Hybrid Indoor Tracking Using Crowdsourced Measurements

Zafer Vatansever and Maite Brandt-Pearce
Charles L. Brown Department of Electrical and Computer Engineering
University of Virginia, Charlottesville, VA 22904
Email: zv4xv@virginia.edu, mb-p@virginia.edu

Abstract—Tracking of user equipment in indoor areas is becoming ubiquitous with the development of location-based services. The use of an RF received signal strength map in wireless positioning systems, a so-called site survey, is a well-studied technique. In this paper, we propose a hybrid method that combines the trilateration technique used in Wi-Fi and an extended Kalman filtering for visible light with crowdsourced measurements to create a received light-signal strength map. How to efficiently create such a site survey is an open challenge. Our aim is to offer a solution to this problem that is able to automate the map creation step using data sourced from the users. The results show that the proposed method outperforms Wi-Fi-only based trilateration methods. The accuracy of the proposed method is around 5 cm when the resolution of the fingerprint map is 1 dm.

Index Terms—Wireless communications, visible light communications, indoor positioning, trilateration, Kalman filtering, data fusion, data processing

I. INTRODUCTION

The advent of wireless mobile devices creates an opportunity for location-based services (LBS). The quality of LBS relies on accurate positioning and tracking systems within an enclosed area. For outdoor areas, the Global Navigation Satellite System (GNSS) is the accepted solution. GNSS signals are not accurate in indoor areas due to obstructions and diffusion. For indoor use, there are a variety of solutions proposed that utilize ultrasound, Wi-Fi, visible light communications (VLC), etc.; however, none of them prevails over the other.

The solution we offer in this work uses already deployed systems in the area. We rely on existing Wi-Fi and LED lighting systems and combine their advantages to enhance the positioning accuracy. Wireless RF access points (AP) are present in almost every indoor space, like shopping malls, industrial plants, etc. Light emitting diodes (LED) are taking the place of incandescent and fluorescent-based lamps because of their longer life expectancy and lower power consumption. LEDs have also created the opportunity for VLC systems.

We propose a three-step solution. In the first step, the mobile unit uses the Wi-Fi based trilateration to estimate its rough position. In the second step, the algorithm measures the received light intensity from orthogonally-coded LEDs, and assigns them to the position found by Wi-Fi based trilateration. As a new user moves, the light intensity measurements are updated. This is called the fingerprint map learning process. The third step is to track the user using an extended Kalman filter (EKF).

The trilateration method is probably the easiest and oldest localization method; it relies on estimating the distance between the transmitter and the receiver. Already deployed wireless access points can be used for indoor tracking and positioning. [1] uses Wi-Fi received signal strength (RSS) for trilateration, averaging the signal strength to improve performance. Android-based mobile devices are used for trilateration with Wi-Fi in [2]. The blocking of Wi-Fi signals decreases the RSS; specific reference points are used to improve performance in [3]. A zero-configuration Wi-Fi trilateration method is proposed in [4], where the process initiates with a protocol messaging over the Wi-Fi. Trilateration is also used in VLC positioning [5]–[7].

Bayesian filtering, especially Kalman filters (KF) and particle filters (PF), are widely used techniques in target tracking literature [8]. The EKF is used in [9] for indoor tracking using VLC. Sigma-point Kalman filter is used for a Wi-Fi-based fingerprinting method in [10]. An angle of arrival (AOA) based KF tracking method is proposed in [11]. The computational complexity of KF and PF is discussed in [12].

Fingerprinting based methods require accurate knowledge of the fingerprint distribution. The information collection step in the targeted indoor area is called a site survey. Performing the site survey is an expensive, time-consuming, labor-intensive work if it is done manually. Crowdsourcing of the fingerprints is a way of reducing the site survey costs; the users collect and inform the fingerprints to the system [13]. The concept of Wi-Fi crowdsourced data collection is gaining more attention. An on-demand approach to select the best fingerprints in the map is proposed in [14]. In [15], the RSS is used as fingerprints; users collect and pass on the RSS information to the system. A probabilistic method that overcomes noisy samples is also proposed. Crowdsourced approaches suffer from the time-data problem, meaning that the system needs time and data from the users to reach a feasible fingerprint map. This problem is discussed and a hybrid solution is proposed in [16].

