Wireless Communications and Mobile Computing

Security, Privacy, and Trust Issues in 6G-enabled Sensing and Localization

Lead Guest Editor: Mu Zhou
Guest Editors: Ying-Ren Chien and Qiao Zhang

Security, Privacy, and Trust Issues in 6G-enabled Sensing and Localization

Wireless Communications and Mobile Computing

Security, Privacy, and Trust Issues in 6G-enabled Sensing and Localization

Lead Guest Editor: Mu Zhou
Guest Editors: Ying-Ren Chien and Qiao Zhang

Chief Editor

Associate Editors

Academic Editors

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7774166, 11 pages
https://doi.org/10.1155/2022/7774166

Research Article

An AOA and Orientation Angle-Based Localization Algorithm for Passive RFID Tag Array

Liangbo Xie, Yan Ren, Yong Wang, Wei Nie, and Mu Zhou

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400000, China

Correspondence should be addressed to Liangbo Xie; xielb@cqupt.edu.cn

Received 15 December 2021; Revised 7 April 2022; Accepted 4 May 2022; Published 23 May 2022

Academic Editor: Daniel G. Costa

This paper proposes an indoor localization algorithm based on AOA (Arrive of Angle) and orientation angle (OA) for passive UHF RFID (Ultra-High Frequency Radio Frequency Identification) tag array. By utilizing a uniform linear tag array (ULA), the AOA, which is defined as the angle of the reader relative to the tag array, can be estimated by the Multiple Signal Classification (MUSIC) algorithm. The relationship between AOA, OA of the tag array, and the direction angle (DA) between each reader antenna to the geometric center of the tag array is analyzed, and an OA retrieval method based on rotating the coordinate system is proposed to calculate DA. The tag array localization is finally acquired by the DA of each antenna. Simulation and experiment results show that the proposed algorithm can achieve decimeter-level localization accuracy. The proposed algorithm achieves a mean accuracy of 0.216 m, which is a significant improvement compared with the traditional AOA localization algorithm.

1. Introduction

With the advantages of low cost and high data rate, radio frequency identification (RFID) technology plays an important role in the perception layer of the internet of things (IoT) [1], which has been widely used in many fields, such as supply chain management (SCM) [2], logistics management, and warehouse management [3]. Combining target recognition and accurate position information is generally beneficial to these applications. The researches on RFID-based localization technology [4, 5] have important practical significance.

In recent years, scholars have conducted a lot of research on how to use RFID tags to achieve localization. In the RFID-based localization system, one or more tags are attached to an object. The received signal strength (RSS) and the phase information of tags are exploited by various localization systems. In [6, 7], they deploy reference tags with fixed position in advance. The target's position is estimated by k reference tags whose RSS are the most similar to that of target tag. However, the RSS is not a reliable indicator, it is susceptible to the tag's orientation, antenna gain, and multipath environment. Thus, the RSS-based system is difficult to achieve fine-grained localization. Although some RSS-based methods can achieve high localization accuracy, these methods need to deploy a considerable number of reference RFID tags [8, 9]. Compared with RSS, the phase of the signal has higher resolution and noise tolerance, which is a better choice to realize high accuracy localization. In [10], the phase differences between two antennas are collected to build hyperbolas. The position of the tag can be estimated by the intersections of these hyperbolas. Another localization method based on the direction of tag is proposed in [11]; the phase difference collected by antenna array is used to estimate the direction of the tag. In the phase-based methods, to solve the ambiguity introduced by the phase, the spacing between two antennas should be within half wavelength, and it is quarter wavelength for the commercial off-the-shelf (COST) UHF RFID with monostatic antennas as it is a round-trip in this system. However, the

size of commonly-used circular polarization antenna is hard to meet this stringent requirement. Such a COST reader cannot satisfy this spacing constraint. To solve this problem, RFID tag array is undoubtedly a better choice.

Compared with a reader antenna, the size of RFID tags is smaller; the distance between two tags can be easily restrict within quarter wavelength or narrower. Moreover, benefitting from the low cost of the RFID tag, the tag array has a lower cost contrasted with the antenna array [12, 13]. In recent years, tag arrays have been widely used in 3D reconstruction [14], gesture recognition [15], and human motion sensing [16, 17]. Besides, tag array is also a great choice for indoor localization systems. Yang et al. localized the target by the AOAs of two antennas which are estimated from the phase information of the tag array [18]. However, when the orientation angle (OA) of the tag array is unknown, the localization accuracy will deteriorate dramatically. Therefore, how to estimate the OA tag array is a key issue. To solve this problem, this paper proposes a novel AOA-based indoor localization system, which can achieve an accuracy of decimeter-level. The main contributions of this paper are as follows:

(i) Different from the traditional tag array localization model, the OA of the tag array is taken into consideration in our localization model, and targets can be localized with an unknown OA

(ii) An orientation angle (OA) retrieval algorithm for the tag array is proposed, which uses AOA estimation results of multiple reader antennas to accomplish the OA retrieval for tag arrays

2. Phase Difference in RFID Tag Array

Most COST RFID readers can collect the phase information of the tag when inventorying the tag, as shown in Figure 1. The measured phase value is the offset of the signal sent and received by the reader antenna which includes the phase introduced by the round-trip propagation between the reader and tag and the phase error introduced by the reader transmitter circuit, the receiver circuit, and the tag. Besides, the measured phase is a value between 0 and 2π due to phase wrapping. The measured value cannot directly represent the distance information between the antenna and the tag. The received phase can be modelled as [19]

$$\varphi_i = \mathrm{mod}\left(\frac{4\pi r}{\lambda} + \varphi_{Tx} + \varphi_{Rx} + \varphi_{\mathrm{tag}}, 2\pi\right), \quad (1)$$

where λ is the wavelength, r is the distance between tag and reader antenna, φ_{tag} is the phase error introduced by tag's reflection characteristics, and φ_{Tx} and φ_{Rx} are the phase errors introduced by transmitter and receiver, respectively.

In general, most of COTS RFID readers work with monostatic antennas that simultaneously transmit RF signals to power up RFID tags and then receive their backscatter signals. For the measured phases of two tags collected by one antenna, the errors introduced by transmitter and receiver are the same. Meanwhile, to eliminate the phase ambiguity between two tags, we restrict the distance d between two adjacent tags within $\lambda/4$, which means that the difference of two tags' round-trip distance is less than λ [20]. According to equation (1), the phase difference can be represented as

$$\Delta\varphi_{i,i+1} = \varphi_i - \varphi_{i+1} = \frac{4\pi(r_i - r_{i+1})}{\lambda} + \varphi_{\mathrm{tag}_i} - \varphi_{\mathrm{tag}_{i+1}}. \quad (2)$$

The errors introduced by the tag arise from the tag's circuit, mutual coupling, and the signal incident direction [20], which can be regarded as

$$\varphi_{\mathrm{tag}} = \varphi_{\mathrm{sel}} + \varphi_{\mathrm{in}} + \varphi_{\mathrm{cop}}, \quad (3)$$

where φ_{sel}, φ_{in}, and φ_{cop} are the errors introduced by tag's circuit, signal incident direction, and mutual coupling between tags. Tags of the same type have similar circuit construction, so $\varphi_{\mathrm{sel}i} \approx \varphi_{\mathrm{sel}j}$. For the same reader antenna, the signal incidence direction for two very close tags (for the UHF RFID $d \le \lambda/4 \approx 8.11$ cm) can be regarded as equal. From Tagyro [21], strong mutual coupling in the tag array will affect the measured phase significantly and make it hard to detect the tags. When the spacing of the tags increases, the mutual coupling on measured phase difference is limited. For the same pair of tags, their mutual coupling is identical [22]; the measured phase difference between two tags is independent to the coupling voltage and impedance [18]. Although mutual coupling has a great effect on the measured phase, the effect of mutual coupling on the phase difference can be attenuated significantly, which has less impact on localization.

3. System Design

3.1. Localization Model. The OA of the tag array is not considered in the traditional tag array-based localization models [18], in which it is assumed that the AOA estimation result θ_i' of the tag array is equal to the direction angle β_i of the tag array relative to the antenna i. According to the geometric relationship, the positioning of the tag array can be estimated by the AOA estimation of the two reader antennas of the tag array. However, in a real localization scenario, the OA of the tag array cannot be known in advance, and using this localization model will generate a large localization error. To solve this problem, a localization method based on pose retrieval of tag arrays is proposed in this paper.

In our system, we deployed M antennas and a RFID tag array in the environment. N tags are uniformly attached to the target to form a uniform linear RFID tag array (ULA), and all tags can be inventoried from both sides. As shown in Figure 2, the tag array is used to acquire the AOA (θ_i') between the center of the array and antenna-i, the OA of the tag array is α, and the DA of the tag array for antenna-i is β_i. Theoretically, based on $M\,(M \ge 3)$ AOA values and

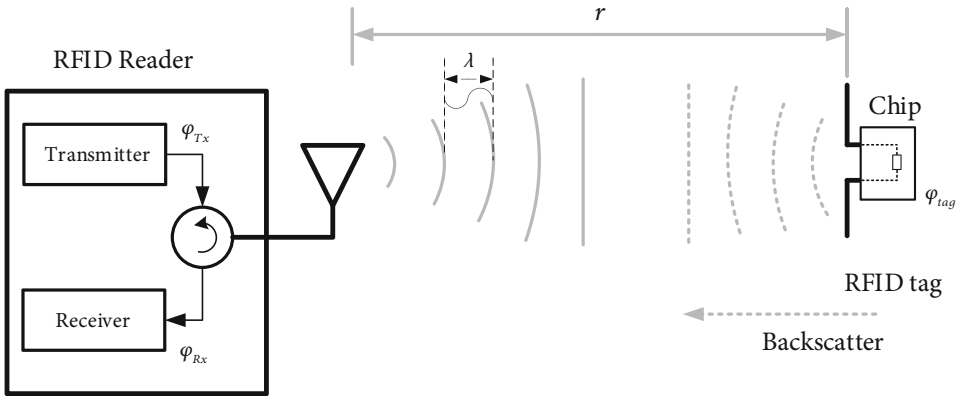


Figure 1: Illustration of the RFID backscatter link.

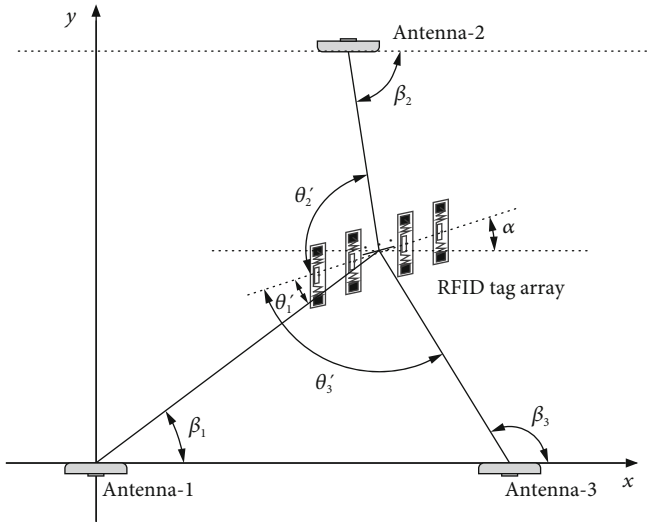


Figure 2: Localization model.

the fixed reader antenna positions, the OA of tag array α can be accurately determined. The DAs of antennas are

$$\begin{cases} \beta_1 = \theta_1' + \alpha, \\ \beta_2 = \theta_2' - \alpha, \\ \beta_3 = \theta_3' + \alpha. \end{cases} \tag{4}$$

If the positions of antennas are fixed, the idea is to use the DAs of each antenna to determine the position of tags. Supposing that the position of the antenna is (x_i, y_i) and the tag array's position is (x_c, y_c), the DAs of antennas satisfy

$$\begin{cases} \cot(\beta_1) = \dfrac{y_c - y_1}{x_c - x_1}, \\ \cot(\beta_2) = \dfrac{y_c - y_2}{x_c - x_2}, \\ \cot(\beta_3) = \dfrac{y_c - y_3}{x_c - x_3}. \end{cases} \tag{5}$$

Since the phase measurements would be inevitably influenced by various noise sources, three lines defined by β_1, β_2, and β_1 will not intersect at one point. Three different solutions $\{(x_{c1}, y_{c1}), (x_{c2}, y_{c2}), (x_{c3}, y_{c3})\}$ can be obtained from Equation (5), and the average value is taken as the localization result of the tag array.

The main challenge of the system is how to effectively estimate the AOA values of antennas and the OA of tag array. In this paper, a new tag array-based localization system is designed, which can estimate the OA of the array. The details of tag array AOA estimation and OA estimation will be discussed in the following sections.

3.2. Tag Array AOA Estimation. The scenario of AOA estimation for single antenna is shown in Figure 3. The spacing d between adjacent tags in the array is restricted to be within $\lambda/4$, which is much less than the distance r between reader antenna and tag array ($d \ll r$) and satisfies the far-field condition. The signal arrived at the tag array can be regarded as plane wave. Assuming that there are K paths in the indoor

environment, the received signal backscattered from tag array at time t can be modelled as

$$X(t) = \sum_{i=1}^{K} a(\theta_i)s_i(t) + n_i(t) = As(t) + n(t) \quad (6)$$

where θ_i is xx, $A = [\alpha(\theta_1), \quad \alpha(\theta_2), \cdots, \alpha(\theta_K)]$ is the manifold matrix, $s(t) = [s_1(t), s_2(t), \cdots, s_K(t)]^T$ is the source signal vector, and $n(t)$ is the noise. Taking the first element of the array as the reference, the received signal can be reconstructed by the phase difference:

$$\widehat{X}(t) = \left[1, e^{-j\varphi_{12}(t)}, \cdots, e^{-j\varphi_{1k}(t)}, \cdots, e^{-j\varphi_{1N}(t)}\right]^T, \quad (7)$$

where $\varphi_{1k}(t)$ is the phase difference of the k -th tag to the first reference tag. As the tags in the array are close to each other, the multipath effect of each tag is similar. Thus, the phase difference can reduce both the multipath effect and the coupling effect. The RFID communication is a round-trip link, so the steering vector for the uniform tag array can be expressed as

$$\alpha(\theta) = \left[1 \quad e^{-j4\pi(d/\lambda)\sin\theta} \quad , \cdots, \quad e^{-j4\pi((N-1)d/\lambda)\sin\theta}\right]^T. \quad (8)$$

Due to the coherent multipath signal, the steering vector and the noise subspace are not completely orthogonal, which affects the estimation of the signal source direction. To further reduce the effect of coherent multipath, spatial smoothing method is adopted. The uniform linear tags array is divided into p subarrays, and the number of elements in each subarray is M $(M \geq K + 1)$. The covariance matrix of l -th subarray is given by

$$R_l = A_M(\theta)D^{(l-1)}R_s\left(D^{(l-1)}\right)^H A_M^H + \sigma^2 I, \quad (9)$$

where $D = \text{diag}\left[e^{-j4\pi d\sin\theta_1/\lambda}, e^{-j4\pi d\sin\theta_2/\lambda}, \cdots, e^{-j4\pi d\sin\theta_K/\lambda}\right]$ is called displacement operator. A_M is a $M \times K$ manifold matrix of the sub-array, R_s is the covariance matrix of signal, and I is the unit matrix and σ^2 is the power of noise. The covariance matrix after spatial smoothing is

$$R_s^f = \frac{1}{p}\sum_{i=1}^{p} D^{(i-1)}R_s\left(D^{(i-1)}\right)^H. \quad (10)$$

Performing eigendecomposition on R_s^f, equation (8) can be rewritten as

$$R_s^f = U_s\Sigma_s U_s^H + U_N\Sigma_N U_N^H, \quad (11)$$

where U_s is the signal subspace formed by the eigenvector associated with the large eigenvalue and U_N is the noise subspace corresponding to the small eigenvalues. By searching all arrival vectors that are orthogonal to the noise subspace,

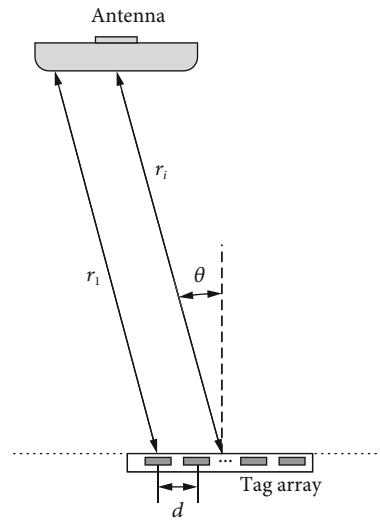


Figure 3: Antenna's AOA estimation.

the direction of antenna can be estimated by

$$\theta_{Tx} = \max_{\theta} \frac{1}{A_M(\theta)^H U_N \mathrm{U}_\mathrm{N}^\mathrm{H} A_M(\theta)}. \quad (12)$$

According to the geometric relationship, the AOA of antenna is $\theta' = \pi/2 - \theta_{Tx}$.

3.3. OA Estimation. In this part, we will provide the details of our proposed OA retrieval algorithm based on a tag array. In an $x - y$ coordinate system, OA is the angle of intersection between a line parallel to the positive x-axis and the line going through tag array. Instead of viewing the antennas A_1, A_2, A_3 in $x - y$ coordinate system, we construct virtual antennas VA_1, VA_2, and VA_3 in $x' - y'$ coordinate system. As shown in Figure 4, the $x' - y'$ coordinate is obtained by rotating $x - y$ coordinate counterclockwise through the origin by $\alpha(m) = (m - 1) \times \Delta\alpha - \pi/2$, where $\Delta\alpha$ is the rotation interval and m is an integer. Assuming that the OA of tag array is limited, the range of $\alpha(m)$ is $[-\pi/2, \pi/2]$, m is in $[1, M]$. $\alpha(m)$ represents the searching start point when $m = 1$; M is inversely proportional to the rotation interval $\Delta\alpha$ and equals to $[\pi/\Delta\alpha]$, where $[\cdot]$ is the rounding operation.

As shown in Figure 4, assuming that the position of antennas in $x - y$ coordinate is $A_i = (x_i, y_i)$, after rotating an angle $\alpha(m)$, the new position in $x' - y'$ coordinate is $A_i(m) = (x_i(m), y_i(m))$, and the transformation from A_i to $A_i(m)$ can be written as

$$\begin{bmatrix} x_i(m) \\ y_i(m) \end{bmatrix} = R[\alpha(m)] \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad (13)$$

where $R[\alpha(m)]$ is the counterclockwise rotation matrix, which can be described as

$$R[\alpha(m)] = \begin{bmatrix} \cos a(m) & \sin\alpha(m) \\ -\sin\alpha(m) & \cos\alpha(m) \end{bmatrix}. \quad (14)$$

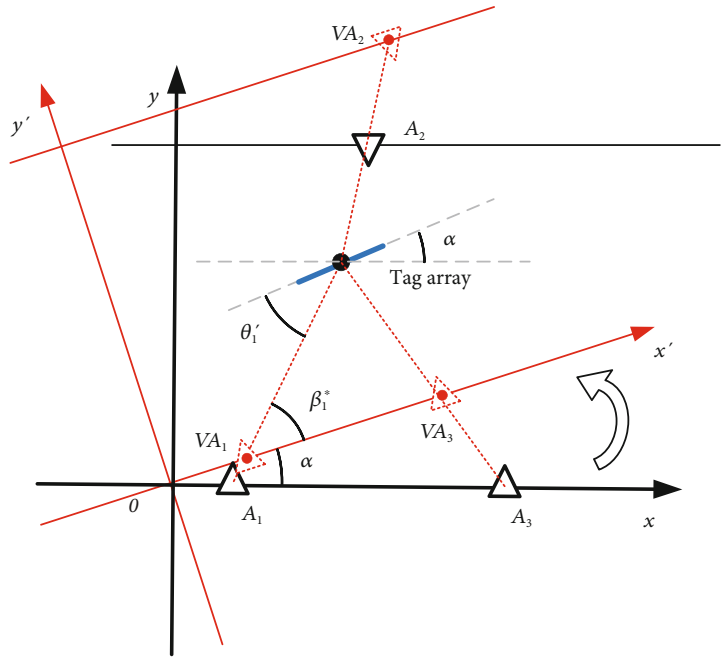


Figure 4: Rotation of coordinate.

The position of virtual antenna $VA_i(m)$ is defined as the intersection of x'-axis and the line $l_i(m)$. $l_i(m)$ is built by antenna $A_i(m)$, and it is AOA value θ_i', which is obtained by equation (12). In the $x' - y'$ coordinate, it can be expressed as

$$\begin{cases} l_1(m)\colon y' = \tan\theta_1' \times \left[x' - x_1(m)\right] + y_1(m), \\ l_2(m)\colon y' = \tan\left(\pi - \theta_2'\right) \times \left[x' - x_2(m)\right] + y_2(m), \\ l_3(m)\colon y' = \tan\theta_3' \times \left[x' - x_3(m)\right] + y_3(m). \end{cases} \tag{15}$$

As shown in Figure 5, we deployed three antennas in the localization system. The three lines $\{l_1(m), l_2(m), l_3(m)\}$ built from Equation (15) will have three intersections $\{C_1(m), C_2(m), C_3(m)\}$. The sum of Euclidean distance between the three intersections is defined as $dist(m)$:

$$\text{dist}(m) = \|C_1(m) - C_2(m)\| + \|C_1(m) - C_3(m)\| + \|C_2(m) - C_3(m)\|. \tag{16}$$

Theoretically, x'-axis is parallel to the tag array when the rotation angle of the $x' - y'$ coordinate equal to the OA of tag array. In this case, as shown in Figure 4, we have $\beta_i(m) = \theta_i'$; the DA is equal to AOA for the same antenna; the three lines will have a unique intersection $\text{dist}(m) = 0$. However, when taking the noise effects into consideration, the three lines will no longer intersect to one point. Thus, a method

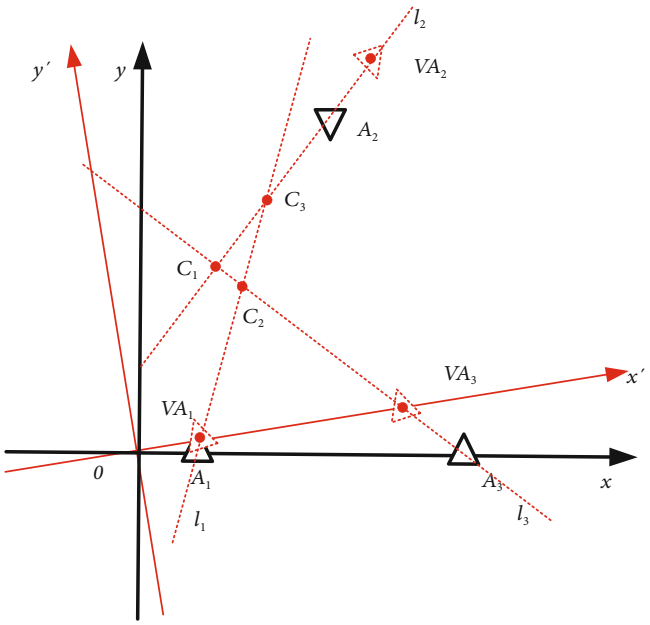


Figure 5: Orientation angle estimation.

based on the rotation of coordinate is employed to estimate the index of OA

$$m^* = \underset{m\in[-90,90]}{\text{argmin}}\left(\text{dist}(m)\right). \tag{17}$$

4. Experiments Results

Experiments were carried out in a typical indoor scenario with Impinj R420 UHF RFID reader and C90G tags (with Monza 4QT chip). The reader can extract information from

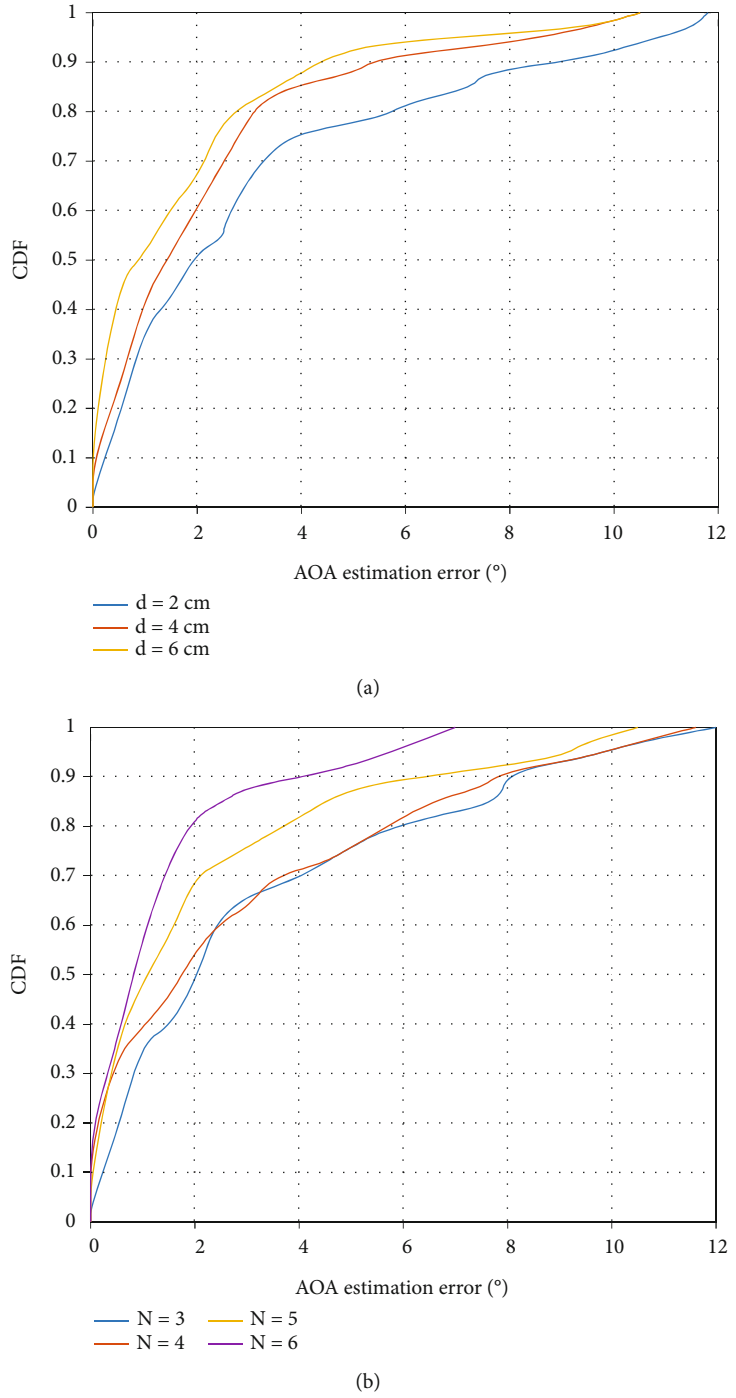


Figure 6: Evaluation of AOA estimation: (a) AOA estimation error of different tag spacing; (b) AOA estimation error of different number of tag array.

tags, including phase angle, RSSI, and Doppler shift based on a Low-Level Reader Protocol (LLRP). A Lenovo laptop with i7-5500u CPU is used for signal processing; the software used in the experiment is developed based on official library, which allows the computer to communicate with the RFID reader. Similar to [23], the experiment considered $2\pi - \varphi_i{}'$ as φ_i, where $\varphi_i{}'$ is the phase value measured by Impinj R420. Besides, to determine some of the fixed variables, several detailed benchmarks are established.

4.1. AOA Estimation. The AOA estimation result of the RFID tag array is the basis for localization. In order to evaluate impacts of the tag spacing and tag number of the RFID tag array on the AOA estimation performance, experiments

Figure 7: Experiment setup.

were carried out by placing a reader antenna at a fixed distance of 1.5 m in front of the tag array, and the tag array was rotated in the range of [-70°, 70°] with 5° interval.

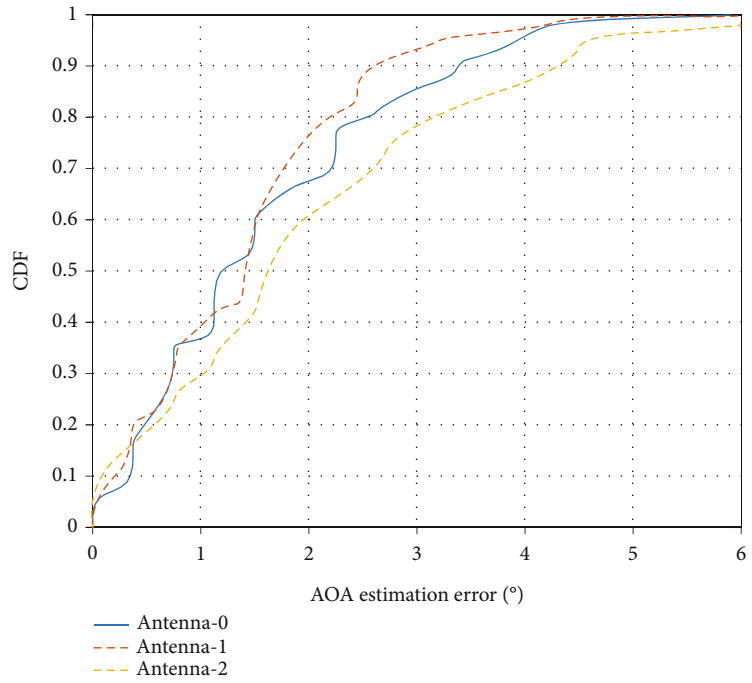
4.1.1. Impact of Tag Spacing. The mutual coupling between the tags in the RFID tag array is an important factor affecting the phase of the backscattered signal, and this effect will gradually decrease with the increase of the tag spacing. To explore the effect of tag spacing on the AOA estimation, four RFID tags were constructed with different spacings to construct a tag array. Experiment results are shown in Figure 6(a); the array with larger spacing has higher AOA estimation accuracy than the case with smaller spacing. The average AOA estimation error corresponding to the spacing of 2 cm, 4 cm, and 6 cm is 3.3°, 2.8°, and 2.19°, respectively.

4.1.2. Impact of Tag Number. The number of array elements in the MUSIC algorithm is also a key factor affecting the accuracy of AOA estimation. In the experiment, the tag spacing was set to 4 cm. The results of AOA estimation with different tag number are shown in Figure 6(b). It can be seen that with the increase of the tag number, the error of the AOA estimation of the tag array gradually decreases. The average errors of the AOA estimation for the tag number of 3, 4, 5, and 6 are 3.45°, 3.15°, 2.42°, and 1.77°, respectively.

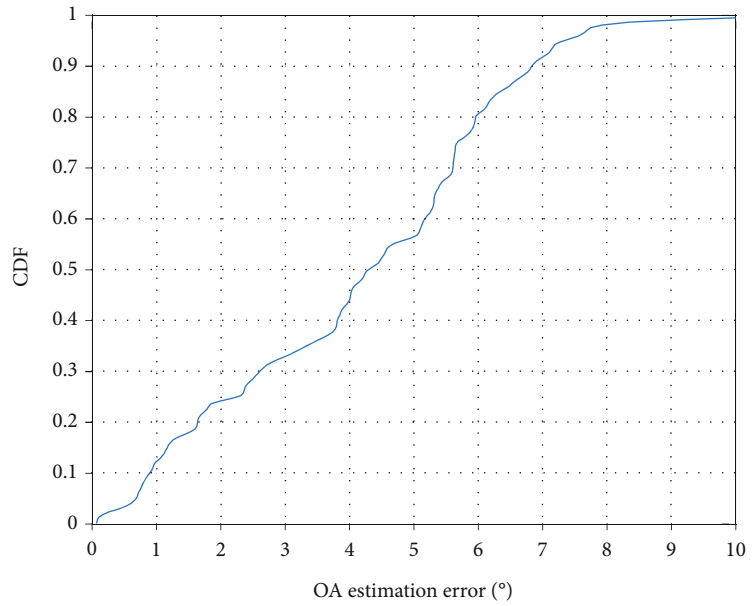
4.2. Localization. To evaluate localization algorithm based on RFID tag array, we built a real localization system in a 4 m × 4 m indoor area. The reader part consists of three circular polarized antennas and a Impinj R420 RFID reader. As shown in Figure 7, considering the width of the target object, six tag are adopted for the RFID tag array, where the spacing d is 4 cm. The experiment environment is shown in Figure 7; the positions of antennas are $A_0(0.5\,\text{m}, 0\,\text{m})$, $A_1(2\,\text{m}, 4\,\text{m})$, and $A_2(3.5\,\text{m}, 0\,\text{m})$, which are distributed on the boundary of the measurement area.

Under the condition that each tag in the tag array can be read correctly, we carried out a series of experiments in different positions with different OAs, the statistical result is shown in Figure 8. The CDFs of the AOA and OA estimation error are shown in Figure 8; for the three antennas, the mean AOA estimation error is about 2.7°. Antenna-1 is at the center of the boundary, so the AOA estimation is slightly better than other two antennas. For the OA estimation result shown in Figure 8(b), 80% OA estimation error is less than 6°, and the mean OA error is 4.35°. The CDF of localization error is shown in Figure 9(a), the mean x-axis error and y-axis error are 0.15 m and 0.13 m, respectively. Furthermore, the average localization error is 0.216 m. As shown in Figure 9(b), compared with the traditional AOA localization algorithm in [18], the OA estimation-based localization algorithm has higher positioning accuracy. The mean localization errors for the proposed algorithm and Ref. [18] are 0.216 m and 0.342 m, respectively. The proposed method can achieve higher localization accuracy by taking the OA of the tag array into consideration when compared to Ref. [18].

4.3. Cost of Localization. In the RFID localization system proposed in this paper, an ULA is constructed by RFID tags. Benefiting from the low cost of UHF RFID tags, the cost of the proposed system is lower than those of other localization systems. As shown in Table 1, we compared the hardware cost and average localization accuracy of several different localization systems. Compared with commercial equipment, the cost of customized equipment is generally higher.



(a)



(b)

Figure 8: Results of angle estimation: (a) CDF of AOA estimation error; (b) CDF of OA estimation error.

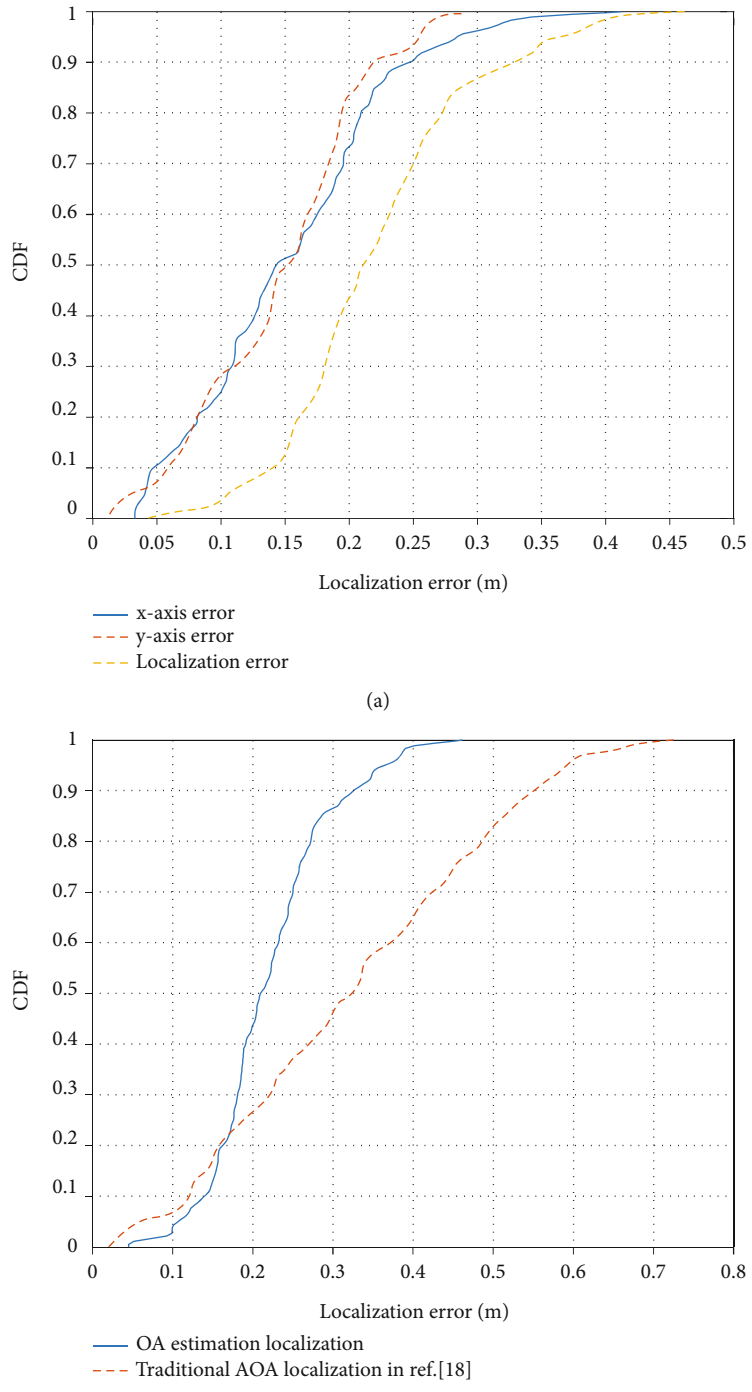


Figure 9: Localization results: (a) CDF of localization error; (b) comparison with the traditional localization algorithm.

Table 1: Comparison of localization systems.

Localization system	Hardware cost	Mean accuracy
LANDMARC [7]	4 off-the-shelf reader antennas, 1 commercial RFID reader	1~2 m
Azzouzi et al. [24]	3 custom antenna arrays, 1 commercial RFID reader	0.21 m
Povalac [25]	1 custom RFID front-end prototype, 1 commercial RFID reader	0.14 m
Peng et al. [26]	8 cost RFID readers, 2 commercial RFID readers	0.38 m
This paper	3 COST RFID antennas, 1 commercial reader	0.21 m

Among them, [24, 25] use customized special equipment, which undoubtedly increases the cost required for positioning. Systems in [7, 26] and this paper only require simple commercial equipment, and the localization cost is lower. The localization accuracy of the system proposed in this paper is higher than those in [7, 26]. Moreover, the system in this paper does not need to perform complex reference tag library construction in advance, which also greatly reduces the human cost.

5. Conclusions

This paper proposes a localization method based on AOA and orientation angle for passive tag array. By analyzing the relationship among AOA, OA of the tag array, and DA between each reader antenna and the geometric center of the tag array, an OA retrieval method based on rotating coordinate system is proposed to estimate DA. And the tag array position can be determined by antennas and their DAs. Simulation and experiment results show that the mean accuracy of the proposed algorithm is 0.216 m.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors do not have any possible conflicts of interest.

Acknowledgments

This work was supported partly by the General Program of Chongqing Natural Science Foundation (Special Program for the Fundamental and Frontier Research) (cstc2019jcyj-msxmX0108) and the National Natural Science Foundation of China (61704015).

References

[1] X. Wang, J. Zhang, Z. Yu, E. Mao, S. C. G. Periaswamy, and J. Patton, "RFThermometer: a temperature estimation system with commercial UHF RFID tags," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, Shanghai, China, 2019.

[2] P. Podduturi, T. Maco, P. Ahmadi, and K. R. Islam, "RFID implementation in supply chain management," *International Journal of Interdisciplinary Telecommunications and Networking*, vol. 12, no. 2, pp. 34–45, 2020.

