Cooperative Multi-Point Vehicular Positioning Using Millimeter-Wave Surface Reflection

Zezhong Zhang, Student Member, IEEE, Seung-Woo Ko, Member, IEEE, and Kaibin Huang, Fellow, IEEE

Abstract—Multi-point vehicular positioning is an essential operation for autonomous vehicles. However, the state-of-the-art positioning technologies, relying on reflected signals from a target (i.e., RADAR and LIDAR), cannot work without line-of-sight (LoS). Besides, it takes significant time for environment scanning and object recognition with potential detection inaccuracy, especially in complex urban situations. Some recent fatal accidents involving autonomous vehicles further expose such limitations.

In this article, we aim at overcoming these limitations by proposing a novel relative positioning approach, called Cooperative Multi-point Positioning (COMPOP). The COMPOP establishes cooperation between a target vehicle (TV) and a sensing vehicle (SV) if a LoS path exists, where a TV explicitly lets an SV to know the TV’s existence by transmitting positioning waveforms. This cooperation makes it possible to remove the time-consuming scanning and target recognizing processes, facilitating real-time positioning. One prerequisite for the cooperation is a clock synchronization between a pair of TV and SV. To this end, we use a phase-differential-of-arrival (PDoA) based approach to remove the TV-SV clock difference from the received signal. With clock difference correction, the TV’s position can be obtained via peak detection over a 3D power spectrum constructed by a Fourier transform (FT) based algorithm. The COMPOP also incorporates nearby vehicles, without knowing their locations, into the above cooperation for the case without a LoS path. Specifically, several strong non-LoS (NLoS) links from the TV to the SV can be generated via mirror-like reflections over the neighboring vehicles’ metal surfaces. Following the same procedures in the LoS case, virtual TVs mirrored by nearby vehicles can be detected. By exploiting the geometric relation between the virtual and actual TVs, COMPOP can be achieved by intelligently combining the virtual TVs to position the actual TV. The effectiveness of the COMPOP is verified by several simulations concerning practical channel parameters.

Index Terms—Vehicular sensing, multi-point positioning, hidden vehicle detection, synchronization, Fourier transform, 3D imaging, resolution.

I. INTRODUCTION

VEHICULAR positioning is one of the most important operations for autonomous driving and a challenging one as it requires high accuracy and low latency [2], [3]. Presently, for long-range positioning (e.g., hundreds of meters), a target vehicle (TV) is abstracted as a single point, and its GPS location is communicated to the sensing vehicle (SV) over either a vehicle-to-vehicle link or across a wireless network [4], [5]. For medium to short-range positioning (tens to several meters), the single-point abstraction of the TV no longer suffices, and its geometric information (e.g., size and orientation) is also required for safe and accurate driving. In these ranges, the popular positioning technologies include RADio-Detection-And-Ranging (RADAR) and LIght-Detection-And-Ranging (LIDAR). They face the challenges of long computation latency caused by complex signal processing and computer vision, potential inaccuracy of identifying TVs from their background environments, and ineffectiveness in the presence of blockages between a TV-SV pair. Their drawbacks have contributed to many accidents involving self-driving cars. To overcome these limitations of current technologies, we present in this article a new technology for relative positioning at the SV, called Cooperative Multi-point Positioning (COMPOP). Essentially, by detecting the cooperative signals broadcast by multi-antennas distributed at a TV, the SV estimates the antenna positions representing the TV skeleton, thereby performing multi-point TV positioning. Algorithms based on Fourier transform (FT) are proposed for fast and accurate COMPOP not only when there are lines-of-sight (LoS) but also when they are blocked. The latter exploits the signals reflected by the surfaces of nearby vehicles.

A. Single-Point Positioning

Single-point positioning techniques were originally developed for locating mobile devices and recently also applied to autonomous driving. The most popular and simplest is to use a built-in Global Positioning System (GPS) receiver for computing the receiver’s position and sharing the information.
to peer devices over wireless links. However, in an urban environment, the required LoS links between GPS receivers and satellites are often blocked by e.g., buildings or tunnels. This issue has motivated researchers to develop alternative techniques relying on base stations or access points in wireless networks as anchors to estimate the position of a mobile by either triangulation or measuring signal power [6]. This requires accurate distance measurements from the mobile to different anchors. The negative effect of mobility on estimation accuracy can be coped with by utilizing sampled temporal measurements and motion models [5]. Nevertheless, due to the unreliability of fading channels, such positioning techniques assisted by a wireless network cannot reach the level of precision required for autonomous driving. The required high-resolution positioning can be realized by ultra-wideband (UWB) radios leveraging the fact that a large bandwidth overcomes multi-path fading and thereby enables accurate time-of-arrival (ToA) measurements [7]. To avoid the need of transmitter-receiver clock synchronization, time-difference-of-arrival (TDoA) and phase-difference-of-arrival (PDoA) based methods are proposed which cancel at a receiver the clock difference of a transmitter (anchor) by observing received signals from different antennas or frequencies [8], [9]. Most recently, TDoA-based positioning over non-LoS (NLoS) links is made possible by separating multi-paths and locating the source via exploiting the paths’ geometric relation [10]. The deployment of such techniques enables the positioning of a hidden TV (one without an LoS link). Despite having a rich literature, the outputs of single-point positioning are insufficient for complex maneuvers in autonomous driving such as platooning and overtaking, for which multi-point TV positioning is required.

B. Vehicular Sensing

Vehicular sensing can be treated as an extreme form of multi-point positioning. The sensing process involves scanning the surrounding environment and then recognizing, imaging, and positioning objects useful for autonomous driving (such as nearby pedestrians and vehicles). Usually vehicular sensing does not rely on the GPS link, but detects relative locations of objects independently with an on-board sensing system. Relevant technologies can be grouped as passive or active, depending on whether they require radiation. Typical passive sensors include infrared sensors, cameras, and passive millimeter waves (mmWave) sensors. They exploit ambient and unintended infrared radiation, light, and mmWave to image the sources and reflectors [11]. The two most popular types of active vehicular sensors are RADAR and LIDAR [2], [12]. A RADAR scans the environment by steering a microwave beam using an antenna array and observes reflected signals with varying attenuation to image the environment [13], [14]. Subsequently, targeted objects are detected and positioned using signal processing and computer vision. LIDAR operates based on a similar principle except for replacing the microwave beam with a mechanically steered sharp laser beam, thereby achieving a higher resolution [15]. Among the sensing technologies, those based on light (i.e., cameras) or infrared (e.g., infrared sensors) are exposed to severe performance degradation caused by hostile weather such as heavy rains and thick fogs [16]. Overall, existing multi-point positioning technologies share two drawbacks that present key challenges for vehicular sensing and positioning, as described below.

- **Latency and accuracy:** Environmental scanning (for RADAR and LIDAR) and object recognition (for RADAR, LIDAR, and cameras) are time-consuming. For example, a 3D beam scanning by a RADAR using a large-scale phased array can incur around ten-second latency [14]. This is unacceptable for autonomous driving in a crowded urban environment or at high speeds. On the other hand, objective recognition using a well-trained deep neural network with an onboard GPU typically takes several seconds. Furthermore, existing objective recognition techniques relying on offline training are easily affected by a variation on object features. As a result, the detection accuracy is usually in the range of 70% ~ 90% [17]. Their application to auto-driving presents safety threats, as exemplified by recent fatal accidents.

- **Hidden vehicle detection:** Besides TVs in sight, detecting hidden vehicles in the sensing blind spots (e.g., a TV around a street corner) can avoid many potential accidents. Though some progress has been made on NLoS single-point TV positioning, the desired detection of hidden TVs as multi-point objects is still an uncharted area and the theme of this work.

C. Main Contributions

Multi-point TV positioning refers to positioning a TV as a multi-point object where each point corresponds to a transmit antenna. To tackle the two challenges discussed in the preceding section, we propose in this work the framework of cooperative multi-point positioning (COMPOP) at a SV building on the cooperation where a TV broadcasts signals that facilitate relative TV positioning at the SV. In this framework, latency reduction is achieved in two ways.

- The first is to retrieve from the total received signal the desired TV signal for further processing using the embedded signature. This avoids the time-consuming conventional method of environmental scanning and learning-based TV identification.

- The second is to apply the low-latency FT to efficiently compute a power spectrum distributed over the 3D space. This allows direct positioning by peak detection.

To tackle the second challenge of multi-point positioning without an LoS link, we propose a novel technique of using reflected TV signals over the smooth surfaces of nearby vehicles to position multiple “virtual TVs”. Then we exploit the geometric relation between the virtual and actual TVs to position the actual TV by combining the “virtual TVs” without any prior knowledge of the nearby vehicles. By overcoming these two key challenges to the existing positioning approaches, we believe the proposed technique can be well integrated with the current positioning system for improving the safety of autonomous driving.
The specific designs of the proposed COMPOP framework are summarized as follows.

- **COMPOP over a LoS link**: The framework for this case comprises two key algorithms operating in separate bandwidths. First, due to transceiver separation, clock synchronization between the transceivers is necessary for the clock difference removal \(^1\) at the SV. By using two reserved single tones and given the knowledge of TV signature waveform, we propose to use a PDnA-based iterative algorithm for TV-SV synchronization by accurately estimating their clock difference from the received TV signals in the presence of channel noise. Second, the remaining bandwidth is used by the TV to transmit a multi-tone waveform with a uniform frequency gap between tones, which is similar to the conventional stepped-frequency-continuous-wave (SFCW) waveform [13]. After correcting the clock difference, the relation between the TV position and the received multi-tone signal is established. As a main feature of COMPOP, we propose the application of FT to transform the received multi-tone signal into the spectrum over the 3D space for estimating the TV position by peak detection.

- **COMPOP over mirror-reflection links**: For the case without LoS, but with nearby vehicles as reflectors, a COMPOP technique is developed as follows. The combined use of antenna array and TV signature allows the SV to resolve the received signals reflected by different vehicular surfaces. The application of the preceding technique for LoS COMPOP on the resolved signals yields multi-point positions of multiple “virtual TVs”. Without prior knowledge of the nearby vehicles’ locations, an intelligent combining approach is proposed to position the actual TV by exploiting the geometric relation between the virtual and actual TVs.

