Accurate Indoor Localization With Zero Start-up Cost

Swarun Kumar   Stephanie Gil   Dina Katabi   Daniela Rus
Massachusetts Institute of Technology
{swarun, sgil, dk, rus}@mit.edu

ABSTRACT

Recent years have seen the advent of new RF-localization systems that demonstrate tens of centimeters of accuracy. However, such systems require either deployment of new infrastructure, or extensive fingerprinting of the environment through training or crowdsourcing, impeding their wide-scale adoption.

We present Ubicarse, an accurate indoor localization system for commodity mobile devices, with no specialized infrastructure or fingerprinting. Ubicarse enables handheld devices to emulate large antenna arrays using a new formulation of Synthetic Aperture Radar (SAR). Past work on SAR requires measuring mechanically controlled device movement with millimeter precision, far beyond what commercial accelerometers can provide. In contrast, Ubicarse's core contribution is the ability to perform SAR on handheld devices twisted by their users along unknown paths. Ubicarse is not limited to localizing RF devices; it combines RF localization with stereo-vision algorithms to localize common objects with no RF source attached to them. We implement Ubicarse on a HP SplitX2 tablet and empirically demonstrate a median error of 39 cm in 3-D device localization and 17 cm in object geotagging in complex indoor settings.

Categories and Subject Descriptors C.2.2 [**Computer Systems Organization**]: Computer-Communications Networks

Keywords Wireless; Localization; Wi-Fi; SAR; PHY

1. INTRODUCTION

Recent years have witnessed a major interest in developing accurate RF-based localization systems that enable users to navigate indoor spaces much like what GPS provides for outdoor environments [26, 19, 37]. Many advances have been made, leading to localization solutions that deliver an accuracy of tens of centimeters [37, 15, 33]. Unfortunately, none of these solutions have actually reached today's users. This is because past work that provides accurate RF-localization falls in one of two categories: 1) It either requires the deployment of new infrastructure (specialized access points [15], antenna arrays [37], acoustic beacons [19], etc.), or, 2) it needs exhaustive fingerprinting of the environment to learn the spatial distribution of signal strength, either in a training phase [28, 18] or via crowdsourcing [26, 5] to achieve meter accuracy at best. Ideally, we would like to achieve tens of centimeters accuracy for RF localization of mobile devices, without any fingerprinting or specialized infrastructure.
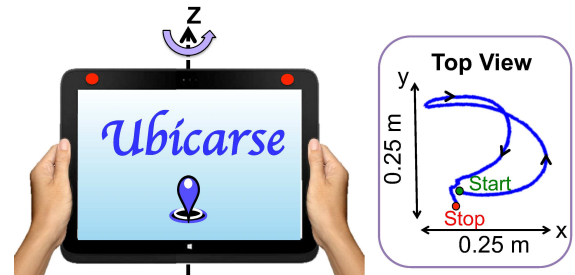
Figure 1: Twisting the Mobile Device. The user twists the device about its vertical (z) axis, as shown in the figure on the left. The red circles indicate antenna locations. On the right, we depict the top view of a candidate device trajectory as measured by a Vicon motion capture system.

In principle, if mobile devices can be outfitted with a large antenna array [37, 15], they could accurately identify the spatial direction of incoming RF signals. Thus, they could localize themselves relative to the locations of neighboring Wi-Fi access points without requiring fingerprinting or modified infrastructure. In fact, public databases of access point locations are already available for a large number of buildings [34]. Of course, it is infeasible to mount a large antenna array on a small low-power handheld device. But what if today's mobile devices could somehow emulate the localization capability of a large antenna array?

We present Ubicarse[1], an indoor geo-location system that enables mobile devices to indeed emulate a large antenna array. At first, it might seem that one can use Synthetic Aperture Radar (SAR) to mimic an antenna array on a handheld device. Unfortunately, SAR in its existing form is unsuitable for handheld devices. To understand why this is the case, note that a moving antenna performs SAR by collecting signal snapshots as it moves along its trajectory, and jointly processing these snapshots to emulate a large antenna array traced out by that trajectory. Therefore for SAR to mimic a particular antenna array geometry, it must first know the position of the moving antenna at every point along its trajectory. Satisfying this requirement is relatively easy in radar systems where the speed and trajectory of the antenna movement are finely controlled. Unfortunately, when a user moves her handheld device, neither its speed nor its trajectory can be accurately controlled. To get a feel for the accuracy required, in 802.11a/n, even a small error of 2 cm in the relative position of the moving antenna leads to an error of 60 degrees in identifying the direction of the source.

One might consider using motion sensors (accelerometer, gyroscope, compass) present in most mobile devices to estimate antenna positions as it moves. Unfortunately, since SAR requires sub-centimeter accuracy in the device position along its trajectory, commercial motion sensors are virtually unusable to measure such fine-grained translation [22] and can at best be used to measure orientation from their gyroscopes.[2] Requiring the user to rotate the device

---

[1]The Spanish word Ubicarse derives from the Latin root *ubi* and signifies to locate or situate oneself with precision.

[2]This is because accelerometers report the net acceleration of the device including gravity, which must be subtracted out and the result integrated twice to obtain translation.

on a perfect arc with zero translation of the device center would preclude measuring this translation but is impractical to enforce.

Ubicarse addresses this challenge by developing a new formulation of SAR that is translation-resilient. Specifically, we allow the user to twist her mobile device about its vertical axis, as shown in Fig. 1. Even if the twisting involves unknown trajectories (that include translation), we can accurately compute SAR knowing only the rotation estimate from the gyroscope. To do so, we exploit the MIMO capability of modern wireless cards. Suppose the mobile device has two antennas. The distance between these antennas is fixed independent of how the user moves her device. Thus, whenever the user translates the device, the relative position vector of both its antennas remains the same. In contrast, as the device rotates, the relative position vector of the antennas, also rotates. Leveraging these observations, we develop a new SAR formulation that operates on *relative* wireless channels to estimate the direction of the access point's signal, purely based on the device's orientation with no information required on its exact position or translation. We analytically demonstrate that Ubicarse's translation-resilience holds under the same assumptions made for standard SAR and antenna arrays, both in single path and in the presence of multipath effects. We also implement the system and demonstrate a localization accuracy of the mobile device within few tens of centimeters, with no specialized infrastructure or fingerprinting.

Ubicarse is not limited to localizing RF devices; it can also localize common objects that have no RF source attached to them. Specifically, we leverage the camera available on most mobile devices to enable a user to learn the location of an object of interest by simply snapping photographs of the object from multiple perspectives. At first, it might seem that this application can be satisfied simply by tagging the object in the photograph with the position of the mobile device, obtained from RF-based localization. Such an approach however would yield poor tagging accuracy, particularly when the user cannot get too close to the object, for e.g., tagging a broken overhead lamp to report it to the facilities department.

To achieve high-precision geotagging, Ubicarse exploits the synergy between its WiFi-based localization capability and stereo vision algorithms used for 3-D reconstruction. Specifically, today's vision algorithms can accurately localize an object with respect to the set of camera coordinates from which the object was imaged [29], but it cannot find the coordinates of the object with respect to a global reference. In contrast, Ubicarse's SAR localization can identify the global positions of the device's camera accurately, but has no information about the object. Therefore, by combining the two methods, we can localize an object to within tens of centimeters in global coordinates – a promising step towards a future where we can tag or search for objects in the physical world using only our mobile devices, much the way we currently search for information on the web.

We implemented Ubicarse on a HP SplitX2 Tablet equipped with Intel 5300 wireless cards and Yei Technology motion sensors. We build on the 802.11 CSI tool [13] to obtain the wireless channels on the tablet. We conducted our experiment in a large university library, with books arranged in multiple shelves and racks that emulate the complex multipath characteristics of large warehouses and departmental stores, where indoor localization has been of particular interest. Our experiments reveal the following:

- **Translation Resilient SAR:** Ubicarse's SAR is resilient to unknown translations. We move the device along complex trajectories including several random walks and curves. We use Ubicarse to localize the direction of access point placed at distances from 2 to 12 m without knowledge of the trajectory of the movement.

Our results show a median error in angle of $3.4°$ across trajectories, demonstrating the translation-resilience of Ubicarse's SAR.
- **Device Localization:** We ask a population of users to twist the tablet to localize themselves relative to the access points in the environment. The results show that Ubicarse localizes the tablet with a median error of 39 cm in full three dimensional space. Thus, Ubicarse achieves its goal of accurate localization without requiring either new infrastructure or signal fingerprinting.
- **Object Localization:** We integrate Ubicarse with the VisualSFM computer vision toolkit [35] and evaluate its object geotagging capability. We use Ubicarse to localize books in the library relative to the access points, by taking a few pictures of them. Our results show a median error of 17 cm in full 3-D space, which, surprisingly, is superior to Ubicarse's device localization. Indeed we show that a joint optimization of Ubicarse and vision has an interesting side-effect: we can significantly refine Ubicarse's device (i.e. camera) location estimates simply by studying the relative camera locations output by vision algorithms, corresponding to pictures snapped from different perspectives. In doing so, we show that Ubicarse's localization accuracy improves to a median of 15 cm from ground truth.

Contributions Ubicarse is an indoor localization system that achieves tens of centimeters of accuracy on commodity mobile devices, with no specialized infrastructure or fingerprinting. Ubicarse achieves this through a novel formulation of Synthetic Aperture Radar that is translation resilient, making it practical to deploy on handheld devices. Further, Ubicarse uses the cameras on mobile devices to help users geo-tag objects around them to within tens of centimeters. We implement Ubicarse on commodity tablets and experimentally validate its accuracy in complex indoor settings.

