High Accurate Time-of-Arrival Estimation With Fine-Grained Feature Generation for Internet-of-Things Applications

Guangjin Pan, Tao Wang, Senior Member, IEEE, Shunqing Zhang, Senior Member, IEEE, and Shugong Xu, Fellow, IEEE

Abstract—Conventional schemes often require extra reference signals or more complicated algorithms to improve the time-of-arrival (TOA) estimation accuracy. However, in this letter, we propose to generate fine-grained features from the full band and resource block (RB) based reference signals, and calculate the cross-correlations accordingly to improve the observation resolution as well as the TOA estimation results. Using the spectrogram-like cross-correlation feature map, we apply the machine learning technology with decoupled feature extraction and fitting to understand the variations in the time and frequency domains and project the features directly into TOA results. Through numerical examples, we show that the proposed high accurate TOA estimation with fine-grained feature generation can achieve at least 51% root mean square error (RMSE) improvement in the static propagation environments and 38 ns median TOA estimation errors for multipath fading environments, which is equivalently 36% and 25% improvement if compared with the existing MUSIC and ESPRIT algorithms, respectively.

Index Terms—Time of arrival estimation, neural networks, Internet-of-Things, narrowband, multipath channels.

I. INTRODUCTION

VARIOUS localization applications, together with the increasing number of smart terminals, trigger the rapid progress in the positioning technologies. Traditional global navigation satellite system (GNSS) and inertial navigation system (INS) are able to provide a continuous localization accuracy up to sub-meter level or even centimeter level with sufficient power supply. The recently proposed schemes, including enhanced Cell ID (E-CID), direction of arrival (DOA), observed time difference of arrival (OTDOA) [1], and fingerprint based solution [2], provide alternative approaches including enhanced Cell ID (E-CID), direction of arrival (DOA), observed time difference of arrival (OTDOA) [1], and fingerprint based solution [2], provide alternative approaches.

Among the above cellular signal based schemes, OTDOA only requires a single antenna observation with a reasonable localization accuracy, which is of great interest for low power IoT applications, especially in the recent deployed narrowband IoT (NB-IoT) systems [1]. Although the aforementioned schemes can improve the TOA estimation performance at the expense of complexity, there are still many practical challenges in the low power IoT scenarios. For example, ESPRIT and MUSIC algorithms require real-time fast Fourier transform (FFT) and channel frequency response (CFR) estimation before the subspace manipulation, which becomes a heavy burden if the TOA estimation needs to be performed frequently. For the DCNN based solution, the adaptation capability to fit different channel environments is usually limited, since it highly relies on the training data to obtain satisfied estimation results [8]. Moreover, the MPLR algorithm can achieve reasonable estimation accuracy only if the reference signals are uniformly distributed.

In this letter, we propose a high accurate TOA estimation scheme with fine-grained feature generations for IoT applications. Specifically, we generate both the full band and resource block (RB) based reference signals, and calculate the cross-correlations with sampled observations to generate the feature map. Using this spectrogram-like cross-correlation feature map, we apply the machine learning technology with decoupled feature extraction and fitting to understand the variations in the time and frequency domains. In this way, we project the features directly into TOA results. Numerical results show...
that the proposed high accurate TOA estimation method can achieve more than 51% and 25% RMSE improvement in the static and multi-path fading environments, respectively.

II. BACKGROUNDS

Consider a general IoT system using orthogonal frequency division multiplexing (OFDM) scheme with \( K \) subcarriers. Denote \( d_k^p \) to be position reference signal (PRS) of baseband for the \( k \)-th subcarrier and the \( p \)-th time slot, and the corresponding time domain representation, \( s_p(t) \), is simply given by, \( s_p(t) = \sum_{k=0}^{K-1} d_k^p e^{j2\pi k\Delta f(t-N_{cp}T_s)} \), for \( t \in [0, (K+N_{cp}-1)T_s] \). \( N_{cp} \) and \( T_s \) denote the length of cyclic prefix (CP) and the sampling period, respectively, and \( \Delta f \) is the subcarrier spacing. Denote \( \tau_i \) to be the propagation delay of the \( l \)-th fading path, the received signal at the baseband can be expressed as,