II. SYSTEM MODEL

Consider the scenario with multiple vehicles on the road, including one pair of TV and SV. The TV is equipped with an antenna array with \(N_w\) elements distributed over the vehicle body which sketches its shape. The SV equipped with a distributed array of \(N_r\) antennas performs multi-point positioning of the TV by locating its antennas from their broadcast waveforms in a mmWave spectrum. As illustrated in Fig. 1, we consider both the scenarios where the TV and SV broadcast waveforms in a mmWave spectrum. As illustrated in Fig. 1, we consider both the scenarios where the TV and SV broadcast waveforms in a mmWave spectrum. As illustrated in Fig. 1, we consider both the scenarios where the TV and SV broadcast waveforms in a mmWave spectrum. As illustrated in Fig. 1, we consider both the scenarios where the TV and SV broadcast waveforms in a mmWave spectrum.

A. Signal Models

Each TV antenna transmits the superposition of two waveforms for different operations of the COMPOP. One is a signature waveform transmission enabling the SV to estimate the TV-SV system clock difference for synchronization.

\(^1\)The clock difference removal aims at removing the clock phase difference in the received signals so that the propagation distance can be deduced from the phase of the resultant signals.

The second is a multi-tone waveform in a separate bandwidth from the signature, which facilitates COMPOP at the SV. The waveforms are described as follows.

1) **Signature Waveform**: For clock-difference estimation, it is sufficient to transmit different signature waveforms over two antennas, whose indices are denoted as \(a\) and \(b\), using two single tones for each antenna. The antenna coordinates are represented as \(x_a = (x_a, y_a, z_a)\) and \(x_b = (x_b, y_b, z_b)\). The two waveforms for the two antennas, defined as \(s_a\) and \(s_b\), are given as

\[ s_a(t + \sigma) = e^{j2\pi f_a(t + \sigma)} + e^{j2\pi (f_a + \Delta)(t + \sigma)}, \]
\[ s_b(t + \sigma) = e^{j2\pi f_b(t + \sigma)} + e^{j2\pi (f_b + \Delta)(t + \sigma)}, \]

where \(f_a\) and \(f_b\) are two orthogonal frequencies specifying the signatures, \(\Delta\) is a given frequency separation, and \(\sigma\) is the TV-SV system clock difference (in sec).

First, consider the scenario with \(L\) surface-reflection links \((L \geq 1)\). Let \(\Gamma(t)\) denote the complex reflection coefficient of the \(t\)-th link given as \(\Gamma(t) = |\Gamma(t)| \exp(j\Gamma(t))\). Moreover, let \(\tau_{n,m}(t)\) denote the signal flight time from TV’s antenna \(n\) to SV’s antenna \(m\) proportional to the propagation distance \(d_{n,m}(t)\), i.e., \(d_{n,m}(t) = c \cdot \tau_{n,m}(t)\) where \(c\) is the speed of light. Given the notations above, the raw received waveform at each SV antenna, say antenna \(m\), is given as

\[ W_m(t) = \sum_{\ell=1}^{L} w_{m}(t) = \sum_{\ell=1}^{L} w_{m}(t), \]

where \(w_{m}(t) = \Gamma(t) s_a(t + \sigma - \tau_{a,m}(t)) + \Gamma(t) s_b(t + \sigma - \tau_{b,m})(t)\).

We assume that the AoAs of different signal arrivals are separable using a classic technique, e.g., MUSIC [18], where the AoAs can be accurately detected in the angular domain by searching the power spectrum of the received signals.\(^2\)

After AoA detection, the received signals can be decomposed by coherent detection in the angular domain, and regrouped as a vector \(w_m(t) = [w_{m1}(t), \ldots, w_{mL}(t)]\), where each element in \(w_m(t)\) comprises waveforms at four different frequencies. By exploiting frequency orthogonality, different frequency components in the received waveform antenna \(m\), can be separated and grouped to form two \(K\) by \(L\) matrices.

\(^2\)The effect of antennas on AoA detection accuracy is well studied by simulations in [19] such that AoA error is significantly reduced as the number of antennas increases.
to facilitate the algorithmic design in the sequel:
\[
\begin{align*}
A_m(t) &= [a_m^{(1)}(t), \ldots, a_m^{(L)}(t)], \\
B_m(t) &= [b_m^{(1)}(t), \ldots, b_m^{(L)}(t)],
\end{align*}
\]
where
\[
\begin{align*}
a_m^{(l)}(t) &= \Gamma^{(l)} \left[ e^{j2\pi f_s(t+\tau_{n,m}^{(l)})} e^{j2\pi (f_s+\Delta)(t+\tau_{n,m}^{(l)})} \right], \\
b_m^{(l)}(t) &= \Gamma^{(l)} \left[ e^{j2\pi f_s(t+\tau_{n,m}^{(l)})} e^{j2\pi (f_s+\Delta)(t+\tau_{n,m}^{(l)})} \right].
\end{align*}
\]

Next, two matched filters are designed as
\[
\begin{align*}
D_a(t) &= \text{diag}\left\{ e^{-j2\pi f_s t}, e^{-j2\pi (f_s+\Delta) t} \right\}, \\
D_b(t) &= \text{diag}\left\{ e^{-j2\pi f_s t}, e^{-j2\pi (f_s+\Delta) t} \right\}
\end{align*}
\]
for clock difference removal, turning \(A_m(t)\) and \(B_m(t)\) from time functions into matrix symbols:
\[
\begin{align*}
A_m &= D_a(t)A_m(t) = \left[ a_m^{(1)}, a_m^{(2)}, \ldots, a_m^{(L)} \right], \\
B_m &= D_b(t)B_m(t) = \left[ b_m^{(1)}, b_m^{(2)}, \ldots, b_m^{(L)} \right],
\end{align*}
\]
where
\[
\begin{align*}
a_m^{(l)} &= \Gamma^{(l)} \left[ e^{j2\pi f_s(\tau_{n,m}^{(l)})} e^{j2\pi (f_s+\Delta)(\tau_{n,m}^{(l)})} \right], \\
b_m^{(l)} &= \Gamma^{(l)} \left[ e^{j2\pi f_s(\tau_{n,m}^{(l)})} e^{j2\pi (f_s+\Delta)(\tau_{n,m}^{(l)})} \right].
\end{align*}
\]

It is worthwhile to notice that all equations above also hold if one of the \(L\) links is a LoS link.

Next, consider the other scenario with a LoS link to the TV, where unresolvable reflection links possibly exist but are neglected due to the significant power difference between LoS and NLoS paths. Without loss of generality, let the channel gain be normalized as \(\Gamma = 1\). Then the LoS counterparts of \(A_m\) and \(B_m\) can be simplified from (5) and (6) as
\[
\begin{align*}
A_m^{\text{LOS}} &= \left[ e^{j2\pi f_s(\tau_{n,m})} e^{j2\pi (f_s+\Delta)(\tau_{n,m})} \right], \\
B_m^{\text{LOS}} &= \left[ e^{j2\pi f_s(\tau_{n,m})} e^{j2\pi (f_s+\Delta)(\tau_{n,m})} \right].
\end{align*}
\]
where \(\tau_{n,m}\) denotes the signal flight time from TV’s antenna \(n\) to SV’s antenna \(m\) proportional to the propagation distance \(d_{n,m}\) in LoS.

2) Multi-Tone Waveform: A multi-tone waveform generated by orthogonal-frequency-division-multiplexing (OFDM), which is similar to the stepped-frequency-continuous-wave (SFCW) waveform [13], [20] used in RADAR, is broadcast by each TV antenna. The waveform, denoted as \(s(t)\), comprises multiple single-tone continuous-waves with equally separated frequencies by a fixed frequency gap \(\Delta\). Mathematically,
\[
s(t) = \sum_{k=1}^{K} \exp(j2\pi f_k t),
\]
where \(\mathcal{F} = \{ f_k \}_{k=1}^{K}\) represents the set of frequencies such that
\[
f_k = f_1 + (k-1)\Delta\text{ for } k = 1, \ldots, K.
\]
The vector form of the multi-tone waveform is also provided as
\[
s(t) = [\exp(j2\pi f_1 t), \ldots, \exp(j2\pi f_K t)]^T,
\]
since the single-tone continuous-waves are naturally separated in frequency.

First, consider the scenario with \(L\) surface-reflection links \((L \geq 1)\). With frequency decoupling, the received signal at the SV’s antenna \(m\) is
\[
r_m(t) = \sum_{\ell=1}^{L} r_m^{(\ell)}(t),
\]
where \(r_m^{(\ell)}(t)\) represents the signal vector from the \(\ell\)-th surface-reflection link as
\[
r_m^{(\ell)}(t) = \Gamma^{(\ell)} \sum_{n=1}^{N_\ell} r_{n,m}^{(\ell)}(t) = \Gamma^{(\ell)} \sum_{n=1}^{N_\ell} s(t+\sigma-\tau_{n,m}^{(\ell)}).
\]

By assuming perfect AoA detection using the MUSIC algorithm [18], we can decompose (11) into individual \(r_m^{(\ell)}(t)\) by coherent detection in the angular domain, which is rewritten as a \(K\) by \(L\) matrix \(R_m(t)\) as follows:
\[
R_m(t) = \left[ r_m^{(1)}(t), \ldots, r_m^{(L)}(t) \right].
\]
Assuming that the TV-SV system clock difference \(\sigma\) is estimated as \(\hat{\sigma}\), the initial phase in received signal (13) is removed by multiplying the \(K\) by \(L\) matched filtering matrix \(D(t) = \text{diag}\{s(t+\hat{\sigma})\}^T\) as
\[
Y_m = [y_m^{(1)}, \ldots, y_m^{(L)}] = D(t)R_m(t),
\]
where \(y_m^{(\ell)} = D(t)y_m^{(\ell)}(t) = [y_{m,1}^{(\ell)}, y_{m,2}^{(\ell)}, \ldots, y_{m,K}^{(\ell)}]^T\) with the component \(y_{m,k}^{(\ell)}\) being
\[
y_{m,k}^{(\ell)} = \Gamma^{(\ell)} \sum_{n=1}^{N_\ell} \exp \left[ j2\pi f_k(\sigma-\hat{\sigma}-\tau_{n,m}^{(\ell)}) \right].
\]
With accurate estimation of the clock difference of the algorithm in Section III-A, the \((\ell,k)\)-th received signal component is simplified as
\[
y_{m,k}^{(\ell)} = \Gamma^{(\ell)} \sum_{n=1}^{N_\ell} \exp \left[ -j2\pi f_k^{(\ell)} \tau_{n,m}^{(\ell)} \right], \quad \forall \ell, k.
\]
Next, consider the other scenario with a LoS link. The received signal component in (16) can be further simplified with \(\Gamma = 1\) as
\[
y_m^{(k)} = \sum_{n=1}^{N_\ell} \exp \left[ -j2\pi f_k^{(\ell)} \tau_{n,m}^{(\ell)} \right], \quad \forall k.
\]

Moreover, without loss of generality, we define the origin \(O\) of the coordinate system at the center of the SV, \(Z\)-axis in the direction of AoA. The \(X, Y\)-axes are parallel and vertical to the ground, respectively, perpendicular to the \(Z\)-axis, as shown in Fig. 1.

