Hindawi Publishing Corporation
International Journal of Antennas and Propagation
Volume 2015, Article ID 453157, 11 pages
http://dx.doi.org/10.1155/2015/453157

Research Article

Source Geolocation in Urban Environments Using Multipath Fingerprinting

Ram M. Narayanan,[1] Brian R. Phelan,[1] and Erik H. Lenzing[2]

[1]*Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802, USA*
[2]*The Applied Research Laboratory, State College, PA 16803, USA*

Correspondence should be addressed to Ram M. Narayanan; ram@ee.psu.edu

Received 16 December 2014; Revised 18 March 2015; Accepted 18 March 2015

Academic Editor: Ana Alejos

A method for determining the location of Global Systems for Mobile Communications (GSM) mobile transmitters is proposed. Our approach estimates the location of a source without the use of multilateration or Line-of-Sight (LOS) techniques. A Multipath Characteristic Database (MCD) containing the multipath feature vectors, for each possible transmitter location within an area of interest, is populated via ray-tracing software simulations. The multipath characteristics of interest are angle-of-arrival (AOA) (azimuth) and time-of-arrival (TOA). By minimizing the "distance" between estimated and simulated multipath feature vectors, an estimate for the actual source location can be obtained. The development of the estimation method is presented, followed by a detailed analysis of its estimation accuracy. Since the proposed method utilizes a simulated multipath signature database based upon the knowledge of the environment and the terrain, the need for *a priori* soundings from the area of interest is eliminated, thus making this location estimation system suitable for application in denied territories. Location accuracies compare favorably with the requirements for the location of wireless 9-1-1 callers as recommended by the Federal Communications Commission (FCC).

1. Introduction

Spatial localization of cellular emitters in dense urban environments can provide a valuable tool for a variety of users such as emergency services, law enforcement, and military personnel. Navigation, social media, and location-dependent searching would also benefit from advances in localization techniques. The ability to use Non-Line-of-Sight (NLOS) and nonmultilateration techniques would allow the end user to perform the localization from an arbitrary position anywhere within a given range of the target.

The concept of using the multipath fingerprint composed of time- and angle-of-arrival data for wireless location finding in urban environments was developed using electromagnetic ray-tracing techniques and validated using computer-aided design (CAD) models of a real city [1]. The fundamental premise of this approach is to extract the features of the multipath signals to create a unique fingerprint, such as angle-of-arrival, time delay, and signal strength, which is then compared to a database of known fingerprints, each corresponding to a known location. A matching fingerprint found in the database provides an estimate for the correct location of the transmitter. A fingerprinting technique using the channel's impulse response information combined with an artificial neural network was developed for geolocation in mines or other confined environments with rough sidewall surfaces with a location accuracy of 2 m [2]. A fingerprinting technique exploiting the spatial-temporal characteristics of the multipath signals received by the base station antenna array was proposed in [3]. The spatial-temporal fingerprint was based on a lower dimensional subspace of the spatial-temporal covariance matrix capturing the AOAs and the differential delays of the dominant multipath reflections. Localization accuracies of about 1 m were achieved in typical indoor environments. Similar fingerprinting techniques were also proposed and developed for emergency location services [4], Global Navigation Satellite System (GNSS) indoor positioning [5], indoor geolocation for ultrawideband (UWB) systems [6], and channel estimation in multipath-rich mobile communication scenarios [7].

The method discussed here compares simulated and measured signal characteristics based only upon multipath propagation to provide an estimate for the location of a GSM emitter. Our approach is different from the one adopted in [1] in that while their base station is located solely within their area of interest, ours may be located both inside and outside of the area of interest, the latter generally occurring in denied environments. Having the base station in the center yields higher variability in the AOA. For base stations located within the area of interest, AOAs are distributed over a 360-degree spread. However, for base stations located outside the area of interest, AOAs are mostly restricted to be within a 180-degree spread.

The Federal Communications Commission (FCC) recently issued its Third Further Notice of Proposed Rulemaking and proposed the following specific measures in their E911 location accuracy rules to ensure accurate indoor location information [8]. According to their guidelines, location accuracies must be within 100 meters for 67 percent of calls and 300 meters for 90 percent of calls for network-based technologies and within 50 meters for 67 percent of calls and 150 meters for 90 percent of calls for handset-based technologies. Network-based technologies have less stringent requirements compared to handset-based technologies. Our approach addresses the network-based case and, therefore, aims to satisfy the first set of requirements stated above. In addition, our approach avoids the use of received signal strength (RSS) as these are dependent on various unknown factors, such as transmit power, reflectivity of walls, and propagation characteristics, which do not largely impact TOA and AOA.

Through the use of ray-tracing software, a MCD is populated via simulations, which contains a feature vector for possible transmitter locations within an area of interest. The MCD is then used to find possible transmit locations that have similar multipath characteristics to that of the estimated parameters. The multipath characteristic estimators used here are joint angle delay estimation-multiple signal classification (JADE-MUSIC) and JADE Estimation of Signal Parameters via Rotational Invariance Techniques (JADE-ESPRIT). A K-Weighted Nearest Neighbor (KWNN) distance metric between the estimated and simulated feature vectors is used to determine the final geolocation estimate. An analysis on the geolocation algorithm's performance is presented which shows the effect of SNR, receiver location, oversampling rate, and number of antenna elements. The computational complexity of the JADE-MUSIC and JADE-ESPRIT is also analyzed.

Although GSM telephony was used in the analysis presented in subsequent sections, our proposed estimation techniques can be applied to all narrowband Time Division Multiple Access (TDMA) systems (which operate in multipath-rich environments). This paper discusses the results of an investigation of our proposed localization technique and extends the work previously reported by us in [9]. New results reported here include geolocation estimates derived from the JADE-ESPRIT algorithm, as well a comparison between estimates using a receiver located inside and outside the area of interest. A summary of the implemented algorithms is included, as well as result comparisons for various SNRs, oversampling rates, and number of antenna elements used.

This paper is organized as follows. Section 2 provides the linear approximation of GSM signals used in the estimation portion of the localization technique. Section 3 provides an introduction to the JADE estimation techniques. Section 4 discusses the creation of the MCD via ray-tracing software. Section 5 discusses the geolocation fingerprint matching technique. An analysis of the performance of the geolocation estimator is given in Section 6. Finally, Section 7 contains a summary and future work for the development of the proposed method.

