A Configurable Fingerprint-Based Hidden-Markov Model for Tracking in Variable Channel Conditions

John D. Roth∗, Murali Tummala†, John McEachen‡, and James Scrofani†

∗Electrical and Computer Engineering Department
U.S. Naval Academy
Annapolis, MD, USA
Email: jroth@usna.edu
†Electrical and Computer Engineering Department
Naval Postgraduate School
Monterey, CA, USA

Abstract—A novel scheme for mobile subscriber positioning is proposed based on the hidden-Markov model (HMM) and the cell-ID maximum-likelihood database correlation method also known as fingerprinting. Using a simulated channel environment, based on the Clearwire deployment of WiMAX base stations in San Jose, CA, we show that matching the right configuration of the model to the deployment environment can realize significant gains in performance. The proposed scheme balances the scalability inherent in hidden-Markov-based motion models deployed in large areas of interest against the existing channel conditions and computational capability. By utilizing a simulated channel this paper demonstrates the effect of base station deployment and shadowing on the fingerprint-based HMM motion model. Further, the benefits gained through scaling the HMM are explored.

Index Terms—Hidden-Markov Model, channel conditions, non-line-of-sight (NLOS), shadowing, positioning, geolocation, maximum-likelihood, fingerprinting, database correlation, cell-ID

I. INTRODUCTION

Wireless geolocation has a ubiquitous presence in modern day life and a myriad of applications. It has even been mandated by the Federal Communications Commission (FCC) in 1996 with specific performance bounds in order to facilitate the location of the growing number of mobile subscribers requiring emergency services [1]. In addition to facilitating emergency services, wireless geolocation also enables a litany of Location-Based Services (LBS) like location sensitive billing, fraud protection, asset tracking, fleet management, surveillance [2], and various services for autonomous vehicles and wireless sensor networks. Additionally, LBS have significant marketing implications such as the ability to deliver location-based advertisements and promotions [3]. Other applications in health care and environmental habitat monitoring abound [4]. Also, mobile internet applications where a user can check into a geographic location can be enhanced such as Facebook or Foursquare. Wireless geolocation techniques can be generally divided into geometric methods and pattern-based methods.

Geometric methods are labeled as such as they heavily rely on the physical geometry of the mobile subscriber (MS) and associated base stations (BS). Their physical geometry is leveraged via either lateration or angulation techniques to arrive at a position estimate.

Multiangulation is accomplished through determining the angle of arrival (AOA) of a signal from several local BSs. In this case, three or more BSs provide a location in 3. Generally, this problem can be seen as solving a system of linear equations [5]–[7]. AOA measurements require an array of antennas that are able to determine the direction of arrival. This may be a prohibitive requirement in some technologies [2], [8].

Multilateration is accomplished by determining the distance of a MS from the known location of a number of BSs. Time of arrival (TOA) is one way to gather this data. In this method the time it takes a signal to travel from known BS locations to the MS is measured. This time is then converted to distance assuming speed of light propagation. Each of these distances defines a circle around the BS of interest. The intersection of circles becomes the MS position estimate [5], [7]. Consequently, the accuracy of this method relies heavily on overall network synchronization, a serious obstacle for some technologies [2], [7], [8]. Time difference of arrival (TDOA) can be seen as an extension to the idea of TOA. Instead of considering the absolute time of travel between a MS and a known BS, rather the time difference of arrival between a MS and any two BSs is recorded. The time difference describes a hyperbola of which the two BSs are foci [6]. The intersection of any three hyperbolae become the position estimate of the MS [5]. A major advantage to this technique is that total network synchronization is not required [2], [7], [8]. Finally, multilateration may be achieved from received signal strength (RSS) measurements. This method attempts to determine the distance from a known BS location via the strength of a received signal. This method assumes that the transmitted signal strength is known and also that the channel can be modeled properly. Because of the difficulty involved in modeling a channel in a multipath environment this method is generally not preferred when a high level of accuracy is required [7], [8].

In practical scenarios there is usually some level of noise
involved in each measurement required for the above geometric techniques. Noise is present in many different forms but most notably when there is a non-line of sight (NLOS) environment between a transmitter and receiver. A NLOS situation occurs when there is a physical blockage in the line of sight (LOS) between a receiver and a transmitter requiring the signal of interest to travel some non-minimal distance to arrive at the receiver [2], [9], [10]. This positive bias [11] in the measurements usually, but not always, results in a set of inconsistent equations that can induce errors an order of magnitude higher than if no NLOS condition existed [10]. Finding a unique solution from among a set of inconsistent equations requires some level of sophistication to address.

