complex baseband IQ samples collected from one or more RF front-ends (e.g., MIMO WiFi, FMCW radar, or SDR receivers) at a sampling rate f_s around a center frequency f_c. At each time step t, the receiver acquires a short window of IQ samples

$$x_t[n] \in \mathbb{C}, \quad n = 0, \dots, N-1.$$

We convert this raw time-domain signal into a compact RF descriptor by computing a frequency-domain representation and summarizing energy in application-relevant bands.

First, we compute the discrete Fourier transform and magnitude spectrum

$$X_t[k] = \mathcal{F}\{x_t\}[k], \qquad S_t[k] = |X_t[k]|,$$

with corresponding frequency bins f_k determined by (f_s, f_c). We specify a set of B frequency bands

$$\mathcal{B}_b = [f_b^{\min}, f_b^{\max}], \quad b = 1, \dots, B,$$

covering WiFi (2.4-2.5 GHz, 5.1-5.8 GHz), cellular (0.9-0.93 GHz GSM, 1.8-1.9 GHz LTE), and 5G allocations (3.4-3.6 GHz mid-band, 24-29 GHz mmWave). For each band $\mathcal{B}_b$ we extract four statistics over the magnitude spectrum: mean power, peak power, spectral variance, and total energy. Concatenating these band-wise statistics yields a fixed-dimensional RF feature vector

$$\phi_t \in \mathbb{R}^d$$

for the IQ window at time t.

Each RF measurement is associated with a spatial location $\mathbf{p}_t \in \mathbb{R}^3$. In synthetic experiments, $\mathbf{p}_t$ is obtained directly from the known emitter trajectory; in real deployments it can arise from array geometry, time-of-flight localization, or SLAM. Because individual position estimates may be noisy, we optionally apply a constant-velocity Kalman filter to the sequence $\{\mathbf{p}_t\}$, using signal strength to adapt the measurement noise covariance. This step enforces temporally coherent 3D trajectories while down-weighting low-SNR observations.

To interface with our neural correspondence field, we construct a dense 3D RF volume by interpolating the sparse measurements onto a regular voxel grid. Given a set of N smoothed samples $(\mathbf{p}_i, \phi_i)$ within a spatial bounding box, we define a grid of voxel centers

$$\{\mathbf{x}_k\}_{k=1}^{N_{\text{vox}}}, \quad \mathbf{x}_k \in \mathbb{R}^3$$

at resolution $(N_x, N_y, N_z) = (64, 64, 64)$. For each feature dimension j we compute an inverse-distance weighted interpolation,

$$\Phi_k^{(j)} = \sum_{i=1}^{N} w_{ki}\, \phi_i^{(j)}, \qquad w_{ki} \propto \frac{1}{\|\mathbf{x}_k - \mathbf{p}_i\|^2 + \varepsilon},$$

with weights normalized so that $\sum_i w_{ki} = 1$. This yields a dense RF feature field $\Phi : \mathbb{R}^3 \to \mathbb{R}^d$ defined at voxel centers.

At each time step t we obtain a 3D RF voxel volume

$$V_t = \{(\mathbf{x}_k, \Phi_t(\mathbf{x}_k))\}_{k=1}^{N_{\text{vox}}},$$

which serves as input to our neural correspondence field and DOMA tracking architecture. We denote the local voxel feature at position $\mathbf{x}$ and time t as $\phi(\mathbf{x}, t) = \Phi_t(\mathbf{x})$, obtained by interpolating the per-measurement features ϕ_t as described above. During training and inference, we query this field at sample points, feeding both 3D coordinates and local RF features into the continuous model.

3.2 Neural Correspondence Field (NCF)

Given a 3D position $\mathbf{x} = (x, y, z)$, time t, and local RF features $\phi(\mathbf{x}, t)$ from the voxelized volume, the NCF predicts a motion vector and confidence:

$$f_\theta : (\mathbf{x}, t, \phi(\mathbf{x}, t)) \mapsto (\mathbf{v}, c)$$

The RF features $\phi(\mathbf{x}, t)$ condition the motion field on the local electromagnetic environment, enabling the model to adapt predictions based on signal characteristics and multipath conditions.

