Hierarchical Learning For FM Radio Based Aerial Localization Using RSSI

Tathagata Mukherjee, Andreas Adolfsson TATHAGATA|ANDREAS@INTELLIGENTROBOTICS.ORG
Intelligent Robotics Inc., 3697 Longfellow Road, Tallahassee, FL 32311

Eduardo Pasiliao EDUARDO@AFRL.ORG
Air Force Research Labs, Shalimar FL

Piyush Kumar PIYUSH@COMPGEOM.COM
CompGeom Inc., 3748 Biltmore Ave., Tallahassee, FL 32311

Abstract

Received Signal Strength Indicator (RSSI) based large scale positioning systems are beginning to gain traction as coarse positioning systems when GPS is unavailable. In this paper we present a system for automatic positioning of an unmanned aerial system using broadcast FM radio. Our method is data driven, and uses machine learning techniques to improve its accuracy. The techniques are easy to extend to other terrestrial static radio transmitters. Using our algorithms, we can localize with a minimum error of 172 meters and mean error of less than 3000 meters.

1. Introduction

Localization using ambient wireless signals have generated a lot of interest recently, both for indoor, as well as outdoor localization (Chu & Jan, 2007; Popleteev, 2011; Krumm et al., 2003). Use of wireless signals have been extensively studied for indoor localization (Popleteev, 2011; Zheng et al., 2016; Martin et al., 2010) and the achievable accuracy continues to improve with time. Though there are no accepted standards for indoor localization, WiFi based localization is most common (Martin et al., 2010) and other methods continue to be studied in detail, examples of other modalities being GSM (Varshavsky et al., 2007) and FM (Chen et al., 2012). For outdoor localization, however the state of the art is GPS (Misra & Enge, 2006). Though GPS has been in use for a long time and is the de-facto standard for large scale outdoor localization, there are situations where GPS becomes unreliable or is unavailable altogether (Cameron, 2016; gps, 2016). As a result, there is a need for research into augmenting the existing GPS systems with external assistance. Assisted GPS (LaMance et al., 2002) and differential GPS (Zhao et al., 2014) are examples of systems that aim to augment and assist existing GPS systems for accurate and reliable positioning. They assume that though the GPS signal is degraded, it is still available and can be augmented and used. This is not always the case and hence there is need for alternative methods of positioning.

Several different techniques have been used to implement radio frequency (RF) based localization systems. Some of these techniques include anchor based approaches (Srinivasan & Wu, 2007; Popleteev, 2011; Fang et al., 2009; Otsason et al., 2005; Savvides et al., 2001), those using Time of Arrival (TOA) (Fuller, 2009), Time Difference of Arrival (TDoA) (Cong & Zhuang, 2001), and Angle of Arrival (Savvides et al., 2001). However one of the most common approach for building very large scale localization systems using FM and other electromagnetic waves is based on the analysis of the received signal strength (RSS) (Kumar et al., 2015; Mukherjee et al., 2017).

Received Signal Strength Indicator (RSSI) at the receiver is dependent on the hardware of the receiver (Zheng et al., 2016), the location and power of the transmitter, and the ambient medium. Given a transmitter t, the RSSI at the receiver, which is at a distance d from the transmitter, is given by

$$r_d^T = r_{d_0}^T - \beta \log(\frac{d}{d_0}) + \eta_d \tag{1}$$

where η_d is the noise at the location, which is at a distance d from the transmitter. This equation assumes that a RSSI measurement of $r_{d_0}^T$ is available at a reference location, at a distance d_0 from the transmitter. (Rappaport et al., 1996). Under the model, η_d has a Gaussian distribution with zero mean and an unknown variance and models the uncertainty in the environment. RSSI based localization is easy to implement, since the RSSI can be read directly from the hardware, and the computation complexity is low. On the flip side, the accuracy of such methods are usually low if the RSSI is used directly for the localization.

Proceedings of the 2^{nd} GNU Radio Conference, Copyright 2017 by the author(s).

This is primarily because the RSSI is affected by the state of the ambient medium, the quality of the device used for measurements and the multi-path. Thus, for improving the accuracy, RSSI based systems require careful feature engineering (Popleteev, 2011; Zheng et al., 2016; Kumar et al., 2015; Mukherjee et al., 2017).

Large scale RSSI based FM localization systems have been shown to have errors of around 4 miles on average, when used in conjunction with estimated FM signal maps (Kumar et al., 2015; Mukherjee et al., 2017). However the experiments reported in (Kumar et al., 2015; Mukherjee et al., 2017) were conducted with data collected from a moving vehicle, traveling on a highway at high speeds and over a very large area without explicitly modeling the noise. It is known that wireless signals are affected by multi-path, which can alter the distribution of the RSSI and hence if the effect of the multi-path is not factored in or eliminated, the localization accuracy can suffer. In this work we explore the effects of the latter, that is, the effect of eliminating the multi-path, as much as possible, on the positioning accuracy. To that effect we collect data high up in the air, where there is *line of sight* with the FM transmitters and the multi-path is low because of lack of reflecting surfaces. In this paper we report the data collection methods, a supervised learning approach for FM signal estimation at a location and the positioning algorithms that use the estimated FM signal for localization. We report the results of using our localization algorithms on data collected in and around Tallahassee.

Notation: We denote locations (and vectors) in any region by lowercase *bold* letters as $\mathbf{x}$, $\mathbf{y}$. *Bold* lowercase *Greek* letters like $\boldsymbol{\psi}$ are used for the power spectrum and its subsets. Scalars including indexes are represented by *lowercase* letters like i, j and k.

Next we briefly discuss the previous attempts to tackle similar problems.