2. Linear Approximation of Gaussian Minimum-Shift Keying (GMSK)

GSM systems utilize a GMSK modulation scheme [10]. A method for representing digital phase modulations via superposition of amplitude modulated pulses was introduced by Laurent in [11]; this method was further investigated for signals with modulation index 1/2 in [12] and for GMSK in [13]. Wiesler et al. show that a GMSK signal with the GSM parameters (BT = 0.3, $L = 4$) can be decomposed into linear and nonlinear parts $s^{\text{lin}}(t) + s^{\text{nlin}}(t)$ [13]:

$$\begin{aligned} s(t) &= \sum_{n=0}^{\infty} \exp\left[\frac{j\pi}{2}\sum_{i=0}^{n} a_i\right] C_0(t - nT) \\ &\quad + \sum_{n=0}^{\infty}\sum_{K=1}^{7} e^{j\pi A_{K,n}} C_K(t - nT) \\ &= s^{\text{lin}}(t) + s^{\text{nlin}}(t), \end{aligned} \tag{1}$$

where

$$C_0(t) = \begin{cases} S(t)\prod_{l=1}^{4} S(t + l \cdot T), & 0 \le t \le 5T, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

$S(t)$ is defined as

$$S(t) = \begin{cases} \sin\left[\pi\int_0^t h_{\text{gauss}}(\tau)\, d\tau\right], & 0 \le t \le 4T, \\ \sin\left[\left(\frac{\pi}{2}\right) - \pi\int_0^t h_{\text{gauss}}(\tau)\, d\tau\right], & 4T \le t \le 8T, \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

and a_i is the NRZ data stream. Since C_K and $A_{K,n}$ (for $K \ge 1$) are only found in the s^{nlin}, we have omitted their definitions here. General expressions for C_K and $A_{K,n}$ are given in equations 11 and 13 of [11], respectively. Wiesler et al. show that approximating $s(t)$ solely by its linear term results in a BER performance as good as or (for good SNR) even better than the exact GMSK representation. We further simplify our representation of the GMSK modulated signal by setting the

exponential term in s^{lin} of (1) equal to z_n. Therefore, our linear approximation yields

$$s(t) \approx s^{\text{lin}}(t) = \sum_{n=0}^{\infty} z_n C_0 (t - nT), \tag{4}$$

where C_0 can be considered to represent the pulse shaping modulation function of a linear modulation scheme. This result is used in JADE techniques presented in subsequent sections.

3. Joint Angle Delay Estimation (JADE)

The coherence time of the (Rayleigh) fading in the mobile channel is roughly given by $t_{\text{coh}} = c/(v \cdot f_c)$ [14]. This yields a coherence time of 160 ms for a mobile moving at a walking speed and 5.6 ms for a mobile moving at a highway speed. The fading can, therefore, be modeled as time-invariant for a single GSM frame at highway speeds and for up to 30 frames for walking speeds. Subsequently, for a transmitter moving at highway speeds or lower, the wireless channel has path fadings that are constant over multiple GSM time slots. An adaptation to the JADE techniques discussed here is presented in [15], which considers channels with path fadings varying within the duration of a time slot. The angles and delays of the received multipaths vary much slower than the mobile channel fading; therefore, we model these as time-invariant over many GSM frames [14].

Vanderveen et al. have developed JADE techniques based on the above assumptions. The q-multipath channel model for a P-element antenna array takes the following form (see [14, 16] for a complete analysis):

$$\begin{aligned}\mathbf{H} &= \left[\mathbf{a}(\theta_1) \cdots \mathbf{a}\left(\theta_q\right)\right] \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_q \end{bmatrix} \begin{bmatrix} \mathbf{g}^T(\tau_1) \\ \vdots \\ \mathbf{g}^T\left(\tau_q\right) \end{bmatrix} \\ &= \mathbf{A}(\boldsymbol{\theta}) \operatorname{diag}[\boldsymbol{\beta}] \mathbf{G}^T(\boldsymbol{\tau}),\end{aligned} \tag{5}$$

where $\mathbf{a}(\theta_i)$ is the steering vector of the ith multipath and $\mathbf{A}(\boldsymbol{\theta})$ is the $P \times q$ steering matrix. The diagonal matrix of β values denotes the complex attenuations associated with each multipath, $\mathbf{g}(\tau_i)$ denotes the ith delayed pulse shaping modulation function, as shown in (2), and $\mathbf{G}^T(\boldsymbol{\tau})$ is a $q \times LO$ matrix containing the q time-delayed pulse shaping functions. L defines the length of the channel impulse response in symbol periods, and O is the oversampling factor.

Unlike GSM's modulation scheme which can be decomposed into both linear and nonlinear parts, the pulse-shaping modulation function is associated with an entirely linear modulation scheme. The aforementioned linear approximation of the GSM signal was used to accommodate this requirement.

3.1. JADE-MUSIC Algorithm. Vanderveen et al. eventually arrive at a vectorized noisy channel estimate of the following form [16]:

$$\begin{aligned}\operatorname{vec}\left(\mathbf{H}_{\text{est}}^{(n)}\right) &= \mathbf{y} \\ &= \left[\mathbf{G}(\boldsymbol{\tau}) \circ \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\beta}^{(n)} + \mathbf{v}^{(n)} \\ &= \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau}) \boldsymbol{\beta}^{(n)} + \mathbf{v}^{(n)}, \quad n = 1, \ldots, N,\end{aligned} \tag{6}$$

where $\circ$ denotes a column-wise Kronecker product, $\mathbf{v}^{(n)}$ denotes the estimate noise, and the superscript $\cdot^{(n)}$ denotes the nth channel estimate. Here $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})$ is time-invariant over N channel estimates, where N is determined based on the stationarity of the AOAs and TOAs for a given multipath scenario. The channel estimation rate is determined by the temporal coherence of the channel path fadings. The channel should be estimated once over the coherence time of the channel path fadings (i.e., the channel can be estimated on the order of a frame for a mobile moving at highway speeds and once every 30 frames for a mobile moving at a walking speed). After estimating N channels, (6) can be represented in matrix form as

$$\mathbf{Y} = \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau}) \mathbf{B} + \mathbf{V}, \tag{7}$$

where $\mathbf{B} = [\boldsymbol{\beta}^{(1)} \cdots \boldsymbol{\beta}^{(N)}]$ and $\mathbf{Y}$ and $\mathbf{V}$ have similar form. A 2-dimensional MUSIC algorithm can now be applied to find $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$. The MUSIC algorithm will result in a joint delay and angle estimation under the following restrictions.

