

MULAN-WC: Multi-Robot Localization Uncertainty-aware Active NeRF with Wireless Coordination

Weiying Wang^{*}, Victor Cai[†], Stephanie Gi^{*},

^{*}John A. Paulson School Of Engineering And Applied Sciences
Harvard University

Email: weiyingwang, sgil@g.harvard.edu

[†]victorcai@college.harvard.edu

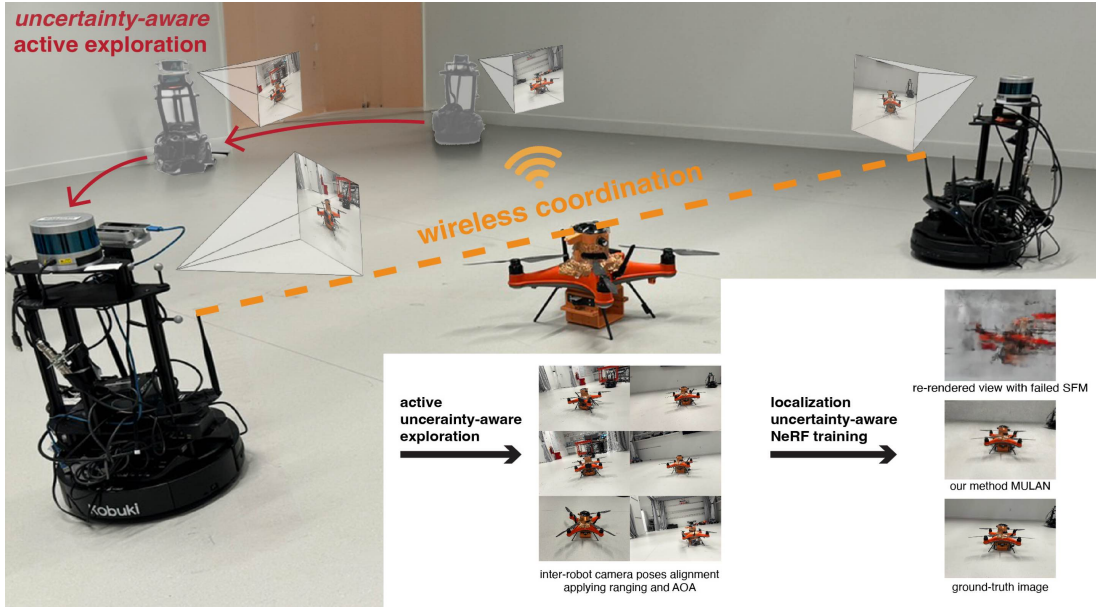


Fig. 1: Overview: We propose a collaborative, localization uncertainty-aware NeRF framework for a team of robots, employing wireless coordination and active best-next-view selection for novel view finding.

Abstract—This paper presents MULAN-WC, a novel multi-robot 3D reconstruction framework that leverages wireless signal-based coordination between robots and Neural Radiance Fields (NeRF). Our approach addresses key challenges in multi-robot 3D reconstruction, including inter-robot pose estimation, localization uncertainty quantification, and active best-next-view selection. We introduce a method for using wireless Angle-of-Arrival (AoA) and ranging measurements to estimate relative poses between robots, as well as quantifying and incorporating the uncertainty embedded in the wireless localization of these pose estimates into the NeRF training loss to mitigate the impact of inaccurate camera poses. Furthermore, we propose an active view selection approach that accounts for robot pose uncertainty when determining the next-best views to improve the 3D reconstruction, enabling faster convergence through intelligent view selection. Extensive experiments on both synthetic and real-world datasets demonstrate the effectiveness of our framework in theory and in practice. Leveraging wireless coordination and localization uncertainty-aware training, MULAN-WC can achieve high-quality 3d reconstruction which is close to applying the ground truth camera poses. Furthermore, the quantification of the information gain from a novel view enables consistent

rendering quality improvement with incrementally captured images by commending the robot the novel view position. Our hardware experiments showcase the practicality of deploying MULAN-WC to real robotic systems.

I. INTRODUCTION

Vision-based 3D reconstruction in previously unseen environments is pivotal in a broad spectrum of robotics applications, ranging from autonomous navigation[1], mapping and localization[2] to scene understanding[3]. The conventional process typically involves: 1) collecting multi-modal sensory information from onboard sensors such as RGB-D cameras and inertial measurement units, 2) extracting geometric features to compute relative pose information, and 3) applying pose graph optimization to produce a 3D environment representation using geometrically constrained spatial feature information[4, 5, 6]. Scaling up this capability to a fleet of robots could enable better coverage and faster exploration in large-scale environments. Nevertheless, it is nontrivial to scale

up conventional methods to a fleet of robots. This is due to challenges in effectively obtaining relative poses between robots that are needed to align inter-robot frames and form a global understanding of the scene. Furthermore, another problem in deploying a multi-robot system is the question of how to actively command the robot to acquire visual information so as to maximize the information gain in the 3D reconstruction [7, 8]. Active image acquisition is even more critical in a multi-robot setting to fully leverage the advantages of the fleet over a single robot. To address these challenges, we introduce a multi-robot collaborative framework utilizing Neural Radiance Fields (NeRF) for reconstruction, and using on-board wireless signal-based coordination to provide relative positional information between robots.