2. Previous Work

In this work we are concerned with absolute positioning techniques. In particular, we focus on the problem of finding the absolute coordinates of a point in a fixed reference frame. Absolute positioning can be done using two methods. The first approach relies on communications with a Global Positioning System (GPS) whereas the second achieves its objective without any such communication. Traditional GPS based localization (Misra & Enge, 2006) uses GPS receivers to communicate with several GPS satellites. The received data is used to compute the distance of the object from at least four known GPS satellites using the idea of time of arrival (TOA) (Fuller, 2009). The final position is found using trilateration. GPS based systems suffer from several limitations, namely, lack of precision (Khattab et al., 2015), jamming (Waterman, 2012), disruption and spoofing (Psiaki & Humphreys, 2016). To get around these problems researchers have used the idea of assisted GPS (Djuknic & Richton, 2001) and differential GPS (Kaplan, 1996). More recently, work has been done in order to achieve centimeter level accuracy with GPS (Farrell et al., 2000; Talbot et al., 1996; Parkinson et al., 2000; OConnor, 1997; Zhao et al., 2014).

Apart from GPS based positioning methods, there are absolute positioning techniques that do not depend on GPS. These methods are usually called GPS-*free* localization techniques. One of the most common forms of GPS-free positioning is called *Network based Geolocation* (Djuknic & Richton, 2001; Gustafsson & Gunnarsson, 2005; Borkowski et al., 1996). These methods are almost exclusively based on technologies that depend on wireless networks and use signal processing heavily. They use techniques such as time of arrival, time difference of arrival, angle of arrival, timing advance and multipath fingerprinting (Vander Stoep, 2009; Fuller, 2009; Savvides et al., 2001; Oguejiofor et al., 2013; Mao et al., 2007; Ibrahim & Youssef, 2012). Examples include AM based localization (McEllroy et al., 2001) and Locata (Barnes et al., 2003) which uses a type of *Network based geo-location*.

Fingerprint based localization systems have been extensively studied for indoor (Yang et al., 2012) as well as outdoor localization. The fingerprints can be received signal strengths for WiFi based localization (Atia et al., 2013; Chen et al., 2013; Haeberlen et al., 2004) or FM based localization (Kumar et al., 2015; Mukherjee et al., 2017). They can also be readings obtained from inertial sensors, which may have unique characteristics at given locations (Abdelnasser et al., 2015). These methods has been extensively used for building indoor localization systems (Otsason et al., 2005; Chen et al., 2012; Abdelnasser et al., 2015; Chen et al., 2013).

Smartphones are being increasingly used for building localization systems. Laoudias et al. (Laoudias et al., 2012; Li et al., 2012) built such a system using WiFi fingerprints collected from smartphones. A crowdsourced version of a similar system was implemented by Petrou et al. (Petrou et al., 2014). Konstantinidis et al. have studied privacy preserving indoor localization using smartphones (Konstantinidis et al., 2015). Azizyan et al. and Aly et al. have also used cell phones for fingerprint based localization (Azizyan et al., 2009; Aly & Youssef, 2013). Abdelnasser et al. (Abdelnasser et al., 2015) implemented a system for indoor localization with fingerprints constructed from different sensor data, on a mobile phone.

An important variation of network based geolocation, is called *signals of opportunity* (SOO) based localization

(Counselman III & Hall, 2002; Yang et al., 2014). FM fingerprint based localization in a small area using *correlation* as a distance metric was studied in (Fang et al., 2009). In general SOO based positioning systems use all the different types of available RF signals in the environment, to create a fingerprint database. Different types of RF signals such as the Global System for Mobile Communications (GSM) (Otsason et al., 2005; Varshavsky et al., 2007), WiFi signal (Ocana et al., 2005; Martin et al., 2010), FM (Chen et al., 2012; 2013; Moghtadaiee et al., 2011) or TV signals (Engelbrecht & Weinberg, 1996) can be used for the positioning. Unlike GPS, these systems can be used for indoor localization and are known to give errors of less than 3m (Moghtadaiee et al., 2011; Martin et al., 2010).

One challenging aspect of RSS fingerprint based localization systems is the fact that the RSS distribution changes over time and with devices. Hence there is need for device calibration and regular updates of the underlying fingerprint database, over time. These are usually laborious processes. To alleviate these problems, *Transfer Learning* methods have been studied for RSS fingerprint based systems (Zheng et al., 2016; 2008b;a).

3. Data Collection

Our data collection system consists of a modified DJI S1000+ octocopter, fitted with a Pixhawk autopilot running PX4 1.6.5 firmware, 3DR GPS module, an i7 NUC computer, a bluetooth speaker, and a Logitech c920 camera. A RTL-SDR dongle is connected to the NUC and is used for FM RSSI data acquisition. Our FM Antenna is mounted vertically on the top of the octocopter. Figure 1 shows our octocopter while ascending. We would like to mention that when mounted horizontally, the results of localization were considerably bad, and hence we chose to fix the orientation of the antenna to be vertical.

The NUC is connected to the Pixhawk autopilot (Meier et al., 2011) using the telemetry port. The NUC runs Ubuntu 16.04 LTS and connects to the Pixhawk using mavlink protocol (Meier et al., 2013). Using this protocol, we can read and control the octocopter. Our code is written in Python and uses multiple libraries and tools like redis, pymavlink, mavproxy and espeak.

The data collection system is autonomous. Once armed, it first checks the accuracy of the GPS. If the GPS error is tolerable, it makes an automatic RSSI reading on the ground and then lifts off to 120 meters in the air, stays there till another RSSI reading is collected. Once this is done, the octocopter lands autonomously. For interested readers, a demonstration of this process is shown in the video available on YouTube at: `https://www.youtube.com/watch?v=DYP22RmxbQ8`

Figure 1. A picture of our assembled drone in the air.