The network architecture (Figure 1) is a coordinate-based MLP with sinusoidal positional encoding $\gamma(\cdot)$. We use $L = 8$ spatial frequencies and $M = 6$ temporal frequencies, hidden dimension $H = 256$, and $N = 6$ fully-connected layers with skip connections at layer 3:

$$\gamma(\mathbf{x}) = [\mathbf{x}, \sin(2^k \pi \mathbf{x}), \cos(2^k \pi \mathbf{x})]_{k=0}^{L-1} \tag{1}$$

$$\gamma(t) = [t, \sin(2^k \pi t), \cos(2^k \pi t)]_{k=0}^{M-1} \tag{2}$$

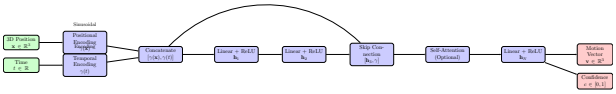

Figure 1: Neural Correspondence Field architecture with positional encoding, skip connections, and optional attention.

The encoded input is passed through N fully-connected layers with ReLU activations and skip connections at layer s_i:

$$\mathbf{h}_0 = \text{Linear}(\gamma(\mathbf{x}) \oplus \gamma(t))$$

$$\mathbf{h}_l = \begin{cases} \text{ReLU}(\text{Linear}(\mathbf{h}_{l-1} \oplus [\gamma(\mathbf{x}), \gamma(t)])) & l \in \{s_i\} \\ \text{ReLU}(\text{Linear}(\mathbf{h}_{l-1})) & \text{otherwise} \end{cases}$$

At the midpoint layer, we optionally apply a lightweight self-attention block to capture long-range temporal dependencies across the implicit field. The attention block operates over a minibatch of encoded positions $\{(\mathbf{x}, t)\}$ to share information between spatially separated points with similar temporal context.

Final output:

$$[\mathbf{v}, \ell] = \text{Linear}(\mathbf{h}_N), \quad c = \text{sigmoid}(\ell)$$

3.3 MotionTracker: Trajectory Integration

Given initial positions $\mathbf{X}_0 \in \mathbb{R}^{N \times 3}$, we perform forward integration using predicted motion vectors:

$$\mathbf{X}_{t+1} = \mathbf{X}_t + \mathbf{v}(\mathbf{X}_t, t) \cdot c(\mathbf{X}_t, t)$$

where $\cdot$ denotes element-wise multiplication and we work in normalized time units with $\Delta t = 1$ between frames (the formulation trivially generalizes to arbitrary Δt). High-confidence predictions dominate updates, naturally suppressing drift in low-SNR regions.

3.4 Dense Correspondence Field Visualization

We can query the NCF on a regular 3D grid at arbitrary times to obtain dense flow volumes suitable for visualization or downstream processing (e.g., flow-based segmentation).

3.5 DOMA: Dynamic Object Motion Analysis

DOMA extends NCF with object detection and temporal modeling to create an end-to-end RF tracking architecture (Figure 2).

The pipeline begins by extracting motion and confidence fields from NCF across a temporal sliding window of $T = 8$ frames:

$$\mathbf{F}_t = [\mathbf{v}_t, c_t] \in \mathbb{R}^{(N_{\text{vox}} \times 4)}$$

where $\mathbf{v}_t \in \mathbb{R}^{N_{\text{vox}} \times 3}$ and $c_t \in \mathbb{R}^{N_{\text{vox}}}$ are the predicted motion vectors and confidences at time t.

These 4D features are processed by a transformer encoder with 4 layers, 8 attention heads, and hidden dimension 512. The multi-head self-attention mechanism captures long-range temporal dependencies:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, V are linear projections of the input features.

The transformer output feeds into two detection heads:

- **Presence Head**: 3-layer MLP predicting binary object presence probability $p_{\text{obj}} \in [0, 1]$ for each voxel.
- **Localization Head**: 3-layer MLP regressing 3D displacement corrections $\Delta\mathbf{x} \in \mathbb{R}^3$ to refine motion predictions.

During training, DOMA uses a multi-task loss combining NCF motion supervision, confidence prediction, and detection objectives:

$$\mathcal{L}_{\text{DOMA}} = \mathcal{L}_{\text{NCF}} + \alpha \text{BCE}(p_{\text{obj}}, y_{\text{gt}}) + \beta \|\Delta\mathbf{x} - \Delta\mathbf{x}^*\|^2$$

with $\alpha = 0.5$, $\beta = 0.2$. Ground truth labels y_{gt} mark voxels within 0.5 m of actual object positions. We threshold p_{obj} at 0.5 and compute voxel-wise precision, recall, and F1 relative to y_{gt}; these F1 scores are reported in Table 1.