\[
r_p(t) = \sum_{l=0}^{L-1} h(\tau_l) s_p(t - \tau_l) + w(t),
\]

where \( h(\tau_l) \) is the channel response corresponding of the \( l \)-th fading path, and \( w(t) \) denotes the additive white Gaussian noise (AWGN) with zero mean and variance \( \sigma^2 \). At the receiver side, by multiplying the conjugate of transmitted signal \( s_p(t) \), the cross-correlation \( R(\cdot) \) can be calculated by,

\[
R(mT_s) = \sum_{n=m}^{N_l+m-1} r_p(nT_s) s_p^*(n - m) T_s
\]

\[
= \sum_{l=0}^{L-1} h(\tau_l) \gamma(mT_s - \tau_l) + \hat{w}(mT_s),
\]

where \( \gamma(mT_s - \tau_l) = \sum_{n=m}^{N_l+m-1} s_p(nT_s) s_p^*(n - m) T_s - \tau_l \). \( N_l \) denotes the size of searching window, and \( \hat{w}(mT_s) \) is the remaining noise term.

Theoretically speaking, an exhaustive based scheme to search over all the possible \( L \), \( \{\tau_l\} \), and \( \{h(\tau_l)\} \) is usually required to minimize the localization error [9]. However, to reduce the search dimension in the practical implementation, we often fix the number of \( L \) by observing the signal characteristics in the cross-correlation function \( R(mT_s) \) [7] or some pre-defined numbers [3]. Thus, the optimal delay profiles as well as channel responses can be estimated via\(^2\)

\[
\{\tau_l^*\}, \{h^*(\tau_l)\} = \arg\min_{\{\tau_l\}, \{h(\tau_l)\}} \sum_{l=0}^{M-1} |R(mT_s) - \sum_{l=0}^{L-1} h(\tau_l) \gamma(mT_s - \tau_l)|^2,
\]

where \( M \) indicates the observation window length of the cross-correlation function \( R(mT_s) \). By taking the minimum value of the delay profile \( \{\tau_l\} \), we obtain the estimated TOA through \( \tau^* = \min_l \{\tau_l^*\} \).

III. PROPOSED TOA ESTIMATION SCHEME

In general, the above optimization problem for TOA estimation requires a joint search of unknown number of fading paths, channel coefficients, and the possible TOAs. Conventional estimation algorithms, such as SAGE [3], are able to obtain rough results via iterative methods, while the estimation accuracy is difficult to improve, especially when the sampling rate is limited. In this section, we propose a deep learning based TOA estimation scheme together with several augmentation techniques to solve this issue.

A. Overview of the Proposed Scheme

The general schematic diagram of the proposed TOA estimation scheme is shown in Fig. 1, where two main improvements are elaborated below.

1) Fine-Grained Feature Generation: In order to improve the resolution of cross-correlation observations, we generate the reference signals in RB based manner\(^3\) at the receiver side. The time domain representation is given by, \( s_p(t) = \sum_{k \in \Omega^v} d_k^p e^{j2\pi k\Delta f(t-N_{cp}T_s)} \), where \( \Omega^v \) denotes the set of subcarrier indices in the \( v \)-th RB and the cardinality of \( \cup_v \Omega^v \) equals to \( K \). Following the same procedures as introduced in Section II, we have the RB based cross-correlations, \( \{R^v(mT_s)\} \), as follows.

\[
R^v(mT_s) = \sum_{l=0}^{L-1} h(\tau_l) \gamma^v(mT_s - \tau_l) + \hat{w}^v(mT_s),
\]

where \( \gamma^v(mT_s - \tau_l) = \sum_{n=m}^{N_l+m-1} s_p(nT_s) s_p^*(n - m) T_s - \tau_l \). With the above manipulations, the input spaces of the proposed TOA estimation scheme can incorporate both the full band cross-correlation \( R(mT_s) \) and the RB based cross-correlations \( \{R^v(mT_s)\} \). Compared with the conventional approach as defined in (3), it has more fine-grained cross-correlation observations and greatly improves the resolution of cross-correlation based feature maps as shown in Fig. 1.