In the problem of indoor tracking with VLC, the drawback with fingerprinting is the site survey. How to perform the site survey in an automated fashion is not well investigated in VLC. In Wi-Fi methods, the crowdsourced information becomes valuable with research on automating the site survey. The VLC fingerprints in our study are the optical RSS. We combine Wi-Fi trilateration with VLC-RSS data collection to create a fingerprint map from the measurements, and use the
KF for indoor tracking.

The rest of this paper is organized as follows. In Section II, our fingerprint map creation method is described in detail; the VLC channel is described in Section II-A, the noise sources and the signal-to-noise ratio analysis are explained in Section II-B, the Wi-Fi signal propagation model and how to estimate the distance between the transmitter and the receiver are explained in Section II-C, and finally the update scheme used to construct the fingerprint map is discussed in Section II-D. In Section III, the use of the fingerprint map and the target tracking with the EKF are explained. Section IV present our numerical results. Conclusions are drawn and summarized in Section V.

II. FINGERPRINT MAP LEARNING

In this section, we explain how the fingerprint map is generated. The basic idea is to use the rough location of the user obtained through Wi-Fi trilateration to generate an optical power site survey, and use the inherent smoothness expected from optical power distributions to enhance the accuracy of the map.

A. Visible Light Communication Model

The visible light channel is an optical wireless channel that uses LEDs as the transmitters. For simplicity, we only consider the line-of-sight (LOS) propagation. The channel DC gain of the LOS propagation is given in [17] as:

\[ H_{LOS} = \begin{cases} \frac{A_r(m+1)}{2m} \cos^m(\phi) \cos(\psi), & \text{if } \Psi_c \geq \psi \geq \Psi_c, \\ 0, & \text{otherwise,} \end{cases} \]

where \( A_r \) is the area of the receiver; \( m \) is the Lambertian mode of the transmitted beam, which is related to the semiangle of the LED, \( d \) is the distance between the LED and the photodetector, \( \phi \) is the radiation angle for the transmitter; \( \psi \) is the incident angle at the photodetector; \( \Psi_c \) is the field-of-view of the photodetector. Assuming no delay spread due to multipath, the received power \( P_r \) can be calculated as:

\[ P_r = H_{LOS} \times P_t, \]

where \( P_t \) is the transmitted power from the light source. Each lamp is encoded with an orthogonal code so that the received power from each lamp can be determined by correlating with each code before computing the RSS.

In our previous work, we showed that Bayesian filtering can be used for indoor tracking using VLC [18]. As explained in the introduction, the received light power is observed by the user. Yet the relationship between the received light and the user position is nonlinear; this dictates the use of an EKF. The challenge with the EKF lies in the VLC channel model given by (1). The EKF requires linearization of nonlinearity in the system. However, the Jacobian of (1) relies on perfect knowledge of irradiance and incidence angles, which is unrealistic in a real system. Instead of calculating the derivatives, we use a finite difference method, which requires a matrix of expected received powers, i.e., a fingerprint map. Henceforth, this map is denoted by \( P \).

B. Signal-to-Noise Ratio Analysis For Visible Light Communications

The noise sources in a VLC system are receiver noise (shot and thermal). There is also optical interference noise caused by sunlight or other light sources; these additional light sources create a background current in the photodetector, and are accounted for the shot noise [19]. Shot noise depends on the random nature of photon absorption and electron-hole recombination. Thermal noise results from the thermal fluctuations in the electron density.

A third noise term we use in our signal-to-noise ratio (SNR) calculations is “uncertainty” noise. The source of uncertainties may be errors in the power map creation, shadowing, other physical effects like objects in the area, uneven dimming of lights, etc. Considering these effects, the SNR equation is modified as:

\[ SNR = \frac{\gamma^2 P^2_r}{\sigma^2_{\text{shot}} + \sigma^2_{\text{thermal}} + \sigma^2_{\text{uncertainty}}}, \]

where \( \gamma \) is the receiver responsivity, \( P_r \) is the average received power. When the \( P_r \) is low, “uncertainty” noise is dominant, because the power map estimation error is high. For high \( P_r \), the shot and thermal noise are dominant.