We collected 30 data points at a height of 120 meters in Tallahassee. A plot of the GPS coordinates of these locations is shown in Figure 2. Note that the preparation needed to take one aerial reading is upwards of an hour because of the time it takes to charge batteries, wait for suitable weather conditions and get permissions for flying in an area.

Before moving on to describe our algorithm for position estimation, we describe in a nutshell the work of (Mukherjee et al., 2017; Kumar et al., 2015), which forms the first step of the learning algorithm. This is important as our algorithm bootstraps on earlier methods (Kumar et al., 2015; Mukherjee et al., 2017).

4. Bootstrapping Method

Our positioning algorithm is based on two phases: the first phase uses the methods described in (Kumar et al., 2015; Mukherjee et al., 2017) for computing a coarse position from the observed data, which is improved upon in the second phase of the algorithm. The first phase of computing the coarse location consists of three sub-phases:1) Model Estimation 2) Feature Extraction and 3) Coarse Localization. As with any learning based method (Bishop, 2006; O'Shea & West, 2016) the *model estimation* phase creates a model for the expected FM spectrum for the entire contiguous United States (COTUS). This model is used for

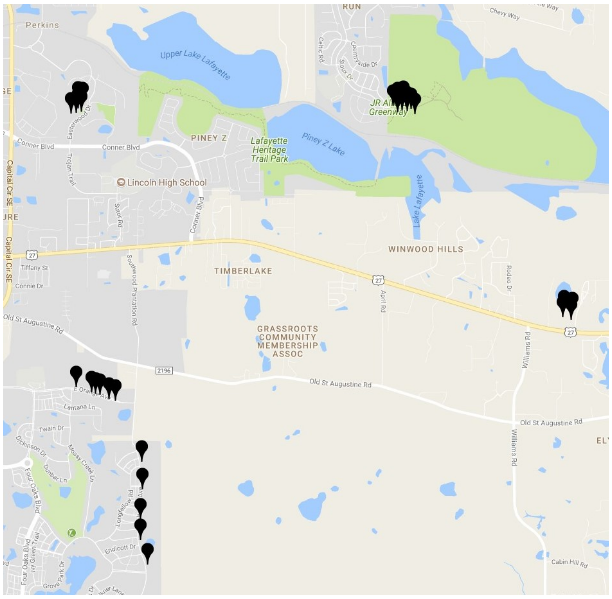

Figure 2. Thirty locations where aerial measurements were taken in Tallahassee, FL.

computing the coarse location of a point of interest based on the observed FM spectrum at that point. Given the observed FM spectrum, the next step in determining the coarse location consists of computing the *dominant channel descriptor* (DCD) features, which are finally used for coarse positioning. We now describe each of these steps in short. Interested readers are referred to the original papers for more detailed descriptions and analysis of the methods.

4.1. Model Estimation

Here the goal is to be able to learn a model that predicts the expected power across the FM channels at a point of interest **x** based on the knowledge of nearby FM transmitters, the power at which they are transmitting and the description of a region around the transmitter that receives a fixed power from the transmitter. Informally, to estimate the expected FM power spectrum in the region of interest, the region is divided into geohashes (Fox et al., 2013) of a fixed precision. The data for estimating the expected FM spectrum consists of information about FM transmitters in the region of interest. More specifically the algorithm assumes that it knows the geo-location of the transmitter denoted by **t**, the radius of its influence denoted by r , and the p-dbu contour plot for the transmitter denoted by **p**. Typically, this contour plot is a *star* polygon (O'Rourke, 1998) with 360 vertices. Given this information, for every transmitter, the algorithm can estimate the expected power at all geohashes within the radius of influence of the transmitter. For points that are in the radius of influence of several transmitters, it gets the total estimated power by adding up contributions from each transmitter affecting the area. For a given transmitter **t** and a point **x** in the circle centered at **t** of radius r the algorithm first computes the intersection of the line joining the points **x** and **t** with the p-dbu polygon **p**. This intersection can be computed using a line sweep algorithm (De Berg et al., 2000). After this step, the problem reduces to that of interpolating or extrapolating the power at **x**, using the value of the power at the intersection, which is known. It must be noted that this step is analogous to the *training phase* of a learning algorithm (Bishop, 2006). The training data, in this case, is the information about the transmitters. For more information about the different ways of creating the so called "fingerprint" databases please refer to (Kjærgaard, 2007).

4.2. Feature Extraction

Given a point of interest **x** that is to be localized, the algorithm starts by looking at the RSS values of the FM signals received at that point. Using the observed RSS values for the 101 channels directly might be problematic. The data is very *high* dimensional, the observed RSSI may be different from device to device, on the same device under different environmental conditions and finally the observed RSSI is corrupted by noise which is uncalibrated. All of these precludes the use of the observed RSSI directly. Intuitively, to find the DCD features the extraction algorithm looks for channels that significantly "dominate" its local neighborhood, thereby making sure that these channels are significantly above the noise level and also have enough received power to be discriminative for location inference.

Given the observed spectra denoted by $\boldsymbol{\psi}$, the i^{th} observed value is selected as a DCD feature if and only if it satisfies the condition:

$$min(\boldsymbol{\psi}_i - \boldsymbol{\psi}_{i-1}, \boldsymbol{\psi}_i - \boldsymbol{\psi}_{i+1}) > \nu$$

for $i \in [2, \ldots, |\boldsymbol{\psi}| - 1]$ and for some constant $\nu > 0$, that should ideally depend on the data and the device used for sensing the spectrum. Note that there are two boundary cases: the first one occurs when $i = 0$ and the second one occurs when $i = |\boldsymbol{\psi}| - 1$. In case, when $i = 0$, the algorithm checks whether $\boldsymbol{\psi}_i - \boldsymbol{\psi}_{i+1} > \nu$. Similarly for the case where $i = |\boldsymbol{\psi}| - 1$ it checks $\boldsymbol{\psi}_i - \boldsymbol{\psi}_{i-1} > \nu$ to determine whether $\boldsymbol{\psi}_i$ is a DCD feature or not.