[3] J. Zhao, F. Xue, and D. A. Li, "Intelligent management of chemical warehouses with RFID systems," *Sensors*, vol. 20, no. 1, pp. 123–123, 2020.

[4] J. Lai, C. Luo, J. Wu et al., "TagSort: accurate relative localization exploring RFID phase spectrum matching for Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 389–399, 2020.

[5] Y. Ma, C. Tian, and Y. Jiang, "A multitag cooperative localization algorithm based on weighted multidimensional scaling for passive UHF RFID," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6548–6555, 2019.

[6] A. Chattopadhyay and A. R. Harish, "Analysis of low range indoor location tracking techniques using passive UHF RFID tags," in *2008 IEEE Radio and Wireless Symposium*, pp. 351–354, Orlando, USA, 2008.

[7] L. M. Ni, Y. Liu, Y. Lau, and A. P. Patil, "LANDMARC: indoor location sensing using active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2003.

[8] J. Wang and K. D. Dude, "Where's my card? RFID positioning that works with multipath and non-line of sight," in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pp. 51–62, Hong Kong, China, 2013.

[9] Y. Zhao, Y. Liu, and L. M. Ni, "VIRE: active RFID-based localization using virtual reference elimination," in *2007 International Conference on Parallel Processing (ICPP 2007)*, pp. 55-56, Xi'an, China, 2007.

[10] F. Xue, J. Zhao, and D. Li, "Precise localization of RFID tags using hyperbolic and hologram composite localization algorithm," *Computer Communications*, vol. 157, pp. 451–460, 2020.

[11] Y. Ma, B. Wang, S. Pei, Y. Zhang, S. Zhang, and J. Yu, "An indoor localization method based on AOA and PDOA using virtual stations in multipath and NLOS environments for passive UHF RFID," *IEEE Access*, vol. 6, pp. 31772–31782, 2018.

[12] Y. Mohamedatni, B. Fergani, J. M. Laheurte, and B. Poussot, "New methodology for the short range localization of UHF RFID tags using a linear uniform array," *International Journal of RF Technologies*, vol. 10, no. 1-2, pp. 9–26, 2019.

[13] E. DiGiampaolo and F. Martinelli, "Multiple baseline synthetic array for UHF RFID localization," in *2019 IEEE International Conference on RFID Technology and Applications (RFID-TA*, pp. 348–352, Pisa, Italy, 2019.

[14] Y. Bu, L. Xie, J. Liu, B. He, Y. Gong, and S. Lu, "3-dimensional reconstruction on tagged packages via RFID systems," in *2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1–9, San Diego, USA, 2017.

[15] S. Amendola, V. Di Cecco, and G. Marrocco, "Numerical and experimental characterization of wrist-fingers communication link for RFID-based finger augmented devices," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 1, pp. 531–540, 2019.

[16] Z. Wang, F. Xiao, N. Ye, R. Wang, and P. Yang, "A see-through-wall system for device-free human motion sensing based on battery-free RFID," *ACM Transactions on Embedded Computing Systems*, vol. 17, no. 1, pp. 1–21, 2018.

[17] Y. Bu, X. Lei, Y. Gong, C. Wang, and S. Lu, "RF-dial: an RFID-based 2D human-computer interaction via tag array," *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 837–845, Honolulu, USA, 2018.

[18] C. Yang, X. Wang, and S. Mao, "SparseTag: high-precision backscatter indoor localization with sparse RFID tag arrays," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1–9, Boston, MA, USA, 2019.

[19] Impinj Support Portal, "Low level user data support," *Impinj Speedway Revolution Reader Application*, 2013, https://support.impinj.com.

[20] T. Liu, Y. Liu, L. Yang, Y. Guo, and C. Wang, "BackPos: high accuracy backscatter positioning system," *IEEE Transactions on Mobile Computing*, vol. 15, no. 3, pp. 586–598, 2016.

[21] T. Wei and X. Zhang, "Gyro in the air," *GetMobile: Mobile Computing and Communications*, vol. 21, no. 1, pp. 35–38, 2017.

[22] F. Lu, X. Chen, and T. T. Ye, "Performance analysis of stacked RFID tags," in *2009 IEEE International Conference on RFID*, pp. 330–337, Orlando, USA, 2009.

[23] Y. Zeng, X. Chen, R. Li, and H. Z. Tan, "UHF RFID indoor positioning system with phase interference model based on double tag array," *IEEE Access*, vol. 7, pp. 76768–76778, 2019.

[24] S. Azzouzi, M. Cremer, U. Dettmar, R. Kronberger, and T. Knie, "New measurement results for the localization of UHF RFID transponders using an Angle of Arrival (AoA) approach," in *2011 IEEE International Conference on RFID*, pp. 91–97, Orlando, USA, 2011.

[25] A. Povalac and J. Sebesta, "Phase difference of arrival distance estimation for RFID tags in frequency domain," in *2011 IEEE International Conference on RFID-Technologies and Applications*, pp. 188–193, Sitges, Spain, 2011.

[26] C. Peng, H. Jiang, and L. Qu, "Deep convolutional neural network for passive RFID tag localization via joint RSSI and PDOA fingerprint features," *IEEE Access*, vol. 9, pp. 15441–15451, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1539056, 10 pages
https://doi.org/10.1155/2022/1539056

Research Article

Pole Extraction of Radar Target in Resonant Region Based on Sliding-Window Matrix Pencil Method

Zhian Deng[1,2], **Tianbao Zhang**[1,2], **Na Li**[3], **Chunjie Zhang**[1,2] and **Weijian Si**[1,2]

[1]*College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China*
[2]*Key Laboratory of Advanced Marine Communication and Information Technology, Ministry of Industry and Information Technology, Harbin 150001, China*
[3]*Sichuan Jiuzhou Electric Group Co., Ltd, Mianyang 621000, China*

Correspondence should be addressed to Na Li; 58630950@qq.com

Received 7 December 2021; Accepted 13 April 2022; Published 9 May 2022

Academic Editor: Andrej Hrovat

In the complex battlefield environment, stealth radar target recognition has been paid increasing attentions. Previous studies have demonstrated that the stealth target can be identified well by pole extraction based on matrix pencil method (MPM). However, MPM suffers from the difficulty in setting model order and the time-domain resonant response aliasing problem. This paper proposes a new sliding-window matrix pencil method (SW-MPM) based on dynamic order setting and sliding window. Dynamic order setting scheme is used to overcome the difficulty of setting the order in matrix pencil method, while sliding window can avoid the aliasing problem of time-domain resonant response to some extent. The time-domain scattering data used in SW-MPM are obtained by inverse Fourier transform of the frequency domain scattering data in the resonance region. Taking typical stealth aircraft identification as example, the simulation results verify that the proposed method may extract more number of poles with better azimuth consistency, which is beneficial to improve the accuracy of pole-based target identification.

1. Introduction

In modern electromagnetic spectrum warfare, radar target recognition has attracted more and more attentions [1–4]. Identification of the enemy targets accurately and timely, especially the stealth aircrafts, is the prerequisite for destroying the enemy targets, and any confusion between civilian or friendly targets and enemy targets may cause tragedy [5]. Radar target recognition is closely related to target scattering echoes, and for different targets, the excited characteristics are also different. For typical stealth aircrafts in the optical region, it is difficult to identify them, since radar cross section (RCS) is small according to the target scattering theory [6, 7]. In contrast, for target in the resonance region [8], with the radar signal wavelength close to the target size, RCS is always larger and easy to identify. Moreover, poles extracted from target scattered echoes are very beneficial to target recognition. Poles are mainly related to the size and shape of the target and thus are not sensitive to the target relative attitude and radar polarization [9, 10]. Therefore, poles are considered as robust and effective for the recognition of stealth targets [11].

The first pole extraction algorithm is Prony algorithm [12] presented in 1975. Prony algorithm combined with the singular value decomposition (SVD) can improve the antinoise ability [13]. Prony-based method is sensitive to noise, because it relies on an accurate estimation of the number of target poles, which is difficult to estimate. Matrix pencil method (MPM) [14] is the most widely used method for pole extraction, which has better accuracy, stability, and antinoise performance; compared with Prony method, MPM only needs to construct Hankel matrix with time-domain echo data of the target and then calculate the generalized eigenvalues of the matrices to obtain the target poles. In order to improve the antinoise performance of MPM, Sarkar et al. conducts SVD on the data matrix [15, 16].

The low-rank matrix approximation can suppress the impact of noise and reduce algorithm calculation cost significantly.

In recent years, a lot of research has been devoted to how to apply MPM to pole extraction for radar target [17–19] in real applications. In general, it is difficult to get the actual time-domain echo data, which can be obtained by transforming frequency-domain excited data through simulation. The data transformation from frequency domain to time domain involves the setting of sampling interval, which can be adaptively determined to reduce the computational complexity and time-domain data required [20]. Moreover, in order to get more poles, the wide-band data are required, while in real applications, the data are often narrowband. Thus, Chauveau et al. propose a method [21] based on narrowband data, which avoids extracting target poles using wide-band data. The number of target poles extracted from narrowband data is small, but all of them are main poles with high precision. Some researches make full use of the azimuth consistency of poles to improve the antinoise performance of poles. These methods construct correlation matrices from echo data of multiple directions and then extract target poles by MPM.

Though MPM is one of the most widely used method, it still faces two problems in the pole extraction of complex target [22, 23]. Firstly, the order of the model, namely, the number of target poles, is difficult to determine. The number of poles can be known for simple ideal conductor targets with theoretical solutions, but for complex targets, the number of poles is unknown in advance. An inaccurate estimation of the number of poles may degrade the accuracy of pole extraction greatly. If the number of pole is too small, the real poles of the target will be missed and the extracted poles will have a big deviation from the real poles. If the number of pole is too large, many false poles will be generated and it is impossible to determine which poles are the true ones.

Secondly, it is difficult to distinguish the early time and the late time [24–26] of the target resonant response in time domain. The early time occurs when the wave front of the incident signal interacts with the target and ends when the incident signal completely leaves the target. After the incident signal completely passes through the target, the late time is generated by the gradual decay of the excited current on the target. Due to the propagation delay, there will be an aliasing of early and late time.

In this paper, we propose a new sliding-window matrix pencil method (SW-MPM) based on dynamic order setting and sliding window. Simulations are carried to verify the effectiveness and better azimuth consistency of the proposed method. Our main contributions can be summarized as the following two aspects.

Firstly, we design a sliding window to capture time-domain resonant response for initial pole extraction. The start time of sliding window is estimated based on model of late-time stage and calculation of late-time commencement for various incident wave directions. Then, multiple sets of poles are extracted by sliding through the signal sequence and applying MPM for each sliding window. Among them, the pole whose statistical frequency exceeds a certain threshold is considered as an initial extracted pole. The diversity of different Hankel matrices generated by sliding window can reduce negative effect of time-domain resonant response aliasing, thus improving the reliability and robustness of pole extraction.

Secondly, we obtain the final stable poles by dynamic order setting on initial extracted pole sets. Relevant studies show that the true poles will converge with the variation of order, while false poles are irregularly distributed in the complex plane. Therefore, we traverse model order and generate another set of poles on the basis of initial pole extraction. Histogram statistics of poles extracted from diverse orders are carried out, and any pole whose related occurrence frequency exceeds another threshold is considered as a true pole.

2. Target Time-Domain Resonant Response Acquisition

The radar target time-domain resonant response can be considered as an output of a linear time-invariant radar system. In actual radar system, we use a narrow pulse signal to illuminate the target, and the scattering response signal at this time is the time-domain resonant response. However, it is difficult to get the actual measurement data of military targets, especially for stealth aircraft. In this paper, a professional electromagnetic simulation software, is used to generate the frequency-domain scattering data based on geometry model of radar target. Then, the time-domain resonant response is obtained through time-frequency data transformation. Target time-domain resonant response can be subdivided into early time and late time. The singularity expansion method (SEM) proves that the late time contains the target pole information. This section first briefly introduces singularity expansion theory and then describes the process of the time-frequency data transformation. Finally, a physical model describing early-time and late-time stages of a simple target is given.

2.1. Singularity Expansion Theory. Singularity expansion theory is an important theory when the late-time stage of radar echo is studied. The theory systematically describes the echo characteristics, transfer function, and properties of radar targets in the resonance region. Time response of the radar target in the resonance region can be expressed as the sum of attenuation exponents on the complex plane [27, 28],

$$y(t) = \sum_{i=1}^{M} R_i e^{s_i t}, \tag{1}$$

where M is the number of poles, R_i is the residue, and $s_i = \sigma_i + j\omega_i$ is the pole. σ_i and ω_i represent the attenuation factor and the attenuation frequency, respectively.

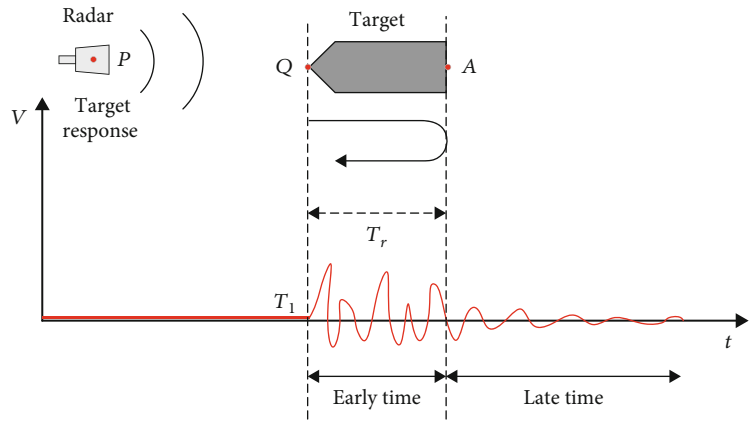


Figure 1: Early- and late-time boundaries.

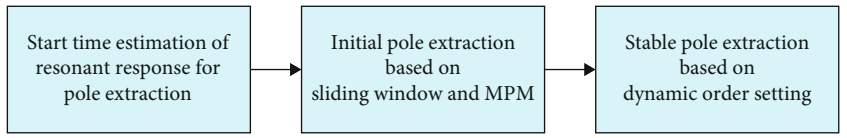


Figure 2: Overview of the proposed sliding-window matrix pencil method.

In Laplace-domain, the transfer function of radar target scattering can be expressed as the sum of a rational function and an integral function.

$$H(s) = \sum_{i=1}^{M} \frac{R_i}{s - s_i} + C(s). \tag{2}$$

The integral part $C(s)$ corresponds to the early-time stage of the system in time domain, which describes the scattering characteristics of the radar target. The rational part corresponds to the late-time stage, which describes the pole characteristics of the radar target.

2.2. Frequency-Domain Data. The frequency-domain response of the radar target can be obtained either by simulation calculation or actual measurement. It is usually difficult to obtain actual measurement data for complex military targets, so electromagnetic simulation software, such as FEKO, is generally used to calculate scattering data by moment method [29–31].

The scattering data in the frequency-domain can be expressed as

$$\widehat{E}_s(\omega) = H(\omega), 0 < \omega < \omega_c, \tag{3}$$

where ω_c is the maximum frequency and $H(\omega)$ is the impulse response of the target in the frequency domain. If the time-domain incident signal can be expressed as a Sinc function

$$e_i(t) = Sa(\omega_c t), \tag{4}$$

the frequency-domain expression of scattering echo is

$$E_s(\omega) = E_i(\omega) \cdot H(\omega) = \begin{cases} \dfrac{\pi}{\omega_c} H(\omega) & |\omega| \le \omega_c, \\ 0 & |\omega| > \omega_c. \end{cases} \tag{5}$$

We consider that the time-domain impulse response in the fact is a real signal, and $H(\omega)$ with conjugate symmetry can be expressed as

$$E_s(\omega) = \begin{cases} \dfrac{\pi}{\omega_c} \cdot \widehat{E}_s(\omega) & 0 \le \omega \le \omega_c, \\ \dfrac{\pi}{\omega_c} \cdot \widehat{E}_s^*(-\omega) & -\omega_c < \omega \le 0. \end{cases} \tag{6}$$

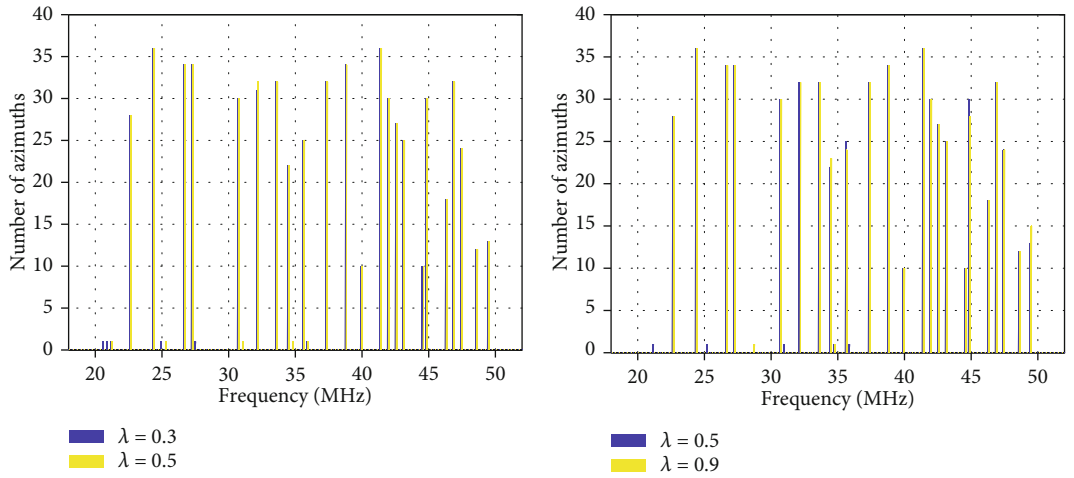
Thus, the time-domain scattering echo signal is

$$e_s(t) = e_i(t) * h(t) = \frac{2\pi}{\omega_c} \cdot \mathrm{Re}\{\widehat{e}_s(t)\}, \tag{7}$$

$$\widehat{e}_s(t) = \frac{1}{2\pi} \int_0^{\omega_c} \widehat{E}_s(\omega) e^{jwt} d\omega. \tag{8}$$

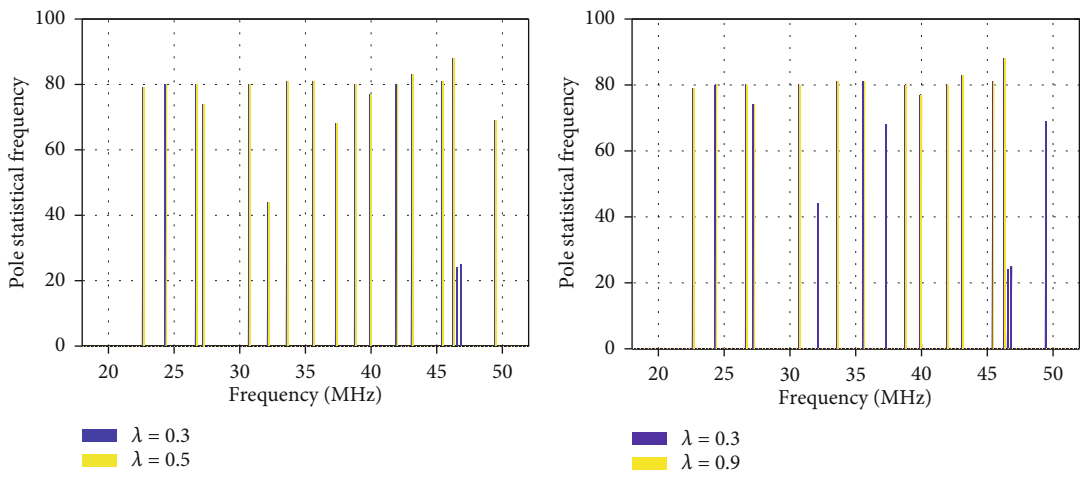
The above derivation shows that time-domain resonant response is obtained directly through the inverse Fourier transform of frequency-domain truncated data. That is to say, when the incident signal of the target is a Sinc pulse, the reflected echo is time-domain resonant response. In order to suppress time-domain data energy leakage caused by data truncation in the frequency domain, we smooth the time-domain resonant response by hamming window.

2.3. Early-Time and Late-Time Stage. The interaction between incident signal and target is shown in Figure 1.



(a) $\lambda = 0.3, 0.5$ (b) $\lambda = 0.5, 0.9$

Figure 3: Number of azimuths for pole frequency with different sliding window probability thresholds.



(a) $\lambda = 0.3, 0.5$ (b) $\lambda = 0.3, 0.9$

Figure 4: Comparison of pole statistical frequency in single azimuth for different sliding window probability thresholds.

Assuming that the signal radiates from the position P, the receiving position is also P. The incident wave reaches the target at time T_1, and the early-time stage begins at this time. The late-time stage start time is often defined as

$$t_1 = T_1 + 2T_r + T_e. \tag{9}$$

T_e is the pulse width of the incident signal. Due to the propagation delay, the response generated at position Q and the response generated at position A have a time difference of T_r. The advantage of this definition is that the late-time stage is not aliased with the early-time stage, but the late-time sequence is incomplete which causes the loss of pole information. More importantly, for complex radar target, the start time of late-time stage always changes with the incident direction and difficult to estimate for noncooperative target.

The late-time stage has already occurred at position Q; thus, we consider the late-time stage starts at

$$t_2 = T_1 + T_r + T_e. \tag{10}$$

This definition of the late-time stage can contain all information about poles, but there are also drawbacks. Obviously, one major drawback is that time delay causes an aliasing of early-time and late-time stage. The aliasing is difficult to eliminate, which may lead to pole loss and pole extraction errors. In addition, the variation of the incident signal direction may also render different start times of the late-time stage. To overcome these drawbacks, sliding window is adopted to capture the echo data used for pole extraction. Through the diversity effect of sliding window on the start of late-time stage, the aliasing data between late-time and early-time stage are preserved and explored, while interference from early-time data is reduced to some extent.

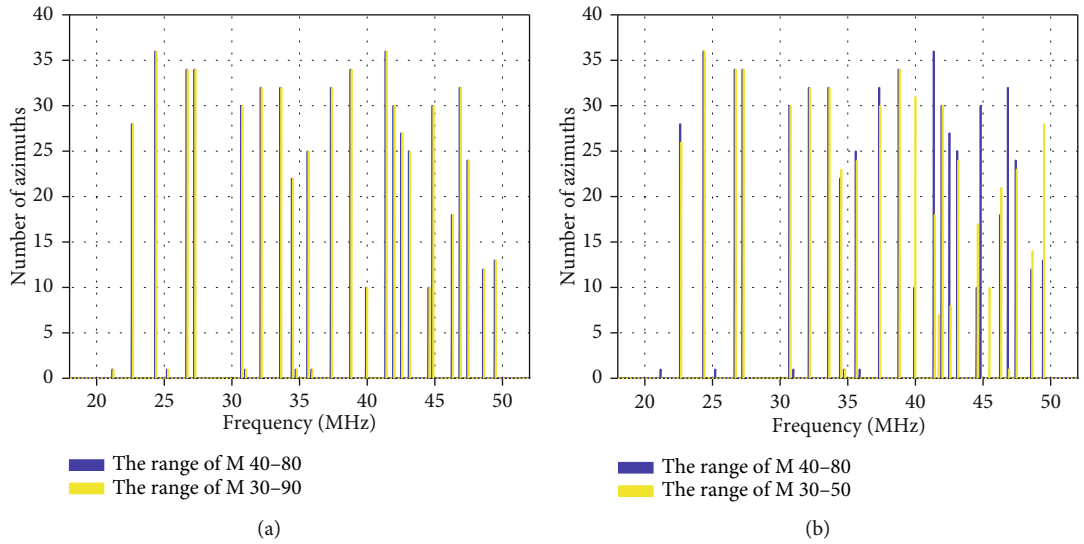


Figure 5: Number of azimuths for pole frequency with different model order ranges.

The typical stealth target studied in this paper belongs to the complex target. Combining with the above discussions, the method for obtaining the start time for pole extraction is given as follows. Firstly, we obtain the time-domain resonant response for each discrete azimuth, which is evenly distributed and spaced by 10 degrees. Secondly, t_2 is calculated based on the geometric relationship between aircraft and incident electromagnetic wave from different azimuth. Finally, the smallest t_2 is selected as the start time for pole extraction.

3. Proposed Sliding-Window Matrix Pencil Method

An overview of the proposed SW-MPM method based on sliding window and dynamic order setting is shown in Figure 2. Firstly, start time of resonant response for pole extraction is obtained based on the method proposed in the last paragraph of Section 3.2. Then, a sliding window is used to generate the signal sequence, and a set of poles is obtained based on the signal sequence in the window using MPM. The window with a fixed width slides through the signal sequence, and multiple sets of poles can be obtained upon different start sliding times. For each fixed model order, we can calculate the statistical histogram to extract a corresponding initial set of poles, whose statistical frequency exceed a certain threshold. Finally, we traverse model order and generate another set of poles on the basis of initial pole extraction for each fixed model order. The final stable poles are calculated by choosing the set of poles, whose statistical frequency exceed another certain threshold.

3.1. Initial Pole Extraction Based on Sliding Window and MPM. To overcome aliasing of early-time stage and late-time stage and improve the azimuth consistency of poles, we deploy the sliding window to capture required time-domain data for initial pole extraction. The diversity of sliding window can improve the reliability and robustness of

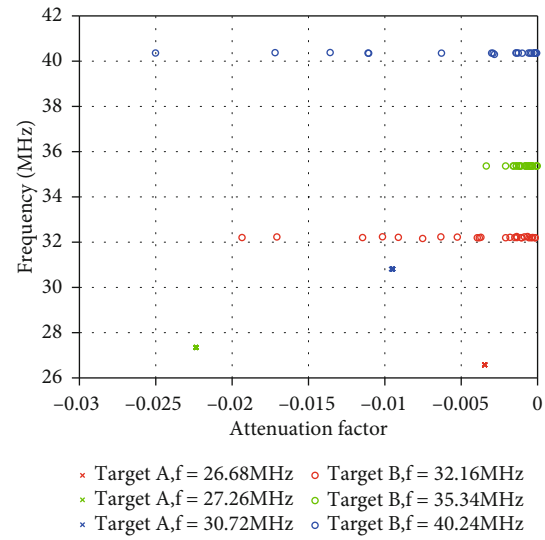


Figure 6: Pole aggregation phenomenon under different model orders.

pole extraction. Firstly, the sliding window is used to generate signal sequence and construct the related Hankel matrix. Then, we establish the matrix pencil based on the special relationship between the eigenvalues of the Hankel matrix and extract initial poles by solving the generalized eigenvalues of the matrix.

Denote the whole signal sequence as $y(t)$, and the related discrete form is

$$y(n) = \sum_{i=1}^{M} R_i z_i^n, \quad n = 1, 2, \cdots N. \tag{11}$$

A set of poles is extracted through a sliding window for the signal sequence $y(n)$ with a total length of N. Each sliding window corresponds to a Hankel matrix, and a total

Table 1: Pole frequency comparison of target A.

Pole frequency of SW-MPM (MHz)	4.76; 7.64; 14.57; 18.03; 24.38; 26.68; 27.26; 30.72; 32.16; 33.61; 37.36; 38.80; 41.39
Pole frequency of MPM (MHz)	14.57; 18.03; 24.38; 33.61; 38.80; 41.39

Table 2: Pole frequency comparison of target B.

Pole frequency of SW-MPM (MHz)	32.16; 35.34; 40.24; 52.64; 65.63; 67.36; 69.38; 71.68; 78.89
Pole frequency of MPM (MHz)	32.16; 35.34; 52.64; 67.36; 69.38; 71.68; 78.89

number N_c sliding window is applied. The window slid from the beginning of the sequence $n = 1$ to the moment $n = N_c$. Define the Hankel matric $Y^{(i)}$ with the ith sliding window:

$$Y^{(i)} = \begin{bmatrix} y_i(1) & y_i(2) & \cdots & y_i(L) \\ y_i(2) & y_i(3) & \cdots & y_i(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ y_i(N-L-1) & y_i(N-L) & \cdots & y_i(N) \end{bmatrix}. \tag{12}$$

y_i represents a sample of the used signal sequence, and L is the pencil parameter.

In order to reduce noise contained in the received echo signal, we deploy SVD to remove the small nonzero singular value and reconstruct Hankel matrix,

$$Y^{(i)} = \begin{bmatrix} U\, U' \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma' \end{bmatrix} \begin{bmatrix} V^H \\ V'H \end{bmatrix}, \tag{13}$$

where $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_M\}$ contains the corresponding larger eigenvalues of the pole information. $\sigma_1, \sigma_2, \cdots, \sigma_M$ is the maximum M singular values of $Y^{(i)}$. Σ' is a diagonal matrix containing small eigenvalues of noise information. $\tilde{Y}^{(i)} = U\Sigma V^H$ is the low rank approximation of $Y^{(i)}$. We can get $\tilde{Y}_1^{(i)}$ with the last column of $Y^{(i)}$ deleted and $\tilde{Y}_2^{(i)}$ with the first column deleted,

$$\tilde{Y}^{(i)} = \begin{bmatrix} \tilde{Y}_1{}^{(i)} & y_0 \end{bmatrix} = \begin{bmatrix} y_L & \tilde{Y}_2{}^{(i)} \end{bmatrix}, \tag{14}$$

where $\tilde{Y}_1^{(i)} = U\Sigma V_1^H$ and $\tilde{Y}_2^{(i)} = U\Sigma V_2^H$. Target pole z_i is the generalized eigenvalue of the matrix with respect to $\{\tilde{Y}_1{}^{(i)}, \tilde{Y}_2{}^{(i)}\}$. As a result, target pole z_i can be obtained by solving the eigenvalue of $\tilde{Y}_1^{(i)+}\tilde{Y}_2^{(i)}$, where $\tilde{Y}_1^{(i)+}$ is the generalized inverse of $\tilde{Y}_1^{(i)}$.

For each sliding window, we can obtain M poles according to the above description. Walk through all the sliding windows to get a set of poles containing $N_c * M$ poles. Then, histogram statistics are carried out on the pole frequency of the pole set obtained by the sliding window. The statistical histogram is divided into multiple regions according to the extracted highest pole frequency $F_{\max}$. Each region has a length of len, indicating the allowable fluctuation range of the pole frequency. If the statistical frequency of a pole falling within a certain interval N_{pole} exceeds the threshold value λ, then the median value of all the poles in the interval is considered as an initial extracted pole frequency.

3.2. Stable Pole Extraction Based on Dynamic Order Setting. The selection of pole number (model order) M has a great influence on the pole extraction accuracy. For simple ideal conductor targets, the number of poles can be calculated according to related theory. For complex radar target, such as typical stealth aircrafts, the number of poles is unknown and difficult to estimate.

One of the most widely used methods for determining M is

$$\min_M \sum_{k=1}^{N} (y_{rec}(k) - y_{cal}(k))^2. \tag{15}$$

Among them, $y_{\text{rec}}(k)$ is the time-domain signal of pole reconstruction, $y_{\text{cal}}(k)$ is the actual signal, and N is the sampling points of target echo signal used for pole extraction. Due to noise interference, it is difficult to choose an appropriate pole number for this method. Moreover, for any fixed pole number, it may render false poles caused by response aliasing between early-time and late-time stage for real applications.

In the proposed SW-MPM, instead of using a fixed M, a dynamic range of M is given to overcome the order ambiguity [32, 33]. Our research results show that the real target poles will converge with the variation of M. The real poles obtained by each simulation are densely distributed, while the false poles are scattered irregularly in the complex plane. Through the statistics of poles extracted from different M, the real poles can be determined and false poles are removed. Related steps are given as follows:

(1) Set the dynamic range of order $M_{\min}$~$M_{\max}$. The value of the minimum $M_{\min}$ is at least 10, and the value of the maximum $M_{\max}$ is generally less than $N/6$

(2) For each fixed M, a set of initially stable poles is obtained as described in Section 3.1

(3) Histogram statistics of poles for all dynamic orders are carried out. Histogram intervals are divided according to the highest frequency $F_{\max}$ extracted

at $M = M_{\max}$. Each interval has a length of len, and there are $F_{\max}$/len intervals in total. $N_p(i)$ represents the number of poles falling in the i interval

(4) If $N_p(i)$ exceeds a certain threshold value ζ, then the median of all pole frequencies in the interval $N_p(i)$ is considered as one ultimate stable pole frequency. Otherwise, all poles in this region are considered as false poles and discarded. The threshold value ζ is generally set as the smallest integer exceeds $0.8 * (M_{\max} - M_{\min} + 1)$

4. Simulation Results and Analysis

In this section, the time-domain echo data are obtained by inverse fast Fourier transform of frequency-domain RCS data, which are generated by FEKO simulation software on two typical stealth aircrafts, called A and B. We firstly discuss the effect of the sliding window probability threshold on pole extraction. Next, we study the effect of dynamic order setting on pole extraction. Then, we compare the pole extraction performance of MPM and the proposed SW-MPM. Finally, we compare the azimuth consistency of poles extracted by MPM and SW-MPM. For two typical stealth aircrafts, we test the pole extraction performance for 36 discrete azimuths, which are evenly distributed within the 360 degrees and spaced by 10 degrees.

4.1. Effect of Sliding Window Probability Threshold on Pole Extraction. Due to the variation of radar target shape and feature size, the late-time stage for different radar targets is different. Even for the same complex target, the late-time stage varies with the different incident wave directions (azimuths). The efficient acquisition of late-time stage data has great influence on pole extraction. If the start of late-time stage is too early, the early-time signal will be introduced, which will generate false poles. If the start of late-time stage is too late, the precision and antinoise performance of pole extraction will degrade due to the rapid attenuation of signal energy. Thus, sliding windows are introduced to explore adversity of different start times of late-time stage data and eliminate the negative effect of late-time stage and early-time stage aliasing problem.

Figures 3 and 4 show the effect of sliding window probability threshold on pole extraction. Figure 3 focuses on the influence of sliding window probability threshold on the azimuth of pole occurrence, while Figure 4 shows the relationship between pole statistical frequency and sliding window probability threshold when the azimuth is constant. Sliding window probability threshold λ is set to select alternate poles. That is, with the movement of the whole N_c sliding window, the poles extracted appearing more than $\lambda * N_c$ are considered as the alternative poles. As can be seen from Figure 3, when λ is selected as 0.3 and 0.5, the extraction of main poles is basically the same; when λ is selected as 0.9, the extraction of individual poles will be slightly worse. As can be seen from Figure 4, the smaller λ is, the more poles meet the conditions. Too large λ will lead to the missing of some real poles. If λ is too small, some interference and false

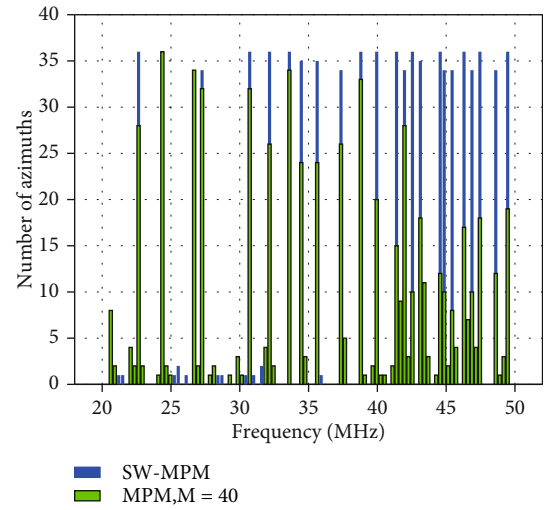


Figure 7: Comparison of pole extraction between MPM and SW-MPM for target A.

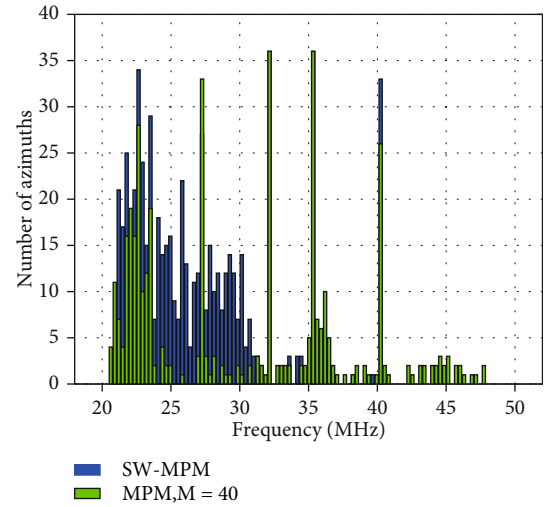


Figure 8: Comparison of pole extraction between MPM and SW-MPM for target B.

poles are extracted and the computational complexity of the system increases. Therefore, it is suitable to choose the probability threshold between 0.5 and 0.8.

4.2. Effect of Dynamic Order Setting on Pole Extraction. Figure 5 shows the effect of dynamic order setting on pole extraction. The horizontal axis represents the pole frequency, and the vertical axis represents the number of azimuths when the related poles appear. Taking coordinate (24.38 MHz, 36) for example, it means that the frequency of 24.38 MHz can be extracted in 36 evenly divided azimuths. As seen in Figure 5(a), when the range of model order is set to 30 : 90, the extraction result is similar to that of 40 : 80, but the former range renders higher computational complexity. As seen in Figure 5(b), when the range is set to 30 : 50, the pole extraction result between 40 MHz and 50 MHz is obviously inferior to 40 : 80. For example,

the frequency of 41.39 MHz can be extracted from all 36 azimuths when the range is set to 40 : 80, while 41.39 MHz can only be extracted from less than 20 azimuths in range 30 : 50. If the model order range is set too large, unnecessary interference and higher computational complexity may be introduced. If the range is set too small, the diversity characteristics of model order are not fully utilized, leading to the loss of real poles. Therefore, it can be seen that pole extraction performance is robust to the dynamic order setting range, and the range of 40 : 80 is reasonable in our simulations.

Figure 6 shows the pole aggregation phenomenon of some poles extracted, which will provide a theoretical basis for dynamic order setting. The figure shows the two-dimensional distribution of some poles, where the horizontal axis represents the attenuation factor and the vertical axis represents the pole frequency. The model order varies from 40 to 80, and each pole frequency can occur up to 41 times with the variation of model order. For pole with frequency 27.26 MHz of target A, 38 similar poles are counted in the simulations. And for pole with frequency 35.34 MHz of target B, 35 similar poles can also be counted even if the attenuation factor is slightly different. Thus, it is reasonable to use the diversity of model order to improve the performance of pole extraction because of the aggregation of poles under different model orders.

4.3. Comparison of MPM and the Proposed SW-MPM. In this section, we compare the pole extraction results of MPM and SW-MPM. Tables 1 and 2 show the pole extraction results of targets A and B. It is shown that the number of poles extracted by SW-MPM is significantly more than that by MPM. SW-MPM can extract 13 pole frequencies for target A and 9 pole frequencies of target B. In contrast, MPM can only extract 6 pole frequencies of A and 7 pole frequencies of B. Furthermore, the poles extracted by the SW-MPM method contain all poles extracted by the MPM method.

Figures 7 and 8 compare pole extraction of target A and target B, respectively, by MPM and SW-MPM. As can be seen from these figures, for common poles, poles extracted by SW-MPM appear in more azimuths, which indicates that these poles have better robustness and antinoise performance. As can be seen from Figures 7 and 8, SW-MPM performs better than MPM in pole extraction for the same target. The main reason is that MPM suffers from the difficulty of setting the model order and the influence of early-time stage and late-time stage aliasing. In contrast, the proposed SW-MPM avoids these problems by dynamic order setting and sliding window, which both can explore diversity advantages.