The NLOS problem can be addressed via three general means: identify transmissions received via NLOS propagation paths and discard them, use a localization error database, or to find an optimal solution while working within the constraints imposed by NLOS.

Identify and discard techniques assume that NLOS conditions increase error variance (an assumption that has been shown to be incorrect [10]) and thus use error variance as a metric for identifying the corrupted information. In [12] each combination of BSs is examined by means of an error residual. When the error residual drops the offending BS(s) is (are) assumed to be found and discarded. In [13] the principles of hypothesis testing are used to compare the current error variance with historical values to determine if a NLOS environment is present. Channel statistics such as kurtosis, mean excess delay, and root mean square (RMS) delay spread have also been shown to be good metrics by which to gauge the validity of a measurement [14]. These approaches can also be made using the principles of maximum-likelihood [15]–[17].

Using a location error database to draw an a priori empirically obtained location error adjustment is another way to address the NLOS problem. This solution obtains an error correction constant from a database to mitigate the effect of NLOS [18]. This method, however, suffers from a lack of flexibility in adapting to changing channel conditions and a large up front cost in establishing a database and maintaining its accuracy.

Lastly, the NLOS problem can be addressed by working within the constraints of NLOS. In [19] and [20] the problem is formulated as an minimization problem with the geometry of the BSs and MS added as a constraint. A weighted least squares (WLS) approach can be taken by weighting more heavily measurements which are more trusted [21]. Another approach, assuming perfect a priori knowledge of which BSs represent NLOS transmissions, first defines a feasible region and linearizes it as a constraint on the original linear set of equations [22].

Alternatively, pattern-based approaches have the distinct advantage of being inherently robust in NLOS environments. Indeed, this class of techniques exploits the conditions created in this type of environment. Because of the robustness of this approach pattern-based methods are normally used in indoor and dense urban environments. This approach takes measure-

ments from surrounding beacons (perhaps local area network access points or BSs) observed by a MS and compares them to a database of measurements. The database measurement that most closely approximates the observed measurement is then mapped to a location, which is used as the position estimate. This process is accomplished in two phases. First, in the offline phase, the measurement database is collected [5], [23]. This is accomplished primarily through wardriving or radio planning tools [24]. The large up front cost associated with database construction is a significant disadvantage to using pattern-based methods of positioning. However, new research is showing that this cost to database construction can be mitigated with the principles of compressive sensing [25]. Signals collected and recorded in the database could be the full channel impulse response (CIR) [26], [27], actual RSS of the BS [28], [29], or only a binary cell-ID measurement [24], [30], [31]. CIR methods are robust in dense multipath environments [26], but require a higher computational cost in the database search and hardware adjustments. RSS methods also demonstrate promising results but may also require hardware adjustment to MSs that do not currently have the ability to accurately measure signal strength [29]. RSS based on SCORE (for WiMAX networks) has also been investigated and shown to have poor performance relative to more exact RSS methods [24], [29]. Cell-ID, while inherently less accurate than the aforementioned database types, shows promising results with little requirement for hardware update. The mobile must only determine if a cell is within reception range.

Pattern-based methods, or fingerprinting, is also frequently augmented with a probabilistic motion model such as the Kalman filter [32], particle filter [29], hidden-semi-Markov model (HSMM) [30], and the hidden-Markov model (HMM) [24], [33]. Kalman filter approaches increase accuracy but have difficulty modeling nonlinear motion [3] like MS mobility. Particle filters are excellent at modeling non-linear motion and will arrive at the optimal Bayesian solution as the number of particles increases [34], but require significant computational prowess relative to other motion models. The HMM, on the other hand, provides an excellent stochastic motion model, capable of modeling nonlinear motion at a reasonable computational cost. Indeed, the contribution of the HMM to various types of state estimation are ubiquitous [35].

This paper proposes a configurable fingerprinting-based MS positioning scheme able to adapt to a multitude of channel environments. The HMM is adopted as a motion model and is leveraged when appropriate to enhance positioning accuracy while minimizing computational burden. Additionally, its inherent and novel scalability (introduced by Henderson [31]) makes its implementation in large target areas computationally practical. This work is the first to study motion model assisted fingerprinting in simulated channel environments. Scalability and HMM behavior is studied in multiple channel environments to address their effect on accuracy.