4.3. Coarse Localization

Given the observed FM spectrum at a given point of interest, the algorithm first extracts the DCD features and then finds candidate locations in the COTUS where there is a high probability of observing the pattern of the observed DCD features. This is done using a subset filtering mechanism which is based on the well known subset query problem (Charikar et al., 2002). Finally the coarse location is computed using an Euclidean nearest neighbor search in the space of DCD features amongst these selected candidates.

Given this background, we are now ready to describe our algorithm, which takes the coarse location obtained from this first phase and attempts to compute a more accurate position using the estimated model and the DCD features.

5. Positioning Algorithm

Our positioning system bootstraps by invoking the algorithms discussed in the work of (Kumar et al., 2015; Mukherjee et al., 2017), on the spectra acquired by the unmanned aerial vehicle (UAV) at a height of 120 meters. This results in the first approximation to the unknown location of the UAV. The system assumes that we have *a priori* knowledge of the transmitted power and location, of all FM transmitters across the region of interest. For every transmitter t at location $\mathbf{t}$, it needs the radius of influence r and the p-dbu polygon $\mathbf{p}$. Given this information, it computes a model that can take a FM power spectrum and output an approximate localization using an Euclidean nearest neighbor search in the *Dominant Channel Descriptor* (DCD) feature space as described in Section 4.

We had access to both the ground and aerial spectra for each of the data points. As a result we compared the accuracy of the first level approximate positioning for the ground spectra vs aerial spectra. The average localization errors for aerial spectra were much better than the ground, and hence we chose to use localization of the aerial spectra as our first step in the algorithm. For example, the localization of the 30 ground data points has a mean error of 31.6 miles whereas the average error for the aerial data is 3.74 miles for the same location. The main reason for this discrepancy is the fact that one of the places where we collected data had a different distribution of noise in the FM channels, than that assumed by the approximate localization system of Kumar et al. (Kumar et al., 2015). Hence the number of DCD features (Mukherjee et al., 2017) obtained were nominal, resulting in low positioning accuracy (Mukherjee et al., 2017).

For this work, we assume that our UAV can be anywhere in the continental United States. After we use the approximate localization system (Kumar et al., 2015) to compute a position for the aerial FM spectra, we consider a 16 km^2 region around this position, and then construct a grid over this region with one grid cell every 100 meters. Each grid cell is a candidate for the second level fine grained positioning system. Thus, using this process we generate 160×160 candidates from which we need to select one as the final position. We denote each candidate by $\mathbf{g}_i, i \in [160 \times 160]$. Before we describe the algorithm for selecting a candidate for final positioning, we explain a supervised learning algorithm for estimating the distance of a location from a given FM transmitter, using RSS values at the receiver and information about the transmitter.

5.1. Distance Estimation for Positioning

As mentioned above, for the aerial FM spectrum, we first use an approximate localization system from Kumar et al (Kumar et al., 2015). This involves the computation of DCD features (Mukherjee et al., 2017). The first step of our algorithm is to determine for each of the computed DCD features, the transmitter responsible for generating the feature. This is done using a hierarchical algorithm. The intuition driving this algorithm is the fact that the channel selected as a DCD feature must be the result of the nearest transmitter, to the approximate position, transmitting at the DCD channel frequency. In cases where there is ambiguity with regards to the transmitter, we refrain from processing the corresponding DCD feature. This simple heuristic identifies a set of transmitters that, with high probability, explains the observed DCD features at the location of interest. Let these transmitters be denoted by $\mathbf{t}_i, \forall i \in \{1, \ldots, k\}$. Now, given the information about the transmitters, one can easily use the *free space path loss model* (Rappaport et al., 1996) to convert RSSI for each of the DCD feature, in the observed spectrum, to a distance estimate from the transmitter, whose location is known *a priori* in GPS coordinates.

We would like to point out that for getting the initial approximation to the position using the algorithms from (Kumar et al., 2015), we use a large scale map estimate that predicts the expected FM spectrum at each location in the continental United States. However the map that we have used was generated in March 2017, and hence the estimates are old. This step for estimating the distances, in essence updates this map and can be interpreted as a *transfer learning* (Pan & Yang, 2010) step.

Instead of using the free space path loss model, we learned a model using supervised learning techniques (a random forest) to estimate this distance (Liaw et al., 2002). This worked consistently better than the free space path loss model (Rappaport et al., 1996) for our problem, most likely because of errors induced by the measuring instruments and the fact that we were dealing with the speed of light which can amplify small measurement errors. Our model learns to estimate the distance to the transmitters from a given location, given the transmitted power, the received power at the location, the height of the receiver, and the *height above average terrain* (HAAT) of the transmitter. Using this model, we can estimate a vector of distances, where each component corresponds to the channel frequency of a DCD feature and gives the estimated distance of the location of interest from the corresponding transmitter. Note that this vector might not have estimates for certain DCD features because of transmitter ambiguity. We denote this vector by $\mathbf{v} \in \mathbb{R}^k$.