2) Decoupled Feature Extraction and Fitting: The straight forward method for TOA estimation using the above fine-grained features is to compute TOAs based on the collected \( R(mT_s) \) and \( R^v(mT_s) \) for \( m \in [0, M - 1] \) using (3) and ensemble results through majority voting or simple averaging. However, this type of scheme fails to explore the interrelationship among different cross-correlation observations. Therefore, a more reasonable framework is to directly minimize the mean square errors (MSE) of TOA estimation through,

\[
\min_{\mathcal{F}(\cdot)} \mathcal{F}(R(mT_s), \{R^v(mT_s)\}, \Sigma_{\{h(\tau_l)\}, \{\tau_l\}} - \tau^2),
\]

subject to \( \tau = \min_l \{\tau_l\} \).

\[\text{(5)}\]

where \( \Sigma_{\{h(\tau_l)\}, \{\tau_l\}} \) denotes the available statistical information of channel fading coefficients and propagation delays, and \( \mathcal{F}(\cdot) \) represents the mapping function from the obtained cross-correlations and the historical information to the estimating TOA result. A straight forward idea for easy deployment is to decouple the environment dependent parameters \( \Sigma_{\{h(\tau_l)\}, \{\tau_l\}} \) and the observation dependent parameters \( R(mT_s), \{R^v(mT_s)\} \). Inspired by [11], we construct two concatenated representative functions \( H_\theta \) and \( H_f \) with auxiliary parameters \( \theta_c \) and \( \theta_f \) to approximate the original function

\[\text{2}\]

\[\text{3}\]
\( F(\cdot) \). \( H_e \) extracts general features of TOA from the cross-correlation functions of the received signal, and \( H_f \) uses the features extracted by \( H_e \) to fit the TOA estimation result. Applying this notation, the original minimization becomes,

\[
\begin{align*}
\min_{\theta_e, \theta_f} \quad & \left| H_f(H_e( R(mT_s), \{R^v(mT_s)\}; \theta_e) \right| \\
\text{subject to} \quad & \tau = \min_i \{\tau_i\}.
\end{align*}
\]

Through the above scheme, we can isolate the deployment of the feature extraction process \( H_e(\cdot; \theta_e) \) and the TOA fitting process \( H_f(\cdot; \theta_f) \) as depicted in Fig. 1, and the corresponding advantages are listed below. First, we can adopt a common implementation strategy for IoT sensors with shared feature extraction. After obtaining the extracted features, a customized TOA estimation module at the central cloud or mobile edge units is applied to analyze them using environment dependent models. Through this approach, it can perfectly solve the aforementioned issues. Second, the data security and communication issues can be partly solved using the extracted feature patterns rather than the original cross-correlation observations. Last but not least, the adaptation to new wireless environments is much easier, since only the fitting modules need to be updated in order to achieve satisfactory performance.

### B. Neural Network Design

Different from the conventional approach to predict the entire propagation delay profile before TOA estimation, we directly model the functions \( H_e(\cdot; \theta_e) \) and \( H_f(\cdot; \theta_f) \) through neural networks. The corresponding network configuration and loss function are elaborated below.

1) **Network Configuration:** The detailed network configuration contains two-dimensional convolutions (Conv2D) and fully connected (FC) layers with Rectified Linear Unit (ReLU) activation, as shown in Table I. Meanwhile, in order to meet different input dimensions, we use a parameterized output dimension with respect to the total observation window length, \( M \), and the total number of RB, \( N_{RB} \), rather than some specific numbers.

2) **Loss Function:** Due to the limited sampling rate, the resolution of TOA estimation is limited. Meanwhile, a more fine-grained classification will increase the dimension of output vector, which increases the neural network complexity as well as the computational overhead. To address this issue, we directly adopt the regression model to predict the TOA result \( \tau \) and the associated loss function, \( L_\tau \), is chosen to be the MSE between predicted results and true TOA values,

\[
L_\tau = \left| H_f(H_e( R(mT_s), \{R^v(mT_s)\}; \theta_e) \right| \\
\sum \{h(\tau_i), \{\tau_i\}; \theta_f \} - \tau \right|^2.
\]

### C. Training Skills

In order to train two concatenated blocks in the proposed neural network architecture, we propose a two-stage training scheme with data augmentation technique to reduce the training data requirement.