The rest of this paper is organized as follows. The three phases and theory of the proposed scheme are presented
in Section II. Experimental setup and the validity of the proposed model is demonstrated through simulation under various channel conditions in Section III. A discussion of the implication of the results is also presented. The proposed scheme and its attributes are summarized in Section IV.

II. HMM ASSISTED FINGERPRINTING SCHEME

The proposed method augments a cell-ID maximum-likelihood fingerprinting scheme with a scalable HMM. The scalability of the HMM allows for flexibility in the face of large areas of interest (AOI) that would otherwise be computationally untenable. The model consists of three phases: offline database training, online fingerprinting, and HMM augmentation.

A. Offline Database Training

Database training is largely conducted as in [24] and is summarized as follows. Training inputs to the database matrix, \( X \), are empirically obtained via methods such as wardriving or estimated using radio planning tools where

\[
X \doteq \{ \chi_{i,j} \} \tag{1}
\]

and

\[
\chi_{i,j} = \Pr[\text{receiving} \ ith \ BS \ | \ MS \ at \ the \ jth \ tile]. \tag{2}
\]

From (2) we see that \( X \) is a matrix of size \( N_{BS} \times N_t \). The \( jth \) tile’s fingerprint (also termed the “\( jth \) tiles diversity set”) is defined as \( X_{i,j} \).

B. Scalable Online Maximum-Likelihood Fingerprinting

Once the database \( X \) is created the online fingerprinting can be accomplished. Part of the novelty of this scheme is its scalability. The scheme considers either the entire AOI at once or scales the AOI to a subset of masked tiles \( \Upsilon \), to be considered independently from the remaining tiles. By considering only those tiles in \( \Upsilon \), where \( |\Upsilon| \leq N_t \), the computational load can be reduced. As the position estimate moves throughout the AOI the tiles considered in \( \Upsilon \) change such that the position estimate always lies at the center of a \( \sqrt{|\Upsilon|} \times \sqrt{|\Upsilon|} \) subset of considered tiles. This paper considers \( |\Upsilon| = 9, 25 \) and 144 where \( |\Upsilon| = 144 = N_t \) is adapted from the model presented in [24].

Once \( |\Upsilon| \) is selected a likelihood vector of length \( |\Upsilon| \) at time \( k \) is created

\[
L_k \doteq \{ l_j \} \tag{3}
\]

where

\[
 l_j = \Pr[\text{MS at the} \ jth \ tile] \tag{4}
\]

calculated by

\[
l_j = \| h_{i,k} - X_{i,j} \|_2^{-1}, \ j \leq |\Upsilon| \ \text{and} \ j \in \Upsilon \tag{5}
\]

where \( h_{i,k} \) is the set of observations made at time \( k \). This method was shown by Bshara, et al. to be robust to measurement errors [24]. The fact that this scheme can accommodate \( |\Upsilon| < N_t \) means that positioning can be achieved in extremely large environments with a lower computational cost than previously established methods. It will be shown later that computational cost is traded for some level of accuracy in positioning and robustness against initialization error.

The maximum-likelihood position estimate from the online fingerprinting phase of the scheme at time \( k \) is calculated by

\[
\hat{p}_{k} = \arg \max \{ L_k \}. \tag{6}
\]

If \( |\Upsilon| < |\text{AOI}| \) and the HMM is not used to augment this phase then \( \hat{p}_{k} \) must be mapped back to the global solution space via a function \( f_{map} = \{ \hat{p}_{k}, \Upsilon \} \). Further, if there is not a unique solution to (6) then a \( \hat{p}_{k} \) is chosen at random from the available solutions. No road network information in included in the fingerprinting phase and a \( \hat{p}_{k} \) that maps to a tile with no road in it is possible.

C. Scalable Hidden-Markov Model

Similar to the fingerprinting phase, the complimentary hidden-Markov motion model is also scalable in order to extend the flexibility to the user of trading computational complexity for accuracy and robustness against initialization error. Two points are important to understand about this phase and its overall integration with the tracking scheme. First, the HMM is complimentary, but not necessary. We will see later that under certain circumstances the HMM actually degrades positioning accuracy. Second, the number of states considered in the HMM must equal \( |\Upsilon| \) in order to provide any sort of meaningful contribution to the overall scheme. The position estimate of the HMM will be later fused with the position estimate of the fingerprinting phase to arrive at an overall position estimate.