5.2. Positioning

Now we are ready to describe the process of selecting one of the 160×160 candidates as a better approximate localization. For each candidate $\mathbf{g}_i$, we compute a vector $\mathbf{v}_i \in \mathbb{R}^k$, each element of which is given by $d(\mathbf{t}_j, \mathbf{g}_i) \forall j \in \{1, \ldots, k\}$, where $d()$ computes the distance between the position of the transmitter and that of the candidate. The distance function is implemented using *Vincenty distance*, which uses an ellipsoidal model of the earth (from `geopy` package). To compute the approximate position of the UAV, we minimize the function $m(\mathbf{v}, \mathbf{v}_i, \rho = 0.75)$ where i iterates through all candidates and $\mathbf{v}$ is the vector of distances between the approximate location obtained from Kumar et al. (Kumar et al., 2015) and the transmitters $\mathbf{t}_i, i \in \{1, \ldots, k\}$ obtained from section 5.1. The candidate that minimizes this distance metric is returned as the new approximate localization for the UAV. The metric $m()$ is described in Algorithm 1.

Algorithm 1 Distance function m

Require: $\mathbf{v}, \mathbf{w} \in \mathbb{R}^k$
Require: $\rho \in (0, 1)$
$\mathbf{x} \leftarrow$sort$(|v_j - w_j| \; \forall j \in [1, k])$
return $\sum_{j=1}^{\lceil k\rho \rceil} x_j$

The model for estimating the distances had 4 parameters that were optimized experimentally. The distance function in Algorithm 1 makes sure that it only takes into account the distances that are similar and throws away the distances that do not match well from the prediction. Experimentally we determined that the mean error (and max error) was minimized at $\rho = 0.75$.

There are three more parameters from (Kumar et al., 2015) that we optimized for our algorithm to bring the errors down. The first was the m, c calibration parameters. We took approximately 50% of the aerial data and minimized the mean localization error, using various m, c values generated randomly in the range of $(-1, 2)$ and $(-40, -75)$ respectively. m, c were optimized over a course of 1000 experiments. This gave us calibration parameters $m = 0.634714, c = -47.175549$, which we fixed for all our experiments. This is a different method of calibration compared to what was used on the ground in (Kumar et al., 2015) and (Mukherjee et al., 2017). Another parameter that we optimized was the threshold ν that determines a DCD feature. We set it to 16dB after extensive experiments to minimize mean error, similar to the calibration parameters.

6. Experiments

We first show how the distribution of the errors in localization for the fixed calibration parameters that we determined, using the algorithms in (Kumar et al., 2015) in Figure 3. The average of these errors is 2.54 miles, which is considerably better than the average error of 4.98 miles reported in (Kumar et al., 2015) and the average error of 3.4 miles reported in Mukherjee et al (Mukherjee et al., 2017). Note that these are not fair comparisons since our experiment has only 30 spectra whereas the others have upward of 900 and 100 spectra respectively. Nevertheless, it verifies our intuition that the aerial spectra are better for localization than the ground spectra.

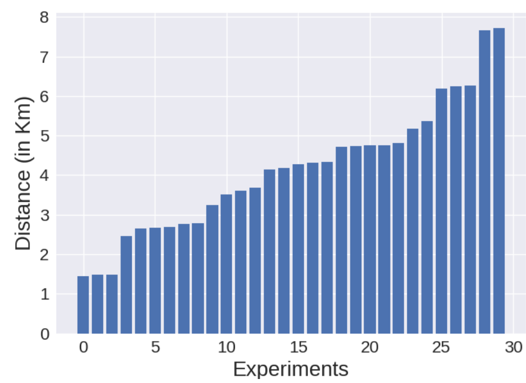

Figure 3. The error in localization for 30 aerial spectra collected in Tallahassee using the algorithm of (Kumar et al., 2015), when m, c, and ν are optimized and fixed.

Next we show the distribution of calibration errors when different m, c values are chosen for calibration in Figure 4. Note that the m and c values are between $(-1, 2)$ and $(-40, -75)$ respectively and that these bounds were determined experimentally.

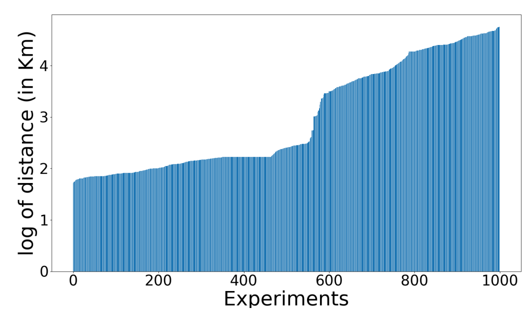

Figure 4. The error in localization for 1000 randomly generated m, c values. The value of m, c for our experiments is fixed using this experiment at $m = 0.634714$ and $c = -47.175549$.

In Table 1 we show the effectiveness of our supervised learning model for the problem of predicting the distance of a given location, from the FM transmitters, given the received power. This table was obtained by using 604 selected DCD features with the learner, with a test train split of $30/70$ respectively.

Errors (in meters)	**SVM**	**Neural Network**	**Random Forest**
Mean	8797	2246	2057
Median	5566	2022	1707
Min	109	1	0
Max	30937	8814	7075
Average $\%$ error	55.9$\%$	15.1$\%$	15.5$\%$

Table 1. Errors for distance estimation

Our neural network (Demuth et al., 2014) implementation was done using the Sequential model in Keras (Chollet et al., 2015). After optimizing for the number of layers and the number of nodes at each layer in the Sequential model, we finally settled down on the following neural network architecture $5 \Rightarrow 13 \Rightarrow 8 \Rightarrow 4 \Rightarrow 1$. The optimization was done using the Tensorflow backend (Abadi et al., 2016). The loss function was chosen to be mean squared logarithmic error (Bishop, 2006) and the optimization algorithm was rmsprop (Tieleman & Hinton, 2012). Both are built into Keras. We chose to use random forests for all our future experiments as it gave better results compared to neural networks on two metrics: 1) Time for model computation and 2) Median error. Both these metrics are important for our algorithm.