Remark 1 (Feasible Ranging Distance of Multi-Tone Waveform): Due to the periodicity of phases, the maximum ranging distance of the multi-tone waveform should be limited by \(R_{\text{max}} = \frac{\pi}{\Delta}\) (in meters) to avoid ambiguity [13].

Remark 2 (Channel Fading): The sampling process takes \(T_s = \frac{1}{B}\) seconds, where \(B\) is the system bandwidth at the SV. Since the sampling duration \((3.3 \times 10^{-10} \text{ s} \text{ with } B = 3 \text{ GHz})\) is
much shorter than the coherence time (approximately $T_c = 0.5 \text{ ms}$ at 60 GHz with velocity 10 m/s [21]), the channel fading, comprising the pathloss and small-scale fading, is considered as a constant and omitted in the presentation for convenience.

**B. Procedure Design**

For their non-overlapping spectrums, the two waveforms can be separated by the SV using filtering.

1) **Clock Synchronization**: To establish direct relation between the TV’s multi-point position and the superimposed signal $y_{m}^{\ell,k}$ (16) or $y_{m}^{k}$ (17), the SV aims at compensating the system clock difference $\sigma$ by using (5), (6) or (8), divided into two cases as follows.

- **A LoS link**: When a LoS link exists, the SV is expected to detect the system clock difference $\sigma$ from the received signature waveforms $\{A_m^{\text{LOS}}, B_m^{\text{LOS}} | \forall m\}$, where $A_m^{\text{LOS}}, B_m^{\text{LOS}}$ are matrix signals in (8) received over the $m$-th antenna.

- **Reflection links**: The clock synchronization needs to be achieved for each reflection link. Take the $\ell$-th reflection link as an example, the clock difference $\sigma$ needs to be detected based on signature waveforms $\{a_m^{(\ell)}, b_m^{(\ell)} | \forall m\}$, where $a_m^{(\ell)}, b_m^{(\ell)}$ are given in (7).

2) **Multi-Point Positioning**: Define the set of transmit-antenna locations at a TV as $X = \{x_n\}$ with $x_n \in \mathbb{R}^3$ and $|X| = N_t$. Then, an indicator function $I_{\{x \in X\}}$ to represent the transmit antenna’s distribution over the 3D spatial domain is defined as

$$I_{\{x \in X\}} = \sum_{n=1}^{N_t} \delta(x - x_n),$$

where $\delta(\cdot)$ is a delta function satisfying $\delta(x) = 0, \forall x \neq 0$ and $\int_{\mathbb{R}^3} \delta(x) dx = 1$. Let $D(x_n, p_m) = \|x_n - p_m\|$ measures the LoS distance between $x_n$ and the location of receive antenna $m$ denoted by $p_m$. For reflection links, $D^{(\ell)}(x_n, p_m)$ measures the propagation distance over the $\ell$-th link. Accordingly, we directly have $\tau_{n,m}^{(\ell)} = D(x_n, p_m)$ and $\tau_{n,m}^{(\ell)} = D^{(\ell)}(x_n, p_m)$.

Also, the received multi-tone waveforms (16) and (17) can be rewritten as

$$y_{m}^{\ell,k} = \sum_{n=1}^{N_t} \exp \left( -j \frac{2\pi f_k}{c} D(x_n, p_m) \right), \quad \forall \ell, k.$$  

for reflection links and

$$y_{m}^{k} = \sum_{n=1}^{N_t} \exp \left( -j \frac{2\pi f_k}{c} D(x_n, p_m) \right), \quad \forall k.$$  

for the LoS link if it exists.

As aforementioned, the point set $X$ represents the multi-point position information of the TV. Estimating $X$ can be divided into two cases as follows.

- **A LoS link**: The multi-point TV position $X$ is directly retrieved from the multi-tone signals received in LoS, i.e., $\{y_{m}^{k} | \forall m, k\}$.

- **Reflection Links**: The real multi-point TV position $X$ needs to be detected by combining all multi-tone signals from different reflection links, i.e., $\{y_{m}^{(\ell)} | \forall m, \ell\}$.

**III. COMPOP OVER A LoS LINK**

In this section, we consider a case where a LoS link between the TV and the SV exists, making it reasonable to ignore other reflection links due to the significant power difference between LoS and NLoS paths. The scheme we propose for LoS case is illustrated in Fig. 2, consisting of two steps: 1) synchronization; and 2) multi-point positioning. The overview and algorithm description for each step are presented in the following.

For clarification, we firstly summarized the assumptions used in this section:

- **Perfect AoA Detection**: Accurate AoA detection can be realized by applying classical MUSIC algorithm [18] with an appropriate number of antennas at the receiver.

- **Constant Channel Fading**: The channel fading is considered constant during the sampling process.

**A. Step 1: Synchronization**

1) **Overview**: As illustrated in Sec. II-A, synchronization is necessary to compensate the system clock difference $\sigma$, enabling correct clock difference removal (14) at the receiver and the subsequent multi-point positioning. Note that the clock difference is contained in the signal phases in (8), which can be observed at the receiver. However, to estimate the clock difference $\sigma$ directly is challenging because it is coupled with the propagation delay $\tau_m$ as shown in (8). Therefore, it is necessary to separate these parameters by estimating the propagation delay first, which can be translated into the estimation of one representative transmit antenna, say $x_a$, given the knowledge of the receive antennas’ locations. To this

**Fig. 2. Diagram of COMPOP in LoS.**
end, we first estimate \( x_a \) in the presence of noise by applying a phase-difference-of-arrival (PDoA) based method [9], [22].

2) Algorithm Description: Here we give an approach to estimate the location \( x_a \) of the representative transmit antenna \( a \) from the received signature waveform \( A_{m}^{\text{LOS}} \) in (8). The index of the transmit antenna is omitted for brevity, i.e. \( \tau_m = \tau_{a,m} \).

- **Detection of TX Location:** Let \( F_m(x_a) \) denote the propagation distance difference from the antenna \( a \) to the SV’s antennas \( m \) and 1, given as

\[
F_m(x_a) = D(x_a, p_m) - D(x_a, p_1), \quad m = 2, \cdots, N_r, \quad (21)
\]

At the SV’s antenna \( m \), the phase difference \( \eta_m = 2\pi\Delta(\tau_m - \sigma) \) between the two components of the received signature waveforms can be directly measured from \( A_{m}^{\text{LOS}} \) in (8) with noise in the presence, denoted as \( \tilde{\eta}_m \). Then based on the relation \( \eta_m = 2\pi\Delta(\tau_m - \sigma) = 2\pi\Delta(D(x_a, p_m) - D(x_a, p_1)) \), a noisy measurement of \( F_m(x_a) \) is given as

\[
\hat{F}_m = c \frac{(\tilde{\eta}_m - \eta_1)}{2\pi\Delta} + \Delta \phi_m, \quad m = 2, \cdots, N_r, \quad (22)
\]

where \( \Delta \phi_m \) is the additional Gaussian noise. Since \( F_m(x_a) \) is univariate w.r.t. the location \( x_a \), the TX location can be optimized based on the minimum-mean-square-error (MMSE) criterion to minimize the gap between \( F_m(x_a) \) and the measurement \( \hat{F}_m \) as follows.

**Problem 1 (TX Location Optimization):**

\[
\min_{x_a} \sum_{m=1}^{N_r} \left\| \hat{F}_m - F_m(x_a) \right\|^2_2, \quad (23)
\]

Then the optimal solution \( x_a^* \), which naturally coincides with \( x_a \), can be achieved by using the iterative Gauss-Newton method [23] with an arbitrary initial point \( \bar{x}_a \) as follows

\[
h = \left( G(\bar{x}_a)^T G(\bar{x}_a) \right)^{-1} G(\bar{x}_a)^T b(\bar{x}_a),
\]

\[
x_a \leftarrow \bar{x}_a + h, \quad (24)
\]

where

\[
G(\bar{x}_a) = \begin{bmatrix}
\frac{\partial F_2(\bar{x}_a)}{\partial x_a} & \frac{\partial F_2(\bar{x}_a)}{\partial y_a} & \frac{\partial F_2(\bar{x}_a)}{\partial z_a} \\
\vdots & \vdots & \vdots \\
\frac{\partial F_{N_r}(\bar{x}_a)}{\partial x_a} & \frac{\partial F_{N_r}(\bar{x}_a)}{\partial y_a} & \frac{\partial F_{N_r}(\bar{x}_a)}{\partial z_a}
\end{bmatrix},
\]

\[
b(\bar{x}_a) = \begin{bmatrix}
\hat{F}_2 - F_2(\bar{x}_a) \\
\vdots \\
\hat{F}_{N_r} - F_{N_r}(\bar{x}_a)
\end{bmatrix}. \quad (25)
\]

- **Detection of Propagation Time:** With the knowledge of \( x_a \), the propagation time \( \tau_m \) can be detected for signals at receive antenna \( m \) by \( \tau_m = D(x_a, p_m) \).