2. PRIMER ON SYNTHETIC APERTURE RADAR

Synthetic Aperture Radar has been widely used in aircrafts and satellites to map the topography of the Earth's surface [9]. Recent systems [33, 32] have also leveraged SAR for RFID localization.

SAR's primary goal is to allow a single-antenna wireless receiver to isolate and analyze the multiple signal paths emanating from a wireless transmitter. Specifically, it computes a *multipath profile*, which measures the relative signal power received from the transmitting source along different spatial directions (see Fig. 2). To do this, SAR leverages a simple principle: a moving receiver antenna snapshots the wireless signals at different spatial locations. Combined, these channel snapshots mimic a multi-antenna array (Fig. 2). Thus, one can simply apply standard antenna array equations [24] on the signals received at each antenna position to compute the multipath profile. For instance, one can apply equations for a linear antenna array [37] if the receive antenna moves on a line, or a circular antenna array [7], if the receiver rotates about a circle.

To understand SAR more formally, lets examine a special case of SAR for a received single path.[3] We focus on emulation of a circular antenna array, a scenario that is particularly relevant for this paper. Let's suppose the receive antenna rotates about the origin along a circle, taking n snapshots of the wireless channel from a source along a direction α_T as shown in Fig. 2. From basic channel models, we can write the the wireless channel h_i measured by the receiver at the ith snapshot (where $i = 1, \ldots, n$) as the complex number [30]:

$$h_i = \frac{1}{d} e^{\frac{-j2\pi}{\lambda}(d + r\cos(\alpha_T - \phi_i))} \tag{1}$$

Where (r, ϕ_i) is the antenna's polar coordinates at snapshot i (Fig. 2), d is its distance to the source, and λ the signal wavelength.

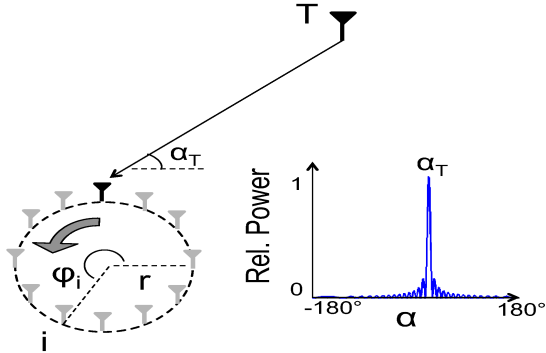[3]We explicitly deal with the more general multipath case in §4.2

Figure 2: Circular SAR. Depicts a single antenna moving in a circle of radius r to emulate a circular antenna array. Gray antennas depict points where channel snapshots were taken. At snapshot i, the antenna is at an azimuthal angle ϕ_i. The multipath profile for a line-of-sight transmitter T has a sharp peak at the direction of the source α_T.

At this point, SAR computes the relative signal power along each spatial direction to generate a multipath profile. To do this, it needs two inputs: The measured channel snapshots h_i along the antenna trajectory, and the positions (r, ϕ_i) of the antenna along this trajectory. It then finds the relative power $P(\alpha)$ along direction α as:

$$P(\alpha) = \left| \frac{1}{n} \sum_{i=1}^{n} h_i e^{\frac{\pm j 2\pi}{\lambda} r \cos(\alpha - \phi_i)} \right|^2 \qquad (2)$$

Notice that the above multipath power profile $P(\alpha)$ is maximum precisely if $\alpha = \alpha_T$, i.e. along the true direction of the source. This is because the terms $h_i e^{\frac{\pm j 2\pi}{\lambda} r \cos(\alpha - \phi_i)}$ are identical in phase, and therefore add up constructively when $\alpha = \alpha_T$ (see Eqn. 1), and tend to add up destructively as α deviates from α_T. Therefore, the multipath profile can be used to accurately ascertain the physical direction of a transmitter, a property crucial for indoor localization.

SAR makes an important assumption: It assumes accurate knowledge of the position of the device at different points in time relative to some fixed origin in space (e.g. its initial location). Clearly, from Eqn. (2), in order to compute $P(\alpha)$ one must know the polar coordinates (r, ϕ_i) of the antenna positions. To get a feel for the accuracy in position required, in 802.11a/n, even a small error of 2 cm in the relative position of the moving antenna leads to an error of about 3 m in localization, when the transmitting source is 6 m away. The assumption of accurately known antenna trajectories is valid in radar systems where the speed and trajectory of the movement are finely controlled. In the following sections, we describe how this assumption poses an important challenge to our system, where users cannot be expected to move their devices along accurately known trajectories.

3. OVERVIEW OF UBICARSE

Ubicarse enables accurate indoor localization of commercially available mobile devices with no prior infrastructure or fingerprinting. At the heart of Ubicarse's device localization is the ability to perform SAR using today's off-the-shelf Wi-Fi cards and motion sensors (i.e. chips containing accelerometers, magnetometers and gyroscopes). In its traditional form, SAR assumes that channel models have a clear dependence on both the position and orientation of the antennas (see Eqn. (1) and Fig. 2). Therefore computation of an accurate multipath profile requires precise knowledge of how the antennas are located in space. However, users of Ubicarse cannot be expected to move their devices along precise trajectories. Consequently, Ubicarse must rely on the device's motion sensors

to infer its position and orientation over time. Unfortunately, applying SAR using commodity motion sensors is challenging for the following reasons: (1) They are virtually unusable to measure device translation at fine granularity [22], as necessitated by SAR; (2) While the orientation they measure is accurate for short time intervals, the accuracy reduces over time due to integration drift [11].

In §4 and §5, we describe how Ubicarse resolves the above challenges to accurately perform SAR even if a user twists her device along unknown trajectories to obtain multipath profiles. In §6, we explain how Ubicarse uses these multipath profiles to accurately localize the mobile device. In §7, we show how Ubicarse integrates with vision algorithms to accurately geotag objects of interest in the environment, enabling a new class of mobile applications.

4. TRANSLATION-RESILIENT SAR

In this section, we describe how Ubicarse performs SAR even if the user moves her device along complex unknown trajectories. Specifically, we address the challenge that commodity motion sensors only provide accurate orientation information and do not provide accurate translation data that is crucial to perform SAR (see §2). Therefore, Ubicarse must perform SAR in a manner resilient to any device translation the user causes as she twists her device, yet correctly obtains the multipath signal profile.

To achieve this goal, Ubicarse leverages the MIMO capability of modern wireless cards. In particular, we consider mobile devices with just two MIMO antennas that measure the wireless channels from any single antenna of a wireless access point. As the user twists her device, the two antennas sample the wireless channel at each point in the device trajectory. The channel snapshots emulate a virtual array of antennas, where the number of virtual antennas depend on the number of channel samples. The main challenge however, is to perform SAR using the channel samples gathered from the two moving physical MIMO antennas, using only the orientation of the device and without its translation information.

Ubicarse's key idea to achieve this is as follows: Instead of performing SAR by plugging in the wireless channels measured at the antennas, Ubicarse feeds into SAR a quantity we call the *relative wireless channel* between the two physical antennas. Unlike channels on individual antennas that depend on the absolute antenna positions, the relative channel only depends on the relative position vector between the two antennas. Recall that the distance between the two antennas on a device is fixed, regardless of how the user moves her device. Thus, whenever the user translates her device, the relative position vector of both its antennas remains the same. In contrast, as the device rotates, the relative position vector of the antennas, also rotates. Hence, by plugging in relative channels into the SAR formulation, Ubicarse obtains multipath profiles with no information on device center translation.

4.1 Description of Translation Resilient SAR

We explain how Ubicarse satisfies two properties: 1) It is translation-resilient; 2) It correctly estimates SAR profiles.

Translation-Independence: We begin by presenting our solution for line-of-sight scenarios (i.e. free-space scenarios), where the signal from the transmitter arrives along one dominant path. For simplicity, we assume that both the transmitter and our device lie on a two-dimensional plane (extension to 3D is described in §4.2). We consider a transmitting source T placed at the origin and a two-antenna receiver device, as shown in Fig. 3. As the user twists the receiving device, it takes n different snapshots at the wireless signals, each at different orientations and positions. The two antennas of the device are separated by a distance of r (typically, 6 to
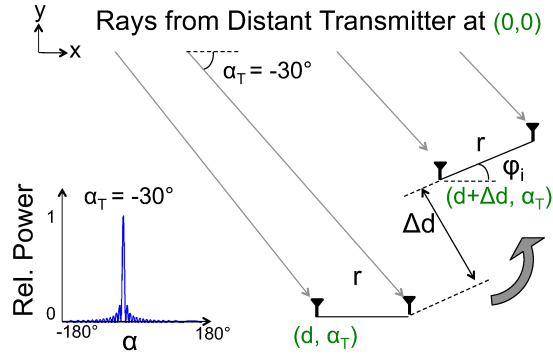
Figure 3: Ubicarse's SAR. Shows polar coordinates (in green) of the device at an offset Δx_i and Δy_i along the x and y-axis and orientation ϕ_i. Note that $\Delta d = \Delta y_i / \sin \alpha_T$. The origin of the system $(0,0)$ is at the transmitter.

20 cm).[4] Denote the initial polar coordinates of the first receive antenna as (d, α_T) relative to the transmitter where d is the distance between the transmitter and receiver. Note here that the main goal of SAR is to estimate α_T which captures the direction of the transmitting source relative to the device.

As the user moves the device, its antenna coordinates at any snapshot i (where $i = 1, \ldots, n$) depend on two quantities: 1) The translation of the device along the x and y axes at the snapshot relative to its initial location denoted by Δx_i and Δy_i. 2) The orientation of the device ϕ_i, relative to the x-axis is shown in Fig. 3. Hence, our goal is to perform SAR purely based on orientation ϕ_i, with no information on translation along the *x* and *y* axes: $\Delta x_i, \Delta y_i$.