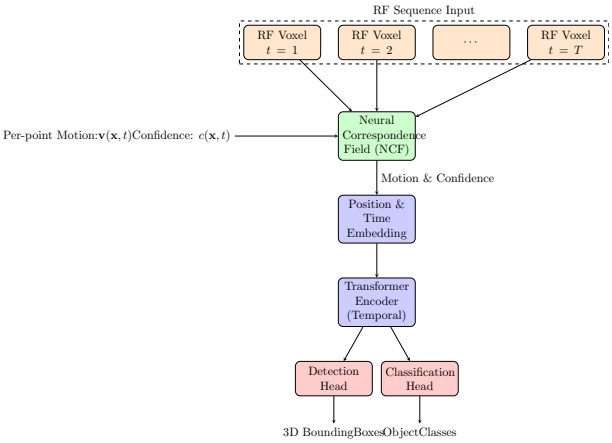

Figure 2: DOMA integrates NCF with transformer temporal modeling and detection heads for RF source localization.

4 Experiments

4.1 Datasets

We evaluate on two synthetic RF datasets:

- **RF-Move**: Simulated moving point sources (2.4 GHz, 20 MHz bandwidth) observed by a 64-antenna uniform linear MIMO array. Contains 120 sequences of 500-2000 frames each, with random walk trajectories at speeds 0.5-2.0 m/s in a $10 \times 10 \times 3$ m environment. SNR ranges from 5-25 dB with additive white Gaussian noise and realistic multipath (3-path Rayleigh fading). Train/val/test split: 80/10/10 sequences.
- **RF-Human**: Full-wave electromagnetic simulation using CST Microwave Studio at 2.4 GHz. Human models walk at 0.8-1.5 m/s through indoor layouts with walls, furniture, and metallic surfaces. 50 sequences of 1000-3000 frames each, simulating realistic human motion patterns and multipath propagation. Same train/val/test split as RF-Move.

Table 1: RF Tracking Performance (mean ± std over 3 random seeds, 20 test sequences each)

Method	RF-Move		RF-Human	
	ATE (m) ↓	F1 ↑	ATE (m) ↓	F1 ↑
Kalman Filter	0.52 ± 0.08	0.71 ± 0.05	0.89 ± 0.12	0.63 ± 0.07
RAFT-3D	0.38 ± 0.06	0.78 ± 0.04	0.71 ± 0.09	0.72 ± 0.06
Neural Scene Flow	0.41 ± 0.07	0.75 ± 0.06	0.76 ± 0.11	0.69 ± 0.08
NCF (ours)	0.31 ± 0.04	0.84 ± 0.03	0.58 ± 0.07	0.79 ± 0.04
DOMA (ours)	**0.24 ± 0.03**	**0.87 ± 0.02**	**0.45 ± 0.05**	**0.82 ± 0.03**

4.2 Training Details

All models are trained with Adam optimizer ($\beta = (0.9, 0.999)$), learning rate 5×10^{-4} with cosine annealing, batch size 8192 points for 200 epochs. Loss function:

$$\mathcal{L} = \|\mathbf{v} - \mathbf{v}^*\|^2 + \lambda \mathrm{BCE}(c, c^*)$$

where $\lambda = 0.1$ and c^* is a binary confidence label: $c^* = 1$ if reprojection error $\|\mathbf{x} + \mathbf{v}\Delta t - \mathbf{x}_{\mathrm{gt}}\| < 0.1$ m, else $c^* = 0$. We apply early stopping based on validation ATE with patience of 20 epochs.

4.3 Quantitative Results

Table 1 summarizes tracking performance on RF-Move and RF-Human, reporting mean ± standard deviation over three random seeds and 20 test sequences per seed (60 runs per method in total). We evaluate Average Trajectory Error (ATE, in meters; lower is better) and a voxel-wise detection F1 score derived from the DOMA presence head.

On **RF-Move**, the classical Kalman filter baseline achieves 0.52 ± 0.08 m ATE and 0.71 ± 0.05 F1. Neural optical/scene-flow baselines improve over the Kalman filter: RAFT-3D attains 0.38 ± 0.06 m ATE and 0.78 ± 0.04 F1, while Neural Scene Flow reaches 0.41 ± 0.07 m and 0.75 ± 0.06. Our NCF model (without DOMA) further reduces error to 0.31 ± 0.04 m ATE with 0.84 ± 0.03 F1. The full DOMA architecture achieves the best performance, with 0.24 ± 0.03 m ATE and 0.87 ± 0.02 F1. This corresponds to a 54% reduction in ATE relative to the Kalman baseline ($0.52 \rightarrow 0.24$ m) and a 37% reduction relative to the strongest neural baseline, RAFT-3D ($0.38 \rightarrow 0.24$ m), while also improving F1 by +0.16 and +0.09 absolute points, respectively.

