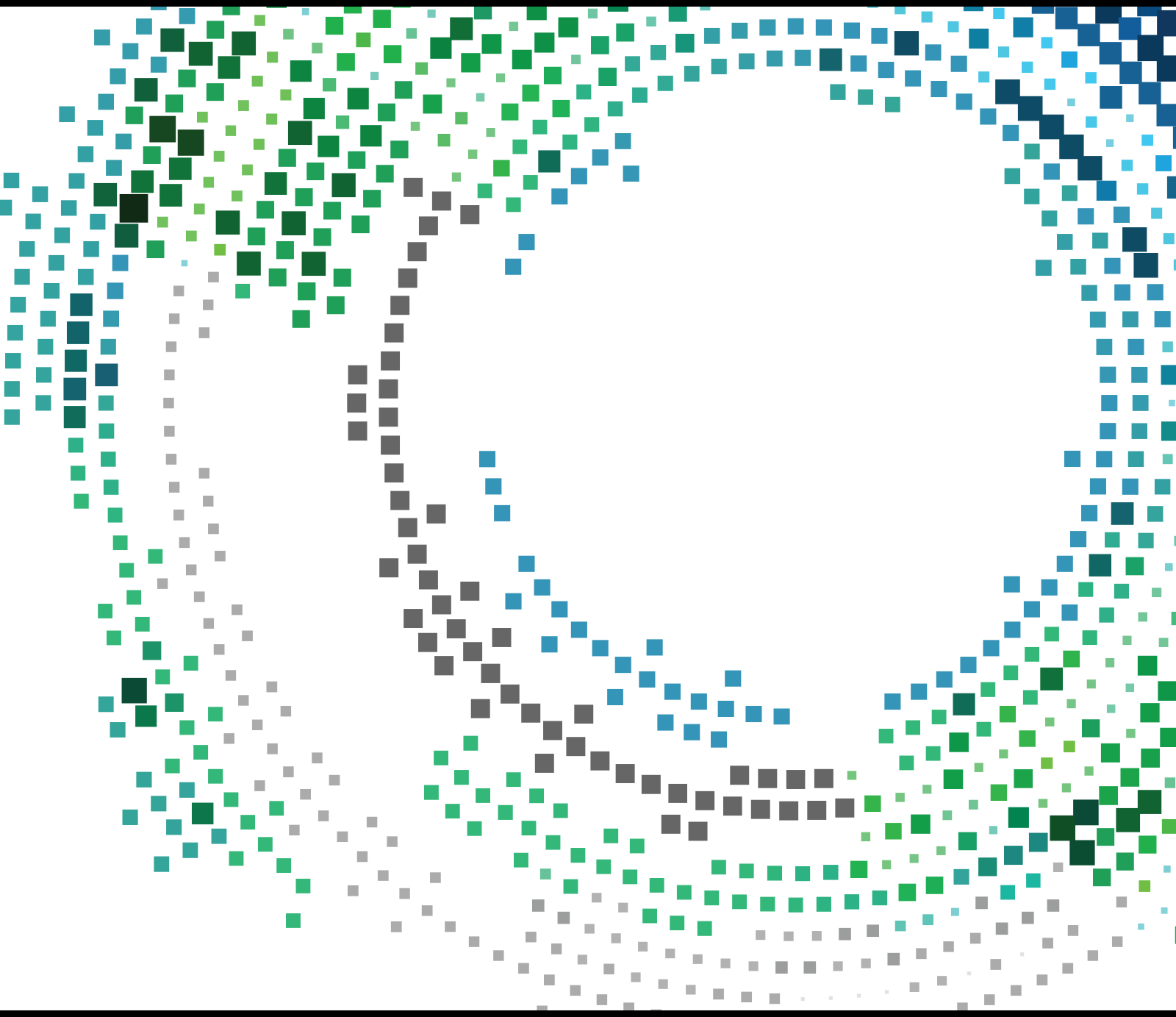
Mobile Information Systems

Intelligent Control in the Industrial Internet

Lead Guest Editor: Jianqing Li
Guest Editors: Wang Wenyong, Wei Ni, and Sai Zou



Intelligent Control in the Industrial Internet

Mobile Information Systems

Intelligent Control in the Industrial Internet

Lead Guest Editor: Jianqing Li
Guest Editors: Wang Wenyong, Wei Ni, and Sai Zou

Contents

Intelligent Testing and Analysis of Dissimilar Steel Welds for Industrial Throttle Flowmeter
Lianghuai Tong, Hui Xu, Xiaojie Xu, Huaye Cheng, Feng Li, and Jiahui Zhou
Research Article (11 pages), Article ID 4006715, Volume 2023 (2023)

An Underwater Target Tracking Algorithm Based on Extended Kalman Filter
Jian Huang
Research Article (12 pages), Article ID 9916531, Volume 2023 (2023)

Network Intrusion Anomaly Detection Model Based on Multiclassifier Fusion Technology
Feilu Hang, Wei Guo, Hexiong Chen, Linjiang Xie, Xiaoyu Bai, and Yao Liu
Research Article (11 pages), Article ID 1594622, Volume 2023 (2023)

Learning Identity-Consistent Feature for Cross-Modality Person Re-Identification via Pixel and Feature Alignment
Sixian Chan, Feng Du, Yanjing Lei, Zhounian Lai, Jiafa Mao, and Chao Li
Research Article (9 pages), Article ID 4131322, Volume 2022 (2022)

Sampled Characteristic Modeling and Forgetting Gradient Learning Algorithm for Robot Servo Systems
Hongbo Bi, Dong Chen, Yanjuan Li, and Ting You
Research Article (11 pages), Article ID 7259504, Volume 2022 (2022)

A Real-Time UWB Location and Tracking System Based on TWR-TDOA Estimation and a Simplified MPGA Layout Optimization
Yanping Zhu, Lei Huang, Jing Liu, Zhongkang Cao, Jinli Chen, and Zijian Mu
Research Article (10 pages), Article ID 1194169, Volume 2022 (2022)

Development of Wireless Transmission System for Microseismicity in Complex Mountainous Area
Qingming Xie, Chunling Wu, Hongliang Liao, Lichuan Chen, Yunbin Hu, Guilan He, and Yueming Kang
Research Article (12 pages), Article ID 7049377, Volume 2022 (2022)

A High-Performance Energy-Balanced Forwarding Strategy for Wireless Sensor Networks
Zhangxiang Hu, Xiaodan Jiang, Xiajun Ding, Kai Fang, and Xiaolong Zhou
Research Article (10 pages), Article ID 3058499, Volume 2022 (2022)

Prediction of Industrial Network Security Situation Based on Noise Reduction Using EMD
Guanling Zhao, Lisheng Huang, Lu Li, and Yongfeng Zhang
Research Article (14 pages), Article ID 2594000, Volume 2022 (2022)

Hindawi
Mobile Information Systems
Volume 2023, Article ID 4006715, 11 pages
https://doi.org/10.1155/2023/4006715

Research Article

Intelligent Testing and Analysis of Dissimilar Steel Welds for Industrial Throttle Flowmeter

Lianghuai Tong,[1] **Hui Xu**,[2] **Xiaojie Xu**,[1] **Huaye Cheng**,[1] **Feng Li**,[3] **and Jiahui Zhou**[4]

[1]*Quzhou Academy of Metrology and Quality Inspection, Quzhou, China*
[2]*Hangzhou Special Equipment Testing and Research Institute, Hangzhou, China*
[3]*Zhejiang Province Metallurgic Products Quality Test Station Co.,Ltd., Hangzhou, China*
[4]*Zhejiang University, Hangzhou, China*

Correspondence should be addressed to Xiaojie Xu; csccpv@163.com

Received 30 August 2022; Revised 6 October 2022; Accepted 16 April 2023; Published 25 May 2023

Academic Editor: Wang Wenyong

The throttling flowmeter used in utility boilers has been used as a measuring instrument for a long time. However, its safety performance, such as welding, lacks enough attention. The manufacturing welds of the throttling flowmeter are welded with austenitic stainless steel and pearlitic heat-resistant steel. Cracks and other defects are generally found in the inspection and detection of the manufacturing welds of throttling flowmeters in service, which have serious potential accidents. To find out the relationship between weldingdissimilar steels and defects, the microhardness indentation method was used to measure the residual stress of welding. Combined with the self-developed calculation software of microscopic indentation residual stress distribution, we used the difference between the color of the indentation area and the surrounding area to quickly measure the indentation strain collected by the microscope using the computer image recognition algorithm, which greatly improved the accuracy and speed of the microscopic indentation residual stress test. The results show that the residual stress at the weld of dissimilar steel is relatively large, and the mechanical property test is generally unqualified. The microindentation residual stress analysis method based on computer image recognition is used to quickly, intuitively, and accurately reveal the direct relationship between the combination of dissimilar steel welding materials and the generation of cracks and other defects.

1. Introduction

The throttling flowmeter used in utility boilers is ubiquitous and is applied in high-temperature and high-pressure environments. The welding seam is generally welded between austenitic stainless steel and pearlitic heat-resistant dissimilar steel. The quality of the welding seam is very important, which is related to whether the throttling flowmeter can be used safely. Many scholars have carried out research on the welding problems of austenitic stainless steel and pearlitic heat-resistant dissimilar steel and the detection and analysis methods of welding defects.

Defects in the form of pores and slag inclusions usually appear in the welds of austenitic stainless steel and pearlitic heat-resistant dissimilar steel, thus causing local strain concentration. In austenitic steel and pearlitic steel, pores of 1 to 3 mm and slag inclusion will reduce the thermal fatigue life of welded joints by more than 10 times [1]. Scanning electron microscopy, transmission electron microscopy, and energy dispersive X-ray spectroscopy adopted by Gotalskii et al. [2] directly observed the weld interface of austenitic steel and pearlitic steel dissimilar alloy welds, and they found that there are two types of interfaces: austenite/martensite and martensite-like/ferrite. Pan and Zhang [3] proposed a method to reduce the width of such martensite interfaces. Schmidova et al. [4] performed hardness measurements in critical areas of pearlitic and austenitic steel welds and translated them into carbon concentration changes, which confirmed the presence of large concentrations and structural inhomogeneity at the fusion line, and such fusion zones

were considered to be unstable [5]. Li et al. [6] and Nelson et al. [7] also analyzed the fusion zone and found it to be characterized by chemical inhomogeneity between the weld and the base metal. Danielewski et al. [8] conducted metallographic analysis of the weld and found that in stainless steel, the heat-affected zone was very narrow, and the growth of austenite grains was not observed, while the slender grains of ferrite formed a discontinuous network around the austenite grain. Another problem with dissimilar steel welds between austenitic steel and pearlitic steel is the presence of decarburization and carbide interlayers, which have an adverse effect on the mechanical properties of the welding seam [9]. Nikulina et al. [10] studied the flash butt welding of high-carbon pearlitic steel and chromium-nickel austenitic steel, and the results showed that layered pearlite colonies containing chromium and nickel and thin austenite interlayers were formed in the weld zone. Vishniakas [11] found that the failure of austenitic electrode welds was viscous, and individual parts were quasibrittle failures, while the failures of ferritic and pearlitic electrode welds were moderate. Elagin et al. [12] found that during long-term high-temperature heating, the formation of nitride particles and grain refinement help to improve the microstructure stability of the weld, inhibit the development of carbide reactions, and reduce the microstructure inhomogeneity of the weld. Therefore, the carbides are distributed more uniformly in the fusion zone with pearlitic steel, thereby reducing the hot brittleness of dissimilar steel welds. Li and Yan [13] used plasma arc welding to weld dissimilar steels of pearlitic steel and austenitic steel and also found that a plate-like martensitic hardened layer was formed in the welding zone, carbon migrated on both sides of the welded joint, and there was a central area of lower hardness. Chu et al. [14] analyzed the failure causes of a certain weld and found that there were intergranular holes and cracks around the main crack, and the main failure mechanism was creep cracking. Dupont [15] summarized the welding of dissimilar steels, where premature failure is usually caused by the following factors: abrupt changes in the microstructure and mechanical properties at the fusion line; large differences in coefficient of thermal expansion (CTE) between ferritic and austenitic alloys; the formation of interfacial carbides, which leads to the formation of creep voids; and the preferential oxidation of ferritic steels near the fusion line.

For the inspection of welds, some standards and manuals have been issued for reference [16, 17]. In addition, except for some traditional inspection methods, new inspection methods are also emerging. Thuvander et al. [18] combined magnetic force microscopy (MFM) with scanning electron microscopy to successfully study austenitic and duplex stainless steel weld metals with different ferrite levels. Hempel [19] conducted an in-depth study on the residual stress state of dissimilar steel welds using X-ray and neutron diffraction and confirmed that the residual stress on the surface of austenitic tubes during sample preparation was greatly affected by the machining technology. Barat et al. [9] experimentally confirmed that the acoustic emission method can be used to detect both typical welding defects in welded joints of different structural grades of steel as well as diffusion interlayers. Lyubimova et al. [20] proposed that the combination of X-ray fluorescence analysis, X-ray diffraction, and microhardness determination with traditional inspection methods (visual or ultrasonic inspection [21], etc.) can improve the operational reliability of dissimilar steel welds. Finite element analysis of welding deformation and residual stress was performed by Bouchard [22] and Rong et al. [23], which can be used to predict the transient behavior of welding deformation and residual stress. Woo et al. [24] and Eisazadeh et al. [25] used the neutron diffraction method to measure the residual stress of different welding layer thicknesses and different welds. The transverse residual stress caused by welding was related to the existence of the martensite phase in dissimilar welds. Rathod et al. [26] concluded that radiography techniques could not detect small inclusions in welds. Therefore, the quality of detection should be checked by additional NDT detection such as ultrasound [27]. Ultrasonic testing requires materials that are acoustically isotropic, whereas austenitic welding materials are highly anisotropic due to the dendritic structure created by the cooling process during welding. Bulavinov [28] addressed this issue with an in-depth understanding of sound propagation in welded structures through elastodynamic simulations to support the assessment of the local structure of the weld. Juengert [29] developed two ultrasonic inspection techniques and validated them on planar specimens with artificial and real defects; both reconstruction techniques gave quantitative inspection results and allowed the determination of defect sizes. Lugin [30] proposed a new method to detect hidden defects using lateral heat flow, which can find hidden defects/cracks that cannot be detected by traditional thermal detection methods. It is a common method to collect welding stress data using Internet of Things technology [31, 32].

Xia et al. [33] developed a method based on microindentation, which could predict the length of the crack as a function of the residual thermal stresses. Mulford et al. [34] described a procedure for extracting simple constitutive parameters from microindentation tests to construct the entire stress versus strain curve. Frutos et al. [35] addressed the determination of residual stresses in sandblasted austenitic steel by ultramicroindentation techniques using a sharp indenter. The results showed good agreement with those obtained by synchrotron radiation on the same specimens. Yonezu et al. [36] proposed a method to evaluate the residual stress and plastic strain of an austenitic stainless steel using a microindentation test. A numerical experiment with the finite element method (FEM) was carried out to simulate an indentation test for SUS316L with various plastic strains (prestrains) and residual stresses. Liu et al. [37] investigated the microstructure and residual stress of laser rapid formed (LRFed) nickel-base superalloy Inconel 718, and residual stress evaluation in microstructure scale by Vickers microindentation method indicates that the residual thermal stress is unevenly distributed in the LRFed sample.

Although dissimilar steel welding of austenitic stainless steel and pearlitic heat-resistant steel is widely used in utility boiler pipes, and domestic and foreign research studies have also been carried out on the problems of dissimilar steel

welding and the detection and analysis methods of welding defects, less research has been done on the welding of dissimilar steels on the throttling flowmeter. For a long time, the throttling flowmeter was only used as a measuring instrument, and its safety performance, such as welding, lacked attention. It was not until 2016 that the explosion of the flowmeter of the Dangyang Power Plant in Hubei caused serious casualties, and all parties in the society began to pay attention to its safety. In this paper, the standard nozzle flowmeter (hereinafter referred to as the nozzle flowmeter) in the throttling flowmeter is taken as the research object, the structure and process of the welding seam of the throttling flowmeter are simulated, and different combinations of dissimilar steel welding materials are selected to make the processing test pieces. Then, the microhardness method combined with the microindentation residual stress distribution calculation software was used to test the residual stress and mechanical properties of the weld of the specimen. The relationship between welding of dissimilar steels and defects is analyzed, and the reasons why throttling flowmeters are pervasive in deficiencies are revealed, which provides a direction for the next step of the flowmeter's structural transformation and processing technology improvement.

2. Problems Existing in the Welding Seam of Throttling Flowmeter

The throttling flowmeters used in utility boilers are similar in structure. The example used in this case is the standard nozzle flowmeter, also known as the nozzle flowmeter, as shown in Figure 1. In order to ensure strength as a pressure-bearing part in a high temperature and high pressure environment, the front and rear clamping rings are generally made of pearlitic heat-resistant steel. To ensure the geometric size of the throttling part and guarantee the measurement accuracy as a measuring element as well as maintain the cleanliness of the surface and prevent oxidation at high temperature, the throttling parts are generally made of austenitic stainless steel. Due to the inconsistent materials of the front and rear clamping rings and the throttling parts, the welding seam has a problem of dissimilar steel welding. To be specific, the welding of austenitic stainless steel and pearlitic heat-resistant steel, and the welding seam is V-shaped.

The author collected 9 out of 53 nozzle flowmeters from 8 thermal power companies in a targeted manner, involving 4 user units and 5 manufacturing units, also taking into account the different materials, operating parameters, different media, and operating times of the nozzle flowmeter. After taking advantage of the convenient conditions of the laboratory, by means of dissection, nondestructive testing, scanning electron microscope, energy spectrum analysis, mechanical performance tests, etc., it was found that none of the nozzle flowmeters survived. To be specific, cracks and other associated defects were found in the manufacturing welds of the nozzle flowmeters that were sampled, as shown in Figures 2–5. The defect rate is 100%, and the high incidence of flowmeters is shocking.

3. The Basic Situation of Welding Specimens of Throttling Flowmeter

To verify the relationship between the welding of dissimilar steel for the throttling flowmeter of a utility boiler and the defects such as cracks commonly found on the weld, the structure and process of the welding seam of the throttling flowmeter were simulated, and different combinations of dissimilar steel welding consumables were selected to make processing test pieces. Then, the residual stress and mechanical properties of the welds of the specimens were tested by appropriate test methods. The structure and welding process of the throttling flowmeter weld were simulated, and different combinations of dissimilar steel welding consumables were selected, including process test pieces 1 to 4 (see Figures 6–9). The hardness and welding residual stress of the specimens under different welding processes were analyzed, and the mechanical properties of the specimens were also tested; the results are shown in Table 1.

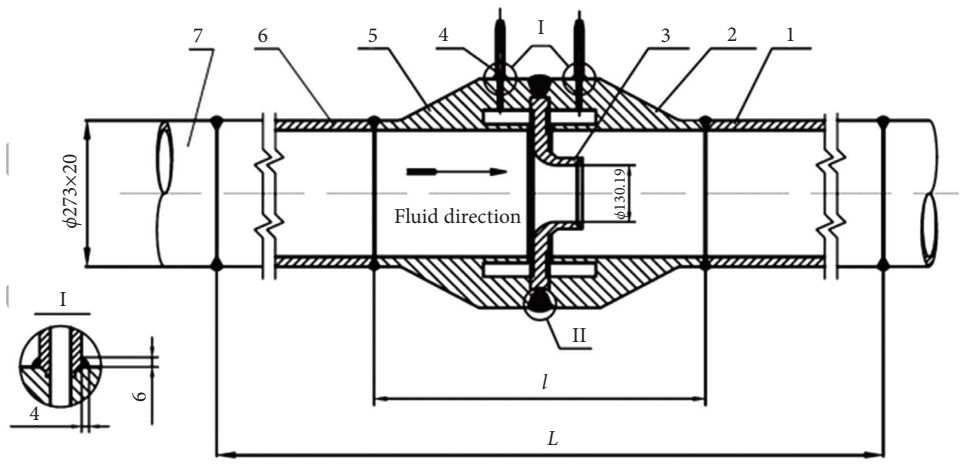
4. Selection of Residual Stress Test Method

The traditional measurement techniques of residual stress of welded components can be roughly divided into two categories: the mechanical release measurement method with certain damage and the nondestructive physical measurement method. These measurement methods are basically only macroscopic measurement methods, which means the test process is difficult to repeat and the experimental data are widely distributed. In addition, mechanical methods, such as the blind hole method will cause greater damage to the measurement samples. In recent 10 years, due to the development of thin film materials and nanotechnology, traditional measurement methods have been unable to meet the experimental requirements, and a new residual stress measurement technology has emerged: the use of microhardness indentation to measure residual stress.

The basic measurement principle of residual stress measured by the microhardness indentation method is that there is a linear relationship between residual stress (strain) and indentation area ratio. When there is tensile stress in the sample, depressions will occur around the indentation, and the area of the indentation will be relatively small. When there is compressive stress, there will be bulges around the indentation, and the area of the indentation will be relatively large. That is why the hardness method for measuring the residual stress of the material is determined by the ratio of the indentation area on the surface of the sample.

According to the theory of OliverWC and PharrGM, the residual stress is sensitive to the amount of metal accumulated around the indentation during the pressing process, so the primary condition for accurate residual stress measurement is to accurately measure the area change of the indentation. The introduction of the parameter indentation area ratio C^2 is given below:

$$c^2 = \frac{A}{A_{\text{nom}}}. \tag{1}$$



1,6-Measuring tube, 2-Rear clamping ring, 3-Nozzle, 4-Pressure tube,
5-Front clamping ring, 7-Process piping

Figure 1: Assembly drawing of $\Phi 273 \times 25$ mm standard nozzle flowmeter.

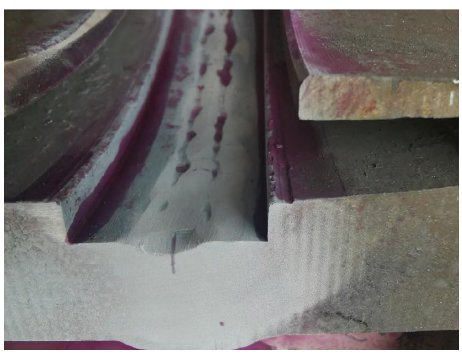


Figure 2: Crack at the root of the weld seam of throttling flowmeter.

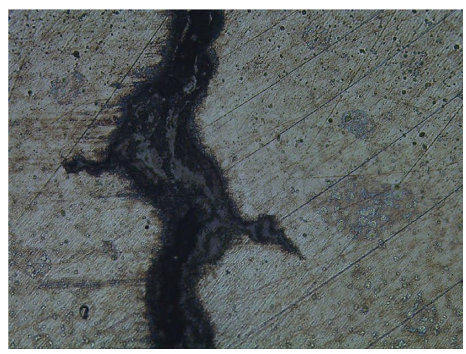


Figure 4: Metallographic diagram of the first layer crack of weld (200 times).

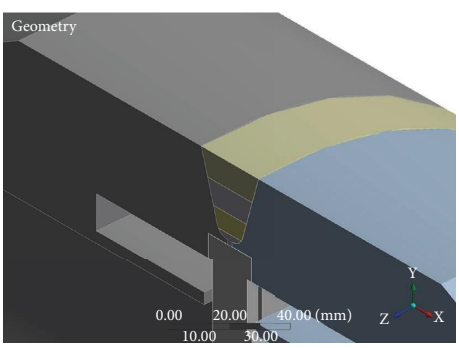


Figure 3: Weld structure diagram of throttling flowmeter.

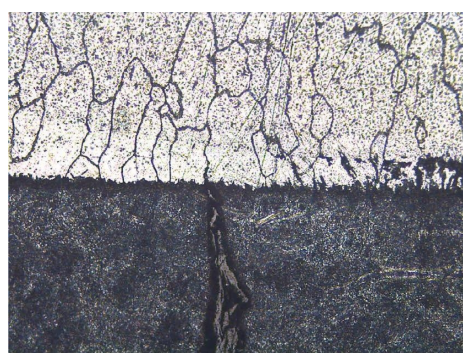


Figure 5: Metallographic diagram of cracks at the junction of the first layer and the second layer of weld seams.

A_{nom} is the projected area of indentation under stress state; A is the area under stress-free state.

The residual stress is finally obtained by calculating the residual strain due to the area change caused by the residual stress.

However, the microhardness indentation method also has some limitations. For example, in order to test the residual stress on the cross section of the weld, it is necessary to take a sample of the weld section, which destroys the integrity of the weld inevitably and only the residual stress in the two-dimensional direction of the cutting plane can be measured, which means the actual residual stress at the test point cannot be completely reflected. Due to the irregularity of the indentation shape affected by residual stress, the accuracy and time-consuming nature of traditional dimensional measurement methods limit the application of this test method in practical engineering. In addition, the

Figure 6: 1# pipe sample.

Figure 7: 2# pipe sample.

Figure 8: 3# pipe sample.

Figure 9: 4# pipe sample.

Table 1: Welding specimen.

No.	Shell material	Shell thickness (mm)	Throttling material	Bottom welding material	Cover welding material	Heat treatment
Specimen 1	12Cr1MoVG	28	304	R31	R317	Yes
Specimen 2	12Cr1MoVG	28	304	ER308	R317	No
Specimen 3	12Cr1MoVG	28	304	ER309	R317	No
Specimen 4	12Cr1MoVG	28	304	ER309	R317	Yes

residual stress is calculated through the relationship between the residual stress and the residual strain. The residual stress between adjacent test points sometimes has a large difference, with peak values and some discontinuities.

To find out the distribution rules of welding residual stress on the welding seam cross section of the throttling flowmeter, the microhardness indentation method was used in this welding residual stress test. The author used the self-developed "microindentation residual stress distribution calculation software" and utilized the difference between the color of the indentation area and the surrounding area. The flow of the algorithm is shown in Figure 10. The computer image recognition algorithm was used to quickly measure the indentation strain collected by the microscope, which greatly improved the testing accuracy and speed of residual stress by microindentation. The software functions included image recognition of indentation topography, automatic calculation of actual area of indentation, theoretical area of indentation, residual strain, and residual stress. It can perform one-click batch identification and calculation of indentation photos.

The key to improving the accuracy and speed of microindentation measurement lies in the rapid and accurate determination of the indentation area of complex shapes that are actually deformed by residual stress. The traditional microscope cannot measure the size and calculate the area of the actual indentation with concave-convex arc features. Using the light and shade difference between the indentation and the surrounding material in the metallographic photo, the computer image recognition method is used to directly extract the sum of the pixels of the dark indentation as the actual indentation area, and the diagonal pixel length between the four endpoints is extracted to obtain the theoretical Indentation area. The ratio of the two can cancel the influence of the pixel unit and the length unit and avoid the difficulty of the actual indentation size measurement and area calculation.

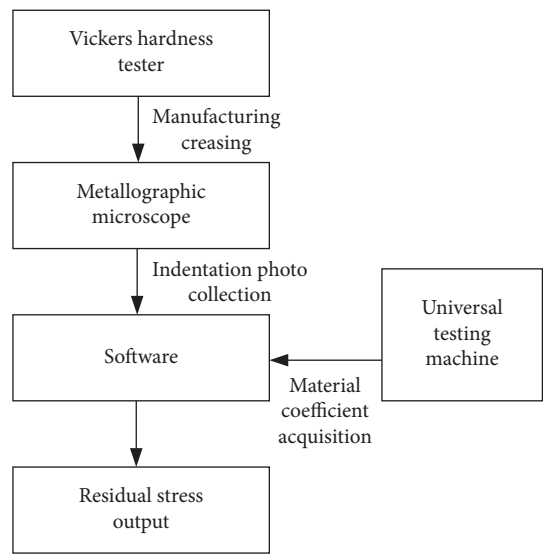


Figure 10: Microindentation residual stress based on computer image recognition algorithm.

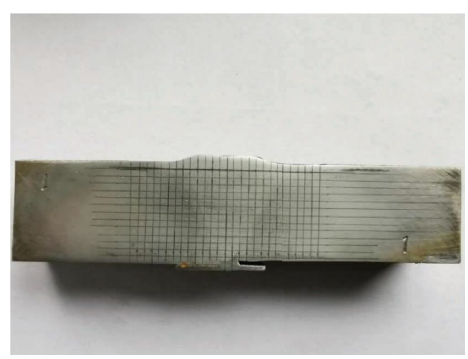


Figure 11: 1#-1 sample.

5. Measurement of Welding Residual Stress on Welding Specimens of Throttling Flowmeter

Weld seam section samples were taken from samples 1#, 2#, 3#, and 4#, respectively, and the samples taken to avoid the arc starting point were recorded as 1#-1, 2#-1, 3#-1, and 4#-1. They were lightly corroded, respectively, and weld distribution can be seen. On the 1#-1 sample, a test point was taken every 2 mm, and a total of 27 columns and 13 rows were taken, totaling 351 points (Figure 11). The microhardness indentation method was used for stress testing, and the microindentation residual stress distribution calculation software was used too. The schematic diagram of the stress distribution is shown in Figure 12. The minimum stress is 97 MPa, and the maximum stress is 291 MPa. On the 2#-1 sample, a test point was taken every 2 mm, with a total of 23 columns and 13 rows, for a total of 299 points (Figure 13. The schematic diagram of stress distribution is shown in Figure 12; the minimum stress is 106 MPa, and the maximum stress is 386 MPa. On the 3#-1 sample, a test point was taken every 2 mm, with a total of 23 columns and 13 rows, for a total of 299 points (Figure 14. The stress distribution diagram is shown in Figure 12. The minimum stress is 122 MPa, and the maximum stress is 398 MPa. On the 4#-1 sample, a test point was taken every 2 mm, with a total of 21

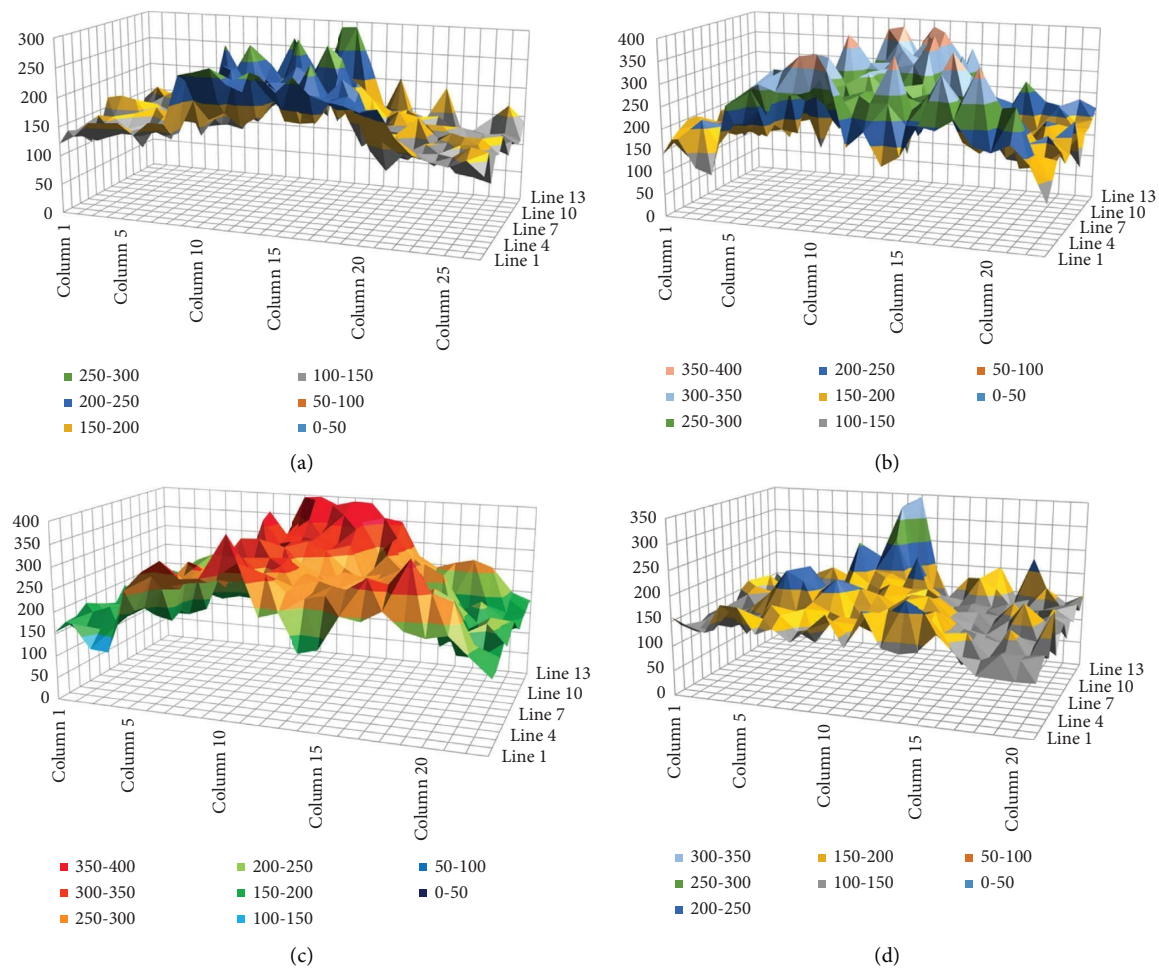


Figure 12: Schematic diagram of stress distribution: (a) sample 1#-1, (b) sample 2#-1, (c) sample 3#-1, and (d) sample 4#-1.

columns and 14 rows, totaling 294 points (Figure 15. The stress distribution diagram is shown in Figure 12. The minimum stress is 102 MPa, and the maximum stress is 346 MPa.