Firstly, to address the need for photometrically and geometrically accurate 3D reconstruction, a large number of works in NeRF[9, 10, 11] offer a revolutionary technique in synthesizing photorealistic 3D implicit representations from sparse 2D images. This is attributed to NeRF’s unique capability to model the volumetric density and color of light in a scene, enabling highly detailed and accurate reconstructions from diverse viewpoints. A crucial input that enables real-time NeRF in [11] is the camera pose corresponding to each image in the same frame. Using optical flow based feature tracking described in [12], the relative camera poses can be computed in real-time and fed to the NeRF training. However, acquiring accurate relative positional information in a multi-robot system is nontrivial, especially in the absence of global localization systems like GNSS or motion capture systems. In the traditional multi-robot coordination or localization approach, multi-robot SLAM is often dependent on the alignment of individual maps and subsequent pose estimation from overlapped appearance-based feature observations [5, 13], namely loop closure. However, inter-robot loop closure brings significant complications and computational overhead [14, 5]. To satisfy the need for relative positional information, we instead use phased array-based wireless sensing between robots, building upon our previous work in [15]. Here, the off-the-shelf WiFi chip, which is native for communication on most robotics platforms, can be used to obtain inter-robot positional information independent of appearance-based environmental features. This positional information thus can be utilized to compute the relative translation between any pair of robots and thus their image frames. Inspired by other works in incorporating depth information from the SLAM pipeline[12] where depth information is used as supervision, we also design our training loss to be aware of which regions of data are more certain than others based on the uncertainty in wireless sensing. Like other perception modalities, wireless sensing also encodes probabilistic perception due to environmental and hardware noise. This work develops a method to quantify the wireless localization uncertainty based on Angle of Arrival (AoA) profile reconstruction and correlation with the received AoA profile. This ensures that the training loss is informed about data regions with higher localization uncertainty, which brings in more accurate visual information. Integrating this quantified

uncertainty information into the NeRF training process allows us to bias the training loss, enhancing the accuracy and reliability of the 3D NeRF-generated reconstructions.

Moreover, a multi-robot system offers advantages beyond merely improving the efficiency of scene coverage from different viewpoints[16]; we can also naturally enable active image acquisition by determining the most informative next view for the NeRF model and controlling robots to acquire these additional images. Most works in applying NeRF to 3D reconstruction only passively process the images given by the perception pipeline. In resource-constrained large-scale deployment, it is beneficial to actively plan robots to acquire the most informative next image. There are some works [17, 18] that acquire images that can maximally cover the scene of interest, leading to higher information gain or better quality of reconstruction. However, to the best of our knowledge, this is the first time that active information acquisition has been applied to a coordinated multi-robot system for NeRF-based 3D reconstruction. The work in [17] proposes a promising approach to evaluate the potential information gained from a novel view by quantifying the reduction of the variance for the rendering. However, this quantification doesn’t consider the localization uncertainty of the camera, which is particularly common in multi-robot or multiple-camera setups, making it inefficient while dealing with localization uncertainty across robots that use wireless coordination. We address this specifically, by considering the inter-robot camera pose uncertainty in the characterization of the color posterior in 3D space. Our work integrates localization uncertainty quantification into the evaluation of novel-view information gain by deriving the reduction of the variance. Subsequently, we can direct the robots to actively capture images from a set of feasible next positions, for the team of robots to achieve the highest information gain in the NeRF model.

In summary, this work makes three main contributions to multi-robot 3D reconstruction integrated with NeRF:

- **Framework for integrating SAR-based wireless coordination for Multi-Robot NeRF:** We present a framework that leverages multi-robot collaboration and SAR-based wireless coordination to enable multi-robot localization uncertainty-aware NeRF.
- **Collaborative Active Image Acquisition:** Our system introduces a framework for active image acquisition, utilizing uncertainty quantification and novel-view location sampling to direct robots for optimal data collection, maximizing information gain for NeRF.
- **Extensive Experiments on hardware robots:** We conducted experiments on our customized hardware robot demonstrating that our framework does not only effectively achieve the same quality of rendering faster and higher quality, but also can actively command the robot to unvisited place by reducing the variance of rendering.

II. PROBLEM FORMULATION

In this section, we briefly review some background knowledge of the NeRF and introduce the wireless coordination from

our previous work [15] as a basis for our approach.

A. NeRF formulation

NeRF implicitly represents a scene using a fully connected neural network. In the ideal propagation ray-tracing model, the scene is modeled as a continuous function that maps any viewing angle in 5D input coordinates to a color $c(r, g, b)$ and a volume density σ . The 5D input coordinate D consists of the position in Cartesian coordinates (x, y, z) and the viewing angle (θ, ϕ) . NeRF renders the color of the sampling ray passing through the environment with classical volume rendering. Suppose we sample a ray from a position \mathbf{o} . The points along the ray can be parameterized by

$$r(t) = \mathbf{o} + t\mathbf{d}$$

The color projection of the ray back to the projection plane is

$$\mathcal{C}(r) = \int_{t_n}^{t_f} T(t) \sigma(r(t)) c(D) dt,$$

where $T(t) = \exp(-\int_{t_n}^t \sigma(r(s)) ds)$ is the accumulated transmittance along the sampling ray, and t_n and t_f are the artificial sampling box. In a realistic setup, the computation of the full integral of the color through the ray can be intractable. Instead, [19] discretized the integral as the linear combination of multiple sample points along the ray. NeRF optimizes the approximated discrete function by minimizing the squared reconstruction error between the ground truth color of each pixel captured in training RGB images and the reconstructed rendering pixel colors. The loss function is then defined as

$$\mathcal{L} = \sum_i \|\mathcal{C}(r_i) - \bar{\mathcal{C}}(r_i)\|_2^2 \quad (1)$$

B. Collaborative NeRF

To achieve 3D reconstruction with more than one robot, one of the fundamental requirements is having a common frame of reference or known camera extrinsic or relative transformation between cameras even if the data is collected from different robots from different views. Instead of being given a set of poses \mathcal{T} in the same frame of reference, we instead focus on the problem of having the sets of poses from all robots in the team in $\mathcal{T}_\alpha, \mathcal{T}_\beta$ and so on. Without loss of generality, we only focus on the observation from two robots α and β . In order to align a pose T_k^α of robot α at local time k and another pose T_p^β of robot β at local time p , we need to obtain the inter-robot camera extrinsic $T_{k\alpha}^{p\beta} = (t_{k\alpha}^{p\beta}, \theta_{k\alpha}^{p\beta})$.