- **Subtraction & Averaging:** By subtracting the propagation time \( \tau_m \) component from the noisy phase measurement \( \tilde{\eta}_m \), the system clock difference can be differently calculated depending on the choice of the SV’s antenna \( m \), denoted by \( \tilde{\sigma}_m = \tau_m - \frac{r_m}{2\pi\Delta} \). Averaging these values gives an accurate estimate of \( \sigma \) such that \( \tilde{\sigma} = \frac{1}{N_r} \sum_{m=1}^{N_r} \tilde{\sigma}_m \).

**Proposition 1 (Synchronization Feasibility Condition):** At least four SV’s antennas are required \( (N_r \geq 4) \) to detect the TV-SV system clock difference, according to the solution (24).

Moreover, one more antenna location \( x_a \) can be detected by applying the same algorithm above on the signature waveform \( B_{m}^{\text{LOS}} \) in (8). The detected locations \( x_a, x_b \) of the representative transmit antennas help the COMPOP over the reflection links illustrated later in Section IV-B.

**Remark 3 (TV Recognition in LoS):** Considering the signature waveform and multi-tone waveform transmissions share the same signal paths, the clock difference detected from signature waveforms will be used for multi-tone waveforms with the same AoAs, and the multi-point TV position detected in Sec. III-B will be mapped to the same AoAs as well.

**Remark 4 (Initial Value Selection):** It is recommended to use a solution satisfying any three equations in

\[
\hat{F}_m = D(x_a, p_m) - D(x_a, p_1), \quad m = 2, \cdots, N_r. \quad (26)
\]

as the initial \( \bar{x}_a \), where the convergence to the global optimal is verified by simulations.

**Remark 5 (Sampling Requirement):** The synchronization procedures are based on the assumption that the phase gap estimated at each two adjacent receive antennas is no larger than \( \bar{\pi} \). Thus the distance \( \Delta_d \) between each two adjacent receive antenna at the SV needs to satisfy \( \Delta_d < \frac{\bar{\pi}}{2\Delta} \).

**Proposition 2 (Error Covariance):** We use the error covariance matrix \( \text{cov}(\tilde{x}_a) \) as the performance metric of the above synchronization approach, defined as

\[
\text{cov}(\tilde{x}_a) = \mathbb{E} \left[ (\tilde{x}_a - \mathbb{E}[\tilde{x}_a]) (\tilde{x}_a - \mathbb{E}[\tilde{x}_a])^T \right]. \quad (27)
\]

For tractability, we assume that the phase error follows an independent identically distributed (i.i.d.) Gaussian distribution where \( \Delta \phi_m \sim N(0, \sigma^2 I) \), \( \forall m \). As the number of SV’s antennas \( N_r \) becomes larger, the covariance matrix \( \text{cov}(\tilde{x}_a) \) scales with \( O\left(\frac{1}{N_r-1}\right) \).

**Proof:** See Appendix A.

According to Proposition 2, we assume perfect synchronization in the following steps. However, the phase noise and the resultant estimation errors are kept through the entire process.
Remark 6 (Comparison With Existing Synchronization Methods): Such a clock synchronization method clearly differs from conventional GPS-based approaches [24] since it does not rely on the GPS. Compared to consensus-based synchronization approaches [25], the method we proposed can be processed in real time and thus is more suitable for vehicular sensing.

B. Step 2: Multi-Point Positioning

1) Overview: As shown in Fig. 2, the system clock difference $\sigma$ can be removed by the preceding step, facilitating the following procedures. The main idea of the multi-point positioning step in LoS is briefly illustrated as follows. We first show that the received signal $y_m^k$ can be presented as a function of the indicator $I_{\{x \in X\}}$, denoted as $y_m^k = g(I_{\{x \in X\}})$. Therefore, the position information $X'$ of the TV can be retrieved from the received signals through an inverse function and a following peak detection. We give the estimation of $X'$ in the following algorithm, divided into two phases 1) Fourier transform; and 2) peak detection, as described in Fig. 2.

2) Algorithm Description: The clock difference removal (17) enables to express $y_m^k$ as (20), which can be rewritten in a 3D surface integral form as

$$y_m^k = y(p_m, f_k) = \int_{\mathcal{R}^3} I_{\{x \in X\}} \exp \left(-\frac{2\pi f_k}{c} D(x, p_m)\right) dx,$$

where $y(x, f) : \mathcal{R}^4 \rightarrow \mathcal{R}$ is a continuous function, and

$$D(x, p_m) = \|x - p_m\|$$

represents the Euclidean distance between point $x = (x, y, z)^T$ and the location of the SV’s antenna $m$, denoted by $p_m = (p_{mx}, p_{my}, p_{mz})$. Then based on (28), we have the following lemma based on the scalar diffraction idea in [26].

Lemma 1: Consider the indicator function of transmit antennas $I_{\{x \in X\}}$ and the function $y(x, f)$ representing signals at the receiver. The following equality holds in the frequency domain as

$$\text{FT}^\text{2D} \{s(x, y, f)\} |_{f = |f_k|} = \text{FT}^\text{3D} \{I_{\{x \in X\}}\},$$

where $f = (f(x), f(y), f(z))^T$ is a spatial frequency vector, and $s(x, y, f) = y(x, y, 0, f)$ is a continuous function, representing received signals at the $X - Y$ plane $z = 0$. The $\text{FT}^\text{2D}$ and $\text{FT}^\text{3D}$ are 2D and 3D Fourier transforms defined in Table I.

Proof: Please refer to Appendix B.

Remark 7 (Sampling at the Receiver): Based on the received signals, we are only able to collect samples of $y(x, f)$ discretely over the receive antenna array. Recall that Z-axis is in the direction of AoA known at the SV, leading to the following approximation

$$s(p_m^x, p_m^y, f_k) = y(p_m^x, p_m^y, 0, f_k)
\approx y(p_m, f_k) \exp \left(-\frac{2\pi f_k}{c} p_m^z\right),$$

which holds tightly when the TV-SV distance is much larger than the SV’s size. Therefore, based on the approximation in (30), discrete samples on $s(x, y, f)$ are also available.

Remark 8 (Linear Interpolation): Since receive antennas may not be regularly distributed, a linear interpolation [27] is adopted to provide samples of $s(x, y, f)$ uniformly spaced along $X$ and $Y$ directions. Take samples at two adjacent receive antennas in $X$ direction as an example, whose coordinates are supposed to be $p_m^1$ and $p_m^2$, and $p_m^y = p_m^y$. Then the interpolation generates samples between these two points as

$$s(x, p_m^1, p_m^y, f) = \frac{x - p_m^x}{p_m^x - p_m^2} s(p_m^x, p_m^y, f)
+ \frac{p_m^x - x}{p_m^x - p_m^1} s(p_m^1, p_m^y, f),$$

where $p_m^x < x < p_m^2$. By using interpolation in (31) along $X$ and $Y$ directions sequentially, continuous samples of $s(x, y, f)$ are generated on the plane $z = 0$, and uniformly spaced samples are available as well.

The samples of $s(x, y, f)$ after the interpolation are denoted as $\{s[x, y, f]\}$. Since multi-tone signals are used, the samples are naturally discrete and uniformly spaced in frequency. It is worthwhile to notice that common interpolation methods, e.g., linear, spline and polynomial interpolations are all capable of accurate resampling here. We simply use linear interpolation for low-complexity.
Moreover, to preserve complete information in the frequency domain with discrete samples, the distances between two adjacent receive antennas along \( X \) or \( Y \) directions are both required to be less than \( \frac{c}{2(1+f_k)} \) according to the Nyquist sampling criterion, which will be elaborated later in Sec. III-C. Here we simply assume that the samplings at the receiver satisfy the Nyquist criterion and the interpolation is accurate. Then the LHS of (29) with discrete inputs turns to be

\[
\text{FT}_{2D} \left( \{(s[x, y, f] \} \right) = \mathcal{S}(f^{(x)}, f^{(y)}, f),
\]

where \( \text{FT}_{2D} \) refers to the 2D discrete-time Fourier transform defined in Table I, and the output \( \mathcal{S}(f^{(x)}, f^{(y)}, f) \) is a function continuous in \( f^{(x)} \) and \( f^{(y)} \) domain while discrete in \( f \) direction. By considering \( \mathcal{S}(f^{(x)}, f^{(y)}, f) \) as a function of \( f \) according to the relation \( f = \|f\| \), the continuous function \( \{\mathcal{S}(f) | f \in \mathbb{R}^3 \} \) over the 3D frequency domain can be estimated by an interpolation [27]. Then the LHS of (29) can be approximated as

\[
\text{FT}_{2D} \left( s[x, y, f] \right) \bigg|_{f=\|f\|} \approx \mathcal{L} \left( \mathcal{S}(f^{(x)}, f^{(y)}, f) \bigg|_{f=\|f\|} \right)
= \mathcal{L} \left( \text{FT}_{2D} \left( \{s[x, y, f] \} \right) \bigg|_{f=\|f\|} \right),
\]

(33)

where \( \mathcal{L}(\cdot) \) represents the linear interpolation process [27].

On the other hand, the RHS of (29) can be represented in the discrete case as \( \text{FT}_{3D} \left( \{l_{\{x \in X \}} \} \right) \), where \( \text{FT}_{3D} \) is the 3D discrete-time Fourier transform. Then according to (33), the equality in (29) directly gives

\[
\mathcal{L} \left( \text{FT}_{2D} \left( \{s[x, y, f] \} \right) \bigg|_{f=\|f\|} \right) \approx \text{FT}_{3D} \left( \{l_{\{x \in X \}} \} \right),
\]

(34)

where the approximation in (34) comes from the interpolation process.

- **A. Fourier Transform:** According to (34), the indicator function \( l_{\{x \in X \}} \) can be estimated in a discrete form via a 3D inverse discrete-time Fourier transform as

\[
\Phi[x] = \text{FT}_{3D}^{-1} \left\{ \mathcal{L} \left( \text{FT}_{2D} \left( \{s[x, y, f] \} \bigg|_{f=\|f\|} \right) \right) \right\} \\
\approx \{l_{\{x \in X \}} \},
\]

(35)

- **B. Peak Detection:** Locations of the transmit antennas can be estimated by peak detection over the power spectrum \( \{\|\Phi[x]\|\} \) after a power normalization, namely,

\[
X = \left\{ x \bigg| \frac{\|\Phi[x]\|}{\|\Phi[x]\|_{\text{max}}} \geq \nu \right\},
\]

(36)

where \( \nu \) represents the detection threshold appropriately selected, and \( \|\Phi[x]\|_{\text{max}} \) is the maximum value of the detected power spectrum \( \|\Phi[x]\| \).