6. Mechanical Properties Test of Welding Specimens of Throttling Flowmeter

According to the requirements of GB/T 228.1-2010, GB/T 2653-2008, and GB/T 229-2007, the mechanical properties of the welds of the four specimens were tested. The test results are shown in Tables 2–4.

7. Analysis of Test Results of Welding Specimens of Throttling Flowmeter

Welding residual stress and mechanical properties test situations are given below:

(1) The welding material of specimen 1 is heat-resistant steel as the base and heat-resistant steel welding rod cover. Except for the welding of dissimilar steel at the welding place between the bottom layer and the stainless steel throttling piece, other welding materials are of the same type. The mechanical properties test reflects that the tensile strength of the welded joint is 539 MPa. The tensile test and impact test results of the welded joint meet the requirements of the specification. However, cracks appeared in some bending specimens of the bottom weld fusion line position of the bending test. The welding residual stress of the specimen is larger than that of the base metal. Most of the residual stress of the weld is between 150 and 250 MPa, and a small part of the point stress is between 250 and 300 MPa. The closer the weld is to the bottom, the greater the welding stress, which is close to 60% of the tensile strength, and there is a certain risk of cracking.

(2) The welding material of specimen 2 is heat-resistant steel and stainless steel transition welding wire as the base and the heat-resistant steel welding rod cover. Due to the transition of welding wire, the heat-resistant steel welding rod will not be in contact with the stainless steel throttling parts. The mechanical performance test reflects that the tensile strength of the welded joint is 500 MPa, and the results of the tensile test, bending test, and the impact

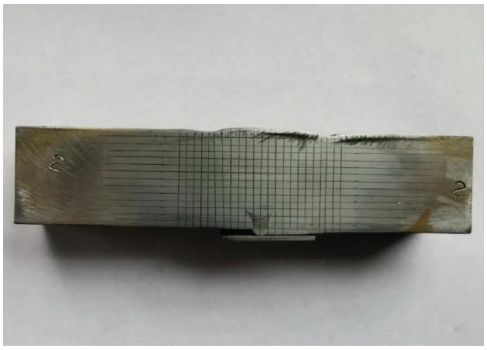


Figure 13: 2#-1 sample.

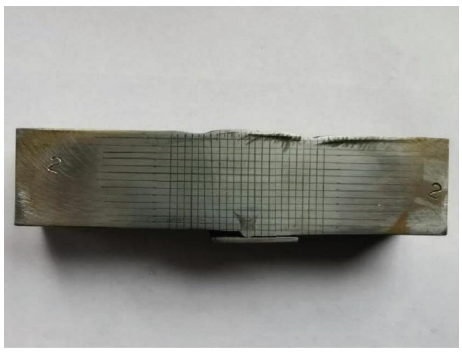


Figure 14: 3#-1 sample.

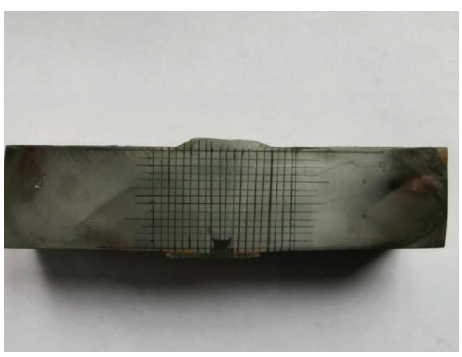


Figure 15: 4#-1 sample.

test of the welded joint all meet the requirements of the specification. The welding residual stress of the specimen is larger than that of the base metal. Since there was no heat treatment performed, most of the residual stress of the weld is between 250 and 350 MPa, and a small part of the point stress is between 350 and 400 MPa. The closer the weld is to the bottom, the greater the welding stress, which is close to 80% of the tensile strength, and there is a greater risk of cracking.

(3) The welding material of specimen 3 is the stainless steel wire as the base and the heat-resistant steel welding rod cover. There is a problem of direct welding of dissimilar steel between the stainless steel wire and the heat-resistant steel welding rod. The mechanical performance test reflects that the tensile strength of the welded joint is 500 MPa, and the tensile test of the welded joint meets the requirements, while the bending test and impact test do not meet the specification requirements. The welding residual stress of the specimen is larger than that of the base metal. Since no heat treatment is performed, most of the residual stress of the weld is between 250 and 350 MPa, and a small part of the point stress is between 350 and 400 MPa. The closer the weld is to the bottom, the greater the welding stress, which is close to 80% of the tensile strength, and there is a great risk of cracking.

(4) The welding material of specimen 4 is the stainless steel wire as the base and the heat-resistant steel welding rod cover. There is a problem of direct welding of dissimilar steel between the stainless steel wire and the heat-resistant steel welding rod. The mechanical performance test reflects that the tensile strength of the welded joint is 500 MPa, and the tensile test of the welded joint meets the requirements, while the bending test and impact test do not meet the specification requirements, but the impact test result is higher than that of specimen 3. The welding residual stress of the specimen is greater than that of the base metal, and the specimen has been heat treated. Most of the residual stress of the weld is between 150 and 250 MPa, and a small part of the point stress is between 250 and 300 MPa. The closer the weld is to the bottom, the greater the welding stress, which is close to 60% of the tensile strength, and there is a certain risk of cracking.

Table 2: Tensile test of welded joints of specimens.

No.	Tensile strength (MPa)		Tensile test of welded joint: Fracture location	
Specimen 1	539	539	Broken base metal	Broken base metal
Specimen 2	503	498	Broken base metal	Broken base metal
Specimen 3	503	498	Broken base metal	Broken base metal
Specimen 4	503	498	Broken base metal	Broken base metal

Table 3: Specimen bending test.

No.	Bending test side bend $D = 40$ mm, $\alpha = 180^\circ$			
	−1	−2	−3	−4
Specimen 1	No opening defect was found on the curved outer surface	A 1.6 mm long crack on the fusion line	A 3.1 mm long crack on the fusion line	Cracks with lengths of 1.5 mm, 2.8 mm, and 1.7 mm on the fusion line
Specimen 2	No opening defect was found on the curved outer surface	No opening defect was found on the curved outer surface	No opening defect was found on the curved outer surface	No opening defect was found on the curved outer surface
Specimen 3	No opening defect was found on the curved outer surface	A 9.3 mm long crack on the fusion line	A 8.1 mm long crack on the fusion line	Cracks with lengths of 2.3 mm, 4.0 mm and 1.0 mm on the fusion line
Specimen 4	No opening defect was found on the curved outer surface	A 9.3 mm long crack on the fusion line	A 8.1 mm long crack on the fusion line	Cracks with lengths of 2.3 mm, 4.0 mm and 1.0 mm on the fusion line

Table 4: Specimen impact test.

No.	Impact test									
	Weld joint 20°C KV2						Heat affected zone 20°C KV2			
Specimen 1	64	35	43	44	67	183	Unbroken	Unbroken	Unbroken	Unbroken
Specimen 2	84	85	66	77	141	233	183	209	223	221
Specimen 3	11	13	14	26	17	157	183	164	189	153
Specimen 4	18	21	35	23	32	197	204	208	187	175

8. Conclusion

Through the testing of welding residual stress and mechanical properties of specimens with four different welding methods, the test results are comprehensively analyzed below:

(1) There are differences in residual stress between the weld passes of the flowmeter welding specimen. The residual stress near the bottom of the weld is larger, and the residual stress near the end of the weld is smaller. The maximum residual stress is located at the root of the weld, and the local strength is close to 60–80% of the tensile strength of the material, indicating that there is a risk.

(2) If the stainless steel throttling parts are directly welded with heat-resistant steel welding materials or stainless steel welding material base and heat-resistant steel welding material cover, the mechanical properties tests are generally unqualified, and the residual stress at the joint of dissimilar steel is large, indicating that this welding process should be avoided.

(3) The stress of the welded seam of the specimen after heat treatment is obviously lower than that of the welded seam without heat treatment, and the test results of mechanical properties are also better than those of the specimen without heat treatment, indicating that the post-weld heat treatment process can effectively improve the performance of the flowmeter and can be popularized and applied.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Quzhou City Science and Technology Project under grant nos. 2022K162, 2021K19.

References

[1] A. F. Malygin, "Thermal cyclic strength of welded joints in an austenitic and a pearlitic steel with weld defects," *Welding Production*, vol. 25, no. 8, pp. 25–30, 1978.

[2] Y. U. N. Gotalskii, V. V. Snisar, and D. P. Novikova, "Methods of reducing the width of the martensite interlayer in the fusion zone of a pearlitic steel with an austenitic weld," *Welding Production*, vol. 28, no. 6, pp. 3–5, 1981.

[3] C. Pan and Z. Zhang, "Characteristics of the weld interface in dissimilar austentic-pearlitic steel welds," *Materials Characterization*, vol. 33, no. 2, pp. 87–92, 1994.

[4] E. Schmidova, L. Benes, and K. Stransky, "Structural stability of austenitic surfacing welds of rail profiles part II (experiment)," *Kovove Materialy-Metallic Materials*, vol. 40, no. 2, pp. 113–123, 2002.

[5] Y. Gotalsky, *Welding of Dissimilar Steels*, Kiev, Tekhnika, 1981.

[6] Y. Li, Z. Zou, and B. Zhou, "Microstructure in the weld metal of austenitic-pearlitic dissimilar steels and diffusion of element in the fusion zone," *Journal of Materials Science & Technology*, vol. 17, no. 3, pp. 338–342, 2001.

[7] T. W. Nelson, J. C. Lippold, and M. J. Mills, "Nature and evolution of the fusion boundary in ferritic-austenitic dissimilar weld metals, Part 1 - nucleation and growth," *Welding Journal*, vol. 78, no. 10, pp. 329–s, 1999.

[8] H. Danielewski, A. Skrzypczyk, M. Hebda et al., "Numerical and metallurgical analysis of laser welded, sealed lap joints of s355j2 and 316l steels under different configurations," *Materials*, vol. 13, no. 24, pp. 5819–5822, 2020.

[9] V. Barat, A. Marchenkov, V. Bardakov, M. Karpova, D. Zhgut, and S. Elizarov, "Features of acoustic emission in tensile testing of dissimilar welded joints of pearlitic and austenitic steels," *Applied Sciences*, vol. 11, no. 24, Article ID 11892, 2021.

[10] A. A. Nikulina, A. I. Smirnov, I. A. Bataev, A. A. Bataev, and A. I. Popelyukh, "Growth of lamellar pearlite in the weld zone between dissimilar steels," *The Physics of Metals and Metallography*, vol. 117, no. 1, pp. 54–60, 2016.

[11] I. Vishniakas, "Special features of breaking the welded connections of the ferritic steels," *Mechanika (Kaunas, Lithuania*, vol. 71, no. 3, pp. 66–71, 1995.

[12] V. P. Elagin, V. Lipodaev, and G. Gordan, "Peculiarities of development of structural heterogeneity in the fusion zone of pearlite steel with austenitic nitrogen-containing weld metal," *Paton Welding Journal*, vol. 2016, no. 8, pp. 23–28, 2016.

[13] M. A. Li and Y. Yan, "Microstructure of welded joint of 1Cr17Mn6Ni5N/Q235 dissimilar steel," *Hot Working Technology*, vol. 40, pp. 168–167, 2011.

[14] Q. Chu, M. Zhang, J. Li, Y. Chen, H. Luo, and Q. Wang, "Failure analysis of a steam pipe weld used in power generation plant," *Engineering Failure Analysis*, vol. 44, pp. 363–370, 2014.

[15] J. N. Dupont, "Review of dissimilar metal welding for the NGNP helical-coil steam generator," 2010.

[16] E. V. Deutsches Institut Fur Normung, *Non-destructive Testing of Welds - Visual Testing of Fusion-Welded Joints*, 2017.

[17] P. O. Moore, *Radiographic Testing in Nondestructive Testing Handbook*, American Society for NDT, Columbus, OH, USA, 3 edition, 2005.

[18] M. Thuvander, L. Karlsson, and O. Kazakova, *Magnetic Force Microscopy as a Tool for Weld Metal Studies*, 2002.

[19] N. Hempel, "Residual stress analysis in girth-welded ferritic and austenitic steel pipes using neutron and X-ray diffraction," *Residual Stress: Icrs*, 2017.

[20] L. L. Lyubimova, R. N. Fisenko, R. B. Tabakaev, A. A. Tashlykov, and A. S. Zavorin, "X-ray investigation of a heterogeneous steel weld," *Materials Science and Engineering A*, vol. 682, pp. 248–254, 2017.

[21] E. M. El-Banna, M. S. Nageda, and M. M. Abo El-Saadat, "Study of restoration by welding of pearlitic ductile cast iron," *Materials Letters*, vol. 42, no. 5, pp. 311–320, 2000.

[22] P. J. Bouchard, "Validated residual stress profiles for fracture assessments of stainless steel pipe girth welds," *International Journal of Pressure Vessels and Piping*, vol. 84, no. 4, pp. 195–222, 2007.

[23] Y. Rong, J. Xu, Y. Huang, and G. Zhang, "Review on finite element analysis of welding deformation and residual stress," *Science and Technology of Welding & Joining*, vol. 23, no. 3, pp. 198–208, 2018.

[24] W. Woo, V. Em, C. R. Hubbard, H. J. Lee, and K. S. Park, "Residual stress determination in a dissimilar weld overlay pipe by neutron diffraction," *Materials Science and Engineering A*, vol. 528, no. 27, pp. 8021–8027, 2011.

[25] H. Eisazadeh et al., "A residual stress study in similar and dissimilar welds," *Welding Journal*, vol. 95, no. 4, pp. 111–119, 2016.

[26] D. W. Rathod, S. Pandey, P. Singh, and R. Prasad, "Experimental analysis of dissimilar metal weld joint: ferritic to austenitic stainless steel," *Materials Science and Engineering A*, vol. 639, pp. 259–268, 2015.

[27] Y. M. Gofman, "Estimation of the reliability of testing welded joints of steam pipelines of thermal power plants," *Russian Journal of Nondestructive Testing*, vol. 39, no. 3, pp. 230-231, 2003.

[28] A. Bulavinov, "Ultrasonic inspection of austenitic and dissimilar welds," *E Journal of Nondestructive Testing*.

[29] A. Juengert, "Advanced ultrasonic techniques for nondestructive testing of austenitic and dissimilar welds in nuclear facilities," *44TH ANNUAL REVIEW OF PROGRESS IN QUANTITATIVE NONDESTRUCTIVE EVALUATION*, vol. 37, 2018.

[30] S. Lugin, "Detection of hidden defects by lateral thermal flows," *NDT & E International*, vol. 56, pp. 48–55, 2013.

[31] T. Wang, J. Li, W. Wei, W. Wang, and K. Fang, "Deep learning-based weak electromagnetic intrusion detection method for the zero touch industrial Internet of Things," *IEEE Network*, 2022.

[32] K. Fang, T. Wang, X. Yuan, C. Miao, Y. Pan, and J. Li, "Detection of weak electromagnetic interference attacks based on fingerprint in IIoT systems," *Future Generation Computer Systems*, vol. 126, pp. 295–304, 2022.

[33] Z. Xia, W. A. Curtin, and B. Sheldon, "A new method to evaluate the fracture toughness of thin films," *Acta Materialia*, vol. 52, no. 12, pp. 3507–3517, 2004.

[34] R. Mulford, R. J. Asaro, and R. J. Sebring, "Spherical indentation of ductile power law materials," *Journal of Materials Research*, vol. 19, no. 9, pp. 2641–2649, 2004.

[35] E. Frutos, M. Multigner, and J. L. González-Carrasco, "Novel approaches to determining residual stresses by ultramicroindentation techniques: application to sandblasted austenitic stainless steel," *Acta Materialia*, vol. 58, no. 12, pp. 4191–4198, 2010.

[36] A. Yonezu, R. Kusano, T. Hiyoshi, and X. Chen, "A method to estimate residual stress in austenitic stainless steel using a microindentation test," *Journal of Materials Engineering and Performance*, vol. 24, no. 1, pp. 362–372, 2015.

[37] F. Liu, X. Lin, G. Yang, M. Song, J. Chen, and W. Huang, "Microstructure and residual stress of laser rapid formed Inconel 718 nickel-base superalloy," *Optics & Laser Technology*, vol. 43, no. 1, pp. 208–213, 2011.

Hindawi
Mobile Information Systems
Volume 2023, Article ID 9916531, 12 pages
https://doi.org/10.1155/2023/9916531

Research Article

An Underwater Target Tracking Algorithm Based on Extended Kalman Filter

Jian Huang

Communication and Countermeasures Division, Sichuan Jiuzhou Electric Group Co. Ltd, Mianyang 621000, China

Correspondence should be addressed to Jian Huang; hj.steven@163.com

Received 4 June 2022; Revised 8 September 2022; Accepted 29 September 2022; Published 3 May 2023

Academic Editor: Wang Wenyong

The technology of ocean monitoring is more advanced while the continuous development of industrial Internet. Unmanned underwater vehicle (UUV) is one of major ways for underwater environment monitoring, which makes high-precision positioning, and tracking of it is one of the key problems and needs to be solved urgently. An underwater acoustic positioning and tracking algorithm based on multiple beacons is proposed to reduce the positioning error of underwater acoustic positioning system caused by uncertain sound speed. The system consists of multiple GPS intelligent buoys floated on the sea surface and acoustic signal generator installed on the UUV. The effective sound speeds between the UUV and different buoys are considered to be unequal and estimated as the state parameters, together with the kinematic parameters of the UUV. Based on the kinematic equations of the UUV, the tracking model is obtained under the framework of the extended Kalman filter. Simulation results show that the proposed algorithm can correct the sound speed and improve the stability and accuracy of underwater acoustic positioning system.

1. Introduction

In the context of the continuous development of industrial Internet, the development of ocean monitoring in the direction of information construction has become inevitable. Due to its wide range of activities, small size, low cost, and many other advantages, unmanned underwater vehicle (UUV) has many important applications in ocean monitoring. In order to ensure the completion of underwater missions and to obtain accurate underwater measurement data, the system is required to obtain accurate position information of the UUV. Therefore, UUV's high-precision positioning and tracking is one of the key technologies for ocean monitoring.

The fact that electromagnetic signals are severely attenuated in water makes GPS hardly used for underwater positioning; however, the good propagation characteristics of sound waves make acoustic positioning an effective alternative. The underwater acoustic positioning system (UAPS) provides position information for underwater target without cumulative errors is an effective underwater positioning method. Classical UAPS includes long baseline (LBL) system [1], short baseline (SBL) system [2], and ultrashort baseline (USBL) system [3, 4].

The error source of UAPS mainly consists of four factors: the calibration error of hydrophones, the propagation time estimation error, the sound speed error, and the random error, where the most damaging error source in LBL system is the sound speed error. In previous studies, sound speed correction techniques usually use the ray acoustics theory to trace the propagation path of acoustic signal, and the straight distance is calculated from the measured propagation time on the basis of the accurately measured sound speed profile (SSP) [5–7]. In order to reduce the impact of sound speed errors on positioning, most researchers usually use different techniques to measure the SSP. However, there is an unavoidable error in the SSP measured by sound velocity profiler (SVP) or derived from conductivity, temperature, and density (CTD) measurements [8]. In addition, SSP in a certain area will also change slowly over time. All these will unfortunately lead to the degradation of sound speed correction techniques based on fixed SSP. In other sound speed

correction algorithms, there are some studies [9, 10] considered the sound speed as an unknown variable and solved the effective sound speed (ESS) in the course of positioning. However, the ESS between each hydrophone in the LBL system to the target is regarded as equal, which is very discrepant from the reality.

Experts and scholars have already done a lot of other researches on underwater acoustic positioning technology. Combining different kinds of traditional UAPSs can provide positioning redundancy and take advantages of each system, the most common of which is the USBL/LBL system [11, 12]. Xu et al. [13] first proposed the application of the difference method to underwater acoustic positioning, where the proposed single-difference method can eliminate long-term systematic errors, and the double-difference method can almost completely eliminate all system errors that depend on depth and space. In [14], a multifunctional system that combines rigidly mounted "fixed" USBL transceivers placed under water surface and "free" cable-mounted LBL stations deployed at a relatively large depth. Moreover, each baseline transponder is equipped with a high-speed communication device to provide a real-time control link for underwater vehicles and navigate them based on the current positioning data.

In UAPS, the Kalman filter (KF) is a commonly used technique to reduce system positioning errors. De Palma et al. [15] measure the distance between a single beacon and the target and locate the target by KF or the extended Kalman filter (EKF). In [10], an unscented Kalman filter (UKF) algorithm based on uncertain least squares (ULS) is proposed for underwater target positioning in moving long baseline systems, under the premise that the underwater sound speed is unknown. In the integrated navigation system, KF is also commonly used to compensate the position information provided by the UAPS to the position information provided by the inertial navigation system [16] or dead reckoning system [17], thereby, eliminating the cumulative error of the latter two systems. In [18], the integration of an USBL acoustic modem and positioning device in a two-parallel EKF multisensory navigation schema for an autonomous underwater vehicle (AUV) is presented. In [19], a constrained form of a square-root unscented Kalman filter (SRUKF) is developed, where the sigma points of the unscented transformation are projected onto the feasible region by solving constrained optimization problems.

In UAPS, the measured values of time of arrival (TOA) or time difference of arrival (TDOA) to the target are usually used to calculate the distance and bearing angle, while in the tightly coupled INS/UAPS integrated navigation system, the distance measurement value is also used as the measurement vector. Such processes usually do not take into account the refraction and multipath effects and assume that the sound speed is a known constant. Other researches [2, 9, 10] regard the underwater sound speed as an unknown variable and estimate it while positioning. However, this kind of method treats the sound speed between the target and different hydrophones as equal, which is an unrealistic assumption. Considering that the uncertain sound speed is one of the serious factors causing underwater acoustic positioning error, this paper proposes a multibeacon based UUV tracking system. The effective sound speed (ESS) between the UUV and different beacons is considered to be unequal and estimated as the state parameters, together with the kinematic parameters of the UUV. The acoustic signal propagation time between the UUV and each beacon is taken as the measurement vector, and the tracking model of UUV is obtained under the framework of EKF. Simulation experiments show that the proposed algorithm is able to correct the sound speed and improve the accuracy and stability of UAPS.

2. GPS Intelligent Buoy System and Geometric Positioning Principle of LBL System

The GPS intelligent buoy (GIB) system [20, 21] consists of several buoys equipped with GPS receivers and submerged hydrophones at the sea surface. The GIB can obtain its own absolute position information through the GPS signal. The UUV carries an acoustic signal generator that periodically broadcasts the acoustic signal. This period is determined by a high precision clock synchronized with the GPS prior to system deployment. Each hydrophone receives acoustic signals and records their arrival time with different latencies. Through the spread spectrum acoustic communication technology, the depth information of the UUV measured by itself can be transmitted to the buoys. The buoy communicates via radio with the central station, where the position of the UUV can be calculated.

The GIB system is depicted in Figure 1. The coordinate system is defined as follows: the north east down (NED) is established by selecting a point in the polygon area surrounded by the N buoys as the origin $O(0,0,0)$. The UUV's coordinate is $T(x,y,z)$, and the coordinates of the hydrophones on each buoy are $GIB_1(x_1,y_1,z_1),\cdots,GIB_N(x_N,y_N,z_N)$, respectively. The distances between the GIBs and the UUV are defined as $R_1,\cdots,R_N$.

In each time period, the acoustic signal generator broadcasts a signal. Since each buoy is time synchronized with the UUV, they can calculate the propagation time after receiving the acoustic signal. Multiply the propagation time by the sound speed can obtain the relative distance between the UUV and each buoy:

$$R_i = \bar{c}_i \cdot t_i, \tag{1}$$

where $\bar{c}_i$ represents the average sound speed, t_i is the propagation time of acoustic signal from the UUV to the i-th buoy, and R_i is the slant range between them. $i = 1,\ldots,N$ are the identification of the buoy, and N is the number of buoys.

After measuring the distance between the UUV and the buoy, the 3-D spatial relationship between them can be expressed as follows:

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

Since the distance between the buoy and the UUV is usually from a few hundred meters to tens of kilometers. Therefore, it can be assumed that the buoys are at the same horizontal plane, which can be expressed as follows:

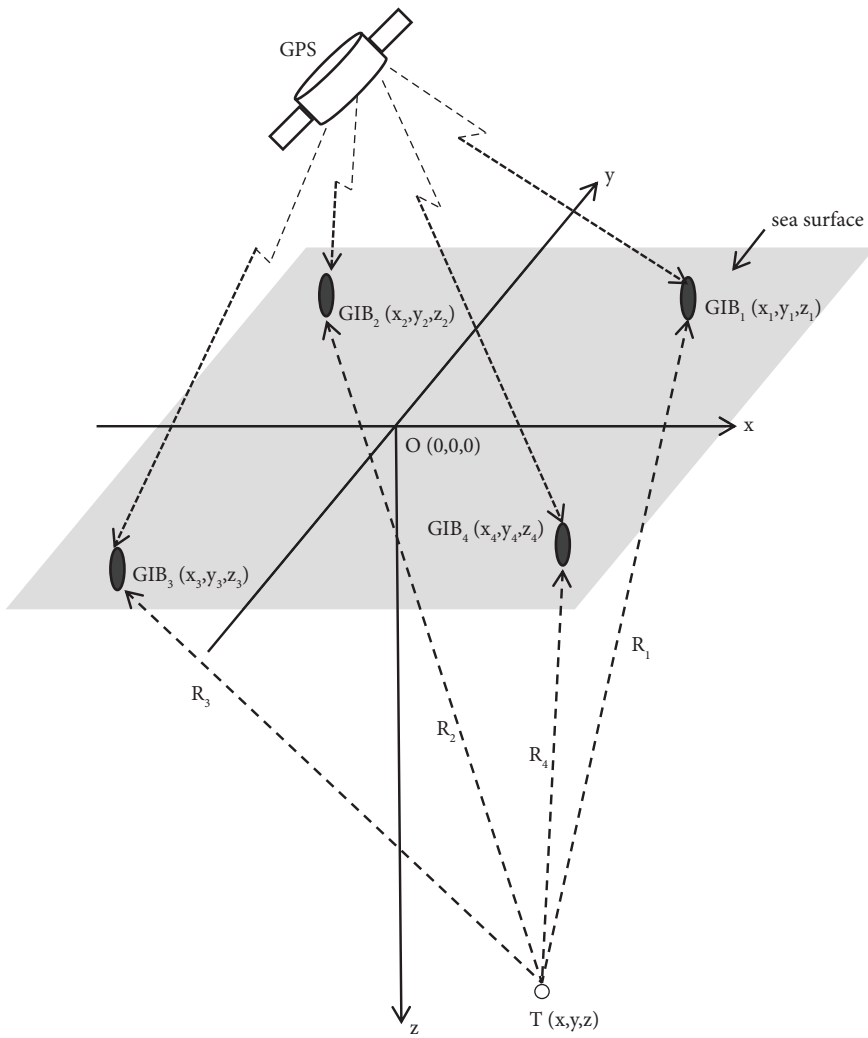


Figure 1: GIB system.

$$z - z_i = z - z_j; i \neq j. \tag{3}$$

Subtracting the equations established by two different buoys can eliminate the quadratic term of the unknown variables, resulting in the following simplified form:

$$\begin{aligned} &x_i^2 + y_i^2 - x_j^2 - y_j^2 - 2\left(x_i - x_j\right)x - 2\left(y_i - y_j\right)y \\ &= \left(\bar{c}_i t_i\right)^2 - \left(\bar{c}_j t_j\right)^2. \end{aligned} \tag{4}$$

Under the premise that the sound speed is known, the geometric positioning model of the LBL system can be described as the following function:

$$\begin{aligned} f_{i,j}(x, y) = {} & 2\left(x_i - x_j\right)x + 2\left(y_i - y_j\right)y \\ & + \left(\bar{c}_i t_i\right)^2 - \left(\bar{c}_j t_j\right)^2 - x_i^2 - y_i^2 + x_j^2 + y_j^2. \end{aligned} \tag{5}$$

Define $\hat{x}$ and $\hat{y}$ are the estimation of the UUV's coordinates x and y. We can calculate them by the least square (LS) method:

$$(\hat{x}, \hat{y}) = \operatorname{argmin} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \left[f_{i,j}(x, y)\right]^2. \tag{6}$$

After obtaining $\hat{x}$ and $\hat{y}$, the coordinate z of the UUV can be calculated by substituting $\hat{x}$ and $\hat{y}$ into (2).

2.1. Underwater Target Tracking Algorithm Based on EKF. When UUV performs underwater surveying or other works, it usually maintains a certain depth and horizontal attitude while performing target detection and information collection by changing its heading angle. Therefore, the 3-D motion of the UUV can be simplified to 2-D form [22], and the depth information of the UUV can be accurately measured by the depth sensor and transmitted to GIBs through the spread spectrum acoustic communication technology. To simplify the description, we limit the movement of the UUV in a plane at a known depth where

$z = z_o$, but the derived solution in this paper can be easily extended to the case where the UUV moves in 3-D space.

2.2. Process Model. In the process of establishing the UUV's kinematic equation, we assume that the coordinates of the UUV are (x, y), the UUV moves at a constant speed V, the angle between V and the x axis is φ, and the derivative of φ is r. The transmitted acoustic signal contains a time stamp $t_k = kh, k \in Z_+$, where h is the emission period of the acoustic signal. After GIBs receiving the signal and transmitting it to the central station, the central station uses the signals with the same time stamp to track the UUV. The discrete time kinematics model of the UUV is expressed as follows:

$$\begin{cases} x(k+1) = x(k) + hV(k)\cos(\varphi(k)), \\ y(k+1) = y(k) + hV(k)\sin(\varphi(k)), \\ V(k+1) = V(k) + \omega_V(k), \\ \varphi(k+1) = \varphi(k) + hr(k) + \omega_\varphi(k), \\ r(k+1) = r(k) + \omega_r(k), \end{cases} \tag{7}$$

where the process noise $\omega_V(k)$, $\omega_\varphi(k)$, and $\omega_r(k)$ are stationary, independent, zero-mean, and Gaussian, with constant standard deviations.

The ESS is defined as the ratio of the slant range between the transmitter and the receiver to the propagation time of the fastest arriving acoustic ray. Yang et al. [9] pointed out that the ESS is related to the spatial relationship (horizontal distance, depth difference, etc.) of the transmitter and receiver. Thus, we consider the ESS between the UUV and each GIB to be unequal since the distance between them is not equal. At the same time, we assume that the ESS keeps equal during the emission interval since the distance the UUV moves during the emission interval of the acoustic signal is much smaller than the slant range between the UUV and GIBs. In this paper, the ESSs between the UUV and each GIB are taken as the state parameter, which are estimated while tracking. Let c_i represents the ESS between the UUV and the i-th GIB, and then, we have

$$c_i(k+1) = c_i(k) + \omega_{c_i}(k); i = 1, \ldots, N, \tag{8}$$

where the process noise $\omega_{c_i}(k)$ is stationary, independent, zero-mean, and Gaussian, with constant standard deviations.