4.4. Analysis of Azimuth Consistency. For 36 azimuths in our simulations, we calculate the statistical frequency of poles occurring in each azimuth. If a certain pole frequency appears in more azimuths, we consider that the pole has higher reliability and better azimuth consistency, which is better for radar target recognition. We express the degree of azimuth consistency of poles by the so-called probability

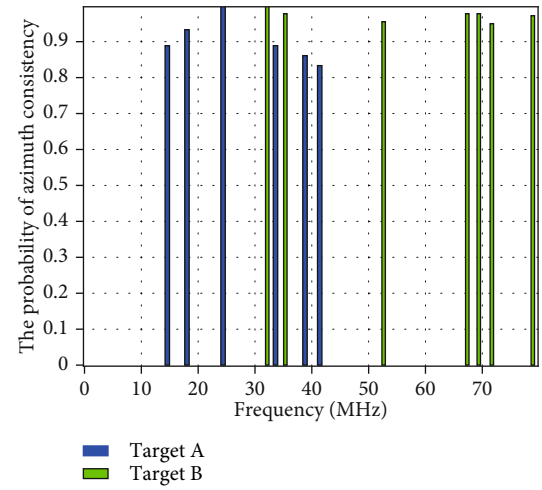


Figure 9: The probability of pole azimuth consistency distribution of MPM.

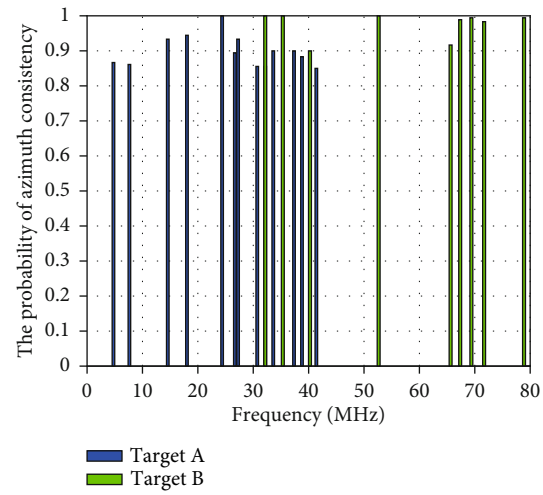


Figure 10: The probability of pole azimuth consistency distribution of SW-MPM.

of azimuth consistency. The probability is defined as the ratio of the number of azimuths that the pole appears to the total number of discrete azimuths.

Figures 9 and 10 show the azimuth consistency probability of the pole extracted in Section 4.3. It can be seen that for the same pole, the azimuth consistency extracted by SW-MPM is better than the MPM. For pole frequency 14.57 MHz of target A, the azimuth consistency probability of SW-MPM and MPM is 94.28% and 89.02%, respectively; for another pole frequency 33.61 MHz, the azimuth consistency probability of the two methods is 90% and 88.58%, respectively. For pole frequency 52.64 MHz of target B, the azimuth consistency probability of the two methods is 100% and 96.98%, respectively. For another pole frequency of 71.68 MHz, the probabilities of two methods are 97.14% and 95.88%, respectively. Therefore, the azimuth consistency of the poles extracted by the proposed SW-MPM is significantly better than that of MPM.

5. Conclusions

We proposed SW-MPM based on sliding window and dynamic order setting for pole extraction. Compared with traditional MPM method, SW-MPM can avoid the aliasing problem of time-domain resonant response to some extent and extracts more reliable poles by exploring diversity advantages of sliding window and dynamic order setting.

Our simulations on two typical stealth aircrafts verify the effectiveness and improvement of the proposed SW-MPM in pole extraction. SW-MPM can extract more poles than the MPM method for the same target. Moreover, for common poles, the azimuth consistency of poles obtained by SW-MPM is better than that of MPM. In our future work, we will test the performance of the proposed SW-MPM in more complicated battlefield electromagnetic environment, including more kinds of complex radar target and noise interference.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Grant No. 61971155 and 61801143), Natural Science Foundation of Heilongjiang Province of China (Grant No. JJ2019LH1760), the Fundamental Research Funds for the Central Universities (Grant No. 3072020CF0814), Aeronautical Science Foundation of China (Grant No. 2019010P6001), and Heilongjiang Postdoctoral Foundation (Grant No. LBH-Z2009).

References

[1] Y. X. Sun, H. Q. Xiong, D. K. P. Tan, T. X. Han, and R. Du, "Moving target localization and activity/gesture recognition for indoor radio frequency sensing applications," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24318–24326, 2021.

[2] D. B. Zhao and H. Li, "Radar target recognition based on central moment feature and GA-BP neural network," *Infrared and Laser Engineering*, vol. 47, no. 8, article 0826005, 2018.

[3] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: fine-grained indoor localization," in *2012 Proceedings IEEE INFOCOM*, pp. 2210–2218, Orlando, FL, USA, March 2012.

[4] M. Zhou, Y. X. Lin, N. Zhao, Q. Jiang, X. L. Yang, and Z. S. Tian, "Indoor WLAN intelligent target intrusion sensing using ray-aided generative adversarial network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 1, pp. 61–73, 2020.

[5] B. Borden, "Radar scattering centre localization by subspace fitting," *Inverse Problems*, vol. 17, no. 5, pp. 1483–1491, 2001.

[6] X. W. Liu, J. Z. Li, Y. Zhu, and S. J. Zhang, "Scattering characteristic extraction and recovery for multiple targets based on time frequency analysis," *Applied Computational Electromagnetics Society Journal*, vol. 35, no. 8, pp. 962–970, 2020.

[7] H. Jia, X. Li, and X. Meng, "Rigid and elastic acoustic scattering signal separation for underwater target," *Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 653–665, 2017.

[8] O. Lalakulich, E. A. Paschos, and G. Piranishvili, "Resonance production by neutrinos: The second resonance region," *Physical Review D*, vol. 74, no. 1, article 014009, 2006.

[9] D. L. Moffatt, "Transient response characteristics in identification and imaging," *IEEE Transactions on Antennas and Propagation*, vol. 29, no. 2, pp. 192–205, 1981.

[10] Y. Gong, S. Q. Xiao, and B. Z. Wang, "Synthesis of sparse planar arrays with multiple patterns by the generalized matrix enhancement and matrix pencil," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 2, pp. 869–881, 2021.

[11] A. Ponsford, L. Sevgi, and H. C. Chan, "An integrated maritime surveillance system based on high-frequency surface-wave radars. 2. Operational status and system performance," *IEEE Antennas and Propagation Magazine*, vol. 43, no. 5, pp. 52–63, 2001.

[12] M. L. Blaricum and R. Mittra, "A technique for extracting the poles and residues of a system directly from its transient response," *IEEE Transactions on Antennas and Propagation*, vol. 23, no. 6, pp. 777–781, 1975.

[13] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proceedings Of IEEE*, vol. 70, no. 9, pp. 975–989, 1982.

[14] Y. B. Hua and T. K. Sarkar, "Generalized pencil-of-function method for extracting poles of an EM system from its transient response," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 2, pp. 229–234, 1989.

[15] T. K. Sarkar and O. Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," *IEEE Antennas and Propagation Magazine*, vol. 37, no. 1, pp. 48–55, 1995.

[16] T. K. Sarkar, S. Park, and J. Koh, "Application of the matrix pencil method for estimating the SEM (singularity expansion method) poles of source-free transient responses from multiple look directions," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 4, pp. 612–618, 2000.

[17] L. Bernard, S. Goondram, B. Bahrani, A. A. Pantelous, and R. Razzaghi, "Harmonic and interharmonic phasor estimation using matrix pencil method for phasor measurement units," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 945–954, 2021.

[18] M. Bhuiyan, E. V. Malyarenko, M. A. Pantea, D. Capaldi, A. E. Baylor, and R. G. Maev, "Time-0," *Journal of Electrical and Computer Engineering*, vol. 2015, 10 pages, 2015.

[19] Y. Terriche, S. Golestan, J. M. Guerrero, D. Kerdoune, and J. C. Vasquez, "Matrix pencil method-based reference current generation for shunt active power filters," *IET Power Electronics*, vol. 11, no. 4, pp. 772–780, 2018.

[20] I. S. Choi, H. Lee, and H. T. Kim, "Natural frequency extraction using late-time evolutionary programming-based CLEAN," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 12, pp. 3285–3292, 2003.

[21] J. Chauveau, N. D. Beaucoudrey, and J. Saillard, "Determination of resonance poles of radar targets in narrow frequency

bands," in *2007 European Radar Conference*, pp. 122–125, Munich, Germany, Oct 2007.

[22] L. Man, X. Wei, C. Dong, and Z. Xiao, "Poles extracting and analyzing of complex stealth target based on matrix pencil method," in *2014 IEEE International Conference on Computer and Information Technology*, pp. 894–898, Xi'an, China, September 2014.

[23] H. Hu and K. L. Wu, "A generalized coupling matrix extraction technique for bandpass filters with uneven-Qs," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 2, pp. 244–251, 2014.

[24] O. H. Chad, L. C. Vaughan, and L. Hoi-Shun, "Late-time estimation for resonance-based radar target identification," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 11, pp. 5865–5871, 2014.

[25] C. Hargrave, A. Abbosh, V. Clarkson, and N. Shuley, "Radar target identification: estimating the start of the late time resonant response," in *2013 International Conference on Radar*, pp. 335–340, Adelaide, SA, Australia, Sept 2013.

[26] L. Carin, H. Yu, Y. Dalichaouch, and A. R. Perry, "On the wideband EMI response of a rotationally symmetric permeable and conducting target," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 6, pp. 1206–1213, 2001.

[27] A. M. Alzahed, S. Mikki, and Y. Antar, "Electromagnetic machine learning for inverse modeling using the spatial singularity expansion method," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 5, pp. 59–71, 2020.

[28] S. Licul and W. A. Davis, "Unified frequency and time-domain antenna modeling and characterization," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 9, pp. 2882–2888, 2005.

[29] S. Clarke and U. Jakobus, "Dielectric material modeling in the MoM-based code FEKO," *IEEE Antennas and Propagation Magazine*, vol. 47, no. 5, pp. 140–147, 2005.

[30] S. R. Chai, L. X. Guo, K. Li, and L. Li, "Combining CS with FEKO for fast target characteristic acquisition," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 5, pp. 2494–2504, 2018.

[31] J. P. Sijabat and T. Indriyanto, "Radar Cross section analysis of unmanned combat aerial vehicle (UCAV) using FEKO software," *AVIA*, vol. 2, no. 2, 2021.

[32] W. B. Deng, *Poles extraction and characteristic analysis of radar target in resonant region*, Harbin Institute of Technology, Harbin, China, 2012.

[33] C. Malzer and M. Baum, "Constraint-based hierarchical cluster selection in automotive radar data," *Sensors*, vol. 21, no. 10, article 3410, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7499836, 11 pages
https://doi.org/10.1155/2022/7499836

Research Article

An Elliptic Curve Signcryption Scheme and Its Application

Ping Zhang, Yamin Li, and Huanhuan Chi

School of Mathematics and Statistics, Henan University of Science and Technology, China

Correspondence should be addressed to Yamin Li; 2390043823@qq.com

Received 24 November 2021; Revised 21 February 2022; Accepted 8 April 2022; Published 6 May 2022

Academic Editor: Mu Zhou

Two basic security requirements in communication are confidentiality and authentication. Signcryption is an ideal technique to transmit encrypted and authenticated data. In view of the shortcomings of existing signcryption schemes and the high security of elliptic curve cryptography (ECC), we design a ECC-based signcryption scheme and evaluate it in terms of security, computational overhead, and communication overhead. Finally, we consider the application of our secure and efficient signcryption scheme in the smart lock key management system and analyze the bit-oriented performance of the designed key management scheme.

1. Introduction

With the rapid development of Internet, there are an increasing number of smart devices, among which the smart lock is one of the typical representatives. Compared with other smart devices, the smart lock requires higher security. When designing the smart lock, the security is the first problem to be considered.

Confidentiality and authentication are two basic security requirements in communication. In general, encryption can ensure the confidentiality of the message, and digital signature can ensure the authentication of the message. In order to meet these two requirements at the same time, the traditional method is either "Encrypt before signing" or "sign before encryption". However, these will result in a large amount of computation and communication costs. In 1997, Zheng [1] firstly proposed the notion of signcryption. Signcryption not only meets these two security requirements at the same time, but also its computational and communication costs are much lower than the traditional methods described above. Signcryption is an ideal way to transmit information encrypted and authenticated. Therefore, it also can be used for mobile device authentication. The information on which authentication is based generally includes the following three categories: (1) information known to the user, such as passwords; (2) things owned by the user, such as smart cards; and (3) biometrics of the user, such as fingerprints. Single-factor authentication generally refers to password-based authentication. Two-factor authentication refers to the smart-card-based password authentication. Multifactor authentication refers to authentication that uses two or more pieces of information. Signcryption has broad application prospects in e-commerce, e-government, and key management.

At present, the secure and practical public key cryptosystems include RSA cryptosystem (based on the big integer factorization problem), DSA cryptosystem (based on the discrete logarithm problem in the finite field), and ECC cryptosystem (based on the ECDLP). Among them, ECC cryptosystem has the highest security when the key length is the same.

The ECC cryptosystem was independently proposed by Neal Koblitz [2] and V. S. Miller [3] in 1985. It uses the elliptic curve whose variables and coefficients are elements in the finite field. The security of ECC is based on the ECDLP. Different from the discrete logarithm problem in the finite multiplication group, the ECDLP on the finite field is more difficult to solve, which cannot be solved by all known algorithms in polynomial time. In the general discrete logarithm problem, the algebraic operation on the finite field includes two operations, field addition and field multiplication, which makes the general discrete logarithm problem can be solved in subexponential time. However, in the ECDLP, the algebraic operation only includes the point addition operation

on the elliptic curve. Therefore, all the discrete logarithm algorithms cannot solve the ECDLP in subexponential time except some very special elliptic curves.

In view of the shortcomings of the existing key management scheme of the smart lock, the advantages of signcryption scheme, and the high security of ECC, this paper designs a signcryption scheme based on elliptic curve and firstly applies the signcryption scheme to the key management scheme of the smart lock system.

1.1. Related Works. Since the signcryption scheme was put forward in 1997, there have been several specific schemes based on different difficult assumptions ([1, 4–6]). In addition to the basic security objectives, some new features are introduced in the study of signcryption schemes, such as identity-based signcryption scheme ([6–11]), hybrid signcryption scheme [12], key encapsulation mechanism (KEM) and data encapsulation mechanism (DEM-)-based signcryption scheme [13], certificateless signcryption scheme [14], verifiable signcryption scheme [10], attribute-based signcryption scheme ([15, 16]), functional signcryption scheme [17], or key invisible signcryption scheme [18].

Malone-Lee [7] defined the security model of identity-based signcryption scheme in 2002 and constructed the first identity-based signcryption scheme using bilinear pairings. In 2003, Nalla et al. [19] proposed an identity-based signcryption scheme on bilinear pairings of elliptic curves. This scheme is an improvement of Lee's [7] signcryption scheme. In 2004, with the difficulty of q-Diffie-Hellman problem (q-DH) in Gap-Diffie-Hellman group, Libert et al. [20] proposed a new public key authenticated signcryption scheme. This scheme is particularly efficient. The cost of signcryption operation is almost the same as that of ElGamal encryption, and the inverse operation only needs one pairing evaluation and three power calculations. Under the assumption of q-strong Diffie-Hellman, they proved the unforgeability of this scheme. In 2009, based on the encryption scheme of water [21], Yu et al. [22] proposed the first identity-based signcryption scheme without random oracle.

In 2012, Kar [23] proposed a provably secure signcryption scheme in the random oracle model by modifying the scheme of Libert et al. [24]. This scheme is safer and more reliable than the scheme of Libert et al. In the random oracle model, they use two hypotheses, strong Diffie-Hellman (SDH) and Diffie-Hellman inversion (DHI), to prove the security of the scheme. In the same year, S. Sharmila et al. [11] firstly proposed an identity-based signcryption scheme with provable security under the standard model. The unforgeability of the scheme is based on the difficulty of computational Diffie-Hellman problem (CDH), and the indistinguishability is based on the difficulty of decisional bilinear Diffie-Hellman problem (DBDH). In 2013, Kar [25] proposed an aggregate signcryption scheme with provable security. The security of the scheme is based on the computational reliability of DBDH and discrete logarithm problem (DL). In 2014, Liu Zhenhua et al. [26] proposed a new revocable identity-based signcryption scheme to revoke malicious users in the signcryption system. In this scheme, the master key is randomly divided into two parts, one is used to construct the initial key, and the other is used to update the key. In the standard model, they proved the IND-CCA2 security based on DBDH difficult problem and the EUF-CMA security based on CDH difficult problem. In 2015, Braeken et al. [27] pointed out some problems of existing pairing-free signcryption scheme. Then, they modified the scheme and extended it to a multiuser signcryption scheme. In 2016, Kar and Naik [28] proposed an effective certificateless signcryption scheme based on bilinear mapping in the random oracle model. They proved the security of the scheme based on the assumptions of the k-CAA, Inv-CDH, q-BDHI, and CDH. In the same year, Han Yiliang et al. [29] combined Niedereiter public key cryptography with CFS signature scheme and constructed a signcryption scheme. This scheme can resist quantum attack and has a small amount of key data. They proved the IND-CCA2 security and EUF-CMA security of the scheme in the random oracle model. In 2017, Zhou Yanwei et al. [30] proposed an efficient certificateless signcryption scheme without bilinear mapping and proved the security of the scheme based on CDH and DL in the random oracle model. Tsai et al. [31] proposed a new multidocument blind signature scheme based on ECC. This scheme adds the design of the signature encryption paradigm to the blind signature scheme to enhance high-level security. In 2018, for the security of hybrid signcryption schemes, Dai et al. [32] studied the replayable CCA security (RCCA) of SKEM+DEM [33] and Tag SKEM+DEM [13]. If the scheme SKEM is RCCA secure and the scheme DEM is RCCA secure, the hybrid signature scheme SKEM+DEM is RCCA secure. If the scheme Tag-SKEM is RCCA secure and the scheme DEM is RCCA secure, the Tag SKEM + DEM hybrid encryption scheme is RCCA secure. In the single-factor authentication research area, He Debiao et al. [34] proposed a password-based remote user authentication scheme without smart cards. The scheme can resist various attacks, such as device stolen attack and privileged insider attack. In the two-factor authentication research area, Wang Ding et al. [35] proposed a smart-card-based password authentication scheme that kills two birds with one stone. By integrating "honeywords" with their proposed "fuzzy-verifiers," the scheme not only not only eliminates the long-standing security-usability conflict that is considered intractable in the literature, but also achieves security guarantees beyond the conventional optimal security bound. Our signcryption scheme has highly efficient and satisfies multiple security properties; we believe it can be used as a building block for the authentication phase of a single-factor authentication scheme. When the server and user authenticate each other and generate a session key, they can use our scheme to signcrypt their own messages, respectively, which not only achieves authentication but also provides additional confidentiality.

At present, there have been many works on the key management system of smart locks. For data security in narrow band Internet of things (NB-IOT) application environment, Jia Rongyuan et al. [36] proposed a lightweight encryption algorithm and encryption model based on AES [37] and chaos sequence. However, they did not explain how to transmit the key. There are problems such as difficult monitoring,

high power consumption requirements, and insecure wireless transmission of wireless smart lock. In order to solve the problems, Zhang Huanlan et al. [38] proposed a 433 MHz wireless module based on Diffie-Hellman key exchange algorithm and corrected block tiny encryption algorithm for double encryption smart lock system. In 2019, under the unreliable UDP data transmission of NB-IoT, Liu Mengjun [39] designed a key transmission interaction scheme to complete the reliable update of the user's unlock key with as little calculation and communication as possible. However, this scheme will continue to use the old key for unlocking when the unlock key update fails, which is not applicable to public rental housing. Because if the user loses the qualification to rent a house, the unlock key must be updated as soon as possible. In addition, in this work, the session key used between the server and the smart lock has low security. Sha Tao et al. [40] designed an identity verification mechanism based on position proof. They also proposed a timestamp encryption mechanism to prevent remote unlocking and replay attacks by malicious users. However, this work did not explain how the server issued the unlock key to the smart lock, and the smart lock did not upload operating information to the server. Wang et al. [41] designed a complementary multidimensional feature fusion network-based hand gesture recognition (CMFF-HGR) to extract features and achieve hand gesture recognition. The smart lock key management system based on hand gesture recognition is different from the key management scheme proposed in this paper. The smart lock system based on hand gesture recognition requires to memorize the gestures manually, and the hand gesture is easy to be known by others during the unlocking process. However, the key management scheme in this paper does not require manually memorizing the unlock key, and every time the unlock key is different, not being fixed. Therefore, the key management scheme in this paper has higher security.

1.2. Contribution. This paper proposed an efficient and secure ECC-based signcryption scheme and applied it to a smart lock key management system. To the best of our knowledge, this work is the first to consider the application of a signcryption scheme in a smart lock key management system. Compared with other smart lock key management schemes, our scheme is more efficient and secure due to the confidentiality and authentication by the signcryption itself, as well as the efficiency and other security properties of our signcryption scheme. In addition, in our key management scheme, the unlock key is delivered to the smartphone by the server, and then, the smartphone unlocks the smart lock through Bluetooth. Therefore, the unlock key is different every time, and the user does not need to memorize a fixed unlock key, which makes our key management scheme more secure and convenient.

1.3. Organization. This paper is organized as follows. The first section is the introduction of this paper. The second section introduces the basic knowledge, including elliptic curve discrete logarithm problem, the formal definition, and the security model of signcryption scheme. In the third section, we design a signcryption scheme based on elliptic curve and analyze the correctness, security, and performance of our signcryption scheme. In the fourth section, we apply our signcryption scheme to the key management system of smart lock. Finally, in the fifth section, we summarize the full text and give an outlook for future work.

2. Preliminaries

2.1. Basic Notation. In the following sections, if $|\mathrm{negl}(\lambda)| < 1/\mathrm{poly}(\lambda)$ for all polynomials $\mathrm{poly}(\lambda)$ and all sufficiently large λ, we call $\mathrm{negl}(\lambda)$ is negligible. In this paper, "PPT" represents probabilistic polynomial time.

2.2. Elliptic Curve Discrete Logarithm Problem

Definition 1 (Elliptic curve discrete logarithm problem). Given an elliptic curve $E(GF(q))$, P is a point on this elliptic curve and its order is a large prime number n $(ord(P) = n)$. For any random number d, $Q = dP$ can be easily calculated. However, if P and Q are known, it is very difficult to find d.

2.3. The Definition of Signcryption Scheme

2.3.1. Syntax. Given the key space $\mathcal{K}$, message space $\mathcal{M}$, and signcryption space $\mathcal{S}$, for any sender and receiver, a signcryption scheme SC = (setup, keygen, signcrypt, unsigncrypt) is a collection of the following four algorithms.

(i) Setup $(1^\lambda) \longrightarrow cp$: This is system initialization algorithm. This algorithm requires a security parameter λ as the input of the algorithm and requires common parameters cp as the output of the algorithm

(ii) Keygen $(cp, r) \longrightarrow (PK, SK)$: This is key generation algorithm, which is a random algorithm. This algorithm requires common parameters cp and random number r as the input of the algorithm and requires key pair (PK, SK) $(PK, SK \in \mathcal{K})$ as the output of the algorithm

(iii) Signcrypt $(cp, SK_S, PK_R, m) \longrightarrow \sigma$: This is signcryption algorithm. This algorithm requires common parameters cp, private key SK_S $(SK_S \in \mathcal{K})$ of sender, public key PK_R $(PK_R \in \mathcal{K})$ of receiver, and message m $(m \in \mathcal{M})$ as the input of the algorithm and requires signcryption $\sigma(\sigma \in \mathcal{S})$ as the output of the algorithm

(iv) Unsigncrypt $(cp, SK_R, PK_S, \sigma) \longrightarrow m$: This is unsigncryption algorithm. This algorithm requires common parameters cp, private key SK_R $(SK_R \in \mathcal{K}$) of receiver, and public key PK_S $(PK_S \in \mathcal{K})$ of sender and signcryption $\sigma(\sigma \in \mathcal{S})$ as the input of the algorithm. This algorithm outputs message m $(m \in \mathcal{M})$ or symbol "$\perp$" ("$\perp$" indicates that the unsigncryption failed)

Definition 2 (Correctness). For any message $m \in \mathcal{M}$, any sender (his key pair (SK_S, PK_S) was generated by $Keygen(c$

p, r)), any receiver (his key pair (SK_R, PK_R) was generated by $Keygen(cp, r)$), and the following formula holds

$$\text{Unsigncrypt}(cp, SK_R, PK_S, \text{Signcrypt}(cp, SK_S, PK_R, m)) = m. \quad (1)$$

2.4. The Security Model of Signcryption Scheme

Definition 3 (Confidentiality). The confidentiality security can be seen as a game between the adversary $\mathcal{O}$ and the challenger $\mathcal{C}$. This game is divided into five phases.

(i) Keygen phase: Challenger $\mathcal{C}$ runs algorithm Keygen(cp, r) to generate a sender key pair (SK_S, PK_S) and a receiver key pair (SK_R, PK_R), and sends (PK_S, PK_R) to adversary $\mathcal{O}$

(ii) Query phase 1: The adversary $\mathcal{O}$ sends multiple signcryption queries and unsigncryption queries to the challenger $\mathcal{C}$

(1) Signcryption query: The adversary $\mathcal{O}$ submits the message m and the public key (PK_S, PK_R) to the challenger $\mathcal{C}$. The challenger $\mathcal{C}$ calculates $\sigma = \text{signcrypt}(cp, SK_S, PK_R, m)$ and sends the result σ to the adversary $\mathcal{O}$

(2) Unsigncryption query: The adversary $\mathcal{O}$ submits the legitimate signcryption result σ and the public key (PK_S, PK_R) to the challenger $\mathcal{C}$. The challenger $\mathcal{C}$ calculates unsigncrypt (cp, SK_R, PK_S, σ) and sends the message m or symbol "$\perp$" to the adversary $\mathcal{O}$

(iii) Challenge phase: The adversary $\mathcal{O}$ submits two messages m_0, m_1 (m_0, m_1 have the same length) to the challenger $\mathcal{C}$. The challenger $\mathcal{C}$ randomly selects $i \in \{0, 1\}$, calculates $\sigma^* = \text{signcrypt}(cp, SK_S, PK_R, m_i)$ and sends the result σ^* to the adversary $\mathcal{O}$

(iv) Query phase 2: Similar to the query phase 1, the adversary $\mathcal{O}$ continues to send multiple signcryption queries and unsigncryption queries to the challenger $\mathcal{C}$ (the adversary $\mathcal{O}$ is forbidden from sending unsigncryption query for the result σ^*)

(v) Guess phase: The adversary $\mathcal{O}$ outputs a value i' as the guess for i. If $i' = i$, the adversary $\mathcal{O}$ wins this game

In this game, the advantage of the adversary $\mathcal{O}$ is $Adv(\mathcal{O}) = |\Pr[i' = i] - 1/2|$.

Definition 4 (Unforgeability). The unforgeability security can be seen as a game between the adversary $\mathcal{O}$ and the challenger $\mathcal{C}$. This game is divided into three phases.

(i) Keygen phase: Challenger $\mathcal{C}$ runs algorithm keygen (cp, r) to generate a sender key pair (SK_S, PK_S) and a receiver key pair (SK_R, PK_R) and sends (PK_S, PK_R) to adversary $\mathcal{O}$

(ii) Query phase: The adversary $\mathcal{O}$ sends multiple signcryption queries and unsigncryption queries to the challenger $\mathcal{C}$

(1) Signcryption query: The adversary $\mathcal{O}$ submits the message m and the public key (PK_S, PK_R) to the challenger $\mathcal{C}$. The challenger $\mathcal{C}$ calculates $\sigma = \text{signcrypt}(cp, SK_S, PK_R, m)$ and sends the result σ to the adversary $\mathcal{O}$

(2) Unsigncryption query: The adversary $\mathcal{O}$ submits the legitimate signcryption result σ and the public key (PK_S, PK_R) to the challenger $\mathcal{C}$. The challenger $\mathcal{C}$ calculates unsigncrypt (cp, SK_R, PK_S, σ) and sends the message m or symbol "$\perp$" to the adversary $\mathcal{O}$

(iii) Forgery phase: The adversary $\mathcal{O}$ submits the challenging content, including challenging message m^* and the forged signcryption σ^*. The challenger $\mathcal{C}$ submits the above input to the oracle, and the oracle returns the unsigncryption of signcryption σ^* to the challenger $\mathcal{C}$. If the result is message m^*, and the adversary $\mathcal{O}$ has not used this message as the input for signcryption query before, the adversary $\mathcal{O}$ wins this game

In this game, the advantage of the adversary is his probability of winning the game.

3. Our ECC-Based Signcryption Scheme

3.1. Construction. In this section, we define and construct our elliptic curve signcryption scheme SC = (setup, keygen, signcrypt, unsigncrypt).

(i) Setup: Let $GF(q)$ be a finite field of order q (the length of q is l), $E: y^2 = x^3 + ax + b (\text{mod } q)$ $(a, b \in GF(q), 4a^3 + 27b^2 \neq 0)$ be an elliptic curve in finite field $GF(q)$, P be the base point of the elliptic curve E. $ord(P) = n$, where n is a large prime number. Let $h = \#E(GF(q))/n$ $(h \ll n)$ is the cofactor. $\#E(GF(q))$ represents the number of points of the elliptic curve E defined on the finite field $GF(q)$). G_1 is an elliptic curve cyclic multiplication group of order q generated by point P. We suppose the plaintext space is $\{0, 1\}^l$ and select two hash functions $H_1: G_1 \longrightarrow \{0, 1\}^l$, $H_2: \{0, 1\}^* \longrightarrow Z_q$. Then, we expose parameters $D = \{q, l, a, b, P, G_1, n, h\}$ and hash function H_1, H_2

(ii) Keygen: The sender randomly selects SK_S as his private key, and his public key is $PK_S = SK_S P$. The receiver randomly selects SK_R as his private key, and his public key is $PK_R = SK_R P$. Then, they keep the private key SK_S, SK_R secret and expose the public key PK_S, PK_R

(iii) Signcrypt: The sender uses PK_R and SK_S to signcrypt message m

(a) Select a random number $k \in [1, n-1]$

(b) Compute $kPK_R = K$

(c) Compute $b = H_1(K)$.

(d) Compute $c = b \oplus m$

(e) Compute $e = H_2(m, K, PK_S, PK_R)$.

(f) Compute $s = k^{-1}(e + SK_S)$. If $s = 0$, return to step 1

(g) Get the signcryption $\sigma = (c, e, s)$, and send it to the recriver

(iv) Unsigncrypt: The receiver gets the signcryption $\sigma = (c, e, s)$ and uses PK_S and SK_R to unsigncrypt it

(a) Compute $w = s^{-1}$

(b) Compute $X = ewPK_R + wPK_SSK_R$

(c) Compute $b' = H_1(X)$.

(d) Compute $m = b' \oplus c$

(e) Compute $e' = H_2(m, X, PK_S, PK_R)$.

(f) If $e' = e$, return m, otherwise return "$\perp$".

3.2. Correctness. Because $s = k^{-1}(e + SK_S)$, we have $s^{-1} = k(e + SK_S)^{-1}$. Therefore, the following formula holds

$$\begin{aligned} X &= ewPK_R + wPK_SSK_R = es^{-1}PK_R + s^{-1}PK_SSK_R \\ &= es^{-1}SK_RP + s^{-1}SK_SSK_RP = (e + SK_S)s^{-1}SK_RP \\ &= (e + SK_S)k(e + SK_S)^{-1}SK_RP = kSK_RP = kPK_R = K. \end{aligned} \tag{2}$$

So, we have $b' = b$, $e' = e$. Here, $b' = b$ ensures that the receiver can restore the sender's message m; that is, the decryption process is correct. $e' = e$ ensures that the receiver can verify the correctness of the sender's signature; that is, the verification process is correct. Therefore, our signcryption scheme is correct.

3.3. Security

3.3.1. Confidentiality. Confidentiality means that information can only be used by authorized users and cannot be disclosed to unauthorized users. Confidentiality is a required property of encryption. Since signcryption needs to realize both signature and encryption, the signcryption scheme must also have confidentiality. According to Theorem 5, our signcryption scheme has confidentiality.

Theorem 5. *In the random oracle model, if there is an adversary $\mathcal{O}$ who can win the game of Definition 3 with the advantage of ε, there is a challenger $\mathcal{C}$ who can solve the ECDLP problem with the advantage of at least $\varepsilon' \geq \varepsilon/(q_{H_2} + q_{Sig} + q_{Uns})$. q_{H_2}, q_{Sig}, and q_{Uns} represent the number of times the adversary initiates H_2 query, signcryption query, and unsigncryption query, respectively.*

Proof. At the beginning of the game, the challenger $\mathcal{C}$ runs algorithm keygen (cp, r) to generate a sender key pair (SK_S, PK_S) and a receiver key pair (SK_R, PK_R) and sends (PK_S, PK_R) to adversary $\mathcal{O}$. The challenger $\mathcal{C}$ manages four lists $L_{H_1}, L_{H_2}, L_{Sig}, L_{Uns}$, which are initially empty. L_{H_1}, L_{H_2} are used to track the adversary's queries to oracle H_1, H_2, respectively, L_{Sig} is used to simulate signcryption oracle, and L_{Uns} is used to simulate unsigncryption oracle. □

Next, the adversary $\mathcal{O}$ sends queries to the challenger $\mathcal{C}$.

(1) (H_1 query) If (K, b) already exists in the list L_{H_1}, the challenger returns b. Otherwise, the challenger selects b from $\{0,1\}^l$ randomly, stores b in list L_{H_1}, and returns b

(2) (H_2 query) If (m, K, PK_S, PK_R, e) already exists in the list L_{H_2}, the challenger returns e. Otherwise, the challenger selects e from Z_q randomly, stores (m, K, PK_S, PK_R, e) in list L_{H_2}, and returns e

(3) (Signcrypt query) The public key of sender is PK_S, the public key of receiver is PK_R, and the message is m. The challenger selects k from $[1, n-1]$ randomly and computes $K = kPK_R$, $b = H_1(K)$. $H_1(K)$ can be obtained from the above H_1 query. Then, the challenger computes $c = b \oplus m$, $e = H_2(m, K, PK_S, PK_R)$. $e = H_2(m, K, PK_S, PK_R)$ can be obtained from the above H_2 query. The challenger computes $s = k^{-1}(e + SK_S)$ and returns (c, e, s)

(4) (Unsigncrypt query) The public key of sender is PK_S, the public key of receiver is PK_R, and the signcryption is $\sigma = (c, e, s)$. The challenger computes $w = s^{-1}$, $X = ewPK_R + wPK_SSK_R$. If $X \notin L_{H_1}$, the challenger returns "$\perp$", else computes $b' = H_1(X)$, $m = b' \oplus c$. If $(m, X, PK_S, PK_R) \notin L_{H_2}$, the challenger returns "$\perp$", else computes $e' = H_2(m, X, PK_S, PK_R)$. If $e' \neq e$, the challenger returns "$\perp$", else computes m

After the above-mentioned queries are initiated polynomial times, the game enters the challenge phase. The adversary $\mathcal{O}$ outputs two messages $\{m_0, m_1\}$. The challenger $\mathcal{C}$ randomly selects i from $\{0,1\}$, b^* from $\{0,1\}^l$, and e^* and s^* from Z_q and computes $c^* = b^* \oplus m_i$ and $w^* = (s^*)^{-1}$. When H_1 is queried at $K^* = (e^*w^* + w^*SK_S)PK_R$, the value b^* is returned directly. When H_2 is queried at $(m_i, K^* = (e^*w^* + w^*SK_S)PK_R, PK_S, PK_R)$, the value e^* is returned directly. The challenger $\mathcal{C}$ returns challenging signcryption $\sigma^* = (c^*, e^*, s^*)$ to $\mathcal{O}$. The adversary $\mathcal{O}$ initiates the second round of query, which is same as the first round of query, but the adversary $\mathcal{O}$ cannot send unsigncryption query for the signcryption result σ^*. At the end of the simulation, the adversary $\mathcal{O}$ outputs i' as the guess for i. If $i' = i$, the

challenger $\mathscr{C}$ outputs $k = ew + wSK_S$ as an answer to the ECDLP, else the challenger $\mathscr{C}$ fails to solve the ECDLP.

In the view of the adversary $\mathcal{O}$, the challenger $\mathscr{C}$ provides a simulation environment similar to the actual environment. However, in the challenge phase, the answer of H_2 to the query $(m_i, K^* = (e^*w^* + w^*SK_S)PK_R, PK_S, PK_R)$ is different. This is because m_i can only be determined at the end of the challenge phase. At this point, $q_{H_2} + q_{Sig} + q_{Uns}$ is the maximum number that H_2 is queried. Therefore, the challenger $\mathscr{C}$ has an advantage of at least $\varepsilon' \geq \varepsilon/(q_{H_2} + q_{Sig} + q_{Uns})$ to solve the ECDLP problem.

3.3.2. Unforgeability. Unforgeability is a required property of signature. Since signcryption needs to realize both signature and encryption, the signcryption scheme must also have unforgeability. According to Theorem 6, our signcryption scheme has unforgeability.

Theorem 6. *In the random oracle model, if there is an adversary $\mathcal{O}$ who can win the game of Definition 4 with the advantage of ε, there is a challenger $\mathscr{C}$ who can solve the ECDLP problem with the advantage of $\varepsilon/(q_{H_1} + q_{H_2} + q_{Sig} + q_{Uns})$. q_{H_1}, q_{H_2}, and q_{Sig} represent the number of times the adversary initiates H_1 query, H_2 query, and signcryption query, respectively.*

Proof. At the beginning of the game, the challenger $\mathscr{C}$ runs algorithm Keygen(cp, r) to generate a sender key pair (SK_S, PK_S) and a receiver key pair (SK_R, PK_R) and sends (PK_S, PK_R) to adversary $\mathcal{O}$. The challenger $\mathscr{C}$ manages three lists $L_{H_1}, L_{H_2}, L_{Sig}$, which are initially empty. L_{H_1}, L_{H_2} are used to track the adversary's queries to oracle H_1, H_2, respectively, L_{Sig} is used to simulate signcryption oracle. □

Suppose the public key of the receiver is PK_R, the adversary uses the oracle described in the proof of Theorem 5 to send various queries. After these queries, in the forgery phase, the adversary outputs the forged signcryption result. It can be seen from the proof of Theorem 5 that our simulation is equivalent to the actual attack environment. In order to forge successfully, the adversary must send H_1 query and H_2 query to get $\sigma^* = (c^*, e^*, s^*)$ corresponding to message m^*. The probability that the adversary chooses the correct record in the list L_{H_1}, L_{H_2} is $1/(q_{H_1} + q_{H_2} + q_{Sig} + q_{Uns})$, so the challenger $\mathscr{C}$ has the advantage of $\varepsilon/(q_{H_1} + q_{H_2} + q_{Sig} + q_{Uns})$ to solve ECDLP problem.

3.3.3. Integrity. Integrity means that information cannot be accidentally or maliciously deleted, modified, forged, replayed, and inserted during transmission and storage.