The HMM position estimate at time \( k \) is represented with a state vector \( \varphi_k \) of length \( |\Upsilon| \) where

\[
\varphi_{j,k} = \Pr[\text{MS at} \ jth \ tile], \ j \leq |\Upsilon| \ \text{and} \ j \in \Upsilon \tag{7}
\]

such that

\[
\| \varphi_{.,k} \|_1 = 1, \ \forall k. \tag{8}
\]

The update to the HMM state vector is given by

\[
\varphi_{t+1,k} = \frac{M^T \Pi_k \varphi_{t,k}}{\|\Pi_k \varphi_{t,k}\|_2} \tag{9}
\]

where \( M \) is a transition matrix and \( \Pi_k \) is a permutation matrix both of size \( |\Upsilon| \times |\Upsilon| \). The transition matrix is defined by

\[
M \doteq \{ m_{i,j} \} \tag{10}
\]

where

\[
m_{i,j} = \Pr[\text{transition to tile} \ j \ | \ MS \ at \ tile \ i] \tag{11}
\]

such that

\[
\| m_{i,.} \|_1 = 1, \ i \leq |\Upsilon|. \tag{12}
\]

The function of the permutation matrix is to ensure the correct orientation of the state vector relative to \( \Upsilon \) as the position estimate moves throughout the AOI. When \( |\Upsilon| = N_t \) or the overall position estimate does not change from time.
where \( k - 1 \) to time \( k \) then \( \Pi_k = I \), where \( I \) is the identity matrix, and (9) simplifies to

\[
\varphi'_{i,k+1} = M^T \varphi_{i,k}.
\]  

An important effect of the permutation matrix is the loss of information in the state vector under conditions where \( \Pi_k \neq I \). As the position estimate transitions and a new set of tiles is considered in \( \Upsilon \), \( \Pi_k \) discards the \textit{a priori} state information in \( \varphi_{i,k} \) corresponding to the tiles considered at time \( k - 1 \) but not at time \( k \). Because of this, the divisor in (9) is required in order to normalize the probability mass function represented by \( \varphi_{i,k} \) satisfying (8).

The transition matrix \( M \) allows the introduction of three different types of information into the model: mobility information in the form of probability of transition, \textit{a priori} state information, and road network information. Probability of transition is adjusted using a user input parameter, probability of self-transition, given by \( p_0 \). A higher \( p_0 \) models a more mobile MS while a lower \( p_0 \) models a MS with a more sedentary nature. In any case \( p_0 \) can be adapted to a specific MS behavior.

\textit{A priori} state information can be valuable to optimal MS location and achieves its maximum impact when \( |\Upsilon| = N_t \). When \( |\Upsilon| < N_t \) information is discarded by \( \Pi_k \) and the value added by this \textit{a priori} distribution drops as \( |\Upsilon| \) decreases. This will later be shown to be a very powerful consideration when choosing the correct HMM configuration.

Finally, the road network information is entirely contained in the transition matrix. The road information mapped to the transition matrix by

\[
r_k = f_{\text{road}}(\hat{p}_k, \Upsilon)
\]

where

\[
r_k \subseteq \Upsilon
\]

and \( r_k \) is the set of tiles which contain a road. The individual values chosen for \( M \) are then given by

\[
m_{i,j} = \begin{cases} 
    p_0 & \text{if } f_{\text{map}}(j, \Upsilon) = \hat{p}_k \text{ and } i \in \Lambda_j \\
    p_i & \text{if } f_{\text{map}}(j, \Upsilon) \in r_k \text{ and } i \in \Lambda_j \\
    0 & \text{otherwise}
\end{cases}
\]

where \( p_i \) is chosen such that (12) is satisfied and \( \Lambda \subseteq \Upsilon \) where \( \Lambda_j \) includes the \( j \)th tile and all tiles that physically border the \( j \)th tile. No matter what value used for \( |\Upsilon| \), (9) allows for transitions from one state to another state that are not immediately connected by the road network as long as both states contain a portion of the road network. This is preferable since we can expect the HMM to make prediction errors and model recovery time is minimized when the previous position estimate is not used as another constraint when mapping the road network information to the conditional transition probabilities.