C. Wi-Fi Signal Based Trilateration

There is an abundance of Wi-Fi access points in our target indoor areas, like shopping malls, industrial facilities, and museums. Trilateration from Wi-Fi based signals is widely studied in the literature. Wi-Fi based trilateration relies on accurate estimation of the distance between the transmitter (AP) and the receiver, referred to as the user equipment (UE). In [20], it is shown that modeling the path loss with the Friis free space propagation model is not accurate. Therefore, we use the following equation given in [20] to model the propagation loss:

\[ P_r(d)[\text{dBm}] = P_0(d_0)[\text{dBm}] - 10n_p \log_{10} \left( \frac{d}{d_0} \right) + X_\sigma \]

\( P_r(d) \) is the RSS at a distance \( d \) from the transmitter. \( P_0(d_0) \) is a known reference power at distance \( d_0 \) from the transmitter. \( n_p \) is the propagation-environment-dependent path loss exponent, and \( X_\sigma \) is a zero mean Normal distributed random variable that models the random effects in the propagation medium with standard deviation of \( \sigma \). Given the received power, \( P_r \), the distance between the AP and the receiver, \( d \), can be estimated from (4). The unbiased estimate of the distance between the transmitter and the receiver is given in [20] as:

\[ \hat{d} = d_0 \left( \frac{P_r}{P_0(d_0)} \right)^{-1/n_p} \exp \left( -\frac{\sigma^2}{2 \left( \frac{10}{\ln(10)} \right)^2 n_p^2} \right) \]

When the distance between the transmitter and the receiver, and the positions of the transmitters are known, the UE positions can be estimated using the least squares method. This method is known as trilateration.
The trilateration method does not by itself provide accurate position estimation. The estimates will have errors caused by the random effects in the propagation environment; these effects were captured by the random variable \( X_\sigma \) in (4). Fig. 1 shows the effect of \( X_\sigma \) on the estimated distance in a typical room for one AP. The AP is placed at \((x = 1 \text{ m}, y = 1 \text{ m})\) in the 5 m \( \times \) 5 m room. To do trilateration, at least three APs are needed. 

As briefly explained in Section I, we propose to use the matrix \( \hat{P} \) to update this power map surface using new crowdsourced data. The data is used to update and learn the fingerprint map. The received light intensity measurements will not be assigned to the true positions but the positions found from trilateration. The initial fingerprint map is updated, and is a sparse matrix now; the nonzero values are in the positions found by trilateration.

(iv) According to (1), the light distribution is a smooth function across the room and from the light source to the floor. In this step, the power map data is smoothed to create a smooth surface, which is updated in the subsequent steps. The surface smoothing methodology chosen is to duplicate the nearest non-zero value.

D. Fingerprint Map Updating

The fingerprint map is a matrix in which the average received power values are stored. The physical area of interest is divided into \( K \times J \) rectangles. The matrix is denoted by \( \mathbf{P} = [P_{ij}] \), where \( i = 1, \ldots, K \) and \( j = 1, \ldots, J \) are the indices of the Cartesian coordinates. The size of the map is application specific. If a high accuracy is needed for sensitive applications, like autonomous robots used in industrial plants, a high-resolution map is needed.

We have examined two approaches of using the trilateration results to generate \( \mathbf{P} \). The first is where the initial map is updated sequentially as users walk into the room, the second is to use the matrix \( \hat{T} \) in a batch process. In the first approach, the initial fingerprint map, \( \mathbf{P}^0 \), is a matrix of zeros with same dimensions as \( \mathbf{P} \). After the first user moves in the surveillance area, four steps are followed:

(i) The position of the UE is found by Wi-Fi trilateration as explained in Section II-C.

(ii) The received light intensity is measured by the UE at the UE’s true position.

(iii) The trilateration positions and measurements are combined. The trilateration introduces positioning errors, which cause a mismatch in the power map. The light intensity measurements will not be assigned to the true positions but the positions found from trilateration. The initial fingerprint map is updated, and is a sparse matrix now; the nonzero values are in the positions found by trilateration.