C. Wireless Coordination

In our previous work [14], we extract the Angle of Arrival (AoA) information between any two robots by measuring the phase difference in the Wi-Fi channel. Suppose we have two robots α and β in communicating range at time t and their poses in local frame T_α and T_β . We can measure relative position between two robots using ranging from the ultrawideband (UWB) as well as Angle-of-Arrival (AoA) from our SAR-based framework output [15] with a probability density

function of ranging and AoA annotated by $f_{uwb}(d|T^\alpha, T^\beta)$ and $f_{aoa}(\phi|T^\alpha, T^\beta)$ respectively, are defined as:

$$f_{uwb}(d|T^\alpha, T^\beta) = c_1 \exp\left(\sigma_{k,p}^{-2}(d - \|t_{k\alpha}^{p\beta}\|_2)^2\right) \quad (2)$$

$$f_{aoa}(\phi|T^\alpha, T^\beta) = c_2 \exp\left(\kappa_{k,p}^2 \cos(\phi_{k\alpha}^{p\beta} - \theta_{k\alpha}^{p\beta})\right) \quad (3)$$

where $c_1 = \frac{1}{\sqrt{2\pi\sigma_{k,p}^2}}$, $c_2 = \frac{1}{2\pi I_0(\kappa_{k,p})}$. Here, $\sigma_{k,p}^2$ and $t_{k\alpha}^{p\beta}$ the variance and mean of the distance measurement; and $\kappa_{k,p}^2$ and $\theta_{k\alpha}^{p\beta}$ are the concentration parameter computed as the inverse of the AoA variance and the mean of the AoA distribution.

III. APPROACH

In this section, we present our multi-robot NeRF framework that addresses the challenges of inter-robot pose localization, uncertainty quantification, and active best-next-view finding. Our approach leverages wireless signals, specifically Angle of Arrival (AoA) and ranging measurements, to estimate the relative poses between robots. We develop a novel method to quantify the uncertainty of AoA estimates by reconstructing the AoA profile and correlating it with the received AoA profile. This uncertainty quantification is then integrated into the NeRF training process to mitigate the impact of inaccurate poses on the reconstruction quality. Furthermore, we propose an active view-finding approach that accounts for the position uncertainty of the robots when selecting the most informative views for NeRF training. By incorporating localization uncertainty into the novel view selection process, our framework can more accurately determine the best next views for each robot, even in the presence of pose uncertainty arising from wireless coordination.

A. Inter-robot Pose Localization and Uncertainty Quantification

As described in Section II-B, accurate and fast NeRF reconstruction relies heavily on the known transformation between cameras on different robots. In a multi-robot setup, we propose using wireless signals to obtain accurate inter-robot poses by leveraging Angle of Arrival (AoA) and ranging measurements. Suppose we have a pose T_k^α in SE(3) of robot α at local time k and another pose T_p^β of robot β at local time p . We then can obtain a wireless measurement between the two robots using onboard WiFi and UWB which can be annotated by a tuple $(t_{k\alpha}^{p\beta}, \theta_{k\alpha}^{p\beta})$. If we aim to use robot α 's frame as the global frame, then the extrinsic or the rigid transformation of robot β 's camera pose can be represented by $T_{kp}^{\alpha\beta} \oplus T_p^\beta$, where the annotation \oplus denotes rigid transformation. However, the accuracy of the resulting pose estimates is subject to the uncertainties in the AoA and ranging measurements as described in Eq 2 and Eq 3. To mitigate the impact of inaccurate poses on the NeRF training process, we propose applying a weight to each training example based on the uncertainty of the associated robot pose's AoA and ranging measurements from the other robots.

Quantifying the uncertainty of AoA estimates is particularly challenging, since there is a lack of standard error

quantification methods applicable from previous works. To address this, we develop a novel approach to quantify the uncertainty of AoA estimates by reconstructing the AoA profile and correlating it with the received AoA profile. We propose the AoA uncertainty quantification methodology with the following outline. The receiving robot α 's wireless channel and pose $\langle \bar{h}_{\alpha\beta}(t), \bar{T}_\alpha(t) \rangle$ are collected over some set of signal packet timestamps $t = t_k, \dots, t_l$ in robot α 's timeframe from a transmitting robot β . We want to measure AoA of robot β at robot α 's local time t_k . Following [20], the poses of robot α trace out a virtual multi-antenna array that enables AoA measurement. First, we apply multiple signal classification (MUSIC) to the measured channels over the timestamps to obtain a measured AoA profile $\bar{F}(\phi, \theta)$ and AoA measurement $(\bar{\phi}, \bar{\theta})$. Then, we reconstruct the wireless channel phases in $h'_{\alpha\beta}(t)$ at each timestamp taking the transmitter to be at the measured AoA, apply MUSIC to obtain the reconstructed AoA profile $F'(\phi, \theta)$, and calculate the AoA uncertainty by comparing two profiles to measure the similarity of the profiles near $(\bar{\phi}, \bar{\theta})$.