The results of the fingerprinting phase and the HMM are then fused by

\[
\varphi_{i,k+1} = L_k \odot \varphi'_{i,k+1}
\]  

where \( \odot \) denotes the Haddamard product and the position estimate is found via

\[
\hat{p}_k = \arg \max \{ \varphi_{i,k} \}.
\]

If there is no unique answer the scheme selects an available answer at random. If \( |\Upsilon| < N_t \) then \( f_{\text{map}}(\hat{p}_k, \Upsilon) \) is used to map the scaled estimate, \( \hat{p}_k \), back to the original solution space.

III. SIMULATION RESULTS

A. Simulated Database Training

In this study database training was conducted entirely in simulation. A 144 square kilometer section of a Clearwire WiMAX network deployment consisting of \( N_{\text{BS}} = 46 \) base stations (BS) was selected as the foundation for the simulated database. The AOI was then divided into one square kilometer tiles creating a total of \( N_t = 144 \) total tiles. Measurements are then simulated in each tile by defining a set of BSs for each tile, which are physically close enough to the tile that a MS could feasibly receive that BS. This set of BSs is termed \textit{local}. The set of BSs that are local to a given tile are then further broken into a center-local and a peripheral-local subset. Center-local BSs are the most proximate to the tile and are close enough that a MS will always be able to receive that BS. Peripheral-local BSs are more distal to the tile and will thus only be received with a certain probability. The full database matrix is then constructed from the sum of the two subset matrices

\[
X = X^{cl} + X^{pl}
\]

where \( X, X^{cl}, \) and \( X^{pl} \) are defined as in (1).

The non-zero elements of the peripheral-local database matrix \( X^{pl} \) are drawn from a pseudo-Gaussian distribution approximated by

\[
f_x(x) \cong N(m_\tau, \sigma_{n,\tau}) I_{[0,1]}(x)
\]

where \( I_{[0,1]}(x) \) is the indicator function. The input parameters to the normal distribution are determined by a user input parameter, threshold \( (\tau) \), defined subsequently.
During actual database training that may take place the time varying nature of the channel will create non-deterministic fingerprints. This phenomenon is modeled with the threshold parameter \( \tau \). The model begins with a uniform random variable \( A \) on the open interval \((0, 1)\). \( A \) is then transformed to a discrete random variable \( Y \) via a non-linear function
\[
Y = f(A)
\] defined by
\[
y = \begin{cases} 
1 & \text{if } A > \tau \\
0 & \text{otherwise}
\end{cases}
\] where each \( y \) is interpreted as a independent and identically distributed cell-ID measurement. The \( n \) measurements are then transformed to \( \chi \) via
\[
\chi = \frac{1}{n} \sum_{i=1}^{n} Y_i.
\]
(23)
By the central limit theorem, a sufficiently large \( n \) will result in an \( \chi \) given by (20) [36], [37]. For this simulation \( n = 50 \). From (21)-(23) it can be shown that the variance and mean of \( \chi \) are given by
\[
\sigma_{\tau,n} = \frac{\tau - \tau^2}{n}
\] and
\[
m_{\tau} = 1 - \tau
\]
(24)
(25)
respectively. Therefore, as can be seen in (24), \( \tau \) controls the strength of the shadowing effect or the uncertainty that a MS will or will not be able to receive a specific BS in the peripheral-local subset of a given tile by adjusting the input parameters to the distribution (20). As shown in Figure 2 and seen in (22) and (24) a low \( \tau \) means high certainty that BSs in the peripheral-local subset will be reachable at a given tile. A high \( \tau \) means high certainty that BSs in the peripheral-local subset will not be reachable at a given tile. Finally, as \( \tau \) approaches its median value the uncertainty of reaching a peripheral-local BS approaches its maximum.