Therefore, according to the algorithm above, the multi-point TV position can be retrieved from signals over the antenna array at the receiver, for the LoS case.

**C. Resolution Analysis**

This subsection provides analysis on resolution of the multi-point position retrieved by the above algorithm. Here we consider the 3D antenna array at the receiver as an ‘equivalent aperture’ located in the \( X-Y \) plane \( z = 0 \), as shown in Fig. 4(a).

For ease of understanding, we first introduce the following terminologies.

**Definition 1** (Bandwidth): A bandwidth \( B \) in a direction of the frequency domain (at the receiver) is defined as the maximum frequency difference of the received signal along the corresponding direction in the spatial domain.

**Definition 2** (Resolution): A resolution \( \delta \), representing the positioning accuracy, is defined as the minimum distance to differentiate multiple objects, TV’s antennas in our work. The resolution is said to be better when the minimum distance is smaller. Accordingly, resolutions in \( X \) or \( Y \) directions and \( Z \) direction are respectively called the **azimuth** and **range** resolutions.

The direct relation between bandwidth \( B \) and resolution \( \delta \) is established as \( \delta = \frac{B}{c} \) [12], where \( c \) is the light speed. Based on this relation, the azimuth and range resolutions are firstly analyzed, and then sampling requirements to achieve a given resolution are provided next.

1) **Azimuth Resolution:** Suppose the detected TV position \( X \) of (36) is projected on the equivalent aperture’s center denoted by \( m_c \) as shown in Fig. 4(b). We focus on the spatial resolution in \( Y \) direction since resolutions in \( X \) and \( Y \) directions become equivalent when the aperture is a \( D \)-by-\( D \) square. Consider signals at frequency \( f_k \), the phase difference between received signal at \( m_1 \) and \( (m_1 - \Delta_y) \) is approximately \( d \frac{2\pi k}{4\sqrt{R^2+(D/2)^2}} \Delta_y \) w.r.t. a subtle distance \( \Delta_y \), corresponding to frequency \( \frac{D}{4\sqrt{R^2+(D/2)^2}} f_k \). Moreover, the phase difference between received signal at \((m_c + \Delta_y)\) and \( m_c \) is 0, corresponding to frequency 0. According to Definition 1, the bandwidth in \( f^{(y)} \) direction, denoted by \( B_y \) is approximately

\[
B_y \approx 2 \times E_k \left[ \frac{D}{\sqrt{4R^2+D^2}} f_k - 0 \right] = \frac{2D}{\sqrt{4R^2+D^2}} f_c,
\]

(37)

where \( f_c = \frac{B}{2c} \). The azimuth resolution in \( Y \) direction can be straightforward obtained as

\[
\delta_y = \frac{c}{B_y} \approx \frac{c\sqrt{4R^2+D^2}}{2f_cD},
\]

(38)
It is shown to be proportional to the term $\sqrt{4R^2 + D^2}$, meaning that higher azimuth resolution can be achieved when the TV is closer and the aperture size becomes larger.

2) **Range Resolution:** Samplings at a fixed location $\left(\frac{s}{m}, \frac{t}{m}, 0\right)$ consist of signals at all frequencies used. Thus bandwidth in the $f_z$ direction is $B_z \approx (f_K - f_1)$, and the range resolution can be obtained as

$$\delta_z = \frac{c}{B_z} \approx \frac{c}{(f_K - f_1)},$$

where $f_K$ and $f_1$ are maximum and minimum frequencies in the multi-tone waveforms specified in (10).

**Remark 9 (Sampling Requirements):** To achieve the above resolutions, there exist two kinds of sampling requirements on spatial and frequency domains.

- **Spatial Sampling:** The spatial sampling represents the distance between two adjacent receive antennas. To achieve the resolution in (38), the receive antennas deployment over the “equivalent aperture” needs to meet the Nyquist sampling criterion such that the required sampling intervals are less than $\Delta x$ and $\Delta y$ to avoid aliasing. Therefore, the distances between two adjacent receive antennas along $X$ or $Y$ directions are

$$\Delta x = \Delta y \leq \min_R \left\{\frac{c}{2f_c} \sqrt{4R^2 + D^2} \right\} (a) \frac{c}{4f_c},$$

where (a) follows for the worst case with $R = 0$.

- **Frequency sampling:** The frequency sampling interval refers to the frequency gap $\Delta$ to achieve the maximum ranging distance $R_{\text{max}}$, given the resolution $\delta_z$ in (39). Specifically, the number of the minimum samples for the resolution $\delta_z$ is given as $\frac{R_{\text{max}}}{\delta_z}$, which provides the following upper bound as

$$\frac{R_{\text{max}}}{\delta_z} \leq \frac{f_K - f_1}{\Delta}.$$  \hfill (41)

Plugging (39) into (41) and with some manipulations, we have $\Delta \leq \frac{c}{R_{\text{max}}^2}$.

**D. Propagation Loss Analysis**

The signal propagation loss determines the power level of the received signals, so as the performance of the detected position against noise. In this subsection, we analyze the propagation loss of the proposed COMPOP and existing multi-point positioning techniques, i.e. RADAR or LIDAR, over a LoS link for signal power comparison. It is obvious that the proposed technique experiences less power loss than RADAR-based techniques due to the half propagation distance. Besides, the orientation of the target surface is another factor to affect the reflected signal power as shown in Fig. 5(a).

On the other hand, according to (42), the received signal power of our proposed technique can be expressed as $P_r = \frac{P_t G_t \lambda^2}{(4\pi)^2 R^2}$, where the RCS has no influence on the received signal power of our proposed COMPOP since the transmit antennas are isotropic. Moreover, reducing the signal propagation distance to half brings a power gain of $4\pi R^2$. Therefore, the proposed COMPOP is able to function well regardless the orientation of the TV and guarantee a signal power gain compared with conventional RADAR systems.

**IV. COMPOP USING SURFACE REFLECTION IN NLoS**

Consider a NLoS scenario where the TV’s LoS path to the SV is blocked but several NLoS paths reflected by nearby vehicles are available. We exploit mmWave spectrum’s specular reflection on the smooth surfaces of nearby vehicles, establishing mirror-reflection links between the TV and the SV. Based on signals reflected from different paths, given as (16), multiple virtual TVs symmetric to the actual TV can be detected via LoS COMPOP. Since SV has no prior information of the reflection surfaces, an additional step should be required to combine multiple virtual TVs into the actual one. To this end, we design a two-stage approach following 1) LoS COMPOP and 2) virtual TV combining, as illustrated in Fig. 6.

Additional assumptions used in this section are summarized for clarification in the sequel:

3) mmWave spectrum has two reflection properties on vehicles’ metal surfaces, where 1) its reflection coefficient is almost 1 unless the incident angle is near $\frac{\pi}{2}$ or 0, and 2) mmWave signals experience the mirror-like specular reflection [30].
Satisfying the following equations.

The surfaces of nearby vehicles or buildings are able to provide specular reflections for mmWave signals as mentioned above.

**Reflection Surfaces Vertical to the Ground:** The reflection surfaces are vertical to the ground. This assumption can be easily relaxed as illustrated in Remark 12 later.

### A. LoS COMPOP

Firstly, we assume that the signals reflected from different paths can be perfectly resolved at the SV by the AoA differentiation. Then based on signals received from each path, the LoS COMPOP can be accomplished by the following two steps.

1) **Synchronization:** The operation is similar to the synchronization under LoS in Sec. III-A. Consider the signals from path \( \ell \) as an example. Let \( \sigma \) denote the system clock difference satisfying the following equations:

\[
\sigma = \tau_{a,m}^{(\ell)} - \frac{\tau_{b,m}^{(\ell)}}{2\pi \Delta} - \tau_{b,m}^{(\ell)} - \frac{\tau_{b,m}^{(\ell)}}{2\pi \Delta},
\]

where \( \{\tau_{a,m}^{(\ell)}, \tau_{b,m}^{(\ell)}\} \) and \( \{\tau_{b,m}^{(\ell)}, \tau_{b,m}^{(\ell)}\} \) represent the propagation time and phase difference of the signature waveforms from the antennas \( a \) and \( b \) through the \( \ell \)-th mirror-reflection link, respectively. Following the procedures in Sec. III-A gives the locations of antennas \( a \) and \( b \) on the virtual TV \( \ell \), denoted by \( x_a^{(\ell)} \) and \( x_b^{(\ell)} \), which are different from the real locations \( x_a \) and \( x_b \). As in the case with an LoS path, the points \( x_a^{(\ell)} \) and \( x_b^{(\ell)} \) are used to compensate the system clock difference, which enables virtual positioning in Sec. IV-A.2. Besides, \( x_a^{(\ell)} \) and \( x_b^{(\ell)} \) help estimate the reflection surface elaborated in Sec. IV-B.2.

**Remark 10 (TV Recognition in NLoS):** Considering that multiple reflection links exist for the same TV in NLoS case, the detected system clock difference can help recognize the TVs. Specifically, signature waveforms transmitted from the same TV share the same system clock difference \( \bar{\sigma} \) regardless of different signal paths. Therefore, the SV can recognize signals from the same TV or different TVs according to the detected system clock difference \( \bar{\sigma} \).