(1) The number of multipaths, q, must be less than $P \cdot L \cdot O$.

(2) The number of channel estimates used must be greater than the number of multipaths.

(3) The collection time must be longer than the coherence time (for fadings) of the mobile channel.

A significant consequence of these restrictions is that the number of antenna elements needed can be reduced to less than the number of multipaths, if an adequate number of samples are obtained.

3.2. JADE-ESPRIT Algorithm. The JADE-ESPRIT algorithm manipulates the form of the channel estimate matrix so that a 2D ESPRIT-like algorithm can be applied. First, the discrete Fourier transform is applied to the $\mathbf{H}$ in (5). The pulse shaping function is then deconvolved from the Fourier-transformed channel estimate (see [16] for details), whereby the channel now satisfies the following model:

$$\begin{aligned}\check{\mathbf{H}} &= \left[\mathbf{a}(\psi_1) \cdots \mathbf{a}\left(\psi_q\right)\right] \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_q \end{bmatrix} \begin{bmatrix} \mathbf{f}^T(\phi_1) \\ \vdots \\ \mathbf{f}^T\left(\phi_q\right) \end{bmatrix} \\ &= \mathbf{A}(\boldsymbol{\psi}) \operatorname{diag}[\boldsymbol{\beta}] \mathbf{F}^T(\boldsymbol{\phi}),\end{aligned} \tag{8}$$

where $\psi_i = \exp(-j2\pi(d/\lambda)\sin(\theta_i))$, $\phi_i = \exp(-j(2\pi/L)\tau_i)$ and $\mathbf{a}^T(\psi_i) = [1\psi_i^1 \cdots \psi_i^{P-1}]$, and $\mathbf{f}(\phi_i) = [1\phi_i^1 \cdots \phi_i^{LO-1}]$.

Applying a similar vectorization and stacking operation as in the JADE-MUSIC algorithm, the vectorized noisy channel estimate becomes

$$\check{\mathbf{Y}} = \check{\mathbf{U}}(\boldsymbol{\psi}, \boldsymbol{\phi})\,\mathbf{B} + \check{\mathbf{V}}, \tag{9}$$

where $\check{\mathbf{U}}(\boldsymbol{\psi}, \boldsymbol{\phi}) = \mathbf{F}(\boldsymbol{\phi}) \circ \mathbf{A}(\boldsymbol{\psi})$. Next, a basis of the column span of $\mathbf{U}(\boldsymbol{\psi}, \boldsymbol{\phi})$ is estimated via the q left singular vectors which correspond to the largest values of covariance matrix, $\mathbf{R}_{\check{\mathbf{Y}}}$. A 2D ESPRIT algorithm is now applied and estimates for $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ can be obtained.

The restrictions of the JADE-ESPRIT algorithm are slightly stricter than those of the JADE-MUSIC algorithm, in addition to the aforementioned JADE-MUSIC restrictions. The number of multipaths must be less than $\min(P(LW - 1), (P - 1)LW)$, where W is a factor used in the DFT (taken as 1 here).

4. Multipath Characteristic Database Analysis

Remcom's Wireless InSite ray-tracing software was used to populate the MCD, which contains location-based multipath feature vectors to be used as a basis for a fingerprint matching algorithm. A grid of transmitters was placed over the area of interest and then various multipath characteristics of the received signal are collected for each transmitter location. For the analysis here, only the AOA (azimuth) and TOA were considered for utmost simplicity and for minimizing processing and time requirements; however, the multipath feature vector could be expanded to contain (a) number of received multipaths; (b) AOA (azimuth and elevation) of each multipath; (c) TOA of each multipath; (d) received signal strength (RSS) of each multipath; (e) delay spread amongst rays; and (f) mean and standard deviation of all aforementioned multipath characteristics. Although adding additional features may doubtlessly increase the location accuracy, this would entail higher latency owing to excessive processing requirements.

A model of downtown State College, Pennsylvania, was used in the ray-tracing simulation. The model consisted of concrete buildings placed on top of the appropriate elevation (i.e., landscape). Transmitters were spaced in 3.5-meter increments along the x- and y-axes at a height of 2 meters (with respect to the ground). The grid covered a 1 km^2 area and consisted of 81,796 transmitters. The receiver was placed at two different locations represented by Scenario A located outside the area of interest and Scenario B located within the area of interest. For Scenario A, the receiver was placed on top of the Applied Research Laboratory building of The Pennsylvania State University, at a height of 10 meters (approximately 2 meters above the building height). For Scenario B, the receiver was placed on top of the HUB-Robeson Center of The Pennsylvania State University, at a height of 12 meters (approximately 2 meters above the building height). Figure 1 shows the 2D representation of the Wireless InSite model overlaid on a satellite image of State College, PA. The transmit antennas were dipoles aligned perpendicular to the ground, and isotropic antennas were used in the receiver antenna array.

Figure 1: 2D representation of the Wireless InSite model overlaid on satellite image of State College, PA. The transmitter grid covers the blue overlay. Rx1 corresponds to the receiver location of Scenario A, and Rx2 corresponds to the receiver location of Scenario B, © 2013 Google Earth.

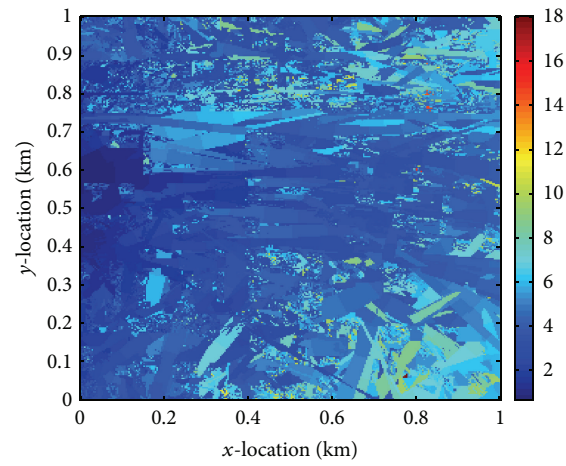

Figure 2: Simulated TOA (μs) data from the MCD for downtown State College, PA, for Scenario A. Data plotted for the dominant multipath.