(iv) According to (1), the light distribution is a smooth function across the room and from the light source to the floor. In this step, the power map data is smoothed to create a smooth surface, which is updated in the subsequent steps. The surface smoothing methodology chosen is to duplicate the nearest non-zero value.

The site survey in the second approach takes longer. In the second approach, we wait until \( N \) users have collected trilateration measurements and optical RSS data in the area for us. The data is used to update and learn the fingerprint map just like in the first approach but in a batch process instead of sequential processing. A linear KF is used to smooth the trilateration results after step (i) of the first approach. The KF reduces the error between the estimated position and the true position. The rest of the algorithm is the same as for the first approach.

Once we have an initial surface from one of the approaches described above, the survey map can be used by the EKF as described below. However, environments are not static. We envision an adaptive implementation where the next step is to update this power map surface using new crowdsourced data. The mixing rule we use in this work is given as:

\[
P_{ij}^N = \alpha \omega P_{ij}^{N-1} + (1 - \alpha \omega) \hat{P}_{ij}^{N-1}
\]
$P_{ij}^N$ is the updated power values at index $ij$ to be updated from the new trajectory. $P_{ij}^{N-1}$ is the power values in the surface from the previous crowdsourcing. $P_{ij}^{N-1}$ is the measured power received from the LEDs at that moment. $\alpha \in (0, 1)$ is the learning rate. It decides which one should be trusted more: the previously computed surface or the measurements. $\omega$ is the VLC channel coherence, allowing the previous power estimates more or less longevity. As the channel impulse response changes this variable changes. The superscript $N$ is the trajectory number.

In the map learning process, the trilateration and light power measurements are done in the UE, then uploaded via Wi-Fi to a central processor. After learning the map $\mathbf{P}$, assuming that the conditions change slowly, the map can be updated based on user-sourced data.

### III. Tracking with Extended Kalman Filter

Once $\mathbf{P}$ is sent to the UE, the user tracking begins. Considering the limited resources of the UE, we use an extended Kalman filter (EKF) for tracking instead of a particle filter [21].

The discrete time EKF is suboptimal under nonlinear process or measurement models [22]. A discrete-time state-space Kalman filter (EKF) for tracking instead of a particle filter considering the limited resources of the UE, we use an extended and $\mathcal{H}$ and $\mathcal{r}$ and $\mathcal{q}$ number of LED lamps, the process noise is

\[
\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{q}_{t-1}
\]

\[
y_t = h(\mathbf{x}_t) + \mathbf{r}_t
\]

where $\mathbf{x}_t \in \mathbb{R}^4$ is the state vector, $\mathbf{y}_t \in \mathbb{R}^v$ is the measurement vector (which is also used as $\mathbf{P} = [\hat{P}_{ij}^{N-1}]$ in (6)), $v$ is the number of LED lamps, the process noise is $\mathbf{q}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ (a zero mean, Gaussian distributed noise with covariance $\mathbf{Q}_t$), and $\mathbf{r}_t \sim \mathcal{N}(0, \mathbf{R}_t)$ is the measurement noise with covariance $\mathbf{R}_t$. $t$ is the time index. $f(\cdot)$ is the dynamic model function and $h(\cdot)$ is the measurement model function.

Success in state estimation requires an accurate representation of the target motion. We do not expect the UE to make sudden direction changes or accelerate. Therefore, the discrete-time constant-velocity (CV) model defined in [23] is used here.

The aim is to track the position of the target. The target state is defined as $\mathbf{x} = [x, \dot{x}, y, \dot{y}]^T$. Where $x$ and $y$ are the positions on the Cartesian coordinates, and $\dot{x}$ and $\dot{y}$ are the corresponding velocities. The initial state of the UE is found by Wi-Fi trilateration, and the initial error covariance matrix is chosen according to the error introduced in the Wi-Fi trilateration process.

As mentioned above, (1) shows a nonlinear relationship between the power measurements and the state variables. Furthermore, the calculation of the derivatives for the linearization step in the EKF is intractable when the angular information is unknown. Instead, we use a finite difference method to approximate the derivatives. We have the fingerprint map, $\mathbf{P}$, already in hand. We use $\mathbf{P}$ as a look-up table in the following way. First, using the CV model in the prediction step of the EKF, the predicted state, $\hat{\mathbf{x}}$, is found. When the light power in the predicted state is measured, the algorithm uses $\mathbf{P}$ to obtain the stored power levels at the predicted location and in the four adjacent locations. These are used to generate a finite difference approximation of the Jacobian of $h(\mathbf{x}_t)$ in (8) needed for the EKF.