2) **Virtual TV Positioning:** Recall that plugging the perfectly estimated \( \sigma \) into the signal (16) gives the following 3D surface integral form of \( y_{m}^{(\ell)} \) as

\[
y_{m}^{(\ell)} = \frac{1}{(\pi^2)^{3/2}} \int_{\mathbb{R}^3} \Gamma^{(\ell)}(x, p_m) \exp \left( \frac{-2\pi i}{c} D^{(\ell)}(x, p_m) \right) dV,
\]

where \( y_{m}^{(\ell)} : \mathbb{R}^3 \to \mathbb{R} \) is a continuous function, and \( D^{(\ell)}(x, p_m) \) represents the Euclidean distance from point \( x^{(\ell)} \) in \( X^{(\ell)} \), which is symmetric to point \( x \) in \( X \) w.r.t. the \( \ell \)-th reflection surface, to the location of SV's antenna \( m \) denoted by \( p_m = (p_{m,1}, p_{m,2}, p_{m,3})^T \). Compared with (28) in the LoS scenario, the reflection coefficient \( \Gamma^{(\ell)} \) can be considered as a constant scaling factor, which does not affect the positioning procedure. Following the same steps in Sec. III-B, the virtual TV \( \ell \)’s power spectrum \( M[x^{(\ell)}] \) can be calculated. After the peak detection in (36), we can obtain the virtual locations of the transmit antennas \( X^{(\ell)} \), which is referred to as virtual TV \( \ell \). The estimated virtual TVs are used to position the actual TV explained in Sec. IV-B.2.

### B. Virtual TV Combining

1) **Overview:** This subsection aims at positioning the actual TV with multiple virtual TVs detected in Sec. IV-A.2 under the assumption that the reflection surfaces are vertical to the ground, which is common in practice. Since the actual TV can be directly obtained by shifting the virtual TVs w.r.t. their corresponding reflection surfaces, the key step becomes the detection of the reflection surfaces. To this end, an algorithm is proposed to position the actual TV, consisting of two steps: 1) detection of reflection surfaces; and 2) combining virtual TVs. The detailed procedures are illustrated in the following.

2) **Algorithm Description:**

- **A. Detection of Reflection Surfaces:** Consider the virtual locations of representative transmit antennas \( a \) and \( b \) detected in Sec. IV-A.1 with coordinates \( x_a^{(\ell)} = (x_{a,1}^{(\ell)}, y_{a,1}^{(\ell)}, z_{a,1}^{(\ell)}) \) and \( x_b^{(\ell)} = (x_{b,1}^{(\ell)}, y_{b,1}^{(\ell)}, z_{b,1}^{(\ell)}) \), which are symmetric to the counterpart points on the actual TV denoted by \( x_a = (x_{a,2}, y_{a,2}, z_{a,2})^T \) and \( x_b = (x_{b,2}, y_{b,2}, z_{b,2})^T \) w.r.t. the reflection surface \( \ell \). It can be inferred from Fig. 7 that given the locations \( (x_a^{(\ell)}, x_b^{(\ell)}) \), the reflection surfaces can be straightforwardly obtained if \( (x_a, x_b) \) is estimated. Therefore, the problem is translated into the detection of representative transmit antennas \( (x_a, x_b) \). To this end, we summarize the geometric relation between \( (x_a^{(\ell)}, x_b^{(\ell)}) \) and \( (x_a, x_b) \) in the following lemma.

**Lemma 2:** Consider virtual TVs \( \ell_1 \) and \( \ell_2 \) whose representative points are \( \{x_a^{(\ell_1)}, x_b^{(\ell_1)}\} \) and \( \{x_a^{(\ell_2)}, x_b^{(\ell_2)}\} \) respectively, which have geometric relations with the counterpart points on the actual TV denoted by \( (x_a, x_b) \) as follows.

1) Let \( \theta_1 \) denote the directed angle from the \( X \)-axis (parallel to the ground) to the line between \( x_b^{(\ell_1)} \) and \( x_a^{(\ell_1)} \) in \( X \). According to the geometric relation illustrated in Fig. 7, \( x_b \) can be given in terms of \( \theta_1 \) and \( \theta_2 \) as (46) shown at the bottom of the next page, and similarly the expression of \( x_b \) in terms of \( \theta_1 \) and \( \theta_2 \) is obtained by replacing all \( a \) in (46) with \( b \). Let \( \phi_1 \) denote the directed angle from the \( X \)-axis to the line segment of virtual TV \( \ell_1 \), from \( x_a^{(\ell_1)} \) to \( x_b^{(\ell_1)} \), as shown in Fig. 7. The angles \( \theta_1 \) and \( \phi_1 \) of two
virtual TVs $\ell_1$ and $\ell_2$ follows the relation

$$\theta_{\ell_1} - \theta_{\ell_2} = \frac{\phi_{\ell_1} - \phi_{\ell_2}}{2}. \quad (47)$$

**Proof:** Please refer to Appendix C. \qed

Based on Lemma 2, the representative points $(x_a, x_b)$ are estimated as follows. First, all angles \{\theta_{\ell}\}_{\ell=2}^3 can be expressed in terms of $\theta_1$ as $\theta_{\ell} = \theta_1 + \frac{\phi_{\ell} - \phi_{\ell_1}}{2}$, using (47). Next, plugging each pair of $(\theta_{\ell_1}, \theta_{\ell_2})$ into (46) makes it possible to express the locations $a$ and $b$ w.r.t. $\theta_1$, denoted by $(v_{a}^{(\ell)}(\theta_1), v_{b}^{(\ell)}(\theta_1))$. If $\theta_1$ is correct, $(v_{a}^{(\ell)}(\theta_1), v_{b}^{(\ell)}(\theta_1))$ naturally coincides with $(x_a, x_b)$. In other words, estimating $(x_a, x_b)$ is translated into finding $\theta_1$ minimizing the following squared Euclidean distance as

$$\theta_1^* = \arg \min_{\theta_1} \sum_{\ell=2}^3 \sum_{\ell=2}^3 \left( \left\| v_{a}^{(\ell_1)}(\theta_1) - v_{b}^{(\ell_2)}(\theta_1) \right\|^2 + \left\| v_{b}^{(\ell_1)}(\theta_1) - v_{b}^{(\ell_2)}(\theta_1) \right\|^2 \right). \quad (48)$$

The optimal $\theta_1^*$ is computed by 1D search over $[-\pi, \pi]$, and the resultant $x_a^* = \frac{1}{L_1-1} \sum_{\ell=2}^L v_{a}^{(\ell)}(\theta_1^*)$ and $x_b^* = \frac{1}{L_1-1} \sum_{\ell=2}^L v_{b}^{(\ell)}(\theta_1^*)$ can be directly obtained by using the optimal $\theta_1^*$. Then with $(x_a^*, x_b^*)$, the reflection surface $\ell$, which is located on the middle of $x_a^*$ and $x_b^*$, can be expressed by a line because it is perpendicular to $X-Z$ plane such that

$$z = -\frac{1}{\cos(\theta_1^*)} \left( x - x_a^* \right) + \frac{z_a^*}{2}, \quad (49)$$

where $x_a^* = (x_a^*, y_a^*, z_a^*)^T$, and $\phi_a^* = \phi_a - \frac{\phi_{\ell} - \phi_{\ell_1}}{2}$. \qed

**Proposition 3 (Feasibility Condition for NLoS Position Combining):** To detect the actual TV in NLoS, at least three reflection surfaces are required: $L \geq 3$.

**Proof:** Please refer to Appendix D. \qed

**B. Combining Virtual TVs:** With the knowledge of the reflection surfaces, the actual TV $\mathcal{X}$ can be obtained as elaborated in the following proposition.

**Proposition 4 (Position Combining):** Consider the virtual TV $\ell$ represented by $\mathcal{X}^{(\ell)}$. Given $\theta_1^*$ and $x_a^*$ (or $x_b^*$), the actual TV $\mathcal{X}$ can be obtained by the following mapping function:

$$\mathcal{X} = \frac{1}{L} \sum_{\ell} G^{(\ell)}(\mathcal{X}^{(\ell)}) \quad (50)$$

where function $G^{(\ell)}(\cdot)$ is given in (51), shown at the bottom of the page.

**Proof:** Please refer to Appendix E. \qed

**Remark 11 (Existence of LoS path):** The LoS case is a special realization of the NLoS case, where one couple of representative points $(x_{a,LoS}^{(\ell)}, x_{b,LoS}^{(\ell)})$ are equivalent to the exact location $(x_a, x_b)$. Therefore, all mathematical expressions for the virtual TV combining still hold when the LoS link exists, and the resultant $(x_a^*, x_b^*)$ can be obtained in the same way.

**Remark 12 (Arbitrary Reflection Surfaces):** The assumption that all reflection surfaces are vertical to the ground can be easily relaxed with one more antenna transmitting signature waveform $s_c$. In this case, representative points $(x_a^{(\ell)}, x_b^{(\ell)}, x_c^{(\ell)}) \forall \ell \notin \mathcal{E}$ are detected in the clock synchronization step. Analogous to the result in (49), reflection surfaces can then be located, without the assumption, by exploiting the information of these representative points, so as the actual TV position.

**C. Propagation Loss Analysis**

Similar to the analysis in Sec. III-D, we also give signal power comparison between the proposed COMPOP and

\begin{align*}
&x_a = \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} = \begin{pmatrix} \frac{z_a^2 - z_a^{(\ell_2)}}{\tan(\theta_{\ell_2}) - \tan(\theta_{\ell_1})} \\ \frac{y_a^{(\ell_1)} - y_a^{(\ell_2)}}{\tan(\theta_{\ell_2}) - \tan(\theta_{\ell_1})} \\ \frac{z_a^2 + z_a^{(\ell_2)}}{\tan(\theta_{\ell_2}) - \tan(\theta_{\ell_1})} \end{pmatrix} \quad (46)
\end{align*}

\begin{align*}
G^{(\ell)}(x^{(\ell)}) = (x^*, y^*, z^*)^T &= \begin{pmatrix} x^{(\ell)} + \frac{\tan(\theta_1^*)}{1 + \tan^2(\theta_1^*)} (x_a^{(\ell)} + x_a^*) + z_a + z_a^* - \frac{2x_a^{(\ell)} + z_a^{(\ell)}}{\tan(\theta_1^*)} - 2z_a^{(\ell)} \\ y^{(\ell)} + \frac{\tan^2(\theta_1^*)}{1 + \tan^2(\theta_1^*)} (x_a^{(\ell)} + x_a^*) + z_a + z_a^* - 2y_a^{(\ell)} \tan(\theta_1^*) - 2z_a^{(\ell)} \\ z^{(\ell)} + \frac{\tan(\theta_1^*)}{1 + \tan^2(\theta_1^*)} (x_a^{(\ell)} + x_a^*) + z_a + z_a^* - \frac{2x_a^{(\ell)} + z_a^{(\ell)}}{\tan(\theta_1^*)} - 2z_a^{(\ell)} \end{pmatrix}, \quad x^{(\ell)} \in \mathbb{R}^3
\end{align*}

(51)
traditional RADAR techniques over the reflection links. Although RADAR techniques do not provide solutions to positioning without a LoS, we still provide the received signal power of a reflection link, say the \( \ell \)-th reflection link, as

\[
P_r^{(\ell)} = \frac{P_i G_i \lambda^2}{64\pi^3 (R_1^{(\ell)})^2 (R_2^{(\ell)})^2} \sigma(\Theta^{(\ell)}),
\]

where \( \Theta^{(\ell)} \) is the incident angle on the reflection surface \( \ell \), \( R_1^{(\ell)} \) is the distance from the SV to the reflection surface \( \ell \), and \( R_2^{(\ell)} \) is the distance from the TV to the reflection surface \( \ell \). For the proposed COMPOP, signals from each reflection link give one virtual TV detected with algorithm in Sec. IV-A.2. For the reflection link \( \ell \), the received signal power is

\[
P_r^{(\ell)} = \frac{P_i G_i \lambda^2}{4\pi^3 (R_1^{(\ell)})^4 (R_2^{(\ell)})^4} \sigma(\Theta^{(\ell)}),
\]

It is obvious that the proposed COMPOP achieves a much lower propagation loss than the conventional RADAR techniques.