Theorem 7. *Our signcryption scheme has integrity.*

Proof. In our signcryption scheme, it is very difficult for an attacker to tamper with the information between the sender and receiver. Because this tampering requires the hash value b, and b corresponds to the hash value of a random point of the elliptic curve, due to the collision resistance of the hash function, the attacker cannot determine the point of the elliptic curve corresponding to the hash value b. Furthermore, every part of ciphertext $c = b \oplus m$ depends on all message blocks. Once a malicious attacker makes any change to a particular block of information, it will cause the ciphertext to change. Therefore, our signcryption scheme has integrity. □

3.3.4. Nonrepudiation. Nonrepudiation in signcryption and signature is the same. Nonrepudiation is preventing a communicating party from denying a previous promise or behavior. In a signcryption scheme, nonrepudiation means that a signer cannot deny that he signed a valid message after signing it.

Theorem 8. *Our signcryption scheme has nonrepudiation.*

Proof. In our signcryption scheme, when the sender signs message m, it first calculates the hash value of message m using its own public key PK_S and receiver's public key PK_R and then signs this hash value with his own private key SK_S. Therefore, the sender cannot deny its signature to message m. In addition, in unsigncryption, the receiver will use the sender's public key PK_S and its own public key PK_R to calculate the hash value. If it is equal to the received hash value, it means that the received signature is indeed signed by the sender. Therefore, our scheme has nonrepudiation. □

3.3.5. Availability. Availability refers to the property that all resources can be accessed by authorized parties at the appropriate time; i.e., information can be accessed by authorized entities and used on demand.

Theorem 9. *Our signcryption scheme has availability.*

Proof. In our signcryption scheme, the recipient, as an authorized entity, can use its own private key to obtain the plaintext m signed by the sender through the unsigncryption after obtaining the signcryption and then use the plaintext m to perform other required operations. Therefore, our signcryption scheme has availability. □

3.3.6. Forward Secrecy. Forward secrecy means that exposure of private key of the encryptor does not affect the confidentiality of previously encrypted messages.

Theorem 10. *Our signcryption scheme has forward secrecy.*

Proof. In our signcryption scheme, if the sender's private key is leaked, the adversary must know the value of b in order to obtain the previous session content, so he must obtain the value k. However, k is randomly selected by the sender. Even if the adversary obtains the sender's private key, he still cannot recover the plaintext information. Therefore, our signcryption scheme has forward secrecy. □

3.3.7. Internal Security. The security model of signcryption can be divided into external security and internal security. External security means that the adversary only knows

public information. Internal security means that the adversary knows the sender's or receiver's private key in addition to the public information. That is, if the sender's private key is exposed, the adversary still cannot recover the plaintext from the ciphertext; if the receiver's private key is exposed, the adversary still cannot forge the ciphertext. Obviously, internal security is stronger than external security.

Theorem 11. *Our signcryption scheme has internal security.*

Proof. On the one hand, in our signcryption scheme, if the adversary wants to recover the plaintext m from the ciphertext c, it must obtain the hash value b. Similar to Theorem 7, due to the collision resistance of the hash function and the randomness of the random number k, the adversary cannot determine the point on the elliptic curve corresponding to the hash value b. Therefore, even if the adversary possesses the sender's private key, the plaintext still cannot be recovered from the ciphertext. On the other hand, in our signcryption scheme, if the adversary possesses the receiver's private key, it is also impossible to forge the valid ciphertext c' of the plaintext m'. The reason is that even if the adversary uses SK_R to compute the value of X, gets the hash value b', and then uses $c = b \oplus m$ to get the ciphertext c' of the plaintext m', the ciphertext c' is invalid. Because the ciphertext c in the signcryption result is the encryption of the plaintext m', and the s in the signcryption result is the signature of the plaintext m, which will make the unsigncryption fail, therefore, our signcryption scheme has internal security. □

We compare the security of our signcryption scheme with Tsai's ECC-based signcryption scheme [31] and Zhou's signcryption scheme [30]. It can be seen from Table 1 that our scheme satisfies the confidentiality, unforgeability, integrity, nonrepudiation, and availability of the other two schemes and also satisfies forward secrecy and internal security. Therefore, compared with the existing signcryption schemes, our signcryption scheme is more secure.

3.4. Performance Evaluation. In this section, we compare the computational and communication overhead of our signcryption scheme with Tsai's scheme [31] and Zhou's scheme [30] in detail. Among them, the computational overhead mainly compares the calculation amount of the signcryption and unsigncryption algorithms, and the calculation amount mainly counts the execution times of the point multiplication operation, point addition operation, number multiplication operation, and inversion operation. The XOR operation, Hash operation, and the number addition operation are not counted. The computational overhead and communication overhead of the three schemes are shown in Table 2. In this table, PM, PA, NM, and IN represent the point multiplication operation, point addition operation, number multiplication operation, and inversion operation, respectively; l_m represents the length of the plaintext message; $|G|$ represents the length of the element on the group; and $|Z_n^*|$ represents the length of the element in Z_n^*.

Among the various operations counted in Table 2, the point multiplication operation takes the most time, followed by the point addition operation. It can be seen from the calculation amount in Table 2 that the computational overhead of our scheme is much less than that of the other two schemes. In addition, the communication overhead of our scheme is comparable to Zhou's scheme and smaller than Tsai's scheme. Our scheme has forward security and internal security in addition to the same confidentiality, unforgeability, integrity, nonrepudiation, and availability as the other two schemes. Overall, our scheme is an efficient and secure ECC-based signcryption scheme.

4. Our Key Management Scheme

In this section, our ECC-based signcryption scheme will be applied to the key management scheme in the smart lock system. In Subsection 4.1, we recall the model of the smart lock system. In Subsection 4.2, we give an overview of the key management scheme for the smart lock system. In Subsection 4.3, we use the above ECC-based signcryption scheme as a building block to construct our key management scheme of the smart lock system. Finally, in Subsection 4.4, we observe the bit-oriented overhead of our smart lock key management scheme through experimental simulations.

4.1. The Model of the Smart Lock System. There are three main parties in a smart lock system [40], that is, smart phone (SP) of user, smart lock (SL), and management server (MS), which is shown in Figure 1. Among them, MS receives the request from the user's SP, reviews the user's qualification, receives the operation information of SL, manages the unlock key, and helps SL and SP exchange the public key. SL communicates with MS through narrow band Internet of Things (NB-IOT) and receives the signcryption for the unlock key. SP of the legitimate user applies for the unlock key to MS and sends the signcryption of this unlock key to SL through Bluetooth.

In the smart lock system, MS is trusted, which cannot disclose the unlock key to the adversary. MS will send correct unlock key to SL and legitimate user and cancel the unlock key of the expired user. SL is safe, controllable, and will not disclose the unlock key. The user is semi-honest. Although he will follow the rule of the key management scheme, he will try to use the obtained information to unlock when his key expires or he has no key.

Each smart lock has a unique international mobile equipment identity (IMEI), which is a 15-digit "electronic serial number." In this paper, the IMEI will be used to generate a session key for the smart lock.

4.2. The Overview of Key Management Scheme. In the key management scheme of the smart lock system, MS and SL generate their own private keys, respectively, and then calculate their own public keys through ECC and send the public key to each other. They realize the key exchange and generate the shared session key between them. MS generates the unlock key, uses the session key to encrypt the unlock key, and sends the encryption result to SL. SL uses the session

Table 1: Security comparison of three signcryption schemes.

	Confidentiality	Unforgeability	Integrity	Nonrepudiation	Availability	Forward secrecy	Internal security
Tsai's scheme	Y	Y	Y	Y	Y	N	N
Zhou's scheme	Y	Y	Y	Y	Y	N	N
Our scheme	Y	Y	Y	Y	Y	Y	Y

Note: "Y" means that the scheme has this property; "N" means that the scheme does not have this property.

Table 2: Performance comparison of three signcryption schemes.

	Computational overhead		Communication overhead
	Signcryption	Unsigncryption	
Tsai's scheme	$4PM + 2NM$	$3PM + 3PA$	$l_m + 3\|G\| + \|Z_n^*\|$
Zhou's scheme	$3PM + 2PA + 2NM + 1\text{IN}$	$6PM + 5PA$	$l_m + 2\|Z_n^*\|$
Our scheme	$1PM + 1NM + 1\text{IN}$	$2PM + 1PA + 1\text{IN} + 2NM$	$l_m + 2\|Z_n^*\|$

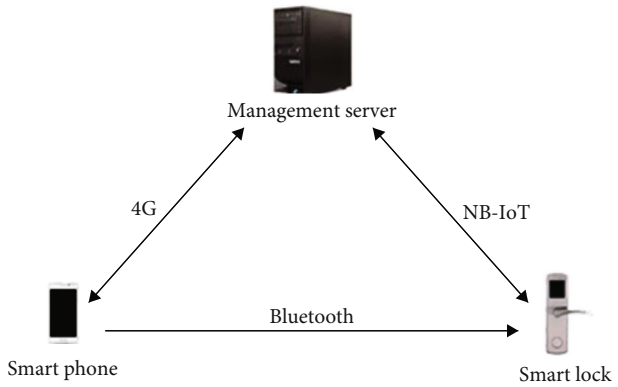


Figure 1: Architecture of smart lock system.

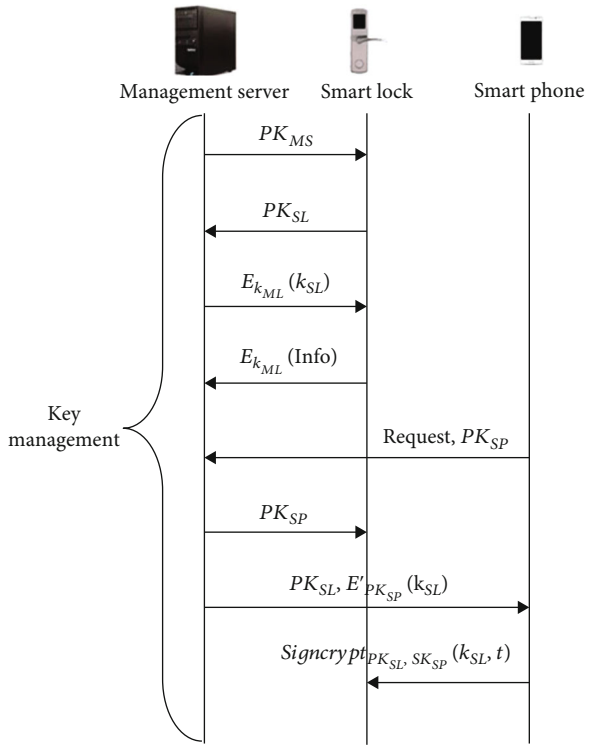


Figure 2: Flow chart for key management of smart lock system.

key to encrypt its operation information and uploads the encryption result to MS. This two communications adopt AES symmetric encryption through the Nb-IOT. After the user applies to MS for the house, MS reviews the user's qualification. If the user does not meet the conditions, MS refuses to send the key to him. If the user meets the conditions, MS sends the user's public key to SL. At the same time, MS uses the user's public key to encrypt the unlock key and sends the encryption result and the public key of SL to the user. The user uses his private key to decrypt the unlock key. This communication uses elliptic curve public key cryptosystem. After that, the user can use the received public key of SL and his own private key to signcrypt the unlock key and send the signcryption result to SL. SL uses the received public key of user and his own private key to de-signcrypt the signcryption, thus obtaining the unlock key. The process above uses our ECC-based signcryption scheme. Finally, SL compares the unlock key from the user with its own unlock key. If the two unlock keys are different, SL cannot be unlocked.

During the lifetime, the smart lock system can periodically update the key according to the security status. If the user loses the housing qualification, MS regenerates the unlock key and sends it to SL. The user cannot unlock with the old key.

4.3. Our Key Management Scheme. Our key management scheme is detailed as follows:

(1) Key exchange between MS and SL. MS selects private key SK_{MS}. SK_{MS} is confidential and satisfying $SK_{MS} < n$. MS computes public key $PK_{MS} = SK_{MS} \times P$ and sends PK_{MS} to SL through NB-IOT. In the transmission, even if PK_{MS} is attacked, the adversary

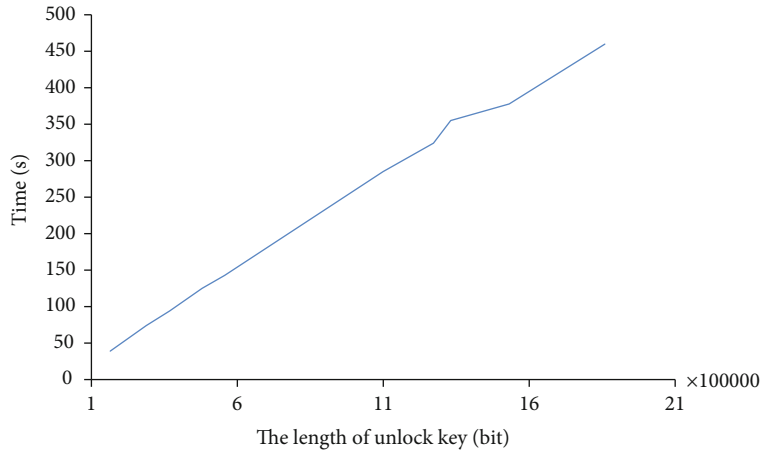


Figure 3: The performance of our smart lock key management scheme.

cannot calculate SK_{MS} by the known P since ECDLP problem

(2) SL selects private key SK_{SL}. SK_{SL} is confidential and satisfying $SK_{SL} < n$. SL computes public key $PK_{SL} = SK_{SL} \times P$ and sends PK_{SL} to MS through NB-IOT. In the transmission, even if PK_{SL} is attacked, the adversary cannot calculate SK_{SL} by the known P since ECDLP problem

(3) After the MS receives PK_{SL}, it uses the private key SK_{MS} and the received PK_{SL} to generate the secret key

$$K = SK_{MS} \times PK_{SL} = SK_{MS} \times SK_{SL} \times P = (x_K, y_K). \quad (3)$$

Similarly, SL uses the private key SK_{SL} and the received PK_{MS} to generate the secret key

$$K = SK_{SL} \times PK_{MS} = SK_{SL} \times SK_{MS} \times P = (x_K, y_K). \quad (4)$$

The two secret keys K are equal, which only are known as MS and SL. Because the secret key K is a pair of numbers (x_K, y_K), MS and SL can select the session key $k_{ML} = x_K + last(IMEI)$ according to the factory agreement, and function $last(IMEI)$ is the last digit of IMEI of SL. Because x_K is known only by MS and SL, the session key k_{ML} is also known only by them.

(4) MS generates 128 *bit* unlock key $k_{SL} = random()$ and sends encryption result $E_{k_{ML}}(k_{SL})$ to SL. At the same time, SL sends the encryption result $E_{k_{ML}}(\text{Info})$ to MS and reports the operation information $Info$, where E is AES symmetric encryption algorithm and $random()$ is random generating function

(5) Users download APP through their SP. SP selects private key SK_{SP}. SK_{SP} is confidential and satisfying $SK_{SP} < n$. SP computes public key $PK_{SP} = SK_{SP} \times P$, sends public key PK_{SP} and the request for unlock key to MS. MS will review the user's qualification after receiving the user's request. If the user does not have the housing qualification, the MS rejects his request

(6) If the user has the housing qualification, MS encrypts the unlock key with the received public key of the user and gets the ciphertext

$$C = E'_{PK_{SP}}(k_{SL}) \quad (5)$$

where E' is elliptic curve public key cryptosystem. Then, MS sends this ciphertext and the public key PK_{SL} to SP and sends the public key PK_{SP} to SL at the same time. In the transmission, even if C is overheard by the adversary, the adversary cannot get k_{SL} by decrypting C since the private key SK_{SP} is not known.

(7) After receiving the ciphertext C and the public key PK_{SL}, the SP uses its private key SK_{SP} to calculate

$$D_{SK_{SP}}(\text{C}) = D_{SK_{SP}}\left[E_{PK_{SP}}(\text{k}_{SL})\right] = \text{k}_{SL}. \quad (6)$$

Thus, the SP obtains the unlock key k_{SL} of SL.

(8) SP uses our ECC-based signcryption algorithm to generate $Signcrypt_{PK_{SL},SK_{SP}}(k_{SL}, \text{t})$ of unlock key k_{SL} and time stamp t, sends the signcryption to SL through Bluetooth. SL uses PK_{SP} and SK_{SL} to calculate

$$DeSigncrypt_{\text{KU}_{SP},SK_{SL}}\left[Signcrypt_{PK_{SL},SK_{SP}}(k_{SL}, t)\right] = (k_{SL}, \text{t}). \quad (7)$$

After getting the unlock key$k_{SL}$and time stamp$t$, SL checks them. If the unlock key is wrong, SL will not unlock. Our ECC-based signcryption scheme plays the role of encryption and authentication at the same time. The addition of time stamp can prevent replay attack.

(9) If the user loses the housing qualification, MS generates a new unlock key and sends it to SL. As a result, the SP cannot unlock with its old key k_{SL}

The flow chart of our key management is shown in Figure 2.

4.4. Performance Analysis. In this subsection, we observe the bit-oriented overhead of our smart lock key management scheme through experimental simulations. Here, AES symmetric encryption is performed in ECB mode, and the q in the elliptic curve used in our signcryption scheme and the order q of the group G_1 are both 160 bit. The experimental environment is as follows: AMD Ryzen 7 5800H, reference frequency 3.20 GHz, memory 16GB (DDR4-3200 MHz), and Windows 11 operating system.

As can be seen from Figure 3, when the length of the unlock key is as high as$21 \times 10^5 bit$, the time consumed of our key management scheme does not exceed$500\,s$. It is worth noting that in the actual deployment of the smart lock key management system, the length of the unlock key is generally not so long. Therefore, our key management scheme is practical and efficient.

5. Conclusion

In this paper, we designed an efficient and secure ECC-based signcryption scheme. Our signcryption scheme has been highly efficient and satisfies multiple security properties; it can also be used for mobile device authentication. Unfortunately, our signcryption scheme is only suitable for single-factor authentication. In the future, it will be interesting to consider applying our signcryption scheme to other application scenarios.

In addition, we proposed a practical and efficient key management scheme of the smart lock using our signcryption scheme firstly. Our key management scheme does not require manually memorizing the unlocking key, and every time the unlocking key is different, not being fixed. However, there has been a recent trend to study smart lock systems using deep learning methods. In addition to hand gesture recognition, face recognition is gradually popular. In the future, we will consider how to use deep learning methods in smart lock key management systems.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China (No.11401172), the Science and Technology Project of Henan Educational Committee of China (No.20A520012).

References

[1] Y. Zheng, "Digital signcryption or how to achieve cost (signature & encryption) << cost(signature) + cost(encryption)," in *Advances in Cryptology – CRYPTO '97, 17th Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 1997, Proceedings*, pp. 165–179, 1997.

[2] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, no. 177, pp. 203–209, 1987.

[3] V. S. Miller, "Use of elliptic curves in cryptography," in *Advances in Cryptology-CRYPTO'85, Santa Barbara, California, USA, August 18-22, 1985, Proceedings*, pp. 417–426, 1985.

[4] Y. Zheng and H. Imai, "How to construct efficient signcryption schemes on elliptic curves," *Information Processing Letters*, vol. 68, no. 5, pp. 227–233, 1998.

[5] R. Steinfeld and Y. Zheng, "A signcryption scheme based on integer factorization," in *Information Security, Third International Workshop, ISW 2000, Proceedings*, pp. 308–322, 2000.

[6] B. Libert and J.-J. Quisquater, "A new identity based signcryption scheme from pairings," in *Proceedings 2003 IEEE Information Theory Workshop, ITW 2003, La Sorbonne, Paris, France, 31 March -4 April, 2003*, pp. 155–158, 2003.

[7] J. Malone-Lee, "Identity-based signcryption," *IACR Cryptology ePrint Archive*, vol. 2002, p. 98, 2002.

[8] X. Boyen, "Multipurpose identity-based signcryption (a swiss army knife for identity-based cryptography)," in *Advances in Cryptology-CRYPTO 2003, Proceedings*, pp. 383–399, 2003.

[9] J. K. Liu, J. Baek, and J. Zhou, "Online/offline identity-based signcryption revisited," in *Information Security and Cryptology -6th International Conference, Inscrypt 2010, Shanghai, China, October 20-24, 2010, Revised Selected Papers*, pp. 36–51, 2010.

[10] S. S. D. Selvi, S. S. Vivek, and C. P. Rangan, "Identity based public verifiable signcryption scheme," in *Provable Security -4th International Conference, ProvSec 2010. Proceedings*, pp. 244–260, 2010.

[11] S. S. D. Selvi, S. S. Vivek, D. Vinayagamurthy, and C. P. Rangan, "ID based signcryption scheme in standard model," in *Provable Security -6th International Conference, ProvSec 2012. Proceedings*, pp. 35–52, 2012.

[12] A. W. Dent, "Hybrid signcryption schemes with insider security," in *Information Security and Privacy, 10th Australasian Conference, ACISP2005, Brisbane, Australia, July 4-6, 2005, Proceedings*, pp. 253–266, 2005.

[13] T. E. Bjørstad and A. W. Dent, "Building better signcryption schemes with tag-kems," in *Public Key Cryptography-PKC 2006, Proceedings*, pp. 491–507, 2006.

[14] M. Barbosa and P. Farshim, "Certificateless signcryption," in *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security, ASIACCS 2008, Tokyo, Japan, March 18-20, 2008*, pp. 369–372, 2008.

[15] T. Pandit, S. K. Pandey, and R. Barua, "Attribute-based signcryption: Signer privacy, strong unforgeability and IND-CCA2 security in adaptivepredicates attack," in *Provable Security -8th International Conference, ProvSec 2014. Proceedings*, pp. 274–290, 2014.

[16] P. Datta, R. Dutta, and S. Mukhopadhyay, "Compact attribute-based encryption and signcryption for general circuits from multilinear maps," in *Progress in Cryptology-INDOCRYPT 2015-16th International Conference on Cryptology in India, Bangalore, India, December 6-9, 2015, Proceedings*, pp. 3–24, 2015.

[17] P. Datta, R. Dutta, and S. Mukhopadhyay, "Functional signcryption: notion, construction, and applications," in *Provable Security -9th International Conference, ProvSec 2015, Proceedings*, pp. 268–288, 2015.

[18] W. Yang, M. Manulis, A. Man Ho, and W. Susilo, "Relations among privacy notions for signcryption and key invisible "sign-then-encrypt"," in *Information Security and Privacy -18th Australasian Conference, ACISP2013, Brisbane, Australia, July 1-3, 2013. Proceedings*, pp. 187–202, 2013.

[19] D. Nalla and K. C. Reddy, "Signcryption scheme for identity-based cryptosystems," *IACR Cryptology ePrint Archive*, vol. 2003, p. 66, 2003.

[20] B. Libert and J.-J. Quisquater, "Improved signcryption from q-diffiehellman problems," in *Security in Communication Networks, 4th International Conference, SCN 2004, Amalfi, Italy, September 8-10, 2004, Revised Selected Papers*, pp. 220–234, 2004.

[21] B. Waters, "Efficient identity-based encryption without random oracles," in *Advances in Cryptology-EUROCRYPT 2005, Proceedings*, pp. 114–127, 2005.

[22] Y. Yong, B. Yang, Y. Sun, and S. Zhu, "Identity based signcryption scheme without random oracles," *Computer Standards & Interfaces*, vol. 31, no. 1, pp. 56–62, 2009.

[23] J. Kar, "An efficient signcryption scheme from q-diffie-hellman problems," *IACR Cryptology ePrint Archive*, vol. 2012, p. 483, 2012.

[24] B. Libert and J.-J. Quisquater, "Efficient signcryption with key privacy from gap diffie-hellman groups," in *Public Key Cryptography-PKC 2004, 7th International Workshop on Theory and Practice in Public Key Cryptography, Singapore, March 1-4, 2004*, pp. 187–200, 2004.

[25] J. Kar, "Provably secure identity-based aggregate signcryption scheme in random oracles," *IACR Cryptology ePrint Archive*, vol. 2013, p. 37, 2013.

[26] L. I. U. Zhenhua, L. I. Juanjuan, and Z. U. Longhui, "Revocable id-based signcryption scheme," *Journal of Sichuan University (Engineering Science Edition)*, vol. 46, no. 2, pp. 79–86, 2014.

[27] A. Braeken and P. Porambage, "Efficient generalized signcryption based on ecc," *International Journal on Cryptography and Information Security (IJCIS)*, vol. 5, no. 2, pp. 1–13, 2015.

[28] J. Kar and S. Naik, "Generic construction of certificateless signcryption scheme," *IACR Cryptology ePrint Archive*, vol. 2016, p. 318, 2016.

[29] H. A. N. Yiliang, L. I. Chong, F. A. N. G. Dingyi, and Y. A. N. G. Xiaoyuan, "New signcryption scheme based on niederreiter cryptosystem," *Journal of Sichuan University (Engineering Science Edition)*, vol. 48, no. 2, pp. 97–103, 2016.

[30] Z. H. O. U. Yan-Wei, Y. A. N. G. Bo, and W. A. N. G. Qing-Long, "Secure certificateless signcryption scheme without bilinear pairing," *Journal of Software*, vol. 28, no. 10, pp. 2757–2768, 2017.

[31] C.-H. Tsai and S. Pin-Chang, "An ecc-based blind signcryption scheme for multiple digital documents," *Security and Communication Networks*, vol. 2017, Article ID 8981606, 14 pages, 2017.

[32] H. Dai, D. Wang, J. Chang, and X. Maozhi, "On the RCCA security of hybrid signcryption for internet of things," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8646973, 11 pages, 2018.

[33] A. W. Dent, "Hybrid signcryption schemes with outsider security," in *Information Security, 8th International Conference, ISC 2005, Singapore, September 20-23, 2005, Proceedings, volume 3650 of Lecture Notes in Computer Science*, pp. 203–217, 2005.

[34] D. He, D. Wang, and W. Shuhua, "Cryptanalysis and improvement of a password-based remote user authentication scheme without smart cards," *Information technology and control*, vol. 42, no. 2, pp. 170–177, 2013.

[35] D. Wang and P. Wang, "Two birds with one stone: two-factor authentication with security beyond conventional bound," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 708–722, 2018.

[36] J. I. A. Rong-yuan, W. A. N. G. Yi-huai, and W. A. N. G. Xiaoning, "Lightweight encryption algorithm for narrowband internet of things," *Computer Engineering and Design*, vol. 39, no. 10, pp. 3039–3044, 2018.

[37] L. Chih-Chung and S.-Y. Tseng, "Integrated design of AES (advanced encryption standard) encrypter and decrypter," in *13th IEEE International Conference on Application-Specific Systems, Architectures, and Processors (ASAP 2002), 17-19 July 2002, San Jose, CA, USA*, pp. 277–285, IEEE Computer Society, 2002.

[38] Z. H. A. N. G. Huan-lan and X. I. A. O. Ming-bo, "Design of intelligent lock system based on 433mhz band security," *Computer Engineering and Design*, vol. 39, no. 9, pp. 2736–2742, 2018.

[39] L. I. U. Meng-jun, S. H. A. Tao, L. I. Dan, and L. I. U. Shu-bo, "Reliable security lock key updating scheme over narrow band internet of things," *Computer Science*, vol. 46, no. 4, pp. 137–143, 2019.

[40] S. Tao, L. Mengjun, L. Dan, and L. Shubo, "Nb-iot security smart lock system solution under background of public rental housing," *Application Research of Computers*, vol. 36, no. 6, pp. 1797–1802, 2019.

[41] Y. Wang, Y. Shu, M. Xiuqian Jia, L. X. Zhou, and L. Guo, "Multifeature fusion-based hand gesture sensing and recognition system," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8991316, 14 pages
https://doi.org/10.1155/2022/8991316

Research Article

A Training Sequence-Based Ranging Method for R-Mode of VHF Data Exchange System

Xiaowen Sun, Qing Hu, and Yi Jiang

College of Information Science and Technology, Dalian Maritime University, Linghai Road No. 1, Dalian 116026, China

Correspondence should be addressed to Xiaowen Sun; sunxiaowen_1984@163.com

Received 15 February 2022; Revised 22 March 2022; Accepted 6 April 2022; Published 5 May 2022

Academic Editor: Mu Zhou

The space-time reference provided by GNSS is the basis of the 6G mobile communication integrating land, sea, air, and space. But GNSS has natural vulnerability. In order to solve this problem at sea, ships should be equipped with both space-based and land-based positioning and navigation systems in the future. The Ranging Mode (R-Mode) of the existing maritime communication system is an economical land-based backup positioning system, which is also an important support for e-Navigation. This paper proposes a pseudorange measurement method based on training sequence with multicorrelators, which can be used in the R-Mode of the VHF data exchange system (VDES). This method uses the correlation property of the ASM and VDE training sequence in the VDES for ranging. It reproduces the training sequence in the receiver, and then, noncoherent correlates with the received signal, meanwhile using the pseudorange measurement method with multicorrelators to reduce the influence of pulse shaping on the correlation curve. This paper also gives a design of receiver to implement the proposed method. Experiments are carried out based on the receiver to evaluate the ranging performance. The results show that, compared with the traditional maximum correlation value and E-L methods, the multicorrelator method can effectively reduce the ranging error. When E_s/N_0 is greater than 15 dB, the ranging accuracy of the proposed method can be better than 3 m. With the proposed method, the VDES signal can be used for both ranging and communication. Compared with the general VDES function, it does not occupy any additional time slot resources. This provides a reference for the future research of the VDES R-Mode, which can be used to solve the vulnerability problem of the GNSS on maritime communication.

1. Introduction

The global internet of things in the land, sea, air, and space, that is, the 6G mobile communication system, will deeply integrate navigation, communication, and remote sensing technologies [1–3]. And the space-time reference provided by GNSS is an important basis for the integration. However, the GNSS is vulnerable to unintentional or intentional interference, which leads to the loss or error of Position, Navigation, and Timing (PNT) information due to the weakness of the signals. The International Maritime Organization (IMO) has been considering the vulnerability of GNSS for a long time and suggests the development of land-based radio navigation systems as an effective and reliable solution to this problem [4, 5]. World Wide Radio Navigation Plan (WWRNP) pointed out that the ships should be equipped with both space-based and land-based positioning and navigation systems in the future to ensure its safety [6]. The R-Mode land-based positioning and navigation system uses electronic measurement to get the distance between the ship and the base station for positioning. It can make full use of the existing shore-based equipment, saving the infrastructure construction costs.

Automatic Identification System (AIS) is a widely used maritime communication system for exchanging relevant information between ships and base stations operating in the very high-frequency (VHF) radio band around 162 MHz [7]. The member states of IMO have been leading in and enforcing the use of AIS in the analysis of ship-to-ship collisions, vessel monitoring, and maritime traffic management offshore [8]. AIS also plays an important role

in avoiding ship collisions and protecting ship safety. However, with the increase of the number of ships, AIS channel overload has become an urgent problem for some ports [9]. In 2012, the International Telecommunication Union (ITU) first proposed the concept of VHF data exchange system (VDES) [10]. It is an enhanced and upgraded system of AIS in the VHF radio band 156.025–162.025 MHz. Based on the existing AIS channel, it adds the special application message (ASM) channel and broadband very high-frequency data exchange (VDE) channel, which can effectively relieve the slot pressure of data exchange in AIS. VDES also adds the function of satellite communication; that is, the VDE channel includes VDE-Terrestrial (VDE-TER) and VDE-Satellite (VDE-SAT), and the ASM channel includes ASM-TER and ASM-SAT [11, 12]. The VDES provides a variety of means for the exchange of data between maritime stations, ship-to-ship, ship-to-shore, shore-to-ship, ship-to-satellite, and satellite-to-ship. It provides an effective auxiliary mean for the safety of ship and comprehensively improves the capability and frequency efficiency of marine data communication [13]. The VDES has been established as an important component of future maritime communication systems in e-Navigation by IMO [14, 15].

The VDES is a communication system, and its R-Mode uses communication signals for ranging and positioning. The signals originally designed for communication also carry pseudorange measurement information, which can be used for ranging. Such signals are signals of opportunity (SOP). In recent years, more and more scholars began to research on the navigation via signal of opportunity (NAVSOP). NAVSOP regards all potential radio signals in the surrounding environment as SOP and extracts location and time information for navigation from them [16]. In the urban and indoor environment where GNSS signals are seriously blocked, the surrounding SOP can be used to assist the positioning. It can greatly improve the positioning accuracy without adding any sensors or transmitters [17–19]. The mobile signals can be SOP for each other in mobile wireless positioning [20–22]. The GNSS signals can combine with the SOP such as WCDMA (Wideband Code Division Multiple Access), LTE (Long-Term Evolution), UWB (Ultra Wideband), and broadcast signals to enhance the GNSS positioning [23–26]. The SOP can also be integrated into the integrated navigation of Inertial Navigation System (INS) and GNSS. It can assist GNSS positioning in complex environment [27–29]. The existence, availability, and combination are the unique attributes of the SOP, which are both advantages and challenges.

Although there are many NAVSOP-related researches on land, there are relatively much less such researches in maritime field. The European Union-funded Accessibility for Shipping, Efficiency Advantages and Sustainability (ACCSEAS) project demonstrated the feasibility of R-Mode using MF DGPS, AIS, and e-Loran transmissions [30–32], and reference [33] extended the theoretical analysis of ranging precision to the VDES R-Mode. The Chinese AIS ship Autonomous Positioning System (AAPS) project realized the positioning function of AIS R-Mode system [34, 35] and carried out the theoretical research of the VDES R-Mode [36]. Based on that, the AIS/VDES R-Mode testbed was built in the Yellow Sea and Bohai Sea in China. Meanwhile, R-Mode Baltic project built the R-Mode testbed in the Baltic Sea and researched on both MF and VHF R-Mode [37–39]. In the studies of R-Mode pseudorange measurement method in the VDES, reference [40] used GMSK modulated AIS communication signal for ranging, which had low accuracy. And reference [41] used the special pseudorandom sequence of VDE channel for ranging, which can get higher accuracy. But it needed specific time slots for the sequence transmission and occupied the communication time slots for the ranging. In the latest researches on VDES R-Mode, reference [42] considered the system security and proposed the concept of authentication, and reference [43] investigated the feasibility of a satellite-based component to VDES R-Mode. These studies are also based on the special pseudorandom sequence of VDE channel.

This paper proposes a pseudorange measurement method of multicorrelators based on training sequence for the VDES R-Mode. It can realize high-accuracy ranging while communicating, without occupying any additional time slot resources. It uses the correlation property of the training sequence of ASM-TER (terrestrial) and VDE-TER communication message in the VDES for ranging. $\pi/4$ QPSK modulation is used for the training sequence that is composed of Barker13 and inverted Barker13 codes with good correlation property. The receiver reproduces the IF signal of training sequence, and noncoherent correlates it with the received signal. And it uses multicorrelator pseudorange measurement method to reduce the influence of pulse shaping on the correlation curve to improve the ranging accuracy.

The main contributions of this paper include the following: (1) we propose a ranging method based on the VDES training sequence, which is normally used to capture and synchronize the communication signal. It can provide a reference for the future research of the VDES R-Mode; (2) we discuss the influence of the pulse shaping in the VDES on the ranging performance. This is a new scientific problem for R-Mode of communication system like VDES; (3) we give a receiver design with a structure of multicorrelator, which can be used to realize the ranging method proposed in this paper. It can reduce the influence of the pulse shaping on the ranging performance and realize high-accurate ranging function without occupying any additional time slot resources.

The proposed method in this paper can be used in VDES R-Mode receiver to measure the pseudorange between the receiver and the base station that transmit the communication signals. Since the location of the base station is known, the receiver can calculate its position by solving the pseudorange equations when it gets the pseudoranges of more than three base stations.

The rest of the paper is organized as follows: Section 2 introduces the theoretical principle of the ranging method based on the training sequence in the VDES; Section 3 introduces the implementation of the training sequence-based multicorrelator ranging method; Section 4 shows the simulation results and analysis of the ranging performance; and Section 5 summarizes the main conclusions.

2. Theoretical Principle of the VDES Ranging Method

2.1. Ranging Signal in the VDES

2.1.1. Terrestrial Subsystems of the VDES. The VDES includes three subsystems: AIS, ASM, and VDE. This paper only deals with the terrestrial component of the VDES, because the current satellite component is not suitable for positioning for their location cannot be accurately provided to the users all the time [33]. The technical specifications of the VDES terrestrial component are summarized in Table 1.

The AIS uses two 25 kHz channels (AIS 1 and AIS 2) for ship position reporting and other safety-related applications. It uses Gaussian Minimum Shift Keying (GMSK) modulation. The symbol rate, $R_S = 1/T_S$, is 9600 symbols/s (sps), where each symbol has one bit, that is, 9600 bits/s (bps).

The ASM-TER uses two 25 kHz channels (ASM 1 and ASM 2) and $\pi/4$ Quaternary Phase Shift Keying ($\pi/4$ QPSK) modulation. It gives a high reliability of message delivery and message acknowledgement support. The symbol rate, R_S is 9600 sps, where each symbol has two bits, that is, 19200 bps.

The VDE-TER has 100 kHz bandwidth which can optionally be configured to work as 50 kHz or 25 kHz bandwidth. It can also be configured to use $\pi/4$ QPSK, Eight-state Phase Shift Keying (8-PSK), or 16-state Quadrature Amplitude Modulation (16-QAM) modulations [44]. The symbol rate R_S increases from 19200 sps to 76800 sps with the increase of bandwidth. The bit rate depends not only on symbol rate but also on the modulation mode.

2.1.2. $\pi/4$ QPSK Modulation. The proposed method in this paper is based on the $\pi/4$ QPSK modulated training sequence in ASM channel and VDE channel.

The $\pi/4$ QPSK is a modulation technique which is developed in QPSK and offset QPSK (OQPSK) and is often used in differential coding [45, 46]. It has twice the bandwidth efficiency of the BPSK, since two bits are transmitted in a single modulation symbol. The kth symbol of the $\pi/4$ QPSK modulated baseband signal s_k in the interval of $kT_s \le t \le (k+1)T_s$ can be expressed as

$$s_k = \sqrt{2\frac{E_s}{T_s}}e^{j\theta_k}, \tag{1}$$

where T_s is the symbol interval, E_s is the energy per symbol, and θ_k is the phase of the kth symbol.

In $\pi/4$ QPSK, the maximum phase change is limited to $\pm 3\pi/4$, as compared to π for QPSK and $\pi/4$ for OQPSK. Therefore, the bandlimited $\pi/4$ QPSK signal preserves the constant envelope property better than bandlimited QPSK but is more susceptible to envelope variations than OQPSK [47].

A more attractive feature of $\pi/4$ QPSK is that it can be noncoherently detected, which greatly simplifies receiver design [47–50]. In this paper, we make use of this feature and implement the proposed method with a noncoherent receiver structure.

2.1.3. Signal Used for Ranging. A frame in the VDES equals one minute and is divided into 2250 slots; that is, each slot lasts about 26.67 milliseconds (ms). The general slot format of ASM-TER and VDE-TER is shown in Table 2. Each slot consists of six parts: ramp up, training sequence, link ID, data, ramp down, and guard. The ramp up time from −50 dBc to −1.5 dBc of the power is 416 microseconds (μs), to provide spectral shaping and reduce interference, and the modulation is not specified for the ramp up. The training sequence is the focus of this paper, which will be introduced in detail in the next paragraph. The link ID follows the training sequence and uses $\pi/4$ QPSK modulation to define the channel configurations. The data payload with its Cyclic Redundancy Check (CRC) is interleaved encoded scrambled and bit mapped. The ramp down time from full power to −50 dBc should be no more than 416 μs. The rest guard time is for delay and jitter.