Fig. 2 shows the finite difference method employed. $\hat{P}(\hat{x})$ is the predicted power, and $\hat{P}(x_{t+1}, j \pm 1)$ are the four adjacent powers found from $\mathbf{P}$. The Jacobian for $h(\mathbf{x}_t)$ is approximated as:

\[
H(x) \approx \left[ \frac{P(i+1,j) - P(i-1,j)}{2\Delta x}, \frac{P(i,j+1) - P(i,j-1)}{2\Delta x} \right] 
\]

where $\Delta x$ is the granularity of the power map.

### TABLE I: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room dimension</td>
<td>$5 \times 5 \times 3 \text{ m}^3$</td>
</tr>
<tr>
<td>Lambertian mode (m)</td>
<td>1</td>
</tr>
<tr>
<td>LED bulb elevation and azimuth</td>
<td>-90° and 0°</td>
</tr>
<tr>
<td>Positions of the LED bulbs</td>
<td>I. (1.25, 1.25, 3)</td>
</tr>
<tr>
<td>in the room(x, y, z) (m)</td>
<td>II. (3.75, 1.25, 3)</td>
</tr>
<tr>
<td>in the room((x, y, z) (m)</td>
<td>III. (3.75, 3.75, 3)</td>
</tr>
<tr>
<td>Height of the PD</td>
<td>0.75 m</td>
</tr>
<tr>
<td>Field of view ($\Psi_{cy}$) of the PD</td>
<td>70°</td>
</tr>
<tr>
<td>Physical area of the PD ($A_z$)</td>
<td>1 mm</td>
</tr>
<tr>
<td>Receiver elevation and azimuth</td>
<td>90° and 0°</td>
</tr>
<tr>
<td>Gain of optical filter and refractive index of the lens at the PD</td>
<td>1.0 and 1.0</td>
</tr>
<tr>
<td>Positions of the Wi-Fi APs</td>
<td>I. (1, 1, 1.5)</td>
</tr>
<tr>
<td>in the room (x, y, z) (m)</td>
<td>II. (5, 1, 1.5)</td>
</tr>
<tr>
<td>in the room (x, y, z) (m)</td>
<td>III. (5, 5, 1.5)</td>
</tr>
<tr>
<td>Transmitted power from each Wi-Fi AP ($P_t$)</td>
<td>40 mW</td>
</tr>
<tr>
<td>Propagation-environment-dependent path loss exponent ($n_p$)</td>
<td>4</td>
</tr>
<tr>
<td>Reference distance ($d_0$) (x, y) m</td>
<td>(1, 1)</td>
</tr>
<tr>
<td>Learning rate ($\alpha$) in (6)</td>
<td>0.5</td>
</tr>
<tr>
<td>VLC channel coherence ($\omega$) in (6)</td>
<td>1</td>
</tr>
</tbody>
</table>
The room is assumed to be an empty room with dimensions $5 \times 5 \times 3 \text{ m}^3$ and the granularity of the power map $\Delta x$ is 1 dm. The user equipment is moving at a fixed height, and the device orientation is 90° degrees to the ceiling. Some of the key simulation parameters are given in Table I.

The process noise covariance $Q_t$ is chosen to be optimal for the particular trajectories used for simulations in this study. The optimal noise parameters are found by keeping the measurement noise constant and tuning the process noise to the level that yields to minimum error as was done in [18].

We assume that the VLC channel impulse response does not change in this study, for simplicity. The power map is updated in batches. We do not update the map every time a user is introduced, but the system waits until a predefined number of trajectories are available. We assume that $\omega$ and $\alpha$ are constant values so that we put equal emphasis on the updated power map values and the real-time measurements.

Wi-Fi trilateration needs at least three APs with known positions. Equations (4) and (5) are used to estimate the distance between the UE and each of the APs, and the least square solution gives the estimated positions. In the simulations, APs with identical path loss parameters are positioned on three corners of the room.