Next we show the errors of our algorithm and compare it with the results obtained from using the algorithm in (Kumar et al., 2015). For this experiment we partition the 30 aerial data points into test train splits. We start with 15 data points in the training set and go up to 29 data points. For each training set, we did 5 experiments, and calculated the localization error using algorithms from (Kumar et al., 2015), and then improved the results using our current algorithm. The average improvements are shown in figure 5.

7. Conclusion & Future Work

In this work we have described a method for reducing the localization error for a passive localization system using broadcast FM transmission, as compared to the results reported in (Kumar et al., 2015; Mukherjee et al., 2017). Our algorithms scale to the entire continental United States. However, even though the minimum possible error is around 172 meters in air, the average error reported by us is still around 3000 meters. One of the reasons for this error is the fact that we are using a low cost RTL-SDR dongle for collecting the RSSI data and this introduces errors in the measurements. In order to improve the accuracy of the

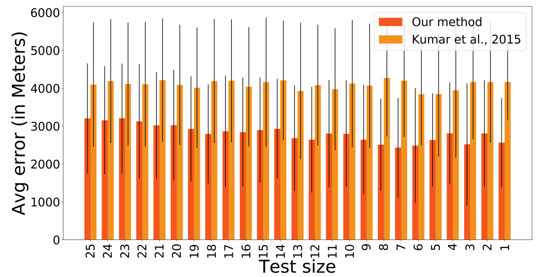

Figure 5. The improvements made by using our current algorithm compared to directly feeding aerial spectra in (Kumar et al., 2015)'s algorithm (after optimizing its parameters). The training size for these experiments is 30 minus the size of the test set. As the training size increases, the difference between the mean errors reported tends to increase (training size = 5, %-age improvement = 21.7%, whereas when training size = 29, %-age improvement jumps to 38.5%.)

positioning system we need to use both properly calibrated and high quality receivers for measuring the RSSI. To this end we plan to use Ettus USRP B210 software defined radio as our receiver for experiments in the near future along with better FM antennas. Increasing the total bandwidth processed for localization is another avenue that we plan to explore for getting better accuracies.

8. Acknowledgments

This work was funded by the Air Force Research Labs, Munitions Directorate, Eglin Air Force Base, Shalimar, FL.

We would like to thank Mr. Daniel Rodriguez from CompGeom Inc. and Ms. Mallory Haulcomb from Intelligent Robotics Inc. for assembling the hardware platform used for the experiments. Special thanks to Mr. Daniel Rodriguez for helping develop and pilot the autonomous UAV that collected the aerial data.

References

Gps is a time bomb! `http://www.locata.com/article/gps-is-a-time-bomb/`, 2016. Accessed: 2016-09-15.

Abadi, Martín, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S, Davis, Andy, Dean, Jeffrey, Devin, Matthieu, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

Abdelnasser, H., Mohamed, R., Elgohary, A., Farid, M., Wang, H., Sen, S., Choudhury, R.R., and Youssef, M. Semanticslam: Using environment landmarks for unsupervised indoor localization. *Mobile Computing, IEEE Transactions on*, PP(99):1–1, 2015. ISSN 1536-1233. doi: 10.1109/TMC.2015.2478451.

Aly, Heba and Youssef, Moustafa. Dejavu: An accurate energy-efficient outdoor localization system. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'13, pp. 154–163, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2521-9. doi: 10.1145/2525314.2525338. URL `http://doi.acm.org/10.1145/2525314.2525338`.

Atia, Mohamed M, Noureldin, Aboelmagd, and Korenberg, Michael J. Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *Mobile Computing, IEEE Transactions on*, 12(9):1774–1787, 2013.

Azizyan, Martin, Constandache, Ionut, and Roy Choudhury, Romit. Surroundsense: mobile phone localization via ambience fingerprinting. In *Proceedings of the 15th annual international conference on Mobile computing and networking*, pp. 261–272. ACM, 2009.

Barnes, Joel, Rizos, Chris, Wang, Jinling, Small, David, Voigt, Gavin, and Gambale, Nunzio. Locata: A new positioning technology for high precision indoor and outdoor positioning. In *Proceedings 2003 International Symposium on GPS\ GNSS*, pp. 9–18, 2003.

Bishop, Christopher M. *Pattern recognition and machine learning*. springer, 2006.

Borkowski, Daniel G, Fung, Hingsum F, Habal, Hadi F, Chao, Kenneth, Kai, Sheng-roan, and Robert, D Packard II. Cellular network-based location system, May 21 1996. US Patent 5,519,760.

Cameron, Alan. Glonass gone ... then back. `http://gpsworld.com/glonass-gone-then-back/`, 2016. Accessed: 2016-03-15.

Charikar, Moses, Indyk, Piotr, and Panigrahy, Rina. New algorithms for subset query, partial match, orthogonal range searching, and related problems. In *Automata, Languages and Programming*, pp. 451–462. Springer, 2002.

Chen, Yin, Lymberopoulos, Dimitrios, Liu, Jie, and Priyantha, Bodhi. Fm-based indoor localization. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pp. 169–182, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1301-8. doi: 10.1145/2307636.2307653. URL `http://doi.acm.org/10.1145/2307636.2307653`.

Chen, Yin, Lymberopoulos, Dimitrios, Liu, Jie, and Priyantha, Bodhi. Indoor localization using fm signals. *Mobile Computing, IEEE Transactions on*, 12(8):1502–1517, 2013.

Chollet, François et al. Keras. `https://github.com/fchollet/keras`, 2015.

Chu, Hung-Chi and Jan, Rong-Hong. A gps-less, outdoor, self-positioning method for wireless sensor networks. *Ad Hoc Networks*, 5(5):547–557, 2007.