The ranging method proposed in this paper is based on the training sequence of ASM-TER and VDE-TER. It is a 27-symbol training sequence and uses $\pi/4$ QPSK modulation. The last 26 symbols are Barker 13 code (1 1 1 1 1 0 0 1 1 0 1 0 1) and inverted Barker 13 code (0 0 0 0 0 1 1 0 0 1 0 1 0) with ideal autocorrelation, which can be used to detect the weak target signal submerged in noise. And this paper uses the ideal autocorrelation of the double Barker 13 code for ranging. In the training sequence, the symbol "1" maps to $\pi/4$ QPSK symbol "3" (1 1), and the symbol "0" maps to $\pi/4$ QPSK symbol "0" (0 0).

Figure 1 shows the bit mapping for $\pi/4$ QPSK used in ASM-TER and VDE-TER and the phase alternating of the training sequence. There are 4 possible phase variations of $\pm\pi/4$ and $\pm 3\pi/4$ when the symbol changes. Since there are only "11" and "00" in the training sequence, without "01" and "10," it has only four kinds of phase alternating as shown in Figure 1. The first the symbol "1" of the training sequence maps to $\pi/4$ QPSK symbol "3" (1 1) is mapped to the constellation defined by the point $(1+j)/\sqrt{2}$; the next symbol "1" is mapped to the constellation defined by point $1+0j$ (shown in blue in Figure 1); the next symbol "1" is mapped to the constellation defined by point $(-1-\mathrm{j})/\sqrt{2}$ (shown in green in Figure 1), and so on.

2.2. Theoretical Principle

2.2.1. Noncoherent Correlation of the Training Sequence. The 27 symbols of the training sequence can be serial-parallel inverted into in-phase (I) and quadrature (Q) branches with the same values; that is, the values of I and Q branches of the kth symbol are $s_{Ik} = s_{Qk} = 0$ or $s_{Ik} = s_{Qk} = 1$. After the bit mapping shown in Figure 1, the signals of I and Q branches can be expressed as follows:

$$I_k = \cos\theta_k = I_{k-1}\cos\phi_k - Q_{k-1}\sin\phi_k, \tag{2}$$

$$Q_k = \sin\theta_k = I_{k-1}\sin\phi_k - Q_{k-1}\cos\phi_k, \tag{3}$$

where

Table 1: The technical specifications of the VDES terrestrial component.

Subsystems	Signal bandwidth	Modulations	Symbol rate	Bit rate
AIS	25 kHz	GMSK	9.6 ksps	9.6 kbps
ASM-TER	25 kHz	$\pi/4$ QPSK	9.6 ksps	19.2 kbps
VDE-TER	25 kHz/ 50 kHz/ 100 kHz	$\pi/4$ QPSK/ 8-PSK/ 16-QAM	19.2 ksps (25 kHz)/ 38.4 ksps (50 kHz)/ 76.8 ksps (100 kHz)	Depends on the modulation and the symbol rate

Table 2: ASM-TER and VDE-TER general slot format.

Ramp up	Training sequence	Link ID	Data	Ramp down	Guard
0.41 ms	27 symbols (1 1111100110101 0000011001010)	16 symbols	Data with CRC	0.41 ms	0.83 ms

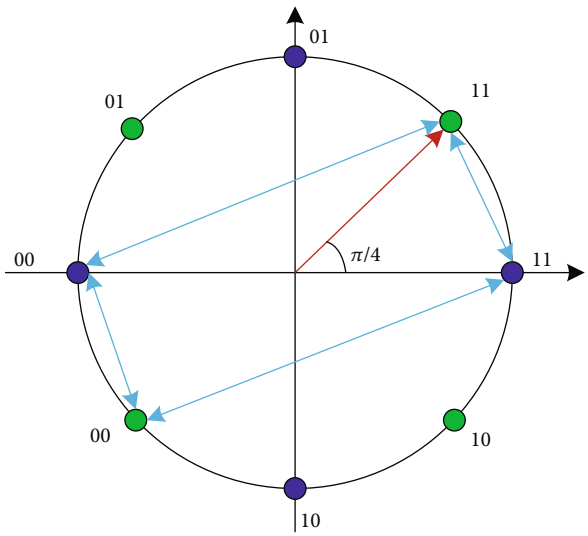


Figure 1: Bit mapping for $\pi/4$ QPSK and phase alternating of the training sequence.

$$\theta_k = \theta_{k-1} + \phi_k, \tag{4}$$

where θ_k and θ_{k-1} are the phases of the kth and the $k-1$st symbols and ϕ_k is the phase shift of the kth symbol compared with the $k-1$st symbol. The value of ϕ_k can be $\pm \pi/4$ and $\pm 3\pi/4$ depending on the input symbols. According to Figure 1, the value of ϕ_k can be determined by the kth symbol s_k, the $k-1$st symbol s_{k-1} and the initial values of s_k and θ_k. The specific values of these $\pi/4$ QPSK modulation parameters of the first 10 symbols in the training sequence are as shown in Table 3. The initial values of s_k and θ_k are $s_0 = 1$ and $\theta_0 = \pi/4$.

The baseband signals of the I and Q branches shown in equations (2) and (3) are, respectively, multiplied by the in-phase carrier signal $\cos \omega_c t$ and the quadrature carrier signal $-\sin \omega_c t$, and then, add the two branches signal to complete the carrier modulation. After that, within the kth symbol duration of $kT_s \le t \le (k+1)T_s$, the $\pi/4$ QPSK modulated signal can be expressed as

$$\begin{aligned} s_k(t) &= I_k \cos \omega_c t - Q_k \sin \omega_c t \\ &= \cos \theta_k \cos \omega_c t - \sin \theta_k \sin \omega_c t \\ &= \cos (\omega_c t + \theta_k), \end{aligned} \tag{5}$$

If the receiver can locally reproduce the modulated wave of $\pi/4$ QPSK, its frequency is the same as the received signal from the transmitter, and the phase is not necessarily the same; then, the locally reproduced signal can also be divided into I and Q branches. The signals of I and Q branches in the time interval of $kT_s \le t \le (k+1)T_s$ can be, respectively, expressed as

$$l_{Ik}(t) = I_k \cos (\omega_c t + \alpha) - Q_k \sin (\omega_c t + \alpha) = \cos (\omega_c t + \alpha + \theta_k), \tag{6}$$

$$l_{Qk}(t) = I_k \sin (\omega_c t + \alpha) - Q_k \cos (\omega_c t + \alpha) = -\sin (\omega_c t + \alpha + \theta_k). \tag{7}$$

By multiplying the received signal from the transmitter as equation (5) and the I and Q signals reproduced by the receiver as equations (6) and (7), respectively, and then integrating them, the results are as follows:

$$ac_{Ik} = \int_{kT_s}^{(k=1)T_s} \cos (\omega_c t + \theta_k) \cos (\omega_c t + \alpha + \theta_k) dt = \frac{T_s}{2} \cos \alpha, \tag{8}$$

$$ac_{Qk} = -\int_{kT_s}^{(k=1)T_s} \cos (\omega_c t + \theta_k) \sin (\omega_c t + \alpha + \theta_k) dt = \frac{T_s}{2} \sin \alpha. \tag{9}$$

Equations (8) and (9) give the integration results of I and Q branches in a symbol time interval of $kT_s \le t \le (k+1)T_s$. If all the 26 symbols of the double Barker codes are integrated, the results are also determined by the autocorrelation of the double Barker codes. Meanwhile, in order to remove the influence of the carrier phase difference between the reproduced signal and the received signal, that is, α, the noncoherent correlation is carried out.

Table 3: $\pi/4$ QPSK modulation parameters in the training sequence.

k	θ_k	ϕ_k	s_k	s_{k-1}
0	$\pi/4$	/	1	/
1	0	$-\pi/4$	1	1
2	$\pi/4$	$\pi/4$	1	1
3	0	$-\pi/4$	1	1
4	$\pi/4$	$\pi/4$	1	1
5	0	$-\pi/4$	1	1
6	$-3\pi/4$	$-3\pi/4$	0	1
7	+/-π	$-\pi/4$	0	0
8	$\pi/4$	$-3\pi/4$	1	0
9	0	$-\pi/4$	1	1
10	$-3\pi/4$	$-3\pi/4$	0	1
…	…	…	…	…

The receiver reproduces the last 26 symbols of the training sequence, which are the double Barker codes. The I and Q branches are multiplied with the received signal from the transmitter and then integrated, respectively. The integral results of I and Q branches are then squared, respectively, and then, the two results are added to get the noncoherent correlation value as follows:

$$\mathrm{Cor} = \mathrm{CF} \cdot \left[\left(\sum_{k=0}^{N-1} ac_{Ik} \right)^2 + \left(\sum_{k=0}^{N-1} ac_{Qk} \right)^2 \right] = \mathrm{CF} \cdot N^2 \cdot \frac{T_s^2}{4}, \tag{10}$$

where $N = 26$ indicates the number of the integrated symbols and CF is the correlation factor, which can be expressed as follows:

$$\mathrm{CF} = \frac{1}{N} \cdot \sum_{i=0}^{N} [\mathrm{ms}(i) - \mathrm{ns}(i)], \tag{11}$$

where ms is the number of matched symbols and ns is the number of mismatched symbols. According to the structure of the training sequence shown in Table 2, if the first symbol of the link configuration ID is 0, as received sequence shown in Figure 2, CF is the maximum value of 1 when the local reproduced 26 symbols exactly match the received signal for ms which is 26 and ns which is 0. And CF is the minimum value of 0 when the locally reproduced 26 symbols are 1 symbol time Ts earlier or later than the received signal, for ms is 13 and ns is 13, as shown in Figure 2.

Assuming that the time deviation between the local reproduced signal and the received signal is $\Delta\tau$, when $-T_s \le \Delta\tau \le T_s$, CF can be expressed as

$$\mathrm{CF}(\Delta\tau) = 1 - \frac{|(\Delta\tau)|}{T_s}. \tag{12}$$

2.2.2. The TOA Ranging. Ranging system based on radio signal is essentially a measurement system of transmission delay [51], so is the VDES R-Mode. The transmitter sends out the radio signal at the start of the slot, expressed as t_T. When the local reproduced sequence in the receiver completely matches the sequence of the received signal, that is, $\Delta\tau$ is 0, the time of arrival (TOA) can be obtained. The difference between TOA and t_T is the transmission delay, expressed as Δt. The distance between the transmitter and the receiver is the result of multiplying the transmission delay Δt by the speed of the electromagnetic wave c, as shown in equation (13). This distance is also known as pseudorange, because the value of distance is not accurate due to the influence of the clock difference between the transmitter and the receiver.

$$d = c \cdot \Delta t = c \cdot (\mathrm{TOA} - t_T). \tag{13}$$

Thus, the time deviation between the local reproduced signal and the received signal $\Delta\tau$ can be determined by the correlation value. When the correlation value is maximum, $\Delta\tau$ is the ideal case of 0. The TOA can be determined by $\Delta\tau$. Since the transmission time t_T is fixed at the beginning of the slot, it is easy to get the pseudorange measurement value according to formula (12).

2.3. Influence of the Pulse Shaping. When rectangular pulses are passed through a bandlimited channel, the pulses will spread in time, and the pulse for each symbol will smear into the time intervals of succeeding symbols. This causes Intersymbol Interference (ISI) and leads to an increased probability of the receiver making an error in detecting a symbol [47]. To solve this problem, spectral shaping is usually done through baseband or IF processing, since it is difficult to directly manipulate the transmitter spectrum at RF frequencies. There are some pulse shaping techniques including Nyquist criterion [52], (root) raised cosine roll-off filter, and Gaussian pulse shaping filter which can be used to reduce the ISI and reduce the bandwidth of a modulated signal.

In order to improve the communication quality of VDES and reduce Intersymbol Interference (ISI), the baseband shall employ a root raised cosine filter with roll-off factor 0.35 in ASM channel and 0.30 in VDE channel for the pulse shaping:

$$H(f) = \begin{cases} 1 & 0 \le |f| \le \dfrac{1-\beta}{2T_s}, \\ \sqrt{\dfrac{1}{2}\left[1 + \cos\left(\dfrac{\pi(|f| \cdot 2T_s - 1 + \beta)}{2\beta}\right)\right]} & \dfrac{1-\beta}{2T_s} < |f| < \dfrac{1+\beta}{2T_s}, \\ 0 & |f| \ge \dfrac{1+\beta}{2T_s}, \end{cases} \tag{14}$$

where $\beta = 0.35$ in the ASM channel and $\beta = 0.30$ in the VDE channel.

After pulse shaping using equation (14), Figure 3(a) shows the correlation curves of in 4 different cases.

Received:	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	
Local 1 symbol Earlier:	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0			ms = 13; ns = 13
Local Prompt:		1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0		ms = 26; ns = 0
Local 1 symbol Later:			1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	ms = 13; ns = 13

Figure 2: Number of matched symbols (ms) and number of mismatched symbols (ns).

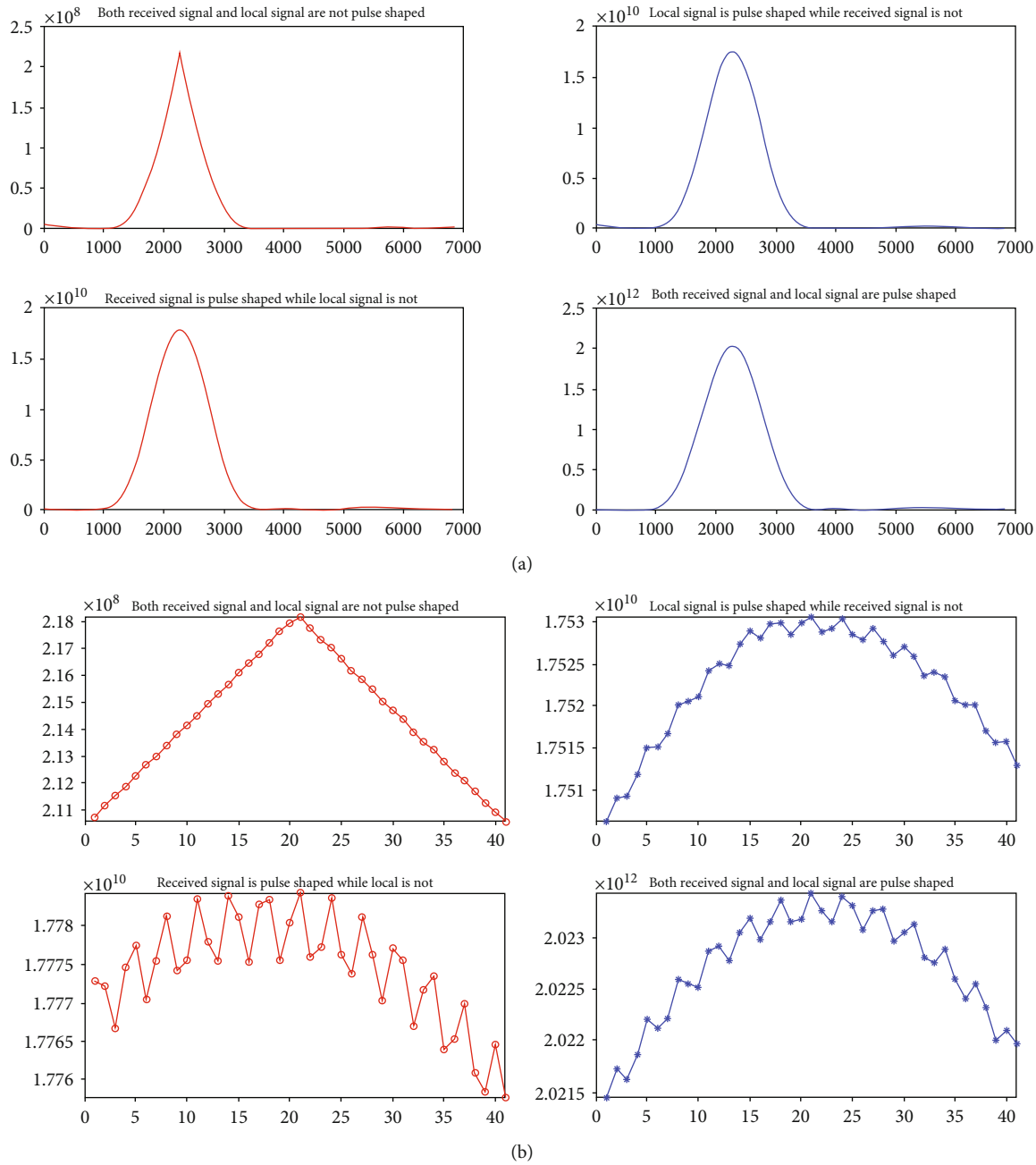


Figure 3: Correlation curves in 4 different cases: (a) the whole correlation curves; (b) the amplificated correlation curves with the maximum correlation value as the center.

Figure 3(b) shows the amplification result with the maximum correlation value as the center.

In the first case, when pulse shaping was not carried out for both the received signal and the local reproduced signal, the correlation value and $\Delta\tau$ are ideal linear relation that met equations (10), (11), and (12). It can be seen from the first figures in Figures 3(a) and 3(b) that the points of correlation values form a symmetrical triangle. The point of maximum value is the synchronous phase.

In the other three cases, when pulse shaping was carried out for the received signal or/and the local reproduced signal, the shape of the correlation curve changed, and the linear relation between the correlation value and $\Delta\tau$ was worse. It can be seen from the rest three figures in Figures 3(a) and 3(b) that the points of correlation values are not asymmetric. The point of maximum value is not the synchronous phase.

3. Implementation of the Proposed Method

3.1. Structure of Proposed VDES R-Mode Receiver. Take the ASM channel receiver as an example. The structure shown in Figure 4 can be used for communication and ranging at the same time. The signal received by the antenna first goes through the Radio Frequency (RF) front-end circuit; after downconversion, Analog-Digital (A/D) sampling, it then becomes the digital Intermediate Frequency (IF) signal, which is connected to the IF circuit. $\pi/4$ QPSK demodulation for ASM messages and correlation processing for TOA are carried out at the same time in the IF circuit. The processor uses the demodulated ASM message for normal data communication and uses the TOA measurement value for positioning.

In the RF circuit, the input signals are sampled by 120 MHz clock signal. The ASM 1 channel signals with 161.95 MHz center frequency are converted into 41.95 MHz digital IF signals. And the ASM 2 channel signals with 162 MHz center frequency are converted into 42 MHz digital IF signals. All the experiments in this paper are based on these digital IF signals and are carried out in the correlator of the IF circuit. The frequency of the master clock used in the experiments is 120 MHz.

3.2. Implementation of the Proposed Method. According to the correlation curve given in Figure 3(b), the linear relation between the correlation value and $\Delta\tau$ is affected by pulse shaping, while the symmetry of the correlation curve is not. In this case, the traditional maximum correlation value and E-L methods can cause a considerable error.

In this paper, the multicorrelator pseudorange measurement method was used and implemented as shown in Figure 5 to realize the proposed training sequence-based ranging method. It was implemented in the correlator design of the IF circuit based on the structure of the receiver shown in Figure 4. The influence of pulse shaping was also considered in this implementation, with the local reproduced baseband signal pulse shaped.

In the implementation diagram shown in Figure 5, the signal mapping module generates the I and Q branches of the 26-symbol double Barker code baseband signals. And then, the signals are pulse shaped by the root raised cosine filter with a roll-off coefficient of 0.35. Meanwhile, the local carrier generator generates the in-phase carrier signal $\cos \omega_c t$ and the quadrature carrier signal $-\sin \omega_c t$ with the same frequency as the transmitted signal. The I and Q branch local carrier signals and the I and Q branch local baseband signals are then multiplied by each other and added to get the I and Q branch local reproduced IF signal. Delay the local reproduced IF signals with the same time intervals t_d to get 1 + 2M duplicated signals. Finally, the received signal and the local signals are noncoherent correlated to get 1 + 2M correlation values.

The 1 + 2M correlation values can be centered with the maximum correlation value by adjusting the delay time of local reproduced IF signals. In addition to the maximum correlation value, there are M groups of correlation values. Each group includes two correlation values that are Ek and Lk. The Ek is $k \cdot t_d$ earlier than the centered maximum correlation value, and the Lk is $k \cdot t_d$ time later. After calculating the average of the earlier M correlation values $E_1, E_2, E_3 \cdots E_M$ and the later M correlation values $L_1, L_2, L_3 \cdots L_M$, respectively, the relation with $\Delta\tau$ is established:

$$\Delta\tau = c \cdot \frac{\sum_{i=1}^{M} E_i - \sum_{i=1}^{M} L_i}{\sum_{i=1}^{M} E_i + \sum_{i=1}^{M} L_i}. \tag{15}$$

$\Delta\tau$ can be used to calculate the value of TOA. Since the transmission time t_T is the beginning of the slot, the pseudorange measurement value can be calculated according to Section 2.2.2.

4. Results and Discussion

4.1. Correlation Results and Discussion. The correlation experiments were carried out to

(1) verify the advantage of using pulse shaping filter for local reproduced signal in the proposed receiver structure

(2) test the correlation property in different noise conditions

There were mainly three steps in the correlation experiments:

(1) Generated the received signal in different noise conditions

(2) Generated local reproduced signal

(3) Did the correlation operation to get the correlation curve

4.1.1. Received Signal. Since the distance between transmitter and receiver is limited in VDES, only the received signals in a limited range are needed for the correlation process. The received signal generated by simulation is a digital IF signal of 32-symbol sequence, which is comprised of 1-symbol

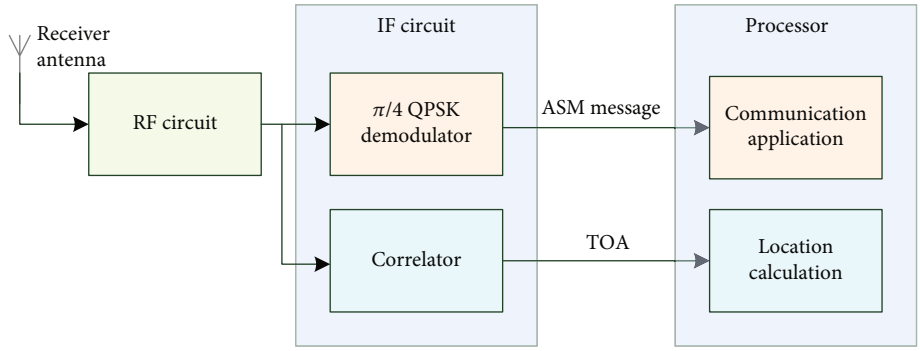


Figure 4: Block diagram of ASM channel receiver structure.

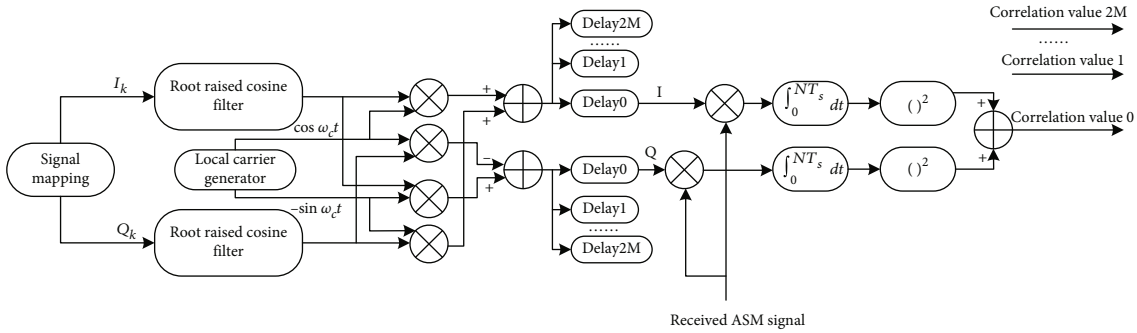


Figure 5: Implementation diagram of multicorrelator pseudorange measurement method.

ramp up (0), 27-symbol training sequence (1 1 1 1 1 1 0 0 1 1 0 1 0 1 0 0 0 0 0 1 1 0 0 1 0 1 0), and the following 4-symbol link ID (0 1 0 1). The frequency of the master clock used in the experiments is 120 MHz as introduced in Section 3.1, the symbol rate of ASM signal used in the experiments is 9600 sps, and then, 1/11 sampling is used to reduce the computation burden. Therefore, the data points of the generated received digital IF signal is as follows:

$$P_R = 32\text{symbols} \cdot \frac{1}{9600\ \text{sps}} \cdot 120\text{MHz} \cdot \frac{1}{11} = 3.6364 \times 10^4. \tag{16}$$

Besides, the received signal should be pulse shaped by root raised cosine filter with a roll-off coefficient of 0.35 in the transmitter and should be added some noise in the transmission channel.

Figure 6 shows the 36364 points of the received digital IF signal, which is pulse shaped in the transmitter and added some Additive White Gaussian Noise (AWGN) in the transmission channel. The Signal-to-Noise Ratio (SNR) decreases with the increase of noise, and it can be seen from Figure 6 that when the SNR is less than -20 dB, the changes in amplitude of the received digital IF signal which caused by the pulse shaping is hidden in the noise.

4.1.2. Local Reproduced Signal. The local reproduced signal generated by simulation is a digital IF signal of the 26-symbol sequence, which is the double Barker 13 code of the training sequence (1 1 1 1 1 0 0 1 1 0 1 0 1 0 0 0 0 0 1 1 0 0 1 0 1 0). The frequency of the master clock is 120 MHz, the symbol rate is 9600 sps, and 1/11 sampling is used. These parameters are exactly same as those in the generation of the received signal. The data points of the local digital IF signal is

$$P_L = 26\text{symbols} \cdot \frac{1}{9600\ \text{sps}} \cdot 120\text{MHz} \cdot \frac{1}{11} = 2.9545 \times 10^4. \tag{17}$$

In order to verify the advantage of using pulse shaping filter for local reproduced signal in the proposed receiver structure, the local reproduced signals with and without pulse shaping are both generated. Figure 7 shows the 29545 points of the local reproduced I and Q branch baseband signals with and without pulse shaping. It can be seen from the figure that the amplitude of the baseband signals changes continuously after the pulse shaping.

The baseband signals shown in Figure 7 are then modulated by the local generated carrier signal to get the local reproduced digital IF signals, which are needed in the correlation experiments.

4.1.3. Correlation Results and Discussion. Twenty experiments were carried out based on the received digital IF signals and the local reproduced digital IF signals. In each

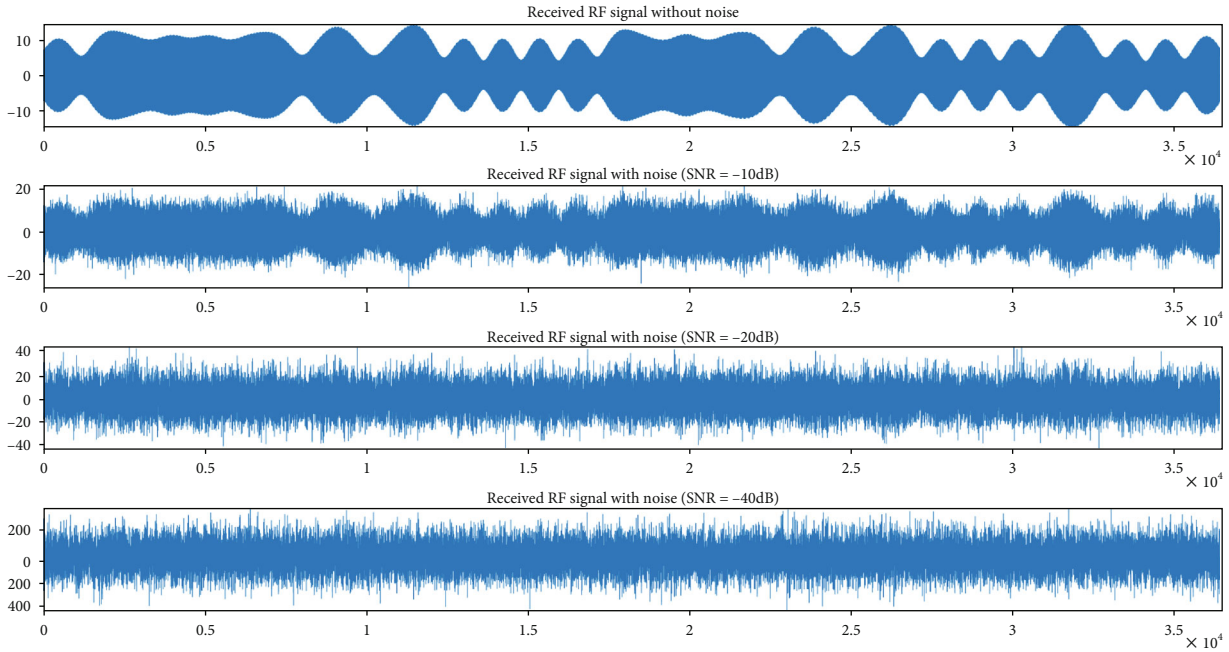


Figure 6: Received RF signal in different noise conditions.

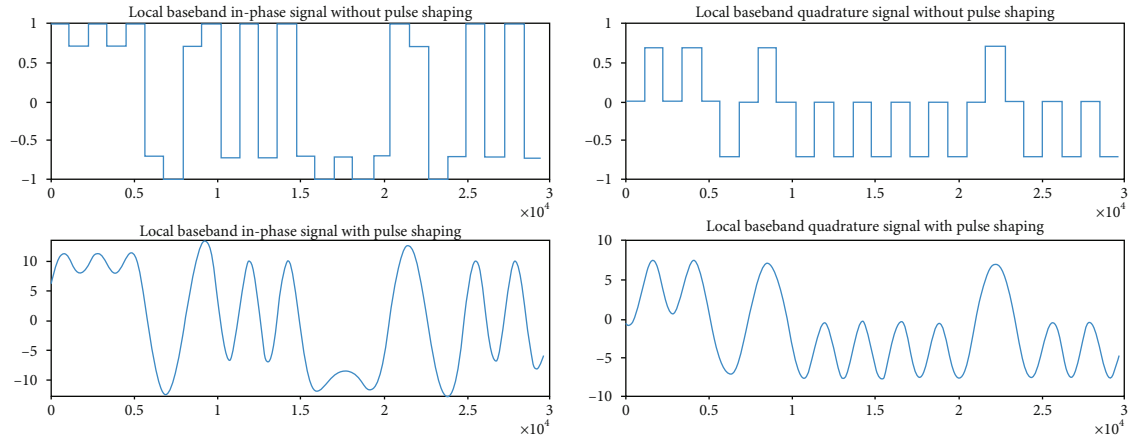


Figure 7: Local reproduced baseband of 26-symbol training sequence before and after pulse shaping.

experiment, the local reproduced digital IF signals with pulse shaping and without pulse shaping are, respectively, noncoherent correlated the same received signal to get two correlation curves at the same time. The shape of the correlation curves changed in each experiment, because the noise added to the received signal in the transmission channel was random.

Three typical experiment results are shown in Figure 8. It can be seen clearly that, if the local reproduced signals are pulse shaped, the correlation curves are obviously distorted when the SNR of the received signals are lower than -30 dB. While the local reproduced signals are not pulse shaped, the correlation curves are obviously distorted when the SNR of the received signals are lower than -20 dB. The results can verify the advantage of using pulse shaping filter for local reproduced signal in the proposed receiver structure. The results also can show the correlation property in different noise conditions.

4.2. Ranging Results and Discussion. Based on the receiver structure given in Section 3.1, the implementation of the multicorrelator pseudorange measurement method in Section 3.2 was used for the ranging experiments, where the number of the multicorrelators was $1 + 2M = 21$.

At the same time, the traditional maximum correlation value method and E-L method were also used for the ranging experiments to compare with the proposed multicorrelator method, in both conditions that the local reproduced

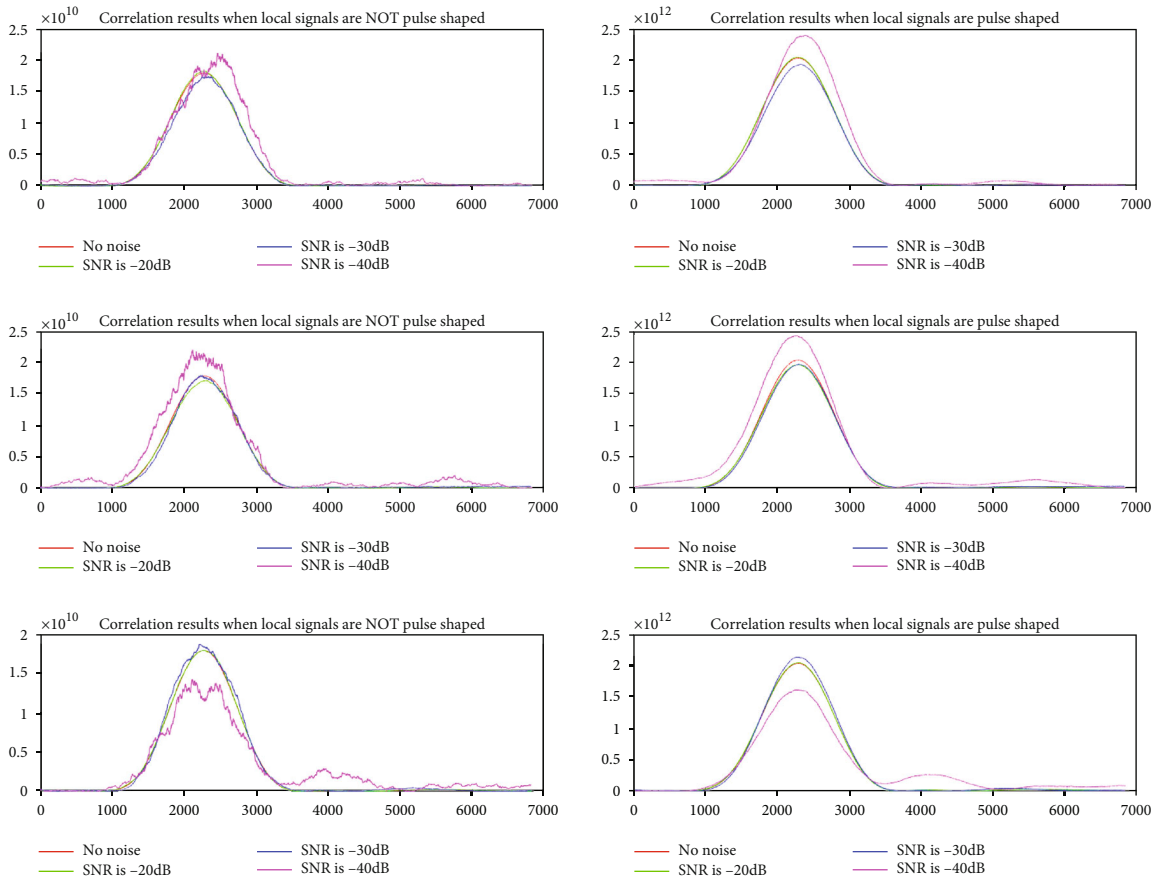


Figure 8: Correlation results of three experiments.

signal was pulse shaped and the local reproduced signal was not pulse shaped. The traditional maximum value method uses one or multiple correlators to get the correlation values in the phase range of [-1, +1] chips between the local reproduced signal and the received signal. The code phase of maximum correlation value is considered to be the synchronous phase. The traditional E-L method uses two correlators with fixed intervals. When the correlation values are equal in the phase range of [-1, +1] chips between the local reproduced signal and the received signal, the middle of the two correlators is considered to be the synchronous phase [53].

The results of ranging experiments using the three methods in the two conditions just described above are shown in Figure 9. It shows the TOA errors in different noise conditions. Figure 9(a) shows the experiment results in a big range of $-30\,\text{dB} < E_S/N_0 < 30\,\text{dB}$. In order to see clearly the experiment results in better noise condition, enlarged results in a smaller range of $-5\,\text{dB} < E_S/N_0 < 30\,\text{dB}$ are shown in Figure 9(b).

Figure 9(a) shows that when E_S/N_0 is less than -5 dB, the proposed multicorrelator method cannot get higher ranging accuracy compared with the other two methods, because the poor signal quality leads to serious distortion of correlation curve. Figure 9(a) also shows that the receiver can get higher ranging accuracy when local reproduced signal is pulse shaped.

Figure 9(b) shows when E_S/N_0 is greater than -5 dB, the proposed multicorrelator method can get higher ranging accuracy compared with the other two methods. And the performance is better when using pulse shaping filter for local reproduced signal. This is also the signal quality required by the communication function in VDES [44]. Figure 9(b) also shows that when E_S/N_0 is greater than 15 dB, the TOA error of the proposed method is less than 10 ns; that is, the accuracy of the pseudorange measurement value is better than 3 m.

According to the results, it can be concluded that when the signal quality is high, the proposed multicorrelator method can realize high accuracy ranging of the communication system. The ranging accuracy can be higher compared with the traditional maximum correlation value method and E-L method. Compared with the general VDES function, it does not occupy any additional time slot resources. This method can be used for VDES R-Mode receiver which generally works under good signal conditions. It can get the pseudorange value between the receiver and the base station. When getting more than three pseudorange values, the receiver can calculate its position. So the VDES R-Mode can be a backup land-based radio navigation system for GNSS, to reduce the impact of GNSS vulnerability, and contribute to the ship safety.