Though our goal is to perform SAR without knowledge of the translation (Δx_i and Δy_i), clearly if the device is translated by a large distance with respect to the source, the angle of arrival itself (the quantity that we are interested in estimating) will change. Therefore in our system we assume that during SAR computation, the source is at a far distance compared to the movement of the tablet (a standard assumption for SAR and antenna arrays in general [37, 9]). The form of translation independence we seek in the current SAR formulation is robustness to local translations of the device center as the person twists the tablet as shown in Fig. 1. To achieve this goal, we first need to characterize the wireless channel of the source's signal measured at the antennas in each snapshot taken by the receiving device. We can apply simple geometry to write the polar coordinates of the two antennas as: $(d + \frac{\Delta y_i}{\sin \alpha_T}, \alpha_T)$ and $(d + \frac{\Delta y_i}{\sin \alpha_T} + r\cos(\alpha_T - \phi_i), \alpha_T)$ (as shown in Fig. 3) where the polar angle α_T is the same for the two antennas under our far distance assumption. From basic wireless channel models [30], we can write the wireless channels $h_{1,i}$ and $h_{2,i}$ measured at the receive antenna 1 and 2 during its ith snapshot as:

$$h_{1,i} = \frac{1}{d + \frac{\Delta y_i}{\sin \alpha_T}} e^{\frac{-j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_T})} \approx \frac{1}{d} e^{\frac{-j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_T})} \quad (3)$$

$$h_{2,i} = \frac{1}{d + \frac{\Delta y_i}{\sin \alpha_T} + r\cos(\alpha_T - \phi_i)} e^{\frac{-j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_T} + r\cos(\alpha_T - \phi_i))} \quad (4)$$

$$\approx \frac{1}{d} e^{\frac{-j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_T} + r\cos(\alpha_T - \phi_i))} \quad (5)$$

Where λ denotes the wavelength of the wireless signal. Note that the approximations hold assuming $\Delta x_i, \Delta y_i \ll d$, i.e. the source is far relative to the movement of the device, as in standard SAR [9].

---

[4]Ubicarse assumes that the distance between the two antennas on the device is known. This is typically available from the manufacturer's specification for a given device.

We now define the relative wireless channel at the ith snapshot $\hat{h}_i = h_{2,i}h_{1,i}^*$, where $(.)^*$ denotes the complex conjugate. Intuitively, the relative wireless channel captures the relative phase between channels of the two antennas. Mathematically, the relative channel is given by:

$$\hat{h}_i = h_{2,i}h_{1,i}^* = \frac{1}{d^2} e^{\frac{-j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_T} + r\cos(\alpha_T - \phi_i))} e^{\frac{+j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_T})} \quad (6)$$

$$= \frac{1}{d^2} e^{\frac{-j2\pi}{\lambda} r\cos(\alpha_T - \phi_i)} \quad (7)$$

Interestingly, note that the relative channel is independent of device translation (Δx_i and Δy_i) and is dependent only on the orientation ϕ_i. In fact, we observe that the relative wireless channel $\hat{h}_i$ is identical in form to the wireless channel of a single antenna rotating about a circle of radius r as discussed in Eqn. 1 of §2 (barring a constant factor $\frac{1}{d} e^{\frac{j2\pi d}{\lambda}}$). Intuitively, this stems from the fact that relative to the frame of reference of the first antenna, the second antenna simply appears to move around a circle.

As a result, we can simply "plug-in" the relative channels into standard circular SAR equations (Eqn. 2 from §2) to derive the multi-path profile of the transmitter's signal as follows:

$$P(\alpha) = \left| \frac{1}{n} \sum_{i=1}^{n} \hat{h}_i e^{\frac{+j2\pi}{\lambda} r\cos(\alpha - \phi_i)} \right|^2 \quad (8)$$

Hence, even if the user twists the device in arbitrary ways, causing both rotation and translation to the device, its two antennas never translate relative to each other, and can therefore emulate a circular antenna array.

We make the following important observations on computing the multipath profile as in Eqn. 8 above:

- The complexity of our algorithm is $O(n * \log(m))$ where n is the number of channel snapshots, and m is the desired resolution in angle of arrival. The complexity is logarithmic in m since one can first perform SAR at a low resolution, and repeat SAR iteratively at higher resolutions only at angles of highest power.
- Note that the wireless channels used in the above equations must capture both magnitude and the phase of the received signal. Our implementation therefore uses the channel state information (CSI) measured by a wireless card that are complex numbers, unlike received signal strength (RSS).
- The above solution can be readily extended to OFDM systems (like 802.11n), by averaging the power-profile computed for each OFDM sub-carrier.
- Our system only requires wireless channels measured from a single antenna at the 802.11 access point and two antennas at the client. If channels from more than one antenna on a Wi-Fi access point are available, the power profiles can be measured from each individual antenna, and the results averaged.

Correctness of Multipath Profile: One might wonder if the multipath profiles generated by the above technique are indeed correct. Specifically, do the power profiles indicate a sharp peak at precisely the direction of the source α_T to help localize the source? Indeed, it is easy to observe from Eqn. 7 and 8 that $P(\alpha)$ achieves a maximum value of $1/d^4$, when $\alpha = \alpha_T$, since all components $\hat{h}_i e^{\frac{+j2\pi}{\lambda} r\cos(\alpha - \phi_i)}$ add up constructively in phase, just as in a circular antenna array. As α deviates from α_T, these components gradually begin to add up destructively. Of course, the precise shape of $P(\alpha)$ over α depends on how the user physically re-orients the device, i.e. the distribution of orientations ϕ_i. In fact, one can show that if we expect the user to simply twist the device randomly, so that ϕ_i's
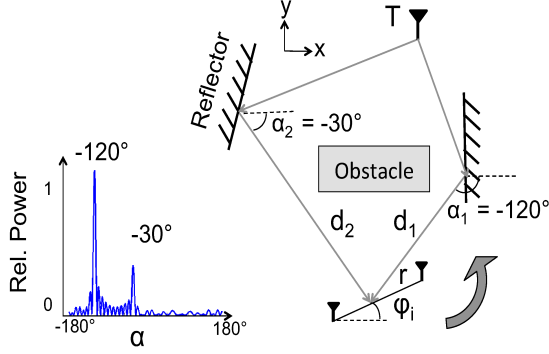
Figure 4: Multipath Scenarios. Depicts the path lengths d_k's and direction of the sources α_k's for a simple multipath scenarios. The corresponding power profile reveals peaks at the desired α_k values.

at least span a semi-circle (i.e. an angle of π)[5], the following lemma holds (Proof in Appendix A):

LEMMA 4.1. *Suppose the orientation $\phi_i \sim \mathcal{U}(\gamma - \pi/2, \gamma + \pi/2)$, at snapshot $i = 1, \ldots, n$ is uniformly distributed for some constant γ, the expected value of the multipath power profile $E[P(\alpha)]$ in line-of-sight, at each $\alpha \in [0, 2\pi]$ over the distribution of $\{\phi_1, \ldots, \phi_n\}$ is given by:*

$$E[P(\alpha)] = \frac{1}{d^4}\left[J_0(\kappa)^2 + \frac{1 - J_0(\kappa)^2}{n} \right] \qquad (9)$$

$$\textit{Where, } \kappa = \frac{4\pi r \sin\left(\frac{\alpha_T - \alpha}{2}\right)}{\lambda} \qquad (10)$$

And $J_0(\kappa)$ is the Bessel-function of the first kind[4]. □

Notice that as the number of snapshots n increases, the first term dominates and the expected power profile resembles a squared Bessel function $J_0(\kappa)^2$. Visually, this function resembles a squared sinc-function with a sharp peak when $\kappa = 0$, i.e. $\alpha = \alpha_T$. Fig. 3 plots the expected profile for a line-of-sight source at $\alpha_T = -30°$. The profile indeed resembles the squared Bessel function with a clear peak at $\alpha = -30°$.

4.2 Extending to Multipath Scenarios

While our discussion so far has considered a source at line-of-sight (i.e. free-space scenarios), in this section, we focus on multipath scenarios (i.e. non-free space scenarios). Specifically, we show that Ubicarse continues to be resilient to translation in these scenarios, while generating correct multipath profiles, indicating the direction of the various paths of the signal.

Assume that the device receives signals from m distinct paths, of length $d_1 \ldots, d_m$, arriving along directions $\alpha_1, \ldots, \alpha_m$, as in Fig. 4. Hence, SAR's goal is to infer the set of signal directions $\alpha_1, \ldots, \alpha_m$ based on the wireless channels and device orientation.

Suppose the device captures n snapshots of the wireless channels as the user twists it. As before, we denote by $(\Delta x_i \text{ and} \Delta y_i)$, the translation of the device along the x and y axes and by ϕ_i, the orientation of the device relative to its initial position and orientation during the ith snapshot. We then express the wireless channels on

---

---

the two receive antennas as:

$$h_{1,i} \approx \sum_{k=1}^{m} \frac{s_k}{d_k} e^{\frac{-j2\pi}{\lambda}(d_k + \frac{\Delta y_i}{\sin \alpha_k})} \qquad (11)$$

$$h_{2,i} \approx \sum_{k=1}^{m} \frac{s_k}{d_k} e^{\frac{-j2\pi}{\lambda}(d_k + \frac{\Delta y_i}{\sin \alpha_k} + r\cos(\alpha_k - \phi_i))} \qquad (12)$$

Where s_k is a constant complex number that refers to the attenuation and phase shift along signal path k ($k = 1, \ldots, m$) and λ is the wavelength of the signal, as before.