The above kinematic equation can be written as a linear parameter variation system model:

$$\begin{aligned} \mathbf{X}(k+1) &= f(\mathbf{X}(k)) + \mathbf{w}(k) \\ &= \mathbf{A}(\mathbf{X}(\mathbf{k})) \cdot \mathbf{X}(k) + \mathbf{L} \cdot \boldsymbol{\omega}(k), \end{aligned} \tag{9}$$

where $\mathbf{X}$ is the state vector, $\mathbf{A}$ is the transfer matrix for state, $\mathbf{L}$ is the transfer matrix for process noise, and ω is the process noise. From (7) and (8), we can obtain

$$\begin{aligned} \mathbf{X}(k) &= \left[x(k), y(k), V(k), \varphi(k), r(k), c_1(k), \ldots, c_N(k)\right]^T, \\ \boldsymbol{\omega}(k) &= \left[\omega_V(k), \omega_\varphi(k), \omega_r(k), \omega_{c_1}(k), \ldots, \omega_{c_N}(k)\right]^T, \\ \mathbf{A}(\mathbf{X}(k)) &= \begin{bmatrix} \mathbf{A}_1(\mathbf{X}(k)) & \mathbf{0}_{5\times N} \\ \mathbf{0}_{N\times 5} & \mathbf{I}_{N\times N} \end{bmatrix}, \\ \mathbf{A}_1(\mathbf{X}(k)) &= \begin{bmatrix} 1 & 0 & h\cos(\varphi(k)) & 0 & 0 \\ 0 & 1 & h\sin(\varphi(k)) & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & h \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{L} &= \begin{bmatrix} \mathbf{0}_{2\times(3+N)} \\ \mathbf{I}_{(3+N)\times(3+N)} \end{bmatrix}, \end{aligned} \tag{10}$$

and the process noise covariance is

$$\begin{aligned} \mathbf{Q} &= E\left[\boldsymbol{\omega}(k)\boldsymbol{\omega}^T(k)\right] \\ &= \operatorname{diag}\left\{\left[\sigma_V^2, \sigma_\varphi^2, \sigma_r^2, \sigma_{c_1}^2, \ldots, \sigma_{c_N}^2\right]^T\right\}. \end{aligned} \tag{11}$$

2.3. Measurement Model. In the calculation process, the measurement is the only connection between KF and the external environment. Whether the filter is polluted or not mainly depends on the accuracy of the input measurement information. Due to the influence of the external environment, the inaccurate sound speed estimation usually introduces a large error to the UAPS. Here, we take the acoustic signal propagation time between the UUV and each GIB as the measurement. Then, the measurement equation is

$$\mathbf{Z}(k) = h(\mathbf{X}(k)) + \mathbf{v}(k), \tag{12}$$

where $\mathbf{Z}$ is the measurement vector that satisfies the following form:

$$\mathbf{Z}(k) = [z_1(k), \ldots, z_N(k)]^T, \tag{13}$$

where $z_i(k)$ is the propagation time of the acoustic signal between the UUV and the i-th GIB at step k:

$$z_i(k) = \frac{1}{c_i(k)} \sqrt{(x_i - x(k))^2 + (y_i - y(k))^2 + (z_i - z_o)^2} + v_i(k); i = 1, \ldots, N, \tag{14}$$

where the measurement noise $v_i(k)$ is stationary, independent, zero-mean, and Gaussian, with constant standard deviations, and the measurement noise covariance is

$$\begin{aligned} \mathbf{R} &= E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right] \\ &= \operatorname{diag}\left\{\left[\sigma_{v_1}^2, \ldots, \sigma_{v_N}^2\right]^T\right\}. \end{aligned} \tag{15}$$

2.4. Design of EKF. The algorithm adopts the basic equation of the standard EKF, and its time update equations are as follows:

$$\begin{aligned} \hat{\mathbf{X}}(k+1/k) &= \mathbf{F}(\hat{\mathbf{X}}(k)) \cdot \hat{\mathbf{X}}(k), \\ \mathbf{P}(k+1/k) &= \mathbf{F}(\hat{\mathbf{X}}(k)) \cdot \mathbf{P}(k) \cdot \mathbf{F}^T(\hat{\mathbf{X}}(k)) + \hat{\mathbf{L}} \cdot \mathbf{Q} \cdot \hat{\mathbf{L}}^T, \end{aligned} \tag{16}$$

where $\mathbf{P}$ is the covariance of the predicted state, and $\mathbf{F}(\hat{\mathbf{X}}(k))$ and $\hat{\mathbf{L}}$ are process Jacobians at step k:

$$\begin{aligned} \mathbf{F}(\hat{\mathbf{X}}(k)) &= \left.\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right|_{\hat{\mathbf{X}}(k)} \\ &= \begin{bmatrix} \mathbf{F}_1(\hat{\mathbf{X}}(k)) & \mathbf{0}_{5\times N} \\ \mathbf{0}_{N\times 5} & \mathbf{I}_{N\times N} \end{bmatrix}, \\ \hat{\mathbf{L}} &= \mathbf{L}, \end{aligned} \tag{17}$$

where

$$\mathbf{F}_1(\hat{\mathbf{X}}(k)) = \begin{bmatrix} 1 & 0 & h\cos(\hat{\varphi}(k)) & -h\hat{V}(k)\sin(\hat{\varphi}(k)) & 0 \\ 0 & 1 & h\sin(\hat{\varphi}(k)) & h\hat{V}(k)\cos(\hat{\varphi}(k)) & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & h \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{18}$$

The EKF measurement update equations are as follows:

$$\begin{aligned} \mathbf{X}(k+1) &= \hat{\mathbf{X}}(k+1/k) + \mathbf{K} \cdot [\mathbf{Z}(k) - h(\hat{\mathbf{X}}(k+1/k))], \\ \mathbf{K} &= \mathbf{P}(k+1/k) \cdot \mathbf{G}^T(\hat{\mathbf{X}}(k)) \cdot \left[\mathbf{G}(\hat{\mathbf{X}}(k)) \cdot \mathbf{P}(k+1/k) \cdot \mathbf{G}^T(\hat{\mathbf{X}}(k)) + \mathbf{R}\right]^{-1}, \\ \mathbf{P}(k+1) &= [\mathbf{I} - \mathbf{K} \cdot \mathbf{G}(\hat{\mathbf{X}}(k))] \cdot \mathbf{P}(k+1/k), \end{aligned} \tag{19}$$

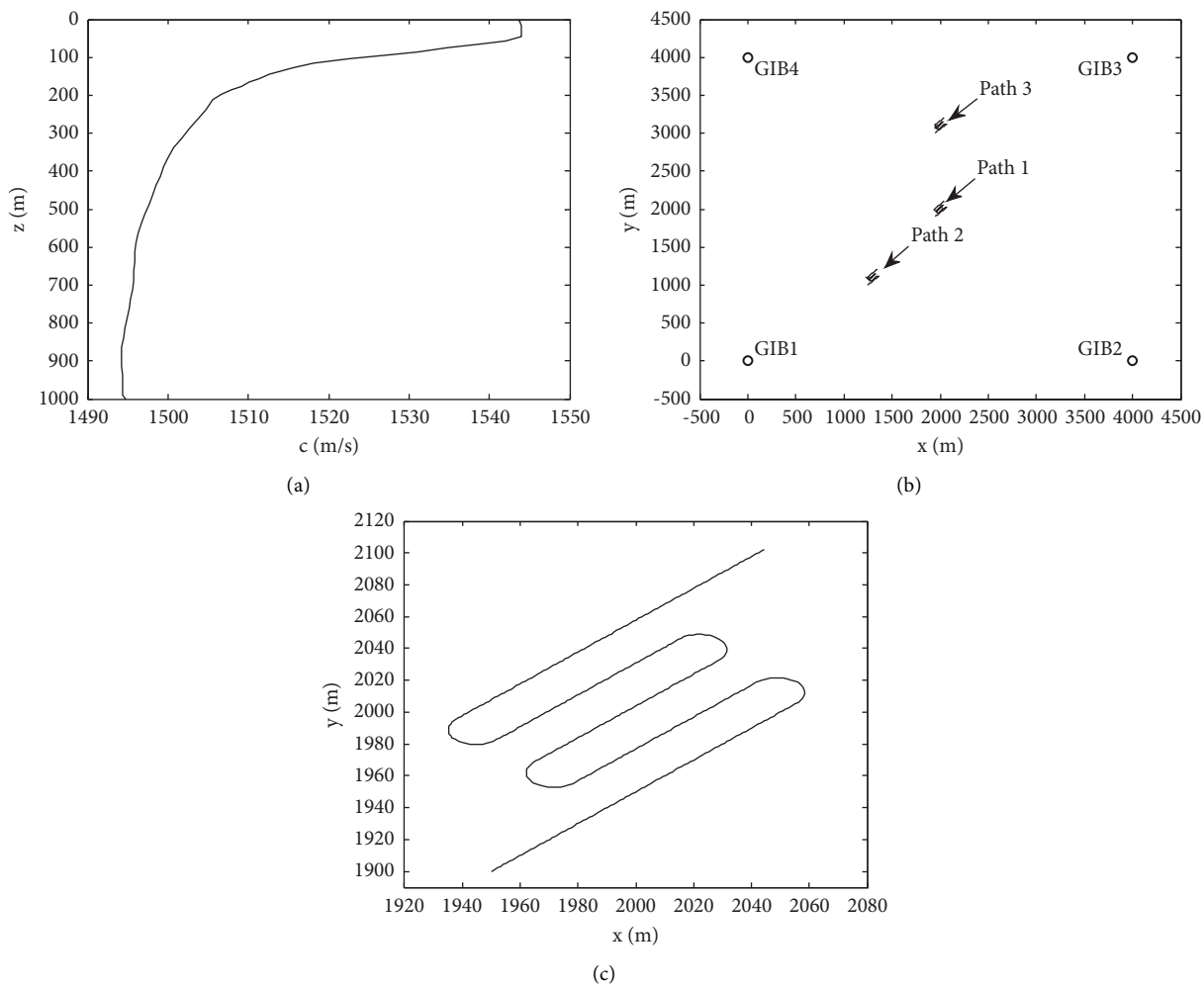


Figure 2: System configuration and the actual path: (a) sound speed profile, (b) system configuration, and (c) detail of the path.

where $\boldsymbol{K}$ is the filter gain, and $\mathbf{G}(\widehat{\mathbf{X}}(k))$ is the measurement Jacobian at step k:

$$\mathbf{G}(\widehat{\mathbf{X}}(k)) = \left.\frac{\partial h(\widehat{\mathbf{X}}(k))}{\partial \mathbf{X}}\right|_{\widehat{\mathbf{X}}(k)} = \begin{bmatrix} -\frac{\left(x_1 - \widehat{x}(k)\right)}{c_1(k)\cdot R_1(k)} & -\frac{\left(y_1 - \widehat{y}(k)\right)}{c_1(k)\cdot R_1(k)} & 0 & 0 & 0 & -\frac{R_1(k)}{c_1^2(k)} & \cdots & 0 \\ -\frac{\left(x_2 - \widehat{x}(k)\right)}{c_2(k)\cdot R_2(k)} & -\frac{\left(y_2 - \widehat{y}(k)\right)}{c_2(k)\cdot R_2(k)} & 0 & 0 & 0 & 0 & -\frac{R_2(k)}{c_2^2(k)}\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\left(x_N - \widehat{x}(k)\right)}{c_N(k)\cdot R_N(k)} & -\frac{\left(y_N - \widehat{y}(k)\right)}{c_N(k)\cdot R_N(k)} & 0 & 0 & 0 & 0 & \cdots & -\frac{R_N(k)}{c_N^2(k)} \end{bmatrix}, \tag{20}$$

Table 1: Initial values setting.

Parameters	Initial values
$\mathbf{X}(0)$	$\left[\,1950\text{m}\ \ 1900\text{m}\ \ 1.5\text{m/s}\ \ \pi/4\text{rad}\ \ 0\text{rad/s}\,\right]^T$
$\hat{\mathbf{X}}(0)$	$\left[\,1970\text{m}\ \ 1880\text{m}\ \ 1.0\text{m/s}\ \ \pi/2\text{rad}\ \ 0\text{rad/s}\,\right]^T$
$\hat{c}_i(0)$	$1500\text{m/s}, i = 1, \ldots, 4$
$\hat{\mathbf{P}}(0)$	$\text{diag}\left\{\left[\,(20\text{m})^2\ \ (20\text{m})^2\ \ (0.5\text{m/s})^2\ \ (0.05\text{rad})^2\ \ (0.005\text{rad/s})^2\ \ (0.01\text{m/s})^2\ \cdots\ (0.01\text{m/s})^2\,\right]\right\}$
σ_V	0.001m/s
σ_φ	0.005rad
σ_r	0.02rad/s
σ_{c_i}	$0.01\text{m/s}, i = 1, \ldots, 4$
σ_{v_i}	$0.0005\text{s}, i = 1, \ldots, 4$

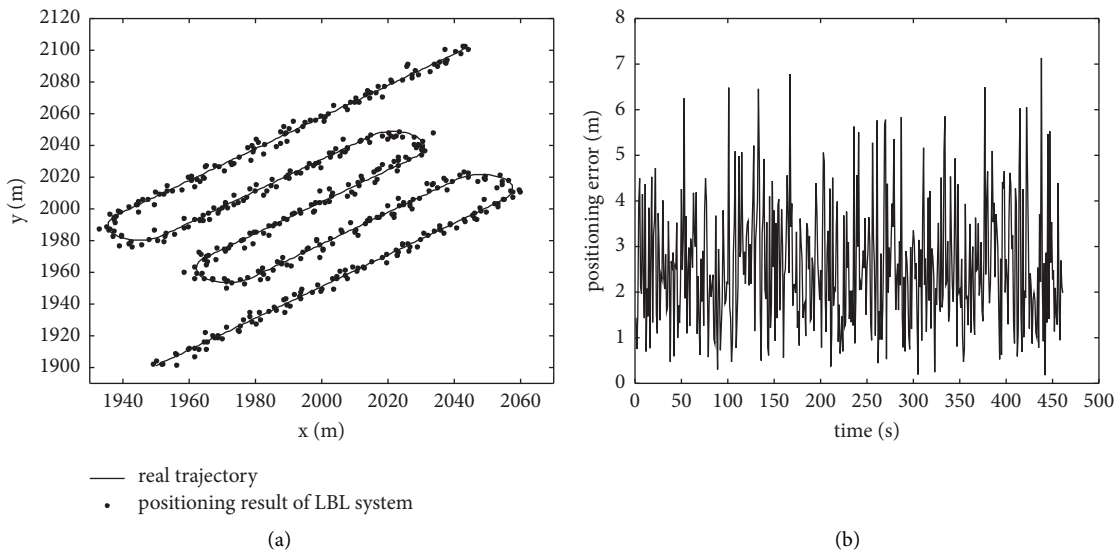


Figure 3: Positioning result of LBL system: (a) estimation path and (b) positioning error.

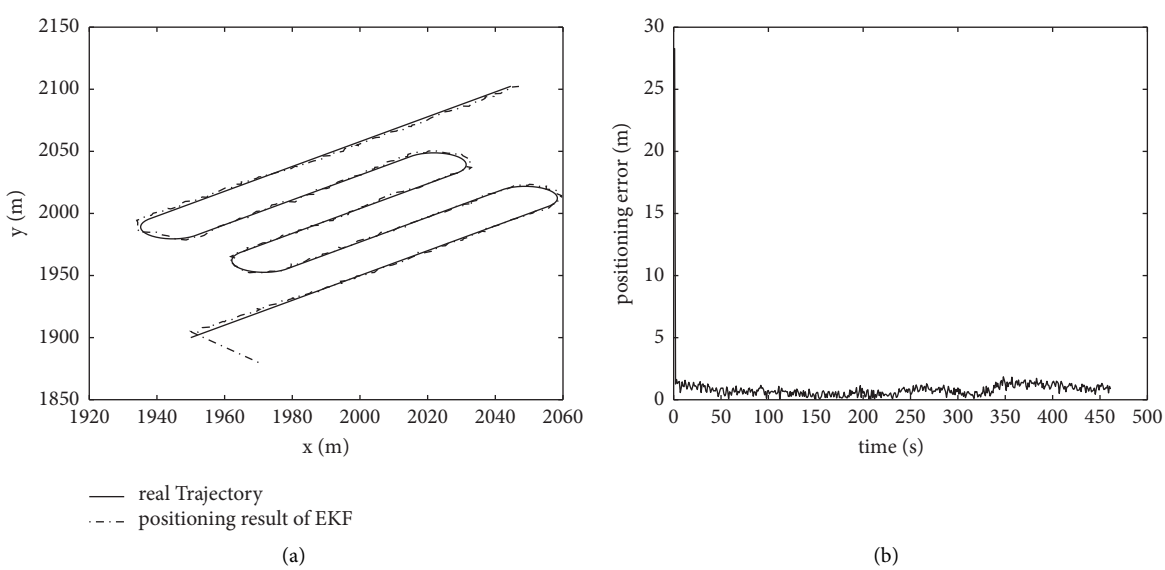


Figure 4: Positioning result of EKF-based algorithm: (a) estimation path and (b) positioning error.

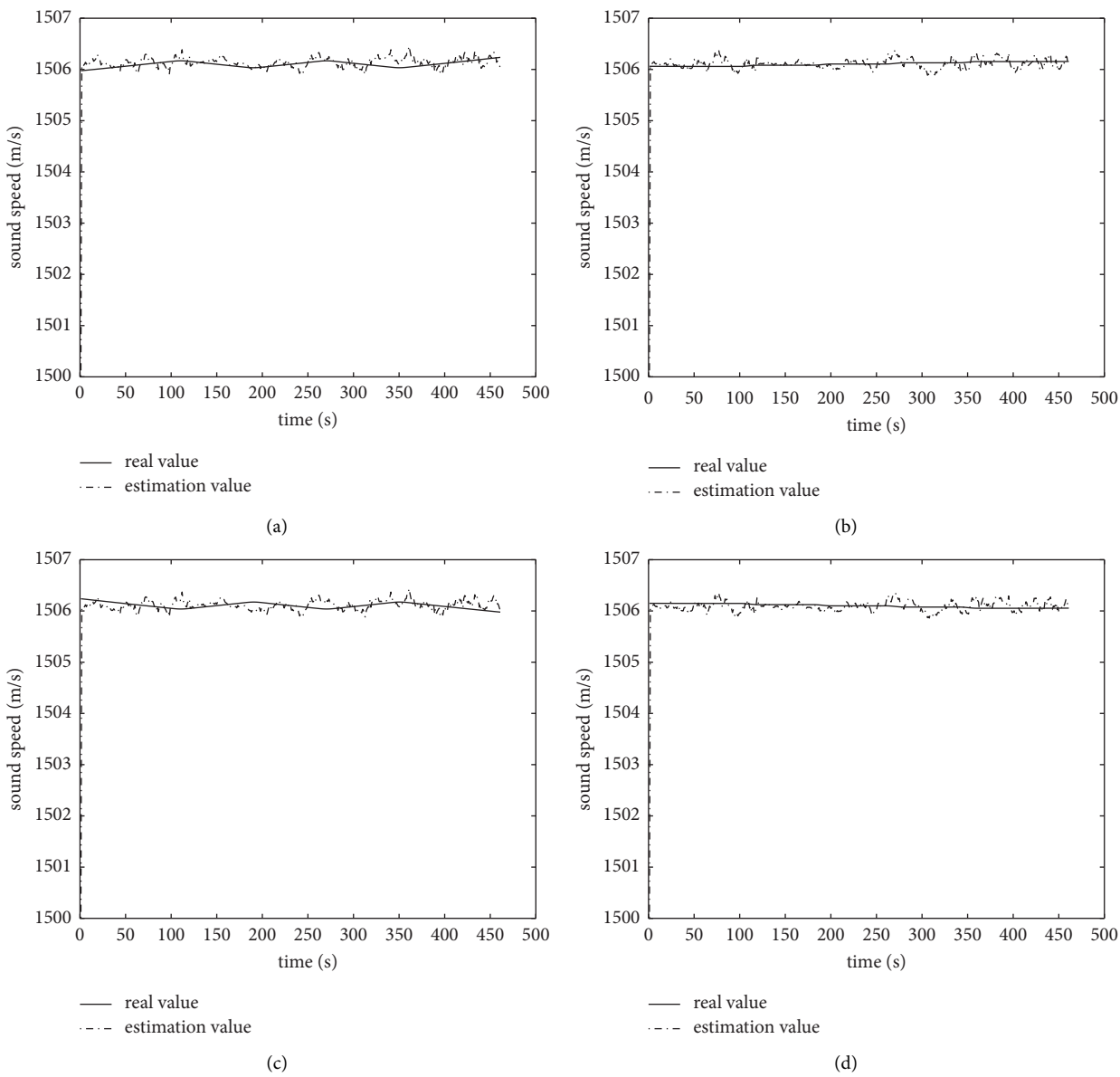


Figure 5: Estimation result of ESSs: (a) c_1, (b) c_2, (c) c_3, and (d) c_4.

where R_i is the slant range between the UUV and i-th GIB:

$$R_i(k) = \sqrt{(x_i - x(k))^2 + (y_i - y(k))^2 + (z_i - z_o)^2}. \quad (21)$$

3. Simulations

This section describes the results of simulations aimed at assessing the efficacy of the algorithms derived. The environmental file used in the simulations is the real measured data of a certain sea experiment, and the sound speed profile is depicted in Figure 2(a). The acoustic signal propagation time is obtained by the BELLHOP [23] model. As shown in Figure 2(b), the GIB system consists of four GIBs, which form a rectangle with the baseline length is 4 km. The UUV is moving at a constant speed in the plane of depth $z_0 = 800m$. Figure 2(c) shows the detail of the real path. The initial values of the EKF algorithm are given in Table 1, where the initial value of first five operating parameters is given as $\mathbf{X}(0)$.

First, we track the UUV travels along the Path 1 depicted in Figure 2(b). The emission period of the acoustic signal is $h = 1s$. It is assumed that the transmitted acoustic signals can be correctly received by all GIBs. For each acoustic signal received by the GIB, the UUV is positioned by the LBL system using geometric method and the EKF based positioning method, respectively. During the LBL positioning, the sound speed is set to 1505.29m/s as the weighted average sound speed [24] at depth 800 m of the sound speed profile as shown in Figure 2(a). The positioning results of the two methods are shown in Figures 3 and 4, respectively. For each positioning result, the distance positioning error is calculated using the following equation:

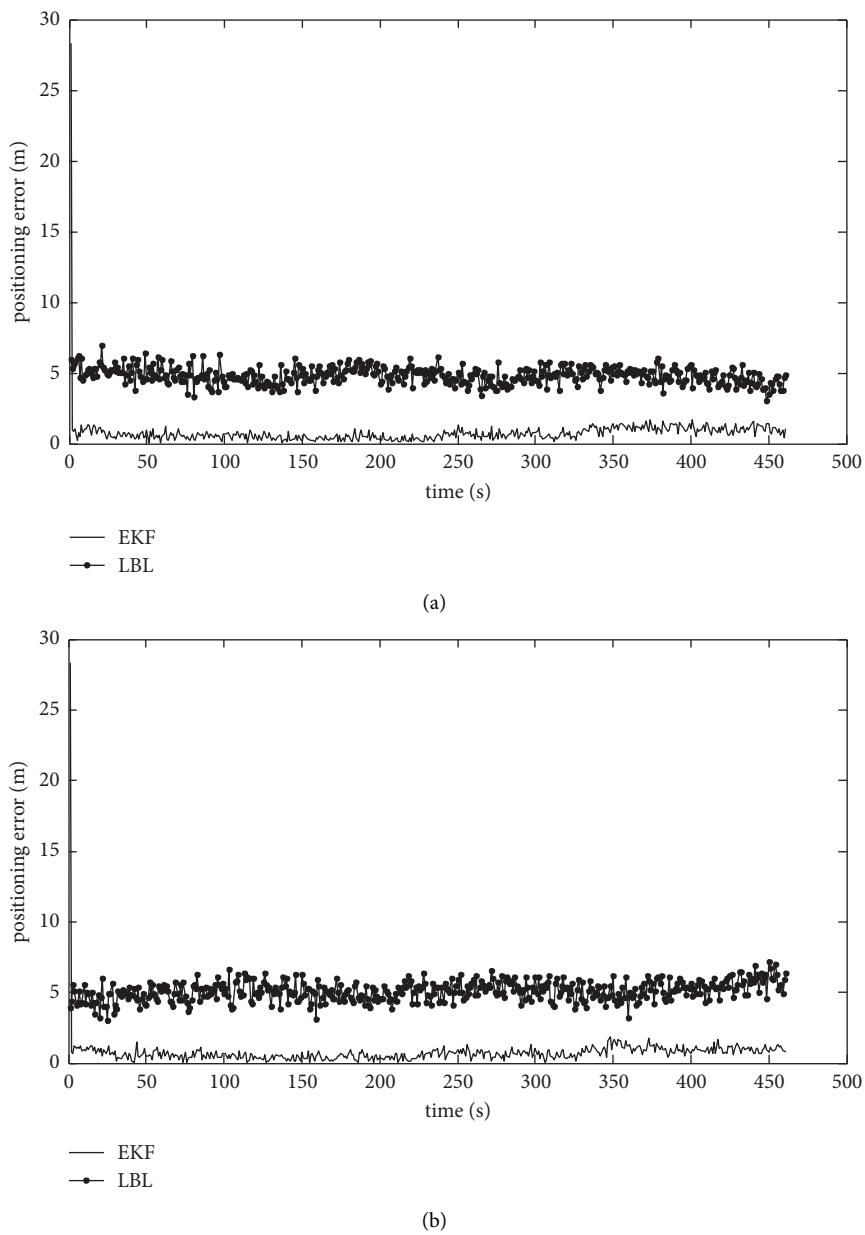


Figure 6: Positioning error of Path 2 and Path 3 by two methods: (a) Path 2 and (b) Path 3.

$$DE = \sqrt{(\hat{x} - x)^2 + (\hat{y} - y)^2}, \tag{22}$$

where (x, y) is the true coordinates of the UUV, and $(\hat{x}, \hat{y})$ is the estimation results. It can be seen that the positioning result obtained by the LBL algorithm deviates far from the real location, and the positioning result calculated by the EKF algorithm is very close to the real path, except for the error of the initial estimation. This is because the accurate sound speed cannot be obtained during the positioning of the LBL system, which results in a large positioning error. However, in the process of EKF-based algorithm, the ESS of acoustic signal propagation can be synchronously estimated (the estimation result of ESSs is shown in Figure 5), and these estimated ESS are used to track the UUV to obtain more accurate positioning result. The final distance positioning error is less than 2 m.

On the basis of the above simulation, the travel path is changed by modifying the start point of the UUV. The first two parameters of $\mathbf{X}(0)$ in Table 1 are set to (1950 m, 1000 m) (Path 2, the distances between the start point and each GIB are 1788.66 m, 3033.04 m, 4147.20 m, and 3346.54 m, respectively) and (1950 m, 3000 m) (Path 3, the distances between the start point and each GIB are 3665.97 m, 3720.12 m, 2416.47 m, and 2332.23 m, respectively). The UUV is tracking by two methods, and the obtained result is shown in Figure 6. The corresponding estimation result of ESSs is shown in Figure 7.

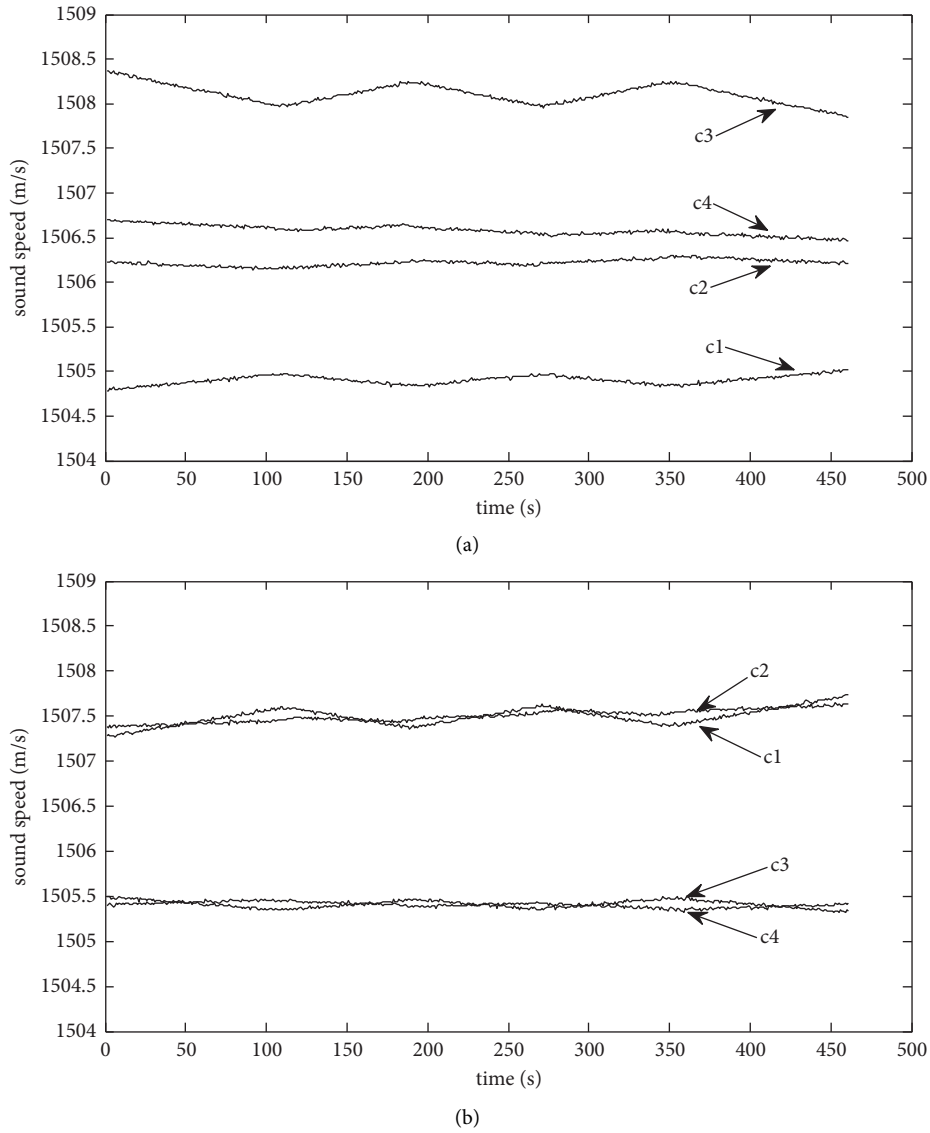


Figure 7: ESSs estimation result of Path 2 and Path 3: (a) Path 2 and (b) Path 3.

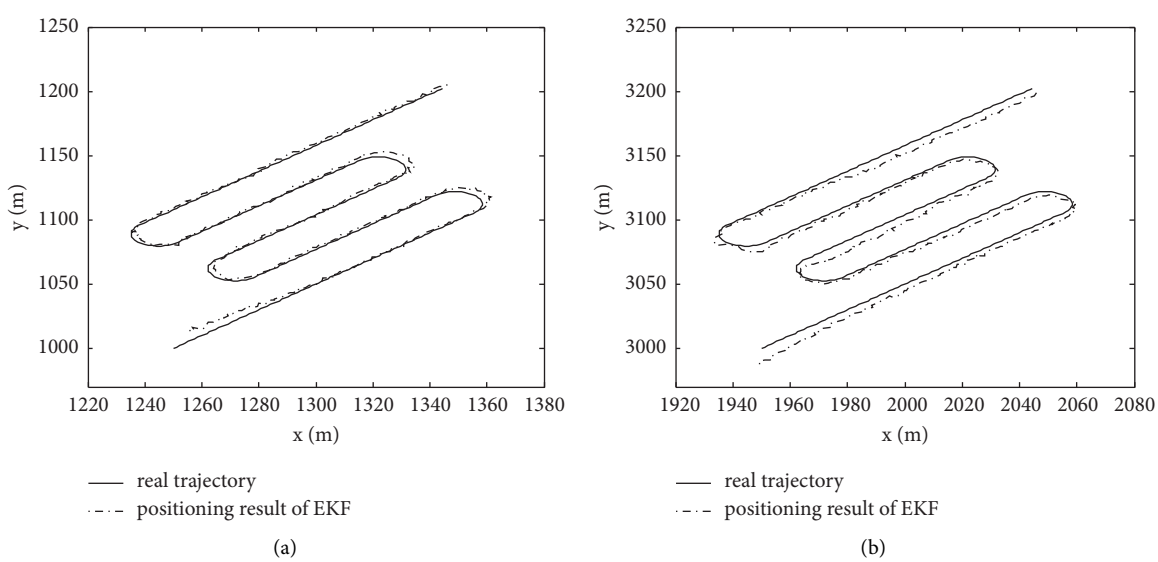


Figure 8: Tracking results of two paths assuming that the ESSs are equal: (a) Path 2 and (b) Path 3.

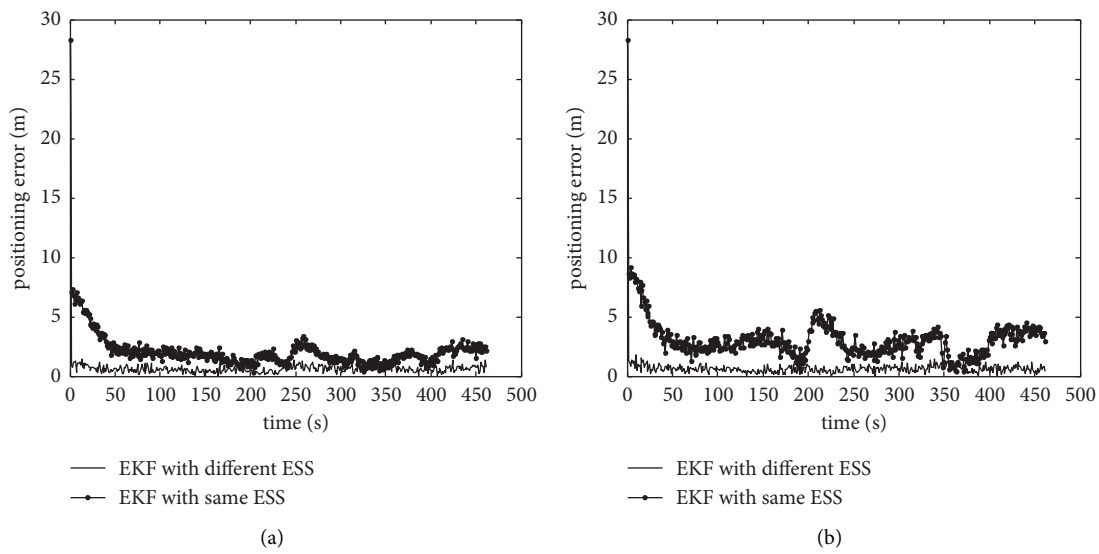


Figure 9: Position error of EKF algorithm with equal or unequal ESSs: (a) Path 2 and (b) Path 3.

For Path 1, the ESSs are almost equal since the difference of the distances between the UUV and each buoy is small. It can be clearly seen from Figure 7 that when the distances between the UUV and each GIB differ greatly, the corresponding ESSs also have a large difference, which indicates that the algorithm proposed in this paper can effectively estimate the ESS in real time. As can be seen from the positioning result from Figure 6, the LBL system cannot obtain an accurate sound speed. Using a single sound speed will cause a large positioning error, and the EKF-based algorithm estimates the ESS while the process of tracking, which can improve the positioning accuracy.

