Research Article

Two-Stage Channel Adaptive Algorithm for Unmanned Aerial Vehicles Localization with Cellular Networks

Chenxi Zeng [1,2] **Zhongliang Deng,**[1] **Jiyang Ma,**[2] **and Shengsong Yang**[3]

[1]*Beijing University of Posts and Telecommunications, Beijing, China*
[2]*China Academy of Information and Communication Technology, Beijing, China*
[3]*University of Science and Technology Beijing, Beijing, China*

Correspondence should be addressed to Chenxi Zeng; 810920991@qq.com

Unmanned aerial vehicle (UAV) is regarded as a powerful tool to expand the existing ground wireless network into aerial space. Since high mobility is an essential characteristic for UAV, it is important to carry out an accurate, real-time, and high-precision localization in terms of safe operation and communication link maintenance. The cellular network-based localization technology has provided UAV a solution with both high coverage and seamless connection. However, the complex channel environment between the UAV and terrestrial base station (BS) would have weakened the localization performance. To solve this problem, a two-stage channel adaptive algorithm for cellular-connected UAV has been proposed. The first stage of the algorithm is to revise the observation error introduced by the complex channel environment using the model of DDPG. The second stage is to locate the UAV position with TDOA algorithm using the revised observation values. Simulation results have demonstrated that the proposed algorithm can achieve the channel adaptive effect by revising the observation errors and improve location performance greatly, especially for UAVs at a relative lower altitude.

1. Introduction

With the large-scale deployment and application of the fifth generation (5G) cellular system, researchers start to focus their studies on the sixth generation (6G) mobile communication network. Compared with 5G, the most impressive improvement of 6G network is the ability to provide an intelligent, seamless, and three-dimensional (3D) aerial access network connectivity with a data rate of several terabits per second (TBPs) and ultralow delay of sub-millisecond [1]. In order to fulfill the innovative objective in wireless communications for the coming 6G communication systems and provide an on-demand connectivity from the sky, new subjects such as satellites, high- and low-altitude platforms, drones, aircrafts, and airships are being included to take the role as aerial base stations. Among them, the unmanned aerial vehicles (UAVs) have been regarded as a powerful tool to expand the existing 5G wireless network into the aerial access network with its flexible on-demand deployment capability [2].

Since high mobility is an essential characteristic for UAVs, it is very important to carry out an accurate, real-time, and high-precision localization in terms of the safe operation and communication link maintenance. According to [3], GNSS (Global Navigation Satellite System), INS (Inertial Navigation System), and visual-based navigation are three technologies used by UAV localization. GNSS is the most widely used one for its global coverage; however the satellite signal is sensitive to obstacles and blockings, leading to accuracy drops in complex environments like city centers. INS does not rely on the satellite signal, but the relatively high cost of equipment makes it not suitable for small aircraft, and the accumulated offset error over time due to integral drift is an unsettled problem [4]. Visual-based localization depends on weather conditions; low visibility environments like dusty or smoking can damage the visual signal seriously, causing significant declines of accuracy [5]. Furthermore, visual-based localization demands large amount of image processing, which requires high computing capability and increases the system complexity.

As the rise of 6G, the aerial-terrestrial integration network has provided an alternative for UAV localization, namely, the cellular-connected UAVs localization solution [6–8]. 6G aerial network can provide cellular connectivity with ubiquitous accessibility for UAVs and not restricted by satellite signal attenuation and visual signal damage. Cellular-connected UAV can also be a research model of user experience for the high-altitude network as well to help improving the service quality of the aerial network [7, 9]. Furthermore, the existing wireless communication technologies such as massive MIMO and millimeter wave communication can also be exploited for UAV localization.

From 1G to 5G, various localization technologies have been designed and implemented [10]. Network topology information and radio signal from wireless network are the main information source for mobile network localization. According to the implementation principle, the location algorithm can be divided into five categories: proximity, angle of arrival (AOA), received signal strength (RSS), time of arrival (TOA)/time difference of arrival (TDOA), and hybrid method [11]. Among them, the proximity method depends on the density of the network transmitter, leading higher cost of the positioning equipment [12]; AOA is easily affected by the external environment, and the need of additional hardware makes it not suitable for the large-scale sensor networks respect to the hardware size and power consumption [13]; the RSS method shows good characteristics in the experimental environment, but its low robustness of the environment temperature, humidity, and propagation mode makes it difficult to implement in the practical world [14]; TOA-based positioning also has a relatively good performance on location accuracy, but it requires accurate time synchronization among all the nodes, which is very difficult to implement and requires a high cost [15]; TDOA is widely implemented in the current mobile network positioning technology, not only for it retains the advantages of TOA's small ranging error, but also for its no need of the strict time synchronization between nodes [16].