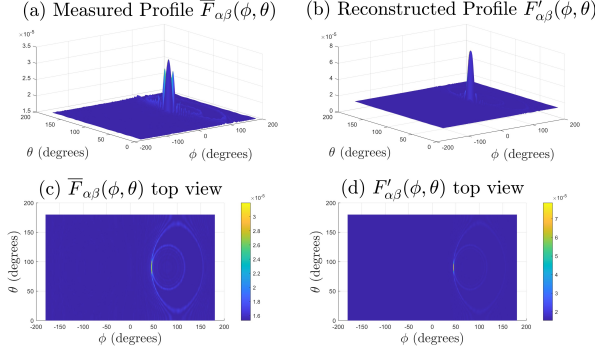


Fig. 2: Simulation of the AoA variance methodology. (a) The Measured AoA profile $\bar{F}_{\alpha\beta}(\phi, \theta)$ from a simulated measured wireless channel $\bar{h}_{\alpha\beta}$ with 0.7 radians standard deviation injected phase noise. (b) The reconstructed AoA profile $F'_{\alpha\beta}(\phi, \theta)$ from the reconstructed channel $h'_{\alpha\beta}$, with 0.5 radians standard deviation phase noise for tolerance. Note both tallest peaks align at $(\bar{\phi} = 45.6^\circ, \bar{\theta} = 90^\circ)$. (c) and (d) are top views of (a) and (b), respectively.

Specifically, the ideal channel of robot α from robot β is

$$h_{\alpha\beta}(t) = \frac{1}{d_{\alpha\beta}} \exp\left(\frac{-2\pi\sqrt{-1}}{\lambda} d_{\alpha\beta}(t)\right) \quad (4)$$

and robot α collects its local pose information, particularly the displacement distance, azimuth, and zenith of robot i from the center of its frame, over $t = t_k, \dots, t_l$. Suppose robot i receives a measured channel $\bar{h}_{\alpha\beta}(t)$. Then the MUSIC AoA profile is constructed as in [20] to be $\bar{F}_{\alpha\beta}(\phi, \theta)$, and the measured AoA is taken to be $(\bar{\phi}, \bar{\theta}) = \arg \max_{(\phi, \theta)} \{\bar{F}_{\alpha\beta}(\phi, \theta)\}$ at the tallest peak in the profile.

From here, we reconstruct the channel over $t = t_k, \dots, t_l$

using $(\bar{\phi}, \bar{\theta})$ and the pose information $\bar{T}_\alpha(t_k), \dots, \bar{T}_\alpha(t_l)$:

$$h'_{\alpha\beta}(t) = \exp\left(\frac{-2\pi\sqrt{-1}}{\lambda} f_\alpha(\bar{T}_\alpha(t), \bar{\phi}, \bar{\theta}) + \sqrt{-1}\nu(t)\right) \quad (5)$$

where f_α is the displacement of robot i projected along the measured AoA direction all in the local frame, relative to the first observation at t_k , and $\nu(t)$ is a zero-mean real random variable that injects Gaussian phase noise into each element of the channel to add a small realistic amount of error tolerance in the measured profile. The same MUSIC algorithm is then run to obtain $F'_{\alpha\beta}(\phi, \theta)$ and its tallest peak at (ϕ', θ') . Since the reconstructed channel is based on an AoA from $(\bar{\phi}, \bar{\theta})$, making the noise $\nu(t)$ small ensures that the reconstruction has $(\phi', \theta') = (\bar{\phi}, \bar{\theta})$, where the ϕ and θ sample spaces are discretized during MUSIC computation.

The AoA uncertainty is then calculated as follows. Both profiles are cropped to a rectangle R with a small radius $\Delta\phi = 5^\circ$ and $\Delta\theta = 5^\circ$ around $(\bar{\phi}, \bar{\theta})$ for comparison near the measured peak, to ignore distant multipath components. Then we define the AoA uncertainty $\kappa_{k,p}$ via

$$\frac{1}{\kappa_{k,p}} = \sum_{(\phi, \theta) \in R} \bar{F}_{\alpha\beta}(\phi, \theta) F'_{\alpha\beta}(\phi, \theta) \quad (6)$$

This measures how concentrated the received profile is around the $(\bar{\phi}, \bar{\theta})$, such that a lower AoA uncertainty $\kappa_{k,p}$ corresponds to a more reliable AoA measurement. This AoA variance can be used in Equation 3 to understand the variability in the pose estimates, as described in the next section.

B. Uncertainty-aware NeRF training

As described in our problem formulation, if we have two sets of camera poses T_α and T_β in their own local frames, we want to find the relative transformation between them to align the poses in the same frame of reference. For each wireless measurement between measuring poses T_k^α and T_p^β , we can quantify the uncertainty of the AoA and ranging estimates using the methods described in the previous section.

Let $\sigma_{k,p}$ and $\kappa_{k,p}$ denote the variance of the ranging measurement and the AoA, respectively, for the wireless measurement between poses T_k^α and T_p^β . We propose incorporating these uncertainty measures into the NeRF training process by modifying the standard pixel loss function \mathcal{L} given in Eq 1 by re-scaling the loss for each training sample. For brevity of the annotation in this section, we omit robot indices α, β , and all local time frames p and k from now on. We apply uncertainty propagation to compute the error ellipse of uncertain measurements. Since the AoA measurements and ranging measurements are taken using different sensing modalities, they can be treated as independent measurements with the error ellipse's axes aligning with x-y axes. The semi-major and semi-minor axes a and b of the error ellipse can be derived as:

$$a^2 = \sigma^2 \cos^2(\theta) + t^2 \sin^2(\theta) \kappa^2 \quad (7)$$

$$b^2 = \sigma^2 \sin^2(\theta) + t^2 \cos^2(\theta) \kappa^2 \quad (8)$$

where t is the ranging measurement, and θ is the AoA estimate. Then the scale factor with confidence interval CI is given by