We can then compute the relative wireless channels $\hat{h}_i = h_{2,i} h_{1,i}^*$ that we plan to feed into SAR as:

$$\hat{h}_i = \sum_{k=1}^{m} \frac{s_k}{d} e^{\frac{-j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_k} + r\cos(\alpha_k - \phi_i))} \sum_{k=1}^{m} \frac{s_k}{d} e^{\frac{+j2\pi}{\lambda}(d + \frac{\Delta y_i}{\sin \alpha_k})}$$

$$= \sum_{k=1}^{m} \frac{s_k}{d_k} e^{\frac{-j2\pi}{\lambda} r\cos(\alpha_k - \phi_i)} \left[\frac{s_k}{d_k} + \sum_{l \neq k} \frac{s_l}{d_l} e^{\frac{+j2\pi}{\lambda}(\frac{\Delta y_i}{\sin \alpha_l} - \frac{\Delta y_i}{\sin \alpha_k})} \right]$$

$$(13)$$

Notice that the first term of the relative channel in Eqn. 13 above is nearly identical to Eqn. 7 in the line-of-sight scenario, and is independent of any translation $\Delta x_i, \Delta y_i$ between channel snapshots. Unfortunately, the second term indeed depends on translation. However, two observations work in our favor: First, if the device's translation remains constant between snapshots, the second term reduces to a constant multiplier, which merely scales the profile. Second, even if the device translation varies between snapshots, any variance (i.e. noise) caused by the second term drops significantly when summing over a large number of snapshots.

Surprisingly, these observations can help us show that by performing SAR over several snapshots, Ubicarse is resilient to device translation in multipath scenarios. Further, this property holds regardless of how the device is actually translated by the user as she twists the device. Of course, we stress that this translation cannot be unbounded, and must be relatively small compared to the distance of the device from the transmitting source, a standard assumption made by SAR and antenna array systems [37, 9]. More formally, for a device that is twisted randomly by the user over an angle of π or more, the following lemma holds (Proof in Appendix B):

LEMMA 4.2. *Suppose the orientation $\phi_i \sim \mathcal{U}(\gamma - \pi/2, \gamma + \pi/2)$, at snapshot $i = 1, \ldots, n$ is uniformly at random for some constant γ, the expected value of the multipath power profile $E[P(\alpha)]$ in multipath scenarios, at each $\alpha \in [0, 2\pi]$ over the distribution of $\{\phi_1, \ldots, \phi_n\}$ is given by:*

$$E[P(\alpha)] = \left(\sum_{k=1}^{m} \left[\frac{s_k^2}{d_k^2} + E_k \right] J_0(\kappa_k) \right)^2 + \frac{1}{n} Var[\hat{h}_i e^{\frac{j2\pi r \cos(\alpha - \phi_i)}{\lambda}}]$$

$$\textit{Where, } \kappa = \frac{4\pi r \sin\left(\frac{\alpha_T - \alpha}{2}\right)}{\lambda}, E_k = E\left[\sum_{l \neq k} \frac{s_k s_l}{d_k d_l} e^{\frac{j2\pi \Delta y_i (\frac{1}{\sin \alpha_l} - \frac{1}{\sin \alpha_k})}{\lambda}} \right]$$

And $J_0(\kappa)$ is the Bessel-function of the first kind[4]. □

Over a large number of snapshots n, the expected multipath profile resembles the squared sum of Bessel functions (which resemble sinc functions) with peaks corresponding to the different multipath directions scaled by their relative signal power. In other words, by inspecting the peaks of the profile $P(\alpha)$ one can infer the different multipath directions $\alpha_1, \ldots, \alpha_m$. Fig. 4 plots the profile for a non-line-of-sight source with two paths along $-30°$ and $-120°$. Notice that we indeed observe distinct peaks at each of the two angles, scaled by their relative signal strength.
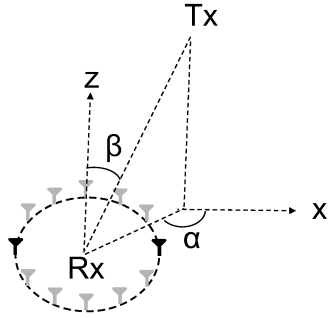
Figure 5: **Generalizing to 3-D.** Depicts the definition of α and β, respectively the azimuthal and polar angle from which the signal arrives.

4.3 Generalizing to Three Dimensions

Our above solution can be readily generalized to three-dimensions. Such a generalization is particularly important, since it is difficult to restrict users of Ubicarse to perfectly rotate their devices in a two dimensional plane. To better understand our solution, consider a two-antenna receiver device, as before and a signal path arriving from azimuthal angle α and polar angle β, as shown in Fig. 5. Suppose the user twists the device about a given axis.[6] Let ϕ_i be the current azimuthal angle of the second receive antenna relative to the first and θ_i denote the polar angle, during the i snapshot. From basic geometry, we can derive the difference between the path lengths to the two antennas is $r\cos(\alpha - \phi_i)\sin(\beta - \theta_i)$, where r is the separation between the two antennas. As a result, we can derive the power-profile along the direction (α, β) by slightly modifying Eqn. 8 as:

$$P(\alpha, \beta) = \left| \frac{1}{n} \sum_{i=1}^{n} \hat{h}_i e^{\frac{\pm j2\pi}{\lambda} r \cos(\alpha - \phi_i) \sin(\beta - \theta_i)} \right|^2 \quad (14)$$

Notice that the SAR power profile equation indeed resembles the expression for circular SAR in 3-D [7], as before. Further, note that much like standard circular antenna arrays [7], even if the user rotates the device entirely on a 2-D plane (e.g., the xy-plane), the above expression returns the angle-of-arrival in a 3-D space.

4.4 Robustness to Frequency Offsets

Recall that a wireless receiver (e.g., the user's tablet) experiences a carrier frequency offset (CFO) and sampling frequency offset (SFO) with respect to a wireless transmitter (e.g., an access point) [25]. The CFO and SFO cause an additional phase rotation, independent of the antenna movement, which unless eliminated can accumulate over time, causing errors in the SAR output. Fortunately, since the two antennas on the mobile device experience the same offsets with respect to the source, taking the relative channel eliminates that effect. Consequently, Ubicarse's SAR is robust to frequency offsets as well.

5. ACTIVE DRIFT COMPENSATION

In this section, we address the following challenge: The error in orientation reported by motion sensors accumulates gradually over time, leading to significant errors in SAR. This is because gyroscopes measure angular velocity and any systematic error in this velocity integrates over time and leads to a drift in orientation [11]. However, over time windows of a few seconds, this drift can be approximated as an unknown but linear shift in orientation [23].

---

[6]In principle, the user can twist the device about any axis, provided the axis of rotation is not parallel to the line joining the two antennas, i.e. so that the relative rotation of the two antennas is non-zero.
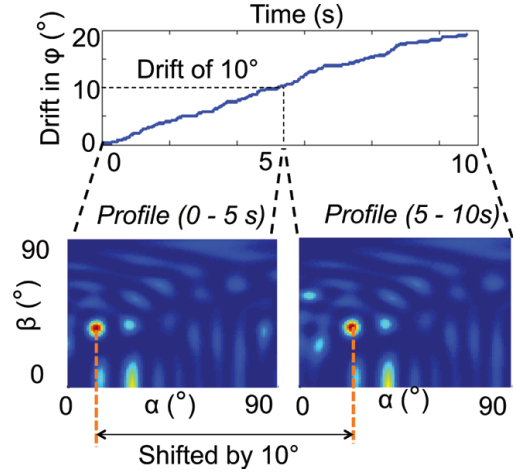

Figure 6: **Drift in Orientation.** The figure (above) plots the drift error over time for 10 seconds. The accumulated drift is $10°$ between two $5s$ intervals. The 3-D multipath profiles corresponding to these intervals (below) indicate a shift of precisely $10°$ (profiles zoom-in over a small range for clarity).

Ubicarse leverages this property to actively estimate and correct for drift in real-time, just as the user twists her device.

Ubicarse resolves orientation drift by making a key observation: A linear drift in orientation leads to a constant shift in SAR multipath profiles generated by Ubicarse. Specifically, suppose we perform SAR over two different time windows to obtain two multipath profiles. Let's say that the gyroscope incurred a drift of δ between these intervals. Recall from Eqn. 8 that any multipath profile $P(\alpha)$ depends on angles of the form $\alpha - \phi_i$, where ϕ_i is the orientation of the device during the ith snapshot. Consequently, if each ϕ_i is shifted by an offset δ between the two time windows, then α is also shifted by the same value δ between them. In other words, any accumulation of drift δ between the two intervals effectively results in a shift of precisely δ between the two multipath profiles. Hence, by estimating the shift δ between the two multipath profiles, we can accurately estimate, and therefore correct for the gyroscope drift between the two intervals.

To illustrate this in an example, Fig. 6 depicts the orientation in drift of $10°$ between two 5 second intervals during one of our experiments. We observe that the corresponding multipath profiles are also shifted by about $10°$. We can then directly estimate and actively correct for this drift in real-time by applying phase correlation [10], a standard technique from computer graphics to estimate shifts between two (noisy) images by calculating their phase difference in the 2-D Fourier domain.

Finally, note that since Ubicarse performs drift correction in real-time, it does not rely on predetermined characteristics of the gyroscope and is thus general to different hardware configurations.

6. ACCURATE DEVICE LOCALIZATION

In this section, describe how Ubicarse leverages SAR multipath profiles to accurately localize the device. Ubicarse first obtains these profiles by asking the user to twist the device about its vertical axis. It then issues beacon frames to multiple neighboring access points, to elicit response frames and estimate channels from these access points. It then simultaneously performs SAR to these access points and obtains multipath profiles. In the absence of multipath, each profile contains a single peak towards the direction of the corresponding access point. Ubicarse can then apply standard triangulation [16] to localize itself relative the access points. In principle,

Ubicarse needs to perform SAR to a minimum of three surrounding access points to localize the device.