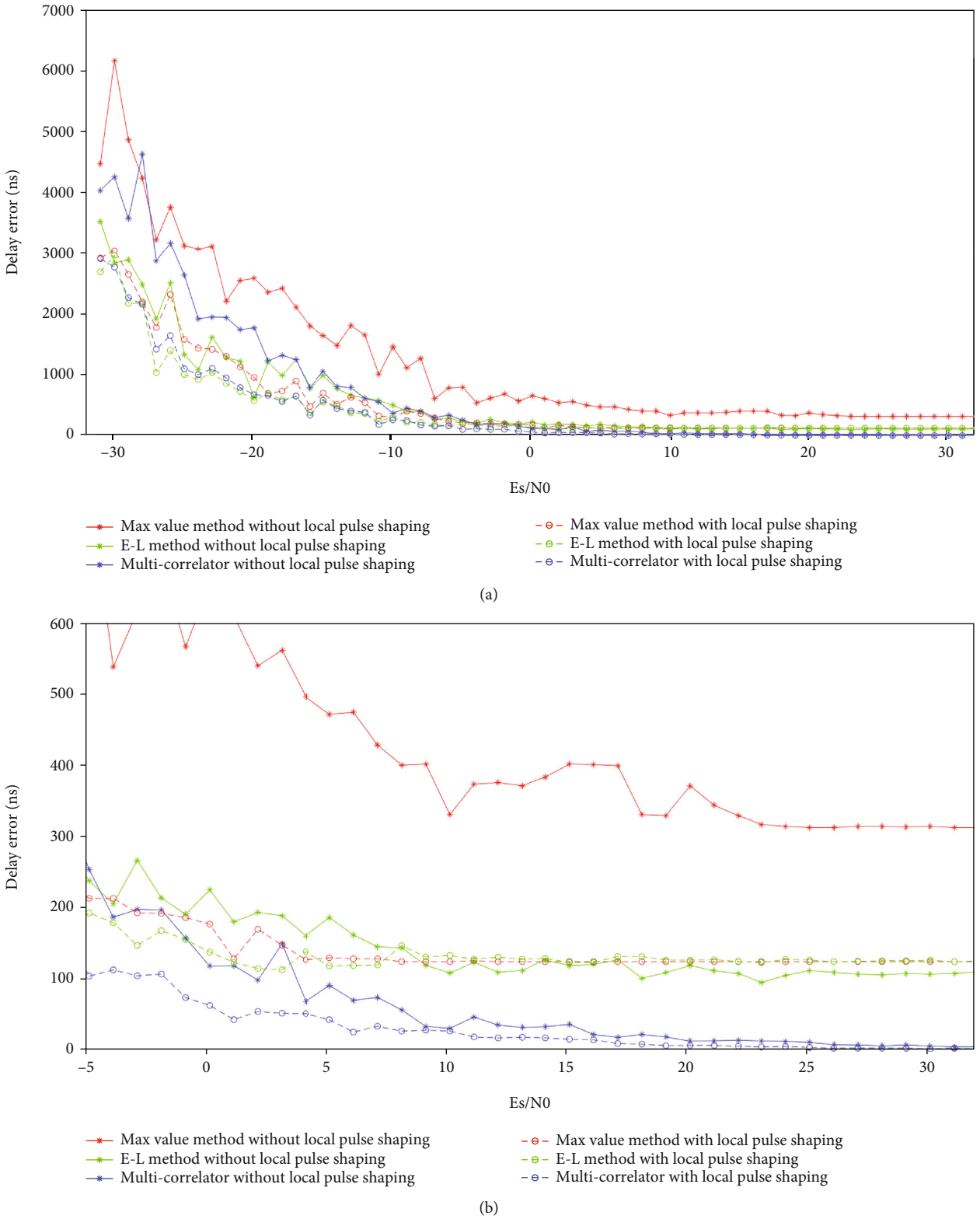


Figure 9: Experiment results of three methods: (a) TOA error when $-30\,\text{dB} < E_s/N_0 < 30\,\text{dB}$; (b) TOA error when $-5\,\text{dB} < E_s/N_0 < 30\,\text{dB}$.

5. Conclusions

This paper proposes a multicorrelator pseudorange measurement method based on the training sequence, which is normally used to capture and synchronize the communication signal. This method can be used for the R-Mode of the VDES. It uses the correlation property of the training sequence in ASM channel and VDE channel for ranging,

so it can realize ranging while communicating without occupying any additional time slot resources.

This paper also gives a receiver design based on the proposed method with a structure of multicorrelator after discussing the influence on the ranging performance of the pulse shaping required in the VDES. The baseband signal of the training sequence is reproduced in the receiver. It is then pulse shaped and becomes digital IF signal after carrier modulation. The IF signal noncoherent correlates with the received signal to get the correlation value which can be used to calculate the TOA, which can be used to get the pseudorange measurement value for positioning. The received signal is also a digital IF signal, which is the RF signal received by the antenna which goes through the downconverting and the A/D sampling in the RF circuit. To reduce the influence of the pulse shaping on the correlation curve, the pseudorange measurement method with multicorrelators is used to effectively reduce the pseudorange measurement error. The experiment results show that, compared with the traditional maximum correlation value method and E-L method, the multicorrelator pseudorange measurement method has higher ranging accuracy. When E_s/N_0 is greater than 15 dB, the ranging accuracy of the proposed method can be better than 3 m. The proposed multicorrelator pseudorange measurement method based on the training sequence can realize high accuracy ranging without occupying any additional time slot resources. It can provide a reference for the future research of the VDES R-Mode, which can be used to solve the vulnerability problem of the GNSS on maritime communication, so as to make a certain contribution to the robustness of 6G network space-time reference.

Data Availability

The data presented in this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

S.X. and H.Q. were responsible for conceptualization. S.X. was responsible for the methodology, formal analysis, writing the original draft preparation, writing, review, and editing. J.Y. and S.X. were responsible for the software. S.X., H.Q., and J.Y. were responsible for the validation. H.Q. was responsible for the investigation, resources, supervision, and project administration. J.Y. was responsible for the data curation, visualization, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the National Program on Key R&D Project of China (2021YFB3901500) and the National Natural Science Foundation of China (52071047). This research was also supported by Joint Fund of the Natural Science Foundation of Liaoning Province of China (2020-HYLH-42) and the Fundamental Research Funds for the Central Universities (No. 017190327).

References

[1] Z. Na, Y. Liu, J. Shi, C. Liu, and Z. Gao, "UAV-supported clustered NOMA for 6G-enabled Internet of Things: trajectory planning and resource allocation," *IEEE Internet of Things Journal*, vol. 8, no. 20, pp. 15041–15048, 2021.

[2] Z. Na, B. Li, X. Liu et al., "UAV-based wide-area Internet of Things: an integrated deployment architecture," *IEEE Network*, vol. 35, no. 5, pp. 122–128, 2021.

[3] M. Zhou, Y. X. Lin, N. Zhao, Q. Jiang, X. L. Yang, and Z. S. Tian, "Indoor WLAN intelligent target intrusion sensing using ray-aided generative adversarial network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 1, pp. 61–73, 2020.

[4] IALA, *Recommendation R-129 GNSS Vulnerability and Mitigation Measures, Edition 3*, IALA, Saint Germain en Laye, France, 2012.

[5] P. Grant, N. Williams, S. Ward, and S. Basker, "GPS jamming and the impact on maritime navigation," *Journal of Navigation*, vol. 62, no. 2, pp. 173–187, 2009.

[6] IALA, *World Wide Radio Navigation Plan*, IALA, Saint Germain en Laye, France, 2nd ed edition, 2012.

[7] ITU, *Recommendation ITU-R M.1371-4 Technical Characteristics for an Automatic Identification System Using Time-division Multiple Access in the VHF Maritime Mobile Band*, ITU, Geneva, Switzerland, 2010.

[8] S. L. Eun, J. M. Amit, Y. M. Sang, and S. K. Geun, "The Maturity of automatic identification systems (AIS) and its implications for innovation," *Journal of Marine Science and Engineering*, vol. 7, no. 9, p. 287, 2019.

[9] ITU, *Automatic Identification System VHF Data Link Loading, Report ITU-R M.2287-0*, ITU, Geneva, Switzerland, 2013.

[10] J. Šafár, C. Hargreaves, and N. Ward, "The VHF data exchange system," in *Antennas, Propagation RF Technology for Transport and Autonomous Platforms*, Curran Associates, Inc., Red Hook, NY, USA, 2017.

[11] ITU, *Recommendation ITU-R M.2092-0 Technical Characteristics for a VHF Data Exchange System in the VHF Maritime Mobile Band*, ITU, Geneva, Switzerland, 2015.

[12] F. Clazzer, A. Munari, and F. Giorgi, "OCEANS 2017-Aberdeen," in *Asynchronous random access schemes for the VDES satellite uplink*, IEEE, Aberdeen, UK, 2017.

[13] J. Yi, Z. Yang, and W. Junsen, "A novel random access algorithm for very high frequency data exchange (VDE)," *Journal of Marine Science and Engineering*, vol. 8, no. 2, p. 83, 2020.

[14] K. M. Kim, Y. Kim, Y. Cho et al., "Performance evaluation of maritime VDES networks with OPNET simulator," in *2018 11th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, Budapest, Hungary, 2018.

[15] IMO, *Resolution MSC.401 (95) Performance Standards for Multi-System Shipborne Radionavigation Receivers*, IMO, London, UK, 2015.

[16] G. M. Huang, T. Jing, and W. Tian, "Survey on navigation via signal of opportunity," *Control and Decision*, vol. 34, no. 6, pp. 1121–1131, 2019.

[17] T. O. Mansfield, B. V. Ghita, and M. A. Ambroze, "Signals of opportunity geolocation methods for urban and indoor environments," *Annals of Telecommunications*, vol. 72, no. 3-4, pp. 145–155, 2017.

[18] G. G. Seco, S. J. López, and B. D. Jiménez, "Challenges in indoor global navigation satellite systems: unveiling its core features in signal processing," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 108–131, 2012.

[19] F. Coluccia, G. Ricciato, and G. Ricci, "Positioning based on signals of opportunity," *IEEE Communications Letters*, vol. 18, no. 2, pp. 356–359, 2014.

[20] S. Dammann and R. R. Sand, "Signals of opportunity in mobile radio positioning," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 549–553, Bucharest, Romania, 2012.

[21] V. Otsason, A. Varshavsky, and A. Lamarca, "Accurate GSM indoor localization," in *International conference on ubiquitous computing*, Springer Heidelberg, Berlin, 2005.

[22] M. Borenovic, A. Neskovic, and N. Neskovic, "Vehicle positioning using GSM and cascade-connected ANN structures," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 34–46, 2013.

[23] M. A. Enright, H. Sridhara, and T. Nguyen, "Advanced GNSS integrity using signals of opportunity," in *Proceedings of the International Technical Meeting of the Institute of Navigation*, pp. 83–89, Newport Beach, CA, 2012.

[24] N. Yang and D. Q. Thao, "Positioning with mixed signals of opportunity," in *Institute of Navigation Satellite Division Proceedings of the International Technical Meeting*, pp. 447–456, Portland, OR, 2011.

[25] L. Chen, L. L. Yang, J. Yan, and R. Chen, "Joint wireless positioning and emitter identification in DVB-T single frequency networks," *IEEE Trans on Broadcasting*, vol. 63, no. 3, pp. 577–582, 2017.

[26] T. A. Webb, P. D. Groves, and P. A. Cross, "A new differential positioning method using modulation correlation of signals of opportunity," in *IEEE/ION Position, Location and Navigation Symposium*, pp. 972–981, New York, 2010.

[27] H. Simkovits, A. J. Weiss, and A. Amar, "Navigation by inertial device and signals of opportunity," *Signal Processing*, vol. 131, no. 2, pp. 280–287, 2017.

[28] J. J. Morales and Z. M. Kassas, "Distributed signals of opportunity aided inertial navigation with intermittent communication," in *Institute of Navigation Satellite Division Proceedings of the International Technical Meeting*, pp. 2519–2530, Portland, Oregon, 2017.

[29] K. M. Pesyna, K. D. Wesson, and R. W. Heath, "Extending the reach of GPS-assisted femtocell synchronization and localization through tightly-coupled opportunistic navigation," in *IEEE Globecom Workshops*, pp. 242–247, Houston, 2011.

[30] G. W. Johnson and P. F. Swaszek, "Feasibility study of R-mode using MF DGPS transmissions," *German Federal Waterways and Shipping Administration, Milestone 2 Report*, IALA, Saint Germain en Laye, France, 2014.

[31] G. W. Johnson and P. F. Swaszek, "Feasibility study of R-mode using AIS transmissions," *German Federal Waterways and Shipping Administration, Milestone 4 Report*, IALA, Saint Germain en Laye, France, 2014.

[32] G. W. Johnson and P. F. Swaszek, "The feasibility of R-mode to meet resilient PNT requirements for e-navigation," in *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+2014)*, pp. 3076–3100, Tampa, Florida, 2014.

[33] J. Šafár, A. Grant, P. Williams, and N. Ward, "Performance bounds for VDES R-mode," *The Journal of Navigation*, vol. 73, no. 1, pp. 92–114, 2020.

[34] Q. Hu, Y. Jiang, J. Zhang, X. Sun, and S. Zhang, "Development of an automatic identification system autonomous positioning system," *Sensors*, vol. 15, no. 11, pp. 28574–28591, 2015.

[35] Y. Jiang, J. Wu, and S. Zhang, "An improved positioning method for two base stations in AIS," *Sensors*, vol. 18, no. 4, 2018.

[36] H. Qing, J. Xiaoyue, and L. Pengfei, "A novel carrier frequency offset algorithm based on a double Barker code in VDE-TER," *Physical Communication*, vol. 40, p. 101059, 2020.

[37] S. Gewies, A. Dammann, R. Ziebold et al., "R-Mode testbed in the Baltic Sea," in *Proceedings of the 19th IALA Conference*, Incheon, Korea, 2018.

[38] M. Hoppe, A. Grant, C. Hargreaves, and P. Williams, "R-Mode: the story so far,," in *Proceedings of the 19th IALA Conference*, Incheon, Korea, 2018.

[39] K. Paul and G. Stefan, "Worldwide availability of maritime medium-frequency radio infrastructure for R-Mode-supported navigation," *Journal of Marine Science and Engineering*, vol. 8, no. 3, p. 209, 2020.

[40] J. Zhang, S. Zhang, and J. Wang, "Pseudorange measurement method based on AIS signals," *Sensors*, vol. 17, no. 5, p. 1183, 2017.

[41] M. Wirsing, A. Dammann, and R. Raulefs, "Investigating R-Mode signals for the VDE system," in *OCEANS 2019 MTS/IEEE SEATTLE*, pp. 1–5, Seattle, WA, USA, 2019.

[42] IALA, *Guideline G1139-The Technical Specification of VDES, Edition 3*, 2019.

[43] F. Lázaro, R. Raulefs, H. Bartz, and T. Jerkovits, "VDES R-Mode: vulnerability analysis and mitigation concepts," *International Journal of Satellite Communications and Networking*, 2021.

[44] J. Šafář, A. Grant, and M. Bransby, "Performance bounds for VDE-SAT R-mode," *International Journal of Satellite Communication Network*, 2021.

[45] X. M. She and D. Weiwei, "Viterbi detection method of $\pi/4$-QPSK signal in VDE," *Procedia Computer Science*, vol. 107, pp. 539–544, 2017.

[46] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Transactions on Communications*, vol. 38, no. 3, pp. 300–308, 1990.

[47] S. R. Theodore and Q. M. Meng, *Wireless Communications Principles and Practice*, Publisher: Publishing House of Electronics Industry Beijing, China, 2nd ed edition, 2018.

[48] C. Sandeep and G. J. Saulnier, "Differential detection of pi /4-shifted-DQPSK for digital cellular radio," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 1, pp. 46–57, 1993.

[49] P. Jiang and X. Hou, "$\Pi/4$-DQPSK demodulation based on DSTFT," *Journal of Telemetry, Tracking and Command*, vol. 28, no. 1, pp. 1–4, 2007.

[50] M. Y. Zhang and Y. Z. Zhang, "An integrated demodulation technology of variable signal of $\pi/4$-DQPSK and GMSK modulation," *Journal of Radio Engineering*, vol. 45, no. 2, pp. 30–33, 2015.

[51] Y. T. Ma, K. Pahlavan, and Y. S. Geng, "Comparison of POA and TOA based ranging behavior for RFID application," in *Proceedings of the 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication*, pp. 1722–1726, Washington, DC, USA, 2014.

[52] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the AIEE*, vol. 47, no. 2, pp. 617–644, 1928.

[53] D. Kaplan, *Understanding GPS: Principles and Applications, Second Edition, Norwood*, Artech House, MA, 2006.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3612073, 11 pages
https://doi.org/10.1155/2022/3612073

Research Article

Research on Intelligent Predictive Analysis System Based on Embedded Wireless Communication Network

Jingwei Sun

Hefei University of Technology, Hefei 230001, China

Correspondence should be addressed to Jingwei Sun; 2014010059@mail.hfut.edu.cn

Received 9 December 2021; Revised 18 January 2022; Accepted 9 February 2022; Published 27 February 2022

Academic Editor: Mu Zhou

In order to improve the practical effect of intelligent prediction system, by analyzing and studying the characteristics of the traffic carried on the network, it can provide an effective way to explore the internal operating mechanism of the network. This paper takes the embedded wireless communication network as the research object to construct an intelligent predictive analysis system and applies the intelligent regression algorithm to the intelligent predictive analysis to construct an intelligent predictive analysis system. Finally, this paper verifies the system model of this paper through experimental research. The research results show that the intelligent predictive analysis system based on the embedded wireless communication network proposed in this paper is very effective and has a positive effect on the construction and development of the embedded wireless communication network.

1. Introduction

Wireless local area network is the product of the combination of computer network and wireless communication technology. Specifically, traditional cables are no longer used when setting up a local area network but are connected wirelessly using infrared rays, radio waves, etc. as the transmission medium to provide all the functions of a wired local area network [1]. The basis of the wireless local area network is the traditional wired local area network, which is the expansion and replacement of the wired local area network. It realizes wireless communication through wireless hubs, wireless access nodes, wireless network bridges, wireless network cards, and other devices on the basis of the wired local area network. Currently, the frequency band used by wireless local area networks is mainly S-band (2.4 GHz-2.4835 GHz). The networking mode of wireless local area network can be roughly divided into two types, one is ad-hoc mode, that is, point-to-point wireless network, and the other is infrastructure mode, that is, centralized control network [2]. The wireless local area network can make up for the deficiencies of wired Ethernet that rely on cables or optical cables in some special application environments and realize the extension of the network. Embedded system integration of wireless local area network technology to achieve wireless communication and data transmission will become a hot spot for future applications. For example, wireless digital set-top boxes, computers, wireless gateways, and household appliances can form a home wireless local area network, and at the same time can be connected to the Internet through an AP, a wireless router, or a wireless bridge. The wireless instrument performs data collection and wireless transmission. Wireless instruments and equipment are arranged in the work site with an ad hoc network for mutual information transmission and remote wireless monitoring, which reduces the trouble and inconvenience of wiring, greatly improves industrial production efficiency and facilitates people's lives [3].

With the further development of embedded operating systems and wireless communication technologies, data terminals relying on wireless network data transmission based on embedded operating systems have become more and more widely used. Embedded operating system is a hot research topic nowadays, and embedded Linux is stable, efficient, easy to customize, easy to cut, extensive hardware support, free, open source, and other characteristics, which

makes Linux widely used in the embedded field. In recent years, the development of global communication technology has been changing with each passing day. Especially in the past two to three years, the development speed and application fields of wireless communication technology have surpassed that of fixed communication technology, showing a trend of development in full swing. The most representative ones are cellular mobile communications, broadband wireless access, as well as trunking communications, satellite communications, and mobile video services and technologies.

Due to the limitation of limited communication and sampling period, the current distributed prediction methods for multiagent systems are mostly noniterative and cooperative algorithms. In each sampling period, there is only one information exchange between the agents, and only local performance indicators are calculated.

This paper studies the application of intelligent network and the behavior characteristics of the network itself and obtains parameter information through the study of network traffic. By analyzing and studying the traffic characteristics carried on the network, combined with the embedded wireless communication network, the intelligent prediction and analysis system is constructed, the intelligent prediction and regression algorithm is improved, the operation effect of the intelligent prediction and analysis system is improved, and the practical effect of the intelligent prediction system is improved.

This article takes the embedded wireless communication network as the research object to construct the intelligent predictive analysis system and verify and analyze its performance, which provides a theoretical reference for the further development of the subsequent wireless communication network technology.

2. Related Work

The modeling research on networked predictive control is as follows:

Literature [4] proposed the concept of networked predictive control, which actively compensates for the communication delay and data packet loss in the control loop through iterative prediction. In this paper, a networked predictive control strategy is designed for a class of linear time-invariant systems to solve the steady time delay and the random communication time delay, respectively. Literature [5] studied the problem of networked predictive control in the presence of random network delays in the feedback channel for a type of multiple-input multiple-output discrete-time general system, carried out numerical simulation research, and based on the NetCon-ARM9 embedded system platform. Verify the effectiveness of the above control strategy. For a type of linear system described by a discrete-time model, some scholars have studied the networked predictive control problem in the case of communication delays in both the forward channel and the feedback channel [6]. Literature [7] gives the necessary and sufficient conditions to ensure the stability of the closed-loop system when the communication delays are constant. At the same time, when the random delay in the communication channel is bounded, if the corresponding switching system is stable, then the obtained closed-loop networked predictive control system is also stable. Simulations and experiments have verified the above conclusions. Literature [8] considered the predictive control problem in the case of disturbances in both the forward channel and the feedback channel of the system model. Literature [9] established a model based on the Markov chain for the communication delay in the two-way channel, using a local Lomberg observer to estimate the state vector that could not be measured, and the measurement output and the estimated state vector through the limited communication capacity. The communication channel is sent to the tracking controller, which solves the problem of networked predictive tracking control. When the random communication delay exists in the control loop of the networked control system, literature [10] proposed an output feedback predictive controller to actively compensate for the communication delay and proved the stability of the closed-loop networked predictive control system.

The research on network predictive control method and data processing is as follows:

In the above research results of the networked predictive control method, the controller is generally required to calculate the control value at each time in the future and then package it and send it to the controlled object via the communication network to actively compensate for various possible time delays. At the end of the controlled object, the most appropriate control value is selected based on the measured delay size, so as to realize the delay compensation in the forward channel [11]. However, this method needs to know the accurate mathematical model of the controlled object in advance and needs to measure the accurate communication time lag, which is difficult to achieve for most applications. Literature [11] proposed a data-driven networked predictive control method for a type of single-input single-output discrete-time autoregressive moving average model. This method does not need to accurately measure the delay in the communication network, and allows the existence of uncertain parameters of the system model.

The related research on the construction of intelligent prediction system for wireless communication network is as follows:

In the past, the research on traffic forecasting used a single model to describe the original characteristics of traffic. However, there are more and more uncertain factors affecting network traffic characteristics (such as new applications and protocol structure changes). When a single model is used for forecasting, large errors will inevitably occur [12]. Therefore, a single prediction model, such as Markov model, ARIMA model, and mock-MA model, cannot predict network traffic well because it only describes some of the characteristics of traffic. Although a single neural network model and support vector machine model can describe the characteristics of network traffic more comprehensively, their prediction accuracy for complex network traffic is still not satisfactory. According to the fact that there are multiple characteristics of network traffic, many scholars use different models to describe the corresponding characteristics and then combine them to predict network traffic and achieve better prediction results than a single model. Literature

[13] has no extraction in Haar. On the basis of wavelet transform, combined with adaptive AR model and sliding window polynomial fitting method, a recursive high-speed network traffic online prediction model based on wavelet transform is established. This model not only improves the accuracy of online traffic prediction but also avoids regular estimation and update of parameters through the recursive automatic adjustment of model parameters. The literature [14] combines wavelet analysis and Kalman filtering and uses Kalman filtering to deal with the linear change part of network traffic, use wavelet analysis to deal with the nonlinear change part, and simulation results show that the model has high prediction accuracy; literature [15] uses the FARtMA model to describe the long correlation and short correlation, and the neural network to describe the nonstationary. Finally, the results of the two models are optimized and combined; literature [16] combines wavelet analysis and neural network to establish a prediction model. First, wavelet decomposition is used to decompose the network traffic data into wavelet coefficients and scale coefficients and the coefficients of these different frequency components. A single branch is reconstructed into high-frequency components and low-frequency components. The FIR neural network is used to predict these components separately, and the synthesized result is used as the prediction of the original network traffic, which has achieved good results; literature [17] uses the wavelet method to predict. The network traffic is preprocessed, and then the linear neural network and the Elman neural network are used to make predictions, respectively, to ensure that the correlation and nonstationarity of the traffic can be described. Finally, the two prediction results are synthesized into the final prediction result through the BP neural network. It shows that the combined model has higher prediction accuracy than the single model.

Literature [18] proposes a novel heuristic to reconstruct application-layer messages in the common case of encrypted traffic. We discuss and experimentally evaluate the suitability of the provided modeling approaches for different tasks.

Literature [19] investigates and specializes a set of architectures selected among convolutional, recurrent, and composite neural networks, to predict mobile-app traffic at the finest (packet-level) granularity.

In order to overcome the problems existing in the intelligent predictive analysis system, this paper combines the embedded wireless communication network to improve the effect of the intelligent predictive analysis system. The first part describes the current situation and background and summarizes the existing problems and the research content of this paper; the second part is the literature review part, which analyzes the research of experts and scholars on related issues. Therefore, the organizational structure of this paper, the third part is the algorithm improvement part, which applies the intelligent regression algorithm to the intelligent prediction analysis, and builds an intelligent prediction analysis system; in the fourth part, the intelligent predictive analysis system is improved and constructed, and the effect of the system is verified by experiments. Finally, the research content of this paper is summarized and prospected.

3. Intelligent Predictive Regression Algorithm

We set the regression function as $m(x) \in C[a, b]$ and assume that $\{\varphi_i\}_{i=0}^{\infty}$ constitutes a set of orthogonal basis on $[a, b]$, namely,

$$\int_a^b \varphi_i(x)\varphi_j(x)dx = \delta_{ij} = \begin{cases} 0, i \neq j, \\ c_i, i = j. \end{cases} \tag{1}$$

Because of $\int \prod_{i=1}^{d} \varphi_{ij}(x_i) \prod_{i=1}^{d} \varphi_{ik}(x_i)dx_1 \cdots dx_d = \prod_{i=1}^{d} \int \varphi_{ij}(x_i)\varphi_{ik}(x_i)dx_i = \prod_{i=1}^{d} \delta_{jk} = \delta_{jk}$, $\{\prod_{i=1}^{d} \varphi_{ij}(x_i)\}_{i=1}^{\infty}$ constitutes a set of orthogonal bases on $\prod_{i=1}^{d}[a_i, b_i]$, then $m(x)$ has an orthogonal sequence expansion $m(x) = \sum_{i=1}^{\infty} \theta_i \prod_{k=1}^{d} \varphi_{ki}(x)$. Therefore, the nonparametric regression model can be approximated as

$$Y_i = \sum_{j=1}^{m} \theta_i \prod_{k=1}^{d} \varphi_{ki}(x_{ki}) + v_i. \tag{2}$$

By performing least squares estimation on the model, we get

$$\widehat{\theta} = \left(Z^T Z\right)^{-1} Z^T Y. \tag{3}$$

Among them, $Z = (Z_1, \cdots, Z_m)$, $Z_i = \left(\prod_{j=1}^{d} \varphi_{ji}(X_{i1}), \cdots, \prod_{j=1}^{d} \varphi_{ji}(X_{\text{in}})\right)^T$.

Then, $m(x)$ has an orthogonal sequence estimate [20]:

$$\widehat{m}_n(x) = z(x)^T \widehat{\theta}. \tag{4}$$

Among them, $z(x) = \left(\prod_{i=1}^{d} \varphi_{i1}(x_i), \cdots, \prod_{i=1}^{d} \varphi_{im}(x_i)\right)^T$.

We set $v(x) = \sigma_n^2 (z(x)^T (Z^T Z)^{-1} z(x))$. When $n \longrightarrow \infty$, $m \longrightarrow \infty$, the orthogonal sequence estimation has the following properties:

(1) $v(x)^{-1/2}(\widehat{m}_n(x) - E\widehat{m}_n(x)) \underline{d} N(0, 1)$

(2) $v(x)^{-1/2}(E\widehat{m}_n(x) - m) \longrightarrow 0$

(3) $\sigma_n^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \widehat{m}_n(X_i))^2$ is a consistent estimate of σ_n^2

In orthogonal sequence estimation, the Legendre polynomial orthogonal basis is often selected, and its representation is as follows:

$$\begin{aligned} P_0(x) &= \frac{1}{\sqrt{2}}, \\ P_1(x) &= \frac{x}{\sqrt{2/3}}. \end{aligned} \tag{5}$$

Other high-order Legendre polynomials can be recursively obtained by the following formula:

$$(m+1)P_{m+1}(x) = (2m+1)xP_m(x) - mP_{m-1}(x). \quad (6)$$

The orthogonal basis $\{P_i(x)\}_{i=0}^{\infty}$ of Legendre polynomial satisfies

$$\int_1 P_i(x)P_j(x)dx = \delta_{ij} = \begin{cases} 0, i \neq j, \\ 1, i = j. \end{cases} \quad (7)$$

If the independent variable X takes a value in the interval $[a, b]$, the variable $Z = 2X - a - b/b - a$ must be replaced so that the value interval of the variable Z is $[-1, 1]$.

Orthogonal sequence estimation will also select Fourier orthogonal basis, which is defined as follows:

$$q_1(x) = 1, \quad q_{2k}(x) = \sqrt{2}\cos(2\pi kx), \quad (8)$$

$$q_{2k+1}(x) = \sqrt{2}\sin(2\pi kx)(k = 1, 2, \cdots). \quad (9)$$

Fourier orthogonal basis $\{q_i(x)\}_{i=1}^{\infty}$ satisfies

$$\int_b^1 q_i(x)q_j(x)dx = \begin{cases} 0, i \neq j, \\ 1, i = j. \end{cases} \quad (10)$$

Similarly, if the independent variable x takes a value in the interval $[a, b]$, the variable must be replaced with $Z = X - a/b - a$ so that the value interval of the variable Z is [0, 1] [21].

In practical problems, the linear relationship between variables will in most cases change with another covariate (such as time, temperature, etc.). Since the data information when the next covariate changes are unknown, a question arises: how to use the known historical data to estimate the parameters of the linear model when the next covariate changes.

Regarding the parameter estimation of the variable coefficient regression model, the local weighted least squares method can be used to introduce the concept of variable coefficient weighted estimable function to construct parameter estimators. On the basis of the local weighted least squares method, this paper combines the relevant content in nonparametric regression to estimate the variable coefficients by weighted least squares.

Nonparametric regression models can be divided into two categories, complete nonparametric regression models (referred to as non-parametric regression models) and semiparametric regression models. The variable coefficient regression model is a special case in the semiparametric regression model.

We assume that the independent variable $x_1, x_2, \cdots, x_p$ and the dependent variable y satisfy a linear relationship at the parameter t.

$$y = \beta_0(t) + \beta_1(t)x_1 + \cdots \beta_p(t)x_p. \quad (11)$$

Among them, $\beta_i(t), i = 0, 1, 2, \cdots, p$ is a bounded continuous function of a one-dimensional (or multidimensional) real variable b and has a continuous derivative, and $\beta_i(t)$ is called a variable coefficient.

We assume that $t_1, t_2, t_3, \cdots, t_n$ is n points of t_0 near a specified point, and the sample observation value $(y_i, t_i, x_i, \cdots, x_{ip})$ is obtained by observing at each point.

$$y_i = \beta_0(t_i) + \beta_1(t_i)x_{i1} + \cdots \beta_p(t_i)x_{ip} + \varepsilon_i. \quad (12)$$

Among them, $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n, E\varepsilon_i = 0, \mathrm{Var}\varepsilon_i = \sigma^2, i = 1, 2, \cdots, n$.

The variable coefficient regression model uses the observation value $(y_i, t_i, x_{i1}, \cdots, x_{ip})$ at the point t_i near t_0 to estimate the parameter $\beta_i(t)$ at t_0. Since the observations at different t_i have different "importance" relative to t_0, the weight function $w_i(t_o)$ needs to be used to measure.

This paper uses the distance between the observation value at t_i and t_0 to define the relationship between them, namely,

$$w_i(t_o) = w(p(v_i, v_o)). \quad (13)$$

Among them, $p(v_i, v_o)$ represents the Euclidean distance from v_i to v_o, $i = 1, 2, \cdots, n$.

The most commonly used weight function estimation is kernel estimation and nearest neighbor estimation.

The probability density function $K(\cdot)$ with symmetric origin is selected:

$$\int K(u)du = 1, \quad (14)$$

is the kernel function and window width $h_n > 0$. The nuclear weight function is defined as

$$W_{ni}(x) = \frac{K_{h_n}(X_i - x)}{\sum_{j=1}^{n} K_{h_2}(X_j - x)}. \quad (15)$$

Among them, $K_{h_n}(u) = h_n^{-1}K(uh_n^{-1})$ is also a probability density function, and the parameter h_n is called the window width.

The Nadaraya-Watson kernel estimate is defined as

$$\widehat{m}_n(x) = \sum_{i=1}^{n} W_{ni}(x)Y_i. \quad (16)$$

The K-nearest neighbor non-parametric estimation algorithm is mainly composed of four parts: database generation, near neighbor sample subset search, near neighbor subset optimization, and forecast quantity estimation. The data flow relationship can be seen in Figure 1.

Four steps of the K-Nearest Neighbor Nonparametric Regression Prediction Algorithm.

(1) The more historical data, the more conducive the nonparametric regression estimation is to more truly and completely express the characteristics of the system state, and the more conducive to obtaining

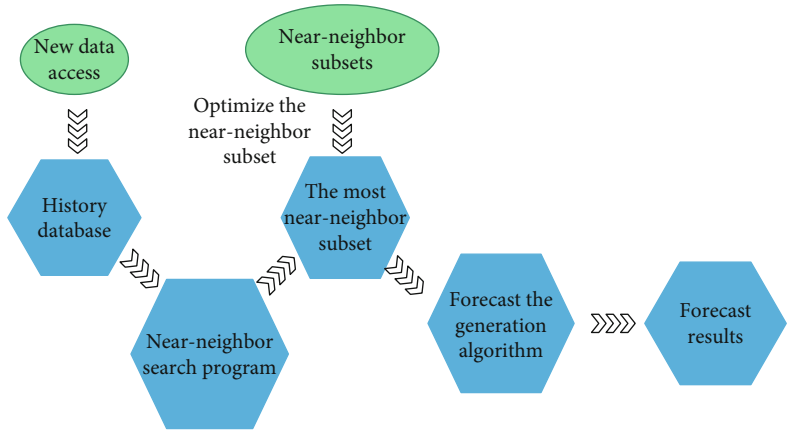


Figure 1: Schematic diagram of K-nearest neighbor nonparametric estimation algorithm.

accurate forecast estimates. Moreover, the historical database should be dynamic. As long as there is new data, it will be put into the database and the database data will be updated continuously

(2) The process of searching for neighbors is to find historical records similar to the current state condition characteristics in the historical database according to the predefined similarity measure and mark the searched historical records with similar characteristics as a neighbor. All the searched neighbors form a subset of neighbors

Step 1. Put the observed historical data into the database to build a complete historical database.

Step 2. Define a state vector that meets the requirements.

Step 3. Select the appropriate distance measurement method to determine the nearest neighbor search rules.

Step 4. Determine the appropriate K value and select K nearest neighbors in the historical database.

Step 5. Input the K-nearest neighbors determined in Step 4 into the prediction algorithm to obtain the predicted value.

The optimal neighbor subset refers to the subset formed by the neighbors that contribute the most to the forecast estimate among all the neighbors obtained by the search. Generally, it is determined by control parameters and optimization indexes. The control parameter of the K-nearest neighbor nonparametric estimation is the sample size of the optimal nearest neighbor subset. The optimization index can be similar to the index in the parameter estimation model. At present, the minimum prediction error sum of squares criterion is frequently used.

(3) After the optimal nearest neighbor subset is determined, the optimal subset can be used to estimate the production forecast. The state feature vector has been divided into condition feature vector and forecast feature vector in the database. In this way, each neighbor in the optimized neighbor subset has a corresponding predictor feature vector, and these predictor feature vectors can be regarded as the output corresponding to the conditional feature vector of each neighbor. The latest input conditional feature vector has a certain "distance" from each neighbor in the nearest neighbor subset. Therefore, it is necessary to synthesize the output vectors in all the best nearest neighbor subsets to get the most likely output vector corresponding to the conditional feature vector to be predicted. It is generally assumed that such a comprehensive operator is a linear operator or a random operator, and it can also be assumed to be a nonlinear operator. The most commonly used is the arithmetic average operator or the weighted average operator

The core problem of the K-nearest neighbor nonparametric estimation algorithm is the determination of the weight function of the nearest neighbor subset. By referring to the method of determining the weight function in the kernel weight estimation, this paper gives the following K-nearest neighbor kernel weight estimation model.

We set $1 < k < n$, and $J_{x,k} = \{i : X_i$ is one of the k nearest predicted values to $x\}$. Moreover, we combine the relevant theories in the kernel weight estimation to obtain the K-nearest neighbor kernel weight estimate of the nonparametric regression model as

$$\hat{m}_n(x,k) = \frac{\sum_{i=1}^{n} K((X_i - x)/R(x,k))Y_i}{\sum_{i=1}^{n} K((X_i - x)/R(x,k))}. \tag{17}$$

Among them, the function $K(\cdot)$ is the kernel function in the kernel weight estimation, and the similarity measure between the data is defined by the Euclidean distance, namely,

$$R(x) = \max\left\{\left[(z-x)^T(z-x)\right]^{1/2} : z \in J_x\right\}. \tag{18}$$

From equation (1), it can be seen that the K-nearest neighbor kernel weight estimation is the weighted average of the k observations closest to x.

The commonly used kernel function for K-nearest neighbor kernel weight estimation is $K(u) = d(d+2)/2S_d (1-u_1^2-\cdots-u_d^2)_+$, where $S_d = 2\pi^{d/2}/\Gamma(d/2)$. The K-nearest neighbor nuclear weight using a one-element kernel is estimated as

$$\widehat{m}_n(x) = \frac{\sum_{i=1}^{n} K\left(R(x)^{-1}\left[(X_i - x)^T (X_i - x)\right]^{1/2}\right) Y_i}{\sum_{i=1}^{n} K\left(R(x)^{-1}\left[(X_i - x)^T (X_i - x)\right]^{1/2}\right)}. \quad (19)$$

The commonly used kernel functions are triangular kernel function $(1-|u|)_+$, parabolic kernel function $0.75(1-u^2)_+$, fourth power kernel function $15/16((1-|u|^2)_+)^2$, and sixth power kernel function $70/81((1-|u|^3)_+)^2$.

(i) In addition to retaining the characteristics of nuclear power estimation, the nearest neighbor nuclear power estimation also has properties such as consistency and asymptotic normality under appropriate conditions. The convergence speed at the interior point can reach $O(n^{-2kd+4})$

From this, the observation value at t_i can be estimated to the k-nearest neighbor kernel weight at t_o as

$$W_i(t_0) = W(\rho(v_i, v_0)) = \frac{K\left(R(v)^{-1}\rho(v_i, v_0)\right)}{\sum_{i=1}^{n} K\left(R(v)^{-1}\rho(v_i, v_0)\right)}. \quad (20)$$

Among them, $R(v) = \max\{\rho(v_i, v_0): v_i \in J_v\}$.

According to regression estimation theory, when the error term of the established regression model has heteroscedasticity, the parameter estimation obtained by the ordinary least squares regression method is biased.

$\widehat{\beta}(t_0)$ is the weighted least squares estimate of $\beta(t_0)$. If $\widehat{\beta}(t_0)$ satisfies

$$Q\left(\widehat{\beta}(t_0)\right) = \min_{\beta(t_0)} Q(\beta(t_0)). \quad (21)$$

Among them,

$$\begin{aligned}(\beta(t_0)) &= \sum_{i=1}^{n} W_i(t_0)\left[y_i - \beta_0(t_i) - \beta_1(t_i)x_{i1} - \cdots - \beta_p(t_i)x_{ip}\right]^2 \\ &= (Y - X\beta(t_0))^T W(t_0)(Y - X\beta(t_0)).\end{aligned} \quad (22)$$

From

$$\frac{\partial Q(\beta(t_0))}{\partial \beta(t_0)} = 2X^T W(t_0)Y - 2X^r W(t_0)X\beta(t_0) = 0. \quad (23)$$

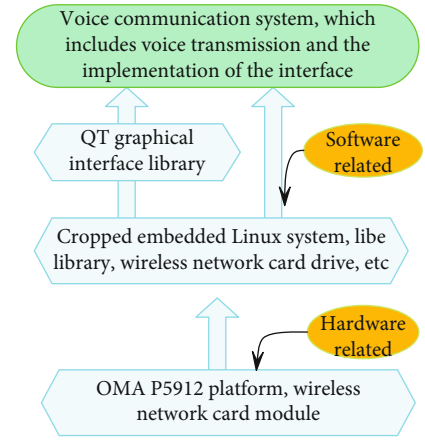


Figure 2: The overall framework of the system.