Related research like drone detection and tracking using technologies mentioned above has achieved positive results. In [17], authors have designed a system in order to detect the unlicensed small-sized drones in 5G mm wave cellular networks. Schloemann et al. [18] have investigated the application of cellular networks for localization of terrestrial mobile terminals with the theory of stochastic geometry. However, those studies have not considered the complex aerial environment UAV works in; the new emerging aerial users (such as cellular-connected UAVs) would have introduced much more complexity to the existing location model as well.

Generally, factors such as the three-dimensional mobility of UAVs, altitude-dependent channel characteristics (between BSs and UAVs), line of sight (LOS)/nonline of sight (NLOS) conditions, and the interference from the neighboring BSs contribute to the complex aerial environment. It is quite difficult to achieve a satisfied result by adjusting only one factor at a time for the coexistence and cooperation among parameters [19]. One creative way to solve this is to observe the problem from a global view instead of regional parts, oriented by factors affecting the localization performance other than reasons complicated UAV working environment. According to [18, 20], localization performance depends on three factors no matter which technology was used. These factors are as follows: number of participating BSs, accuracy of original observations, and relative distance between surrounding BSs and the target device. Among them, location observation is the only option with enough flexibility to be optimized; number of participating BSs and relative distance between BSs and the target device are related to the infrastructures which are relatively difficult to be modified.

To improve the quality of location observations from a global view, reinforcement learning (RF) is introduced to compensate the errors. RF is a machine learning algorithm trained by unlabelled data similar to the semisupervised learning. Since the dynamic aerial model is difficult to describe and the action of TDOA is a continuous behaviour, model-free policy optimization methods are good options under these circumstances. Policy optimization methods can solve problems of continuous behaviour space with controllable computational complexity using a dual neural network (actor-critical network). Gradient algorithm (VPG) [21], trust region policy optimization (TRPO) [22], proximity policy optimization (PPO) [23], and deep deterministic policy gradient (DDPG) [24] are typical policy algorithms. In [25], Zhang et al. used the PPO algorithm to correct the NLOS measurement error in AOA location and achieved good results. However, PPO uses online strategy which cannot take efficient usage of the historical data and increases the cost of training set. However, DDPG has skilfully combined the advantage of Q learning and policy algorithm by using experience playback and a duel-double network. That means DDPG can solve the problem of continuous action space and improve data utilization rate of TDOA at the same time.

In this paper, a channel adaptive algorithm for UAV localization by optimizing the location observations based on DDPG has been proposed. The proposed algorithm has included two stages: firstly, revise the location observations error introduced by the complex aerial channel environment by studying the historical data using DDPG model; secondly, calculate the location of target UAV using the TDOA algorithm. Since the observation values are revised by the learning model of the whole channel environment instead of each single variable, the complexity of the algorithm can be reduced while the location performance increased.

The organization of the paper is as follows: the channel propagation model of aerial environment UAV works in has been analysed in Section 2. The introduction of basic idea in DDPG, observation revised model, and TDOA procedure analysis have been constructed in Section 3. Simulation results have been illustrated in Section 4. Section 5 has concluded the whole paper and the future work plans.

2. Channel Model

Due to the complexity of the aerial wireless channel communication environment, the location observation value will be interfered by many factors such as three-dimensional

mobility of UAVs, altitude-dependent channel characteristics between base stations (BSs) and UAVs, LOS/NLOS conditions, and interference from neighboring BSs, who will lead certain deviations to the final location accuracy. In this section, the wireless transmission channel model of location observation and the mathematical expression of error would be illustrated.

Assume that there are M transmitting antennas at the base station, N receiving antennas in the mobile receiving terminal, one wireless propagation path between the transmitting and receiving points. Then the signal received by the receiving antenna at time t can be expressed as follows [26, 27]:

$$y_j = \sum_{i=1}^{M} x_i(t) h_{i,j}^l(t) + n_{i,j}(t), \qquad (1)$$

where $x_i(t)$ is the transmitted signal, y_j is the received signal, $h_{i,j}^l(t)$ is the channel model between transmitting antenna i and receiving antenna j at time t, and $n_{i,j}(t)$ is the noise.