The user input threshold also has a secondary effect of adjusting the homogeneity of the AOI. The homogeneity of an AOI is defined as a measure of similarity in the resident diversity sets. This paper quantifies the homogeneity via the Manhattan distance of two diversity sets given by
\[
d(X_{i,q}, X_{i,r}) = \|X_{i,q} - X_{i,r}\|_1, \ q \neq r.
\]
(26)
Two diversity sets are said to be in the homogeneous set \( \Omega \) if
\[
d(X_{i,q}, X_{i,r}) \leq \epsilon
\]
(27)
for some small \( \epsilon \). The homogeneity of an AOI is then defined as the cardinality of \( \Omega \). The similarity of the simulated AOI, only considering diversity sets on the road network, is also shown in Figure 2 as a function of \( \tau \). As \( \tau \) increases, a MS at any given tile is able to receive less BSs with more certainty. This effectively increases the homogeneity of the AOI. This phenomenon was also seen in [24] and the weak signal from more distant BSs was necessary in order to achieve the reported performance. Thus, with the user input threshold parameter \( \tau \) the channel environment can be used to adjust the environmental homogeneity and effect of shadowing.

Next an observation matrix
\[
H = \{ h_{i,k} \}
\]
(28)
is created where
\[
h_{i,k} = \begin{cases} 
1 & \text{if the } ith \text{ BS is center-local at time} k \\
0 & \text{if the } ith \text{ BS is peripheral-local at time} k \\
\end{cases}
\]
(29)
where \( k \) is the measurement number and \( y \) is a Bernoulli random variable defined in (22) and drawn independently from the database creation step. For this simulation \( N_k = 34 \). From (29) we see that \( H \) is a \( N_{BS} \times N_k \) matrix. Each measurement \( h_{i,k} \) is taken at some regular sampling interval as the simulated MS moves throughout the AOI via a predetermined route highlighted in Figure 1.

**B. HMM Augmented Results**

First, the HMM augmented scheme is considered. For each value of \( |\mathcal{Y}| \) the simulation was run according to the track shown in Figure 1. The actual track covers all of the available tiles on the road network with the exception of the outermost band of tiles. Additionally, the simulated MS traverses each intersection at least once. Uniform initialization is always applied to \( \phi_{1} \) in order to minimize initialization error within \( \mathcal{Y} \). For this simulation the simulated MS had a tile dwell time of two samples per tile and the model was empirically optimized in the approximate range \( 0.20 \leq p_{10} \leq 0.30 \). Each configuration was evaluated by a figure of merit (FOM), \( \Psi \), defined as the average positioning error over a simulation and given analytically by
\[
\Psi = \frac{\sum_{k} ||p_k - \hat{p}_k||_2}{N_k}.
\]
(30)
The averaged results over 12 trials of the simulation at
Fig. 3. Performance of the HMM augmented model when $|\Upsilon| = 9$. Each value of $|\Upsilon|$ and for each combination of $p_0$ and $\tau$ are shown in Figures 3-5 as a colored contour plot. Red denotes a high average positioning error, and thus poor positioning performance, while blue denotes the opposite. These plots are intuitively satisfying as for some probabilities of self-transition as the threshold increases the model performance generally follows the plot of the normalized threshold in Figure 2. Additionally, the models tend to prefer the lower thresholds where there is less homogeneity in the AOI. Overall, the plot where $|\Upsilon| = N_t$ seems to provide the best performance. This should be expected as the HMM can take the maximum amount of information from time $k$ to $k + 1$ and provide the most considered maximum-likelihood solution. Additionally, the entire fingerprinting database may be considered during each iteration. It is also expected that $|\Upsilon| = 25$ will outperform $|\Upsilon| = 9$ as the larger model will be able to carry more information from time $k$ to $k + 1$ than the smaller model. Higher values of $|\Upsilon|$ also demonstrate less sensitivity to the probability of self-transition. However, these improvements in performance come at a computational cost, especially when a MS must be tracked in AOIs much larger than this one. It is worth considering how much positioning accuracy the computational cost really contributes when choosing the appropriate model for an AOI.

The cumulative distribution function (CDF) of the positioning error over 100 independent simulations of each model configuration is presented in Figure 6. It is shown that the model where $|\Upsilon| = 25$ has the same circular error probable (CEP) of $|\Upsilon| = N_t$ up to approximately CEP 67% while the model where $|\Upsilon| = 9$ only matches the other models’ CEP up to approximately CEP 30%.