If the classical idea is used, that is, the ESSs between the UUV and different buoys are equal (i.e., $c_1 = \cdots = c_N$). The UUVs in Path 2 and Path 3 are tracked by EKF-based algorithm, and the result is depicted in Figure 8. Figure 9 shows the positioning error of the EKF-based algorithm on the UUV in these two paths, assuming that the ESSs are equal or unequal, respectively. It can be seen in both cases that the positioning results have a certain offset as a whole if a single ESS is used, especially for Path 3. This means that the positioning accuracy has some relationship with the location of the UUV if the ESSs are considered as equal. The algorithm proposed in this paper regards the ESS as unequal, so they can be estimated separately, which allows the EKF-based algorithm to achieve good performance wherever the UUV is located.

4. Conclusion

The classical geometric positioning method of UAPS usually ignores the positioning error caused by inaccurate sound speed. In this paper, an UUV positioning and tracking algorithm based on multibeacon is proposed. By setting the ESSs between the UUV and different buoys unequal, the ESSs are used as the state parameter to be estimated, and the propagation time is used as the measurement. Under the framework of EKF, the kinematics equations of the UUV are utilized, and the corresponding formulas are derived. The algorithm is verified by simulations, and the results show that the proposed algorithm can correct the sound speed estimation and improve the stability and accuracy of the UAPS. By using spread spectrum acoustic communication technology, the proposed method can be conveniently implemented in the application of underwater multitarget positioning and tracking.

Data Availability

The data supporting the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Science and Technology Project of Sichuan Province (2021YFG0014).

References

[1] W. Chen, W. Yan, R. Cui, and H. Cui, "Optimal configuration of USVs for moving long baseline positioning system," in *Proceedings of the International Conference on Advanced Robotics and Mechatronics*, pp. 394–398, IEEE, Macau, China, August 2016.

[2] W. Cheng, "Research for enhancing the precision of asymmetrical SBL system for any vessels," *Ocean Engineering*, vol. 33, no. 10, pp. 1271–1282, 2006.

[3] J. Reis, M. Morgado, P. Batista, P. Oliveira, and C. Silvestre, "Design and experimental validation of a USBL underwater acoustic positioning system," *Sensors*, vol. 16, no. 9, p. 1491, 2016.

[4] F. Zhong and W. Zhou, "Optimal method for USBL underwater acoustic positioning by combining TDOA and

TOA," in *Proceedings of the ACM International Conference on Underwater Networks & Systems*, no. 45, October 2016.

[5] H. Ramezani, H. Jamali-Rad, and G. Leus, "Localization and tracking of a mobile target for an isogradient sound speed profile," *IEEE International Conference on Communications*, pp. 3654–3658, 2012.

[6] H. Ramezani, H. Jamali-Rad, and G. Leus, "Target localization and tracking for an isogradient sound speed profile," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1434–1446, 2013.

[7] T. E. Barnard, F. J. Klein, and L. Resca, "Ray theory results and ray wavefront diagrams for the hyperbolic cosine propagation sound-speed profile," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 4, pp. 938–946, 2015.

[8] S. Jamshidi and M. N. Abu Bakar, "An analysis on sound speed in seawater using CTD data," *Journal of Applied Sciences*, vol. 10, no. 2, pp. 132–138, 2010.

[9] F. Yang, X. Lu, J. Li, L. Han, and Z. Zheng, "Precise positioning of underwater static objects without sound speed profile," *Marine Geodesy*, vol. 34, no. 2, pp. 138–151, 2011.

[10] W. Yan, W. Chen, and R. Cui, "Moving long baseline positioning algorithm with uncertain sound speed," *Journal of Mechanical Science and Technology*, vol. 29, no. 9, pp. 3995–4002, 2015.

[11] A. Caiti, F. D. Corato, D. Fenucci et al., "Experimental results with a mixed USBL/LBL system for AUV navigation," in *Proceedings of the Underwater Communications and Networking*, pp. 1–4, IEEE, Sestri Levante, Italy, September 2015.

[12] P. Batista, C. Silvestre, and P. Oliveira, "Tightly coupled long baseline/ultra-short baseline integrated navigation system," *International Journal of Systems Science*, vol. 47, no. 8, pp. 1837–1855, 2016.

[13] P. Xu, M. Ando, and K. Tadokoro, "Precise, three-dimensional seafloor geodetic deformation measurements using difference techniques," *Earth Planets and Space*, vol. 57, no. 9, pp. 795–808, 2005.

[14] K. G. Kebkal, O. G. Kebkal, S. G. Yakovlev, and R. Bannasch, "Experimental performance of a hydro-acoustic USBL-aided LBL positioning and communication system," *IFAC Proceedings Volumes*, vol. 45, no. 5, pp. 249–254, 2012.

[15] D. De Palma, F. Arrichiello, G. Parlangeli, and G. Indiveri, "Underwater localization using single beacon measurements: observability analysis for a double integrator system," *Ocean Engineering*, vol. 142, pp. 650–665, 2017.

[16] G. T. Donovan, "Position error correction for an autonomous underwater vehicle inertial navigation system (INS) using a particle filter," *IEEE Journal of Oceanic Engineering*, vol. 37, no. 3, pp. 431–445, 2012.

[17] Y. Guo, C. Li, D. Zhang, Z. Tiehu, S. Rui, and Z. Yanshun, "The integrated navigation method by underwater towing body based on dead reckoning/hydroacoustic positioning system," *Marine Geology Frontiers*, vol. 31, no. 6, pp. 63–67, 2015.

[18] E. Guerrero-Font, M. Massot-Campos, P. L. Negre, F. Bonin-Font, and G. O. Codina, "An USBL-aided multisensor navigation system for field AUVs," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, September 2017.

[19] S. Yousefi, X. W. Chang, and B. Champagne, "Mobile localization in non-line-of-sight using constrained square-root unscented kalman filter," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 2071–2083, 2015.

[20] H. G. Thomas, "GIB buoys: an interface between space and depths of the oceans," in *Proceedings of the Workshop on Autonomous Underwater Vehicles*, pp. 181–184, IEEE, Cambridge, MA, USA, August 2002.

[21] A. Alcocer, P. Oliveira, and A. Pascoal, "Study and implementation of an EKF GIB-based underwater positioning system," *Control Engineering Practice*, vol. 15, no. 6, pp. 689–701, 2007.

[22] Z. Yan, S. Peng, J. Zhou, J. Xu, and H. Jia, "Research on an improved dead reckoning for AUV navigation," in *Proceedings of the Chinese Control and Decision Conference*, pp. 1793–1797, Xuzhou, China, May 2010.

[23] M. B. Porter, *The BELLHOP Manual and User's Guide: Preliminary Draft*, Heat, Light, and Sound Research, Inc, La Jolla, CA, USA, 2011.

[24] C. Yi, W. Ren, and C. Wang, "Analysis on error of secondary acoustic positioning system," *Oil Geophysical Prospecting*, vol. 44, no. 2, pp. 136–139, 2009.

Hindawi
Mobile Information Systems
Volume 2023, Article ID 1594622, 11 pages
https://doi.org/10.1155/2023/1594622

Research Article

Network Intrusion Anomaly Detection Model Based on Multiclassifier Fusion Technology

Feilu Hang,[1] **Wei Guo**,[1] **Hexiong Chen**,[1] **Linjiang Xie**,[1] **Xiaoyu Bai**,[2] **and Yao Liu**[2]

[1]*Information Center, Yunnan Power Grid Co. Ltd., Kunming, Yunnan 650034, China*
[2]*Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China*

Correspondence should be addressed to Xiaoyu Bai; 202121090126@std.uestc.edu.cn

Received 8 August 2022; Revised 14 September 2022; Accepted 28 September 2022; Published 8 April 2023

Academic Editor: Sai Zou

With the increasing development of the industrial Internet, network security has attracted more and more attention. Among the numerous network security technologies, anomaly detection technology based on network traffic has become an important research field. At present, a large number of methods for network anomaly detection have been proposed. Most of the better performance detection methods are based on supervised machine learning algorithms, which require a large number of labelled data for model training. However, in a real network, it is impossible to manually filter and label large-scale traffic data. Network administrators can only use unsupervised machine learning algorithms for actual detection, and the detection effects are much worse than supervised learning algorithms. To improve the accuracy of the unsupervised detection methods, this study proposes a network anomaly detection model based on multiple classifier fusion technology, which applies different fusion techniques (such as Majority Vote, Weighted Majority Vote, and Naive Bayes) to fuse the detection results of the five best performing unsupervised anomaly detection algorithms. Comparative experiments are carried out on three public datasets. Experimental results show that, in terms of RECALL and AUC score, the fusion model proposed in this study achieves better performance than the five separate anomaly detection baseline algorithms, and it has better robustness and stability, which can be effectively applied to a wide range of network anomaly detection scenarios.

1. Introduction

With the advancement of network technology, an increasing number of users are connected to the Internet. As the Internet grows in size, the threats from attackers and criminal enterprises have also increased accordingly. The increasing number of these threats makes firewall and intrusion detection systems (IDS) become one of the foundational technologies of network security. However, intrusion detection systems on the market primarily exploit the signatures of known attacks to detect anomalies. Such systems require frequent updates of the rule database and signatures, and unknown attacks cannot be detected. Anomaly detection systems can effectively discover the attack behavior by modelling normal behaviors and detecting anomalous behavior which are different from normal behavior. Although anomaly detection systems are conceptually attractive, there are many technological issues to overcome before they can be widely adopted, such as high false positive rates and the inability to scale to gigabit speeds. At present, anomaly detection methods are mainly divided into two categories: anomaly detection based on statistics and anomaly detection based on machine learning.

Haystack [1] is one of the earliest examples of a statistical anomaly intrusion detection system. Haystack defines a series of values that are considered normal for each feature. If, during a session, a feature is outside the normal range, the topic's anomaly score increases, and it is designed to detect six types of intrusions. The Statistics Package Anomaly Detection Engine (SPADE) [2] is a statistical anomaly detection system. SPADE was one of the first systems to propose the use of the anomaly scoring concept to detect

port scans, rather than using the traditional method of looking at p attempts within q seconds. In [2], the authors use a simple frequency-based method to calculate the "anomaly score" of a packet. The less views a given packet has, the higher its anomaly score is.

Machine learning is designed to answer many of the same questions as statistics. However, unlike statistical methods, which tend to focus on understanding the processes that generate data, machine learning techniques focus on building a system that improves performance based on previous results. The Bayes Network is a graph model that encodes the probabilistic relationships between target variables. When combined with statistical techniques, Bayes Network has advantages in data analysis [3]. Some researchers have taken inspiration from Bayes statistics to create anomaly detection models. Valdes [4] has developed an anomaly detection system that uses a naive Bayes Network to perform intrusion detection on traffic bursts. Bayesian techniques are also often used to classify and suppress false positive regions. Kruegel et al. [5] proposed a multisensor fusion method in which the outputs of different IDS sensors are aggregated together to generate a single alarm. While intrusion detection using the Bayes Network is effective in some applications, its limitations should be considered in practice. Unfortunately, typical network structures are complex, and choosing an accurate model of behavior is a daunting task.

To solve the problem of high-dimensionality datasets, the researchers developed a dimensionality reduction technique called principal component analysis (PCA) [6]. Shyu et al. [7] proposed an anomaly detection scheme in which PCA is used as an anomaly detection scheme and applied to reduce the dimension of audit data and obtain a classifier. Extreme Learning Machine (ELM) and NSL Knowledge Discovery Data Mining [8] have been identified as criteria for evaluating network intrusion detection mechanisms. The researchers also implemented a random forest classifier on an IDS dataset sample [9, 10]. In addition, the available datasets need to be continuously updated based on the dynamic characteristics of the malware attack.

Some researchers have proposed implementing deep learning models and deep neural networks (DNN) to develop flexible and dynamic IDS systems that can successfully detect and classify capricious cyberattacks [10–13]. In harmony with the use of DNN, convolutional neural network has been identified as an advanced and superior technique for extracting features from intrusion datasets for classification [14]. Since this method [15] provides visual detection of network intrusions, it can be justified as a real-time solution for deploying intrusion detection systems [10]. The authors in [16] explored the different deep learning methods used in IDS and proposed comparisons and analyses. The authors of [17] detected intrusion based on Spark-Chi-SVM technology.

The authors in [18–21] propose several hybrid classification models that classify datasets using naturally inspired algorithms such as cuckoo search, BAT, fireflies, and genetic algorithms [10]. The differential performance of different classifiers is related to several factors, such as the statistical distributional properties of the categorical data, prior knowledge, the size of the training data samples, and the structure of the classifier itself. In anomaly detection, different intrusion identification results can be obtained using different classifiers, and these results are often highly complementary. Therefore, the fusion of the decisions of multiple classifiers can effectively improve the detection effect of anomalies. Moreover, the fusion of multiple classifiers can also improve the robustness of the classification system. In the selection of fusion methods, Dainotti et al. [22] summarize some of the common fusion methods, including majority voting, weighted majority voting, Naive Bayes, behavioral knowledge space (BKS), Wernecke's (WER) method, and the Oracle (ORA) method.

In view of this, this study proposes a network intrusion anomaly detection model MF based on multiclassifier fusion. Three fusion techniques (majority voting, weighted majority voting, and Naive Bayes) were used to fuse the decisions of different anomaly detection methods based on unsupervised learning that have performed well in recent years, such as Lightweight Online Detector of Anomalies (LODA), AutoEncoder, PCA, Histogram-based Outlier Score (HBOS), and iForest (Isolation Forests). The major contributions and findings of this study are listed below:

(1) In real-world network environments, the large scale of labelled data is rare, and the supervised learning methods are not suitable. However, using only unsupervised learning methods leads to poor performance and low accuracy of the model due to the lack of guidance from labelled data. The MF model proposed in this study provides a framework to ensemble various anomaly detection classifiers to form heterogeneous or homogeneous modelling backgrounds for final decision-making and significantly improves the overall detection performance; in other words, it improves the RECALL and AUC metrics.

(2) In real-world network environments, intrusions change rapidly and new attacks are endlessly emerging. Different detection classifiers tend to be biased in their detection performance, and perhaps one of the anomaly detection classifiers would work well for a particular intrusion detection job. The MF model proposed in this study can remedy the shortcomings of a single anomaly detection method and realize the complementarity of different detection methods by using multiclassifier fusion technology.

The remainder of the study is organized as follows. Section 2 presents the anomaly detection methods based on unsupervised learning used in this study. Section 3 describes the multiclassifier fusion technique, as well as the MF model architecture and procedure. Section 4 presents the three anomaly detection datasets. Section 5 presents the experimental configuration, discussion, and findings. Finally, Section 5 presents our conclusions and future work.

2. Anomaly Detection Methods Based on Unsupervised Learning

This section introduces five anomaly detection algorithms based on unsupervised learning that have the best detection effect and fast detection speed and are suitable for large-scale networks. Unsupervised learning methods have no training or learning phase and do not require data labelling, which is more suitable for the detection requirements of a real network. And these five detection methods all assign an outlier score to each of the data points. This is an advantage compared to binary output methods because, additionally, the outlier score allows for estimating the reliability or certainty of the prediction.

2.1. LODA. A lightweight online detector of anomalies (LODA) [23] can be used to quickly process a large stream of data generated by continuously changing behaviors. It consists of a collection of k one-dimensional histogram $\{h_i\}^k$; each histogram approximates the probability density of input data projected into a single projection vector $\{w_i \in R^d\}_{i=1}^k$. The output of LODA is the average of the logarithms of probability density on multiple histograms, thereby improving the performance of a single histogram detector.

The input to LODA is sample x, and the output $f(x)$ represents the average of the logarithms of the estimated probabilities projected by the sample into different projection vectors. Using $\hat{p}_i$ represents the probability estimated by the ith histogram, W_i represents the corresponding projection vector. The output $f(x)$ of LODA can be written as follows:

$$f(x) = -\frac{1}{k}\sum_{i=1}^{k}\log \hat{p}_i\left(x^T w_i\right). \tag{1}$$

If different histograms are considered to be independent of each other, (2) can be used to integrate the probability densities of multiple histograms:

$$f(x) = -\log p\left(x^T w_1, x^T w_2, x^T w_i\right), \tag{2}$$

where $(x^T W_1, x^T W_2, \ldots, xW_k)$ represents the joint probability of the projection. (2) shows that the output of LODA is proportional to the negative log likelihood of the sample, which means that the less likely the sample is to appear, the higher its outliers are.

2.2. Autoencoder. An autoencoder [24] is an unsupervised learning model that essentially uses a neural network to generate a low-dimensional representation of a high-dimensional input. Autoencoder is similar to principal component analysis (PCA), but autoencoder utilizes a nonlinear activation function, thus overcoming the limitation that PCA can only do feature linear transformation.

An autoencoder is composed of two parts, an encoder and a decoder. The encoding procedure performs a dimension reduction on the training data and projects training data in the latent space, where the features of the training data are preserved. A decoder can be constructed to recover the data using the features in the latent space. The difference between the original input vector and the reconstruction vector is called the reconstruction error. If the features of the sample are all numerical variables, the mean squared error (MSE) or the mean absolute error (MAE) can be used as the reconstruction error. For example, the input sample is $X = (X_1, X_2, \ldots, X_m)$. The result of autoencoder reconstruction is $X^R = (X_1^R, X_2^R, \ldots, X_m^R)$. The reconstruction error is MSE $1/m\sum_{i=1}^{m}(X_i - X_i^R)^2$.

The reconstruction error was used as the anomaly score. Data points with high reconstruction are considered to be anomalies. Only data with normal instances are used to train the autoencoder. After training, the autoencoder will reconstruct normal data very well, while failing to do so with anomaly data, which the autoencoder has not encountered.

2.3. PCA. Principal component analysis (PCA) [25] is the most common method of data dimensionality reduction. Generally, in anomaly detection scenarios, noise, outlier, and anomaly are different representations of the same thing. Since PCA can recognize noise, it can naturally detect anomalies. PCA maps the data to low-dimensional feature space and then checks the deviation of each data point from the other data on different dimensions of the feature space.

For a feature vector e_j, the deviation degree d_{ij} of the data sample x_i in direction j can be calculated by using

$$d_{ij} = \frac{\left(x_i^T \cdot e_j\right)^2}{\lambda_j}, \tag{3}$$

where λ_j mainly plays a role in normalization, which can make the deviation degrees in different directions comparable. After that, the anomaly score of sample x_i is calculated, as shown in Equation (4). If the score is greater than the threshold, the sample x_i is judged as an anomaly:

$$\text{Score}\left(x_j\right) = \sum_{j=1}^{n} d_{ij}. \tag{4}$$

2.4. HBOS. HBOS (Histogram-based Outlier Score) [26] is a combination of univariate methods that cannot model dependencies between features, but is faster and friendly to large datasets. HBOS calculates an outlier score by creating a univariate histogram for every single feature of the dataset. It assumes that the features are independent. The drawback of assuming feature independence becomes less severe when the dataset has a high number of dimensions due to a larger sparsity.

Two different methods can be used to construct histograms: the static bin width and the dynamic bin width [27].

(1) The static bin width: data are grouped using bins of the same width. A rectangle is drawn in each interval, whose height is proportional to the number of points that fall into the interval. Equal binning assumes that

each bin is equally likely, but this assumption is usually not met.

(2) The dynamic bin width: the width of bins depends on the values of data and may not necessarily be equal to the specified number of bins. In the case of long-tail distributions with repetitive integers, some bins may contain more values than specified.

In both the static and dynamic cases, it is necessary to specify the number of bins k. It is recommended that the k value should be equal to the square root of all data points. The height of every single bin in the histogram represents the density estimation. To ensure an equal weight for each feature, the histograms are normalized in such a way that the maximum height of the bin would be equal to one. Then, calculated values are inverted so that anomalies have a high score and normal instances have a low score. The anomaly score of each sample x_i is calculated according to Equation (5). The higher the HBOS score, the greater the anomaly degree of the sample:

$$HBOS(x_i) = \sum_{j=0}^{a} \log\left(\frac{1}{\mathrm{hist_j}(x_i)}\right), \tag{5}$$

where a is the number of features/histograms, x_j is the vector of features, and $\mathrm{hist}_j(x_j)$ represents the density estimation of feature instance x_i in the jth histogram.

2.5. iForest. Isolation Forest (iForest) is a fast outlier detection method based on an ensemble, with linear time complexity and high accuracy, which meets the needs of big data processing. iForest was first proposed in 2008, and then in 2012, an improved version was proposed [28], which is suitable for anomaly detection of continuous data. iForst is different from other anomaly detection algorithms to characterize the degree of alienation between data samples based on quantitative indicators such as distance and density and detects outliers by isolating sample points.

The algorithm utilizes a binary search tree structure called an isolation tree (iTree) to isolate samples. An iForest consists of multiple isolation trees, which are created by choosing attributes and the values of attributes randomly. At each node in the isolation trees, the instance set is divided into two parts based on the chosen attributes and their values. Here, the attributes are selected randomly, and the split value for this selected attribute is selected randomly as well between the minimum value and maximum value of this selected attribute. Commonly, anomalous instances are those objects whose attribute values are very different from the normal instances and are easier to be divided than normal instances. In the process of isolation, they are also closer to the root and more easily divided than normal instances.

3. Network Intrusion Anomaly Detection Model Based on Multiclassifier Fusion

The network intrusion anomaly detection model based on multiclassifier fusion is described in the following section.

3.1. Anomaly Detection Model Based on Multiclassifier Fusion. The anomaly detection method is based on network traffic to extract the content features, essential features, and traffic features from the original network traffic data and detect anomalies from the feature space composed of these three types of features. In intrusion detection, the features associated with different network behavior have different meanings, and it is difficult for a single classifier to effectively deal with the combination of these features with different meanings. A multiple classifier system is needed, whose outputs are combined in some way to obtain a final classification decision under different situations. In an ensemble anomaly detection system, each base classifier will focus on different aspects of the data.

To overcome the limitations of a single classifier, this study proposed a network intrusion anomaly detection model (MF) based on the multiclassifier fusion method. The MF model intelligently combines the decision outputs of multiple classifiers according to selected fusion rules and achieves better detection performance than a single classifier. The overall architecture of the MF model is shown in Figure 1.

The algorithm selection criteria for the MF model are that there is no strong dependence between individual classifiers and that a series of individual classifiers can be generated in parallel. These algorithms are preferably heterogeneous, which means that the types of these algorithms are not the same as the five algorithms chosen in this study. In anomaly detection, different individual classifier results can be obtained using different classifiers, and these results are often highly complementary. Thus, several well-performing heterogeneous anomaly detection algorithms can potentially yield better anomaly detection models.

After data collection and feature extraction, the dataset is divided into two subsets: a training set and a testing set (unsupervised machine learning produces predictions during training). The purpose of the training set is to apply the multiclassifier fusion technique and obtain the confusion matrix of multiple anomaly detection algorithms. The role of the test set is to fuse the outputs of multiple anomaly detection methods using the confusion matrix and obtain the final detection results.

3.2. Multiclassifier Fusion Methods. A confusion matrix is a standard format for expressing accuracy evaluation, which is represented by a matrix with n rows and n columns. Each column of the confusion matrix represents the predicted category, and the total number of data in each column represents the number of data predicted to be in that category. Each row represents the true attribution category of the data, and the total number of data in each row represents the number of data instances in that category.

To ensure the optimal performance, the multiclassifier fusion method should be able to select the subset of classifiers that is optimal in the sense that it produces the highest possible performance for a particular combiner. In this study, three multiclassifier fusion methods are carried out in the MF model, such as majority voting, weighted majority

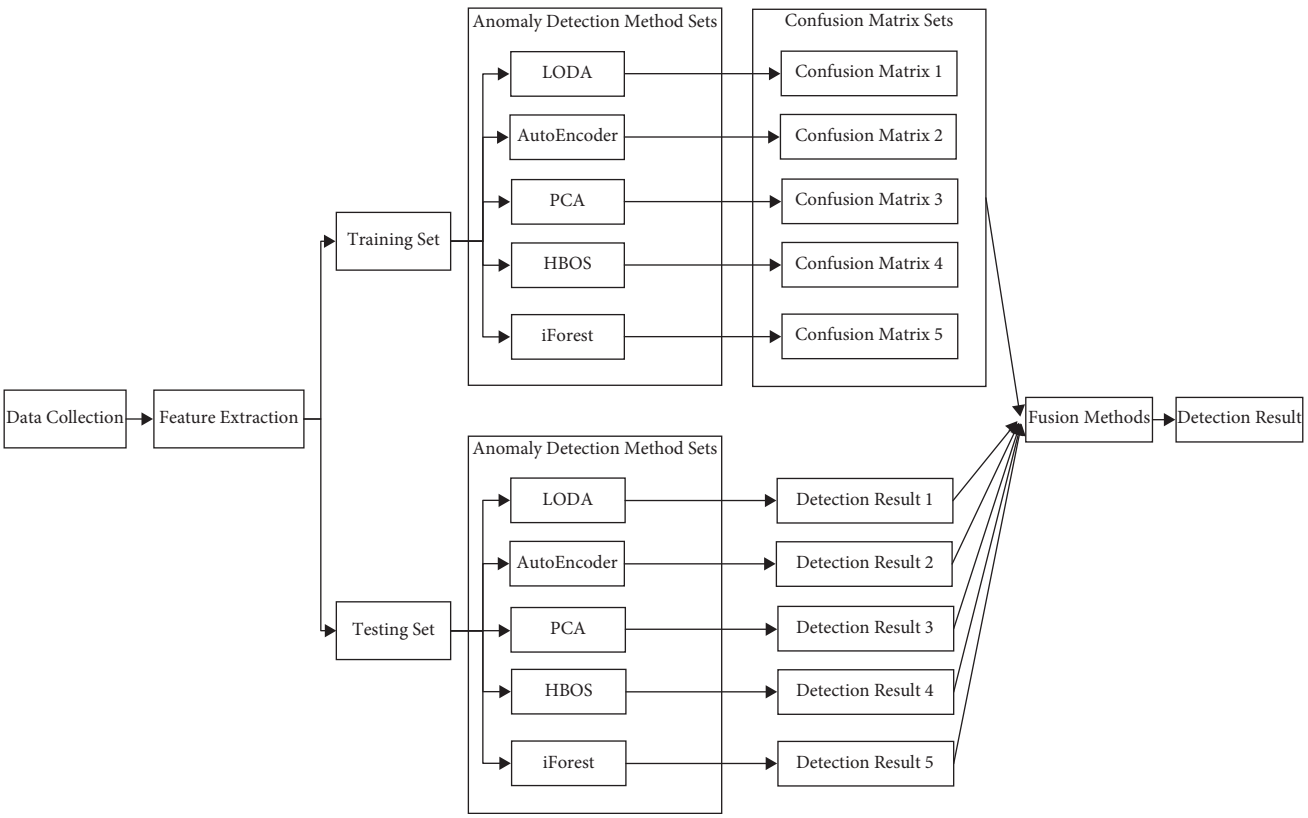


Figure 1: MF model overall architecture.

voting, and Naive Bayes. To better evaluate the performance of these three methods, we assume that the kth classifier gets the training set as input, the predicted confusion matrix is E^k, and $e^k{}_{ij}$ represents the percentage of the samples in class i is predicted as class j in the kth classifier.

3.2.1. Majority Voting. The number of admissible classifiers voted for a special class is assessed. The class with more votes is selected as the ultimate decision. However, in the case of an equal number of votes in two or three classes, the number of admissible classifiers voted against each class is counted and the class with the minimum vote is selected as the final decision. The majority voting can be written formally as follows:

$$C = \underset{i}{\operatorname{argmax}} \sum_k V_i^k, \tag{6}$$

where the combiner classifies the sample into class C in the manner shown in Equation (6). If the kth classifier predicts that the class of the sample is i, then V_i^k is 1, otherwise, it is 0. The algorithm implementation of the MV is shown in Algorithm 1.

3.2.2. Weighted Majority Voting. If many classes receive the same number of votes, $e^k{}_{ii}$ is used for tie-breaking; that is, the voting rights of each classifier are weighted by a number representing the confidence of the classifier for its vote. The vote given by each classifier is weighted by the confidence level assessed by the confusion matrix, and the combiner classifies the sample into class C in the manner shown as

$$C = \underset{i}{\operatorname{argmax}} \sum_k e_{ii}^k \cdot V_i^k. \tag{7}$$

The algorithm implementation of the WMV is shown in Algorithm 2.

3.2.3. Naive Bayes. Based on the priori probability of a sample, the Naive Bayes classifier calculates its posterior probability using the Bayes formula. The posterior probability is the probability that the sample belongs to a certain class. The method selects the class with the largest posterior probability as the decision class for the sample.

When the kth classifier predicts that the class of the sample is j, combined with the prediction results of all classifiers, the probability $p(i)$ that the sample belongs to the class i is shown as

$$p(i) = \frac{M_i \cdot e_{ij}^k}{\sum_{h=1}^m M_h \cdot e_{hj}^k} \tag{8}$$

where M_i is the number of samples that belong to class i. Applying Bayes formula and assuming that the classifiers are independent, the maximizing posterior probability is shown as [29]

```
Input: the results set predicted by each base classifiers Y;
Output: final prediction result C;
a [2] ← {0, 0};
for y in Y do
    a [y] ← a [y] + 1;
end for
C ← argmax (a)
return C
```

Algorithm 1: Majority voting

```
Input: the results set predicted by each base classifiers Y = {y1, y2, ..., yn}, confusion matrix set E = {e^1, e^2, ..., e^n};
Output: final prediction result C;
a [2] ← {0, 0};
for i in {0, 1} do
for j in {1, 2, ..., n} do
if y_j = = i then
v_j ← 1;
else
v_j ← 0;
end if
a [i] ← a [i] + v_j * e^j_ii;
end for
end for
C ← argmax (a);
return C;
```

Algorithm 2: Weighted majority voting

```
Input: the results set predicted by each base classifiers Y = {y1, y2, ..., yn}, confusion matrix set E = {e^1, e^2, ..., e^n}, the number of
samples of each types M = {m0, m1};
Output: final prediction result C;
a [2] ← {0, 0};
for i in {0, 1} do
c_i ← 1;
for j in {1, 2, ..., n} do
k ← y_j;
c_i ← c_i * e^j_ik;
end for
a [i] ← m_i * c_i;
end for
C ← argmax (a);
return C;
```

Algorithm 3: Naive Bayes

$$C = \underset{i}{\operatorname{argmax}}\, M \prod_{k=1}^{N} e_{ij}^{k}. \tag{9}$$

The algorithm implementation of the NB is shown in Algorithm 3.

4. Dataset

In this study, a series of experiments are carried out on three public available datasets which are widely used in the field of network intrusion detection. The three datasets are CICIDS2017 [30], UNSW-NB 15 [31], and KDDCUP 99. A detailed description of the three datasets is shown in Table 1.

4.1. CICIDS2017. CICIDS2017 is an IDS domain dataset that has been collected for a total of 5 days up to 5 p.m. on Friday, July 7, 2017. Monday is a normal day that includes only normal traffic. Some of the most common attacks, such as DoS, DDoS, Brute Force, XSS, SQL injection, infiltration, port scanning, and botnets were performed on Tuesday,

Table 1: Public dataset description.

Dataset	Dimension	Number
CICIDS2017	83	592782
UNSW-NB 15	196	185423
KDDCUP 99	118	50000

Table 2: Experimental configuration environment.

Hardware/software	Configuration/version
Pyod	1.0.0
CPU	2.3 GHz dual core intel core i5
Memory	16 GB 2133 MHz LPDDR3
Operating system	macOS big sur

Table 3: Anomaly detection algorithm performance comparison (anomaly ratio of 0.05).

Dataset	Performance metrics	LODA	AE	PCA	HBOS	iForest	MV	WMV	NB
CICIDS2017	AUC	0.635	0.576	0.538	0.632	0.657	0.763	0.749	0.852
	RECALL	0.497	0.438	0.401	0.531	0.492	0.794	0.821	0.886
	Time (seconds)	0.511	8.369	0.732	0.902	7.436	7.082	8.198	8.073
UNSW-NB 15	AUC	0.675	0.739	0.852	0.724	0.913	0.871	0.912	0.925
	RECALL	0.724	0.791	0.893	0.852	0.937	0.897	0.943	0.961
	Time (seconds)	0.867	10.023	0.992	1.942	8.241	6.924	8.245	8.128
KDDCUP 99	AUC	0.942	0.923	0.924	0.938	0.953	0.910	0.928	0.956
	RECALL	0.964	0.957	0.940	0.964	0.983	0.924	0.965	0.995
	Time (seconds)	0.735	9.710	1.324	1.672	8.092	7.942	8.124	8.206

Wednesday, Thursday, and Friday morning and afternoon, respectively. The dataset is fully labelled and provides the CICFlowMeter software that triages traffic data and extracts more than 80 stream signatures, which are available on the Canadian Cybersecurity Institute website.

4.2. UNSW-NB 15. The UNSW-NB 15 dataset was created by the Australian Center for Cyber Security (ACCS) using the IXIA PerfectStorm tool, which can be used to generate raw network traffic mixed with normal activity and attack behavior. 100 GB of raw traffic (PCAP files) was captured using the tcpdump tool, and 49 stream signatures were extracted using the Argus and Bro-IDS tools, covering 9 attacks: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms.