Cong, Li and Zhuang, Weihua. Non-line-of-sight error mitigation in tdoa mobile location. In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 1, pp. 680–684 vol.1, 2001. doi: 10.1109/GLOCOM.2001.965202.

Counselman III, Charles C and Hall, Timothy D. Instantaneous radiopositioning using signals of opportunity, December 10 2002. US Patent 6,492,945.

De Berg, Mark, Van Kreveld, Marc, Overmars, Mark, and Schwarzkopf, Otfried Cheong. Computational geometry. In *Computational geometry*, pp. 1–17. Springer, 2000.

Demuth, Howard B, Beale, Mark H, De Jess, Orlando, and Hagan, Martin T. *Neural network design*. Martin Hagan, 2014.

Djuknic, Goran M and Richton, Robert E. Geolocation and assisted gps. *Computer*, 34(2):123–125, 2001.

Engelbrecht, Lloyd and Weinberg, Aaron. Location determination system and method using television broadcast signals, April 23 1996. US Patent 5,510,801.

Fang, Shih-Hau, Chen, Jen-Chian, Huang, Hao-Ru, and Lin, Tsung-Nan. Is fm a rf-based positioning solution in a metropolitan-scale environment? a probabilistic approach with radio measurements analysis. *Broadcasting, IEEE Transactions on*, 55(3):577–588, Sept 2009. ISSN 0018-9316. doi: 10.1109/TBC.2009.2025456.

Farrell, Jay A, Givargis, Tony D, and Barth, Matthew J. Real-time differential carrier phase gps-aided ins. *Control Systems Technology, IEEE Transactions on*, 8(4): 709–721, 2000.

Fox, Anthony, Eichelberger, Chris, Hughes, John, and Lyon, Skylar. Spatio-temporal indexing in non-relational distributed databases. In *Big Data, 2013 IEEE International Conference on*, pp. 291–299. IEEE, 2013.

Fuller, Richard. Tutorial on location determination by rf means. In *Mobile Entity Localization and Tracking in GPS-less Environnments*, pp. 213–234. Springer, 2009.

Gustafsson, Fredrik and Gunnarsson, Fredrik. Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *Signal Processing Magazine, IEEE*, 22 (4):41–53, 2005.

Haeberlen, Andreas, Flannery, Eliot, Ladd, Andrew M, Rudys, Algis, Wallach, Dan S, and Kavraki, Lydia E. Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pp. 70–84. ACM, 2004.

Ibrahim, Mohamed and Youssef, Moustafa. Cellsense: An accurate energy-efficient gsm positioning system. *Vehicular Technology, IEEE Transactions on*, 61(1):286–296, 2012.

Kaplan, E.D. *Understanding GPS: Principles and Applications*. Artech House telecommunications library. Artech House, 1996. ISBN 9780890067932.

Khattab, Ahmed, Fahmy, Yasmine A, and Wahab, Ahmed Abdel. High accuracy gps-free vehicle localization framework via an ins-assisted single rsu. *International Journal of Distributed Sensor Networks*, 2015:71, 2015.

Kjærgaard, Mikkel Baun. A taxonomy for radio location fingerprinting. In *International Symposium on Location- and Context-Awareness*, pp. 139–156. Springer, 2007.

Konstantinidis, Andreas, Chatzimilioudis, Georgios, Zeinalipour-Yazti, Demetrios, Mpeis, Paschalis, Pelekis, Nikos, and Theodoridis, Yannis. Privacy-preserving indoor localization on smartphones. *Knowledge and Data Engineering, IEEE Transactions on*, 27(11): 3042–3055, 2015.

Krumm, John, Cermak, Gerry, and Horvitz, Eric. Rightspot: A novel sense of location for a smart personal object. In *UbiComp 2003: Ubiquitous Computing*, pp. 36–43. Springer, 2003.

Kumar, Piyush, Mukherjee, Tathagata, Pasiliao, Eduardo, and Xu, Liqin. Cheap approximate localization using fm radio. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 77. ACM, 2015.

LaMance, Jimmy, DeSalas, Javier, and Jarvinen, Jani. Assisted gps: a low-infrastructure approach. *GPS World*, 13(3), 2002.

Laoudias, Christos, Constantinou, George, Constantinides, Marios, Nicolaou, Silouanos, Zeinalipour-Yazti, Demetrios, and Panayiotou, Christos G. The airplace indoor positioning platform for android smartphones. In *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*, pp. 312–315. IEEE, 2012.

Li, Fan, Zhao, Chunshui, Ding, Guanzhong, Gong, Jian, Liu, Chenxing, and Zhao, Feng. A reliable and accurate indoor localization method using phone inertial sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 421–430. ACM, 2012.

Liaw, Andy, Wiener, Matthew, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

Mao, Guoqiang, Fidan, Barış, and Anderson, Brian DO. Wireless sensor network localization techniques. *Computer networks*, 51(10):2529–2553, 2007.

Martin, Eladio, Vinyals, Oriol, Friedland, Gerald, and Bajcsy, Ruzena. Precise indoor localization using smart phones. In *Proceedings of the international conference on Multimedia*, pp. 787–790. ACM, 2010.

McEllroy, JA, Raquet, JF, and Temple, MA. Use of a software radio to evaluate signals of opportunity for navigation. In *Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2006)*, pp. 126–133, 2001.

Meier, Lorenz, Tanskanen, Petri, Fraundorfer, Friedrich, and Pollefeys, Marc. Pixhawk: A system for autonomous flight using onboard computer vision. In *Robotics and automation (ICRA), 2011 IEEE international conference on*, pp. 2992–2997. IEEE, 2011.