### V. Simulation Results

In this section, the performance of the proposed COMPOP technique is evaluated by realistic settings. Signature waveforms at 4 frequencies are used for the synchronization procedure. The number of frequencies used in the multi-tone waveform (10) is \( K = 256 \) ranging from 57 GHz to 60 GHz with the constant gap \( \Delta = 11.72 \text{ MHz} \). The four frequencies used in the signature waveforms (1) are \((57 - \Delta \cdot i)\) GHz where \( i = 1, 2, 3, 4 \). The numbers of the TV’s and SV’s antennas are \( N_t = N_r = 200 \), which are uniformly deployed on the vehicles bodies. The size of the equivalent receive aperture is \( 1 \times 1 \text{ m}^2 \). The SNR of each received signal is fixed to 10 dB. The number of reflection surfaces are 3 and the distance between TV and SV is 8 m unless stated otherwise.

For the performance metric, we use the Hausdorff distances defined as follows.

**Definition 3 (Hausdorff Distance [31]):** The Hausdorff distance is widely used to evaluate the image retrieval performance by measuring the similarity between two images. Consider an image \( A \) and its retrieved one \( B \), both of which are composed of discrete points. The Hausdorff distance is defined as

\[
H(A, B) = \max \left( h(A, B), h(B, A) \right),
\]

where \( h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b|| \).

**A. Graphical Example of Multi-Point Positioning**

This subsection aims at explaining the entire vehicular positioning process with step-by-step graphical examples. We consider the topology with three reflection surfaces illustrated in Fig. 8. The TV is represented by discrete points, each of which is one TV’s antenna. The TV size is \( 3 \times 1 \times 0.6 \text{ m}^3 \). The equivalent receive aperture is parallel to Y-axis and located at \((0, 0, 0)\). The equations of three reflection surfaces are given as \( z = 1.02x + 3 \), \( z = \frac{x+13}{4} \), and \( z = 3x + 4 \). Using the reflection surfaces, the SV detects three virtual TVs represented by yellow slots in Fig. 9, each of which is differentiable using AoA information. By the intelligent combining algorithm in Proposition 4, each virtual TV can be shifted to its real location represented by green spots, of which the Hausdorff distance is 0.355m, which is relatively small compared to the size of the TV. After graphical rendering process, the final detected position is obtained as in Fig. 10(b) that is similar to the original one in Fig. 10(a).

**B. Error in Clock Synchronization**

The clock difference detection error is omitted after clock synchronization. To justify such an assumption, we check it by simulations given in Fig. 11. It clearly shows that the clock difference detection error is well suppressed by an appropriate number of antennas, e.g., \( N_t = 16 \) or \( N_r = 64 \), at the receiver. Therefore, it is reasonable to assume perfect clock synchronization.

**C. Effect of Distance Between SV and TV**

In Fig. 12, the Hausdorff distances are given under different TV-SV distances, showing that the positioning quality degrades as the TV-SV distance increases. This phenomenon can be explained by (38), where the spatial resolution is poor with large detection range \( R \). Moreover, the relation between the Hausdorff distance and the TV-SV distance is...
not linear due to the $\tan(\cdot)$ function in the solution (46). With noise in consideration, the Hausdorff distance, indicating the comprehensive error level, increases faster when the TV-SV distance becomes larger.

D. Effect of Reflection Surface Number

The relation between the performance and the number of reflection surfaces is also presented in Fig. 12 (a). As explained in Sec. IV-B, signals reflected from any three reflection surfaces give one estimation of $\theta_1$, but the resultant positioning quality is low due to the phase error, and the performance becomes unstable when the TV-SV distance increases. We also consider the cases where more than 3 reflection links exist ($L > 3$). Larger $L$ provides more combinations to estimate $\theta_1$, resulting in more accurate estimation of $\theta_1$. Moreover, the LoS case leads to the minimum Hausdorff distance. The case with mixed LoS and NLoS paths are also investigated here, which is plotted by dotted lines. It can be observed that although the existence of LoS path enhances the performance, the errors from NLoS paths hamper the accuracy.

E. Comparison to MIMO RADAR

The performance of conventional MIMO RADAR [12], [14] is also checked by simulations as shown in Fig. 12 (b), which is considered as a baseline of the proposed COMPOP design. The conventional MIMO RADAR is naturally synchronized with co-located transceivers and thus free from the clock difference detection error. Hence, conventional MIMO RADAR could be a better choice in short distance. However, when the distance becomes larger, the proposed COMPOP has a signal power gain and thus shows better performance as illustrated in Sec. III-D.

VI. CONCLUSION

In this article, a novel Cooperative Multi-point Positioning (COMPOP) approach via mmWave signal transmissions has been proposed to capture the shape and location information of the TVs in both LoS and NLoS. The cooperative transmission between the TV and SV enables the real-time COMPOP without scanning process. The synchronization issue due to transceiver separation has been well addressed by a PDoA-based positioning approach. In NLoS case, COMPOP establishes mirror-reflection links between the TV and SV under the assistance of the nearby vehicles. The geometric relation between the virtual and actual TVs has been exploited to position the actual TV without prior knowledge on the nearby vehicles. In conclusion, the proposed COMPOP is available in both LoS and NLoS situations with ultra-low latency and high accuracy, which is challenging for existing vehicular sensing techniques (e.g., RADAR and LIDAR). Therefore, this technique opens a new area of mmWave-based vehicular positioning and sensing. We believe the proposed COMPOP contributes to more intelligent and safer autonomous driving, and the potential of mmWave-based vehicular sensing can still be activated in the future.

APPENDIX

A. Proof of Proposition 2

With Gaussian phase error, the covariance of the location estimation can be expressed as $\text{cov}(\tilde{x}_a) = X\sigma^2_2 I_3$, where $X$ is a 3-by-3 matrix as (55), shown at the bottom of the next page, where the inverse matrix’s each component converges to its expectation as $N_r$ becomes large, which becomes independent to $N_r$. In other words, $X$ is proportional to $\frac{1}{N_r-1}$.

B. Proof of Lemma 1

We adopt the scalar diffraction idea in [26] that the wave field can be decomposed as an infinite integral of planar waves, given as (56), shown at the bottom of the next page, where $f = [f(x), f(y), f(z)]^T$ denotes a spatial frequency vector. Accordingly, the signal in (28) can be expanded by rewriting the exponential term in terms of $f$ as (57), shown at the bottom of the next page. Then consider the received signal at the 2D plane $z = 0$ where $p_m = (p_m^x, p_m^y, 0)^T$, given as (58), shown at the bottom of the next page.

C. Proof of Lemma 2

In Fig. 7, we have $\theta_{\ell_1} \in [-\pi, \pi]$ and the equations of line $l_{x_1(\ell_1)}$ and $l_{x_2(\ell_2)}$ are

$$
l_{x_1(\ell_1)} : \quad z = z_{2(\ell_1)} + \tan(\theta_{\ell_1}) (x - x_{2(\ell_1)}) ;
$$

$$
l_{x_2(\ell_2)} : \quad z = z_{2(\ell_2)} + \tan(\theta_{\ell_2}) (x - x_{2(\ell_2)}) .
$$
Thus the intersection point can be solved as

\[
\begin{align*}
\begin{cases}
x_a = \left( z_a^{(e)} - z_a^{(o)} \right) + x_a^{(o)} \tan(\theta_{t_e} - \theta_{t_o}) \\
\hat{z}_a = z_a^{(e)} + \tan(\theta_{t_e}) (x_a - x_a^{(e)})
\end{cases}
\end{align*}
\]  
(59)

The general relations in (46) can be derived similarly. Due to the symmetric relation between the virtual and actual TVs, we have

\[
\pi - \varphi = \varphi_{t_e} - 2\theta_{t_e} = \varphi_{t_o} - 2\theta_{t_o}
\]  
(60)

D. Proof of Proposition 3

Plugging (60) into (59), \( x_a \) can be simplified as

\[
x_a = \left( z_a^{(e)} - z_a^{(o)} \right) + x_a^{(o)} \tan(\theta_{t_e} - \frac{\varphi_{t_e} - \varphi_{t_o}}{2}) - x_a^{(e)} \tan(\theta_{t_e})
\]  
\[
\left( \frac{\varphi_{t_e} - \varphi_{t_o}}{2} - \tan(\theta_{t_e}) \right)
\]  
(61)

which indicates that \( x_a \) is only determined by the angle \( \theta_{t_e} \in [-\pi, \pi] \). Given the angle \( \theta_{t_e} \), the estimation of \( x_a \) from (61) is denoted as \( \hat{x}_a \). Similarly, another estimation of \( x_a \) can be obtained from the common points \((x_a^{(e)}, x_a^{(o)})\) and \((x_b^{(e)}, x_b^{(o)})\), which is given as

\[
x_a = \left( z_a^{(e)} - z_a^{(o)} \right) + x_a^{(o)} \tan \left( \theta_{t_e} - \frac{\varphi_{t_e} - \varphi_{t_o}}{2} \right) - x_a^{(e)} \tan \left( \theta_{t_e} \right)
\]  
\[
\left( \frac{\varphi_{t_e} - \varphi_{t_o}}{2} - \tan(\theta_{t_e}) \right)
\]  
(62)

The estimation of \( x_a \) from (62) is denoted as \( \hat{x}_b \), and it can be observed from (62) that \( \hat{x}_a \) is also only determined by \( \theta_{t_e} \). Therefore, the SV can search \( \theta_{t_e} \) in the range of \([-\pi, \pi]\) to minimize \( |\hat{x}_a - \hat{x}_b| \). Two solutions can be obtained with the optimal \( \theta_{t_k} \), denoted as \( \hat{x}_a^* \) and \( \hat{x}_b^* \), then the optimal solution of \( x_a \) is given as \( x_a^* = \frac{\hat{x}_a^* + \hat{x}_b^*}{2} \). With \( x_a^* \), the location \( x_a \) can be calculated according to (46), so as \( x_b \). Therefore, three virtual TVs are enough to detect the actual TV.