Assume t_0 is the time delay caused by multipath propagation, so that each multipath subchannel model of MIMO system can be defined as

$$h_{i,j}^l(\tau - t) = h_{i,j}^l(t) \delta(\tau - t_0). \qquad (2)$$

The propagation delay is usually caused by the signal reflection, refraction, and scattering introduced by the NLOS condition in the environment. Therefore, the observation values used in the location algorithm as input of the second stage will introduce certain errors to the location system. The mathematical expression of the error is shown as follows:

$$\alpha_{i,j} = \alpha_{0,i,j} + \alpha_{n,i,j} + \alpha_{e,i,j}, \qquad (3)$$

where $\alpha_{0,i,j}$ is the location observation value in the ideal environment, $\alpha_{n,i,j}$ is the channel error caused by NLOS, $\alpha_{e,i,j}$ is the measurement error introduced by the complex environment, i is the number of transmitting antennas, and j is the number of receiving antennas. Our goal is to minimize the sum of $\alpha_{n,i,j}$ and $\alpha_{e,i,j}$ so that the value $\alpha_{i,j}$ is as close to $\alpha_{0,i,j}$ as possible.

3. Channel Adaptive Algorithm

3.1. Deep Deterministic Policy Gradient. DDPG is a combination of Q learning and policy gradient. There are four networks in DDPG, namely, the critic, actor, target critic, and target actor network. The critic network responses for the Q function updates using the loss function similar as DQN (deep Q network) [28], except that the action is calculated separately by the actor network. This operation makes DDPG possible to deal with complex and continuous actions. The actor network calculates an action result with the highest Q value according to the policy gradient used in DPG (deterministic policy gradient) [29]. It is deterministic since DDPG does not calculate the probability of every possible action but outputs only one deterministic action.

Since in many cases (for example, our localization scenery), targets keep changing while update values are calculated, which would lead in difficulties for update. Similar as DDQN (deep reinforcement learning with double Q learning) [30], DDPG has implicated the fix network technology to fix the target network before assigning the new parameters. The parameter assignment of the target network has taken usage of a soft update instead of the hard one to assure the stability of learning.

In order to avoid the correlation of samples and different feature unit, DDPG has introduced the replay buffer and batch normalization mechanism. The replay buffer stores a finite set of state, action, and reward pair. The actor and critic network take samples from the replay buffer to calculate the target at each time step. When the buffer is full, the oldest set would be discarded, leaving room for the new comings.

The core of the channel adaptive algorithm proposed here is to revise the location observation with the help of the DDPG model. Since the observation value is time-changing and needs a continuous revised action, it is quite suitable for taking use of DDPG. Once the observations are revised, the UAV location will be calculated using the revised observation value by the TDOA algorithm.

3.2. Location Observation Revise Model. Assume $\{NIdeal_i, i = 1, \ldots, N1\}$ are samples of $N1$ observation values under the real channel environment and $\{Ideal_i, i = 1, \ldots, N2\}$ are samples of $N2$ observation values under ideal channel environment. According to formula (3), the relationship is as follows:

$$NIdeal_i = Ideal_i + \alpha_{n,i} + \alpha_{e.i}. \qquad (4)$$

Let $S = \{NIdeal_i, i = 1, \ldots, N1\}$ and $T = \{TIdeal_i, i = 1, \ldots, N2\}$; denote S as the initial state of the revise model and T as the training target. According to the DDPG algorithm, there are four networks in the system. During the initialization process, four networks and their caches would be assigned initialized values. The initialization process is as follows:

(i) Initialize the critic network $Q(s, a|\theta^Q)$ and actor network $\mu(s|\theta^\mu)$ randomly, with s representing the current state of two networks, a representing the action to be executed, θ^Q and θ^μ are network parameters.

(ii) Initialize the target networks Q' and target actor μ' randomly, with network parameters $\theta^{Qi} = \theta^Q$, $\theta^{\mu'} = \theta^\mu$.

(iii) Empty the replay buffer and denote it as R.

After initialization, for $i = 1, \ldots, M$, denote S to be the initial state set and perform the following steps at time $t = 1, \ldots, T$:

(i) Obtain action $a_t = \mu(s_t|\theta^\mu)$ in the policy network at state S,

where μ is a mapping from state to action with function approximators parameterized by θ^μ and s_t is the state action that a_t starts from.

(ii) Execute action a_t and obtain the new state s_{t+1} and reward r_t.

(iii) Store (s_t, a_t, r_t, s_{t+1}) into the replay buffer R.

(iv) Select N groups of variables $B = \{(s_i, a_i, r_i, s_{i+1})\}$ from R randomly and input B into the actor target network and critical target network, respectively.

(v) Calculate the next action a_{t+1} in the actor target network with

$$a_{t+1} = \mu'\left(s_{t+1} \mid \theta^{\mu'}\right), \tag{5}$$

where μ' is the mapping function of state and action at time $t+1$ whose function approximators parameterized by $\theta^{\mu'}$.