1) **Two-Stage Training:** In the first stage, we simultaneously train two concatenated blocks using the data sets generated under different channel fading environments, which enables the feature extraction module to learn sufficient TOA-related features. In the second stage, we freeze the feature extraction part to be \( H_e(\cdot; \theta_e) \), and fine-tune the fitting module with low cost training samples. It is worth to mention that the training data in the second stage can be collected from remote entities with feature extraction capabilities or generated by the data augmentation scheme as explained later.

\*The true TOA values are obtained by directly dividing the Euclidean distances between transmitters and receivers over the speed of light.
2) Perturbation Based Data Augmentation: Denote \( \delta(mT_S) \) and \( \{ \delta^v(mT_S) \} \) to be the random perturbation on the full band and RB based cross-correlation observations, respectively. The corresponding output vectors of the feature extraction module, \( \eta_b \), are given by \( \eta_b = H_c(R(mT_S) + \delta(mT_S), \{ R^v(mT_S) + \delta^v(mT_S) \}); \theta_e \). By pairing \( \eta_b \) with the original TOA value \( \tau \), we can enlarge the training data set for the second stage optimization, which is defined as,

\[
\text{minimize } J = \| H_f \left( \eta_b, \sum \{ h(\tau_l) \}, \tau_l ; \theta_f \right) - \tau \|^2. \tag{8}
\]

It is worth to mention that we generate \( \delta(mT_S) \) and \( \{ \delta^v(mT_S) \} \) according to zero mean Gaussian distribution with the variations much less than the absolute values of \( R(mT_S) \) and \( \{ R^v(mT_S) \} \).

IV. Numerical Results

In this section, we provide some numerical results to demonstrate the advantages of proposed TOA estimation scheme by comparing with state-of-the-art results. In particular, we consider a typical NB-IoT scenario with sampling rate 1.92 Mbps, which is equivalent to \( T_s = 520 \) ns and \( N_{RB} = 6 \). The numerical simulations are performed under three channel conditions, e.g., static propagation environments (case 1), multipath fading environments including Extended Pedestrian A model with a doppler frequency of 5Hz (EPA-5Hz) (case 2) and Extended Vehicular A model with a doppler frequency of 5Hz (EVA-5Hz) (case 3) [14]. Without the prior statistical knowledge of fading environments, we choose \( \Sigma \{ h(\tau_l) \}, \tau_l \) to be \{ case 1, case 2, case 3 \}, and the mapping scheme of PRS follows 3GPP Release 15 standard [10].

A. Effects of Training Skills

In the first experiment, we compare the validation losses and cumulative distribution function (CDF) of TOA estimation errors using different training skills to demonstrate the advantages of the proposed scheme. As illustrated before, we accumulate the accurate TOA results under three cases to train the proposed neural networks and test them under training strategies. In Fig. 2, we share the training result in terms of the validation loss in the first stage (blue line), and compare the validation losses of the proposed scheme with (red line) and without (green line) data augmentation strategies. Each epoch represents one training process using all training data. As shown in this figure, the validation loss of stage 1 converges to \( 10^4 \) and that of stage 2 can achieve a better result, \( 10^3 \). If we can apply perturbation based data augmentation, we can obtain another 20% reduction, e.g., from \( 10^3 \) to \( 8 \times 10^2 \).

5According to [12], the perturbation can be utilized to make neural networks smoother and more robust without significantly sacrificing the system performance. In a real-life implementation, we also need to filter out abnormal data to eliminate outliers as proposed in [13].

6The data set as well as the source codes in this letter are available at https://github.com/XLPolar/TOA_estimation.

In fact, our proposed algorithm can be extended to other OFDM-based IoT communication systems, such as WiFi or LTE-M, since we only choose extremely low sampling rate, limited RB numbers, and no exact channel statistical knowledge in our simulation settings.
module to show that the decoupled feature extraction and fitting strategy are feasible for different PRSs’ configurations.