Localization under Multipath: In practice, SAR profiles may contain multiple peaks owing to multipath. As a result, there exists an inherent ambiguity in the direction of these access points relative to the device. Ubicarse addresses this challenge using the following two key observations.

First, as noted in [37], the multipath peaks in an antenna-array profile can be differentiated from the direct path peak using the following observation: The direct peak in a profile persists even as the receiving device's position is slightly perturbed, whereas multipath peaks change significantly or disappear altogether (Fig. 9 in [37]). In fact, as Ubicarse explicitly requires a user to twist the device to perform SAR, such perturbations are inherent to our system and this is a fact that we exploit.[7] Therefore, by performing SAR on two different segments of the device's trajectory (emulating two antenna arrays at slightly different locations), one can compare the resulting multipath profiles to eliminate peaks that do not persist between these profiles.

Second, often devices are surrounded by more than three access points. This leads to an over-constrained optimization problem for triangulation, which can be formulated as a least-square fitting problem [16]. Notice that the peaks corresponding to the direct path to each of many different access points agree on a unique location for the device, and therefore lead to a good fit. In contrast, multipath peaks of the access points indicate conflicting locations for the device leading to poor fits. Therefore, by selecting the set of peaks to access points that achieve the best fit in the optimization problem (i.e. agree the most geometrically), Ubicarse can identify the set of direct paths to these access points.

Obtaining Global Device Orientation: One of the key benefits of Ubicarse is the ability to not just find the device's global position, but also its global orientation, relative to the access points' frame of reference. Notice that global orientation is not available from the device's gyroscope, which only measures angular velocity, and therefore can only be used to find relative orientation between two time intervals.[8] The key reason Ubicarse can readily estimate device orientation is that it computes the angle-of-arrival of signals from the access points. In particular, Ubicarse knows the angle of the access point, relative to the axis of its own body . But after performing localization, Ubicarse also knows the direction of this access point relative to the global x-axis. Hence, by simply subtracting the two angles, Ubicarse can compute its orientation ψ relative to the world's x-axis.

7. APPLICATION TO OBJECT GEOTAGGING

In this section, we show how Ubicarse's accurate indoor localization of mobile devices opens the door to a new class of applications: precision geotagging of indoor objects with no RF source attached to them. This would enable users to catalog and share the location of objects of interest they observe around them, i.e., products in a warehouse or artifacts in a museum. To enable this application, we leverage the fact that in addition to Wi-Fi chips and gyroscopes, most wireless devices also have cameras on them. Today, vision toolkits can be applied to overlapping photos taken from

---

[7]Notably however, the direct path peak does experience a shift due to drift as discussed in §5, although the peak's magnitude remains the same, and its shift is consistent.

[8]While the device's compass can be used to compute global orientation, such compasses are fairly inaccurate [14] and easily misled by any surrounding magnetic fields.
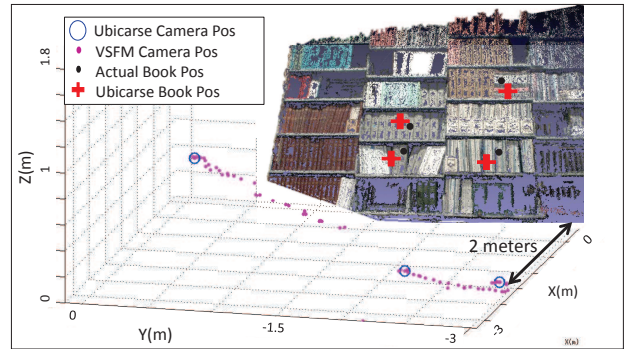

Figure 7: Point cloud of library shelving and VSFM reconstruction with camera positions and anchored camera positions. Anchored camera positions are those camera positions with both Ubicarse localization coordinates in the global frame and relative coordinates in the coordinate frame of the 3D reconstruction. These anchor points are used in Equation (15) to find a suitable transform to a the global frame.

these cameras to produce 3D pointclouds, or *relative* 3D positions, for imaged objects [29].

In particular, these vision toolkits use multiple snapshots taken of a single scene from different vantage points in order to reconstruct the 3D positions of objects in the image as a point cloud. While these tools are impressively accurate at inferring relative distances and orientations in the local coordinate frame of the camera, they cannot provide information on the positions of the camera or photographed objects in the *global* frame that a user cares about. For example, a vision-based reconstruction can accurately determine that the biology magazine stack is one foot away from the physics magazine stack, but cannot determine that these magazine stacks are located in shelf-5 of the mezzanine level room in the NY public library. In contrast, Ubicarse's core ability is that it can determine the global position of the device's camera at a given point in time. Hence we can synergize the fine-grained relative distances of object-to-camera computed by the vision toolkit, with the accurate global camera-position provided by Ubicarse. This allows for a joint optimization problem that provides the global position of the objects of interest. Surprisingly, we show that this joint optimization has an interesting side-effect: one can further refine Ubicarse's device localization by using the relative camera positions between the different snapshots, that are output by the vision toolkit. This leads to an integrated system that capitalizes upon the best of both technologies for device and object localization.

Vision Algorithms: To better understand how object localization works, we provide a briefly introduce vision algorithms for 3-D imaging. The goal of these algorithms is to read a set of 2-D images of the object snapped by the device's camera from multiple perspectives to obtain two outputs:

- *3-D reconstruction of the object*: obtained in the frame of reference of the device camera. This reconstruction is a 3-D point cloud (Fig. 7), containing the (x, y, z) coordinates of points on the object's surface in a 3-D frame of reference whose origin is at the camera's initial position and axes along its initial orientation.
- *Relative Camera Positions and Orientation*: of the vantage points where the 2-D images were taken, defined relative to the camera's initial position and orientation.

At a high level, vision algorithms obtain these quantities in three phases: First, they identify salient features [2] (e.g. corners or distinctive textures) in each of the images. Next, these features are used to identify positions of the same object between these images.

Finally, they apply advanced algorithms such as [29], which uses the distance of the same features to the different camera positions from where the images were taken, to infer relative object positions and camera positions.

Accurate Object Localization: Once a 3D pointcloud of the imaged objects is constructed via the vision algorithm, a suitable transformation must be found in order to map the object locations into the (global) reference frame of the environment. To do this, we employ Ubicarse's translation-resilient localization in N different positions and orientations where the snapshots were taken (assuming N is at least three). The N camera positions can now serve as anchor points, points for which we have both global positions $\{x_{g_1}, \ldots, x_{g_N}\}$ provided by Ubicarse, and relative positions, $\{x_{p_1}, \ldots, x_{p_N}\}$ in the coordinate frame of the reconstructed point cloud. A relationship (or transformation) between the two coordinate systems can be found via the following optimization problem that can readily be solved for the unknown transformation (rotation R and translation t) that minimizes the following error in fit (see Algorithm 1):

$$\min_{R,t} \sum_{i=1}^{N} ||x_{g_i} - (Rx_{p_i} + t)||^2 \qquad (15)$$

In general the more anchor points you have, the more accurately you can find the transformation. [9] The rotation R and translation t are now the key relationship between the point cloud and global coordinate frames and in particular, it can be applied to any object in the point cloud in order to find the global position of that object.

Refining Device Localization: Finally, we note that the above joint optimization has an interesting side-effect: Ubicarse can leverage the relative camera locations output by vision algorithms to greatly refine its own device localization. Specifically, visual algorithms excel at finding relative camera locations, particularly when they observe a rich set features at pixel resolution. Therefore, by applying the transformation to the camera positions returned by VSFM we now have two sets of position estimates for the camera, both in the global reference frame. In this way, one can accurately compensate for relative errors in Ubicarse's indoor positioning. Indeed, our results in §9.5 show that such an optimization greatly improves median localization accuracy by as much as 86% as compared to just using Ubicarse alone.

Algorithm 1 Object Geo-tagging: Finding the transformation, rotation R and translation t, between local and global coordinates

1: **function** GETTRANSFORM(X_g, X_p)
2:     ▷ $X_g = \{x_{g_1}, \ldots, x_{g_N}\}$: Set of N positions in global frame
3:     ▷ $X_p = \{x_{p_1}, \ldots, x_{p_N}\}$: Set of N positions in local camera frame
4:     $C = \sum_{i=1}^{N}(x_{g_i} - \text{mean}(X_g))(x_{p_i} - \text{mean}(X_p))^*$
5:     ▷ Where $(.)^*$: conjugate transpose, and mean($.$): average of elements
6:     $[U, S, V] = \text{svd}(C)$ ▷ Perform SVD of covariance matrix C
7:     $R = VU^*$ ▷ Rotation Matrix
8:     $t = \text{mean}(X_p) - R \times mean(X_g)$ ▷ Translation Vector
9:     **return** R, t
10: **end function**

---

8. IMPLEMENTATION AND EVALUATION

We implemented Ubicarse on a HP SplitX2 Tablet running Ubuntu Linux equipped with Intel 5300 wireless cards and a Yei Technology motion sensor (i.e. an accelerometer, gyroscope and
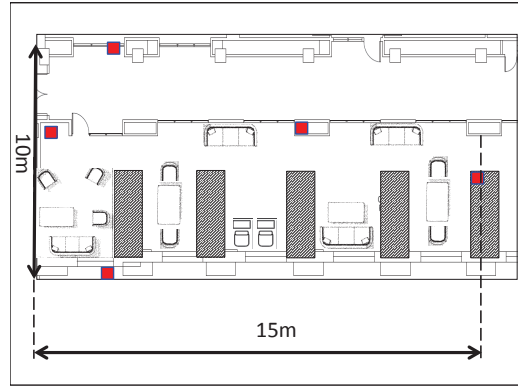

Figure 8: Library floorplan and testbed: APs are red squares; book shelves are shown as hashed rectangles.

compass). We build on the 802.11 CSI tool [13] to obtain the wireless channels on the tablet. We implement Ubicarse's translation resilient SAR completely in software (C++ and Matlab). For each experiment, we allow users to twist their devices randomly along its vertical axis, spanning a wide range of orientations. We configure the tablet to transmit beacon packets (10 times per second) to elicit responses from nearby access points and measure their channels.[10] We then combine this with gyroscope readings from the motion sensor to perform SAR *simultaneously* to all access points, as the user twists the device.