$$k = \sqrt{-2\log(1 - CI)} \quad (9)$$

Hence, the uncertainty of this wireless localization γ can be represented by the area of the error ellipse $\gamma = k^2\pi ab$. Then the new loss function can be re-scaled with γ that is normalized by sigmoid function

$$\mathcal{L}_{uncertainty} = \text{SIGMOID}(\gamma)\mathcal{L} \quad (10)$$

By incorporating the uncertainty-aware scaling factor into the NeRF loss function, our multi-robot NeRF system can effectively learn to reconstruct the 3D scene while accounting for the varying reliability of the pose estimates obtained through wireless coordination. This approach results in a more robust and accurate 3D reconstruction, especially in scenarios where the pose estimates may be subject to significant uncertainties.

C. Active Best-View finding with position uncertainty

In a multi-robot NeRF system, actively selecting the best views for each robot to capture can significantly improve the efficiency and quality of the 3D reconstruction. However, the uncertainty in robot poses obtained through wireless coordination can impact the effectiveness of the view selection process. When a robot attempts to find the best next view location by proposing and evaluating potential new positions, the uncertainty in its current pose can lead to inaccurate assessments of the information gain at these novel view locations.

To address this challenge, we propose an active view finding approach that incorporates the position uncertainty of the robots to guide the selection of the most informative views for NeRF training. Building upon the approach proposed in [1] for evaluating the potential information gain from novel views by quantifying the reduction of variance in rendering, we extend this method to account for the uncertainty in the robot's current position and its propagation to the novel view locations being evaluated. By considering the localization uncertainty during the novel view selection process, we can more accurately determine the most informative next views for each robot, even in the presence of pose uncertainty arising from wireless coordination.

We adopt the assumption that the radiance color of any location along the ray $r(t)$ can be parameterized by a Gaussian distribution with mean $\bar{c}(r(t))$ and variance $\bar{\beta}(r(t))$. To incorporate this uncertainty, we model the origin \mathbf{o} of each ray r following a Gaussian distribution with 0 mean and the variance σ , representing the localization uncertainty.

$$\mathbf{o} \sim \mathcal{N}(0, \sigma) \quad (11)$$

Assuming a NeRF model \mathcal{M} has been trained on an initial collection of data D , the prior distribution $P(c(r(t_k))|D)$ of the color c at location $r(t)$ follows a Gaussian distribution $\mathcal{N}(\bar{c}(r(t)), \bar{\beta}^2(r(t)))$. The accumulated color from a new

ray r passing through can also be modeled as a Gaussian distribution:

$$p(C(r)|c(r(t)), \mathbf{o}) \quad (12)$$

where $C(r)$ is the color of the rendered pixel accumulated from the ray r , and $c(r(t))$ is the color of the location in 3D space. Then if we marginalize over \mathbf{o} ,

$$\begin{aligned} p(C(r)|c(r(t))) &= \int p(C(r)|c(r(t_k)) * p(\mathbf{o}) d\mathbf{o} \\ p(C(r)|c(r(t_k))) &\sim \mathcal{N}\left(\sum_{i=1}^N \alpha_i \bar{c}(r(t)), \sum_{i=1}^N \alpha_i * \sigma^2 + \bar{\beta}^2(r(t_k))\right) \end{aligned} \quad (13)$$

Then apply Bayes' rule to get the posterior:

$$\begin{aligned} P(C(r)|D, r(t_k), \mathbf{o}) &\propto P(C(r)|c(r(t_k))) * P(c(r(t_k))|D) * P(\mathbf{o}) \\ &\propto \exp\left(-\frac{1}{2} * (c(r(t_k)) - \left(\frac{\omega C(r)}{\alpha_k} + (1 - \omega) * \bar{c}(r(t_k))\right))^2\right) * \\ &\quad \left(\frac{\alpha_k^2}{\alpha_k^2 \sigma^2 + \bar{\beta}^2(r)} + \frac{1}{\bar{\beta}^2(r(t_k))}\right)^{-1} \\ \text{where } \omega &= \frac{\alpha_k^2 \bar{\beta}^2(r(t_k))}{\alpha_k^2 \bar{\beta}(r(t_k))^2 + \alpha_k^2 \sigma^2 + \bar{\beta}^2(r)} \end{aligned} \quad (15)$$

Then the variance of the posterior distribution can be extracted as

$$\left(\frac{\alpha_k^2}{\alpha_k^2 \sigma^2 + \bar{\beta}^2(r)} + \frac{1}{\bar{\beta}^2(r(t_k))}\right)^{-1} \quad (16)$$

As the localization variance increases, the uncertainty of the radiance field also increases accordingly. To select the best view, the metric will prefer the novel view with lower localization uncertainty and lower variance. Since we only need to consider the variance reduction given multiple rays from a sampled novel-view position, we can then command the robot to move to the location with highest variance reduction using Equation 16.

IV. RESULTS

In this section, we present a comprehensive evaluation of our methodology through both synthetic datasets collected in synthetic environments, and real-world datasets collected on our hardware robots. Our findings validate the effectiveness of our algorithm in integrating perspectives from multiple robots within an active acquisition framework, showcasing significant improvements in data capture and processing. We implemented our algorithm based on a Pytorch implementation of the SDF and NeRF part described in Instant-NGP [11] with CUDA-accelerated ray marching. We modified the loss function to implement the localization uncertainty-aware loss described in Equation 1. Furthermore, for the active image collection, we re-implemented the rendering variance reduction described in [17] incorporated with the localization uncertainty as described in Eq 16. We use a desktop with NVIDIA RTX 6000 for all evaluations.