A 3D ray-tracing model was used in the creation of the MCD. The ray-tracing model considered building reflections, diffractions, and ground interactions. The model was capable of analyzing LOS, singly diffracted, and doubly diffracted paths. The building faces were chosen to be one-sided, which only allow for transmission from inside the building to outside. This choice was made because the number of transmissions greatly increases the computation time, and multipaths which underwent attenuation through 2 wall transmissions would be significantly weaker than multipaths which did not undergo wall transmissions. The maximum number of reflections allowed was 6, and the maximum number of diffractions allowed was 2; this allows for rays between the receiver and transmitters which are blocked by the height of a building. A simulated GSM signal with center frequency, f_c, of 1.91 GHz was used.

Figures 2–4 display the diversity of the MCD data obtained via our simulations for Scenario A. Figure 2 shows the TOA of the dominant multipath for each transmitter location. The TOA generally increases as the distance between the receiver and the transmitters increases. Some outliers exist, which represent multipaths which underwent multiple

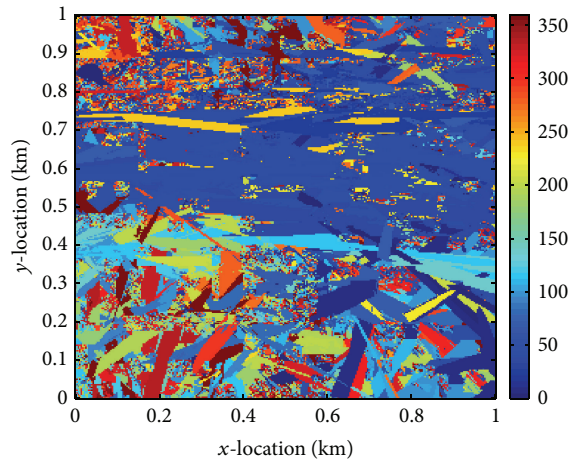

Figure 3: Simulated AOA in azimuth plane (°) data from the MCD for downtown State College, PA, for Scenario A. Data plotted for the dominant multipath.

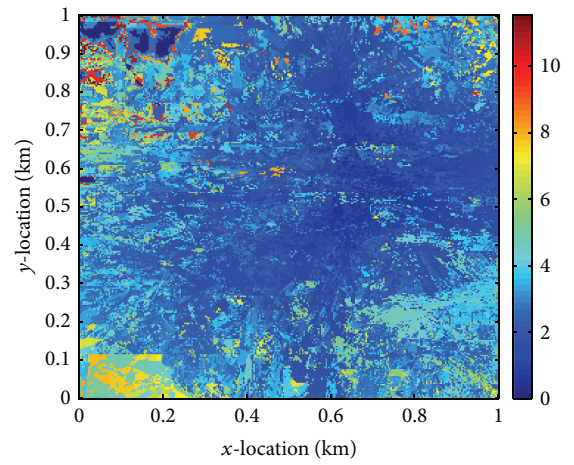

Figure 5: Simulated TOA (μs) data from the MCD for downtown State College, PA, for Scenario B. Data plotted for the dominant multipath.

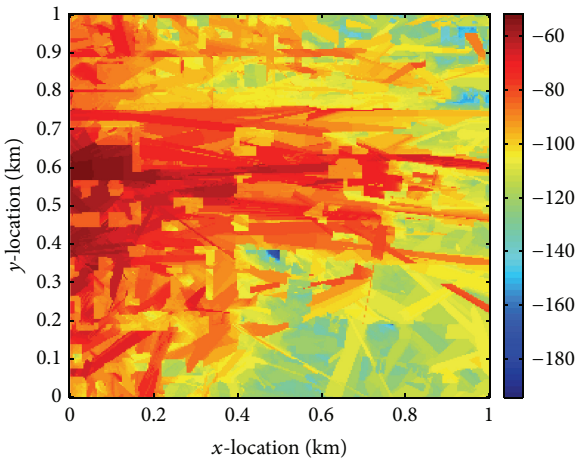

Figure 4: Simulated received power (dBm) data from the MCD for downtown State College, PA, for Scenario A. Data plotted for the dominant multipath.

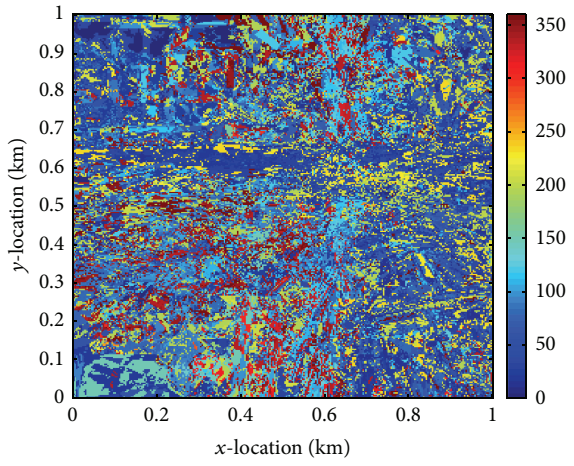

Figure 6: Simulated AOA in azimuth plane (°) data from the MCD for downtown State College, PA, for Scenario B. Data plotted for the dominant multipath.

reflections. Figure 3 shows the AOA in the azimuth plane for the dominant multipath of each transmitter. The AOA (azimuth) is influenced more by the surrounding environment of each transmitter than the distance between transmitter and receiver. Figure 4 displays the received power of the dominant multipath for each transmitter location. Figure 4 shows a strong (negative) correlation between the received signal strength and the distance between the transmitter and receiver, as expected. The properties for each multipath characteristic can be exploited to produce optimal clustering schemes.

Figures 5–7 show similar plots for Scenario B. Figure 5 shows the TOA, Figure 6 the AOA, and Figure 7 the received power. The main differences between these plots and those for Scenario A are the increase in the AOA variation over short distances, and the decrease in the spread of TOA values. Since the transmitter is placed inside the area of interest, the dominant multipaths from adjacent transmitters are less likely to follow similar paths, due to a lower number of obstructions between the transmitter and receiver. This is evident when comparing Figure 3 with Figure 6, wherein the former contains large areas where the dominant paths (from neighboring transmitters) arrive from similar directions, while the latter has greater variation in AOA over short distances. Scenario A is better suited for our geolocation algorithm, as an error in the AOA estimate may result in a geolocation estimate of a transmitter that is close to the actual transmit location. Furthermore, the TOA variability between neighboring transmitters also appears to be lower for Scenario A than Scenario B. Although it is not always possible, Scenario A is preferred.