We implement Ubicarse's object localization by integrating it with the VisualSFM (VSFM) toolkit [35] for 3-D reconstruction of objects. We use the tablet's built-in 5.7 Megapixel camera to take images of objects of interest.

We conducted our experiment in a university library, with books arranged in multiple shelves and racks that emulate the complex multipath characteristics of large warehouses and departmental stores, where indoor localization has been of particular interest, as shown in Fig. 8. We leverage five unmodified 802.11n access points in and around the library, as shown by red squares in Fig. 8 to perform device localization and vary the tablet's location between several randomly chosen locations. The access points are configured to the 5 GHz Wi-Fi frequency band. We obtain the locations of access points and tablet centers to sub-centimeter accuracy in a global coordinate frame through direct measurements and using architectural drawings of the library. To perform object localization, we capture images from multiple perspectives of different randomly chosen books in the library. We then apply a joint optimization using VSFM's reconstructed 3D point cloud and Ubicarse's device localization to locate the object.

Baseline: We compare Ubicarse against an angle-of-arrival scheme that treats the two antennas on the device as a 2-antenna array. We chose this baseline, since it uses only 2-antennas like our system, does not require specialized infrastructure or additional calibration input unlike other schemes, leading to a fair comparison.

9. RESULTS

We present four categories of results to evaluate Ubicarse: 1) Validating Translation Resilience of Ubicarse's SAR for different trajectories. 2) Computing Angle of Arrival in 3-D space. 3) Localizing devices in 3-D. 4) Geotagging Objects of Interest.

---

[9] In practice, the above optimization also needs to account for camera calibration and scaling, whose details we omit for brevity.

[10]The results in this paper sample the wireless channels for a minimum of 25-30 packets over a semi-circular arc of rotation.
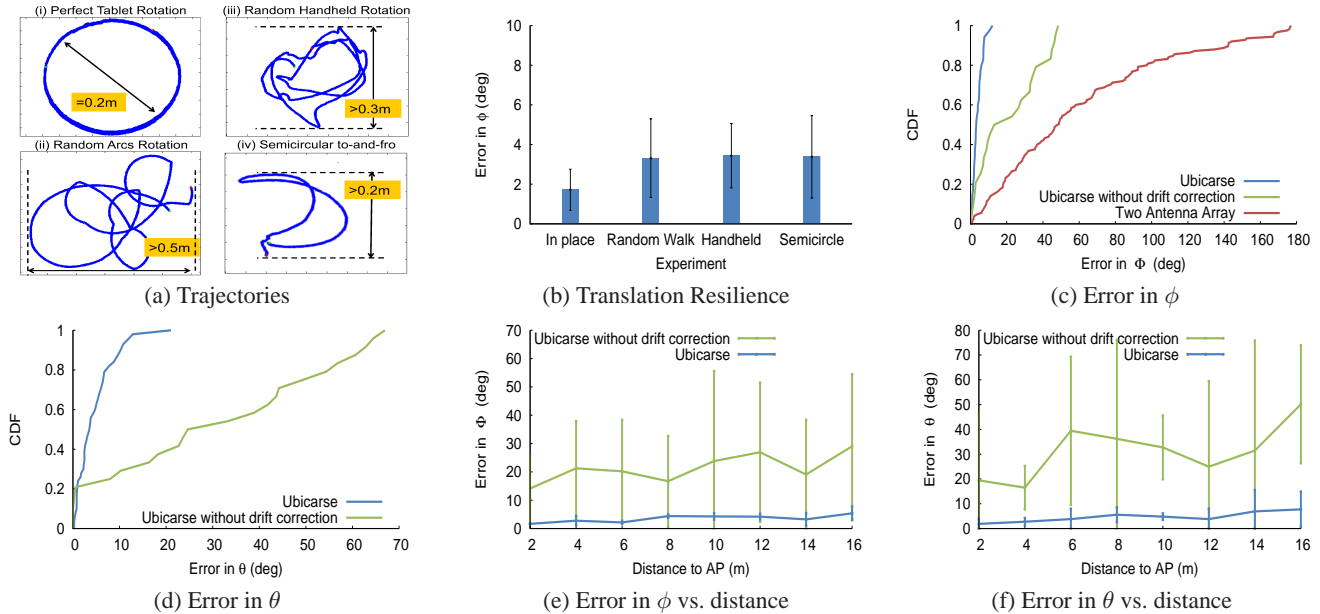
Figure 9: (a) Plots the different trajectories of antenna centers (b) Translation resilience of SAR: Plots mean and Standard deviation of error in azimuthal angle for different trajectories. (c)-(d) CDF of error in Angle of arrival (ϕ and θ), for Ubicarse, Ubicarse without drift correction and 2-Antenna Array. (e)-(f) Error in azimuthal, polar angle versus distance from the access point.

9.1 Translation Resilient SAR

In this experiment, we demonstrate that Ubicarse's new formulation of SAR is resilient to device translation, as described in §4. Specifically, we consider different device trajectories, including random motion of the antenna's center on the order of half a meter.

Method: We perform our experiments in a mm-precision motion capture room [1] where the mobile device was tagged with infrared markers to accurately track the motion of the antenna centers for the purpose of evaluation. SAR was computed entirely using one YEI technologies gyroscope mounted on the tablet and channel measurements acquired from the Wi-Fi card. A single transmitting source was present and the experiment was conducted at distances of 2m to 12m from the source across multiple line-of-sight locations (we consider non-line-of-sight in §9.2).

We moved the tablet in four different types of trajectories: (1) In-place rotation; (2) 2D random trajectory within 0.5 m diameter; (3) A random handheld 3D twisting motion (i.e. including variations in polar angles) up to 0.3 m in diameter (4) A handheld semi-circular to-and-fro twist. For the first two trajectories, we mount the device on a roomba robot for stable trajectories on a 2-D plane.

Fig. 9(a) demonstrates actual trajectory traces from our experiments where these traces are representative of each of the four classes of trajectories described above.

Results: Our results in Figure 9(b) show a mean error of $3.3°$ for the random 2-D trajectory, $3.4°$ even as the center of the antennas is translated randomly by the user in three dimensional space and $3.3°$ if the user traces a simple semi-circular to-and-fro twist. Our results demonstrate that Ubicarse's translation resilient SAR correctly identifies the angle of arrival for the two cases to within $1.6°$, $1.7°$ and $1.6°$ respectively on average as compared to the ideal case of rotation in place.

9.2 Angle of Arrival Estimation on a Tablet

We measure the accuracy of angle-of-arrival estimation in both the azimuthal angle (ϕ) and polar angle (θ) obtained over each handheld twist of the device in line-of-sight (LOS) and non-line-of-sight (NLOS) settings.

Method: We conduct our experiment in the university library as in §8. We place the device at different distances from the access points. In NLOS settings, Ubicarse identifies the direct path using the methods in §6. We compare our results against two important benchmarks: 1) A 2-antenna array composed of the two antennas on the device directly, as opposed to performing SAR. 2) Ubicarse without the gyroscope drift compensation presented in §5.

Results: Fig. 9(c)-(d) shows that our method achieves a median of accuracy of $3.2°$ in the azimuthal orientation of the device and $3.6°$ in polar angle. We observe that gyroscope drift can cause angle inaccuracies of more than $18°$ in ϕ ($29°$ in θ) and as large as $48°$ in ϕ ($67°$ in θ) if left uncompensated. As a result, Ubicarse's drift compensation in §5 plays an important role in achieving high accuracy. Fig. 9(c) also shows that the 2-antenna array has a poor median error of $49°$ in azimuthal angle owing to the lack of resolution from using only two antennas. Note that Figure 9(d) does not compare with the two antenna benchmark since two antennas alone are insufficient to obtain 3-D information [24].

Fig. 9(e)-(f) depicts the error in ϕ and θ respectively, against the distance of the device to the wireless access point. We observe that the error in these angles increases only marginally across distances ranging from $2m$ to $16m$.

9.3 Tablet Localization

In this experiment, we evaluate Ubicarse's accuracy in device localization in the university library setting.