(vi) Calculate the target value y_i of Q in the critical target network:

$$y_i = r_i + \gamma Q'\left(s_{i+1}, \mu'\left(s_{i+1} \mid \theta^{\mu'}\right) \mid \theta^{Q'}\right), \tag{6}$$

where r_i is the reward at time step i, γ is a discounting factor with $\gamma \in [0,1]$, and $Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$ represents the expected return after taking action a_{i+1} at state s_{i+1}. However, y_i is also dependent on $\theta^{Q'}$.

(vii) Calculate the loss function L of Q:

$$L = \frac{1}{N} \sum_i \left(y_i - Q\left(s_i, a_i|\theta^Q\right)\right)^2, \tag{7}$$

where N is the number of sets selected from the replay buffer.

(viii) Update the critical network by minimizing the loss function.

(ix) Update the actor network by the gradient back propagation algorithm of the neural network:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q\left(s, a \mid \theta^Q\right)\Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu\left(s \mid \theta^\mu\right)\Big|_{s_i}. \tag{8}$$

This is the policy gradient calculation model, and the derivation process can be found in [27].

(x) Update critical target network parameters and actor target network parameters:

$$\begin{aligned} \theta^{Q'} &\longleftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \\ \theta^{\mu'} &\longleftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}, \end{aligned} \tag{9}$$

where τ is the soft update parameter with value of $\tau \in [0,1]$.

(xi) Complete the current iteration if s_{t+1} is the termination state, otherwise return the first step.

Figure 1 has illustrated the training process of the observation data using DDPG. In our model, action a_t is a continuous value that can make the location observation samples as close as the samples in the ideal environment. The
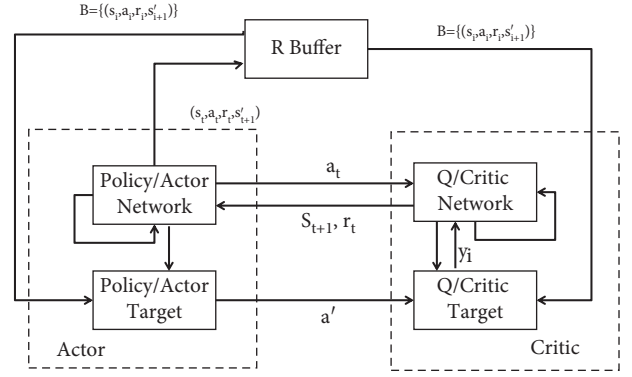

FIGURE 1: The training process of DDPG.

reward information r_t represents the quality of the location observation value of DDPG model, and the policy would adopt corresponding correction actions according to r_t, in order to maximize the reward value.

Since the concept of "Replay Buffer" is quoted in this paper, the data of the previous policy can be used for each training episode instead of only the current cycle, which would reduce the training data scale and improve the data utilization efficiency.

3.3. UAV Location with TDOA. In the location system of TDOA, once the observation value is determined, the distance between UAV and base station can be calculated. Several observation values could constitute a set of hyperbolic equations about the terminal position of the target, and the estimated position would be obtained by solving those equations.

Assume there are N base stations distributed in the three-dimensional (3D) space, the coordinate of the i-th base station is (x_i, y_i, z_i), the location of target UAV is (x, y, z), and the distance between target UAV and the i-th BS is R_i, where $i = 1, 2, 3, \ldots, M$. Then we got

$$R_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}. \tag{10}$$

Denote signal from BS1 as signal 1, and mark it as the reference signal. Record the time difference between signal i ($i \in m$) and signal 1 as t_{i1} and the distance as R_{i1}; then

$$\begin{aligned} t_{i1} &= t_i - t_1, \\ R_{i1} &= c * t_{i1} \\ &= c(t_i - t_1) \\ &= R_i - R_1, \end{aligned} \tag{11}$$

where c is the propagation velocity of electromagnetic wave. After transposition, we got

$$\begin{aligned} R_i^2 - R_1^2 &= (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 \\ &\quad - (x - x_1)^2 - (y - y_1)^2 - (z - z_1)^2. \end{aligned} \tag{12}$$