C. Multipath Fading Environments

Fig. 5 compares the CDF of TOA estimation errors in the multipath fading environments, where PRSs with two subframe durations are configured. The proposed TOA estimation scheme with fine-grained feature generation can achieve the median errors of 38 ns and 126 ns for case 2 and case 3, respectively. In other words, the proposed TOA estimation scheme can achieve 36% and 47% improvement if compared with the MUSIC algorithm, 25% and 57% improvement if compared with the ESPRIT algorithm, or 47% and 44% improvement if compared with the DCNN scheme. In the above numerical simulations, we also plot the estimation errors of the conventional scheme as defined by (3) for reference purpose. Meanwhile, we list the running time of different algorithms in Table II for the sake of completeness. Based on the above results, we conclude that the proposed method can outperform the existing baseline schemes and achieve a significant TOA estimation accuracy improvement for static propagation as well as multipath fading environments. Meanwhile, the computational complexity of proposed scheme is also lower than other baseline algorithms.

V. CONCLUSION

In this paper, we propose a high accurate ToA estimation algorithm based on fine-grained feature generation using a spectrogram-like cross-correlation feature map. The idea of decoupled feature extraction and fitting is then applied to improve the estimation accuracy and system robustness. A two-stage training strategy with perturbation based data augmentation scheme is utilized to fine-tune the feature extraction and fitting modules. Through some numerical results, we believe the proposed estimation scheme is promising to achieve more accurate TOA results than conventional methods.

REFERENCES

[1] X. Lin et al., “Positioning for the Internet of Things: A 3GPP perspective,” IEEE Commun. Mag., vol. 55, no. 12, pp. 179–185, Dec. 2017.
[2] J. Gante, G. Falcão, and L. Sousa, “Deep learning architectures for accurate millimeter wave positioning in 5G,” Neural Process. Lett., vol. 47, no. 1, pp. 487–514, Aug. 2020.
[3] S. Hu, X. Li, and F. Rusek, “On time-of-arrival estimation in NB-IoT systems,” in Proc. IEEE Wireless Commun. Network. Conf. (WCNC’19), Marrakech, Morocco, Apr. 2019, pp. 1–6.
[4] C. Roy and T. Kailath, “ESPRIT estimation of signal parameters via rotational invariance techniques,” IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 7, pp. 984–995, Jul. 1989.
[5] X. Li and K. Pahlavan, “Super-resolution TOA estimation with diversity for indoor geolocation,” IEEE Trans. Wireless Commun., vol. 3, no. 1, pp. 224–234, Jan. 2004.
[6] Z. He, Y. Ma, and R. Tafazolli, “Improved high resolution toa estimation for OFDM-WLAN based indoor ranging,” IEEE Wireless Commun. Lett., vol. 2, no. 2, pp. 163–166, Apr. 2013.
[7] S. Hu, A. Berg, X. Li, and F. Rusek, “Improving the performance of OTDOA based positioning in NB-IoT systems,” in Proc. Global Commun. Conf. (GLOBECOM’17), Singapore, Jan. 2017, pp. 1–7.
[8] O. Bialer, N. Garnett, and D. Levi, “A deep neural network approach for time-of-arrival estimation in multipath channels,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP’18), Calgary, AB, Canada, Apr. 2018, pp. 2936–2940.
[9] J.-J. Fuchs, “Multipath time-delay detection and estimation,” IEEE Trans. Signal Process., vol. 47, no. 1, pp. 237–243, Jan. 1999.
[10] Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation, V15.7.0, Standard 3GPP TS 36.211, May 2019.
[11] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS’14), Dec. 2014, pp. 3320–3328.
[12] C. Xiang et al., “Robust sub-meter level indoor localization with a single WiFi access point—Regression versus classification,” IEEE Access, vol. 7, pp. 146309–146321, 2019.
[13] H. Wang, M. J. Bah, and M. Hammad, “Progress in outlier detection techniques: A survey,” IEEE Access, vol. 7, pp. 107964–108000, 2019.
[14] Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTra); Base Station (Bs) Radio Transmission and Reception, V16.4.0, Standard 3GPP TS 36.104, Dec. 2019.