A. Wireless variance

First, we present the performance benchmark with our proposed wireless variance metric as formulated in Section III-A. With a simulated trajectory, the metric is tested over 16 random trials with various amounts of injected Gaussian channel noise as defined in Equation 5, ranging from 0.01 to 3 radians phase noise standard deviation. As shown in Figure 3, the AoA error from the ground truth scales up quickly and becomes unstable as our proposed AoA variance metric increases. To further demonstrate the relationship between the variance of the AoA Error and our proposed profile variance, Figure 4 clearly shows that the variance of the AoA error will exponentially increase as our metric increases. This result indicates that our AoA uncertainty quantification is well aligned as an indicator of the variance of AoA measurements, which can be further used to quantify the uncertainty of the camera position.

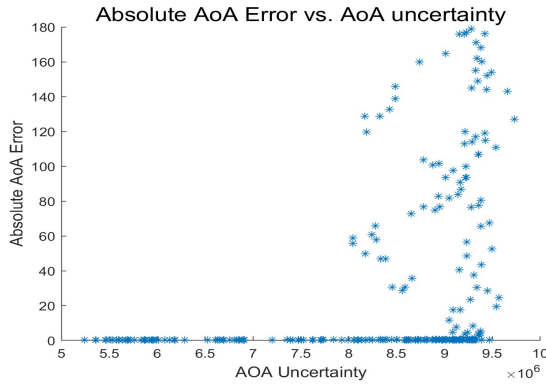


Fig. 3: The Absolute AoA error from ground truth plotted against our AoA uncertainty metric. We see that nonzero AoA error grows as our AoA uncertainty metric grows. This indicates that our metric successfully captures true error in measured AoA.

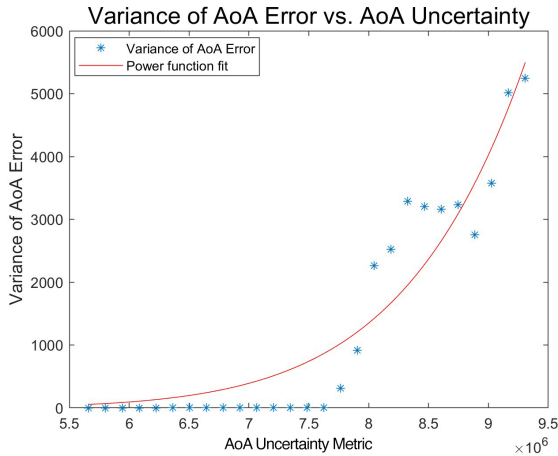


Fig. 4: The variance of the AoA error here as a function of AoA uncertainty is calculated empirically by finding the variance of the AoA error on the y-axis within a sliding window of $\Delta\kappa_{k,p} = 8.4 \times 10^5$ along the x-axis. It is fit with a power curve of the form $y = ax^b$, with $r^2 = 0.8942$.

B. Simulation Experiment

Our algorithm localization uncertainty-aware framework described in Sec III-B is first assessed using a synthetic dataset lego released with the original NeRF work [19]. The dataset is commonly used for evaluating NeRF frameworks. The dataset is partitioned into two subsets to simulate the data acquired from two robots. This approach allows us to mimic the real-world scenario of capturing images from different angles and positions, thereby testing the robustness and adaptability of our algorithm in synthesizing and analyzing data from varied viewpoints. There are total 3 setups are evaluated:

- A [Oracle]: Camera poses from the dataset in the global frame (known as extrinsic between cameras) and images from the dataset
- B [Normalized camera poses with AOA and ranging simulated]]: Camera poses from the dataset but normalized by the first pose in each partition. Then the AOA and ranging are simulated with noise whose standard deviation are 0.05 meters and 5 degrees respectively.
- C [Normalized camera poses with AOA and ranging simulated with variance as supervision]: The same setup in B but the training loss is incorporated with the noise variance.

A	PSNR	30.47
	LPIPS	0.062
B	PSNR	26.48
	LPIPS	0.092
C	PSNR	28.69
	LPIPS	0.071

TABLE I: Performance comparison between different setups where larger PSNR values are better, and smaller LPIPS values indicate better quality. The comparison demonstrate that applying the uncertainty-aware loss can effectively improve the quality of the model

C. Hardware Experiment

For the real-world application, we deployed our algorithm on two of customized Locobot PX100 robots. These robots were equipped with Oak-D Pro Cameras, operating at 1080p 20Hz, along with DWM1001 UWB modules, 5dBi Antennas, and Intel NUC 10 computers for onboard processing. The experimental setup involves a drone object centrally relative to the two robots, which are programmed to navigate curved paths around the object to complete data capture.

Both robots utilize onboard Visual Inertial Odometry (VIO) to estimate local camera displacement within their respective frames. At the onset of the experiment, Angle of Arrival (AoA) and ranging measurements are taken to establish an initial estimate of the relative positioning between the robots. Subsequently, the covariance of the VIO data was monitored to identify optimal intervals for refreshing wireless data collection. In the meantime, the testbed is equipped with the Optitrack motion capture system providing the ground truth camera poses for each robot.

The experiments are conducted using five setups:

- A **[Oracle]**: Camera poses captured by motion system in the global frame (known extrinsic between cameras) and images from the onboard camera.
- B **[Best case for our system]**: Camera poses from motion capture system with wireless coordination and images from the onboard camera. The poses are normalized in each robot’s local frame.
- C **[Our system “in-the-wild” (no mocap)]**: Camera poses from onboard VIO, wireless coordination, and images from the onboard camera.
- D **[Our system “in-the-wild” with variance as supervision]**: Camera poses from onboard VIO, wireless perception for coordination, and uncertainty-aware training loss scaled by localization variance.
- E **[Benchmark comparison]**: Camera poses from onboard VIO; COLMAP[21] is used for computing inter-robot relative camera pose extraction.