Depending on the size of the area of interest and spatial density of the transmitters, the MCD can contain an enormous amount of information. One way to reduce

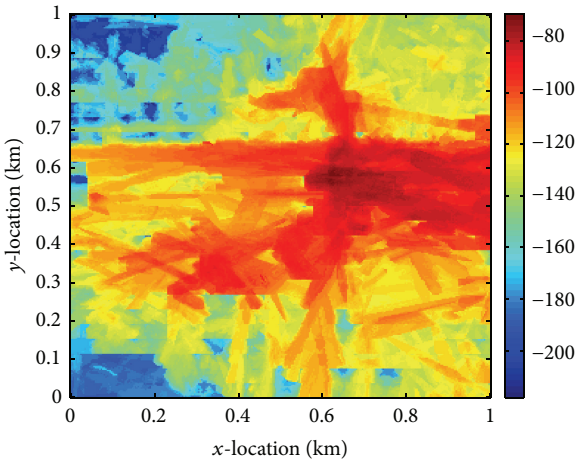

FIGURE 7: Simulated received power (dBm) data from the MCD for downtown State College, PA, for Scenario B. Data plotted for the dominant multipath.

search redundancy and thus improve the location estimation speed is to perform a clustering of the MCD. Owing to the variability in each multipath characteristic over the area of interest, the clustering algorithm should be able to detect clusters of arbitrary shape. If the receiver location is known in advance, the application of clustering techniques on the MCD can be performed prior to taking field measurements. Also, clustering schemes may be more useful than others in different scenarios. For example, if an approximate location prediction is urgently required, then a clustering scheme containing a large amount of data points in each cluster may be more useful. For these reasons, a partitioning clustering algorithm appears to be a good choice to process the MCD, because it can provide 1 to K (number of data points in MCD) different clustering configurations, even though it has the highest computational complexity among the predominant clustering techniques. A more thorough analysis of the clustering of the MCD is given in [17], but it has not been applied to the final geolocation algorithm yet.

5. Geolocation Algorithm

The method presented here is a fingerprint matching technique, in which a comparison between measured multipath feature vectors and the feature vectors in the reference database is used to determine the best estimate for the location of a transmitter. The fingerprinting localization technique can be decomposed into two phases described below [18].

(1) Calibration. Measured (or simulated) reference multipath feature vectors are collected for possible transmit locations within the area of interest. Each feature vector is stored, along with its corresponding location.

(2) Run-Time. The multipath features for a transmitter at an unknown location are estimated via the aforementioned estimation techniques. A distance metric is then calculated between the unknown transmitter's feature vector and the reference feature vectors found in the calibration phase. The minimum values of the distance metric can then be used to provide estimates for the transmitter's location.

Traditionally, fingerprinting techniques only considered RSS features and real-world measurements were taken to construct the reference database [19]. Such an approach was limited for use in indoor or in small areas, because obtaining the reference multipath measurements would be impractical for a large area. Since the method presented here uses ray-tracing software to generate its reference database (the MCD), a larger area of operation is possible. Furthermore, our method allows for the localization of a mobile emitter found in a denied territory or other areas where reference measurements cannot be obtained.

Since the earlier versions of the technique only relied on RSS, multiple stationary access points (or base stations) were needed [18, 19]. Usually the access points used were preexisting in a wireless network. The RSS from transmitters positioned in reference locations were collected for each access point. A test transmitter then emits from an unknown location, and the RSS is once again collected for each access point. A distance metric is then calculated; the minimization of this metric yields the location estimate.

Reference [18] introduced the use of ray-tracing software and propagation models to construct the RSS reference database, which induced a slight increase in location estimate errors compared to the same approach with measured RSS reference data. This greatly reduced the complexity of the calibration phase and allowed the technique to be used in larger areas.

References [20, 21] implemented fingerprinting techniques which consider more multipath characteristics than those found in [18, 19]. By using the channel impulse response (CIR), these techniques essentially utilize all multipath features, except for the phase of the complex attenuation, which varies greatly over small distances due to small-scale fading. These techniques are similar to the one presented here, except that our technique performs a distance metric on each of the estimated multipath parameters.

5.1. Summary of Implemented Algorithm

Initialize Constants

(i) Set SNR values to implement.

(ii) Set oversampling rate.

(iii) Set distance-threshold.

Load MCD

(i) Load 4-dimensional MCD.

(ii) Determine size of MCD dimensions.

(iii) Determine random location indices based on number of iterations.

(iv) Determine weights to normalize TOA and AOA calculations.

(a) The weights allow for appropriate comparison of TOA and AOA distances, as their respective domains vary over different values.

(b) The TOA weight is the maximum TOA in the MCD.

(c) The AOA weight is 180°.

For SNR Values

For n Iterations

(i) Select random location of iteration n.

(ii) Extract AOAs and TOAs from nth location:

(a) if no multipaths exist, select another (unused) random location.

Ignore Multipaths That Are Too Closely Spaced in the AOA/TOA Domain

(i) Ignore only one of the pairs that are closely spaced.

(a) The distance-threshold is a user-defined parameter which can be set to determine which multipaths should be ignored.

(b) These could be an artifact of the ray-tracing technique which would most likely produce a single signal.

(c) They diminish the performance of the JADE-MUSIC and JADE-ESPRIT algorithms.

Create Received Signals Based on Multipath and Antenna Array Parameters

(i) Create antenna array matrix.

(ii) Create pulse shaping matrix.

(iii) Create attenuation matrix.

(iv) Compute channel estimates based on JADE-MUSIC and JADE-ESPRIT algorithms.

(v) Add white Gaussian noise with power calculated based on signal power and current SNR.

Perform MUSIC and ESPRIT Algorithms

(i) The MUSIC and ESPRIT Algorithms are described in Section 3.

Calculate the Distance Metric

(i) The distance-metric is described in Section 5.2.

Record Performance Results for Each SNR and Iteration

(i) Determine radius between each estimate and actual location.