We obtain

$$X^T W(t_0)X\beta(t_0) = X^T W(t_0)Y. \quad (24)$$

When $X^T W(t_0)X$ is reversible, we obtain

$$\widehat{\beta}(t_0) = \left(X^T W(t_0)X\right)^{-1} X^T W(t_0)Y. \quad (25)$$

4. Intelligent Predictive Analysis System Based on Embedded Wireless Communication Network

The wireless voice communication system implemented in this paper uses an embedded design model. After refinement, the entire system must complete the design of the following modules. The overall system framework is shown in Figure 2.

The overall design process of the system is divided into a system migration part, a software development part, and a graphical interface development part. The system migration part is to build an environment for the software part, including transplanting the Linux kernel and so on. The software part is to complete the voice transmission and reception, voice conversation management functions, and so on. The overall design process of the system is shown in Figure 3 below.

In order to assist the upper layer protocol to select a better link for data transmission and effectively improve the transmission efficiency of the network, real-time and accurate link quality prediction is required. Existing research shows that although PRR is the most common and direct link quality indicator for link quality evaluation, the PRR in a small-time window cannot accurately reflect the link quality, and long-term statistics are needed to obtain a more accurate PRR estimate. Therefore, the agility of directly using PRR for link quality prediction is usually very poor. Existing link quality prediction methods usually predict physical layer parameters such as RSSI, LQI, and SNR, and then evaluate the link quality based on the mapping model between the corresponding physical layer parameters and

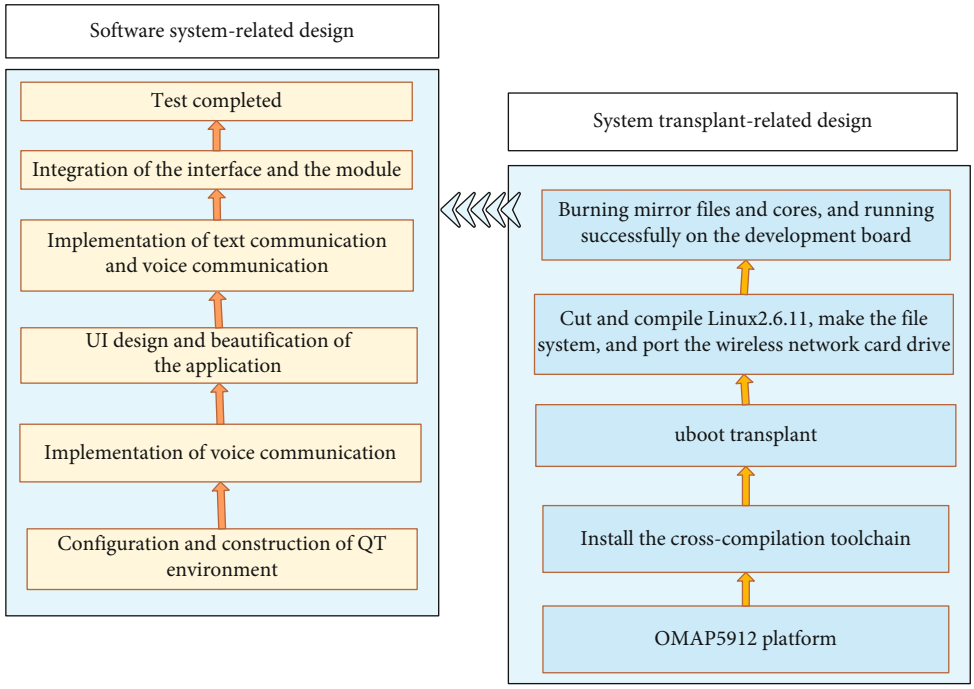


Figure 3: Flow chart of overall system design.

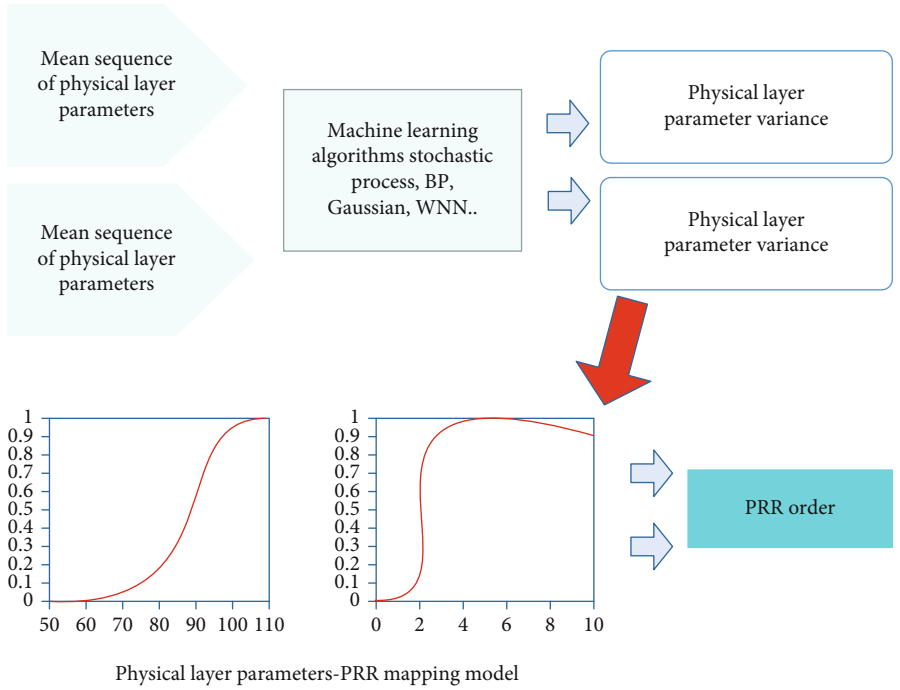


Figure 4: Traditional link quality prediction method model.

PRR. This can effectively solve the problem of poor agility in directly using PRR. Figure 4 summarizes the traditional wireless link quality prediction model.

The RNN-LQP proposed in this paper is shown in Figure 5. The predictor uses a recurrent neural network to predict the LQI counted in a small time window to ensure sufficient agility. Considering that the cyclic neural network has short-term memory characteristics and high prediction accuracy, it can also ensure sufficient accuracy and reliability. In addition, the predictor also selects LQI as the physical layer parameter.

In order to prevent data conflicts and network congestion and cause a large number of data packet loss, which will cause a serious negative impact on the performance of the

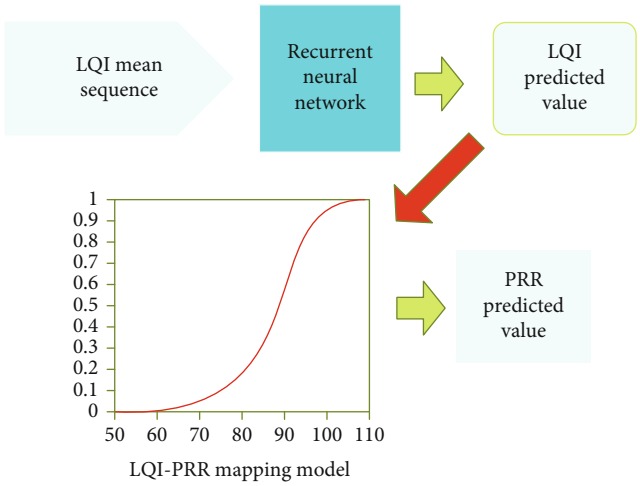


Figure 5: RNN-LQP structure diagram.

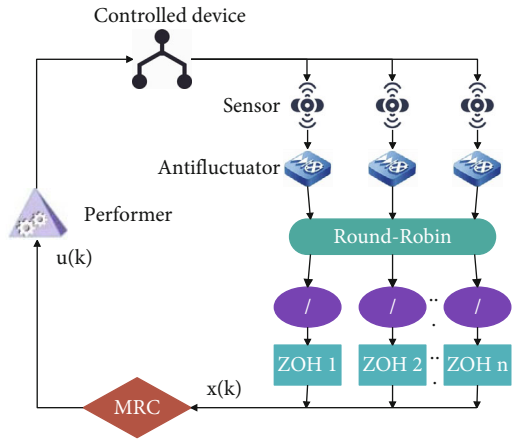


Figure 6: The structure diagram of the closed-loop system under the RR protocol.

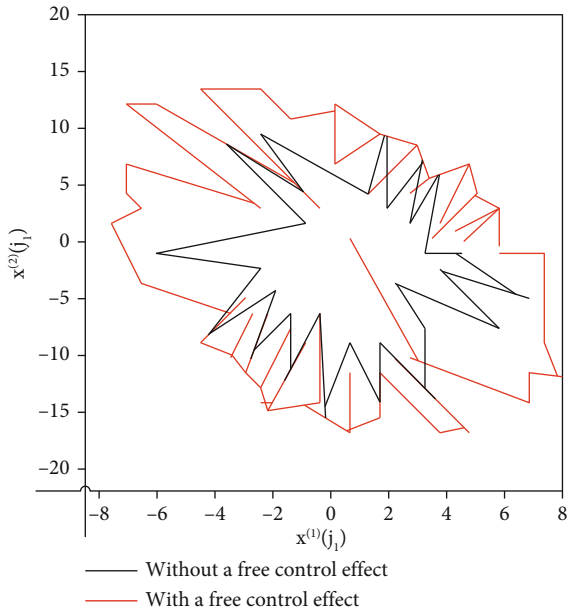


Figure 7: State trajectory of robust MPC based on RR protocol.

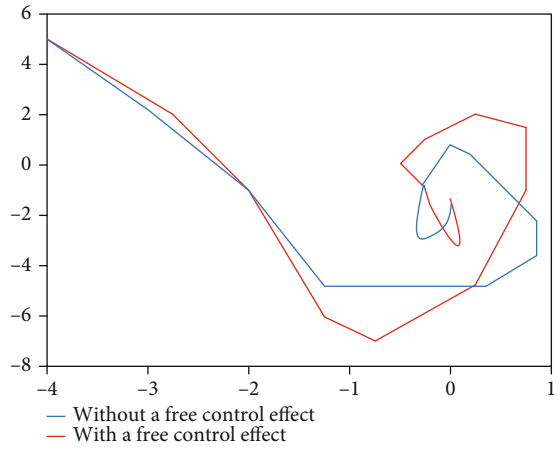


Figure 8: The attractive domain of robust MPC based on RR protocol.

system, this paper will use the RR protocol to schedule the data transmission sequence from the sensor to the controller. The structure of the control system based on the RR protocol is shown in Figure 6.

According to Figure 7, it can be obtained that both of the two robust MPC strategies proposed based on the cyclic communication protocol can make the closed-loop system finally reach an asymptotically stable state. In addition, according to Figure 8, it can be known that the attractive domain of the robust MPC strategy with a free control function is larger than that of the robust MPC strategy without free control function. Therefore, appropriately increasing the free control function can make the optimization of the system more flexible and easier to achieve stability.

The big data prediction model can effectively extract the characteristics of the time series of the core performance indicators of the communication network, but it still has the defect of low data processing rate. In order to solve this problem, this paper models separately according to the classification results, reduces the model data dimension, and improves the prediction rate. Figure 9 shows the flow chart of the improved big data prediction model based on support to the machine and association rules.

Although the analysis and prediction of massive data can improve the accuracy and reliability of the prediction results, the processing of massive data will inevitably bring defects such as slow prediction rate, so appropriate measures must be taken to increase the prediction rate. It is necessary to analyze the change law of the core performance indicators of the mobile communication network, use the change law to extract and classify the core performance indicators, and establish prediction models for different data segments of the performance indicators. On the one hand, data dimensions can be reduced, data processing time can be reduced, and the prediction rate can be improved. On the other hand, different parameters are used to predict different categories of core performance indicators, so that the prediction results

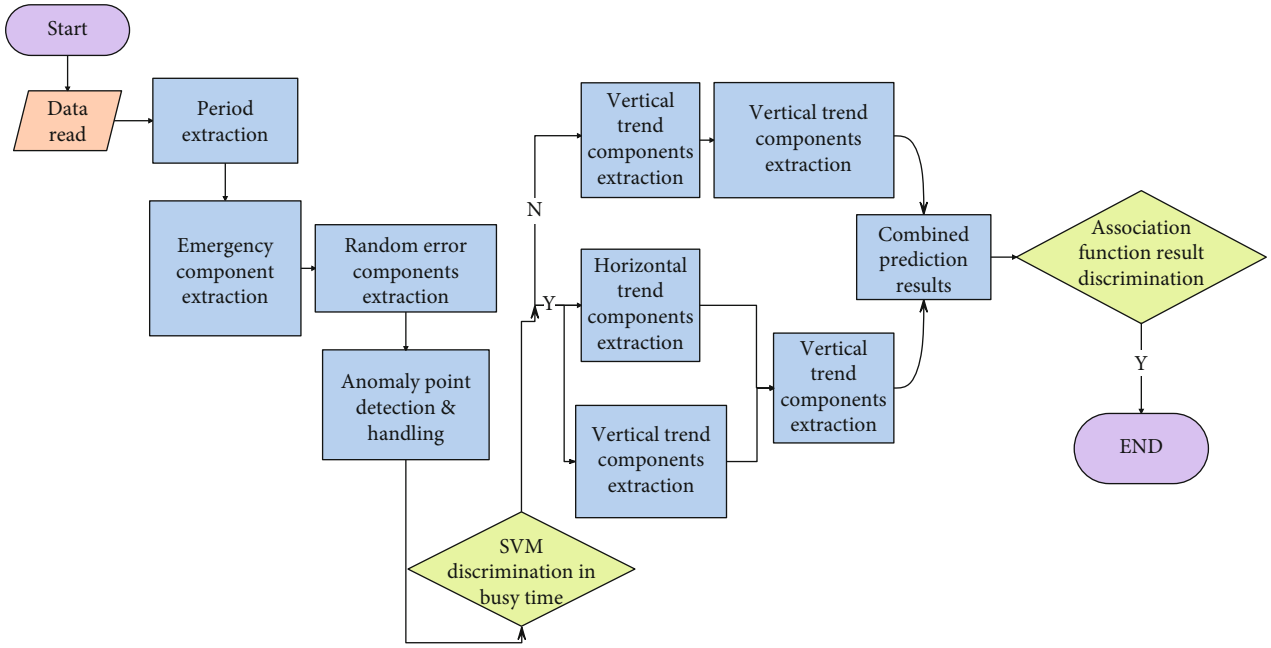


Figure 9: Flow chart of an improved big data prediction model based on SVM and association rules.

have higher accuracy. In time series, the series can be divided into stationary time series and nonstationary time series. A stationary series means that the mean and variance of the series have no systematic changes and strictly eliminate periodic changes. Intuitively speaking, the series have small fluctuations and no trend changes. The nonstationary time series corresponds to the stationary time series, and it is intuitively manifested as the series changes and fluctuates greatly. In the same time series, there are also stationary and nonstationary segments. For the core performance indicators of mobile communication networks in the form of time series, the numerical changes are also divided into stationary time series and nonstationary time series. For the core performance indicators of the mobile communication network, the stationary segment is called the nonbusy hour, and the nonstationary segment is called the busy hour. The so-called busy hour means that the core performance index of the cell fluctuates greatly over time, and the so-called nonbusy hour means that the change of the core performance index of the cell tends to be stable. In mobile communication networks, busy hours and nonbusy hours generally appear in day and night, weekends, and normal hours. Excessive cell network load is the main reason for busy hours, and the main goal of network optimization is to optimize network resources and reduce network load. Therefore, in the prediction of core performance indicators of mobile communication networks, busy hour performance indicators are more important. Figure 10 shows the busy and nonbusy hours of the RRC setting success rate in one cycle. Among them, the blue dot represents the value of the RRC setting success rate, the red area is the busy period, the green area is the nonbusy period, and the black line is the line where the dividing point between busy and nonbusy hours is located.

Since the predictive analysis of busy hour performance indicators is essential for network optimization, and the correct distinction between busy hour and nonbusy hour is the prerequisite guarantee for busy hour prediction, the traditional method of identifying busy hour and nonbusy hour is to find a demarcation point. The algorithm performs differential processing on points in a single cycle and takes the absolute value to obtain L-1 differential values and defines the minimum busy/nonbusy time length N and the busy/nonbusy time threshold xx. The algorithm sequentially calculates the probability $P1$ that the value of the first N consecutive points is greater than the threshold value for point i (range $N + 1 \sim L$) and the probability $P2$ that the value of the next consecutive N points is less than the threshold value. Moreover, the algorithm calculates the comprehensive proportion $P1 + P2$ or $2 - P1 - P2$ and selects the point with the maximum comprehensive proportion among all the points as the busy/nonbusy time boundary point. The length N is user-defined and is at least one-third of the period value. If it is less than one-third of the period value, there

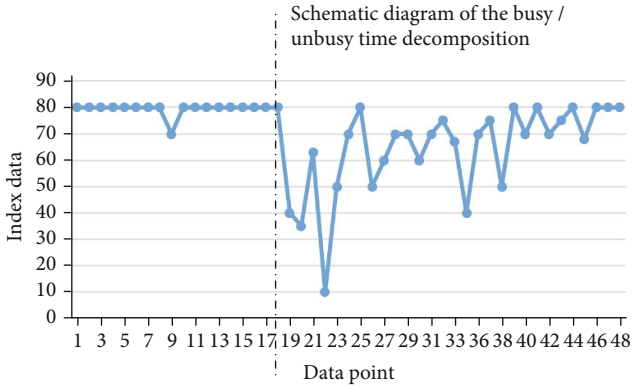


Figure 10: RRC setting success rate during busy/nonbusy time in a single cycle.

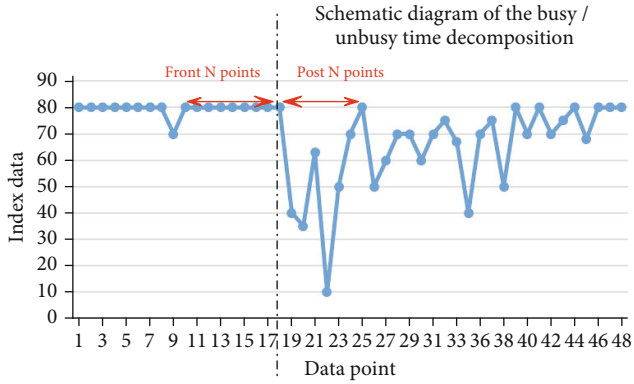


Figure 11: Discrimination rules for single-cycle busy/nonbusy time.

Table 1: The prediction effect of the intelligent prediction analysis system based on the embedded wireless communication network.

No.	The method of this paper	The method of [18]	No.	The method of this paper	The method of [18]
1	88.24	76.24	22	93.76	81.21
2	94.80	87.79	23	89.75	80.83
3	92.19	80.67	24	88.93	83.60
4	95.61	86.90	25	93.63	81.24
5	90.79	85.85	26	93.72	82.40
6	90.34	80.82	27	86.57	81.64
7	92.03	84.27	28	87.84	82.19
8	89.99	79.78	29	88.02	80.66
9	90.31	81.16	30	94.71	83.80
10	92.04	84.58	31	91.93	79.36
11	92.09	82.65	32	89.85	77.48
12	91.54	84.08	33	88.75	82.56
13	95.55	82.80	34	94.19	82.16
14	88.22	83.24	35	86.38	75.30
15	90.36	83.11	36	95.21	81.00
16	89.64	81.32	37	86.80	76.10
17	89.29	81.96	38	86.88	80.43
18	91.78	81.04	39	94.31	80.29
19	86.25	76.18	40	92.11	86.84
20	94.62	84.86	41	92.51	78.70
21	87.94	75.12	42	88.71	77.93

may be multiple busy/nonbusy demarcation points in a single period. The threshold is determined according to the fluctuation range of the performance index, and the general range is 0.1 ~ 1. The figure below shows the rules for distinguishing between busy and nonbusy hours. The discriminant rule of single-cycle busy/nonbusy time is shown in Figure 11.

The method in this paper is compared with the literature 19, and the effect of model prediction is comprehensively evaluated. The performance evaluation of the intelligent predictive analysis system based on the embedded wireless communication network is carried out, and its intelligent predictive effect and intelligent decision-making effect are

Table 2: Decision-making effect of intelligent predictive analysis system based on embedded wireless communication network.

No.	The method of this paper	The method of [18]	No.	The method of this paper	The method of [18]
1	85.72	74.78	22	83.25	76.66
2	81.36	73.41	23	92.42	80.10
3	88.89	78.69	24	81.01	70.97
4	80.13	70.80	25	87.11	75.45
5	92.68	84.75	26	87.70	74.73
6	85.44	78.03	27	82.96	77.31
7	79.71	72.31	28	84.74	76.48
8	80.15	70.75	29	91.16	85.02
9	88.64	81.93	30	92.08	79.88
10	89.11	76.71	31	83.05	71.75
11	80.95	76.68	32	90.59	79.47
12	87.04	79.12	33	82.88	71.64
13	89.74	77.02	34	80.17	71.58
14	81.35	75.06	35	89.89	79.05
15	84.31	79.19	36	81.59	71.31
16	87.58	75.12	37	82.62	73.66
17	79.58	70.11	38	84.82	72.23
18	92.72	86.95	39	88.01	75.30
19	87.48	77.94	40	86.60	82.19
20	90.91	81.84	41	89.08	77.69
21	80.71	76.12	42	83.85	71.71

counted, and the test results shown in Tables 1 and 2 below are obtained.

From the above research, it can be seen that the intelligent predictive analysis system based on the embedded wireless communication network proposed in this paper is very effective and has a positive effect on the construction and development of the embedded wireless communication network.

5. Conclusion

In order to further improve the network performance, it is necessary to study the network traffic and extract the parameters that can characterize the network traffic, so as to pass the modeling and performance analysis of the network traffic. At the same time, it is necessary to find adjustable performance parameters and implement effective control of traffic, thereby improving and optimizing network performance. In addition, today's networks are beginning to carry more and more application services, the scale of networks is getting larger and larger, and the characteristics of network behaviors are becoming more and more complex. This has brought huge challenges to network service quality, flow control, and network management planning, and contradictions have become increasingly prominent. This article takes the embedded wireless communication network as the research object, constructs the intelligent predictive analysis system, and validates and analyzes its performance. The

research results show that the intelligent predictive analysis system based on the embedded wireless communication network proposed in this paper is very effective and has a positive effect on the construction and development of the embedded wireless communication network.

The research work in this paper is mainly to carry out distributed predictive control research for the agent of the nonlinear nominal model, but the influence of uncertainty on the system is not fully considered, such as the uncertainty of the agent model, the uncertainty of external interference, and communication uncertainty between agents. In view of these problems, the traditional robust predictive control method can be used for reference, but it cannot be simply and directly extend to distributed situations, so the follow-up further research needs to focus on the above problems and performance improvement.

Because the forecasting process is affected by many factors, the regression-based forecasting method has complex characteristics, which makes its forecasting have its complex characteristics. At the same time, each forecasting model has its own shortcomings and adaptability. Therefore, as far as the current research methods and prediction models are concerned, a lot of research is still needed, which is also the next step.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This study is sponsored by Hefei University of Technology.

References

[1] A. Al-Halafi and B. Shihada, "UHD video transmission over bidirectional underwater wireless optical communication," *IEEE Photonics Journal*, vol. 10, no. 2, pp. 1–14, 2018.

[2] J. Barowski, M. Zimmermanns, and I. Rolfes, "Millimeter-wave characterization of dielectric materials using calibrated FMCW transceivers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 8, pp. 3683–3689, 2018.

[3] B. Behroozpour, P. A. M. Sandborn, M. C. Wu, and B. E. Boser, "Lidar system architectures and circuits," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 135–142, 2017.

[4] Q. Y. Cheng, X. L. Zhao, Y. X. Weng, Y. D. Li, and J. B. Zeng, "Fully sustainable, nanoparticle-free, fluorine-free, and robust superhydrophobic cotton fabric fabricated via an eco-friendly method for efficient oil/water separation," *ACS Sustainable Chemistry & Engineering*, vol. 7, no. 18, pp. 15696–15705, 2019.

[5] Y. Jiang, S. Karpf, and B. Jalali, "Time-stretch LiDAR as a spectrally scanned time-of-flight ranging camera," *Nature Photonics*, vol. 14, no. 1, pp. 14–18, 2020.

[6] H. C. Kumawat and A. B. Raj, "Extraction of Doppler signature of micro-to-macro rotations/motions using continuous wave radar-assisted measurement system," *IET Science, Measurement & Technology*, vol. 14, no. 7, pp. 772–785, 2020.

[7] Y. J. Ma, A. Tadros, J. du, and E. Y. Chang, "Quantitative two-dimensional ultrashort echo time magnetization transfer (2D UTE-MT) imaging of cortical bone," *Magnetic Resonance in Medicine*, vol. 79, no. 4, pp. 1941–1949, 2018.

[8] N. Maring, P. Farrera, K. Kutluer, M. Mazzera, G. Heinze, and H. de Riedmatten, "Photonic quantum state transfer between a cold atomic gas and a crystal," *Nature*, vol. 551, no. 7681, pp. 485–488, 2017.

[9] Z. Meng, J. Li, C. Yin et al., "Dual-band dechirping LFMCW radar receiver with high image rejection using microwave photonic I/Q mixer," *Optics Express*, vol. 25, no. 18, pp. 22055–22065, 2017.

[10] H. Mohapatra and A. K. Rath, "Detection and avoidance of water loss through municipality taps in India by using smart taps and ICT," *IET Wireless Sensor Systems*, vol. 9, no. 6, pp. 447–457, 2019.

[11] J. Pan, Q. Xie, H. Chiang et al., ""From the nature for the nature": an eco-friendly antifouling coating consisting of poly(lactic acid)-based polyurethane and natural antifoulant," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 3, pp. 1671–1678, 2020.

[12] Z. Sabouri, A. Akbari, H. A. Hosseini, A. Hashemzadeh, and M. Darroudi, "Eco-friendly biosynthesis of nickel oxide nanoparticles mediated by okra plant extract and investigation of their photocatalytic, magnetic, cytotoxicity, and antibacterial properties," *Journal of Cluster Science*, vol. 30, no. 6, pp. 1425–1434, 2019.

[13] A. Seri, G. Corrielli, D. Lago-Rivera et al., "Laser-written integrated platform for quantum storage of heralded single photons," *Optica*, vol. 5, no. 8, pp. 934–941, 2018.

[14] S. Sharaf and M. E. El-Naggar, "Eco-friendly technology for preparation, characterization and promotion of honey bee propolis extract loaded cellulose acetate nanofibers in medical domains," *Cellulose*, vol. 25, no. 9, pp. 5195–5204, 2018.

[15] K. Soga and L. Luo, "Distributed fiber optics sensors for civil engineering infrastructure sensing," *Journal of Structural Integrity and Maintenance*, vol. 3, no. 1, pp. 1–21, 2018.

[16] L. J. Xu, X. Lin, Q. He, M. Worku, and B. Ma, "Highly efficient eco-friendly X-ray scintillators based on an organic manganese halide," *Nature Communications*, vol. 11, no. 1, pp. 1–7, 2020.

[17] T. Zhong, J. M. Kindem, J. G. Bartholomew et al., "Nanophotonic rare-earth quantum memory with optically controlled retrieval," *Science*, vol. 357, no. 6358, pp. 1392–1395, 2017.

[18] G. Aceto, G. Bovenzi, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapé, "Characterization and prediction of mobile-app traffic using markov modeling," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 907–925, 2021.

[19] A. Montieri, G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapè, "Packet-level prediction of mobile-app traffic using multitask deep learning," *Computer Networks*, vol. 200, article 108529, 2021.

[20] F. Zhang, Q. Guo, and S. Pan, "Photonics-based real-time ultra-high-range-resolution radar with broadband signal generation and processing," *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.

[21] T. Zhong and P. Goldner, "Emerging rare-earth doped material platforms for quantum nanophotonics," *Nano*, vol. 8, no. 11, pp. 2003–2015, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5409046, 12 pages
https://doi.org/10.1155/2022/5409046

Research Article

The Application of the Combination of Virtual and Reality in the Film Space Performance

Yi Wang and Yidong Cheng

Sejong University, Seoul 05006, Republic of Korea

Correspondence should be addressed to Yi Wang; wyjjh20190420@163.com and Yidong Cheng; 16115291@bjtu.edu.cn

Received 14 December 2021; Revised 8 January 2022; Accepted 20 January 2022; Published 11 February 2022

Academic Editor: Mu Zhou

In order to improve the effect of film space performance, this paper applies the combination of virtual and reality to the film space performance processing. Moreover, this paper analyzes the light field distribution of the three-dimensional object in space, and modulates the projected image projected on it with the spectral characteristics of the optical screen to finally restore the light field of the original object. In addition, this paper uses the multi-projection method to study the recovery process of the cylindrical microlens array screen to the light field, and divides and reorganizes the multi-view image of the scene shot by the camera array to obtain the correct projected image. Finally, this paper applies the space processing algorithm based on the combination of virtual and reality to the film space performance processing, and uses experimental research to verify the reliability of the method proposed in this paper.

1. Introduction

When filming the film, the most important thing is how to use perspective to enhance and capture the sense of space in the picture. "Perspective is described as drawing various objects on a plane through the convergence of lines and giving people a sense of relative distance between objects in real space." The term perspective originated from Greece and refers to the method or technique of depicting the spatial relationship of objects on a plane or curved surface [1]. Perspective first appeared in painting, and at the beginning, the painter could only outline the outline of the scene with lines. What the painting showed was a pure two-dimensional form, which could not show the three-dimensional and spatial depth of the actual scene. Later in the Middle Ages, grid forms appeared in paintings. When the painters observe, the scene is gridded through a grid, and then the grid is reduced in proportion to the drawing paper, and then the actual scene is drawn using the grid on the drawing paper. The scenery drawn by this method, on a two-dimensional drawing paper, presents a three-dimensional sense of distance and a distinct depth of space, which is actually perspective [2].

The human eye is born with a perspective effect. When observing, all visible objects in the field of vision automatically enter the perspective state. Just as the painter observes the scene through the mesh grid, in the eyes of humans, the objective scene seems to be automatically included in a virtual grid plane. Moreover, the scenery lines and the air present a perspective effect, and show the perspective effects such as the density of virtual reality and the density of light and darkness [3].

Surrealist film creators usually apply the elements of surrealist dreams to movies. The fusion of surrealist space and reality is worth exploring in terms of technique and function. It forms a unique film narrative structure while also highlighting the emotional expression of the characters and showing the inner state of the characters. Digital technology helps the fusion of surrealist space and reality in a variety of ways, and on the basis of ensuring the transition of natural fusion, realizes the rich narrative features of surrealist space.

Based on the combination of virtual and real technology, this paper conducts research on film space performance, explores the detailed application process of film space expression techniques, and proposes the intelligent system of this paper to provide a theoretical reference for subsequent film space performance.

2. Related Work

Digital film and television" is a brand-new type of film and television produced under a new film and television production method that combines certain traditional film production technologies through computer and digital, audio and video, processing disk recording, and network technologies. The complete process from the early stage to the later stage to the release [4]. For example, in the early stage of creation, the computer-aided system is used to design, draw, test and simulate the scenes, plots, pictures, etc. of future films in order to find the best narrative Techniques and solutions for creating visual impact; another example is the use of computer control technology in actual shooting to complete certain shooting that cannot be done with traditional methods; another example is the use of computers to process and process images and sounds in post-production, The real shot material and computer images are synthesized, and the final synthesis processing is performed on the computer [5]. The emergence of digitization not only allows us to see the hope of revival of image creation through computers; but also makes us feel all kinds of In the growing market, computers and software have been used in all aspects of film and television, which has improved production efficiency [6]. "The practice of using computers to change, enhance and reshape the original pictures of films has developed steadily. Computers use digital methods and methods of enhancing film images to create magical images: using data and movies as digital outdoor scenes, digital color correction, filter effects and painting, digital image synthesis, digital animation production, Digital film repair, digital line mark removal, digital image enhancement, digital defect elimination, the integration of computer-generated material and film shooting material [7]. "Digital technology has caused us a strong and dazzling visual impact, and the fake process of its modified images is so simple and clever. With these, we have displayed the power of computer imaging technology at a glance. No wonder some people would say, "Today's movies are no longer "made", but "made". "The digitization of film and television has fundamentally changed the fate of film and television: digital film and television create a new era of entertainment facilities [8]. When film and television audiences are attracted by other media, film and television can only survive by reforming again. Digital computing creates The multi-layer synthesis of the breathtaking scenes, the shocking sound effects, and the extra large screen without the frame, the presence and reality of the cinema projection will be unmatched by any media [9]. Various special effects Entertainment programs, dynamic movies, virtual scenes, and various new types of programs will inject fresh blood into the production of motion video programs. The production and broadcasting of film and television programs has become more diversified, randomized, globalized and accessible due to the addition of digital methods. Pursuing. The development of film and television technology is closely linked to the continuous progress of science and technology. In the early stage of film and television development, the film recorded life at that time, and the development of digital technology has recorded our lives in a rich and colorful [10]. From the early silent black and white movies to the current multi-channel surround sound digital movies, from the original reception of TV signals to the current online video on-demand, the continuous evolution and development of technology and equipment has promoted the change and innovation in the thinking of creators [11]. "At the same time, digital technology has had a revolutionary impact on film and television production methods since its birth. It has incredible expansion of film and television expression space and realizability, creating unprecedented audiovisual wonders and virtual reality that people have never heard of, seen, or even imagined. .Therefore, a new narrative method is produced, and a set of new rules is proposed for us to understand, use, and break through. The new digital technology not only produces new video works, but also cultivates a new generation of video audiences [12].

The production process of digital special effects movies is roughly divided into three main parts: pre-planning, mid-term production, and post-production. The pre-planning is mainly composed of directors, special effects artists, screenwriters and other key creative personnel to jointly determine the story outline, that is, the visualized script. It is necessary to complete a unified planning, design, and determination of all involved image elements, such as characters, environments, and props. Special effect production plan and test, determine personnel's responsibilities and coordination work [13]. Mid-term production is carried out around the production of animation, including model establishment, material and lighting settings, special effects generation, image rendering, etc., and in conjunction with these core productions, some peripheral cooperation work should be carried out, such as providing three-dimensional scanning. Data model production, texture mapping, motion data capture, etc. [14]. The post-compositing work is mainly divided into three aspects: one is to complete the shots that are inconvenient in the animation software at one time by layered synthesis; the second is to add special effects; the third is to improve the overall artistic style and picture quality of the film Process [15]. Faced with the lack of real images in digital images, the anxiety of filmmakers has blurred the content of the ideological film to convey to the audience, because it blurs the difference between the essence of the film and the entertainment of the film. This sense of crisis is expressed in a "loss" argument: digital images used to be thought that the film may present a poor version of the effect. Literature [16] even regards digital special effects as a representative of industrial revolution automation. Digital media is only a representation of symbols. Living in such a formal and diversified world, the degree of

separation from the material of the real world is unprecedented. This has an important trend, which is to liberate the engrossed audience from the so-called real world.

3. Space Image Processing Algorithm Based on the Combination of Virtual and Reality

The plenoptic function is based on the observer's description of light in space and time. The mathematical form of the plenoptic function is [17]:

$$L = P(x, y, z, \theta, \phi, \lambda, t). \tag{1}$$

It is a 7-dimensional plenoptic function. Among them, (x, y, z) represents the coordinates of the light in space, λ and t are the wavelength of light and the time when the light is observed, respectively. Because the wavelength of light changes very little in free space, for a single observer, the plenoptic function of an object observed at a specific position at a certain moment can be expressed in a 5-dimensional form [18]:

$$L = P(x, y, z, \theta, \phi). \tag{2}$$

In the light field three-dimensional display, the three-dimensional object to be displayed can be regarded as composed of a large number of three-dimensional points, and each three-dimensional point actively or passively emits light to the surroundings. The light that enters the human eye causes us to see the corresponding three-dimensional object. Therefore, the plenoptic function can be used to represent the light field of a three-dimensional object:

$$L = L_{o_1}(x_i, y_i, z_i, \theta, \phi). \tag{3}$$

Among them, $i = 1,2, \ldots, N$ is the number of three-dimensional points of the object, as shown in Figure 1.

Due to the huge amount of information of the five-dimensional light field, usually in the process of three-dimensional display of the light field, the light field information in the vertical direction is compressed in combination with the optical screen, and only the light field information in the horizontal direction is processed. The light field at this time is expressed as [19]:

$$L = L_{o_1}(x_i, y_i, z_i, \phi), \quad i = 1, 2, \ldots, N. \tag{4}$$

In Figure 1, the holographic directional scattering screen is located at $z = z_1$, the light emitted by the three-dimensional point O_i intersects the screen with point M, the direction vector is $(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta)$, and the straight line OM can be expressed as:

$$\frac{x_1 - x_i}{\sin\theta\cos\phi} = \frac{y_1 - y_i}{\sin\theta\sin\phi} = \frac{z_1 - z_i}{\cos\theta}. \tag{5}$$

From formula (5), the x and y coordinates of point M can be calculated as:

$$\begin{cases} x_1 = (z_1 - z_i)\tan\theta\cos\phi + x_i, \\ x_1 = (z_1 - z_i)\tan\theta\sin\phi + y_i, \end{cases} \tag{6}$$

$\alpha = \tan\theta\cos\phi$ and $\beta = \tan\theta\sin\phi$ respectively represent the spatial angle information in the x and y directions, and the formula (6) can be written as [20]:

$$\begin{cases} x_1 = (z_1 - z_i)\alpha + x_i, \\ y_i = (z_1 - z_i)\beta + y_i. \end{cases} \tag{7}$$

At this time, the light field of the light O_iM emitted by the three-dimensional point O_i on the holographic directional scattering screen can represent $L_{O_1}(x_1, y_1, \alpha, \beta)$.

Next, in order to better understand the reproduction of the spatial light field by the three-dimensional display of the light field, the light field distribution of the three-dimensional object in the space is analyzed below. We assume two point elements A and B of the object in space, and analyze their light field distribution on the plane P. It can be seen from formulas (5)–(7) that x and y are related to α and β respectively, but they are independent of each other. Therefore, this section selectively analyzes only the x direction of light.

The three-dimensional coordinates of the point elements A and B are (x_A, y_A, z_A) and (x_B, y_B, z_B) respectively. They emit light in all directions, and the plane P is located at z_P, as shown in Figure 2. Generally, for any piece of light, according to formula (7), the light field of point element A and B at plane P can be expressed as:

$$A(\text{Red}): \begin{cases} x = (z_P - z_A)\alpha + x_A, \\ y = (z_P - z_A)\beta + y_A, \end{cases} \tag{8}$$

and

$$B(\text{Green}): \begin{cases} x = (z_P - z_B)\alpha + x_B, \\ y = (z_P - z_B)\beta + y_B. \end{cases} \tag{9}$$

The corresponding light fields are shown in the right figure of Figure 2, which are two oblique straight lines. In the figure on the right, the abscissa represents the spatial position of the light field, and the ordinate represents the angle information of the light in that direction. Since the point element A is far away from the plane P, the light field it produces has a smaller slope.

By doing the above analysis on all the point elements of the object in space, the entire light field of the object at the plane P can be obtained. Therefore, the purpose of the three-dimensional display of the light field is to use the optical screen to restore the light field of the object so that it can be seen by the observer, just like seeing a real object.