E. Proof of Proposition 4

Based on the line function (49), as well as the symmetric geometry relation between virtual TV \( \ell \) and the actual one,
it is easy to establish the mathematical relation between \( x^{(l)} \) and \( x \) as

\[
\begin{align*}
\{ x &= x^{(l)} + \Delta x^{(l)} \\
\theta &= \theta^{(l)} + \tan(\theta^{(l)}) \cdot \Delta x^{(l)}
\end{align*}
\]  

(63)

where \( \Delta x^{(l)} \) is the \( z \)-direction projection of distance between \((x, z)\) and \((x^{(l)}, z^{(l)})\). The middle point of \((x, z)\) and \((x^{(l)}, z^{(l)})\) locates at the surface \( \ell \), meaning that the middle point \( x^{(l)} + \frac{\Delta a^{(l)}}{2}, z^{(l)} + \tan(\theta^{(l)}) \cdot \frac{\Delta a^{(l)}}{2} \) should satisfy function (49). Thus we have

\[
\begin{align*}
z_a^{(l)} + z_a - 1 = \frac{\Delta a^{(l)}}{\tan(\theta^{(l)})} \left[ 2x^{(l)} + \Delta x^{(l)} - x_a^{(l)} - x_a \right] = 2z_a^{(l)} + \tan(\theta^{(l)}) \cdot \Delta x^{(l)}.
\end{align*}
\]

(64)

The result derived from (64) is given as

\[
\begin{align*}
\Delta x_a &= \frac{\tan(\theta^{(l)})}{1 + \tan^2(\theta^{(l)})} \left[ a^{(l)} + a_a^{(l)} + a_a^{(l)} - \frac{2x^{(l)}}{\tan(\theta^{(l)})} - x_a^{(l)} - x_a \right].
\end{align*}
\]

(65)

Bring (65) into (63), the result (50) can be obtained by replacing \((x, z)\) with \((x^*, z^*)\).

REFERENCES

[1] Z. Zhang, S.-W. Ko, R. Wang, and K. Huang, “Millimeter-wave multi-point vehicular positioning for autonomous driving,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[2] H. Li, “Landscape detection by leveraging millimeter wave communication signals,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[3] S.-W. Ko, H. Chae, K. Han, S. Lee, D.-W. Seo, and K. Huang, “V2X-based vehicular positioning: Opportunities, challenges, and future directions,” *arXiv:1908.04606*, [Online]. Available: http://arxiv.org/abs/1908.04606.

[4] K. Liu, H. B. Lim, E. Frazzoli, H. Ji, and V. C. S. Lee, “Improving positioning accuracy using GPS pseudorange measurements for cooperative vehicular localization,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2544–2556, Jul. 2014.

[5] F. Gustafsson and F. Gunnarsson, “Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements,” *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 41–53, Jul. 2005.

[6] S. Zhu and Z. Ding, “Distributed cooperative localization of wireless sensor networks with convex hull constraint,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2150–2161, Jul. 2011.

[7] S. Gezici et al., “Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks,” *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, Jul. 2005.

[8] J. Shen, A. F. Molisch, and J. Salmi, “Accurate passive location estimation using TOA measurements,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2182–2192, Jun. 2012.

[9] K. Fujii, Y. Sakamoto, W. Wang, H. Arie, A. Schmitz, and S. Sugano, “Hyperbolic positioning with antenna arrays and multi-channel pseudolite for indoor localization,” *Sensors*, vol. 15, no. 10, pp. 25157–25175, Sep. 2015.

[10] K. Han, S.-W. Ko, H. Chae, B.-H. Kim, and K. Huang, “Hidden vehicle sensing via asynchronous V2 V transmission: A multi-path-geometry approach,” *IEEE Access*, vol. 7, pp. 169399–169416, 2019.

[11] L. Yujiri, M. Shoucri, and P. Molla, “Passive millimeter wave imaging,” *IEEE Microw. Mag.*, vol. 4, no. 3, pp. 39–50, Sep. 2003.

[12] D. M. Sheen, D. L. Mcmakin, and T. E. Hall, “Three-dimensional millimeter-wave imaging for concealed weapon detection,” *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 9, pp. 1581–1592, Sep. 2001.

[13] C. Nguyen and J. Park, *Stepped-Frequency Radar Sensors: Theory, Analysis and Design*. Cham, Switzerland: Springer, 2016.

[14] X. Zhu and A. G. Yarovoy, “Three-dimensional near-field MIMO array imaging using range migration techniques,” *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3026–3033, Jun. 2012.

[15] H. V. Duong, M. A. Lefsky, T. Ramond, and C. Weiner, “The electronically steerable flash lidar: A full waveform scanning system for topographic and ecosystem structure applications,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4809–4820, Nov. 2012.

[16] R. H. Rasshofer, M. Spies, and H. Spies, “Influences of weather phenomena on automotive laser radar systems,” *Adv. Radio Sci.*, vol. 9, pp. 49–60, Jul. 2011.

[17] D. Maturana and S. Scherer, “VoxNet: A 3D convolutional neural network for real-time object recognition,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 922–928.

[18] J. Li, X. Zhang, R. Cao, and M. Zhou, “Reduced-dimension MUSIC for angle and array gain-phase error estimation in bistatic MIMO radar,” *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 443–446, Mar. 2013.

[19] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[20] S. Lambot, E. C. Slob, I. van den Bosch, B. Stockbroeckx, and M. Vancoostr, “Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2555–2568, Nov. 2004.

[21] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[22] H. Guo, K.-S. Low, and H.-A. Nguyen, “Optimizing the localization of a wireless sensor network in real time based on a low-cost microcontroller,” *IEEE Trans. Ind. Electron.*, vol. 58, no. 3, pp. 741–749, Mar. 2011.

[23] W. F. Mascarenhas, “The divergence of the BFGS and gauss Newton methods,” *Math. Program.*, vol. 147, nos. 1–2, pp. 253–276, Oct. 2014.

[24] M. Weiss, “Synchronisation of bistatic radar systems,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 3, Sep. 2004, pp. 1750–1753.

[25] Y.-C. Wu, Q. Chaudhari, and E. Serpedin, “Clock synchronization of wireless sensor networks,” *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 124–138, Jan. 2011.

[26] O. K. Ersoy, *Diffraction, Fourier Optics and Imaging*. Hoboken, NJ, USA: Wiley, 2006, ch. 4, pp. 41–62.

[27] P. Davis, *Interpolation and Approximation* (Dover Books on Mathematics). New York, NY, USA: Dover, 1975.

[28] M. I. Skolnik, *Introduction to Radar Systems*, 2nd ed. Singapore: McGraw-Hill, 1980.

[29] D. Barrick, “Rough surface scattering based on the specular point theory,” *IEEE Trans. Antennas Propag.*, vol. 16, no. 4, pp. 449–454, Jul. 1968.

[30] K. Sato et al., “Measurements of reflection and transmission characteristics of interior structures of office building in the 60-GHz band,” *IEEE Trans. Antennas Propag.*, vol. 45, no. 12, pp. 1783–1792, Dec. 1997.

[31] D. P. Huttonlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.
Rui Wang (Member, IEEE) received the B.S. degree from the University of Science and Technology of China in 2004 and the Ph.D. degree in wireless communications from The Hong Kong University of Science and Technology in 2008.

From 2009 to 2012, he was a Senior Research Engineer with Huawei Technologies, Co., Ltd. Since 2012, he has been an Associate Professor with the Southern University of Science and Technology, China. He has research experience in both academia and industry. He has contributed over 20 U.S. patent applications and over 30 Chinese patent applications (20 of them have been granted). He has authored over 30 articles in top-level IEEE journals and flagship international conferences, especially in the area of wireless radio resource optimization and interference management. He also involved in the development of interference mitigation technology for 5G systems.

Kaibin Huang (Fellow, IEEE) received the B.Eng. and M.Eng. degrees from the National University of Singapore and the Ph.D. degree from The University of Texas at Austin, all in electrical engineering. He is currently an Associate Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. He received the Outstanding Teaching Award from Yonsei University, South Korea, in 2011. He also received the 2015 Asia-Pacific Best Paper Award, the IEEE Communication Society’s 2019 Best Tutorial Paper Award, and the 2019 Asia-Pacific Outstanding Paper Award, and Best Paper Award from the IEEE/CIC ICCC 2018 and IEEE GLOBECOM 2006. He has served as the Lead Chair for the Communications Theory Symposium of the IEEE GLOBECOM 2014 and the Wireless Communications Symposium of the IEEE GLOBECOM 2017 and the TPC Co-Chair for the IEEE CTW 2013 and IEEE PIMRC 2017. He has also served on the Editorial Board of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has guest edited special issues for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC), the IEEE JOURNAL ON SELECTED TOPICS ON SIGNAL PROCESSING, and the IEEE Communications Magazine. He is an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE JSAC and an Area Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is an IEEE Distinguished Lecturer and an ISI Highly Cited Researcher.