C. Maximum-Likelihood Fingerprinting Only Results

Next, the fingerprint only scheme is considered. For each value of $|\Upsilon|$ the simulation was run for each threshold value according to the same track used in the previous section.
The results are shown in Figure 7. Each data point represents the aforementioned $\Psi$ averaged over 12 independent trials and presented for varying levels of threshold. Again, superior performance is noted in the case where $|\Upsilon| = N_t$, but comparable performance is also noted in the case where $|\Upsilon| = 25$. In the case where the HMM is not used to augment the fingerprint-based positioning, $|\Upsilon| = 25$ strikes an excellent balance between performance, computational cost, and scalability. The increase in $\Psi$ as threshold also increases is expected as the homogeneity of the AOI was previously shown to be proportional to $\tau$. The variance in performance and high average positioning error in the case where $|\Upsilon| = 9$ when $\tau \gg 0.5$ makes the model’s practical applicability suspect.

Finally, a comparison of the performance of the model in the HMM augmented and un-augmented cases is presented in Figures 8-10. Here the performance gains realized by augmenting the fingerprint-only model with an HMM are presented. Positive values represent the improvement in average positioning error an HMM can contribute. A value of zero shows that either the HMM degrades performance or there is no overall change in performance. These results demonstrate that the value added in the HMM occurs in highly homogeneous environments. This makes intuitive sense since in highly homogeneous environments we would expect the fingerprinting scheme to fail. As the effectiveness of fingerprinting decreases the HMM is able to recover performance by adding road map information, a priori state information, and mobility information.

D. Optimizing Model Configuration

Given the results we can deduce what conditions warrant which configuration of the model guided by the premise that the primary goal is maximizing accuracy while balancing a secondary goal of minimizing computational cost. Several factors must be held in tension. First, and most importantly,
TABLE I
CONDITIONS FAVORABLE TO DIFFERENT CONFIGURATIONS OF THE PROPOSED MODEL.

| $|T|$ | 9 | 25 | $|N_t|$ |
|---|---|---|---|
| **Fingerprint Only Model** | - Heterogeneous environments | - Heterogeneous environments | - Heterogeneous environments |
| | - Very low computational capability | - High initialization cost | - High computational capability |
| | - High initialization confidence | | - Low initialization confidence |
| **HMM Augmented Fingerprinting** | - Homogeneous environment | - Homogeneous environment | - Homogeneous environment |
| | - Very low computational capability | - High initialization confidence | - High computational capability |
| | - High initialization confidence | | - Low initialization confidence |

is the nature of the operating environment. If the environment is heterogeneous then a fingerprint only method will realize a greater accuracy than the HMM augmented method while simultaneously reducing computational complexity. If the environment is homogenous then the HMM augmented model must be considered.

Given the model type, the optimal value of $|T|$ may be considered. The smallest model, when $|T| = 9$, only makes sense when the demand for a scheme with very low computational complexity is significant. Otherwise better performance can be achieved with models of larger $|T|$. If the un-augmented model is selected, comparable performance can be achieved when $|T| = N_t$ or $|T| = 25$. Thus, performance can be maintained with lower computational complexity when $|T| = 25$. However, this size model is very sensitive to initialization errors. If the uncertainty of the MS location at $k = 1$ is high, a model where $|T| = N_t$ must be utilized. Alternatively, another method, such as one of the geometric techniques presented in Section I may be used to initialize the model. Once initialization is achieved with some predetermined level of confidence the model should be transitioned to the case where $|T| = 25$ in order to minimize computational costs. If the HMM augmented model is chosen the same methodology can be used to select $|T|$. In all cases the goal is to transition the model to $|T| = 25$ as soon as confidence in the position estimate meets a certain threshold. The results of this comparison are summarized in Table I.

IV. CONCLUSION

This paper has presented a configurable fingerprint-based HMM positioning model. Fingerprinting was done via a maximum-likelihood method on an a priori database. The database and corresponding observation vectors were trained via simulation during the offline phase. Additionally, an optional HMM-based motion model was used to augment the fingerprinting scheme. The configuration of the model was tested in a simulated channel environment based on a Clearwire WiMAX network deployment in San Jose, CA. Simulations were conducted under varying degrees of homogeneity and various shadowing conditions.

The un-augmented model was shown to be preferable in heterogeneous environments while the HMM augmented model was shown to be preferable in homogenous environments. The optimal mask size for balancing performance and computational cost was shown to be $|T| = 25$. The un-scaled model, introduced in [24], where $|T| = N_t$ demonstrated the highest performance and robustness against initialization error purchased at a high computational cost. The model with $|T| = 9$ was shown to be advantageous only when computational capability was scarce. The scheme was shown to be highly adaptable to a wide variety of channel environments delivering high accuracy while minimizing computational cost.

REFERENCES