Considering that the projection distance of the projector is usually much larger than the projection aperture, it can be approximately considered that the light reaching the surface of the cylindrical microlens array screen is parallel light. The modulation mechanism of the cylindrical microlens array screen to the projection light is shown in Figure 3. The red, green, and yellow lights in the figure represent the collection of light projected from three different directions. When the

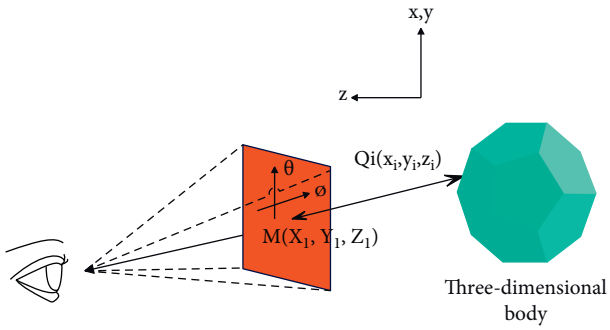


Figure 1: Representation of the plenoptic function at a specific position.

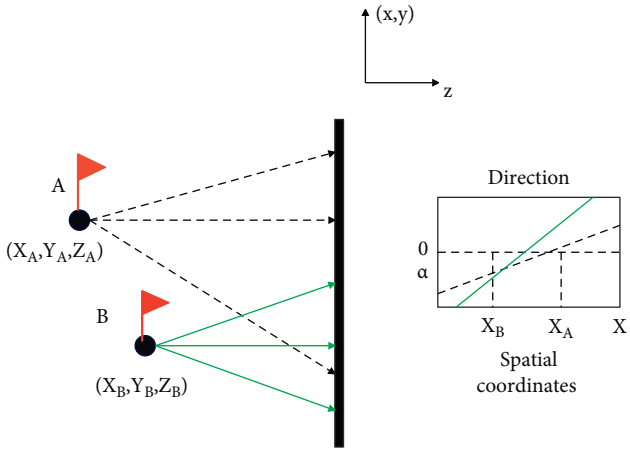


Figure 2: Distribution of light field in space.

light in the horizontal direction is converged, due to the scattering of the scattering film, part of the light will propagate back to the original projection direction according to the incident light path. Therefore, observers at different positions in the observation area will see the image projected by the corresponding projector, as shown in Figure 4.

Next, this paper analyzes the recovery process of the cylindrical microlens array screen to the light field by means of multi-projection. In Figure 4, the projector array is located in front of the screen, and the distance from the screen is d. The modulation function of the cylindrical microlens array on the light field can be expressed as:

$$\alpha' = F(x, y, -\alpha). \tag{10}$$

The light to be projected travels in the opposite direction. The pupil coordinate of projector P_i is $(x_{P_1}, y_{P_i}, z_{zp-d})$, which is regarded as a point light source. The light field of the light projected by P_i on the cylindrical microlens array screen is expressed as:

$$\begin{cases} x = x_{P_i} - d\alpha \ , \\ y = y_{P_i} - d\beta \ . \end{cases} \tag{11}$$

The information of each ray in the light field is determined by the image projected by P. By analogy, the light field corresponding to the projected image of the entire projector array can be obtained. The light field distribution corresponding to multiple projected images is shown in Figure 5(a), which shows the light field distribution in the x direction. It can be seen from Figure 5(a) that the light field of the image projected by the projector array is a series of straight lines with the same slope, and the slope is negative. Therefore, by combining the real light field distribution of the object point element in formula (8) and formula (9), the algorithm uses formula (11) to sample it to calculate the projected image corresponding to each projector in the projector array, as shown in Figure 5(b).

In particular, it can be found from Figure 4 that when the observer is under the projector array, the viewpoint image observed by the human eye is the same as the projected image of the projector. When the human eyes at different positions at this time are imagined as a camera array, it can be inferred that the multi-view image of the scene shot by the camera array is the projected image of the projector array.

We assume that the holographic directional scattering screen is located at zi in Figure 2, and the projector array is located in front of the screen, and the distance from the screen is d_p. The observation area is located behind the screen with a distance of d_v, as shown in Figure 6. The modulation function of the horizontal holographic directional scattering screen can be expressed as:

$$\alpha' = F(x, y, \alpha). \tag{12}$$

That is, the horizontal direction does not change the spread of light. The pupil coordinate of the projector P_i in Figure 6 is $(x_{P_1}, y_{P_i}, z_{zp-d})$, and the light field of the light projected by P_i on the holographic directional scattering screen is expressed as:

$$\begin{cases} x = x_{P_i} - d\alpha \ , \\ y = y_{P_i} - d\beta \ . \end{cases} \tag{13}$$

The x-direction distribution of the light field corresponding to multiple projection images is shown in Figure 7(a), and the light field is a series of straight lines with the same slope and positive values.

Similarly, by combining the real light field distribution of the object point elements in formula (8) and formula (9), when the algorithm uses formula (13) to sample it, the projected image corresponding to each projector in the projector array in the case of using the holographic directional scattering screen can be calculated, as shown in Figure 7(b). However, unlike the light field recovery under the cylindrical microlens array screen, the viewpoint image content of the observation area in Figure 6 is not a projection image of a certain projector, but a collection of multiple image blocks projected by the projector array. For example, the viewpoint image at V_j in the figure is a mosaic of all the projected images after being modulated. Therefore, when performing multi-projection display, the multi-view image of the scene shot by the camera array needs to be

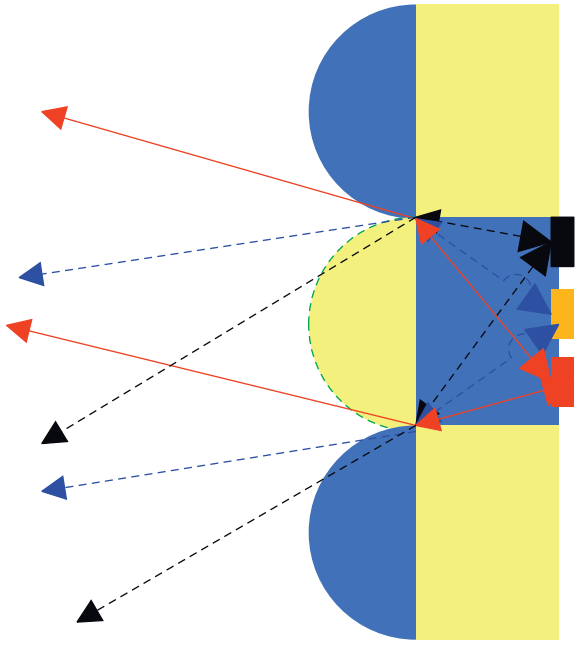


Figure 3: The light modulation mechanism of the cylindrical microlens array screen.

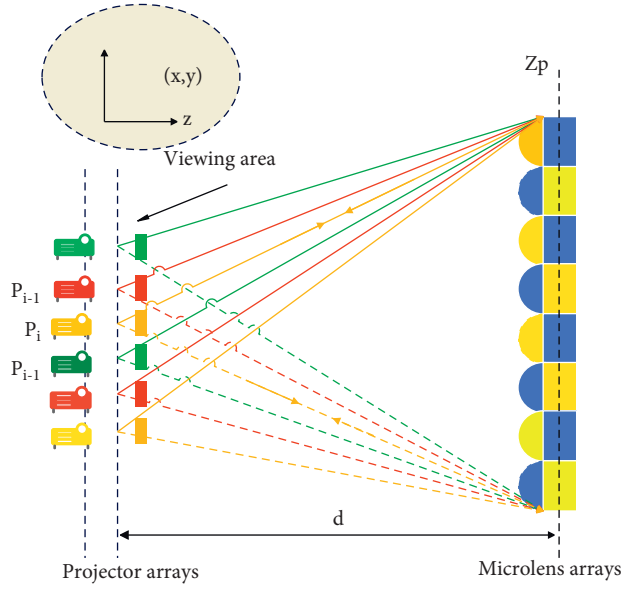


Figure 4: The three-dimensional display imaging model of the light field of the cylindrical microlens array screen.

segmented and reorganized to obtain the correct projected image.

In this paper, a front projection light field three-dimensional display system based on a cylindrical microlens array screen (projection array and the observer are on the same side of the screen) and a rear projection light field three-dimensional display system based on a holographic directional scattering screen are built (the projection array and the observer are on both sides of the screen). The system mainly includes two main parts: light field acquisition and light field display. The hardware structure of the system is shown in Figure 8.

The camera array (GigE interface) of the acquisition part collects the light field of the real scene. The camera array is arranged in an arc and compact, and is connected to the computer through a switch. The display part is mainly composed of a projection control computer, a projector array (HDMI interface) and a display screen. The projector array is staggered and closely arranged, and is connected to the computer through a screen splitter and multiple multi-channel graphics cards. The light field data captured by the collecting terminal is transmitted to the display terminal for display through the UDP protocol. The working resolution of the camera array and the projector array are both set to 1280×800.

The software module of the system is designed according to the structure of the hardware. The functions realized by the collection terminal include the collection of the light field by the camera array, the correction of the collected image, the generation of the projected image and the synchronous control of the collection. The display terminal mainly includes projection correction and synchronization control. The software block diagram of the system is shown as in Figure 9.

Assuming that there are four landmark points A, B, C, and D in the target light field, their coordinates are $(x_i, y_i), i \in (1, 2, 3, 4)$, and the coordinates of target points A', B, C', and D' are $(x_i', y_i'), i \in (1, 2, 3, 4)$, as shown in Figure 10. After extracting the coordinates of the marker point, map it to the target point through perspective projection to obtain the perspective transformation matrix P. The perspective transformation formulas are:

$$\begin{bmatrix} t_i x_i' \\ t_i y_i' \\ t_i \end{bmatrix} = P \cdot \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}. \tag{14}$$

In the formula, P is a 3×3 perspective transformation matrix, and t_i is a constant value coefficient. The form of P is:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}. \tag{15}$$

To simplify, we normalize P_{33}, that is, $P_{33} = 1$. From formula (14) and formula (15), we can get:

$$\begin{cases} x_i' = \dfrac{p_{11}x_i + p_{12}y_i + p_{13}}{p_{31}x_i + p_{32}y_i + 1}, \\ y_i' = \dfrac{p_{21}x_i + p_{22}y_i + p_{23}}{p_{31}x_i + p_{32}y_i + 1}. \end{cases} \tag{16}$$

There are 8 unknowns $p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}, p_{31}, p_{32}$ in the above formula, which can be solved by 8 equations formed by four pairs of coordinate points to obtain the perspective transformation matrix P. By multiplying the light field multi-view image collected by the camera array with the perspective transformation matrix P, the light field image of the same field of view can be obtained. The proposed camera array light field acquisition image correction

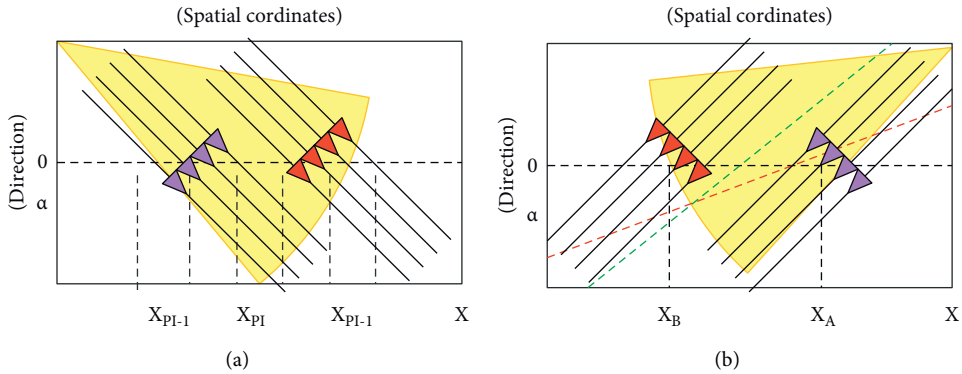


Figure 5: (a) Modulation of the cylindrical microlens array screen to the light field and (b) recovery of the target light field.

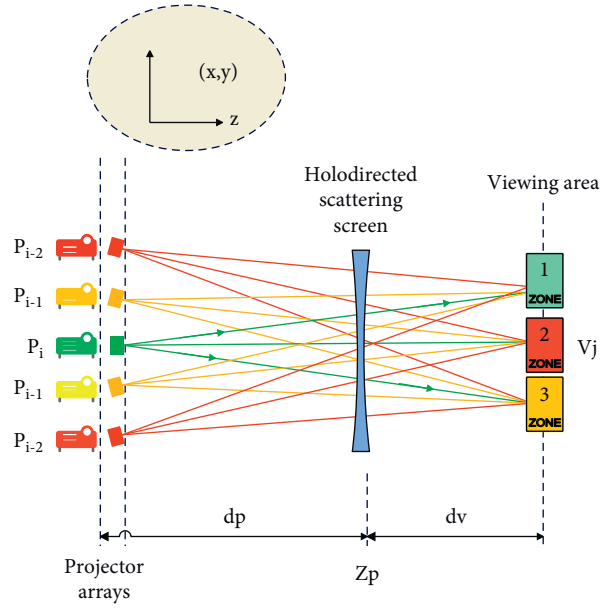


Figure 6: The light field three-dimensional display imaging model of the holographic directional scattering screen.

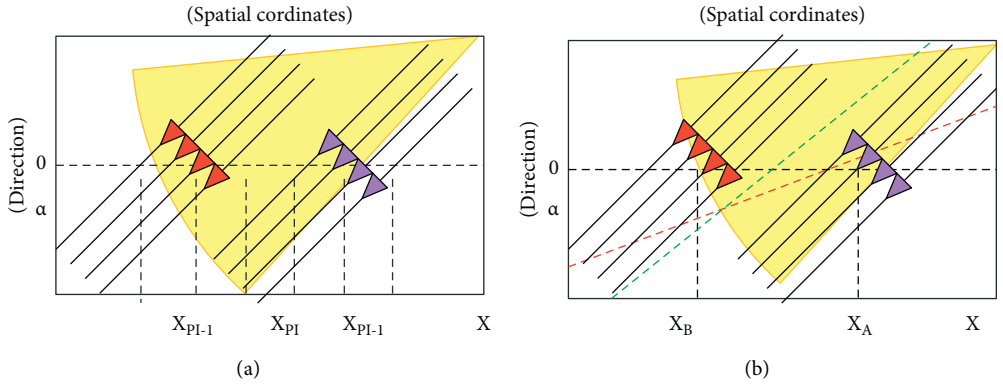


Figure 7: (a) Modulation of the light field by the holographic directional scattering screen and (b) recovery of the target light field.

method will reduce the spatial resolution of part of the light field, and finally the light field data within the range of the mark point will be retained.

When the three-dimensional model of the scene is known, a virtual camera can be used to render the viewpoint image of the scene, and then segmentation and

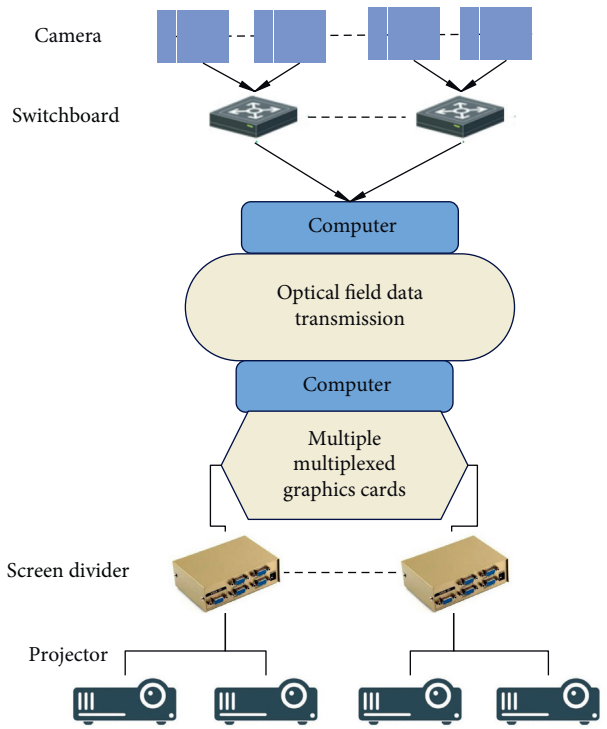


Figure 8: The structure block diagram of the light field three-dimensional acquisition and display system.

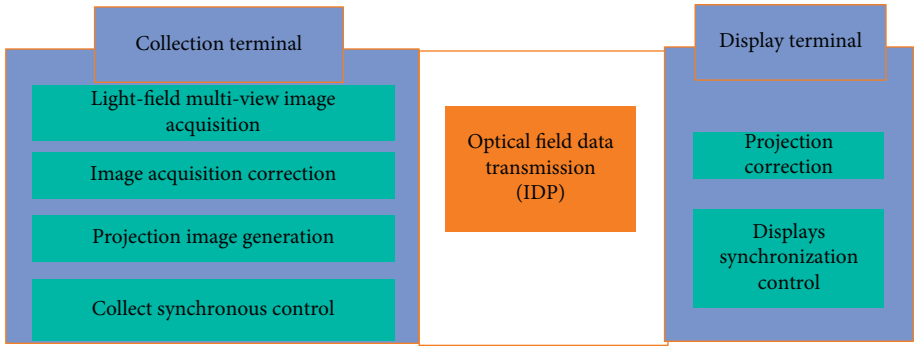


Figure 9: Software block diagram of the light field three-dimensional display system.

reorganization are performed to obtain the projected image. However, the efficiency of selecting this method for three-dimensional display is low. When the 3D model of the scene is known, the projection image is usually rendered directly from the 3D model. We establish a three-dimensional coordinate system with the center of the holographic directional scattering screen as the origin, and assume that the array of projectors in the system are arranged in arc rows, the arc radius is R_p, and the angular interval between adjacent projectors is θ.

The viewpoint distribution of the observation area is a circular arc $\overset{\frown}{l}_v$ with a radius of R_V and a height of H_V, and $S(x_0, y_0, z_0)$ is any point on the surface of the three-dimensional model. The following analyzes the process of using the model rendering method to obtain the projection image corresponding to the projector $P_i(x_{P_i}, y_{P_i}, z_{P_i})$ in the projector array.

As shown in Figure 11, first, the optical center of the projector and the three-dimensional point S are connected and extended to the cylindrical surface where the arc $\overset{\frown}{l}_v$ of the viewpoint distribution of the observation area is located, and the intersection point is denoted as v_i. The equation of light P_iS can be expressed as

$$\frac{x - x_0}{x_{P_i} - x_0} = \frac{y - y_0}{y_{P_i} - y_0} = \frac{z - z_0}{z_{P_i} - z_0} = k_{P_iS}. \tag{17}$$

Among them, k_{P_iS} is the proportional coefficient. The cylinder where the arc 1 of the viewpoint distribution is located can be written as:

$$x^2 + y^2 = R_V^2. \tag{18}$$

Through simultaneous formula (17) and formula (18), the proportional coefficient k_{P_iS} can be calculated as:

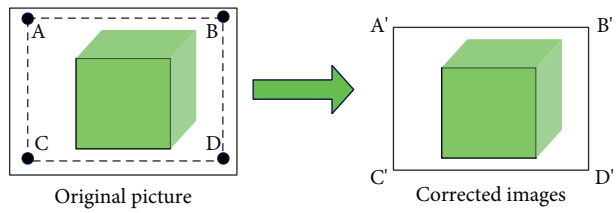


Figure 10: Schematic diagram of image correction method for real scene light field acquisition.

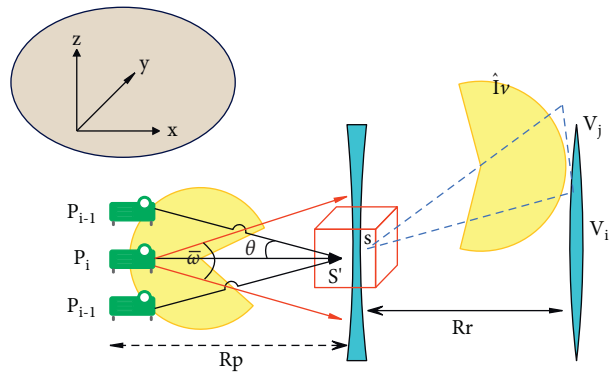


Figure 11: shows the calculation method of the projected image of the system.

$$k_{P_iS} = \frac{-\left(x_{P_i} - x_0\right)x_0 - \left(y_{P_i} - y_0\right)y_0}{\left(x_{P_i} - x_0\right)^2 - \left(y_{P_i} - y_0\right)^2} - \frac{\sqrt{\left[\left(x_{P_i} - x_0\right)^2 + \left(y_{P_i} - y_0\right)^2\right]R_V^2 - \left(x_{P_i}y_0 - y_{P_i}x_0\right)^2}}{\left(x_{P_i} - x_0\right)^2 + \left(y_{P_i} - y_0\right)^2}. \tag{19}$$

Thus, the three-dimensional coordinates of v_i can be expressed as:

$$\begin{cases} x_{v_i} = k_{P_iS}\left(x_{P_i} - x_0\right) + x_0, \\ y_{v_i} = k_{P_iS}\left(y_{P_i} - y_0\right) + y_0, \\ z_{v_i} = k_{P_iS}\left(z_{P_i} - z_0\right) + z_0. \end{cases} \tag{20}$$

Therefore, it can be known that the coordinate of the viewpoint V_j corresponding to v_i is $V_j\left(x_{v_i}, y_{v_i}, H_v\right)$. It can be seen from Figure 11 that the three-dimensional point S observed at the viewpoint.

$$\frac{x - x_0}{x_{v_i} - x_0} = \frac{y - y_0}{y_{v_i} - y_0} = \frac{z - z_0}{z_{v_i} - z_0} = k_{V_jS}. \tag{21}$$

Since the holographic directional scattering screen is parallel to the $y = z$ plane, the normal vector at this time is $\vec{n} = (1, 0, 0)$, and its surface equation can be written as:

$$1 \cdot (x - 0) + 0 \cdot (y - 0) + 0(z - 0) = 0. \tag{22}$$

In the same way, through simultaneous formula (21) and formula (22), the three-dimensional coordinates of S'are calculated as

$$\begin{cases} x_{S'} = k_{V_jS}\left(x_v - x_0\right) + x_0, \\ y_{S'} = k_{V_jS}\left(y_v - y_0\right) + y_0, \\ z_{S'} = k_{V_jS}\left(z_v - z_0\right) + z_0. \end{cases} \tag{23}$$

Among them, $k_{V_jS} = -x_0/x_v - x_0$. Then, through the decoy projection transformation, the two-dimensional pixels in the projection image corresponding to the three-dimensional point S′ are calculated.

We assume that the horizontal and vertical field of view of the projector is o. Since the projection plane of the projector is in the yz plane, the image coordinates $I(u, v)$ in the projector P corresponding to the three-dimensional point S' after transmission projection can be expressed as:

$$(u, v) = \text{round}\left(\left(\frac{y_{S'}}{R_P \cdot \tan(\omega/2)} + 1\right) \cdot \frac{M}{2}, \left(\frac{Z_{S'}}{R_P \tan(\omega/2)} + 1\right) \cdot \frac{N}{2}\right). \tag{24}$$

In the formula, the round symbol means to take the closest integer, and M and N are the horizontal and vertical resolutions of the projected image, respectively. Repeatedly, by using the transmission projection method to traverse the entire three-dimensional model, the light field of the model corresponding to the complete projection image in the

projector P can be obtained. After that, the same processing is performed on the entire projector array to generate projection images corresponding to all the projectors.

When a camera array is used to collect light field data of a real scene for three-dimensional display, the collected light field viewpoint image is the viewpoint image in the observation area. In this case, the viewpoint image needs to be segmented and reorganized to generate the projected image.

In order to facilitate understanding, the three-dimensional display of the light field of the three projection images is used as a column for analysis. As shown in Figure 12, since the holographic directional scattering screen does not change the angle of the projected light in the horizontal direction, viewpoints located at different positions in the observation area on the right side of the screen will see different projected image content. And each viewpoint image is composed of different parts in different projection images. The specific performance is as follows. We assume that the left, center, and right three projection images are denoted as A, B, and C respectively, and they are all divided into three columns of sub-image blocks. When the projected images A, B, C are projected to the same area of the holographic directional scattering screen, the viewpoint images seen by the observer at the left, middle, and right positions of the observation area will be B, C, and A, B, C, and A, B respectively, which are composed of three different sub-image blocks of the projected image.

We assume that the camera array captures N viewpoint images of the light field of the real scene. Each viewpoint image is divided equally into columns, and the number of equally divided columns is N, which is represented by the symbol $V[i,j]$. Among them, i is the serial number of the viewpoint image, and j represents the jth column in the image i. In the process of generating the projected image, in order to better explain the method of projected image generation, all the projected images are regarded as a large image that is horizontally spliced by N columns of equally divided projected images, which is represented by the symbol $P[n]$. The n in $P[w]$ is the column number in the projected large image, and satisfies $1 \le n \le N'$. For example, $P[6]$ represents the sixth column in the projected large image. It is concluded through experiments that the corresponding relationship between the viewpoint image and the projected large image can be expressed as:

$$P[n] = V[i, j]: n = N(N - i) + (N - 1)(j - 1). \tag{25}$$

This paper proposes a projection correction method based on solving the homography matrix. The purpose of projection correction is to correct all projected images to a designated area of the display screen, and its essence is to map the projected image plane coordinates to the world coordinates where the display screen is located. Therefore, the mapping relationship between the planes can be achieved through homography mapping. In particular, in order to improve the efficiency of projection correction, a camera is used as an auxiliary tool for correction.

The correction method involves three plane coordinate systems: the projection coordinate system (x_p, y_p), the camera coordinate system (x_c, y_c), and the display screen coordinate system (x_s, y_s). The homography matrix from the projection plane to the camera plane is denoted as H_{pc}, and the homography matrix from the camera plane to the display plane is denoted as H_{cs}. The homography transformation matrix for projection correction can be expressed as:

$$H = H_{\text{PC}} \cdot H_{\text{CS}} \cdot H_{\text{PC}}^{-1}. \tag{26}$$

The implementation steps of the projection correction method are as follows. First, each projector projects a standard chessboard image on the display screen, uses the camera to record the deformed chessboard image, uses feature extraction algorithms to extract the corner coordinates of the recorded chessboard image, and calculates the homography matrix H_{pc}. Then, the coordinates of the corner points of the deformed chessboard corresponding to each projector recorded by the camera are mapped to the specified screen coordinates, and the homography matrix Hes is calculated. Finally, the homography transformation matrix H is obtained by formula (26). The flow of the projection correction algorithm is shown in Figure 13.

4. Evaluation of the Application Effect of the Combination of Virtual and Reality in the Film Space Performance

On the basis of the above research on the virtual and real combination algorithm, this paper analyzes the application of virtual and real combination technology in film space performance. This article uses expert evaluation method to evaluate the effect, and obtains multiple sets of resource data through the Internet as the research sample of this paper. First, this paper evaluates the spatial image correction and projection process of the system proposed in this paper, and the results are shown in Table 1 and Figure 14 below.

It can be seen from the above research that the virtual-real combination technology proposed in this paper has good spatial image correction and projection effects in the design of film space. Afterwards, the film space performance effect of this system is evaluated, and the results shown in Table 2 and Figure 15 below are obtained.

From the above research, it can be seen that the virtual and real combination technology proposed in this paper performs well in the design of film space, which verifies that the method proposed in this paper has a certain effect.

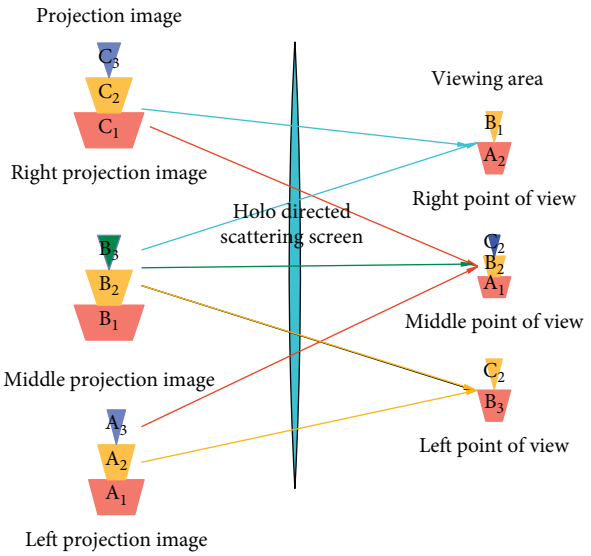


Figure 12: Correspondence between the projected image and the viewpoint under the holographic directional scattering screen.

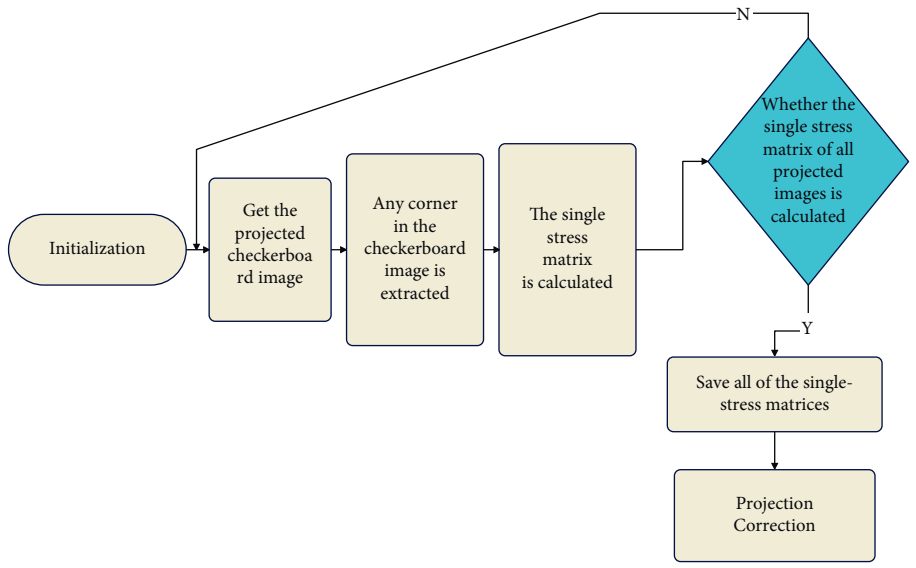


Figure 13: Flow chart of projection correction algorithm.

Table 1: Evaluation data of spatial image correction and projection process.

NO	Correction	Projection	NO	Correction	Projection
1	86.34	90.16	15	88.53	88.83
2	96.23	91.75	16	88.27	91.82
3	85.23	82.16	17	86.92	86.69
4	90.54	92.98	18	94.11	93.13
5	92.21	85.03	19	95.81	82.41
6	86.28	84.84	20	88.94	83.41
7	96.88	84.94	21	94.76	88.61
8	85.58	86.22	22	91.86	92.24
9	86.59	90.80	23	93.82	93.27
10	88.09	86.05	24	93.66	86.91
11	89.97	90.15	25	92.22	93.34
12	94.56	85.21	26	94.29	86.00
13	95.58	85.17	27	89.77	89.37
14	92.34	93.18	28	90.93	88.84

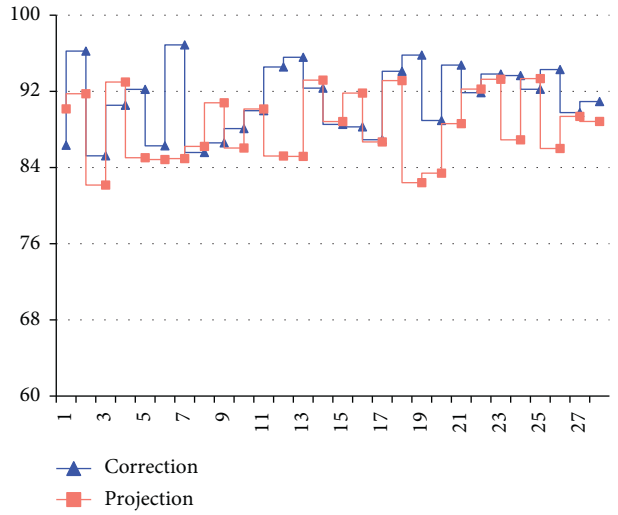


Figure 14: Statistical diagram of evaluation of spatial image correction and projection process.

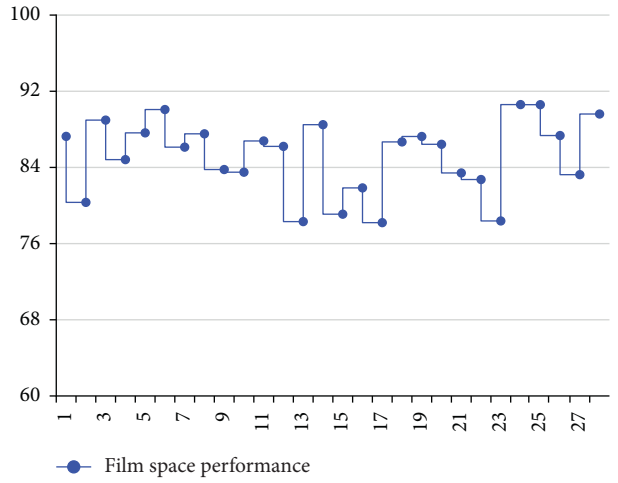


Figure 15: Statistical chart of film space performance effect.

Table 2: Evaluation data of film space performance effect.

NO	Film space performance	NO	Film space performance
1	87.26	15	79.09
2	80.33	16	81.85
3	88.97	17	78.20
4	84.82	18	86.68
5	87.63	19	87.25
6	90.08	20	86.43
7	86.13	21	83.42
8	87.53	22	82.73
9	83.78	23	78.38
10	83.50	24	90.60
11	86.78	25	90.59
12	86.21	26	87.35
13	78.31	27	83.24
14	88.49	28	89.61

5. Conclusion

The perspective effect produced on the film screen is similar to the principle of the perspective effect produced by painting. The objective scene in the form of three-dimensional space can produce a perspective effect through the optical lens of the camera due to the difference of the spatial position in the actual space. Then, three-dimensional images of different sizes, shapes, and colors are formed on a flat photosensitive medium to present images with a sense of space and depth. In film shooting, the lens replaces the human eye. As far as perspective is concerned, there is no essential difference between the perspective of a film picture and the perspective of a conventional painting, but the viewing range is limited by the focal length of the lens. From the perspective law, it can be found that when the human eye puts the scene into the field of view, it can automatically see through, thereby judging the actual size and distance of the object. Based on the combination of virtual and real technology, this paper conducts research on film space performance, explores the detailed application process of film space expression techniques, and proposes the intelligent system of this paper to provide a theoretical reference for subsequent film space performance.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This study is sponsored by Sejong University.

References

[1] J. Steffens, "The influence of film music on moral judgments of movie scenes and felt emotions," *Psychology of Music*, vol. 48, no. 1, pp. 3–17, 2020.

[2] G. Schalk, C. Kapeller, C. Guger et al., "Facephenes and rainbows: causal evidence for functional and anatomical specificity of face and color processing in the human brain," *Proceedings of the National Academy of Sciences*, vol. 114, no. 46, pp. 12285–12290, 2017.

[3] S. Han, B. Liu, R. Wang, Y. Ye, C. D. Twigg, and K. Kin, "Online optical marker-based hand tracking with deep labels," *ACM Transactions on Graphics*, vol. 37, no. 4, pp. 1–10, 2018.

[4] D. Stawarczyk, M. A. Bezdek, and J. M. Zacks, "Event representations and predictive processing: the role of the midline default network core," *Topics in Cognitive Science*, vol. 13, no. 1, pp. 164–186, 2021.

[5] A. G. Sares, N. E. V. Foster, K. Allen, and K. L. Hyde, "Pitch and time processing in speech and tones: the effects of musical training and attention," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 3, pp. 496–509, 2018.

[6] A. K. Fishell, T. M. Burns-Yocum, K. M. Bergonzi, A. T Eggebrecht, and J. P Culver, "Mapping brain function during naturalistic viewing using high-density diffuse optical tomography," *Scientific Reports*, vol. 9, no. 1, pp. 11115–11211, 2019.

[7] E. Peters and C. Muñoz, "Introduction to special issue Language learning from multimodal input," *Studies in Second Language Acquisition*, vol. 42, no. 3, pp. 489–497, 2020.

[8] C. Li, Z. Wang, Y. Lu, X. Liu, and L. Wang, "Conformation-based signal transfer and processing at the single-molecule level," *Nature Nanotechnology*, vol. 12, no. 11, pp. 1071–1076, 2017.

[9] N. Molinaro and M. Lizarazu, "Delta (but not theta)-band cortical entrainment involves speech-specific processing," *European Journal of Neuroscience*, vol. 48, no. 7, pp. 2642–2650, 2018.

[10] Y. Deldjoo, M. F. Dacrema, M. G. Constantin et al., "Movie genome: alleviating new item cold start in movie recommendation," *User Modeling and User-Adapted Interaction*, vol. 29, no. 2, pp. 291–343, 2019.

[11] R. Piryani, V. Gupta, and V. K. Singh, "Movie Prism: a novel system for aspect level sentiment profiling of movies," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 5, pp. 3297–3311, 2017.

[12] S. De Jans, D. Van de Sompel, L. Hudders, and V. Cauberghe, "Advertising targeting young children: an overview of 10 years of research (2006–2016)," *International Journal of Advertising*, vol. 38, no. 2, pp. 173–206, 2019.

[13] F. M. Schneider, "Measuring subjective movie evaluation criteria: conceptual foundation, construction, and validation of the SMEC scales," *Communication Methods and Measures*, vol. 11, no. 1, pp. 49–75, 2017.

[14] M. A. Mizher, M. C. Ang, A. A. Mazhar, and M. A. Mizher, "A review of video falsifying techniques and video forgery detection techniques," *International Journal of Electronic Security and Digital Forensics*, vol. 9, no. 3, pp. 191–208, 2017.

[15] J. T. Fisher, J. R. Keene, R. Huskey, and R. Weber, "The limited capacity model of motivated mediated message processing: taking stock of the past," *Annals of the International Communication Association*, vol. 42, no. 4, pp. 270–290, 2018.

[16] V. Grech, "The application of the Mayer multimedia learning theory to medical PowerPoint slide show presentations," *Journal of Visual Communication in Medicine*, vol. 41, no. 1, pp. 36–41, 2018.

[17] J. Black, M. Barzy, D. Williams, and H. Ferguson, "Intact counterfactual emotion processing in autism spectrum disorder: evidence from eye-tracking," *Autism Research*, vol. 12, no. 3, pp. 422–444, 2019.

[18] P. Tashman, V. Marano, and J. Babin, "Firm-specific assets and the internationalization–performance relationship in the US movie studio industry," *International Business Review*, vol. 28, no. 4, pp. 785–795, 2019.

[19] D. Stawarczyk, C. N. Wahlheim, J. A. Etzel, A. Z. Snyder, and J. M. Zacks, "Aging and the encoding of changes in events: the role of neural activity pattern reinstatement," *Proceedings of the National Academy of Sciences*, vol. 117, no. 47, pp. 29346–29353, 2020.

[20] N. Alp, A. R. Nikolaev, J. Wagemans, and N Kogo, "EEG frequency tagging dissociates between neural processing of motion synchrony and human quality of multiple point-light dancers," *Scientific Reports*, vol. 7, no. 1, pp. 44012–44019, 2017.