(ii) Record number of "hits" for radii of interest.

5.2. Distance Metric. Multiple distance metrics have been used in fingerprinting localization techniques. Each has advantages in different multipath scenarios. Utilization of these metrics is known as instance-based localization as they do not rely on previous location estimates. Intuitively, one may assume that previous location estimates may aid in the estimation of a transmitter in close proximity to a previous estimate. However, Qiu and Kennedy found that implementing a Support Vector Machine (SVM), which is not an instance-based technique, but rather a feed-forward network, did not improve location estimation accuracy [21]. The reason that implementation of the SVM, which is known for its aptitude in solving pattern classification problems, did not improve performance is due to the "small-scale" distribution of multipath characteristics; that is, the wireless channel impulse response fluctuates rapidly over short distances.

The distance metric used here was K-Weighted Nearest Neighbors (KWNN) given as [21]

$$(\widehat{x}, \widehat{y})_{\mathrm{KWNN}} = \frac{\sum_{k=1}^{K} w_k \left(\widehat{x}_k, \widehat{y}_k\right)_{\mathrm{NN}}}{\sum_{k=1}^{K} w_k}, \tag{10}$$

where

$$\left(\widehat{x_k}, \widehat{y_k}\right)_{\mathrm{NN}} = \left(\Delta_x i_k, \Delta_y j_k\right),$$
$$(i, j) = \min_{(i,j)} \left[\sum_{m=1}^{\widehat{q}} \sum_{n=1}^{N} \left| \frac{\mathbf{f}_{m,n} - \mathrm{MCD}_{(i,j,m,n)}}{W_n} \right|^P \right]^{1/P}. \tag{11}$$

In the above equations, $P = 2$ (i.e., the Euclidean distance was measured), i and j represent the "coordinates" of the possible transmit locations, $\widehat{q}$ is the number of multipaths, and N is the number of multipath characteristics considered. $\mathbf{f}_{m,n}$ denotes the estimated feature vector which contains AOA and TOA for the analysis presented here. $\mathrm{MCD}_{(i,j,m,n)}$ denotes the feature vector obtained via the ray-tracing software. W_n is a weight given to each multipath characteristic; since the range of values differs greatly between different multipath characteristics, a weight is needed to ensure that each characteristic's contribution to the final location estimate is appropriate. W_n is user-defined and can be modified based on the geolocation scenario, that is, the expected range of values for each multipath. The weights used to produce the results found in this paper were 10.14 nS for the TOA parameter and 180° for the AOA parameter. Δ_x and Δ_y denote the spatial resolution in the x- and y-directions used when constructing the MCD, respectively. The nearest neighbor metric is the special case of KWNN where $K = 1$ and $w_k = 1$, and K-nearest neighbor metric corresponds to $w_k = 1$. Various values of K were used in the subsequent analysis, and the w_k's, which denote the weights, were varied between 1 (NN and KNN) and the inverse of the distance between measured/simulated and reference feature vectors (KWNN). The subscripted k in $(\Delta_x i_k, \Delta_y j_k)$ represents the kth calculated coordinates if the $k - 1$ coordinates were removed from the MCD.

Table 1: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 30 dB; number of antennas = 8; oversampling factor = 2; distance-threshold = 0.2, Scenario A.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	79	91	94	96	32	52	69	82
4NN	57	80	86	95	19	41	64	82
8NN	48	76	84	93	15	41	63	80
12NN	36	65	78	89	11	29	54	80
4WNN	63	89	94	96	21	43	67	84
8WNN	57	86	92	96	21	43	63	84
12WNN	49	77	91	96	16	36	60	85
	MUSIC				ESPRIT			

Table 2: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 15 dB; number of antennas = 8; oversampling factor = 2; distance-threshold = 0.2, Scenario A.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	63	76	80	85	30	51	67	83
4NN	45	67	72	84	21	43	64	83
8NN	41	62	72	79	19	38	63	83
12NN	28	59	66	74	11	26	55	78
4WNN	47	73	79	84	22	42	63	85
8WNN	45	70	78	82	20	40	64	85
12WNN	37	63	70	81	16	32	60	84
	MUSIC				ESPRIT			

Table 3: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 0 dB; number of antennas = 8; oversampling factor = 2; distance-threshold = 0.2, Scenario A.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	22	44	55	60	11	29	45	59
4NN	25	41	50	58	8	22	40	57
8NN	22	38	46	58	6	19	35	57
12NN	14	37	48	54	8	19	35	55
4WNN	26	42	51	58	8	22	40	58
8WNN	20	39	48	59	6	21	36	57
12WNN	19	37	48	53	8	22	36	57
	MUSIC				ESPRIT			

Table 4: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 30 dB; number of antennas = 8; oversampling factor = 2; distance-threshold = 0.2, Scenario B.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	75	82	83	87	31	42	52	69
4NN	44	54	63	74	15	28	45	65
8NN	33	47	54	74	13	27	41	63
12NN	17	36	43	62	8	23	36	59
4WNN	52	64	73	80	16	29	47	67
8WNN	41	58	71	80	15	27	42	65
12WNN	25	50	64	71	9	23	38	61
	MUSIC				ESPRIT			

6. Geolocation Algorithm Analysis

Several different estimation scenarios are shown in this section, followed by statistics of the geolocation algorithm. Sections 6.1–6.4 provide insight in the geolocation algorithm's performance when varying the following attributes: SNR, receiver location, oversampling rate, and number of receiver antennas.

6.1. Effect of SNR. Tables 1–3 display the number of times a location estimate was within a certain radius of the actual transmit location. An oversampling rate 2 and 2 uniform linear arrays (angled 90° apart) with 8 antennas (spaced a half wavelength apart) were used. The number of trials performed was 100. We see that the ESPRIT algorithm's overall performance is worse than that of the MUSIC algorithm; this is expected since the JADE-ESPRIT algorithm has stricter requirements compared to the JADE-MUSIC algorithm. The results show that the nearest neighbor estimate performs best, but KWNN may be beneficial in real-world experiments since the exact transmitter location may not be included in the MCD, and modeling errors may be present.

Overall, these results show that the geolocation technique is conceptually sound. Although a 85% sucess rate (at SNR = 15 dB) at a 100-meter radius may not meet the requirements of practical systems, with improvements in the technique a reasonable performance level may be achievable.