All five setups are evaluated using standard metrics for NeRF: Peak Signal-to-Noise Ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS)[22]. Each metric is evaluated from samples in the test set ground truth images, along with camera poses. For each setup, there is a total of 100 images with camera poses that are collected continuously from each robot while robots are moving around the drone subject. drone-1 and drone-2 are different images in the testing dataset.

		drone-1	drone-2
A	PSNR	26.4	24.5
	LPIPS	0.351	0.384
B	PSNR	25.45	23.4
	LPIPS	0.382	0.378
C	PSNR	23.32	22.3
	LPIPS	0.41	0.405
D	PSNR	25.04	23.03
	LPIPS	0.389	0.395
E	PSNR	11.5	12.5
	LPIPS	0.79	0.85

TABLE II: Performance comparison between different setups where larger PSNR values are better, and smaller LPIPS values indicate better quality. The comparison between setup A and setup B demonstrates that applying wireless coordination can effectively achieve close performance to having a global coordinate system. Results from setup C show the realistic performance of our system using fully onboard VIO for local positioning, which is degraded but still relatively robust. Setup D shows the scaling with localization uncertainty quantification can improve the quality almost to the best-case scenario in setup B.

As shown in Table II which provides our quantitative results, setup A shows the best performance we can achieve in a two-robot team since it is based on the ground truth camera poses provided by the motion capture system. In setup B, the poses are normalized by the starting pose of each robot captured in the motion capture system. Then wireless coordination is incorporated to provide inter-robot camera extrinsic. Setup C shows the realistic setup, which applies the wireless localization between robots to the local VIO poses

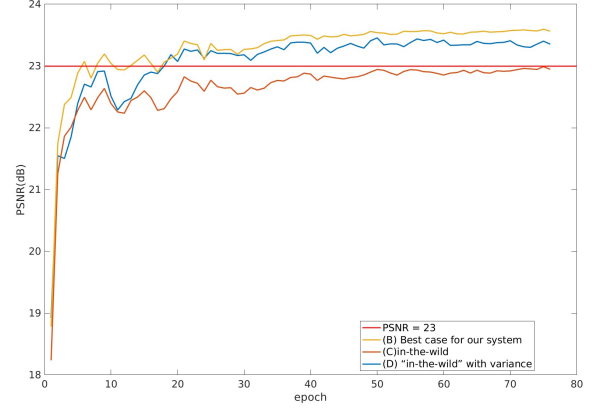


Fig. 5: PSNR improvement over iterations with different setups

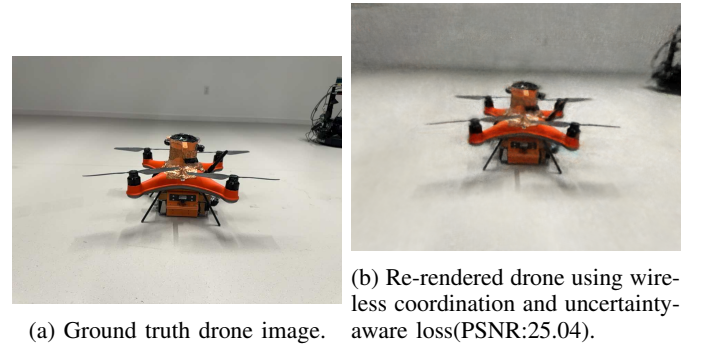


Fig. 6: An example of the drone we reconstructed in the testbed. The left figure is the ground truth image, right figure is the re-rendered image from a trained model.

and is effectively close to the result in setup B. Setup D shows that our framework can achieve better results than C by using the variance-aware loss function defined in Equation 10 with the corresponding localization variance proposed in Equation 16. The benchmark comparison setup E fails to produce a cohesive 3D rendering, due to the discrepancy of the relative camera pose estimation using COLMAP [21]. This is mainly due to the drastic translation change in camera view from different robots.

In many robotics applications, there are possibilities that we need to have a quickly-converging view of the environment before the model training fully converges. In our experiment, we also validate that our methods not only deliver better rendering but also achieve faster PSNR improvement as shown in Fig 5.

D. Active Image Capturing

Following the initial phase of data gathering, a waiting period is observed until the Neural Radiance Field (NeRF) model’s loss stabilized. The robots then execute a series of maneuvers, sampling eight different directions at 0.5-meter intervals. Our evaluation focused on minimizing variance of the rendering posterior, employing Equation 16 to identify

positions yielding the most significant reduction in variance. The application of our algorithm in a hardware setting demonstrates its practical feasibility. Moreover, it underscores the potential of our method to optimize the data capture process through strategic robot positioning and movement.

After selecting the location with the highest variance reduction using our proposed method, the robot is commanded to the new location and observes the environment again. We then let the model train until the loss stabilizes and repeat the process four times to evaluate the efficacy of our method. For comparison, we also randomly selected accessible locations around the robots and controlled the robots to move to those locations. The evaluation-maneuver-training loop was conducted on both our policy and the random policy and the results are reported in Table III. These results demonstrate that our approach provides a principle metric that can improve the quality of the rendering consistently.

observation#		1st	2nd	3rd	4th
Our algorithm	PSNR	19.66	19.80	20.04	20.08
	LPIPS	0.407	0.398	0.394	0.381
Random Exploration	PSNR	19.53	19.63	19.60	19.63
	LPIPS	0.422	0.419	0.421	0.418

TABLE III: Performance comparison between different setups, demonstrating that our method improves the rendering quality metric with consecutive views.