4.3. KDDCUP 99. The KDD CUP 99 dataset is the data used in the 1999 KDD CUP competition. In 1998, the U.S. defense advanced program (DARPA) conducted an intrusion detection assessment program at the Lincoln Laboratory at MIT. Lincoln Labs built a network environment that simulated the U.S. Air Force's local area network, collected network connection and system audit data for 9 weeks, and simulated various user types, various network traffic, and attack methods to make it resemble a real network environment. The raw data collected through tcpdump are divided into two parts: 7 weeks of training data, which contains about 5,000,000 million network connection records; the remaining 2 weeks of test data contains approximately 2,000,000 million network connection records.

5. Results and Discussion

5.1. Performance Metrics. RECALL and AUC (Area under Curve) are two important metrics used to evaluate the performance of anomaly detection algorithms. The time overhead is used to measure the time efficiency of each algorithm running.

(1) RECALL, also known as the detection rate, reflects the ability of the classifier or model to correctly predict positive samples (abnormal samples), that is, the proportion of the total number of positive samples predicted as positive to the total number of positive samples. The higher the value, the better the performance.

(2) AUC (Area under Curve): in a prediction, if it is an anomalous score value, a score is generally selected as a threshold, and the score above this threshold is abnormal, the score below this threshold is determined to be normal. At this time, according to the different values of the threshold, different FPR (False Positive Rate) and TPR (True Positive Rate) values can be obtained, and all values are plotted with FPR as the horizontal axis, TPR as the vertical axis, that is, the ROC curve, and the area covered by the curve downward is the AUC value.

Table 4: Anomaly detection algorithm performance comparison (anomaly ratio of 0.1).

Dataset	Performance metrics	LODA	AE	PCA	HBOS	iForest	WV	WMV	NB
	AUC	0.629	0.643	0.683	0.513	0.753	0.789	0.819	0.842
CICIDS2017	RECALL	0.389	0.421	0.569	0.492	0.628	0.853	0.853	0.896
	Time (seconds)	0.441	6.649	0.484	0.197	7.391	3.724	3.621	3.604
	AUC	0.721	0.691	0.783	0.689	0.927	0.919	0.924	0.943
UNSW-NB 15	RECALL	0.752	0.803	0.902	0.762	0.932	0.934	0.951	0.958
	Time (seconds)	0.529	8.236	0.692	0.384	9.242	4.045	4.578	4.248
	AUC	0.928	0.956	0.931	0.928	0.954	0.894	0.927	0.961
KDDCUP 99	RECALL	0.952	0.969	0.950	0.955	0.985	0.961	0.985	0.993
	Time (seconds)	0.502	7.923	0.669	0.328	8.927	3.928	4.029	4.293

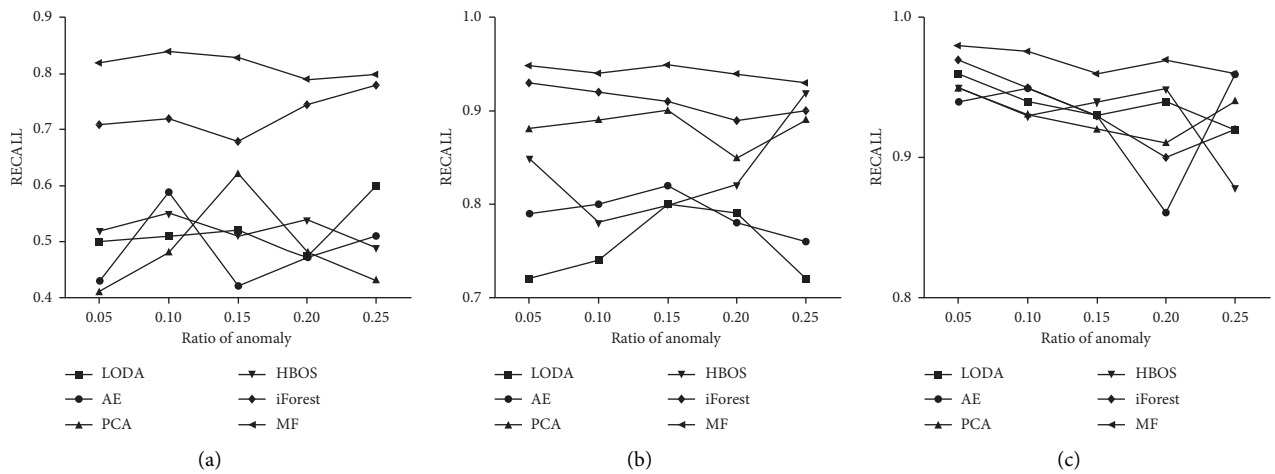


Figure 2: RECALL of different datasets varies with the anomaly rate. (a) CICIDS2017. (b) UNSW-NB 15. (c) KDDCUP 99.

(3) Time is also known as time overhead, during which an algorithm is executed.

5.2. Experimental Configuration Environment. The experimental configuration environment is given in Table 2. This study uses a well-known python third-party library for anomaly detection, PyOD, which runs on a PC machine. The hardware configuration is 2.3 GHz dual core Intel Core i5, 16 GB of 2133 MHz LPDDR3 memory, and the macOS Big Sur operating system without GPU acceleration. In addition, since the base classifier references the third-party python library, Pyod, the anomaly rate is the only parameter setting of the algorithms, which is reflected in Section 5.4.

5.3. Performance Comparison. Most of the real network traffic is normal, with a small percentage of attacks or anomalous traffic. Therefore, the vast majority of datasets widely used in the field of anomaly detection are data imbalances, the proportion of normal samples is very large, and the proportion of abnormal samples is very small. The anomaly rate represents the proportion of anomalous samples to the total number of samples. Since, in real network datasets, the ratio of anomalous data is small, the ratio of normal data is large. This experiment sets the anomaly rate to two smaller values (0.05 and 0.1) to simulate the real environment [32], comparing the performance of five anomaly detection baseline algorithms and three fusion methods on different datasets. The results are shown in Tables 3 and 4, respectively.

The motivation we use the multiclassifier fusion technique is to improve the accuracy of anomaly detection; hence, AUC and RECALL are the two most important metrics. In addition, the time overhead for each fusion technique should also be used as an indicator to measure its performance.

Among the three fusion methods, NB (Naive Bayes) performs better than MV (Majority Voting) and WMV (Weighted Majority Voting) in both RECALL and AUC; besides, the fusion takes a similar amount of time. So, the MF model in this study selects Naive Bayes as the fusion method.

On the contrary, multiclassifier fusion models combine multiple classifiers and must ensure that the overall detection is better than the best single classifier. If the overall performance of the fusion model is slightly better than the worst classifier, but slightly worse than the best classifier, then, in this case, it is better to let the best classifier work on its own and not participate in the mix. From Tables 3 and 4, it can be observed that the overall detection performance of NB is better than that of a single classifier in all three datasets, while the detection

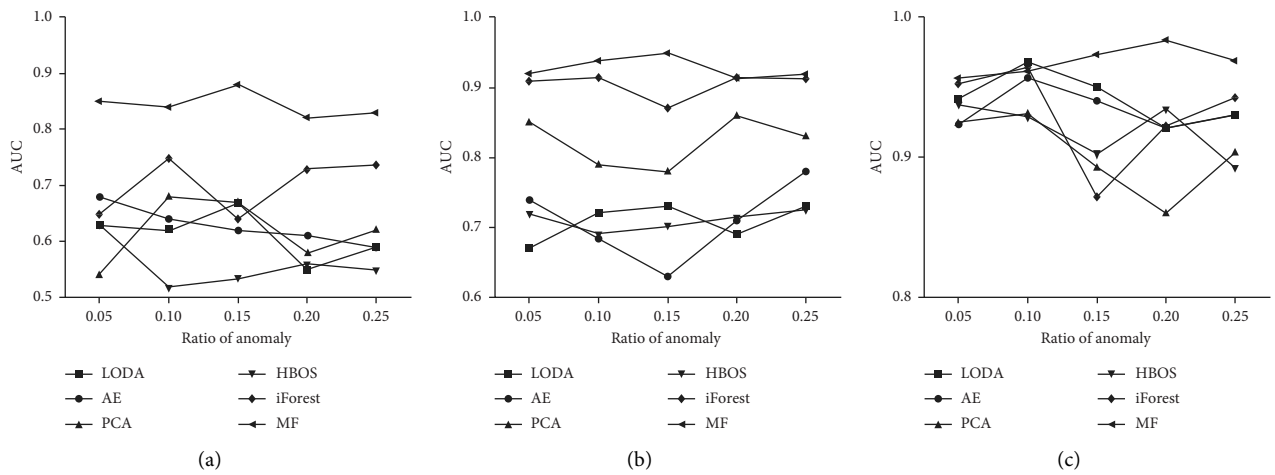


Figure 3: AUC of different datasets varies with the anomaly rate. (a) CICIDS2017. (b) UNSW-NB 15. (c) KDDCUP 99.

performance of MV and WMV is sometimes worse than that of a single detection algorithm.

In terms of time overhead, this study randomly samples 100,000 pieces of data in each of the three datasets, and the time costs of the three basic anomaly detectors, HBOS, LODA, and PCA, are close to each other and significantly faster than AE and iForest. This is related to the fact that the first three algorithms are suitable for large amounts of data. The time of the last two datasets (UNSW-NB 15 and KDDCUP 99) was slightly higher than that of the first dataset, CIDIDS2017, due to the fact that the dimensions of the latter two datasets were higher than those of CICIDS2017. It should be noted that the times of MV, WMV, and NB represent only the fusion time of their respective results and do not include the training time of the base classifier.

As shown in Table 3, when the anomaly rate is set to 0.05, the MF model proposed in this study is superior to the other five anomaly detection baseline algorithms in terms of the performance of both the AUC and RECALL indicators.

On the CICIDS2017 dataset, the RECALL and AUC of the MF were 0.886 and 0.852, respectively, which were 0.355 higher than the HBOS (0.531) with the highest RECALL rate and 0.195 higher than iForest (0.657), which had the best AUC performance.

On the UNSW-NB 15 dataset, the RECALL and AUC of the MF were 0.961 and 0.925, respectively, an increase of 0.24 and 0.12 compared to iForest (0.937 and 0.913), which had the highest RECALL and AUC indicators.

On the KDDCUP 99 dataset, the recall and AUC of MF were 0.995 and 0.956, respectively, an increase of 0.12 and 0.03 compared with iForest (0.983 and 0.953), which had the highest RECALL and AUC indicators.

As shown in Table 4, when the anomaly rate is set to 0.1, a similar conclusion can be seen from Table 4: MF outperforms the other five baseline methods in both AUC and RECALL in three datasets.

The results of Tables 3 and 4 prove that the network intrusion anomaly detection model MF proposed in this study based on multiclassifier fusion technology has more advantages in anomaly detection and performs better than the five state-of-the-art unsupervised learning algorithms.

5.4. Performance Comparison at Different Anomaly Rates. In order to better verify the model proposed in this study, this experiment also compares the detection results of five baseline algorithms for anomaly detection and the MF model in the case of continuous changes in the anomaly rate on three public datasets.

As Figures 2 and 3 shown, with the increase in the anomaly rate, the values of AUC and RECALL of the other five baseline methods are always lower than the MF model and have obvious fluctuations, which means the instability of performance. In contrast, the performance (AUC and RECALL) of the MF model was stable, and the polyline is flatter. This also verifies that the MF model has good robustness and can better adapt to the dynamic changes in attack scale and attack numbers in different network environments.

6. Conclusions

Anomaly detection refers to the detection of data or behaviors that do not conform to normal expected patterns in a large amount of data and is widely used in the field of data security and network security. Aiming at the problem that the detection effect of unsupervised learning methods cannot meet the anomaly detection requirements of the real network environment, this study proposes a network intrusion anomaly detection model MF based on multiclassifier fusion technology. The model can use different fusion methods, such as majority voting fusion, weighted majority voting fusion, and Naive Bayes fusion, to intelligently fuse the detection results of multiple anomaly detection baseline methods (such as LODA, AutoEncoder, PCA, HBOS, and iForest) to obtain a higher detection effect than a single detection method. Finally, the MF model is compared with other anomaly detection baseline methods on three public network intrusion anomaly detection datasets, such as CICIDS2017, UNSW-NB 15, and KDDCUP

99. Experimental results show that the network intrusion anomaly detection model MF proposed in this study based on multiclassifier fusion technology successfully improves the accuracy and effectiveness of detection and has good robustness.

In future research, more advanced intelligent fusion technology will be used to fuse the results of the baseline method to obtain better performance.

Data Availability

All data included in this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 62072074, 62076054, 62027827, 61902054, and 62002047), the Frontier Science and Technology Innovation Projects of National Key R&D Program (no. 2019QY1405), the Sichuan Science and Technology Innovation Platform and Talent Plan (no. 2020TDT00020), and the Sichuan Science and Technology Support Plan (no. 2020YFSY0010).

References

[1] S. E. Smaha, "Haystack: an intrusion detection system," in *Proceedings of the IEEE Fourth Aerospace Computer Security Applications Conference*, pp. 37–44, Orlando, FL, September 1988.

[2] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," *Journal of Computer Security*, vol. 10, no. 1-2, pp. 105–136, 2002.

[3] D. Heckerman, *A Tutorial on Learning with Bayesian Networks*, Technical Report MSRTR-95-06, Microsoft Research, Bengalaru, India, 1995.

[4] K. S. Valdes, "Adaptive model-based monitoring for cyber attack detection," in *Proceedings of the Recent Advances in Intrusion Detection Toulouse*, pp. 80–92, France, October 2000.

[5] D. M. Kruegel, W. Robertson, and F. Valeur, "Bayesian event classification for intrusion detection," in *Proceedings of the 19th Annual Computer Security Applications Conference*, Las Vegas, NV, December 2003.

[6] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[7] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," in *Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop*, pp. 172–179, Melbourne, FL, USA, November 2003.

[8] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018.

[9] K. Park, Y. Song, and Y. G. Cheong, "Classification of attack types for intrusion detection systems using a machine learning algorithm," in *Proceedings of the 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 282–286, Bamberg, Germany, March 2018.

[10] S. Bhattacharya, P. K. R. Maddikunta, P. K. R. Maddikunta et al., "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," *Electronics*, vol. 9, no. 2, p. 219, 2020.

[11] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.

[12] M. Alazab, S. Venkatraman, P. Watters, M. Alazab, and A. Alazab, "Cybercrime: the case of obfuscated malware," in *Global Security, Safety and Sustainability & E-Democracy*, pp. 204–211, Springer, Thessaloniki, Greece, 2011.

[13] S. Huda, J. Abawajy, M. Alazab, M. Abdollalihian, R. Islam, and J. Yearwood, "Hybrids of support vector machine wrapper and filter based framework for malware detection," *Future Generation Computer Systems*, vol. 55, pp. 376–390, 2016.

[14] R. U. Khan, X. Zhang, M. Alazab, and R. Kumar, "An improved convolutional neural network model for intrusion detection in networks," in *Proceedings of the 2019 Cybersecurity and Cyberforensics Conference(CCC)*, pp. 74–77, Melbourne, Australia, May 2019.

[15] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, "Robust intelligent malware detection using deep learning," *IEEE Access*, vol. 7, pp. 46717–46738, 2019.

[16] G. Karatas, O. Demir, and O. K. Sahingoz, "Deep learning in intrusion detection systems," in *Proceedings of the 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, pp. 113–116, ANKARA, Turkey, December 2018.

[17] S. M. Othman, F. M. Ba-Alwi, N. T. Alsohybe, and A. Y. Al-Hashida, "Intrusion detection model using machine learning algorithm on Big Data environment," *J. Big Data*, vol. 5, no. 1, p. 34, 2018.

[18] T. R. Gadekallu and N. Khare, "Cuckoo search optimized reduction and fuzzy logic classifier for heart disease and diabetes prediction," *International Journal of Fuzzy System Applications*, vol. 6, no. 2, pp. 25–42, 2017.

[19] G. T. Reddy and N. Neelu, "Hybrid firefly-bat optimized fuzzy artificial neural network based classifier for diabetes diagnosis," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 4, pp. 18–27, 2017.

[20] N. Khare and G. T. Reddy, "Heart disease classification system using optimised fuzzy rule based algorithm," *International Journal of Biomedical Engineering and Technology*, vol. 27, no. 3, pp. 183–202, 2018.

[21] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evolutionary Intelligence*, vol. 13, no. 2, pp. 185–196, 2019.

[22] A. Dainotti, A. Pescapé, and C. Sansone, "Early classification of network traffic through multi-classification," in *Proceedings of the International Workshop on Traffic Monitoring and Analysis*, Springer, Berlin, Heidelberg, 2011.

[23] D. C. Le and N. Zincir-Heywood, "Anomaly detection for insider threats using unsupervised ensembles," *IEEE*

Transactions on Network and Service Management, vol. 18, no. 2, pp. 1152–1164, 2021.

[24] H. D. Nguyen, K. P. Tran, S. Thomassey, and M Hamad, "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management," *International Journal of Information Management*, vol. 57, Article ID 102282, 2021.

[25] R. Patil, R. Biradar, V. Ravi, P. Biradar, and U Ghosh, "Network traffic anomaly detection using PCA and BiGAN," *Internet Technology Letters*, vol. 5, no. 1, p. e235, 2022.

[26] O. Abdelrahman and P. Keikhosrokiani, "Assembly line anomaly detection and root cause analysis using machine learning," *IEEE Access*, vol. 8, Article ID 189661, 2020.

[27] N. Paulauskas and A. Baskys, "Application of histogram-based outlier scores to detect computer network anomalies," *Electronics*, vol. 8, no. 11, p. 1251, 2019.

[28] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, pp. 1–39, 2012.

[29] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, The Netherland, 2004.

[30] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSp*, vol. 1, pp. 108–116, 2018.

[31] N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Information Security Journal: A Global Perspective*, vol. 25, no. 1-3, pp. 18–31, 2016.

[32] Y. Wu, LiWei, M. Ni, and Z. Xu, "Anomaly detection model based on one-class support vector machine fused deep autoencoder," *Computer Science*, vol. 49, no. 3, pp. 144–151, 2022.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 4131322, 9 pages
https://doi.org/10.1155/2022/4131322

Research Article

Learning Identity-Consistent Feature for Cross-Modality Person Re-Identification via Pixel and Feature Alignment

Sixian Chan,[1,2] Feng Du,[1] Yanjing Lei,[1] Zhounian Lai,[3] Jiafa Mao,[1] and Chao Li[4]

[1]*Zhejiang University of Technology, Hangzhou, China*
[2]*Hangzhou Xsuan Technology Co. Ltd, Hangzhou, China*
[3]*Huzhou Institute of Zhejiang University, Hangzhou, China*
[4]*Zhijiang College of Zhejiang University of Technology, Hangzhou, China*

Correspondence should be addressed to Yanjing Lei; leiyj@zjut.edu.cn

Received 16 August 2022; Accepted 17 September 2022; Published 10 October 2022

Academic Editor: Wang Wenyong

RGB-IR cross-modality person re-identification (ReID) can be seen as a multicamera retrieval problem that aims to match pedestrian images captured by visible and infrared cameras. Most of the existing methods focus on reducing modality differences through feature representation learning. However, they ignore the huge difference in pixel space between the two modalities. Unlike these methods, we utilize the pixel and feature alignment network (PFANet) to reduce modal differences in pixel space while aligning features in feature space in this paper. Our model contains three components, including a feature extractor, a generator, and a joint discriminator. Like previous methods, the generator and the joint discriminator are used to generate high-quality cross-modality images; however, we make substantial improvements to the feature extraction module. Firstly, we fuse batch normalization and global attention (BNG) which can pay attention to channel information while conducting information interaction between channels and spaces. Secondly, to alleviate the modal difference in feature space, we propose the modal mitigation module (MMM). Then, by jointly training the entire model, our model is able to not only mitigate the cross-modality and intramodality variations but also learn identity-consistent features. Finally, extensive experimental results show that our model outperforms other methods. On the SYSU-MM01 dataset, our model achieves a rank-1 accuracy of 40.83% and an mAP of 39.84%.

1. Introduction

Person ReID can be viewed as a cross-camera image retrieval problem, which aims at matching individual pedestrian images in a query set to ones in a gallery set captured by different cameras. Its main challenge lies in the interclass and intraclass variations caused by different lighting, poses, occlusions, and views. Most existing methods [1–5] mainly focus on matching RGB images captured by visible cameras, which can be formulated as an image matching problem under a single modality. However, these methods cannot be applied to images taken in poor lighting conditions, because the visible camera cannot capture pictures with discriminative features. However, in practical application scenarios, the camera should ensure all-weather operation.

Since the visible camera has limited effect on the security work at night, the camera that can switch the infrared mode is being widely used in the intelligent monitoring system. In visible mode and infrared mode, RGB images and infrared images are collected, respectively, which belong to two different modalities. RGB images have three channels but IR images have only one channel, so the ReID problem in a cross-modality setting becomes extremely challenging, which is essentially a cross-channel retrieval problem. First, infrared images of different identities are difficult to distinguish but are easy to distinguish in visible images. In addition, the same person varies greatly in different modalities. It is known as modality discrepancy.

To address visible-infrared person ReID, several approaches [6–10] have been proposed, aiming to mitigate

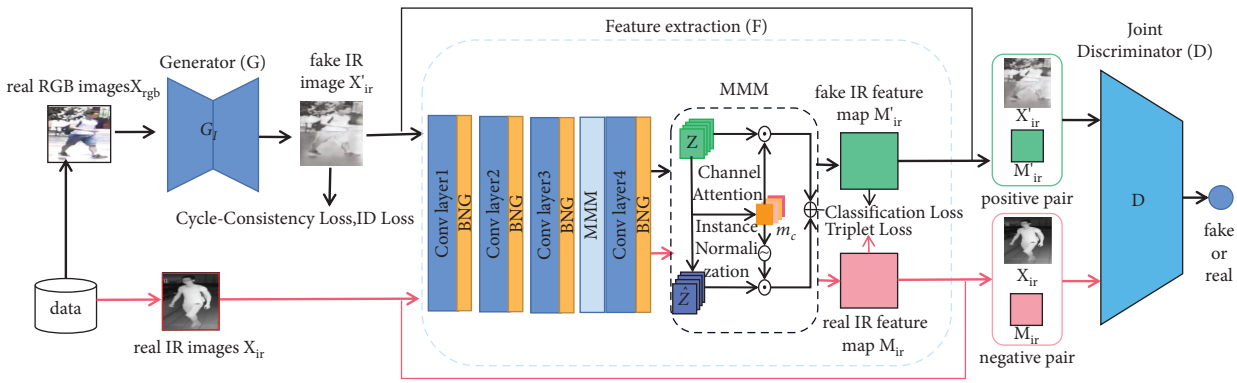


Figure 1: Framework of the proposed model. It consists of an image generation module (G), a joint discriminator module (D), and a feature extraction module (F). The G can generate fake IR images X'_{ir} to mitigate the cross-modality variation, and the F can alleviate the intramodality variation. The F module contains ResNet-50 and BNG attention and MMM module. The BNG module can focus on channel and spatial information, and the MMM module can reduce modality differences.

modal differences by aligning features or pixel distributions. Feature alignment methods [6, 8, 10] mainly focus on bridging the gap between RGB and IR images through features. It is difficult to match RGB and IR images in a shared space due to large cross-modality differences between the two modalities. Different from existing methods that directly match RGB and IR images, we use generative adversarial networks to generate fake IR images based on real RGB images and then match the generated images through a feature alignment network. The generated fake IR images are used to reduce the modality difference between the RGB and IR images. Although the generated fake IR images are very similar to real images, there are still intraclass differences due to pose variations, viewpoint changes, and occlusions.

Inspired by the above discussion, in this paper, we propose a pixel and feature alignment network (PFANet) that simultaneously mitigates cross-modality differences in pixel space and intramodality variation in feature space. As shown in Figure 1, to reduce the modal difference, we apply a generator (G_I) to generate fake IR images. Then, to alleviate the intramodality variation, a feature extraction module (F) is designed to encode fake and real IR images into a shared feature space by exploiting identity-based classification and triplet loss. The batch normalization and global (BNG) attention is added to the feature extraction network (F), which can make the network learn which channel is more important as well as can interact between channels and spaces. Furthermore, to mitigate the modal difference in the feature space, a modal mitigation module (MMM) is proposed, which can significantly mitigate the difference between the two modalities. Finally, to learn identity-consistent recognition, a joint discriminator (D) is utilized. Its input is an image-feature pair.

The major contributions of this work can be summarized as follows:

(i) We propose a generative adversarial network to generate cross-modality images that alleviated modal differences in pixel space. This model consists of a generator and a joint discriminator, by playing a max-min game, our model is able to not only reduce the cross-modality and intramodality variations but also learn identity-consistent features.

(ii) We design a batch normalization and global (BNG) attention, which consists of channel attention and global attention. In the channel attention, we measure the importance of each channel by applying the scale factor of BN to the channel dimension and suppressing insignificant features. As for the global attention module, it can reduce information attenuation and amplify the features of global dimension interaction.

(iii) We apply a modal mitigation module (MMM) to mitigate the modal distribution. The instance normalization (IN) is utilized to mitigate modal differences on a single instance. Moreover, the channel attention is used to guide the learning of IN, which can mitigate modal differences while preserving identity information.

2. Related Works

2.1. RGB-IR Person ReID. RGB-IR cross-modality person ReID can be seen as a multicamera retrieval problem that aims to match pedestrian images captured by visible and infrared cameras, which are widely used in video surveillance, public security, and smart cities. Compared with RGB-RGB single-modality person ReID which only deals with RGB images, the key challenge in this work is to mitigate the large differences between the two modalities. To address the challenge caused by differences in modality distributions, a variety of approaches to cross-modality person re-identification have been proposed. Some early work focused on solving the channel mismatch between RGB images and IR images, due to RGB images having three channels. In contrast, IR images have only one channel. Wu et al. [10] proposed a deep zero-padding network and contributed a new ReID dataset SYSU-MM01. In [11], a dual-path network with a bi-directional dual-constrained top-ranking loss was introduced to learn modality alignment

feature representations for RGB-IR ReID. Feng et al. [12] proposed a framework for solving heterogeneous matching problems using modality-specific networks. Ye et al. [13] proposed a dual-stream network with feature learning and metric learning to convert two heterogeneous modalities into a consistent space where the modalities share a metric. Dai et al. [6] introduced a cross-modality generative adversarial network (cmGAN) to reduce the distribution differences between RGB and IR features. Most of the above approaches mostly focus on reducing intermodality differences by feature alignment, while ignoring the large cross-modality differences in pixel space.

Unlike these approaches, the proposed model in this paper is able to combine feature alignment and pixel alignment, effectively reducing intramodality and cross-modality variations. By training the model, the model is able to learn identity consistency features.

2.2. GAN in Person ReID. A generative adversarial network (GAN) consists of a generator and a discriminator, using the idea of game theory, where the generator tries to generate an image to deceive the discriminator, and the discriminator tries to discriminate whether the image is real or generated. Through multiple adversarial training, generative adversarial networks are able to learn deep representations of data in a self-supervised manner. GAN can generate high-quality images, perform image enhancement, generate images from text, and convert images from one domain to another [14, 15]. GAN was first proposed in 2014's [16]. After that, researchers have proposed a variety of task-specific GAN structures, such as CycleGAN [14], Pix2Pix [17], and StarGAN [15]. There are many works in the field of pedestrian re-identification that also apply GAN to improve accuracy. Li et al. [18] proposed a network that allows querying images of different resolutions to process cross-resolution person ReID. Wang et al. [19] designed an end-to-end alignment generative adversarial network (AlignGAN) for the RGB-IR ReID task. JSIA-ReID [20] implemented a two-layer alignment of pixels and features in a unified GAN framework.

In our work, we apply GAN to generate cross-modality images that mitigate modal differences between RGB-IR image data in pixel space.

2.3. Attention Mechanisms. There is an important feature in the human visual system that allows people to selectively focus on things of interest in order to capture valuable information. Inspired by the human visual system, many works have attempted to employ attention mechanisms to improve the performance of CNNs.

Attention mechanisms enable the network to focus on areas of interest to the human body and better extract useful information. SENet [21]integrated spatial information into the channel-level feature responses and computed the corresponding attention with two MLP layers. Later, bottleneck attention module (BAM) [22] built independent space and channel submodules in parallel and embedded them into each bottleneck block. Considering the relationship between any two positions of the feature map, nonlocal feature attention [23] was proposed to capture the relationship between them. The convolution block attention module (CBAM) [24] sequentially cascaded channel attention and spatial attention. However, these works ignored the information about the weights adjusted from the training; therefore, we wanted to highlight the significant features by using the variance of the trained model weights, which also was able to amplify cross-dimensional interactions and captured important features of all three dimensions. We propose new attention (BNG) to solve the above problem. A modal mitigation module (MMM) is designed to mitigate the modal distribution, using channel attention to guide the learning of instance normalization (IN) for mitigating modal differences while preserving identity information.

3. The Proposed Method

In this part, we introduce the proposed PFANet in detail. Our network will be presented in the following three parts, including (1) RGB-IR images generation module, (2) BNG attention module, and (3) modal mitigation module. To reduce cross-modality variation, we apply generative adversarial networks to convert RGB images to fake IR images, which have IR style while maintaining their original identity.

Then, the features of the two modalities are extracted for feature alignment. The BNG attention is designed to make the network focus on channel and spatial information. In addition, the modal mitigation module (MMM) is proposed to mitigate the differences between the two modalities. The main output of the PFAnet during testing is the feature for person ReID.

3.1. RGB-IR Images Generation Module. There is a large cross-modality difference between RGB and IR images, which significantly increases the difficulty of the task of cross-modality pedestrian re-identification. To reduce cross-modality variation, we apply generative adversarial networks to convert RGB images X_{rgb} to fake IR images X'_{ir}, which has IR style while maintaining their original identities. The generated fake IR image X'_{ir} can mitigate the modality differences between RGB and IR images. The module consists of a generator G_I that generates a fake IR image from an RGB image and a joint discriminator D_I that discriminates whether the image is a real image or a generated image. The input of the generator is the real images X_{rgb}, and its output is the fake IR images $X'_{ir} = G_I(X_{rgb})$. The input of the discriminator is the generated fake IR image X'_{ir}; if the image is real, its output is one, and if the image is the generated image, the output is zero. The goal of the generator is to make the generated image as similar as possible to the real image, and the goal of the discriminator is to discriminate as much as possible whether the input image is real or generated. Unlike ordinary discriminators, the input to our discriminator is a pair of IR images and ReID feature maps. The generator and discriminator play the min-

max game as [16], and the modal can make the fake IR image X'_{ir} as realistic as possible.

The adversarial loss for generating IR images is defined as follows:

$$\mathscr{L}_{G_I} = \mathbb{E}\left[\log_{D_I}\left(X'_{ir}, f^{X'_{ir}}_{map}\right)\right], \tag{1}$$

$$\mathscr{L}_{D_I} = \mathscr{L}^{real}_{D_I} + \mathscr{L}^{fake}_{D_I}, \tag{2}$$

where

$$\mathscr{L}^{real}_{D_I} = \mathbb{E}_{(x,f)\in\left(X_{ir}, f^{X_{ir}}_{map,R}\right)}\left[\log D_I(x,f)\right], \tag{3}$$

$$\mathscr{L}^{fake}_{D_I} = \mathbb{E}_{(x,f)\in M}\left[\log\left(1 - D_I(x,f)\right)\right], \tag{4}$$

$$M = \left(X'_{ir}, f^{X'_{ir}}_{map,R}\right) \cup \left(X'_{ir}, f^{X_{ir}}_{map,R}\right) \cup \left(X_{ir}, f^{X'_{ir}}_{map,R}\right). \tag{5}$$

Among them, $f^{X_{ir}}_{map,R}$ is the extracted image feature of X_{ir} and $f^{X'_{ir}}_{map,R}$ is the extracted image feature of generated image X'_{ir}. Equation (1) is used to train the generator model; after the constraint of the loss function, the generator will generate a more realistic IR image. Equations (3) and (4) are used to train the discriminator, which differs from traditional discriminators in that the input is a pair of image features. It has two advantages, firstly, the fake IR image X'_{ir} will be closer to the real IR image X_{ir} through the max-min game [16], and the distribution of the features $f^{X'_{ir}}_{map,R}$ of the fake IR image will be more similar to the real image features $f^{X_{ir}}_{map,R}$. Secondly, $f^{X'_{ir}}_{map,R}$ is able to maintain the identity-consistency by the corresponding image X'_{ir} constraint. Although $\mathscr{L}_{G_I}$ loss can ensure that the fake IR image X'_{ir} resembles the real IR image X_{ir}, there is no guarantee that the generated fake IR images retain the structure and content of the original RGB images X_{rgb}. To deal with this problem, we introduce a generator G_R for generating IR images into RGB images and the corresponding discriminator D_R. Also we introduce cycle-consistency loss which is defined as follows:

$$\begin{aligned}\mathscr{L}_{cyc} = {} & E\left[\left\|G_R\left(G_I\left(X_{rgb}\right)\right) - X_{rgb}\right\|_1\right] \\ & + E\left[\left\|G_I\left(G_R\left(X_{ir}\right)\right) - X_{ir}\right\|_1\right].\end{aligned} \tag{6}$$

$\mathscr{L}_{\text{cyc}}$ loss enables the G_I generated IR image to be consistent with the input real RGB image. We use the L1 norm instead of the L2 norm because the L1 norm allows the generator to generate better image edges. Specifically, we input the real RGB image X_{rgb} into the generator G_I to generate the fake IR image X'_{ir} and then use the generator G_R to generate the reconstructed RGB image from the fake IR image. We do something similar with IR images.