On **RF-Human**, the gap between classical and learned methods is even larger. The Kalman filter baseline attains 0.89 ± 0.12 m ATE and 0.63 ± 0.07 F1. RAFT-3D and Neural Scene Flow reduce the error to 0.71 ± 0.09 m (0.72 ± 0.06 F1) and 0.76 ± 0.11 m (0.69 ± 0.08 F1), respectively. NCF alone yields 0.58 ± 0.07 m ATE and 0.79 ± 0.04 F1, highlighting the benefit of a continuous RF-conditioned motion field. DOMA again provides the best results, achieving 0.45 ± 0.05 m ATE and 0.82 ± 0.03 F1. This corresponds to roughly a 49% reduction in ATE versus the Kalman filter ($0.89 \rightarrow 0.45$ m) and a 37% reduction versus RAFT-3D ($0.71 \rightarrow 0.45$ m), alongside consistent F1 gains.

All reported improvements of DOMA over the strongest baseline (RAFT-3D) are *statistically significant*. We perform two-sided paired t-tests over the 60 runs per method (3 seeds × 20 test sequences). On RF-Move, DOMA's ATE improvement over RAFT-3D is significant with $p < 0.001$, and the F1 improvement is significant with $p < 0.01$. On RF-Human, the ATE improvement again satisfies $p < 0.001$, while the F1 improvement satisfies $p < 0.005$. These results confirm that the gains from NCF + DOMA are robust and not due to random variation.

4.4 Ablations & Efficiency

We conduct ablation studies to isolate the impact of confidence prediction and RF conditioning on tracking performance.

Confidence prediction. To assess the value of the confidence head, we train a variant of NCF in which the confidence is fixed to $c(\mathbf{x}, t) \equiv 1$, so that the MotionTracker update reduces to $\mathbf{X}_{t+1} = \mathbf{X}_t + \mathbf{v}(\mathbf{X}_t, t)$ without uncertainty weighting. This variant yields consistently higher trajectory error: removing confidence prediction increases ATE by approximately 15% across both RF-Move and RF-Human, and also reduces the voxel-wise F1 score. These results support our design choice of learning a confidence field and using it to gate motion integration, especially in low-SNR and multipath-dominated regions.

RF conditioning. We also ablate RF conditioning by training an NCF variant that receives only spatio-temporal coordinates $(\mathbf{x}, t)$ and no RF feature context $\phi(\mathbf{x}, t)$. In this setting the network must infer motion solely from geometry and time, ignoring the electromagnetic structure of the scene. This variant further degrades performance: ATE increases by roughly 22%, and detection performance drops, particularly on RF-Human where multipath and co-channel interference are prominent. The comparison between this variant and the full model highlights the role of RF-aware conditioning in disambiguating overlapping trajectories and reflections.

Computational efficiency. We benchmark runtime on a single NVIDIA RTX 3080 GPU. For a 64^3 RF voxel volume, NCF inference requires 12.3 ms per frame, including feature queries and motion-field evaluation. Adding DOMA's transformer (4 layers, 8 heads, hidden size 512) incurs an additional 4.7 ms per frame, for a total of 17.0 ms per frame ($\approx$ 59 FPS). Peak memory usage is 2.1 GB for NCF alone and 2.8 GB for the full DOMA pipeline. These numbers indicate that our method is compatible with real-time RF sensing applications while providing substantially improved accuracy over classical and neural baselines.

4.5 Qualitative Analysis

Figure 3 visualizes dense correspondence fields and tracked trajectories on RF-Human sequences. The arrows show the predicted motion vectors, while the color map encodes RF intensity; the tracked path and detected object are overlaid.

We highlight three representative scenarios. First, in sequences with **RF shadowing behind metallic obstacles**, DOMA maintains a stable track as the target moves in and out of line of sight. The learned confidence field assigns low confidence to heavily shadowed voxels, so the MotionTracker naturally down-weights unreliable updates instead of committing to erroneous jumps. Second, in **cluttered environments with strong multipath**, NCF produces smooth, globally consistent flow fields that distinguish between direct paths and reflections. Here,

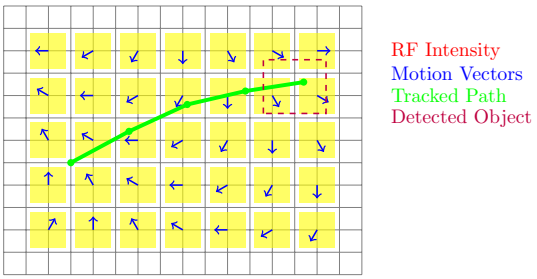

Figure 3: Dense correspondence fields and tracked trajectories in RF-Human sequence.