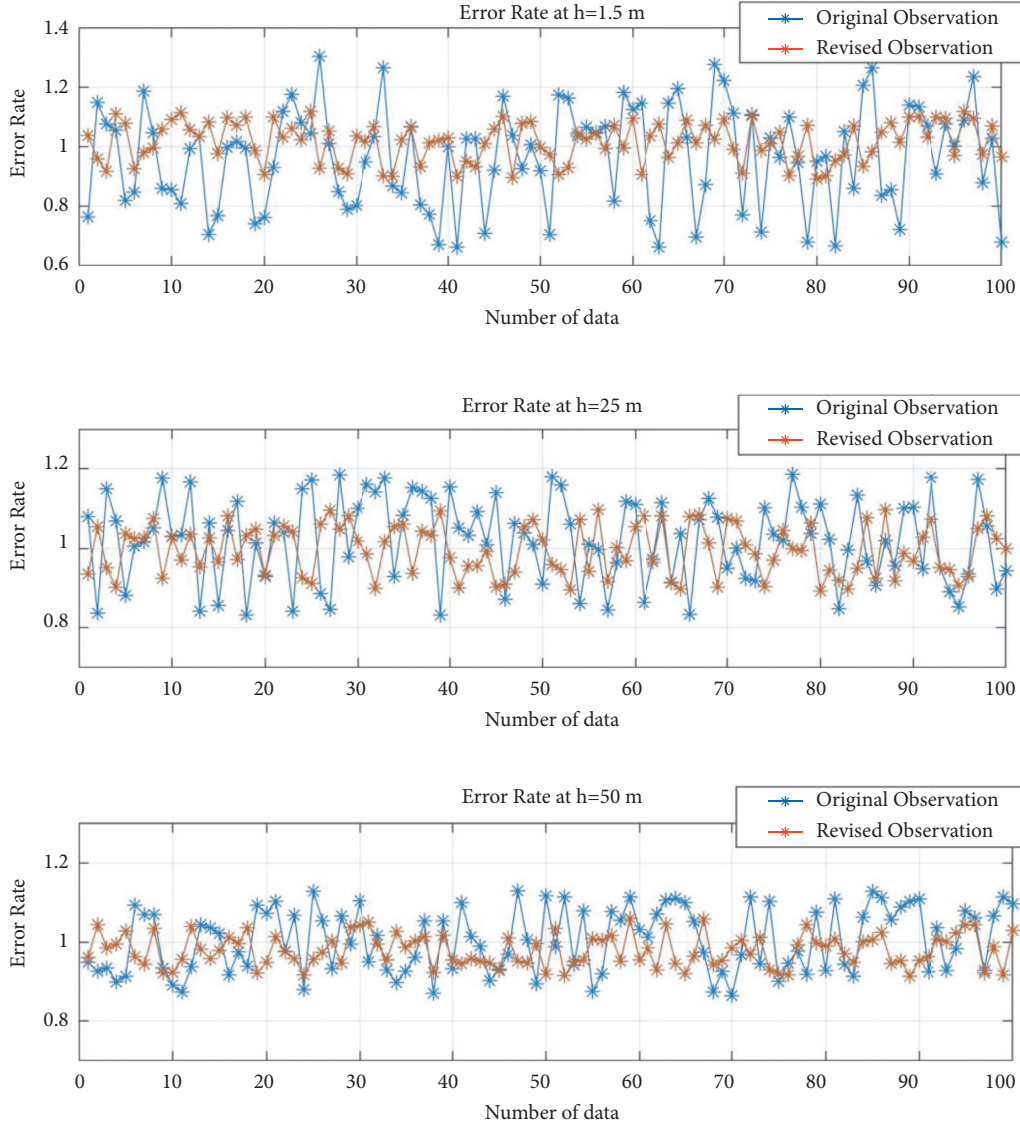After calculation, equation (12) can be presented as

FIGURE 2: Error rates of the observation value before and after using the revised model at different UE altitude.

$$2(x_1 - x_i)x + 2(y_1 - y_i)y + 2(z_1 - z_i)z$$
$$+ (x_i^2 + y_i^2 + z_i^2) - (x_1^2 + y_1^2 + z_1^2). \quad (13)$$

Here, we denote

$$K_i = x_i^2 + y_i^2 + z_i^2,$$
$$X_{i,1} = x_1 - x_i,$$
$$Y_{i,1} = y_1 - y_i, \quad (14)$$
$$Z_{i,1} = z_1 - z_i.$$

After simplification, we got

$$R_i^2 - R_1^2 = 2\begin{pmatrix} X_{i,1} & Y_{i,1} & Z_{i,1} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + K_i - K_1. \quad (15)$$