Now, the loss of the generator can be defined as follows:

$$\mathscr{L}_G = \mathscr{L}_{G_I} + \omega * \mathscr{L}_{cyc}, \tag{7}$$

where ω is the weight of cycle loss and ω is set to 10 as in [14]. By using this loss during adversarial training, we can generate high-quality IR images.

3.2. The BNG Attention Module. Our proposed BNG attention is an efficient and lightweight attention mechanism. The BNG attention can be embedded at the end of any convolutional neural network, for the residual network ResNet-50; the end of the residual structure can be embedded. The structure of BNG is shown in Figure 2.

BNG attention consists of two submodules, as shown in Figure 2(a); the channel attention submodule can use the weight information of the trained model to highlight salient features. We obtain its scale factor from batch normalized (BN [25]) as shown in

$$B_{out} = BN\left(B_{in}\right) = \gamma\frac{B_{in} - \mu_{\mathscr{B}}}{\sqrt{\sigma^2_{\mathscr{B}} + \epsilon}} + \beta, \tag{8}$$

where $\mu_{\mathscr{B}}$ and $\sigma_{\mathscr{B}}$ are the mean and standard deviation of mini batch $\mathscr{B}$ and γ and β are the trainable parameters used to fit the data distribution.

The formula for channel attention can be expressed as follows:

$$\mathbf{F}_1 = \text{sigmoid}\left(W_\gamma\left(BN\left(\mathbf{F}\right)\right)\right), \tag{9}$$

where γ is the scale factor for each channel, and the weights are obtained as $W_\gamma = \gamma_i / \sum_{j=0}\gamma_j$. We measure the importance of each channel by applying the scale factor of BN to the channel dimension and suppressing insignificant features. Since channel attention only focuses on channel information, there is no global space-channel information interaction; to solve this problem, we design a global attention module. It can reduce information attenuation and amplify the features of global dimension interaction. Inspired by CBAM [24], the channel attention and spatial attention are connected in turn. The main structure is shown in Figure 2(b). Given the input feature map $\mathbf{F_1} \in \mathbb{R}^{C\times H\times W}$, the intermediate state F_2 and output F_3 are defined as follows:

$$\begin{aligned}\mathbf{F}_2 &= \mathbf{M}_c\left(\mathbf{F}_1\right) \otimes \mathbf{F}_1, \\ \mathbf{F}_3 &= \mathbf{M}_s\left(\mathbf{F}_2\right) \otimes \mathbf{F}_2,\end{aligned} \tag{10}$$

where $\mathbf{M}_c$ and M_s are the channel and spatial attention maps, respectively. $\otimes$ denotes element-wise multiplication.

The channel attention submodule uses a 3D arrangement to preserve information across three dimensions and then uses a two-layer MLP layer that amplifies the channel spatial dependencies across dimensions. The channel attention submodule is illustrated in Figure 3.

In the spatial attention submodule, to focus on the spatial information, two convolutional layers are used to fuse the spatial information. The size of the convolution kernel is set to $7 * 7$. Since max-pooling reduces information and has a negative influence, we remove the max-pooling operation to retain more features. The same reduction ratio γ is adopted from the channel attention submodule, same as BAM. The spatial attention submodule without group convolution is shown in Figure 4.

3.3. Modal Mitigation Module (MMM). To mitigate the modal distribution, a modal mitigation module (MMM) is

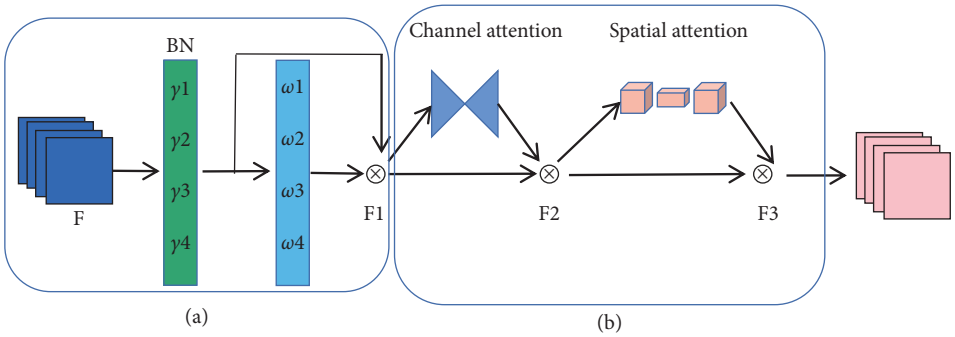


Figure 2: BNG attention.

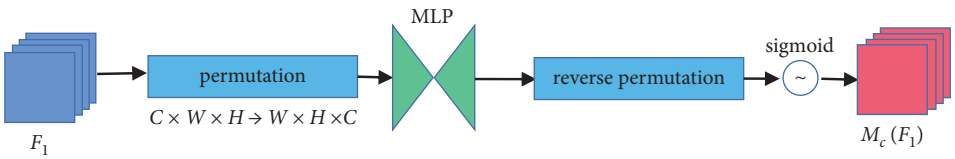


Figure 3: Channel attention submodule.

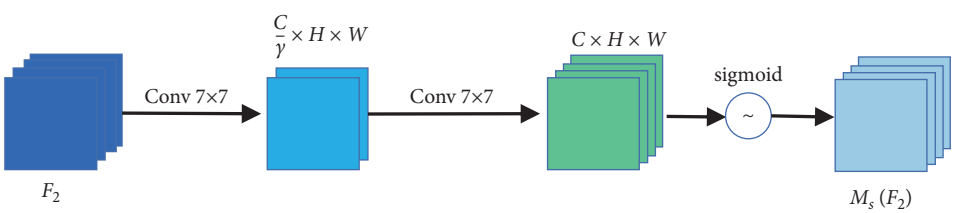


Figure 4: Spatial attention submodule.

designed. For the input image X, we denote the features extracted in the convolution block as $\mathbf{M} \in \mathbb{R}^{h\times w\times c}$ and input it into the MMM, where $h, w, \text{an d}\, c$ represent the height, width, and a number of channels of the feature map $\mathbf{M}$, respectively. The instance normalization (IN) is used to mitigate modal differences on a single instance [27]. Instance normalization (IN) computes the mean and variance in a single instance and reduces the difference between the two data distributions. However, using IN directly may has a negative impact on the ReID task. Because the distribution of image data has changed significantly, some identifying information may be lost.

To overcome these shortcomings, we use channel attention to guide the learning of IN, which mitigates modal differences while preserving identity information. Specifically, we input the feature into a two-layer MLP to downsample the channels and then upsample to the original number of channels and use the activate function to activate the feature as a mask to supervise the IN operation:

$$\mathbf{F} = \mathbf{m}_C \odot \mathbf{M} + (1 - \mathbf{m}_C) \odot \widehat{\mathbf{M}}, \tag{11}$$

where m_C is the channel mask, representing the identity-related channels, and $\widehat{\mathbf{M}}$ is the instance-normalized result of the input $\mathbf{M}$.

Similar to SENet [21], the method of generating a mask with channel dimension can be expressed as follows:

$$\mathbf{m}_C = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 g(\mathbf{M}))), \tag{12}$$

where $\mathbf{W}_1 \in \mathbb{R}^{c/r\times c}$ and $\mathbf{W}_2 \in \mathbb{R}^{c\times c/r}$ are learnable parameters in the two bias-free fully connected (FC) layers, which are followed by ReLU activation function $\delta(\cdot)$ and sigmoid activation function $\sigma(\cdot)$. $g(\cdot)$ denotes global average pooling of features. In order to balance performance and reduce the number of parameters, the downsampling ratio is set to $r = 16$.

The formula for instance normalization is defined as follows:

$$\widehat{\mathbf{M}}_j = \mathrm{IN}(\mathbf{M}_j) = \frac{\mathbf{M}_j - E[\mathbf{M}_j]}{\sqrt{\mathrm{Var}[\mathbf{M}_j] + \epsilon}}, \tag{13}$$

where $E[\cdot]$ is to calculate the mean of each dimension and $\mathrm{Var}[\cdot]$ is to calculate the standard deviation of each dimension. To avoid dividing by zero, we add ϵ to the denominator, and $\mathbf{M}_j \in \mathbb{R}^{h\times w}$ is the j-th dimension of the feature map M.

3.4. Loss Function. In this section, we will introduce the loss we used when training the generator to generate a fake IR image X'_{ir}. On the one hand, X'_{ir} should be classified to the same identity class as the corresponding X_{rgb}; on the other

hand, $X_{ir}^{'}$ should satisfy the triplet loss [28] of the corresponding X_{rgb} identity constraint. We define these two losses as $\mathscr{L}_{cls}^{gan}$ and $\mathscr{L}_{tri}^{gan}$ and denote them in

$$\begin{aligned}\mathscr{L}_{cls}^{gan} &= \mathscr{L}_{cls}\left(X_{ir}^{'}\right) = E_{x\in X_{ir}^{'}}[-\log p(x)],\\ \mathscr{L}_{tri}^{gan} &= \frac{1}{2}\left[\mathscr{L}_{tri}\left(X_{ir}^{'}, X_{ir}, X_{ir}\right) + \mathscr{L}_{tri}\left(X_{ir}, X_{ir}^{'}, X_{ir}^{'}\right)\right],\end{aligned} \tag{14}$$

where $p(\cdot)$ is the predicted probability of belonging to the ground-truth identity; the ground-truth identity of the fake IR image $X_{ir}^{'}$ should be the same as that of the original RGB image X_{rgb}.

Although the generated image X_{ir} can reduce cross-modality differences, there are still large intramodality differences caused by lighting, human pose, and view. We minimize the fake IR image $X_{ir}^{'}$ and the real IR image X_{ir} in a shared space via identity-based classification and triplet loss. We define these two losses as $\mathscr{L}_{cls}^{feat}$ and $\mathscr{L}_{tri}^{feat}$ and denote them in

$$\begin{aligned}\mathscr{L}_{cls}^{feat} &= \mathscr{L}_{cls}\left(X_{ir} \cup X_{ir}^{'}\right) = E_{x\in X_{ir}\cup X_{ir}^{'}}[-\log p(x)],\\ \mathscr{L}_{tri}^{feat} &= \mathscr{L}_{tri}\left(X_{ir}, X_{ir}^{'}, X_{ir}^{'}\right) + \mathscr{L}_{tri}\left(X_{ir}^{'}, X_{ir} X_{ir}\right),\end{aligned} \tag{15}$$

where $p(\cdot)$ represents the predicted probability that the input belongs to the ground-truth identity, and $\cup$ means the union sets. In summary, the overall loss of our module is shown in

$$\begin{aligned}\mathscr{L}_{\mathrm{ReID}} &= \lambda_1 \mathscr{L}_G + \lambda_2 \mathscr{L}_{D_I} + \lambda_3 \mathscr{L}_{cls}^{gan} + \lambda_4 \mathscr{L}_{tri}^{gan}\\ &\quad + \lambda_5 \mathscr{L}_{cls}^{feat} + \lambda_6 \mathscr{L}_{tri}^{feat},\end{aligned} \tag{16}$$

where $\mathscr{L}_G$ and $\mathscr{L}_{D_I}$ are calculated by equations (1) and (2). $\mathscr{L}_{cls}^{gan}$, $\mathscr{L}_{tri}^{gan}$, $\mathscr{L}_{cls}^{feat^I}$, and $\mathscr{L}_{tri}^{feat}$ are calculated by equations (14) and (15), respectively. Among them, $\lambda_1 = 1.0, \lambda_2 = 1.0$, $\lambda_3 = 0.1$, $\lambda_4 = 0.1$, $\lambda_5 = 1.0$, and $\lambda_6 = 1.0$.

4. Experiments

4.1. Datasets and Settings. We evaluate our model on SYSU-MM01 [10]. SYSU-MM01 is a very popular RGB-IR ReID dataset; it contains pedestrian images captured by six cameras, including two infrared cameras (camera3 and camera6), and four natural light cameras (camera1, camera2, camera4, and camera5). For each pedestrian, there are at least 400 RGB images and IR images with different poses and viewpoints. Among them, 296 IDs are used for training, 99 IDs are used for verification, and 96 IDs are used for testing. Following [29], there are two test modes, i.e., all-search mode and indoor-search mode. For the all-search mode, all images are used. For the indoor-search mode, only use indoor images from 1st, 2nd, 3rd, and 6th cameras. Both modes employ single-shot and multishot settings, in which 1 or 10 images of a person are randomly selected to form a gallery setting. Both modes use IR images as probe sets and RGB images as gallery sets.

Evaluation protocols: we use cumulative matching features (CMC) and mean average precision (mAP) as evaluation metrics. Following [29], the results of SYSU-MM01 are evaluated using the official code based on the mean of 10 repeated random splits of the gallery and probe set.

Implementation details: we use the ResNet-50 [30] pretrained on ImageNet as the CNN backbone, use the output of its pool5 layer as the feature map M, and use the average pooling to obtain the feature vector V. We add BNG-attention to each layer of residual blocks in ResNet-50 and MMM module after the third and fourth layers. For triplet loss, we use the FC layer to map the feature vector V into a 256-dimensional embedding vector. For classification loss, the classifier takes the feature vector V as input and includes a 256-dim fully connected (FC) layer, followed by batch normalization [25], dropout, and RELU as the middle layer, and an FC layer with the identity number logit as the output layer. The dropout rate is set at 0.5. We use PyTorch to implement the model, the images are data augmented by horizontal flipping, and the batch size is set to 72 (9 people, each of which has 4 RGB images and 4 IR images). For the learning rate, the learning rate of the generation module and discriminator module is set to 0.0002 and optimized using the Adam optimizer. We set the classifier and the embedder to 0.2 and the CNN backbone to 0.02 and optimize them by SGD.

4.2. Comparison with the Other Methods. In this section, we compare our method with several different cross-modality person ReID methods including the following methods: (1) with different structures and loss functions, two-stream [10], one-stream [10], zero-padding [10], BCTR [13], BDTR [13], D-HSME [26], and DGD + MSR [12] learned modality-invariant features and align them in feature space and (2) cmGAN [6] and JSIA [20] use the generative adversarial networks (GANs) to generate cross-modality IR images; they mitigate modal differences in pixel space. The experimental results are shown in Table 1.

In Table 1, we can find that there are various evaluation protocols, i.e., all-search/indoor-search and single-shot/multishot; firstly, for the same method, indoor-search performs better than all-search, because the images have less background variation in indoor mode, and matching is easier. Secondly, the rank scores of single-shot are lower than ones of multi-shot, but mAP scores of single-shot are higher than ones of multishot. This is because, in multishot mode, there are ten images in the gallery setting, while in single-shot, there is only one image. As a consequence, under the multishot mode, it is much easier to hit an image but difficult to hit all images. This situation is inverse under the single-shot mode.

The R1, R10, and R20 denote Rank-1, Rank-10, and Rank-20 accuracy (%). The mAP denotes the mean average precision score (%), and our model shows good performance. Compared with JSIA, our model achieves over 2.7% on Rank-1 and 2.49% on mAP in the single-shot setting of all-search mode. In the single-shot setting of indoor-search mode, our model achieves a rank-1 accuracy of 44.0% and an mAP of 52.96%. In the multishot setting of indoor search, our model achieves a rank-1 accuracy of 53.40%, and an mAP of 44.35%, which is higher than JSIA by 0.7% and 1.65%, respectively.

Table 1: Comparison of CMC (%) and mAP (%) performances with other methods on SYSU-MM01.

Methods	All-search								Indoor-search							
	Single-shot				multishot				Single-shot				multishot			
	R1	R10	R20	mAP	R1	R10	R20	mAP	R1	R10	R20	mAP	R1	R10	R20	mAP
Two-stream [10]	11.65	47.99	65.50	12.85	16.33	58.35	74.46	8.03	15.60	61.18	81.20	21.49	22.49	72.22	88.61	13.92
One-stream [10]	12.04	49.68	66.74	13.67	16.26	58.14	75.05	8.59	16.94	63.55	82.10	22.95	22.62	71.74	87.82	15.04
Zero-padding [10]	14.80	54.12	71.33	15.95	19.13	61.40	78.41	10.89	20.58	68.38	85.79	26.92	24.43	75.86	91.32	18.86
BCTR [13]	16.20	54.90	71.5	19.2	—	—	—	—	—	—	—	—	—	—	—	—
BDTR [13]	17.1	55.5	72.0	19.7	—	—	—	—	—	—	—	—	—	—	—	—
D-HSME [26]	20.7	62.8	78.0	23.2	—	—	—	—	—	—	—	—	—	—	—	—
cmGAN [6]	27.0	67.5	80.6	27.8	31.5	72.7	85.0	22.3	31.7	77.2	89.2	42.2	37.0	80.9	92.3	32.8
DGD + MSR [12]	37.35	83.40	93.34	38.11	43.86	86.94	95.68	30.48	39.64	89.92	97.66	50.88	46.56	93.57	98.8	40.08
JSIA-ReID [20]	38.10	80.70	89.90	36.90	45.10	85.70	93.80	29.50	43.80	86.20	94.20	52.90	52.70	91.10	96.40	42.70
Ours	**40.83**	**83.40**	**92.38**	**39.84**	**48.13**	**86.0**	93.67	**32.54**	**44.0**	**86.8**	**94.87**	**52.96**	**53.40**	90.52	95.70	**44.35**

Table 2: Ablation study in terms of CMC (%) and mAP (%) SYSU-MM01.

Method	SYSU-MM01			
	Single-shot all-search			
	R1	R10	R20	mAP
Baseline	34.13	78.86	90.07	33.54
B + BNG	39.60	81.95	91.60	37.93
B + MMM	39.97	82.38	92.54	39.52
B + BNG + MMM	40.83	83.40	92.38	39.84

Figure 5: Fake IR images generated by our module. The fake IR images can maintain identities and contents with original real RGB ones and have IR style.

4.3. Ablation Study. In this section, we design ablation experiments to test the effectiveness of the BNG module and MMM module. Our ablation experiments are performed on the dataset SYSU-MM01 and use the single-shot setting of all-search mode.

Influence of BNG module: the results of ablation experiments for BNG attention are shown in Table 2. Compared with the baseline model (B), by adding BNG attention, the rank-1 accuracy and mAP are improved by 5.57% and 4.39%, proving the effectiveness of BNG attention.

Influence of MMM module: as shown in Table 2, the model with MMM (B + MMM) achieves a rank-1 accuracy of 39.97% and an mAP of 39.52%, which are higher than those of the baseline (B) by 5.84% and 5.98%, respectively. It is proved that our proposed MMM module has good performance.

4.4. Visualization of Generated Images. For a more intuitive understanding of the generator model, we show the learned fake IR images in Figure 5. As shown in Figure 5, the first row is the real RGB image, the middle is the fake IR image generated by the generator, and the last row is the real IR image. We can observe that fake IR images have similar content (e.g., pose and view) and maintain the identity of the corresponding real RGB images while having an IR style. Therefore, the generated fake IR images can bridge the gap between RGB and IR images and can reduce cross-modality variation in pixel space.

5. Conclusion

In this paper, we proposed a new pixel and feature alignment network (PFANet) for the RGB-IR ReID task. The model consisted of a feature extractor, a generator, and a joint discriminator. The BNG attention and the MMM module were designed in the feature extraction module. Through these two modules, the model not only mitigated modality differences but also paid attention to channel and global

information. The cross-modality IR images were generated by the generator, which could bridge the gap between RGB and IR images and reduce cross-modality variation. Ablation experiments verified the effectiveness of each module. Extensive experiments on the SYSU-MM01 dataset illustrated that our model achieved state-of-the-art performance.

Data Availability

The SYSU-MM01 data used to support the findings of this study have been deposited in the "Rgb-infrared cross-modality person re-identification" repository (http://isee.sysu.edu.cn/project/RGBIRReID.html).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (grant nos. 51906217, 61906168, and 62176237), Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (grant no. LZJWZ22E090001), Zhejiang Provincial Natural Science Foundation of China under (grant no. LQ20F020024), and the Hangzhou AI Major Scientific and Technological Innovation Project (2022AIZD0061).

References

[1] X. Jin, C. Lan, W. Zeng, Z. Chen, and Li Zhang, "Style normalization and restitution for generalizable person re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3143–3152, Seattle, WA, USA, June 2020.

[2] J. Song, Y. Yang, Yi-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 719–728, Long Beach, CA, USA, June 2019.

[3] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2275–2284, Salt Lake City, UT, USA, June 2018.

[4] Z. Zhong, L. Zheng, S. Li, and Yi Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 172–188, Munich, Germany, August 2018.

[5] Z. Zheng, X. Yang, and Z. Yu, "Joint discriminative and generative learning for person re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2138–2147, Long Beach, CA, USA, June 2019.

[6] P. Dai, R. Ji, H. Wang, Q. Wu, and Y. Huang, "Cross-modality person re-identification with generative adversarial training," in *Proceedings of the IJCAI*, vol. 1, p. 6, Stockholm, Sweden, July 2018.

[7] H. Yi, N. Wang, X. Gao, J. Li, and X. Wang, "Dual-alignment feature embedding for cross-modality person re-identification," in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 57–65, Nice, France, October 2019.

[8] D. Li, X. Wei, X. Hong, and Y. Gong, "Infrared-visible cross-modal person re-identification with an x modality," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 4, pp. 4610–4617, 2020.

[9] Z. Wang, Z. Wang, Y. Zheng, Y.-Yu Chuang, and S.'ichi Satoh, "Learning to reduce dual-level discrepancy for infrared-visible person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 618–626, Long Beach, CA, USA, June 2019.

[10] A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, and J. Lai, "Rgb-infrared cross-modality person re-identification," in *Proceedings of the IEEE international conference on computer vision*, pp. 5380–5389, Venice, Italy, October 2017.

[11] M. Ye, Z. Wang, X. Lan, and P. C. Yuen, "Visible thermal person re-identification via dual-constrained top-ranking," in *Proceedings of the IJCAI*, vol. 1, p. 2, Stockholm, Sweden, July 2018.

[12] Z. Feng, J. Lai, and X. Xie, "Learning modality-specific representations for visible-infrared person re-identification," *IEEE Transactions on Image Processing*, vol. 29, pp. 579–590, 2020.

[13] M. Ye, X. Lan, J. Li, and P. Yuen, "Hierarchical discriminative learning for visible thermal person re-identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, Venice, Italy, October 2017.

[15] Y. Choi, M. Choi, and M. Kim, "Stargan: unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8789–8797, Salt Lake City, UT, USA, June 2018.

[16] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, Honolulu, HI, USA, June 2017.

[18] Yu-J. Li, Y.-C. Chen, Y.-Yu Lin, X. Du, and Yu-C. F. Wang, "Recover and identify: a generative dual model for cross-resolution person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8090–8099, Seoul, Korea, October 2019.

[19] G.'an Wang, T. Zhang, and J. Cheng, "Rgb-infrared cross-modality person re-identification via joint pixel and feature alignment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3623–3632, Seoul, Korea, October 2019.

[20] G.-An Wang, T. Zhang, Y. Yang et al., "Cross-modality paired-images generation for rgb-infrared person re-identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12144–12151, 2020.

[21] J. Hu, Li Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, Salt Lake City, UT, USA, June 2018.

[22] J. Park, S. Woo, L. Joon-Young, and K. In So, "Bam: bottleneck attention module," arXiv preprint arXiv:1807.06514, 2018.

[23] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, Salt Lake City, UT, USA, June 2018.

[24] S. Woo, J. Park, L. Joon-Young, and K. In So, "Cbam: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, Munich, Germany, August 2018.

[25] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International conference on machine learning*, pp. 448–456, PMLR, Lille, France, July 2015.

[26] H. Yi, N. Wang, and J. Li, "Hsme: hypersphere manifold embedding for visible thermal person re-identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8385–8392, 2019.

[27] X. Pan, P. Luo, J. Shi, and X. Tang, "Two at once: enhancing learning and generalization capacities via ibn-net," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 464–479, Munich, Germany, August 2018.

[28] F. Schroff, D. Kalenichenko, and P. James, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, Boston, MA, USA, June 2015.

[29] Y. Yang, Z. Lei, J. Wang, and Z. Stan, "In defense of color names for small-scale person re-identification," in *Proceedings of the 2019 International Conference on Biometrics (ICB)*, pp. 1–6, IEEE, Crete, Greece, June 2019.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 7259504, 11 pages
https://doi.org/10.1155/2022/7259504

Research Article

Sampled Characteristic Modeling and Forgetting Gradient Learning Algorithm for Robot Servo Systems

Hongbo Bi, Dong Chen, Yanjuan Li, and Ting You

College of Electrical and Information Engineering, Quzhou University, Quzhou, Zhejiang 324000, China

Correspondence should be addressed to Ting You; 8991548@qq.com

Received 12 June 2022; Revised 29 July 2022; Accepted 3 August 2022; Published 4 October 2022

Academic Editor: Sai Zou

Servo systems of robotic exhibit nonlinear coupling with multidimensional characteristics, which poses a challenge to existing modeling and identification techniques. According to a kind of robot servo system which runs repetitively operations over a prespecified finite time interval, a low-order sampling characteristic modeling method is derived in this work. Characteristic parameters are allowed to vary from both time axis and iteration one; the forgetting gradient learning algorithm is utilized to estimate characteristic parameters. Furthermore, the effectiveness of the proposed algorithms is proved via theoretical analyses and numerical simulations.

1. Introduction

As an important part of the industrial Internet system, robots are attracting more and more attention. Existing robot control methods mostly depend on models. Because of the dynamic characteristics of controlled objects, the complexity of control tasks and operating environment, it is often difficult to establish accurate mathematical models. Nonnegligible unmodeled dynamics have higher requirements on system models and the robust performance of closed-loop systems needs to be improved. In order to reduce the impact of unmodeled dynamics on the performance of control systems, most existing methods build high-order models of systems. Modeling and control technology for high-order systems is extremely complex and difficult to implement. The feature modeling method provides a predictable way to model complex systems. It does not need to establish a precise dynamic model of the system nor does it use a system reduction or linearization method. Instead, it expresses the model as a low-order time-varying difference equation, taking into account factors, such as system state, external environment, and control variables, and compresses the dynamic information into feature parameters. The Euler approximation is often used to discretize the continuous system, and the equivalent characteristic model [1–6] is given. The feature model can also be performed with exact discretization to give a sampling feature model of the system. Without limiting the sampling interval and control tasks, a feature model with rapidly changing (even abrupt) characteristic parameters may be obtained [7, 8].

Stochastic gradient algorithm and its derivative algorithm are popular for training network weights and avoid the matrix operation compared with the recurrence least squares algorithm and the small online computation, which has attracted much attention [9–17]. The consistent convergence results of the parameters are still available for the stochastic gradient algorithm under weaker incentives. The martingale theory and stochastic process theory are powerful tools to analyze the convergence of recurrence algorithms. From the published results, most methods are targeted for stationary systems, where the estimated parameters are stationary.

When dealing with time-varying systems, it is found that the stochastic gradient algorithms do not have the ability to track the time-varying parameters. In the field of time-varying system identification, more consideration is given to how to correct the recursive algorithms, that is, how to construct the correction algorithm of the recursive algorithms, so as to achieve effective tracking of the time-varying parameters and improve the convergence rate of the

parameters. It is well known that if the time-varying parameter change law is not known, there is no consistent parameter estimation convergence [18–21]. Based on this conclusion, one abandons the attempt to achieve complete tracking of time-varying parameters and instead works on a correction algorithm of how to construct and analyze recursive algorithms, expecting to give lower upper bounds on the estimation error of time-varying parameters. Common correction algorithms such as weighted recursive algorithms include rectangular windows, exponential windows (forgetting factors), etc. For irreducible nonstationary processes, these results can more accurately evaluate the parameter estimation accuracy, which has important applications in practical engineering applications.

When the system parameters are repetitive and independent, the instant change system runs repeatedly on a finite interval, the system parameters change over time, but the next time, it repeats this change, the instant change parameter to a repetition invariant. "Recursive" algorithms are constructed along the repetition axis, and the learning algorithm gives its complete estimates regardless of parameter slow change, fast change, and even mutation [22, 23]. However, in practical situations, there are often system parameters that change not only along the time domain but also with the number of iterations, that is, iteration dependence. If the time-varying parameter changes with the number of iterations are unknown, then similar to the recurrence algorithm, the iterative learning identification algorithm based on the principle of repeated invariant cannot track the system parameters effectively. We consider introducing the forgetting factor in the iterative learning algorithm and propose the forgetting gradient learning algorithm to estimate the iteration-dependent time-dependent system parameters. Then, under the premise of satisfying the repeated continuous incentive conditions, the performance analysis of the proposed algorithm is used and the simulation examples are completed to illustrate the effectiveness of the forgetting gradient learning algorithm.

2. Description of the Problem

Actual robot servo systems are mostly continuous nonlinear time-varying coupling processes, considering the following systems:

$$y^{(n)}(t) = f\Big(y(t), \dot{y}(t), \cdots, y^{(n-1)}(t), u(t), \dot{u}(t), \cdots, u^{(m)}(t), t\Big) + \xi(t), \tag{1}$$

where u stands for input and y for output, and $\xi(t)$ represents external disturbances.

With the development of computer control technology, the analysis and synthesis process of the robot servo system needs a sampling model. The following discrete nonlinear time-varying coupling processes are also considered:

$$y(k+1) = f(y(k), y(k-1), \cdots, y(k-n), u(k), u(k-1), \cdots, u(k-m), k) + \xi(k). \tag{2}$$

Here, we consider relaxing the initial condition that $f(0, 0, \cdots, 0, t) \wedge f(0, 0, \cdots, 0, k)$ can be arbitrary values.

The goal of this paper is to build a sampling feature model for a robot servo system. For iteration-dependent and time-varying feature parameters in the model, the amnestic gradient learning algorithm is used to estimate them. To further validate the effectiveness of the proposed learning algorithm, the random process theory is used to analyze the model and a numerical example is given to verify the effectiveness of the proposed algorithm.

The rest of this paper is arranged as follows: the third part establishes the sampling characteristic model of the servo system, the fourth part puts forward the forgotten gradient learning algorithm, the fifth part gives the convergence analysis process of the learning algorithm, the sixth part verifies by numerical examples, and the seventh part gives the conclusion of this paper.

3. Characteristic Modeling for the Servo System

We consider the following discrete nonlinear system:

$$\begin{aligned} y(k+1) &= f(y(k), y(k-1), \cdots, y(k-n), u(k), u(k-1), \cdots, u(k-m), k) \\ &\quad - f(0, y(k-1), \cdots, y(k-n), u(k), u(k-1), \cdots, u(k-m), k) \\ &\quad + f(0, y(k-1), \cdots, y(k-n), u(k), u(k-1), \cdots, u(k-m), k) \\ &\quad - f(0, 0, \cdots, 0, k) + f(0, 0, \cdots, 0, k) + \xi(k) \\ &= \frac{\partial f(\cdot)}{\partial y(k)} y(k) + \frac{\partial f(\cdot)}{\partial y(k-1)} y(k-1) + \cdots \frac{\partial f(\cdot)}{\partial u(k-m)} u(k-m) \\ &= \varphi^{\mathrm{T}}(k)\theta(k) + v(k), \end{aligned} \tag{3}$$

where $\varphi(k) = [y(k), y(k-1), \cdots, u(k-m)]$, $\theta(k) = [(\partial f(\cdot)/\partial y(k)), (\partial f(\cdot)/\partial y(k-1)), \cdots, (\partial f(\cdot)/\partial u(k-m))]$, $v(k) = f(0, 0, \cdots, 0, k) + \xi(k)$; in the same way, we can obtain the following:

$$y^{(n)}(t) = \frac{\partial f(\cdot)}{\partial y(t)} y(t) + \frac{\partial f(\cdot)}{\partial \dot{y}(t)} \dot{y}(t) + \cdots \frac{\partial f(\cdot)}{\partial u^{(m)}(t)} u^{(m)}(t) + f(0, 0, \cdots, 0, t) + \xi(t), \tag{4}$$

which denotes $a_i(t) = (\partial f(\cdot)/\partial y^{(i)}(t))$, $b_i(t) = (\partial f(\cdot)/\partial y^{(j)}(t))$, where $i = 0, 1, \cdots, n$; $j = 0, 1, \cdots, m$, then

$$y^{(n)}(t) = \sum_{i=0}^{n} a_i(t) y^{(i)}(t) + \sum_{j=0}^{m} b_j(t) u^{(j)}(t) + f(0, 0, \cdots, 0, t) + \xi(t). \tag{5}$$

The characteristic modeling method establishes a characteristic model for a nonlinear higher-order system by compressing the characteristic information to the characteristic parameters. Obviously, the lower the order of the characteristic model, the faster the change of the feature parameters. We consider building a first-order sampling characteristic model for the controlled system [7], and we can see that

$$y(k+1) = a(k) y(k) + b(k) u(k) + v(k). \tag{6}$$

In general, we can also set up second-order and third-order sampling feature models of the controlled system, that is, formula (5) can be written into formula (3).