6.2. Effect of Receiver Location. Tables 4–6 show results with the same simulation parameters as Tables 1–3, except the receiver was placed on top of the HUB-Robeson Center (Scenario B). It is apparent that moving the receiver location inside the area of interest has diminished the performance of the geolocation algorithm; this is most likely a consequence of a reduction in spatial coherence in multipath parameters when the receiver is placed inside the area of interest. The AOA and TOA are constant over larger areas of transmit locations for Scenario A. Therefore, an error in AOA and TOA estimation in Scenario A may still allow for a geolocation estimate that is nearby to the actual transmit location. The success rate was 77% (at SNR = 15 dB) at a 100-meter radius for Scenario B.

6.3. Effect of Oversampling Rate. Table 7 shows the geolocation performance when the oversampling rate is set to

Table 5: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 15 dB; number of antennas = 8; oversampling factor = 2; distance-threshold = 0.2, Scenario B.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	60	72	73	77	26	38	47	66
4NN	33	46	55	67	13	23	35	58
8NN	27	40	52	66	13	21	36	57
12NN	13	31	37	57	6	19	32	55
4WNN	35	56	60	73	13	24	38	62
8WNN	31	52	59	71	14	22	37	59
12WNN	19	39	51	65	8	21	33	59
	MUSIC				ESPRIT			

Table 6: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 0 dB; number of antennas = 8; oversampling factor = 2; distance-threshold = 0.2, Scenario B.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	16	27	35	43	14	21	27	37
4NN	11	19	27	35	2	12	17	31
8NN	10	21	27	38	3	10	17	33
12NN	5	13	20	31	2	8	14	28
4WNN	11	20	27	35	2	12	18	30
8WNN	10	21	27	40	3	10	18	33
12WNN	7	14	22	33	3	8	16	31
	MUSIC				ESPRIT			

Table 7: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 15 dB; number of antennas = 8; oversampling factor = 20; distance-threshold = 0.2, Scenario A.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	71	84	87	90	30	52	71	84
4NN	56	67	73	79	23	44	62	79
8NN	48	63	69	78	19	42	61	73
12NN	35	53	62	74	14	37	52	70
4WNN	61	76	81	84	21	46	65	82
8WNN	58	70	78	84	22	44	65	77
12WNN	47	66	73	78	14	42	57	74
	MUSIC				ESPRIT			

20, compared to Table 2. An SNR of 15 dB was used; the number of antennas in each ULA was 8, and only Scenario A is considered. The geolocation accuracy generally increases with an increase in oversampling rate. However, a 300% increase in oversampling rate results only in an increase of 12.6% accuracy at a radius of 10 meters for NN-MUSIC and an increase of 6.67% accuracy for NN-ESPRIT. Thus, the small increase in estimation accuracy may not justify the increase in computation time for most applications.

Table 8: Number of "hits" within given radius to actual location for each distance metric out of 100 trials. SNR = 15 dB; number of antennas = 12; oversampling factor = 2; distance-threshold = 0.2, Scenario A.

	Radius (meters)							
	10	25	50	100	10	25	50	100
Distance metric								
NN	68	84	88	90	31	45	73	85
4NN	54	67	79	83	30	46	64	83
8NN	46	66	79	83	20	41	63	81
12NN	32	53	66	80	13	36	55	74
4WNN	57	75	83	87	31	45	70	88
8WNN	56	74	84	87	22	42	62	85
12WNN	39	69	79	87	13	39	59	82
	MUSIC				ESPRIT			

6.4. Effect of Number of Antenna Elements in ULA. Table 8 shows the geolocation performance when the number of antennas used in each ULA is set to 12, compared to Table 2. An SNR of 15 dB was used; the oversampling rate used was 2, and only Scenario A is considered. The geolocation accuracy generally increased with an increase in the number of antennas used. It is clear that the ESPRIT algorithm is more affected than the MUSIC algorithm by an increase in the number of antennas used. The MUSIC- and ESPRIT-based geolocation algorithms only increased in accuracy by 8.0% and 3.3% at a radius of 10 meters when the number of antennas used in each ULA was increased from 8 to 12, respectively. Again, as stated in Section 6.3, the slight increase in performance may not warrant the increase in computational complexity.

6.5. Analysis of Computation Time. Table 9 displays the time (in seconds) required to run the geolocation algorithm as a function of oversampling factor and the number of antennas used in each ULA. The channel was sampled 30 times and the path fadings are assumed to be nonstationary over the sample time. The computational complexity of the JADE-MUSIC algorithm is $O((P \cdot L \cdot OF)^3)$ for the eigendecomposition of the covariance matrix, plus an additional $O((P \cdot L \cdot OF)^2)$ for each point in the 2D search (or 3D for AOA in elevation as well). Whereas JADE-ESPRIT has a much lower computational complexity $O((P \cdot L \cdot OF)^3)$, at the cost of reduced performance [15], the JADE-MUSIC algorithm's computational complexity is highly dependent on its 2D search. For the entirety of the results presented in this paper, a temporal resolution of 2 (i.e., a temporal resolution equal to twice the product of the oversampling rate and bit rate) and an angular resolution of 1° were used in the JADE-MUSIC search. These numbers could be reduced to decrease the computation time of the MUSIC-based geolocation algorithm, but the multipath characteristic estimations would suffer.

Table 9: Number of seconds to compute a geolocation estimate as a function of oversampling rate and number of antennas used in each ULA.

		Oversampling factor				
		2		10	20	
Number of antennas	2	171.36	173.43	175.51	195.30	MUSIC
	4	171.55	175.06	187.43	252.90	
	6	172.60	175.12	199.45	345.65	
	8	175.80	182.08	234.34	501.09	
	2	2.11	2.08	2.42	3.23	ESPRIT
	4	2.12	2.16	2.44	3.24	
	6	2.13	2.16	2.46	3.25	
	8	2.13	2.18	2.51	3.31	

7. Summary and Further Considerations

A localization scheme that estimates location based on a comparison between estimated and simulated multipath characteristics was proposed. The main advantage of our proposed approach is that it uses only the AOA and TOA data as inputs, thereby reducing the processing overload, compared to other approaches which use additional data inputs. Although the extensive results and conclusions presented here are based solely on simulations and relate to a single scenario, they provide an understanding of the trade-offs and a good starting point for future studies. It should be emphasized that construction activity or other events that alter the multipath landscape will necessitate a resurvey of the region and/or recalculation of the entire multipath fingerprint database [4]. Furthermore, the inclusion of unexpected scatters (i.e., motor vehicles, people, animals, etc.) will need to be analyzed experimentally to determine the effect on the estimation technique and which countermeasures would be appropriate for such situations.