When there are four base stations participating in the localization process successfully, at least three TDOA observations can be obtained. Assume R_1 is known, and the location of UAV can be calculated as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = -\begin{pmatrix} X_{2,1} & Y_{2,1} & Z_{2,1} \\ X_{3,1} & Y_{3,1} & Z_{3,1} \\ X_{4,1} & Y_{4,1} & Z_{4,1} \end{pmatrix}$$

$$* \left[\begin{pmatrix} R_{2,1} \\ R_{3,1} \\ R_{4,1} \end{pmatrix} R_1 + \frac{1}{2} \begin{pmatrix} R_{2,1}^2 - K_2 - K_1 \\ R_{3,1}^2 - K_3 - K_1 \\ R_{4,1}^2 - K_4 - K_1 \end{pmatrix} \right]. \quad (16)$$
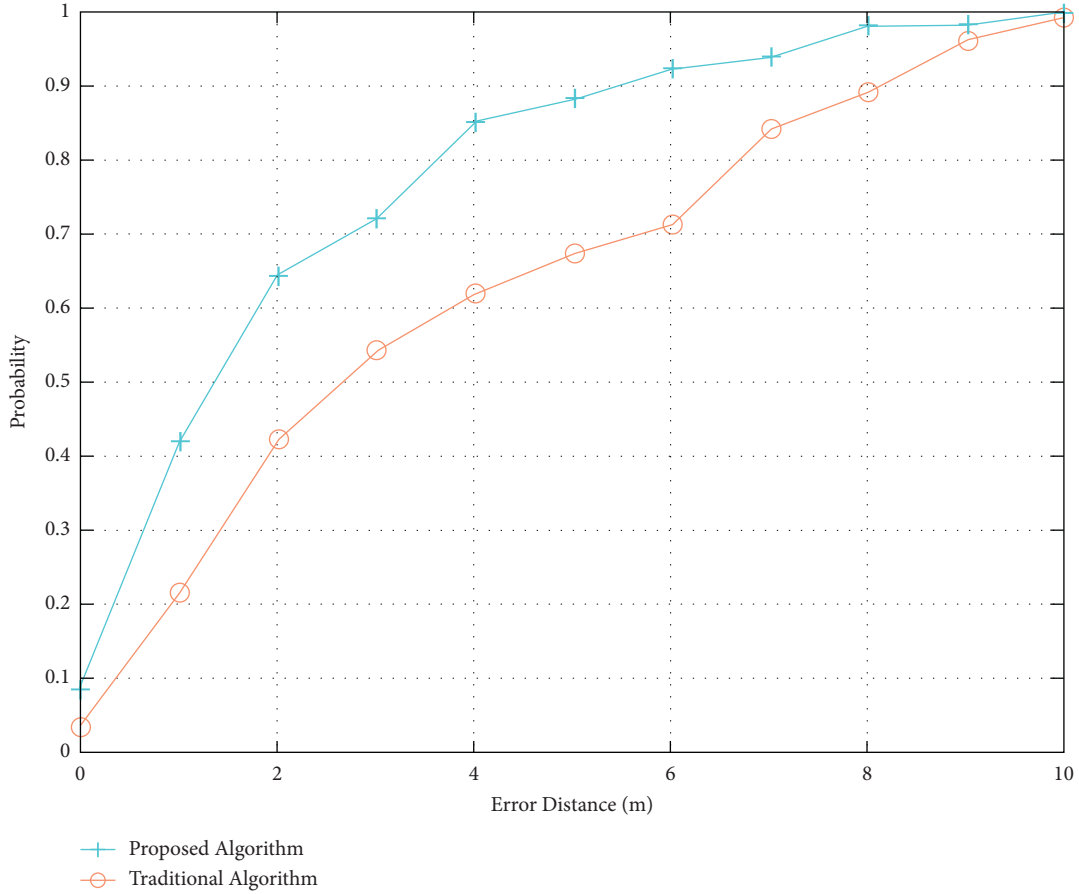
FIGURE 3: The cumulative distribution error of location accuracy of the proposed algorithm and traditionalTDOA.

This is the specific location information of the UAV to be measured. It can be seen that to localize an UAV with TDOA algorithm, at least four base stations are required to participate the localization. For simplification, we discussed only the four base stations case in this paper.

4. Simulation Results

4.1. Simulation Parameters. The 3GPP research project has studied three cellular-connected UAV scenarios, namely, urban-macro with aerial vehicles (UMa-AV), urban-micro with aerial vehicles (UMi-AV), and rural-macro with aerial vehicles (RMa-AV) [7]. The eNodeB antennas mounted above the rooftop levels of surrounding buildings in urban environment are UMa-AV scenarios. UMi-AV represents scenarios where eNodeB antennas are mounted below rooftop. eNodeB antennas mounted on top of towers of larger cells in rural environment are represented by RMa-AV [31]. In our simulation study, the 3GPP channel model of UMa-AV scenario for UAVs is considered.

The simulation environment is set by assuming the intersite distances of 500 m, and the height of BS is 25m [4]. The channel being used is with bandwidth of 10 MHz and the carrier frequency (fc) of 2 GHz. Variance of the shadowing is modelled as 4.64exp (−0.0066hUT) and 6 dB for LOS and NLOS conditions, respectively [4]. The transmitted

power is taken as 46 dBm and the noise figure of the UAV as 9 dB.