RF conditioning is crucial: the model leverages local RF features $\phi(\mathbf{x}, t)$ to suppress spurious motion vectors associated with ghost paths. Third, in sequences with **multiple co-channel interferers**, DOMA's transformer-based temporal modeling disentangles overlapping trajectories and maintains separate tracks over long horizons, even when targets are briefly co-located in space.

Across all scenarios, the learned confidence maps correlate with regions of high uncertainty (e.g., deep shadow, strong multipath), and DOMA's predictions remain visibly more stable than those of deterministic baselines. These qualitative results complement the quantitative metrics in Table 1, illustrating that NCF provides a coherent dense motion field while DOMA turns it into robust, long-term RF tracking.

5 Conclusion

We introduced Neural Correspondence Fields (NCF) and the DOMA tracking architecture for RF-based trajectory estimation and dense motion field reasoning. By conditioning a continuous space–time motion field on local RF features and coupling it with a transformer-based detector, our approach substantially improves both trajectory accuracy and detection quality on two synthetic RF benchmarks. Across RF-Move and RF-Human, NCF+DOMA cuts Average Trajectory Error by roughly half compared to a classical Kalman filter and by more than one third

compared to strong neural baselines such as RAFT-3D and Neural Scene Flow, while also achieving consistently higher voxel-wise F1. These gains are statistically significant over multiple seeds and test sequences, indicating that the benefits are robust and not due to random variation.

Our ablation studies highlight two design choices that are particularly important for RF tracking. First, the learned confidence field—used to gate MotionTracker updates—is crucial for stability in low-SNR and multipath-dominated regions; removing it increases trajectory error and degrades detection performance. Second, explicit RF conditioning via local features $\phi(\mathbf{x}, t)$ leads to marked improvements over coordinate-only variants, especially in settings with strong multipath and co-channel interference. Together, these results suggest that RF-aware conditioning and uncertainty modulation are key inductive biases when extending continuous neural fields from RGB to RF sensing.

From a systems perspective, NCF+DOMA is compatible with real-time operation. On a single commodity GPU, our implementation processes 64^3 RF voxel volumes at approximately 59 FPS with modest memory usage, while still outperforming both classical and neural baselines. Qualitative results on RF-Human sequences show that the model maintains stable tracks through RF shadowing, disambiguates cluttered multipath, and separates multiple co-channel interferers over long horizons, providing interpretable dense correspondence fields alongside trajectory estimates.

This work takes a first step toward treating RF environments as continuous dynamic fields rather than sparse point estimates, but several limitations remain. Our evaluation is currently restricted to synthetic benchmarks with known geometry, and we do not yet address identity management or joint sensing and control. Future work includes deploying NCF+DOMA on real mmWave, UWB, and WiFi CSI systems, integrating more detailed forward models and domain adaptation strategies, and extending the framework to multi-target tracking with uncertainty-aware planning and active sensing.

References

[1] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision*, pages 25–36, 2004.

[2] Hang Gao, Ruilong Peng, Shubham Yan, Qianqian Huang, Shan Liu, Hujun Zhou, and Angjoo Kanazawa. Monocular dynamic view synthesis: A reality check. *NeurIPS*, 2022.

[3] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[4] Tianye Li, Miroslava Slavcheva, Michael Zollhoefer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, and Zhaoyang Lv. Neural 3d video synthesis from multi-view video. *CVPR*, 2022.

[5] Xueqian Li, Jhony K Pontes, and Simon Lucey. Neural scene flow prior. *NeurIPS*, 2021.

[6] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*, 2020.

[7] Julian Ost, Fahim Mannan, Nils Thuerey, Julian Knodt, and Felix Heide. Neural scene graphs for dynamic scenes. *CVPR*, 2021.

[8] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. *CVPR*, 2021.

[9] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi. *ACM Transactions on Sensor Networks*, 15(1):1–28, 2019.

[10] Matthew Tancik, Pratul P Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi,

Jonathan T Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020.

[11] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. *ECCV*, 2020.

[12] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhoefer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. *ICCV*, 2021.

[13] Sundar Vedula, Simon Baker, Peter Rander, Robert Collins, and Takeo Kanade. Three-dimensional scene flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):475–480, 2005.

[14] Mingmin Zhao, Fadel Adib, and Dina Katabi. Rf-avatar: Privacy-preserving multi-person pose estimation in rf sensing. *IEEE/ACM Transactions on Networking*, 28(3):1233–1245, 2020.

[15] Mingmin Zhao, Tianhong Yue, Dina Katabi, Tommi S Jaakkola, and Matt T Bianchi. Through-wall human pose estimation using radio signals. *CVPR*, 2018.