3.1. Forgetting Gradient Learning Algorithm. We consider the following single-input single-output (SISO) discrete time-varying system repetitively operates over a prespecified finite time interval:

$$A\left(q^{-1}, k, t\right) y_k(t) = B\left(q^{-1}, k, t\right) u_k(t) + v_k(t), \tag{7}$$

where $t = (0, 1, \cdots, N)$ denotes time domain, and $k = (1, 2, \cdots)$ denotes iteration domain. $u_k(t)$ and $y_k(t)$ represent the input and output of the system, respectively. $v_k(t)$ is the interference variable. $A(q^{-1}, k, t)$ and $B(q^{-1}, k, t)$ are time-varying polynomials of shift operators of SISO discrete systems, where $A(q^{-1}, k, t) = 1 + a_{1,k}(t) + a_{2,k}(t) + \cdots + a_{n_a,k}(t)$ and $B(q^{-1}, k, t) = b_{1,k}(t) + b_{2,k}(t) + \cdots + b_{n_b,k}(t)$. $a_{1,k}(t), a_{2,k}(t), \cdots, a_{n_a,k}(t)$ and $b_{1,k}(t), b_{2,k}(t), \cdots, b_{n_b,k}(t)$ are the unknown parameters. We denote $\varphi_k(t) = [-y_k(t-1), -y_k(t-2), \cdots, -y_k(t-n_a), u_k(t-1), u_k(t-2), \cdots, u_k(t-n_b)]^T$, $\theta_k(t) = [a_{1,k}(t), a_{2,k}(t), \cdots, a_{n_a,k}(t), b_{1,k}(t), b_{2,k}(t), \cdots, b_{n_b,k}(t)]^T$, and $n = n_a + n_b$.

Equation (6) can be rewritten into the following regression model:

$$y_k(t) = \varphi_k^T(t) \theta_k(t) + v_k(t). \tag{8}$$

Similar to the stochastic gradient algorithm, a forgetting gradient learning algorithm for identifying iteratively dependent time-varying systems is presented:

$$\widehat{\theta}_k(t) = \widehat{\theta}_{k-1}(t) + \frac{\varphi_k(t)}{r_k(t)} \left[y_k(t) - \varphi_k^T(t) \widehat{\theta}_{k-1}(t) \right], \tag{9}$$

$$r_k(t) = \lambda r_{k-1}(t) + \left\| \varphi_k(t) \right\|^2. \tag{10}$$

3.2. Convergence Analysis of Forgetting Gradient Algorithm. The learning algorithm obtains parameter estimation based on the input and output data obtained when the system runs repeatedly in the operating interval. Because the operating interval is limited, we cannot obtain the convergence analysis results in the conventional sense. Only repetitive convergence results can be obtained. That is, for $t \in \{0, 1, \cdots, N-1\}$, the corresponding convergence classification is as follows

Repetitive consistency:

$$\lim_{k \longrightarrow \infty} \widehat{\theta}_k(t) = \theta(t), a.s. \tag{11}$$

Repetitive boundedness:

$$\lim_{k \longrightarrow \infty} \mathrm{E} \left\| \widehat{\theta}_k(t) - \theta_k(t) \right\|^2 \le \varepsilon < \infty, a.s. \tag{12}$$

When the system parameters are iteratively independent, using the learning algorithm, we can obtain the repeated consistency convergence results of the parameters. When the system parameters are iteratively dependent, we analyze the convergence performance of the forgetting gradient learning algorithm represented by equations (9) and (10).

For fixed time $t \in \{0, 1, \cdots, N-1\}$, $F_k(t)$ is denoted as a σ algebra consisting of the input and output data obtained by k repeated operations. In order to analyze the convergence of the proposed learning algorithm, the following assumptions are derived.

Hypothesis 1. $v_k(t)$ satisfies

$$E\left[v_k(t) \mid F_{k-1}(t) \right] = 0, a.s. \tag{13}$$

Hypothesis 2. There exists uniformly bounded $\sigma_v(t)$ with respect to t such that

$$E\left[v_k^2(t) \mid F_{k-1}(t) \right] \le \sigma_v^2, a.s. \tag{14}$$

Hypothesis 3. The following repetitive persistent excitation conditions are established:

$$\alpha(t) I \le \frac{1}{N} \sum_{i=1}^{N} \varphi_{k+i} \varphi_{k+i}^T \le \beta(t) I, a.s, \tag{15}$$

where both $\alpha(t)$ and $\beta(t) > 0$, $N > \mathrm{n}$.

From formula (10), we obtain as follows:

$$r_k(t) = \lambda r_{k-1}(t) + \left\| \varphi_k(t) \right\|^2 = \sum_{i=1}^{k} \lambda^{k-i} \left\| \varphi_i(t) \right\|^2 + \lambda^k r_0(t). \tag{16}$$

Then,

$$\begin{aligned} Nr_k(t) &= N\sum_{i=1}^{k}\lambda^{k-i}\|\varphi_i(t)\|^2 + N\lambda^k r_0(t) \\ &\le \sum_{i=1}^{k}\lambda^{k-i}\|\varphi_i(t)\|^2 + \sum_{i=0}^{k}\lambda^{k-i}\|\varphi_i(t)\|^2 + \cdots \\ &+ \sum_{i=2-N}^{k}\lambda^{k-i}\|\varphi_i(t)\|^2 + N\lambda^k r_0(t) \\ &\le \sum_{i=1}^{k}\lambda^{k-i}\left[\sum_{l=0}^{N-1}\|\varphi_{i-l}(t)\|^2\right] + \|\varphi_k(t)\|^2 + \sum_{i=0}^{1}\lambda^i\|\varphi_{k-i}(t)\|^2 \\ &+ \cdots + \sum_{i=0}^{N-2}\lambda^i\|\varphi_{k-i}(t)\|^2 + N\lambda^k r_0(t), \end{aligned} \tag{17}$$

where Hypothesis 3 is satisfied, $\sum_{l=0}^{N-1}\lambda^i\|\varphi_{i-l}(t)\|^2 \le nN\beta$, then

$$\begin{aligned} Nr_k(t) &\le nN\beta\sum_{i=1}^{k}\lambda^{k-i} + \|\varphi_k(t)\|^2 + \sum_{i=0}^{1}\|\varphi_{k-i}(t)\|^2 \\ &+ \cdots \sum_{i=0}^{N-2}\|\varphi_{k-i}(t)\|^2 + N\lambda^k r_0(t) \\ &\le nN\beta\sum_{i=1}^{k}\lambda^{k-i} + (N-1)nN\beta + N\lambda^k r_0(t) \\ &= \frac{1-\lambda^k}{1-\lambda}nN\beta + (N-1)nN\beta + N\lambda^k r_0(t), \end{aligned} \tag{18}$$

i.e.,

$$r_k(t) \le \frac{1-\lambda^k}{1-\lambda}n\beta + (N-1)n\beta + \lambda^k r_0(t). \tag{19}$$

Take limit of both sides of this inequality

$$\begin{aligned} \lim_{k\longrightarrow\infty} r_k(t) &\le \frac{1}{1-\lambda}n\beta + (N-1)n\beta \\ &= \frac{nN\beta - \lambda(N-1)n\beta}{1-\lambda}. \end{aligned} \tag{20}$$

For system (6), we define the transition matrix as follows:

$$L_{k+1,k}(t) = \left[I - \frac{\varphi_k(t)\varphi_k^T(t)}{r_k(t)}\right]. \tag{21}$$

The upper bound of $\lambda_{\max}[L_{k+N,k}^{\mathrm{T}}(t)L_{k+N,k}(t)]$ is then solved, and this bound is denoted as $A(t)$. Let $x_k(t)$ be the unit eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}(t)$ of the matrix $L_{k+N,k}^{\mathrm{T}}(t)L_{k+N,k}(t)$, and we construct the difference equation

$$x_{k+1}(t) = \left[I - \frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\right]x_k(t), \tag{22}$$

then $x_{k+N}(t) = L_{k+N,k}(t)x_k(t)$

Taking norm on both sides of the abovementioned equation, and it follows that $\|x_{k+N}(t)\|^2 = x_k^{\mathrm{T}}(t)L_{k+N,k}^{\mathrm{T}}(t)L_{k+N,k}(t)x_k(t) \le A(t)$, according to (22) and $r_k(t) > \|\varphi_k(t)\|^2$,

$$\begin{aligned} x_{k+1}^{\mathrm{T}}(t)x_{k+1}(t) &= x_k^{\mathrm{T}}(t)\left[I - 2\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)} + \left[\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\right]^2\right]x_k(t) \\ &\le x_k^{\mathrm{T}}(t)\left[I - 2\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)} + \frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\right]x_k(t) \\ &\le x_k^{\mathrm{T}}(t)x_k(t) - \frac{\|\varphi_k^{\mathrm{T}}(t)x_k(t)\|^2}{r_k(t)}. \end{aligned} \tag{23}$$

We transpose both sides of this inequality and that

$$\frac{\|\varphi_k^{\mathrm{T}}(t)x_k(t)\|^2}{r_k(t)} \le \|x_k(t)\|^2 - \|x_{k+1}(t)\|^2, \tag{24}$$

then

$$\sum_{i=0}^{N-1}\frac{\|\varphi_{k+i}^{\mathrm{T}}(t)x_{k+i}(t)\|^2}{r_{k+i}(t)} \le \|x_k(t)\|^2 - \|x_{k+N}(t)\|^2 \le 1 - A(t), \tag{25}$$

for any $i \in [0, N-1]$,

$$\begin{aligned} \|x_{k+i}(t) - x_k(t)\| &= \left\|\sum_{j=0}^{i-1}\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{r_{k+j}(t)}x_{k+j}(t)\right\| \\ &\le \sum_{j=0}^{i-1}\left\|\frac{\varphi_{k+i}(t)}{\sqrt{r_{k+j}(t)}}\right\|\left\|\frac{\varphi_{k+j}^{\mathrm{T}}(t)x_{k+j}(t)}{\sqrt{r_{k+j}(t)}}\right\| \\ &\le \sqrt{i(1-A(t))}. \end{aligned} \tag{26}$$

Taking trace to repetitive excitation condition A3), we obtain as follows:

$$\sum_{i=0}^{N-1}\|\varphi_{k+i}(t)\|^2 \le nN\beta, a.s. \tag{27}$$

In the condition (A3), we multiply $x_k^T(t)$ to the left and $x_k(t)$ to the right and use the formulas (15), (19)–(22) to obtain as follows:

$$\begin{aligned}
\alpha N &\le x_k^{\mathrm{T}}(t)\sum_{i=0}^{N-1}\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)x_k(t)\\
&\le \sqrt{\frac{nN\beta-\lambda(N-1)n\beta}{1-\lambda}}\left[x_k^{\mathrm{T}}(t)\sum_{i=0}^{N-1}\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}x_k(t)-x_k^{\mathrm{T}}(t)\sum_{i=0}^{N-1}\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}x_{k+i}(t)+x_k^{\mathrm{T}}(t)\sum_{i=0}^{N-1}\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}x_{k+i}(t)\right]\\
&\le \sqrt{\frac{nN\beta-\lambda(N-1)n\beta}{1-\lambda}}\left[\left\|x_k^{\mathrm{T}}(t)\sum_{i=0}^{N-1}\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}\left(x_k(t)-x_{k+i}(t)\right)\right\|+\left\|x_k^{\mathrm{T}}(t)\sum_{i=0}^{N-1}\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}\right\|x_{k+i}(t)\right]\\
&\le \sqrt{\frac{nN\beta-\lambda(N-1)n\beta}{1-\lambda}}\left[\left\|x_k^{\mathrm{T}}(t)\right\|\sum_{i=0}^{N-1}\left\|\frac{\varphi_{k+i}(t)\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}\right\|\left\|x_k(t)-x_{k+i}(t)\right\|+\left\|x_k^{\mathrm{T}}(t)\right\|\sum_{i=0}^{N-1}\left\|\varphi_{k+i}(t)\frac{\varphi_{k+j}^{\mathrm{T}}(t)}{\sqrt{r_{k+i}(t)}}x_{k+i}(t)\right\|\right]\\
&\le \sqrt{\frac{nN\beta-\lambda(N-1)n\beta}{1-\lambda}}\left[\sqrt{nN\beta}\sqrt{\frac{N(N-1)}{2}}\sqrt{1-A(t)}+\sqrt{nN\beta}\sqrt{1-A(t)}\right].
\end{aligned} \tag{28}$$

It is derived as follows:

$$A(t)\le 1-\frac{(1-\lambda)\alpha^2N}{((1/N)+2(\sqrt{(N-1/2N)})+(N-1/2))\left(n^2N\beta^2-\lambda n^2N\beta^2+\lambda n^2\beta^2\right)}. \tag{29}$$

Lemma 1. *Let the nonnegative sequence $x_k(t), a_k(t), b_k(t)$ satisfy the following relation:*

$$x_{k+1}(t)\le\left(1-a_k(t)\right)x_k(t)+b_k(t), \tag{30}$$

where $a_k(t)\in([0,1]), \sum_{k=1}^{\infty}a_k(t)=\infty, x_k(0)<\infty$, then

$$\lim_{k\longrightarrow\infty}x_k(t)\le\frac{b_k(t)}{a_k(t)}, \tag{31}$$

where the right limit is assumed to exist.

Suppose the observed noise $v_k(t)$ and the parameter iteration-dependent system parameter change rate $\omega_k(t)=\theta_k(t)-\theta_{k-1}(t)$ are zero-mean random noise sequences unrelated to the input $u_k(t)$ and the following relationships are satisfied:

$$\begin{aligned}
&\text{(A2) } \mathrm{E}\left[v_k(t)\right]=0, \mathrm{E}\left[\omega_k(t)\right]=0, \mathrm{E}\left[v_k(t)\omega_k(i)\right]=0,\\
&\text{(A3) } \mathrm{E}\left[v_k(t)v_k(i)\right]\\
&\qquad \mathrm{E}\left[v_k^2(t)\right]=\sigma_v^2(t)\le\sigma_v^2<\infty,\\
&\text{(A4)}\\
&\qquad \mathrm{E}\left[\left\|\omega_k(t)\right\|^2\right]=\sigma_\omega^2(t)\le\sigma_\omega^2<\infty.
\end{aligned} \tag{32}$$

If PE condition (A1) is satisfied, the parameter estimation error $\widehat{\theta}_k(t)-\theta_k(t)$ given by the forgotten gradient learning algorithm is repetitively bounded, which is solved below.

Solution: Define the parameter estimation error vector

$$\widetilde{\theta}_k(t)=\widehat{\theta}_k(t)-\theta_k(t). \tag{33}$$

We assume $\widetilde{\theta}_0(t)$ is independent of $v_k(t)$, and $E[\|\widetilde{\theta}_0(t)\|^2]<\infty$, it can be obtained by using the formulas (7) and (9).

$$\begin{aligned}
\widetilde{\theta}_k(t)&=\widehat{\theta}_{k-1}(t)-\theta_k(t)+\theta_{k-1}(t)-\theta_{k-1}(t)\\
&\quad+\frac{\varphi_k(t)}{r_k(t)}\left[\varphi_k^{\mathrm{T}}(t)\theta_k(t)+v_k(t)-\varphi_k^{\mathrm{T}}(t)\widehat{\theta}_{k-1}(t)\right],\\
&=\widetilde{\theta}_{k-1}(t)-\omega_k(t)-\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\widetilde{\theta}_{k-1}(t)\\
&\quad+\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\omega_k(t)+\frac{\varphi_k(t)}{r_k(t)}v_k(t),\\
&=\left[I-\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\right]\widetilde{\theta}_{k-1}(t)-\left[I-\frac{\varphi_k(t)\varphi_k^{\mathrm{T}}(t)}{r_k(t)}\right]\omega_k(t)\\
&\quad+\frac{\varphi_k(t)}{r_k(t)}v_k(t),\\
&=L_{k+1,k-N+1}(t)\widetilde{\theta}_{k-N}(t)-\sum_{i=0}^{N}L_{k+1,k-i+1}(t)\omega_{k-i}(t)\\
&\quad+\sum_{i=0}^{N-1}L_{k+1,k-i+1}(t)\frac{\varphi_{k-i}(t)}{r_{k-i}(t)}v_{k-i}(t).
\end{aligned} \tag{34}$$

Taking norms on both sides of formula (34)

$$\begin{aligned}\left\|\tilde{\theta}_k(t)\right\|^2 \le{}& \tilde{\theta}_{k-N}^{\mathrm{T}}(t)L_{k+1,k-N+1}^{\mathrm{T}}(t)L_{k+1,k-N+1}(t)\tilde{\theta}_{k-N}(t)\\ &+2\tilde{\theta}_{k-N}^{\mathrm{T}}(t)L_{k+1,k-N+1}^{\mathrm{T}}(t)\left[\sum_{i=0}^{N-1}L_{k+1,k-i+1}(t)\frac{\varphi_{k-i}(t)}{r_{k-i}(t)}v_{k-i}(t)\right.\\ &\left.-\sum_{i=0}^{N}L_{k+1,k-i+1}(t)\omega_{k-i}(t)\right]\\ &+\left\|\sum_{i=0}^{N-1}L_{k+1,k-i+1}(t)\frac{\varphi_{k-i}(t)}{r_{k-i}(t)}v_{k-i}(t)-\sum_{i=0}^{N}L_{k+1,k-i+1}(t)\omega_{k-i}(t)\right\|^2.\end{aligned} \tag{35}$$

Taking expectation on both sides of formula (35),

$$\begin{aligned}\mathrm{E}\left(\left\|\tilde{\theta}_k(t)\right\|^2\right) \le{}& A(t)\mathrm{E}\left\|\tilde{\theta}_{k-N}(t)\right\|^2+N\sum_{i=0}^{N-1}A(t)\frac{\left\|\varphi_{k-i}(t)\right\|^2}{\left\|r_{k-i}(t)\right\|^2}\left\|v_{k-i}(t)\right\|^2\\ &+(N+1)\sum_{i=0}^{N}A(t)\left\|\omega_{k-i}(t)\right\|^2\\ \le{}& A(t)\mathrm{E}\left\|\tilde{\theta}_{k-N}(t)\right\|^2+N\sum_{i=0}^{N-1}A(t)\frac{1}{r_{k-i}(t)}\sigma_v^2\\ &+(N+1)\sum_{i=0}^{N}A(t)\sigma_\omega^2\\ \le{}& A(t)\mathrm{E}\left\|\tilde{\theta}_{k-N}(t)\right\|^2+N^2A(t)\frac{1-\lambda}{\lambda^{N-1}n\alpha}\sigma_v^2\\ &+N(N+1)A(t)\sigma_\omega^2.\end{aligned} \tag{36}$$

From Lemma 1 and $A(t)<1$, it is derived as follows:

$$\begin{aligned}\mathrm{E}\left(\left\|\tilde{\theta}_k(t)\right\|^2\right) \le{}& \frac{N^2\left(1-\lambda/\lambda^{N-1}n\alpha\right)\sigma_v^2+N(N+1)\sigma_\omega^2}{1-A(t)}\\ \le{}& \frac{1+2N\sqrt{(N-1/2N)}+\left(N^2-N/2\right)}{(1-\lambda)\alpha^2}\\ &\left(n^2N\beta^2-\lambda n^2N\beta^2+\lambda n^2\beta^2\right)\left(\frac{1-\lambda}{\lambda^{N-1}n\alpha}\sigma_v^2+N(N+1)\sigma_\omega^2\right).\end{aligned} \tag{37}$$

Here, we give the convergence analysis results of the forget gradient learning algorithm in the case of parameter iteration dependence. According to formula (37), we can obtain the bounded convergence effect of the algorithm, that is, the parameter bounds converge to the true values and we give the bounds of the convergence bounds. From formula (37), when $\sigma_\omega^2=0$, which is equal to the system parameters iterate independently, we take $\lambda=1$. A random gradient learning algorithm is obtained.

$$\begin{aligned}\hat{\theta}_k(t)&=\hat{\theta}_{k-1}(t)+\frac{\varphi_k(t)}{r_k(t)}\left[y_k(t)-\varphi_k^{\mathrm{T}}(t)\hat{\theta}_{k-1}(t)\right],\\ r_k(t)&=r_{k-1}(t)+\left\|\varphi_k(t)\right\|^2.\end{aligned} \tag{38}$$

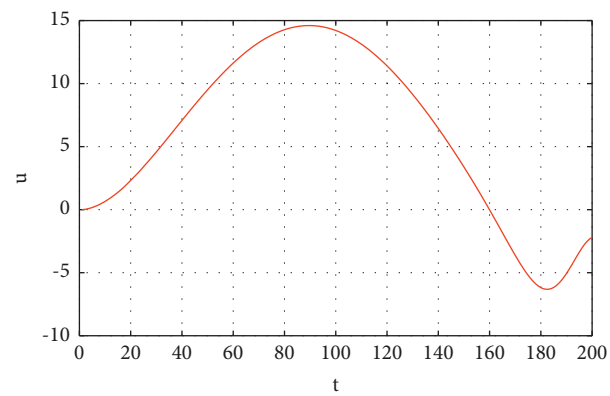


Figure 1: Input values.

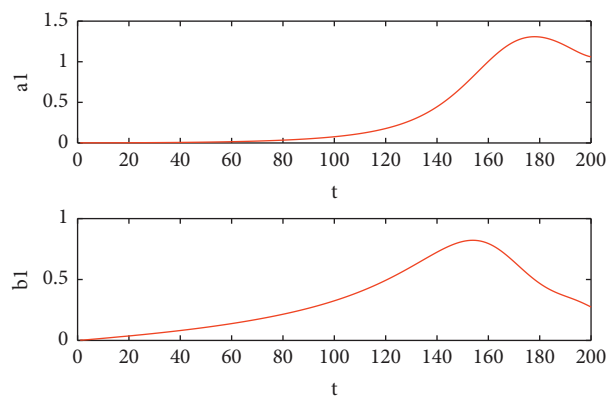


Figure 2: Parameters estimation.

In this case, the consistent convergence result of parameters can be obtained according to formula (37), that is, the parameter estimation converges completely to the parameter truth value.

4. Numerical Results

This section completes numerical examples to demonstrate that the learning identification algorithm can be used to estimate time-varying parameters in dynamic systems as shown in Figures 1 and 2.

Example 1. Consider the following nonlinear system:

$$\begin{cases}\dot{x}_1=x_2,\\ \dot{x}_2=\dfrac{1-e^{-x_1}}{1+e^{-x_1}}+\cos t\sin\left(x_2u\right)+u+\dot{u},\\ y=x_1.\end{cases} \tag{39}$$

The expected trajectory is $y_d(t)=10(10t^3-15t^4+6t^5)$. Using the sampling feature modeling method provided in this paper and the adaptive learning control method provided in reference [7], a first-order sampling feature model is established. Where the sampling time $T=0.005$, the initial

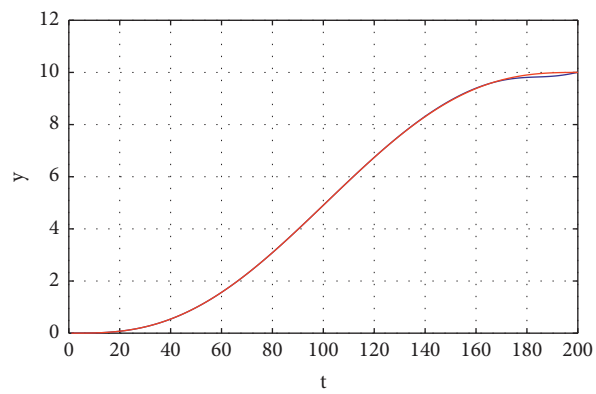


Figure 3: Tracking performance.

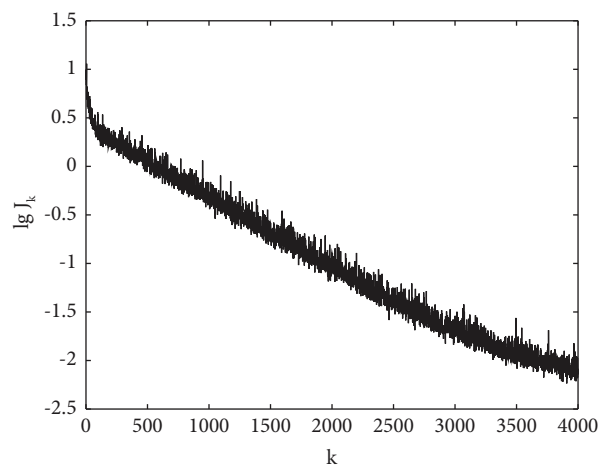


Figure 4: The error with respect to repetition.

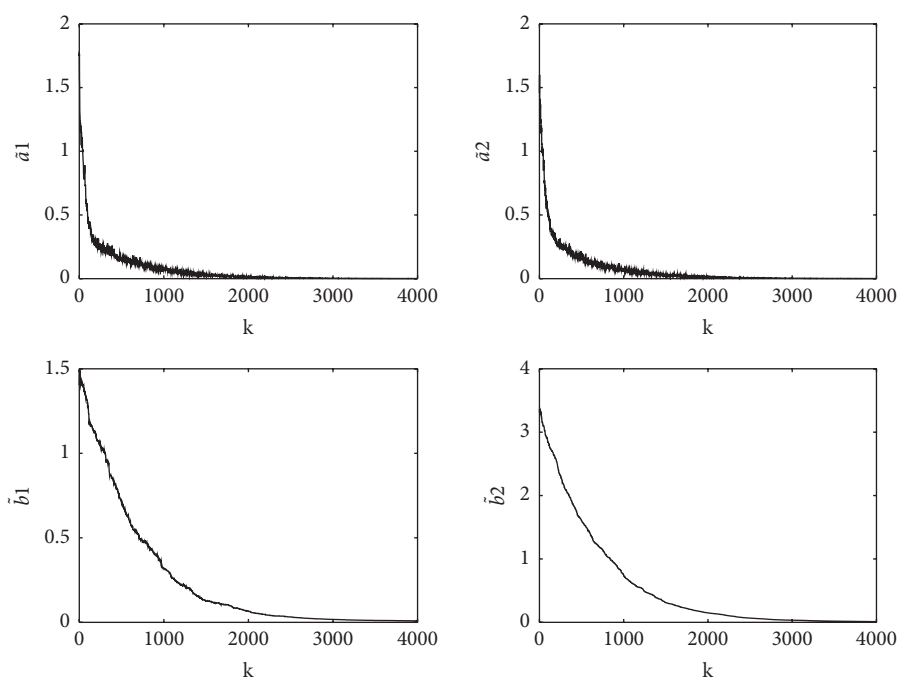


Figure 5: Parameter estimation errors after learning.

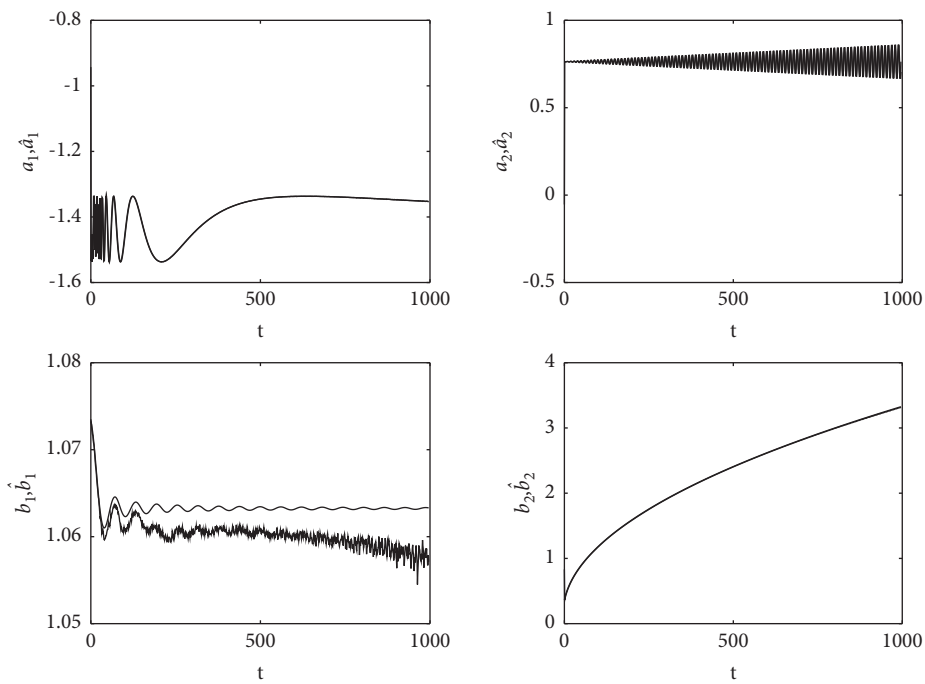


Figure 6: Parameter estimations after learning.

value $r = 1$; the forgetting factor $\lambda = 0.65$, we can obtain as follows:

Tracking performance is shown in Figure 3.

From the simulation results, using the sampled characteristic modeling method provided in this paper, although fast time-varying characteristic parameters are obtained, the output of the system enables efficient tracking of desired trajectories.

Example 2. Consider the following finite interval time-varying system.

$$\begin{aligned} y_k(t+1) + a_{1,k}(t)y_k(t) + a_{2,k}(t)y_k(t-1) \\ = b_{1,k}(t)u_k(t) + b_{2,k}(t)u_k(t-1) + v_k(t+1), \end{aligned} \tag{40}$$

where

$$\begin{aligned} a_{1,k}(t) &= -1.5 + 0.1\sin\left(\frac{50}{t}\right) + 0.001\sqrt{k}, \\ a_{2,k}(t) &= 0.7 + \frac{0.1t}{1000} * \sin\left(\frac{\pi t}{5}\right) + 0.001\sqrt{k}, \\ b_{1,k}(t) &= 1 + 0.1\frac{\sin(2\pi t/61)}{t} + 0.001\sqrt{k}, \\ b_{2,k}(t) &= 0.1 + 0.1\sqrt{t} + 0.001\sqrt{k}. \end{aligned} \tag{41}$$

In the simulation, we set finite interval length $N = 1000$. For $t = 1, 2, \cdots, N, k = 1, 2, \cdots, 4000, u_k(t)$ as uniformly distributed random variables on [-0.5, 0.5], forgetting factor $\lambda = 0.7$, $v_k(t) = 0.01$ *randn*. Here, *randn* is the production function of random variables that obey (0, 1) normal distribution. In the random dependence learning algorithms

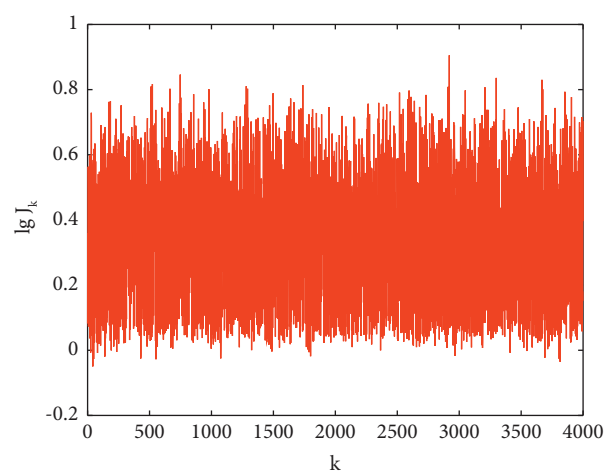


Figure 7: The error with respect to every interval.

(8) and (9), we set the initial value $r_0(t) = 1$, $\widehat{\theta}_0(t) = 0$. To examine the convergence performance, we define $J_k = \max_{1 \le t \le N} \lg|e_k(t)|, e_k(t) = y_k(t) - \varphi_k^T(t)\widehat{\theta}_{k-1}(t)$.

The simulation results are shown in Figure 4–6. The prediction error is shown in Figure 4, and the parameter estimation error is shown in Figure 5, and the parameters estimate values in Figure 6. It can be seen from Figure 4 that the prediction error decreases rapidly with the increase of the number of iterations. In Figure 5, the parameter estimation error asymptotically converges to zero in a small field, and the simulation results show the uniform convergence of the parameter estimation, which can almost converge to real values.

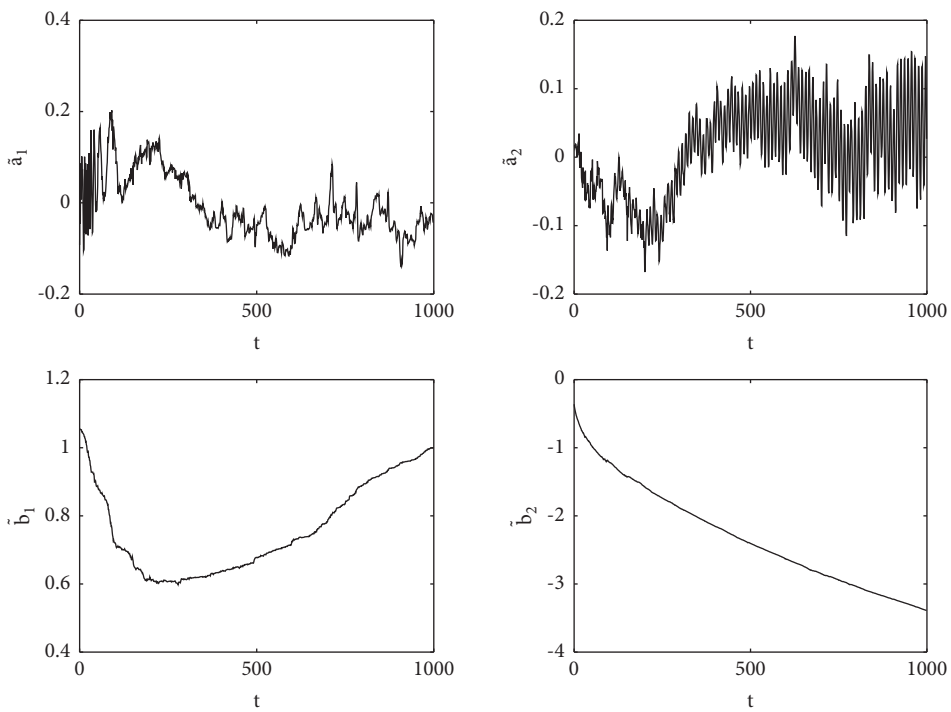


Figure 8: Parameter estimation errors after the last recursive process.