4.2. Numerical Results. Since our proposed algorithm is a two-staged process, there are two parts of numerical results demonstrated in this section, that is, the observation revised results and the final UAV location accuracy analysis.

Before setting the error revised model, it is necessary to collect a set of reference data under the ideal environment to help training and evaluating the revised effect. Define error ratio as the ratio between observation value and the reference data. An error ratio of 1 indicates that the observed data are exactly the same as the reference data.That is, the closer the error ratio is to 1, the better the correction effect will be.

Figure 2 shows the error ratio of observation data of TDOA before and after using the revised model as the UAV working at different altitudes. It is obvious that the revised observations have smaller amplitude than the original ones, which means their error rates are better. This advantage becomes more pronounced when the UAV (UE) stays at a relatively lower altitude. This phenomenon is consistent with the probabilistic simulation results of LOS at different UE heights. The references [7, 20] have shown the same result.

For the location stage, the cumulative distribution of positioning error is used to measure the final location accuracy. The cumulative error of positioning accuracy represents the cumulative probability value of positioning results within a specific range. Within the same distance range, the higher the cumulative probability value is, the better the positioning accuracy is.

Figure 3 shows the cumulative distribution error of location accuracy obtained by the TDOA algorithm before and after optimization. The UAV's altitude (hUE) was set to 25 m here. It can be seen that the proposed algorithm has a probability of 64.4% that the location accuracy is within 2 m, and the probability of positioning error within 4m is 85.4%; compared with the results of using the original TDOA algorithm, the positioning accuracy within 2 m has been improved by 51.9%, and the positioning accuracy within 4m has been improved by 37.7%.

Simulation results shown in this chapter have demonstrated that the proposed algorithm can achieve the channel adaptive effect by revising the observation errors and improving location performance greatly, while the improved effect is especially noticeable for UAVs at a relative lower altitude.

5. Conclusions

In this paper, a two-stage channel adaptive algorithm for cellular-connected UAV has been proposed. The first stage of the algorithm is to revise the observation error introduced by the complex channel environment between UAV and base stations. The second stage is to locate the UAV position with TDOA using the revised observation values. Simulation results have demonstrated that the proposed algorithm can achieve the channel adaptive effect by correcting the observation errors and improving location performance greatly, especially for the UEs at a relatively lower altitude.

The future work would be concentrated on the following aspects: one is to replace the single TDOA algorithm in this paper by the hybrid algorithm composed of TDOA, AOA, or RSS to optimize the location algorithm itself; the second aspect is to introduce the random policy factor represented by the information entropy into the optimization algorithm, in order to improve the randomness of the system and avoid the problem of the local optimal trap. The third aspect is to expand the quality evaluation domain from single location accuracy to a diversified set by introducing other user experience target, leading the model closer to user experience.

Data Availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Disclosure

The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Chenxi Zeng conceived the main idea and the game theory model; all authors contributed to data analysis, simulations, and the writing of this paper.

References

[1] Z. Xiao and Y. Zeng, "An overview on integrated localization and communication towards 6G, signal processing," 2020, https://arxiv.org/abs/2006.01535.

[2] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: a tutorial on UAV communications for 5G and beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327–2375, 2019.

[3] C. S. Y. C. S. Yoo and I. K. A. I. K. Ahn, "Low cost GPS/INS sensor fusion system for UAV navigation," *In IEEE Digital Avionics Systems Conference*, vol. 2, 2003.

[4] A. Nemra and N. Aouf, "Robust INS/GPS sensor fusion for UAV localization using SDRE nonlinear filtering," *IEEE Sensors Journal*, vol. 10, no. 4, pp. 789–798, 2010.

[5] J. P. Queralta, C. M. Almansa, F. Schiano, D. Floreano, and T. Westerlund, "UWB-based system for uav localization in gnssdenied environments: characterization and dataset," 2020, https://arxiv.org/abs/2003.04380.

[6] Y. Zeng, J. Lyu, and R. Zhang, "Cellular-connected UAV: potential, challenges, and promising technologies," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 120–127, 2018.

[7] 3GPP, "Study on enhanced LTE support for aerial vehicles," Technical Report, 3GPP TR, Southwest of Nice, France, 2017.

[8] A. Azari, F. Ghavimi, M. Ozger, R. Jantti, and C. Cavdar, "Machine learning assisted handover and resource management for cellular connected drones," in *Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–7, Antwerp, Begium, May 2020.