Fig. 3 gives an overview of the performance of our algorithm using one of four different methods:

(a) Wi-Fi trilateration: The distances from at least three APs are estimated using (5). The estimated distances are used to solve trilateration equations.

(b) Wi-Fi trilateration and smoothing with a KF: The estimated positions are smoothed using a linear KF.

(c) Wi-Fi trilateration, VLC-EKF: The received light intensity powers at the UEs’ true positions are assigned to the UEs’ estimated positions by Wi-Fi trilateration. This step constructs a fingerprint map of the light intensity after smoothing. The fingerprint map is used for EKF-based tracking.

(d) Wi-Fi trilateration, KF smoothing, and VLC-EKF: The same as (d); the only difference is the use of a linear KF to smooth the estimates from the Wi-Fi trilateration.

The experiments show that as the random effects in the propagation medium, $X_\sigma$, increase, the RMSE increases. The estimates can be more accurate by using a KF on the estimated positions obtained with Wi-Fi trilateration, i.e., the performance of (c) is improved in (d) by introducing a KF. This is because the KF decreases the error introduced while building $P$ by smoothing the Wi-Fi triangulation measurements.

As the main contribution in this paper, we show that the accuracy can be increased by combining a different modality that exists in every indoor space: visible light. When we introduce the visible light measurements and its modified EKF, as explained in Section III, the tracking accuracy increases dramatically. Fig. 3 shows this improvement in tracking accuracy for low to moderate VLC-SNR levels. At a low SNR (25 dB), the uncertainty noise dominates, and the performance of (c) is worse than (d). This is because the error introduced while creating $P$ limits the system. However, if the SNR increases, for example to 45 dB, this error is reduced and the tracking performance of (c) becomes similar to (d). The results in Fig. 3 motivate the use of a linear Kalman filter between the Wi-Fi trilateration and the $P$ building process at the expense of increased computation. It also shows that the results for (c) and (d) are not affected by $X_\sigma$, although $X_\sigma$ affects the Wi-Fi trilateration directly, and $P$ indirectly.
Fig. 4 shows the number of UE-trajectories needed to create a sufficient and accurate $P$ that is to be used in the VLC-EKF when method (d) is employed. The UE-trajectories are the source of crowdsourced visible light intensity used to build the map, $P$. It is clear from the figure that for our particular simulation and room size, after 300 trajectories the performance is indistinguishable.

Fig. 5 shows that the tracking performance is hardly affected by the value of $X$, if crowdsourcing method (d) is used. Thus, the error introduced from the Wi-Fi trilateration during the crowdsourcing step does not affect the final performance. The figure also shows that with a light intensity map $P$ defined over a $\Delta x = 1$ dm grid, the RMSE reaches the quantization level (less than $0.5$ dm) at higher SNR levels.

In Fig. 6, the effect of the number of UE-trajectories and its effect on the tracking accuracy is shown when method (c) is used. If we compare Figs. 4 and 6, we can see that method (d) is slightly better than (c) in terms of accuracy, especially for lower SNR values like 25 to 30 dB. But the difference is not significant for moderate to high SNR values.

Fig. 7 shows the tracking accuracy results for method (c). At low SNR, the performance is far worse than methods (a), (b) and (d). But as the SNR increases, the performance converges to similar values as for (d). The question becomes which one of these two methods to use. The latter one requires more calculations due to the use of a linear KF applied on the Wi-Fi trilateration solutions, but yields a higher accuracy, especially for low SNRs. The first one sacrifices a little accuracy but does not use a KF.

V. CONCLUSION

In this study, we adopt a crowdsourced measurement model that is often-used in Wi-Fi positioning systems to a VLC based target tracking system. We use the Wi-Fi to estimate the rough position of the UE and the associated visible light intensity at the UE’s estimated position. We use this light intensity to create a power map that is needed in the VLC-based EKF tracking algorithm. There are two main contributions in this work: we propose to automate the fingerprint map building process for a VLC-based tracking system, and, we combine two modalities, the Wi-Fi and the VLC for indoor tracking. The results show that the accuracy is better than Wi-Fi only trilateration methods.

REFERENCES