V. CONCLUSION

This work presents MULAN-WC, a multi-robot 3D reconstruction method that uses wireless signal-based coordination between robots. This work presents i) a framework for multi-robot NeRF that uses SAR-based wireless relative position measurements to stitch together views of the environment from multiple robots, ii) uncertainty-based weighting of samples in the NeRF training as a supervision technique, where samples with greater wireless measurement noise are weighted less, leading to better accuracy of the combined rendering, and iii) collaborative active next-image acquisition, where novel-view location sampling incorporates wireless pose uncertainty, and is used to direct robots to better sampling locations that reduce variance during NeRF training. We demonstrate the performance of the multi-robot framework in hardware, where our results show good quality of rendering according to the standard NeRF error metrics of PSNR and LPIPS, and an consistent improvement when we additionally use the uncertainty of the AoA measurement as supervision in the NeRF training. Lastly, we show that AoA measurements can be used to select next-best-view based on regions of better position accuracy and that this results in incremental rendering quality improvement.

ACKNOWLEDGMENTS

The authors gratefully acknowledge partial funding through NSF grant #CNS-2114733 and the Sloan award #FG-2020-13998.

REFERENCES

- [1] M. Meilland, A. I. Comport, and P. Rives, “Dense omnidirectional rgb-d mapping of large-scale outdoor environments for real-time localization and autonomous navigation,” *Journal of Field Robotics*, vol. 32, no. 4, pp. 474–503, 2015.
- [2] Y.-J. Yeh and H.-Y. Lin, “3d reconstruction and visual slam of indoor scenes for augmented reality application,” in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, 2018, pp. 94–99.
- [3] Y. Nie, J. Hou, X. Han, and M. Nießner, “Rfd-net: Point scene understanding by semantic instance reconstruction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4608–4618.
- [4] M. Cao, W. Jia, Y. Li, Z. Lv, L. Li, L. Zheng, and X. Liu, “Fast and robust local feature extraction for 3d reconstruction,” *Computers & Electrical Engineering*, vol. 71, 2018.
- [5] Y. Chang, Y. Tian, J. P. How, and L. Carlone, “Kimera-multi: a system for distributed multi-robot metric-semantic simultaneous localization and mapping,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 210–11 218.
- [6] J. Yin, L. Carlone, S. Rosa, and B. Bona, “Graph-based robust localization and mapping for autonomous mobile robotic navigation,” in *2014 IEEE International Conference on Mechatronics and Automation*. IEEE, 2014, pp. 1680–1685.
- [7] Y. Gao, L. Su, and H. Liang, “Mc-nerf: Multi-camera neural radiance fields for multi-camera image acquisition systems,” 8 2023.
- [8] W. Wang, N. Jadhav, P. Vohs, N. Hughes, M. Mazumder, and S. Gil, “Active rendezvous for multi-robot pose graph optimization using sensing over wi-fi,” in *Robotics Research*, T. Asfour, E. Yoshida, J. Park, H. Christensen, and O. Khatib, Eds. Cham: Springer International Publishing, 2022, p. 832.
- [9] J. Kerr, L. Fu, H. Huang, Y. Avigal, M. Tancik, J. Ichnowski, A. Kanazawa, and K. Goldberg, “Evo-nerf: Evolving nerf for sequential robot grasping of transparent objects,” in *6th annual conference on robot learning*, 2022.
- [10] M. Adamkiewicz, T. Chen, A. Caccavale, R. Gardner, P. Culbertson, J. Bohg, and M. Schwager, “Vision-only robot navigation in a neural radiance world,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4606–4613, 2022.
- [11] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [12] A. Rosinol, J. J. Leonard, and L. Carlone, “Nerf-slam: Real-time dense monocular slam with neural radiance fields,” 2022.
- [13] H. Do, S. Hong, and J. Kim, “Robust loop closure method for multi-robot map fusion by integration of consistency and data similarity,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5701–5708, 2020.
- [14] W. Wang, A. Kemmeren, D. Son, J. Alonso-Mora, and S. Gil, “Wi-closure: Reliable and efficient search of inter-robot loop closures using wireless sensing,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
- [15] N. Jadhav, W. Wang, D. Zhang, S. Kumar, and S. Gil, “Toolbox release: A wifi-based relative bearing framework for robotics,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 13 714–13 721.
- [16] W. Chen, X. Wang, S. Gao, G. Shang, C. Zhou, Z. Li, C. Xu, and K. Hu, “Overview of multi-robot collaborative slam from the perspective of data fusion,” *Machines*, vol. 11, no. 6, 2023.
- [17] X. Pan, Z. Lai, S. Song, and G. Huang, “Activenerf: Learning where to see with uncertainty estimation,” in *European Conference on Computer Vision*, 2022.
- [18] K. Lee, S. Gupta, S. Kim, B. Makwana, C. Chen, and C. Feng, “So-nerf: Active view planning for nerf using surrogate objec-

tives,” *arXiv preprint arXiv:2312.03266*, 2023.

- [19] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” 2020.
- [20] N. Jadhav, W. Wang, D. Zhang, O. Khatib, S. Kumar, and S. Gil, “A wireless signal-based sensing framework for robotics,” *The International Journal of Robotics Research*, vol. 41, no. 11-12, pp. 955–992, 2022.
- [21] J. L. Schönberger and J.-M. Frahm, “Structure-from-motion revisited,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jun 2018, pp. 586–595. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00068>