Meier, Lorenz, Camacho, J, Godbolt, B, Goppert, J, Heng, L, Lizarraga, M, et al. Mavlink: Micro air vehicle communication protocol. *Online]. Tillgänglig: http://qgroundcontrol. org/mavlink/start.[Hämtad 2014-05-22]*, 2013.

Misra, Pratap and Enge, Per. *Global Positioning System: Signals, Measurements and Performance Second Edition*. Lincoln, MA: Ganga-Jamuna Press, 2006.

Moghtadaiee, Vahideh, Dempster, Andrew G, and Lim, Samsung. Indoor localization using fm radio signals: A fingerprinting approach. In *IPIN*, pp. 1–7. Citeseer, 2011.

Mukherjee, Tathagata, Kumar, Piyush, Pati, Debdeep, Pasiliao, Eduardo, and Xu, Liqin. Large scale fm signal strength map estimation for passive approximate localization, 2017.

Ocana, M, Bergasa, LM, Sotelo, MA, Nuevo, J, and Flores, R. Indoor robot localization system using wifi signal measure and minimizing calibration effort. In *Proceedings of the IEEE International Symposium on Industrial Electronics*, volume 4, pp. 1545–1550, 2005.

Oguejiofor, O, Okorogu, V, Adewale, A, and Osuesu, B. Outdoor localization system using rssi measurement of wireless sensor network. *Proceedings of International Journal of Innovative Technology and Exploring Engineering*, pp. 1–6, 2013.

O'Rourke, Joseph. *Computational Geometry in C*. Cambridge University Press, New York, NY, USA, 2nd edition, 1998. ISBN 0521640105.

O'Shea, Timothy J and West, Nathan. Radio machine learning dataset generation with gnu radio. In *Proceedings of the GNU Radio Conference*, volume 1, 2016.

Otsason, Veljo, Varshavsky, Alex, LaMarca, Anthony, and de Lara, Eyal. Accurate gsm indoor localization. In *Proceedings of the 7th International Conference on Ubiquitous Computing*, UbiComp'05, pp. 141–158, Berlin, Heidelberg, 2005. Springer-Verlag. ISBN 978-3-540-28760-5. doi: 10.1007/11551201_9. URL `http://dx.doi.org/10.1007/11551201_9`.

OConnor, Michael Lee. *Carrier-phase differential GPS for automatic control of land vehicles*. PhD thesis, Stanford University, 1997.

Pan, Sinno Jialin and Yang, Qiang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

Parkinson, Bradford W, O'connor, Michael L, Elkaim, Gabriel H, and Bell, Thomas. Method and system for automatic control of vehicles based on carrier phase differential gps, April 18 2000. US Patent 6,052,647.

Petrou, Lambros, Larkou, George, Laoudias, Christos, Zeinalipour-Yazti, Demetrios, and Panayiotou, Christos G. Demonstration abstract: Crowdsourced indoor localization and navigation with anyplace. In *Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on*, pp. 331–332. IEEE, 2014.

Popleteev, Andrei. *Indoor positioning using FM radio signals*. PhD thesis, University of Trento, 2011.

Psiaki, Mark L and Humphreys, Todd E. Gnss spoofing and detection. *Proceedings of the IEEE*, 104(6):1258–1270, 2016.

Rappaport, Theodore S et al. *Wireless communications: principles and practice*, volume 2. Prentice Hall PTR New Jersey, 1996.

Savvides, Andreas, Han, Chih-Chieh, and Strivastava, Mani B. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pp. 166–179. ACM, 2001.

Srinivasan, Avinash and Wu, Jie. A survey on secure localization in wireless sensor networks. *Encyclopedia of Wireless and Mobile communications*, 2007.

Talbot, Nicholas C, Allison, Michael T, and Nichols, Mark E. Centimeter accurate global positioning system receiver for on-the-fly real-time kinematic measurement and control, May 21 1996. US Patent 5,519,620.

Tieleman, Tijmen and Hinton, Geoffrey. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

Vander Stoep, Jeffrey. *Design and Implementation of Reliable Localization Algorithms using Received Signal Strength*. PhD thesis, University of Washington, 2009.

Varshavsky, Alex, De Lara, Eyal, Hightower, Jeffrey, LaMarca, Anthony, and Otsason, Veljo. Gsm indoor localization. *Pervasive and Mobile Computing*, 3(6):698–720, 2007.

Waterman, S. North korean jamming of gps shows systems weakness. *Wall Street Journal*, August 2012.

Yang, Chun, Nguyen, Thao, and Blasch, Erik. Mobile positioning via fusion of mixed signals of opportunity. *Aerospace and Electronic Systems Magazine, IEEE*, 29 (4):34–46, 2014.

Yang, Zheng, Wu, Chenshu, and Liu, Yunhao. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pp. 269–280. ACM, 2012.

Zhao, Sheng, Chen, Yiming, Zhang, Haiyu, and Farrell, Jay A. Differential gps aided inertial navigation: a contemplative realtime approach. In *IFAC World Congress*, pp. 8959–8964, 2014.

Zheng, Vincent W, Cao, Hong, Gao, Shenghua, Adhikari, Aditi, Lin, Miao, and Chang, Kevin Chen-Chuan. Cold-start heterogeneous-device wireless localization. In *AAAI*, pp. 1429–1435, 2016.

Zheng, Vincent Wenchen, Pan, Sinno Jialin, Yang, Qiang, and Pan, Jeffrey Junfeng. Transferring multi-device localization models using latent multi-task learning. In *AAAI*, volume 8, pp. 1427–1432, 2008a.

Zheng, Vincent Wenchen, Xiang, Evan Wei, Yang, Qiang, and Shen, Dou. Transferring localization models over time. In *AAAI*, pp. 1421–1426, 2008b.