Method: We gather aggregate results on device localization by in twenty five randomly chosen locations where users twist a Ubicarse enabled tablet in the university library setting. Ubicarse's localization was performed to find relative position of the device with respect to commodity access points whose locations were known with respect to the library floor plan. For device and object localization we present results in the three-dimensional Euclidean space as well as global orientation (in the X-Y plane) as given by the ψ angle
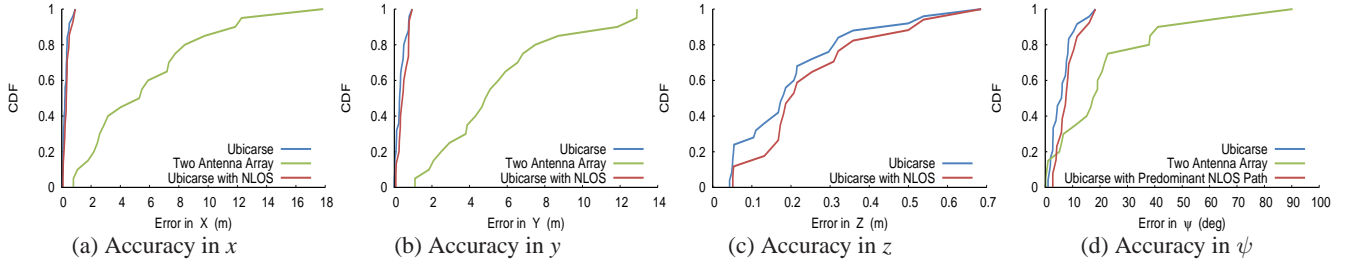
(a) Accuracy in x (b) Accuracy in y (c) Accuracy in z (d) Accuracy in ψ

Figure 10: Tablet Localization: Plots CDF of error in device localization.
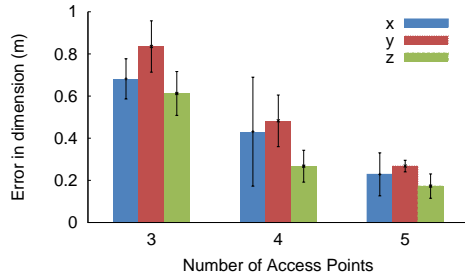

Figure 11: Ubicarse's Device Localization Accuracy measured against the number of wireless access points.

described in §6.[11] These experiments were conducted both in line-of-sight (LOS) and non-line-of-sight (NLOS) settings during peak library hours with pedestrian foot traffic. In particular, we classify locations as non-line-of-sight (NLOS) only when the predominant (highest power) peak in the multipath profile is not the direct path to at least one neighboring AP. Our localization technique explicitly handles multipath as described in section 6.

Results: Our results in Fig. 10 for localization using five surrounding access points demonstrate a median device localization error of 39 cm, where along each dimension we demonstrate a median error of 22 cm in x, 28 cm in y, and 18 cm in z and 6.09° in global device orientation (ψ). Furthermore we show that Ubicarse still performs well in NLOS, incurring a small additional median error of 10 cm in x, 18 cm in y, and 2 cm in z and 7.7° in ψ as compared to the LOS case. Thus, Ubicarse achieves its goal of accurate localization without requiring either new infrastructure or signal fingerprinting.

Fig. 11 plots Ubicarse's error in localization along x, y and z measured against the number of randomly chosen access points used to triangulate the location of the device. We notice that the error in localization reduces with increasing number of surrounding access points. This is because, as described in §6, computing multipath profiles to an increasing number of surrounding access points, can help mitigate ambiguity due to multipath. Further, redundant measurements from over three APs reduces the noise in localization.

9.4 Object Geotagging

In this experiment, we evaluate Ubicarse's object geotagging capabilities (See §7) to localize arbitrarily chosen books on a bookshelf in the library setting.

Method: We conduct our experiments in a library setting with shelving that houses many books, journals, and magazines, a picture of which can be found in Figure 7. This implementation emu-

---

[11]In principle, Ubicarse can be used to find orientations along other planes as well. We omit these results for brevity.

lates similar applications in a warehouse, department store, or other such settings where a person can catalog objects of interest. In particular we show that there is a distinct advantage to localizing far away objects more accurately than simply using the camera's location as a geotag. We compare the object localization we achieve using Ubicarse plus VSFM against the accuracy of using the camera's location directly as a position tag for the book.

Results: Fig. 12(a)-(c) shows that the method of Ubicarse plus VSFM attains a median error in distance of 17 cm, where along each dimension we find an error of 5 cm in x, 15 cm in y and 4 cm in z for localization accuracy of books on library bookshelves whose distance from the camera varies from 1 to 3 m. Therefore simply using the camera's position as a geotag for the books would result in up to 3 m of inaccuracy.

Surprisingly, a comparison of Figures 12 and 10 reveals that *a greater accuracy in object localization* can be obtained by a combination of Ubicarse with VSFM than using either method standalone. This is because these two methods are highly complimentary and a thoughtful integration unlocks an effective method for localization *refinement* (see §7). We provide further results on this concept in the next section.

9.5 Refined Tablet Localization using Vision

The previous section suggests that a method to achieve a fine-scaled object position fix is to integrate an accurate device localization method, Ubicarse, with visual toolkits like VSFM. In this experiment, we study if this improvement applies both ways, i.e. if device localization can be refined using inputs from vision toolkits.

Method: Our method for refinement of camera positions derives from the discussion in §7. We transform the camera locations discovered by VSFM into the global coordinate system, using Algorithm 1. We then output the transformed camera locations as the result of Ubicarse's localization.

Results: Fig. 12(d) shows that a Ubicarse plus VSFM integration achieves a significant improvement in localization performance with a median of 15 cm error from ground truth. We report a median improvement in localization accuracy of 66 % in X, 16% in Y and 86% in the Z dimension. Therefore, integration of Ubicarse with vision algorithms also serves to improve device localization.

10. RELATED WORK

Related work falls under two broad categories:

Indoor RF Localization: RF localization have had three main approaches for indoor localization: The first require deploying specialized infrastructure to perform accurate indoor localization, e.g. acoustic [21, 19], RFIDs [33, 32], specialized access points [15], and antenna arrays [37]. The second mandates signal fingerprinting [3, 27, 39] either in a training phase or via crowdsourcing [26, 38, 18]. The third advocate modeling the wireless signal instead of
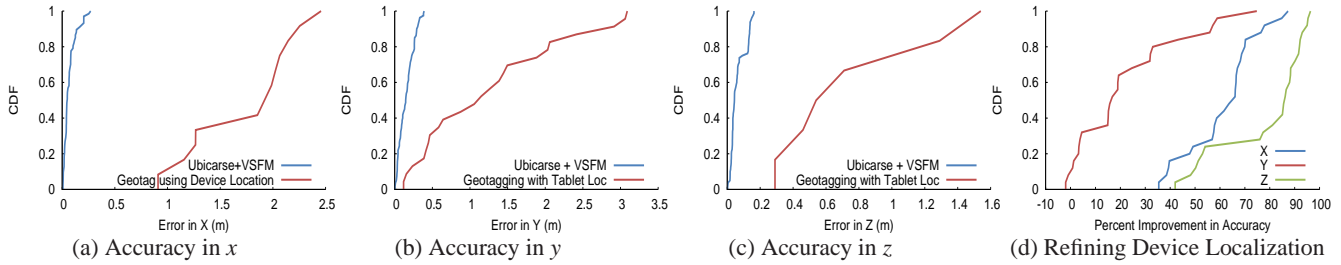
Figure 12: (a)-(c) **Object Localization**: Plots CDF of error in object geotagging. (d) Demonstrates refinement of Device localization using object geo-tagging

measurements [17, 5, 12], but at the expense of accuracy. In contrast to these systems, Ubicarse obviates the need for new infrastructure or fingerprinting to achieve tens of centimeter of accuracy by emulating large antenna arrays using only commodity mobile devices.

Ubicarse builds on past work in Synthetic Aperture Radar (SAR) that leverage precise mechanically controlled antenna movements or advanced motion sensing equipment for radar imaging [9] and RFID localization [33, 32]. Unlike these systems, Ubicarse provides a new formulation of SAR that is translation resilient, making it suitable for mobile devices held by users.

Several RF localization papers use mobile motion sensors as hints for localization, either coupled with Wi-Fi signal strength for fingerprinting [36, 3, 26] or with GPS measurements [6, 31]. Instead, Ubicarse uses the device's gyroscope only to measure its relative orientation to help emulate antenna arrays and achieves significantly improved accuracy.

Object Geo-tagging: There has been much work in computer vision on stereo-imaging and 3D-reconstruction to localize objects relative to known camera locations or to find relative locations of different objects in the same scene [2, 29]. Ubicarse builds on this work, coupled with RF localization, to obtain object locations in the global frame.

Several solutions for indoor device and object localization have been proposed in the literature that use specialized hardware, e.g. depth cameras [20] or odometry/laser-based simultaneous localization and mapping [8], which are unavailable for today's mobile devices, unlike Ubicarse.

11. CONCLUSION AND FUTURE WORK

We presented Ubicarse, an indoor localization system to achieves tens of centimeters of accuracy on commodity mobile devices, without requiring specialized infrastructure or fingerprinting. Ubicarse's provides a new formulation of SAR that allows mobile devices to emulate an antenna array; where this formulation is tolerant to unknown translation of the antenna center of up to half a meter. We implement Ubicarse on a commodity tablet and demonstrate tens of centimeter accuracy in both device localization and object geotagging in complex indoor settings. We believe evaluating our system on a wide-range of mobile devices, particularly smartphones, is an important task for future work.

12. REFERENCES

[1] VICON T-Series. http://www.vicon.com. VICON.

[2] M. Agrawal and K. Konolige. Real-time localization in outdoor environments using stereo vision and inexpensive gps. ICPR. 2006.

[3] M. Azizyan, I. Constandache, and R. Roy Choudhury. Surroundsense: Mobile phone localization via ambience fingerprinting. MobiCom, 2009.

[4] F. Bowman. *Introduction to Bessel Functions*. Dover Publications, 2012.

[5] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan. Indoor localization without the pain. MobiCom, 2010.

[6] I. Constandache, R. Roy, and C. I. Rhee. Compacc: Using mobile phone compasses and accelerometers for localization. In *In INFOCOM*, 2010.

[7] R. Fenby. Limitations on directional patterns of phase-compensated circular arrays. *REE*, 1965.

[8] B. Ferris et al. WiFi-SLAM using gaussian process latent variable models. IJCAI, 2007.

[9] P. J. Fitch. *Synthetic Aperture Radar*. Springer, 1988.

[10] H. Foroosh, J. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *Image Processing, IEEE Transactions on*, 2002.

[11] J. Georgy et al. Modeling the stochastic drift of a mems-based gyroscope in gyro/odometer/gps integrated navigation. *IEEE ITS*, 2010.