With an SNR of 15 dB, the location estimate was within a 25-meter radius of the actual transmit location more than 70% of the time in both Scenarios A and B when the oversampling factor was set to 2 and the ULA's had 8 antenna elements. Increases in the number of antenna elements and oversampling rate resulted in a slight improvement in performance, which leads the authors to believe that when the restrictions in Section 6.3 are satisfied, the estimation technique's performance is limited by the diversity in multipath parameters amongst each transmit location. With the inclusion of more multipath characteristics, the location estimate should improve thus making this a valuable technique for source geolocation estimation. When used in outdoor scenarios, the location estimation technique presented here has an advantage over other techniques which only include RSS comparisons, because they are more susceptible to error due to weather and other changes in the environment. Furthermore, only one base station is needed to obtain a geolocation estimate since our technique does not rely on multilateration.

Although K-weighted distance metrics performed worse than the nearest neighbor estimates, they have been included here for completeness. The K-weighted approach may be more beneficial when the algorithm is implemented in real-world experiments and the multipath model is no longer exact. The K-weighted approach should also be beneficial as the distance between simulated transmitters is reduced. A slight estimation error may move the "max" geolocation estimate to a nearby transmitter, and the average of nearest neighbor estimates may yield a better estimate to the actual transmitter location.

Further work with this technique include (but are not limited to) (a) inclusion of more multipath characteristics in the MCD, (b) utilization of clustering techniques to increase efficiency, (c) development and assessment of other distance metrics, (d) application of the technique within other rich-multipath environments, and (e) implementation of experimental validation of the simulated results.

Conflict of Interests

The authors certify that they do not have a direct financial relationship with the commercial identities mentioned in this paper that may lead to a conflict of interests for any of the authors.

References

[1] I. Y. Kelly, H. Deng, and H. Ling, "On the feasibility of the multipath fingerprint method for location finding in urban environments," *Applied Computational Electromagnetics Society Newsletter*, vol. 15, no. 3, pp. 232–237, 2000.

[2] C. Nerguizian, C. Despins, and S. E. E. Affès, "Geolocation in mines with an impulse response fingerprinting technique and neural networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 3, pp. 603–611, 2006.

[3] E. Kupershtein, M. Wax, and I. Cohen, "Single-site emitter localization via multipath fingerprinting," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 10–21, 2013.

[4] J. Israelsohn, "On the edge of geolocation," *EDN*, vol. 47, no. 5, pp. 35–42, 2002.

[5] M. Andreotti, M. Aquino, M. Woolfson, J. Walker, and T. Moore, "Signal propagation analysis and signature extraction for GNSS indoor positioning," in *Proceedings of the IEEE/ION Position, Location, and Navigation Symposium*, pp. 913–919, San Diego, Calif, USA, April 2006.

[6] G. M. R. I. Godaliyadda and H. K. Garg, "Versatile algorithms for accurate indoor geolocation," in *Proceedings of the 16th International Conference on Digital Signal Processing (DSP '09)*, pp. 1–6, Santorini-Hellas, Greece, July 2009.

[7] P. Singh and P. Sircar, "Time delays and angles of arrival estimation using known signals," *Signal, Image and Video Processing*, vol. 6, no. 2, pp. 171–178, 2012.

[8] *Third Further Notice of Proposed Rulemaking in the Matter of Wireless E911 Location Accuracy Requirements*, 2014.

[9] B. R. Phelan, E. H. Lenzing, and R. M. Narayanan, "Source localization using unique characterizations of multipath propagation in an urban environment," in *Proceedings of the IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM '12)*, pp. 189–192, June 2012.

[10] K. Murota and K. Hirade, "Gmsk modulation for digital mobile radio telephony," *IEEE Transactions on Communications*, vol. 29, no. 7, pp. 1044–1050, 1981.

[11] P. A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses," *IEEE Transactions on Communications*, vol. 34, no. 2, pp. 150–160, 1986.

[12] P. Jung, "Laurent's representation of binary digital continuous phase modulated signals with modulation index 1/2 revisited," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 221–224, 1994.

[13] A. Wiesler, R. Machauer, and F. Jondral, "Comparison of GMSK and linear approximated GMSK for use in software radio," in *Proceedings of the IEEE 5th International Symposium on Spread Spectrum Techniques and Applications*, pp. 557–560, Sun City, South Africa, September 1998.

[14] M. C. Vanderveen, A. J. van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 682–690, 1998.

[15] M. C. Vanderveen, *Estimation of parametric channel models in wireless communication networks [Ph.D. thesis]*, Stanford University, Stanford, Calif, USA, 1997.

[16] M. C. Vanderveen, A.-J. van der Veen, and A. J. Paulraj, "Joint angle and delay estimation using shift-invariance properties," *IEEE Signal Processing Letters*, vol. 4, no. 5, pp. 142–145, 1997.

[17] B. R. Phelan, *Location of GSM transmitters in an urban environment via unique multipath characterizations [Master thesis]*, The Pennsylvania State University, State College, Pa, USA, 2012.

[18] S. Cavalieri, "WLAN-based outdoor localisation using pattern matching algorithm," *International Journal of Wireless Information Networks*, vol. 14, no. 4, pp. 265–279, 2007.

[19] M. Brunato and C. K. Kalló, "Transparent location fingerprinting for wireless services," Tech. Rep. DIT-02-0071, University of Trento, 2002.

[20] T. Nypan, K. Gade, and O. Hallingstad, "Vehicle positioning by database comparison using the Box-Cox metric and Kalman filtering," in *Proceedings of the 55th Vehicular Technology Conference*, pp. 1650–1654, Birmingham, Ala, USA, May 2002.

[21] L. Qiu and R. A. Kennedy, "Radio location using pattern matching techniques in fixed wireless communication networks," in *Proceedings of the International Symposium on Communications and Information Technologies (ISCIT '07)*, pp. 1054–1059, Sydney, Australia, October 2007.