[9] M. M. Azari, F. Rosas, A. Chiumento, and S. Pollin, "Coexistence of terrestrial and aerial users in cellular networks," in *Proceedings of the 2017 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Singapore, December 2017.

[10] J. A. D. P. Rosado, R. Raulefs, J. A. L. O. Salcedo, and G. S. Granados, "Survey of cellular mobile radio localization methods: from 1G to 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2017.

[11] B. Brown and M. Aaron, "The politics of nature," in *The Rise of Modern Genomics*, J. Smith, Ed., Wiley, NY, USA, 3rd edition, 2001.

[12] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and F. Carlo, "A. survey of enabling technologies for networklocalization, tracking, and navigation," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, 2018.

[13] K. Doğançay and H. Hmam, "Optimal angular sensor separation for AOA localization," *Signal Processing*, vol. 88, pp. 1248–1260, 2008.

[14] Z. li, Z. Deng, F. Yang, and A. Hu, "The effect of temperature and humidity on finger print positioning," in *Proceedings of the China Satellite Navigation Conference, CSNC*, Harbin, China, 2018.

[15] J. S. Junyang, A. F. Molisch, and J. Salmi, "Accurate passive location estimation using TOA measurements," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2182–2192, 2012.

[16] A. Dersan and Y. Tanik, "Passive radar localization by time di_
erence of arrival," in *Proceedings of the IEEEMilitary Com-
munications Conference MILCOM*, vol. 2, pp. 1251–1257,
IEEE, Anaheim, CA, USA, October 2002.

[17] D. Solomitckii, M. Gapeyenko, V. Semkin, S. Andreev, and
Y. Koucheryavy, "Technologies for efficient amateur drone
detection in 5G millimeter-wave cellular infrastructure," *IEEE
Communications Magazine*, vol. 56, no. 1, pp. 43–50, 2018.

[18] J. Schloemann, H. S. Dhillon, and R. M. Buehrer, "Toward a
tractable analysis of localization fundamentals in cellular
networks," *IEEE Transactions on Wireless Communications*,
vol. 15, no. 3, pp. 1768–1782, 2015.

[19] I. A. Meer, M. Ozger, and C. Cavdar, "On the localization of
unmanned aerial vehicles with cellular networks," 2020,
https://arxiv.org/abs/2003.04889.

[20] J. Schloemann, H. S. Dhillon, and R. M. Buehrer, "Locali-
zation performance in cellular networks," in *Proceedings of the
IEEE International Conference on Communication Workshop
(ICCW)*, pp. 871–876, London, UK, June 2015.

[21] R. Sutten, D. M. Allen, S. Singh, and Y. Mansour, "Policy
gradient methods for reinforcement learning with function
approximation, 2000-02," *Advances in Neural Information
Processing Systems*, vol. 12, 2020.

[22] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel,
"Trust region policy optimization," 2017, https://arxiv.org/
abs/1502.05477.

[23] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and
O. Klimov, "Proximal policy optimization algorithms,"
https://arxiv.org/abs/1707.06347.

[24] T. P. Lillicrap, J. J. Hunt, A. Pritzel et al., "Continuous control
with deep reinforcement learning," 2019.

[25] Y. Zhang, Z. Deng, and Y. Gao, "Angle of Arrival Passive
Location Algorithm Based onProximal Policy Optimization,"
Electronic, vol. 8, no. 12, 2019.

[26] K. Lee, J. Oh, and K. You, "TDOA/AOA based geolocation
using Newton method under NLOS environment," in *Pro-
ceedings of the 2016 IEEE International Conference on Cloud
Computing and Big Data Analysis (ICCCBDA)*, pp. 373–377,
ChengDu, China, July 2016.

[27] N. Aghaie and M. A. Tinati, "Localization of WSN nodes
based on NLOS identification using AOAs statistical infor-
mation," in *Proceedings of the 2016 24th Iranian Conference on
Electrical Engineering (ICEE)*, pp. 496–501, Shiraz,Iran, May
2016.

[28] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level
control through deep reinforcement learning," *Nature*,
vol. 518, no. 7540, pp. 529–533, 2015.

[29] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and
M. Riedmiller, "Deterministic policy gradient algorithms," in
*Proceedings of the 31st International Conference on Machine
Learning (ICML 2014)*, Beijing China, June 2014.

[30] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement
learning with double q-learning," 2015, https://arxiv.org/abs/
1509.06461.

[31] S. D. Muruganathan, X. Lin, H. LiinaMaattanen,
W. A. H. ZhenhuaZou, and S. Yasukawa, "An overview of
3gpp release-15 study on enhanced lte support for connected
drones. networking and internet architecture (cs.ni)," 2019,
https://arxiv.org/abs/1805.00826.