[12] A. Goswami, L. Ortiz, and S. Das. Wigem: A learning-based approach for indoor localization. CoNEXT, 2011.

[13] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR*, 2011.

[14] M. Hölzl, R. Neumeier, and G. Ostermayer. Analysis of compass sensor accuracy on several mobile devices in an industrial environment. In *EUROCAST*. 2013.

[15] K. Joshi, S. Hong, and S. Katti. PinPoint: Localizing Interfering Radios. NSDI, 2013.

[16] Z. Li et al. Robust statistical methods for wireless localization in sensor networks. In *IPSN*, 2005.

[17] H. Lim, L.-C. Kung, J. C. Hou, and H. Luo. Zero-configuration indoor localization over 802.11 wireless infrastructure. *IEEE Wirel. Netw.*, 2010.

[18] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye. Push the limit of WiFi based localization for smartphones. Mobicom, 2012.

[19] K. Liu et al. Guoguo: Enabling fine-grained indoor localization via smartphone. MobiSys, 2013.

[20] Microsoft. Kinect. http://www.kinect.com.

[21] R. Nandakumar et al. Centaur: Locating devices in an office environment. Mobicom, 2012.

[22] M. Niessner, A. Dai, and M. Fisher. Combining Inertial Navigation and ICP for Real-time 3D Surface Reconstruction. Eurographics, 2014.

[23] H. Nyqvist et al. A high-performance tracking system based on camera and IMU. In *FUSION*, 2013.

[24] S. J. Orfanidis. Electromagnetic waves and antennas. http://www.ece.rutgers.edu/ orfanidi/ewa/.

[25] H. Rahul et al. Jmb: Scaling wireless capacity with user demand. In *SIGCOMM*, 2012.

[26] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen. Zee: Zero-effort crowdsourcing for indoor localization. Mobicom, 2012.

[27] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka. You are facing the mona lisa: Spot localization using phy layer information. MobiSys, 2012.

[28] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik. Indoor localization without infrastructure using the acoustic background spectrum. MobiSys, 2011.

[29] B. Triggs et al. Bundle adjustment – a modern synthesis. In *Vision Algorithms Theory/Practice*, 2000.

[30] D. Tse and P. Vishwanath. *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[31] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury. No need to war-drive: Unsupervised indoor localization. MobiSys, 2012.

[32] J. Wang, F. Adib, R. Knepper, D. Katabi, and D. Rus. RF-compass: Robot object manipulation using rfids. MobiCom, 2013.

[33] J. Wang and D. Katabi. Dude, where's my card?: RFID positioning that works with multipath and non-line of sight. SIGCOMM, 2013.

[34] WiGle. Database. https://wigle.net.

[35] C. Wu. VisualSFM : A Visual Structure from Motion System. http://ccwu.me/vsfm/.

[36] C. Wu, Z. Yang, Y. Liu, and W. Xi. Wireless indoor localization without site survey. In *INFOCOM*, 2012.

[37] J. Xiong and K. Jamieson. Arraytrack: A fine-grained indoor location system. NSDI, 2013.

[38] Z. Yang, C. Wu, and Y. Liu. Locating in fingerprint space: Wireless indoor localization with little human intervention. Mobicom, 2012.

[39] M. Youssef and A. Agrawala. The Horus WLAN location determination system. MobiSys, 2005.

APPENDIX

A. PROOF OF LEMMA 4.1

In this section, we derive the expected value of SAR multi-path power profiles for line-of-sight scenarios as stated in Lemma 4.1. We consider the scenario in Fig. 3, where a transmitting source T is in line-of-sight relative to a two-antenna receiver device. By substituting the relative wireless channel from Eqn. 7 into the multipath power profile from Eqn. 8, and simplifying terms we get:

$$\hat{h}_i = h_{2,i}h_{1,i}^* = \frac{1}{d^2}e^{\frac{-j2\pi r cos(\alpha_T - \phi_i)}{\lambda}} \quad (16)$$

Thus, the multi-path power profile (See Eqn. 8) is:

$$P(\alpha) = \frac{1}{d^4}\left|\frac{1}{n}\sum_{i=1}^{n}e^{-j\kappa\sin\psi_i}\right|^2 = \frac{1}{d^4}\left|\frac{1}{n}\sum_{i=1}^{n}e_i\right|^2 \quad (17)$$

Where $\kappa = \frac{4\pi r \sin\left(\frac{\alpha_T - \alpha}{2}\right)}{\lambda}$, is independent of device orientation, $\psi_i = \phi_i - \frac{\alpha_T + \alpha}{2}$ and $e_i = e^{-j\kappa\sin\psi_i}$.

At this point, our goal is to evaluate the expected value of $P(\alpha)$ over the device orientation between snapshots. To do this, let us consider the case where the snapshots are taken at random device orientations spanning an angle of π (i.e. a semi-circle) so that $\phi_i \sim \mathcal{U}(\gamma - \pi/2, \gamma + \pi/2)$ is chosen uniformly at random. Since γ depends on ϕ_0, we choose ϕ_0 such that, $\psi_i \sim \mathcal{U}(-\pi/2, \pi/2)$ is also a uniformly distributed random variable. Under these assumptions, let us look at the distribution of the quantity $e_i = e^{-j\kappa\sin\psi_i}$. Specifically, the mean and variance of e_i is given by $E[e_i] = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}e^{-j\kappa\sin\psi}d\psi = J_0(\kappa)$ and $\text{Var}[e_i] = E[|e_i|^2] - |E[e_i]|^2 = 1 - J_0(\kappa)^2$ respectively, where $J_0(\cdot)$ is the well-known Bessel-function of the first kind[4].

Next, we look at $\frac{1}{n}\sum_{i=1}^{n}e_i$. Applying central limit theorem for large n, this quantity is normally distributed with mean $J_0(\kappa)$ and variance $\frac{1-J_0(\kappa)^2}{n}$. Specifically, the variance of this distribution:

$$\text{Var}\left[\frac{1}{n}\sum_{i=1}^{n}e_i\right] = E\left[\left|\frac{1}{n}\sum_{i=1}^{n}e_i\right|^2\right] - \left|E\left[\frac{1}{n}\sum_{i=1}^{n}e_i\right]\right|^2$$

$$\frac{1-J_0(\kappa)^2}{n} = d^4 E[P(\alpha)] - J_0(\kappa)^2$$

$$\tilde{P}(\alpha) = E[P(\alpha)] = \frac{1}{d^4}\left[J_0(\kappa)^2 + \frac{1-J_0(\kappa)^2}{n}\right]$$

Which desired expected SAR profile, as in Lemma 4.1. □

B. PROOF OF LEMMA 4.2

We derive the expected SAR power profiles in multipath scenarios as in Lemma 4.2. Recall the relative wireless channel between two antennas of the receiver at snapshot i is given by Eqn. 13, such that the multipath profile $P(\alpha)$, depends on the sum of terms of the form $g_i = \hat{h}_i e^{\frac{j2\pi r \cos(\alpha - \phi_i)}{\lambda}}$ (Eqn. 8). Each term is:

$$g_i = \sum_{k=1}^{m}D_k e^{\frac{-j2\pi}{\lambda}r(\cos(\alpha_k - \phi_i) - \cos(\alpha - \phi_i))} = \sum_{k=1}^{m}D_k e^{-j\kappa_k\sin\psi_{i,k}}$$

Where similar to Appendix A, $\kappa_k = \frac{4\pi r \sin\left(\frac{\alpha_k - \alpha}{2}\right)}{\lambda}$ and $D_k = \frac{s_k}{d_k}\left[\frac{s_k}{d_k} + \sum_{l\neq k}\frac{s_l}{d_l}e^{+j2\pi(\frac{\Delta y_i}{\sin\alpha_l} - \frac{\Delta y_i}{\sin\alpha_k})}\right]$, depend only on translation, while $\psi_{i,k} = \phi_i - \frac{\alpha_k + \alpha}{2}$ depends only on orientation. As a result, D_k and $e^{-j\kappa_k\sin\psi_{i,k}}$ are independent, and we can write: the expectation $E[g_i] = \sum_{k=1}^{m}E[D_k]E[e^{-j\kappa_k\sin\psi_{i,k}}] = \sum_{k=1}^{m}\left[\frac{s_k^2}{d_k^2} + E_k\right]J_0(\kappa_k)$, assuming that ϕ_i's are chosen uniformly at random. Here we denote $E_k = E\left[\sum_{l\neq k}\frac{s_k s_l}{d_k d_l}e^{\frac{j2\pi\Delta y_i(\frac{1}{\sin\alpha_l} - \frac{1}{\sin\alpha_k})}{\lambda}}\right]$.

Finally, let us look at the summation $\frac{1}{n}\sum_{i=1}^{n}g_i$. Applying central limit theorem for large n, this quantity is normally distributed with mean $\sum_{k=1}^{m}\left[\frac{s_k^2}{d_k^2} + E_k\right]J_0(\kappa_k)$ and variance $\frac{Var[g_i]}{n}$. Specifically, the variance of this distribution is:

$$\text{Var}\left[\frac{1}{n}\sum_{i=1}^{n}g_i\right] = E\left[\left|\frac{1}{n}\sum_{i=1}^{n}g_i\right|^2\right] - \left|E\left[\frac{1}{n}\sum_{i=1}^{n}g_i\right]\right|^2$$

$$E[P(\alpha)] = \left(\sum_{k=1}^{m}\left[\frac{s_k^2}{d_k^2} + E_k\right]J_0(\kappa_k)\right)^2 + \frac{1}{n}\text{Var}[\hat{h}_i.e^{\frac{j2\pi r\cos(\alpha - \phi_i)}{\lambda}}]$$

Which desired expected SAR profile, as in Lemma 4.2. □