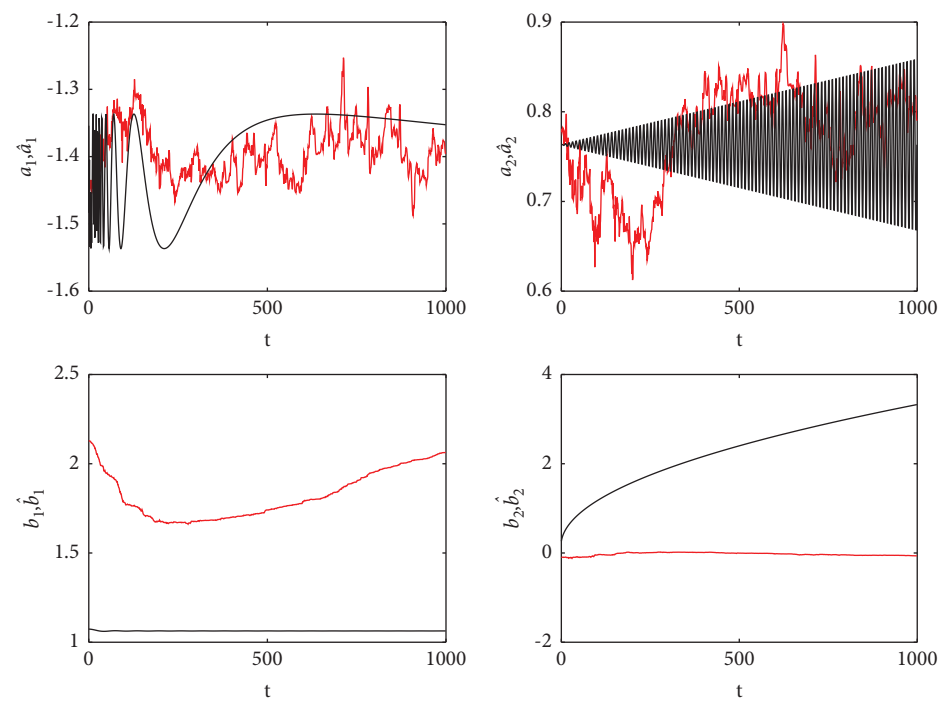


Figure 9: Parameter estimations after the last recursive process.

To demonstrate the effectiveness of the proposed algorithm, the simulation results are compared with stochastic gradient method appeared in reference [17], where the finite interval length $N = 1000 * 4000$ and the other conditions are same to which of the abovementioned simulation. The errors with respect to every interval are shown in Figure 7. Parameter estimation errors after last recursive process and parameter estimations after last recursive process are shown in Figures 8 and 9.

From the simulation results, the bounded convergence result is guaranteed, but the identification result is weaker than the result of the method presented in this paper.

5. Conclusions

As modeling technology is of great importance to robots for industrial Internet systems, the proposed sampled characteristic model and forgetting gradient learning identification method can be used to solve the parameter estimation problem of time-varying systems running round-trip over finite intervals. The forgetting gradient learning algorithm is derived for time-varying systems under finite-interval repeat operations. We prove the repetition boundedness of the learning algorithm under repeated continuous excitation conditions and give the estimation error bounds given by the proposed algorithm. The completed simulations also verify the effectiveness of the learning algorithm. Further, we present the convergence analysis results of the stochastic gradient learning algorithm, which can obtain consistent convergence results for the parameters when parameter iterations are independent. The main purpose of this paper is to propose this learning identification method and to clarify the connection and difference between the learning identification and the existing recursive identification algorithms. For the completeness of the theory and the expression simplicity, for the consistency analysis of the learning algorithm, we learn from the mature results of the learning algorithm. However, there are still differences between the two algorithms, such as the recurrence algorithm requires the PE condition along the time domain, while the learning algorithm requires the repeated PE condition; the assumption of the convergence consistency of the learning algorithm and the estimation of the obtained convergence rate are allowed to depend on time. Systematic results for recursive identification are presented in literature [17], from which we can learn for follow-up studies, including the case of system interference such as colored noise, continuous excitation, improvements of SPR conditions, and convergence rate estimation.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Project for Public Interest Research Projects of Science and Technology Program of Zhejiang Province, China (LGF21F010002, LGN21C130001, LGG21F030002, LGN22C140007).

References

[1] H. X. Wu and J. Hu, *Theory, Methods and Applications of Characteristic Modelling*, National Defense Industry Press, Arlington, VA, USA, 2019.

[2] J. F. Huang, Y. Kang, B. Meng, Y. Zhao, and H. Ji, "Characteristic model based adaptive controller design and analysis for a class of SISO systems," *Science China Information Sciences*, vol. 59, no. 5, Article ID 52202, 2016.

[3] S. G. Gao, H. R. Dong, and B. Ning, "Characteristic model-based all-coefficient adaptive control for automatic train control systems," *Science China Information Sciences*, vol. 57, no. 9, pp. 1–12, 2014.

[4] T. T. Jiang and H. X. Wu, "Sampled-data feedback and stability for a class of uncertain nonlinear systems based on characteristic modeling method," *Science China Information Sciences*, vol. 59, no. 9, Article ID 92205, 2016.

[5] X. Wang, Y. F. Wu, J. Guo, and Q. Chen, "Adaptive terminal sliding-mode controller based on characteristic model for gear transmission servo systems," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 1, pp. 219–234, 2019.

[6] T. T. Jiang, "Adaptive control and stability for characteristic model with unmodeled dynamics," in *Proceedings of the World Congress on Intelligent Control and Automation*, pp. 216–221, Guilin, China, June 2016.

[7] M. X. Sun, H. B. Bi, and J. Zhang, "Characteristic modeling and adaptive iterative learning control for nonlinear time-varying systems," *Journal of Systems Science and Mathematical Sciences*, vol. 36, no. 4, pp. 461–475, 2016.

[8] M. X. Sun, Z. L. Li, and L. J. Yu, "The first-order characteristic models of dynamic systems and adaptive iterative learning control of linear servo systems," *Journal of Systems Science and Mathematical Sciences*, vol. 32, no. 2, pp. 666–682, 2012.

[9] X. Y. Peng, L. Li, and F. Y. Wang, "Accelerating minibatch stochastic gradient descent using typicality sampling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4649–4659, 2020.

[10] Y. W. Lei, T. Hu, G. Y. Li, and K. Tang, "Stochastic gradient descent for nonconvex learning without bounded gradient assumptions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4394–4400, 2020.

[11] N. Costilla-Enriquez, Y. Weng, and B. Zhang, "Combining Newton-raphson and stochastic gradient descent for power flow analysis," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 514–517, 2021.

[12] S. A. M. Bin Al Islam, H. M. Abdul Aziz, and A. Hajbabaie, "Stochastic gradient-based optimal signal control with energy consumption bounds," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3054–3067, 2021.

[13] R. Bitar, M. Wootters, and S. El Rouayheb, "Stochastic gradient coding for straggler mitigation in distributed learning," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 277–291, 2020.

[14] Z. Wang and H. Q. Li, "Edge-based stochastic gradient algorithm for distributed optimization," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1421–1430, 2020.

[15] Z. X. Wu, Q. Ling, T. Y. Chen, and G. B. Giannakis, "Federated variance-reduced stochastic gradient descent with robustness to byzantine attacks," *IEEE Transactions on Signal Processing*, vol. 68, no. 7, pp. 4583–4596, 2020.

[16] D. M. Yuan, D. W. C. Ho, and S. Y. Xu, "Stochastic strongly convex optimization via distributed epoch stochastic gradient algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2344–2357, 2021.

[17] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction, and Control*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1984.

[18] L. Guo, *Time-varying Stochastic Systems Stability and Adaptive Theory*, Science Press, Beijing, China, 2020.

[19] F. Ding, T. Ding, J. B. Yang, and Y. M. Xu, "Convergence of forgetting gradient estimation algorithm for time-varying

parameters," *Acta Automatica Sinica*, vol. 28, no. 6, pp. 962–968, 2002.

[20] J. Chen, Y. J. Liu, F. Ding, and Q. Zhu, "Gradient-based particle filter algorithm for an ARX model with nonlinear communication output," *IEEE Transactions on systems, man and cybernetics: Systems*, vol. 50, no. 6, pp. 2198–2207, 2020.

[21] K. Y. You, "Recursive algorithms for parameter estimation with adaptive quantizer," *Automatica*, vol. 52, pp. 192–201, 2015.

[22] M. X. Sun and H. B. Bi, "Learning identification: least squares algorithms and their repetitive consistency," *Acta Automatica Sinica*, vol. 38, no. 5, pp. 698–706, 2012.

[23] M. X. Sun, H. B. Bi, and B. X. Chen, "Learning identification of a class of stochastic time-varing systems with colored noise," *Control Theory & Applications*, vol. 29, no. 8, pp. 974–984, 2012.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 1194169, 10 pages
https://doi.org/10.1155/2022/1194169

Research Article

A Real-Time UWB Location and Tracking System Based on TWR-TDOA Estimation and a Simplified MPGA Layout Optimization

Yanping Zhu,[1] Lei Huang,[1] Jing Liu,[2] Zhongkang Cao,[1] Jinli Chen,[1] and Zijian Mu[1]

[1]*School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[2]*School of Networks and Telecommunications Engineering, Jinling Institute of Technology, Nanjing 211199, China*

Correspondence should be addressed to Yanping Zhu; 001520@nuist.edu.cn

Received 7 March 2022; Revised 29 May 2022; Accepted 7 June 2022; Published 31 July 2022

Academic Editor: Sai Zou

This article designs and implements a 3D moving target positioning and tracking system by using ultrawideband (UWB) technology. The result of the two-way ranging by the time difference of arrival (TWR-TDOA) positioning algorithm is adopted to result in a new resolution that is to resolve the hyperbolic equations. The proposed algorithm is applied to outdoor and indoor scenarios. To minimize the effect of the sensor layout, this article proposes a simplified multi-population genetic algorithm (MPGA) to obtain the optimum distribution of anchors, which can rapidly reduce the number of search iterations. To resolve the low stability of the TDOA algorithm in outdoor and indoor scenarios, the Kalman filter algorithm is utilized to improve the stability and positioning accuracy of this system and a good simulation effect is achieved. The test results show that the system's positioning error is far less than that of using other methods. The whole system has a feature of high precision, high stability, low complexity, and low cost.

1. Introduction

With the development of communication technology, more and more systems are designed for wireless sensor network node location. The most popular techniques are Bluetooth, ZigBee, radio-frequency identification, and an ultra-wideband (UWB) wireless position [1, 2]. Among them, UWB technology is a promising ranging method that uses an extremely narrow pulse to transmit data with a wide spectrum range and fast communication speed. A UWB wireless position and tracking communication system has the advantages of low complexity, low power consumption, and high positioning accuracy against multipath interference. It can be applied in tunnel vehicle positioning, underground personnel positioning, and indoor positioning. It also has the potential to provide the position and tracking of lunar/Mars rovers and astronauts, where satellite navigation systems (such as Beidou or GPS) are not available [2].

For the location algorithms, there are traditional methods and hybrid algorithms. Received signal strength (RSS), angle of arrival (AOA), a difference of arrival (DOA), and time difference of arrival (TDOA) are four classical methods to realize the location. In recent years, a series of novel hybrid algorithms have been proposed to improve the system precision based on these four basic algorithms with new mathematical models or advanced solutions [3–5]. TDOA-UWB technology is widely developed as a hybrid positioning theory, which has superior localization performance compared with other approaches. The accuracy of most traditional TDOA-related algorithms based on the least square theory in a practical test is between 10 and 30 cm [6, 7]. Whistle, a novel TDOA localization without time synchronization is proposed by Xu et al. [8], which achieves high accuracy for 2D and 3D cases in acoustic source location framework, but this method has imperfections in energy efficiency and scalability. A UWB localization with a

new A-TDOA method is designed in [9], while its precision is not accurate enough and only suitable for 2D cases. These algorithms have bottlenecks in power consumption, precision, and transmission delay.

Some hybrid approaches were proposed to improve the system properties. In [10], a closed-form hybrid TDOA and AOA measurement with two observation stations in 3D space is proposed, which has high accuracy and low CRB. However, this method is only demonstrated in theoretical analysis without practical tests. A combined TDOA/TOF measurement is proposed by Mazraani et al. [11], which can achieve high accuracy and reduce power consumption, while this technique requires an accurate synchronization between anchors and tags. Based on TDOA equations, the Asymmetry Double-Sided Two-Way Ranging (ADS-TWR) algorithm and Chan-assisted Extended Kalman Filter in a 3D indoor positioning system with five anchors are used in Reference [12], which can achieve fast and high accurate localization.

Recently, deep learning algorithms were widely reported to solve the location problems [13–15], such as convolutional neural networks (CNN), K-nearest neighbor (KNN) algorithms, and long short-term memory (LSTM), which bring heavy computational burdens. These new frameworks further improve the positioning accuracy, in theory, however, due to the algorithms' complexities, the system has high requirements on the hardware platform in an actual scene.

The anchor layout is another important factor in the UWB location system. Most of the previous studies are implemented in 2D plane layout positioning. Due to the complexity of indoor and outdoor scenes, 3D positioning is more practical. There are some studies on the displacement of the number of nodes in 2D/3D scenes, which focus on positioning networks [16, 17]; however, to the best of our knowledge, the minimal anchors' layout is less studied. Considering the power consumption and the system cost, the optimum layout for fewer nodes needs to be discussed.

Aiming at high positioning accuracy and favorable stability, this article proposes a new solution based on two-way ranging (TWR) TDOA and a simplified multi-population genetic algorithm (MPGA) to obtain the optimum anchor layout. Effective experiments on indoor and outdoor scenes with specific hardware systems are provided. The followings are the main contributions of this article.

(1) The distance measured by the TWR algorithm is used to calculate the distance difference between the target and the two base stations, then the hyperbolic equation is established by using the distance differences. We proposed a new method for solving TDOA equations that is easy to implement on our platform.

(2) A 3D UWB position model and platform are built to test the proposed algorithm.

(3) An optimum anchor layout is calculated by a simplified 3D MPGA.

(4) Kalman filtering is adopted to solve the instability of the TDOA algorithm in outdoor and indoor scenes.

This article is organized as follows. The proposed 3D UWB model and TWR-TDOA algorithm are given in Section 2; Section 3 analyzes the existing error in the model, proposes the simplified MPGA layout algorithm, and adopts the Kalman filter to improve location precision; Section 4 outlines the hardware platform and the experimental measurement scenario settings, including the scenario tests. Section 5 gives the results and discussion. Finally, Section 6 summarizes the article and discusses the future directions.

2. System Model and Theoretical Location Method

The 3D UWB model is sketched in Figure 1. One tag is measured and four UWB anchors are used to reduce the system complexity and system consumption. The advantage of the system model is that the spatial information of the target can be obtained with fewer anchors [12]. To obtain the target location, firstly, the asymmetric double-sided two-way ranging algorithm is used to calculate the distance difference between the target and the two base stations, which has low requirements for system synchronization. Secondly, according to the measurement distance, the TDOA algorithm is adopted, and a new resolution for hyperbolic equations is proposed to obtain high location accuracy.

2.1. Double-Sided Two-Way Ranging with Three Messages. For the hardware platform, the mainstream UWB device is DW1000, which uses two-way ranging without synchronization. The traditional algorithms can be easily transplanted into this platform. According to the user manual of DW1000, the ADS-TWR algorithm with three messages is used to measure the distance. This algorithm simplifies the DS-TWR from four messages to three messages by using the reply of the first roundtrip measurement as the initiator of the second roundtrip measurement [10], which means lower power consumption. The transmission process between the tag and anchors is schematically shown in Figure 2.

As shown in Figure 2, the tag is named Device A and the anchor is named Device B. Suppose the flight of time is T_{prop}. After the sensor receives signals, the delay time is T_{reply}, which is generated by transmission processing and other reasons. The roundtrip communication time between the anchor and tag is totally T_{round}. The full process of these three messages communication is described as follows: (a) device A sends a packet to device B, and the transmission time is T_{prop}; (b) device B records the time while receiving this packet; (c) device B answers in response to receiving packet after a fixed delay time T_{reply}; (d) device A records the moment when receiving the message from Device B.

It is easy to calculate the roundtrip time T_{round}:

$$T_{\text{round}} = 2T_{\text{prop}} + T_{\text{reply}}. \tag{1}$$

Here, the time of flight (TOF) as ΔToF is given by

$$\Delta\text{ToF} = \frac{T_{\text{round1}} \times T_{\text{round2}} - T_{\text{reply1}} \times T_{\text{reply2}}}{\left(T_{\text{round1}} + T_{\text{reply1}}\right) + \left(T_{\text{round2}} + T_{\text{reply2}}\right)}. \tag{2}$$

In asymmetrical double-sided two-way ranging, it is not necessary that the reply time of the two devices be synchronous, which means reducing the clock requirements of the system. Under the same information transmission, ADS-TWR can save message flow, which means saving battery power and space-time. The clock frequency errors can be controlled at the picosecond level when the quality of the crystal oscillator is not high. The most important error affecting accuracy depends on the following equation:

$$\text{error} = \Delta\text{ToF} \times \left(1 - \frac{k_a + k_b}{2}\right), \tag{3}$$

where the actual frequency of device A is k_a times the expected frequency, and the actual frequency of device B is k_b times the expected frequency. k_a and k_b are close to 1 in the application. The TWR data are exported for positioning processing. Take the 100 m UWB communication range for an example, the TOF is around 300 ns, whereas the error time is about 6 ps, and the corresponding range error is 2 mm.

2.2. TDOA Algorithm. According to the system model in Figure 1, suppose the coordinate of the target tag is $T(x, y, z)$, the number of UWB anchors is M, where there is one primary sensor S_0 and $(M-1)$ secondary sensors S_i, their coordinates are (x_i, y_i, z_i), and $i = 0, 1, \ldots, M-1$. Suppose the time of arrival to each station is $t_i\,(i = 0, 1, 2, \ldots, M-1)$. The time difference between each secondary and primary sensor is written as ΔToF_i, $(i = 1, 2, \ldots, M-1)$, which is determined by

$$\Delta r_i = c\Delta\text{ToF}_i. \tag{4}$$

The distance differences can also be directly written as the Euclidean distance from the target to the secondary station minus the distance from the target to the primary station, which is given by

$$\Delta r_i = D_i - r_0, \tag{5}$$

where $D_i = T - S_{i2}$, $i = 1, 2, 3$ and $r_0 = T - S_{02}$.

Then, we have

$$(\Delta r_i + r_0)^2 = D_i^2. \tag{6}$$

subtracting r_0^2 at both sides, we have

$$\begin{aligned}\Delta r_i^2 + 2\Delta r_i r_0 &= D_i^2 - r_0^2,\\ &= 2x(x_0 - x_i) + 2y(y_0 - y_i) + 2z(z_0 - z_i) + d_i^2 - d_0^2,\end{aligned} \tag{7}$$

where $d_i = S_{i2}$, $i = 1, 2, 3$ and $d_0 = S_{02}$.

We obtain the following equation:

$$x(x_0 - x_i) + y(y_0 - y_i) + z(z_0 - z_i) = \Delta r_i r_0 + \frac{\Delta r_i^2 + d_0^2 - d_i^2}{2}. \tag{8}$$

There are M equations in equation (8), and we rewrite it as a matrix form, where x, y, z are unknown and $l_i = d_i^2 - d_0^2 - \Delta r_i^2/2$.

$$\begin{bmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \\ \vdots & \vdots & \vdots \\ x_M - x_0 & y_M - y_0 & z_M - z_0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\Delta r_1 \\ -\Delta r_2 \\ \vdots \\ -\Delta r_M \end{bmatrix} r_0 + \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_M \end{bmatrix}. \tag{9}$$

Let $\mathbf{AX} = \mathbf{B}$, then

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \\ \vdots & \vdots & \vdots \\ x_M - x_0 & y_M - y_0 & z_M - z_0 \end{bmatrix},\\ \mathbf{B} &= r_0\mathbf{C} + \mathbf{D} = r_0 \begin{bmatrix} -\Delta r_1 \\ -\Delta r_2 \\ \vdots \\ -\Delta r_M \end{bmatrix} + \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_M \end{bmatrix}.\end{aligned} \tag{10}$$

According to the linear properties of linear equations, the solution of $\mathbf{AX} = \mathbf{B}$ is the summation of r_0 times the solution of $\mathbf{AX} = \mathbf{C}$ and the solution of $\mathbf{AX} = \mathbf{D}$. When $M = 3$, $\mathbf{A}$ is a square matrix, according to the Cramer criterion, the solution is $x_{ij} = |\mathbf{A_j}|/|\mathbf{A}|$, where $\mathbf{A}_j$ is the determinant obtained by replacing the jth column element in a constant term. The solution is obtained as follows:

$$\begin{cases} x = a_1 r_0 + b_1, \\ y = a_2 r_0 + b_2, \\ z = a_3 r_0 + b_3, \end{cases} \tag{11}$$

where a_i is the solution of $\mathbf{AX} = \mathbf{C}$, and b_i is the solution of $\mathbf{AX} = \mathbf{D}$.

In equation (11), r_0 is an unknown parameter. We calculate it from the following method: using the definition of r_0, we get a bivariate quadratic equation about r_0, as shown in equation (12):

$$Er_0^2 + 2Fr_0 + G = 0, \tag{12}$$

where

$$\begin{cases} E = a_1^2 + a_2^2 + a_3^2 - 1, \\ F = a_1(b_1 - x_0) + a_2(b_2 - y_0) + a_3(b_3 - z_0), \\ G = (x_0 - b_1)^2 + (y_0 - b_2)^2 + (z_0 - b_3)^2. \end{cases} \tag{13}$$

We can get the roots of this one variable quadratic equation:

$$r_0 = \frac{-F \pm \sqrt{F^2 - EG}}{E}. \tag{14}$$

According to this root, the TDOA algorithm for the 3D situation has three different position results: ambiguity,

precision, and loss. We need to handle these three cases differently.

(1) r_0 has two roots (one of which is a false solution). From the two roots, two location positions can be obtained, which means that location ambiguity needs an increased anchor. In this algorithm, we take the real part of the conjugate complex as the root $r_0 = -F/E$.

(2) r_0 has only one root, and the target position can be uniquely determined.

(3) No solution. This means the target location cannot be determined. For a moving target, we take the position and motion state of the previous time to fill the missing value.

3. Deviation Error Analysis and Improved Algorithms

The core of the TDOA algorithm is associated with the difference between the primary anchor and the secondary anchor. First, the derivation of the difference is expressed as

$$d\mathbf{r} = \mathbf{H}d\mathbf{u} + d\mathbf{s}, \tag{16}$$

where $d\mathbf{r} = [d\Delta r_1 d\Delta r_2 \cdots d\Delta r_4]^T$, $\mathbf{H} = [\ (x - x_1/r_i)\ -(x - x_0/r_0)(y - y_1/r_i) - (y - y_0/r_0)(z - z_1/r_i)\ -(z - z_0/r_0)\ (x - x_2/r_i) - (x - x_0/r_0)(y - y_2/r_i) - (y - y_0/r_0)(z - z_2/r_i) - (z - z_0/r_0)\ \vdots\ \vdots\ \vdots\ (x - x_M/r_i) - (x - x_0/r_0)\ (y - y_M/r_i) - (y - y_0/r_0)(z - z_M/r_i) - (z - z_0/r_0)]$, $d\mathbf{u} = [dx dy dz]^T$ $ds = [k_1 - k_0\ k_2 - k_0 \cdots k_M - k_0]^T$, $k_i = (x - x_i/r_i)dx_i + (y - y_i/\ r_i)dy_i + (z - z_i/r_i)dz_i, i = 1, 2, \ldots, M$.

Making use of equation (16), one can get a tag position error vector:

$$d\mathbf{u} = \mathbf{H}^{-1}(d\mathbf{r} - d\mathbf{s}). \tag{17}$$

Based on this equation, the accuracy of the tag position is correlative with the sensor distribution and the measurement error. In the following part, the accuracy of the TDOA algorithm will be improved in both aspects.

3.1. Optimized Distribution Based on a Simplified MPGA. The different sensor placement strategies influence the geometric dilution of precision (GDOP) of the target positioning. The GDOP can be expressed as [18]

$$\mathrm{GDOP} = \sqrt{\mathrm{trace}(\mathbf{P})}, \tag{18}$$

where $\mathbf{P} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{M}\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}$, $\mathbf{M} = \begin{bmatrix} \sigma_{r1}^2 + \sigma_s^2 & \sigma_s^2 & \cdots & \sigma_s^2 \\ \sigma_s^2 & \sigma_{r2}^2 + \sigma_s^2 & \cdots & \sigma_s^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_s^2 & \sigma_s^2 & \cdots & \sigma_{rM}^2 + \sigma_s^2 \end{bmatrix}$. The variance of the location error is σ_s^2, and the variance of the measurement error is $\sigma_{ri}^2 (i = 1, 2, \ldots, M)$.

When the GDOP of the located target is the smallest, the distribution can reach the best condition. This article adopts the MPGA to find the best anchor layout whose GDOP of the space is nearly optimal.

In this part, $\mathrm{AVER}_{\mathrm{GDOP}}$ is treated as the average GDOP of one plane whose height is h.

$$\mathrm{AVER}_{\mathrm{GDOP}}(h) = \frac{1}{N \times K}\sum_{i=0}^{N}\sum_{j=0}^{K}\mathrm{GDOP}\left(x_i, y_j, h\right), \quad i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, K. \tag{19}$$

The genetic algorithm searches for the best result by producing subsequent generations from a random population with genetic operators. Compared with the genetic algorithm, MPGA uses more populations to produce individual fitness as a multi-population crossover [19]. Whole space-wide searching brings a huge amount of computation, whereas one horizontal plane searching can greatly reduce computing and has similar precision. Data statistics via MATLAB simulation show that the amount of simplified calculation is reduced by 70% compared with the traditional search algorithms. In this part, this simplified MPGA method is adopted to calculate the optimal layout.

The detailed process of this simplified MPGA for optimized distribution is shown in Algorithm 1, where $\mathbf{X}^{2\times 12}$ represents the range matrix of the given space, h is the height of the measured plane, MAXGEN limits maximum iteration times, and N has the information on the population size. In addition, the probability rate of the crossover is P_c and the probability rate of the mutation is P_m.

At the beginning of each experiment, the algorithm can be used to ensure the minimum value of the average GDOP in the measured range. In this MPGA algorithm, the population size N is set as 10 and the maximum iteration time MAXGEN is set as 5. The crossover rate P_c is one value between 0.7 and 0.9. The mutation probability is a random value between 0.001 and 0.05. Considering the convenience of placement and calculation, the anchors are set on either side of the scene and one anchor is fixed in the condition of the algorithm.

The following experiment is designed to test the efficiency of the optimized distribution based on simplified MPGA. Let $h = 80, \mathrm{Genmax} = 10$ and $X^{2\times 12} = \begin{bmatrix} 0 & 0 & 0 & 0 & 470 & 0 & 0 & 470 & 0 & 0 & 470 & 0 \\ 0 & 0 & 0 & 495 & 470 & 370 & 495 & 470 & 370 & 495 & 470 & 370 \end{bmatrix}$.

Table 1 gives the GDOP of random positions at the height of 80 cm and compares the results with simplified MPGA and unoptimized distribution separately.

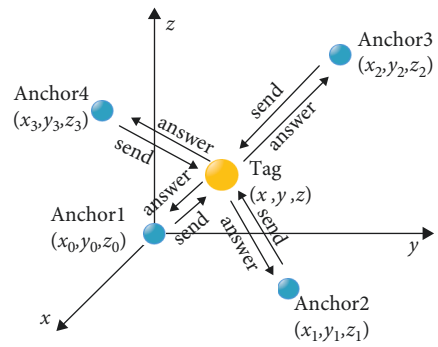


Figure 1: The 3D mathematical model.

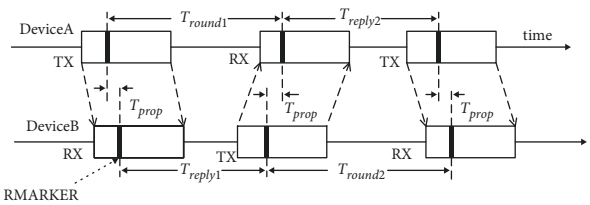


Figure 2: ADS-TWR with three messages.

Here we can see that, if anchors are set at the positions located by the MPGA, the GDOP can be reduced largely. This means the optimized distribution can decrease the deviation of measured data. Thus, this proposed MPGA can effectively decrease the sensor distribution error caused by the system.

3.2. Kalman Filter. The TDOA is unstable in the positioning process, resulting in a measurement error. To resolve this problem, we use the Kalman filter to improve the stability and precision of positioning. Kalman filtering can achieve the best tracking performance when dealing with a target that has a measurement equation and motion equation that are linear and a process noise that obeys Gaussian distribution [20]. When there is no solution in the TDOA algorithm, for the target position and motion state at the previous time, the constant velocity (CV) model is used to fill in the missing position. The CV model is given as

$$X(k+1) = F(k)X(k) + W(k), \tag{20}$$

$$Z(k+1) = H(k+1)X(k+1) + V(k+1). \tag{21}$$

The state equation is given by equation (20): $\mathbf{X}(k)$ is the state vector, which is an N-dimensional column vector. Here, $N = 9$; $\mathbf{F}(k)$ is called the state transition matrix, $\mathbf{W}(k)$ is the process noise, and its covariance matrix is $\mathbf{Q}(k)$, which is an $N \times N$ matrix in this system. Equation (21) is the measurement equation: $\mathbf{Z}(k)$ is called the measurement vector and is the M-dimensional column vector ($M = 3$). $\mathbf{H}(k)$ is the measurement matrix and $\mathbf{V}(k)$ is the measurement noise. The idea of the filling method is to estimate the missing position by using the state transition matrix and measurement matrix, i.e., the last measured value correction process is removed while using the Kalman filter.

The initial test state of the system can be obtained by using the two-point difference method for initialization:

$$\widehat{X}(1|1) = \begin{bmatrix} \widehat{x}(1|1) \\ \widehat{\dot{x}}(1|1) \end{bmatrix} = \begin{bmatrix} Z(1) \\ \dfrac{Z(1) - Z(0)}{T} \end{bmatrix}. \tag{22}$$

The covariance matrix of the initial test state error is as follows:

$$\mathbf{p}(1|1) = \begin{bmatrix} r & \dfrac{r}{T} \\ \dfrac{r}{T} & \dfrac{2r}{T^2} \end{bmatrix}. \tag{23}$$

After initialization, the Kalman filter can start working from the time $k = 2$. The standard deviation of the measured range data is calculated by

$$\widehat{\sigma}_{\mathbf{m}} = \sqrt{\frac{1}{\mathbf{N}} \sum_{\mathbf{k}=1}^{\mathbf{N}} \left(\mathbf{r}(\mathbf{t_k}) - \mathbf{X}(\mathbf{t_k})\right)^2}. \tag{24}$$

The data is collected from the four base stations, and the moving model is the CV. The TDOA data are further processed by the Kalman filter. Figure 3 shows the estimations of the position of Chan-EKF [12], TWR-TDOA, and TWR-TDOA-KF algorithm when the tag moves with a speed of 0.2 m/s under the 3D coordinates. The algorithm and simulation environment are implemented with MATLAB. The simulator implements the process of positioning in the condition of one tag and four anchors. The tag moves on the straight line ((0,0,0) m to (4,4,4) m). Four anchors adopt the proposed MPGA optimal layout. The test records measurement results 20 times.

The root means square error (RMSE) is employed in Figure 4 to measure the accuracy of position in different conditions of noise standard deviation. The RMSE can be calculated by the following equation:

$$\text{RMSE} = \sqrt{E\left[(x_k - x)^2 + (y_k - y)^2 + (z_k - z)^2\right]}, \tag{25}$$

where (x, y, z) represents the actual position of the tag and (x_k, y_k, z_k) represents the estimated location of the tag.

Input: $\mathbf{X}^{2\times12}$ (the range matrix of the given space), h (the height of the measured plane), MAXGEN (maximum iteration times), N (the population size), P_c (the crossover rate), P_m (the mutation probability)
Output: Ax (the anchor coordinates)

```
generation = 0;
Initialize the populations and generate plenty of random coordinates Ax in each population according to X^{2×12};
Evaluate the fitness values via equation (19) and all chromosomes are ranked according to the fitness values;
while generation < MAXGEN
  for i←1 to N
    Calculate the fitness of individuals in population[i] according to X^{2×12};
    Select the better chromosomes as parents that have the more fitness values;
    Cross the chosen parents to produce new offspring at a probability of Pc;
    Mutate the new offspring at a probability of P_m;
    Calculate the fitness of the individual in the offspring;
    Reinserts offspring in the population[i];
  end
  for i←1 to N
    if i = N
      Replace the worst chromosome in population[1] with the best one in population[N];
    else
      Replace the worst chromosome in population[i + 1] with the best one in population(i);
    end
  end
  Select the best individual in each population as the elite population;
  Select all the converted chromosomes to find the minimum value and to reassemble the chromosomes;
  If the new minimum value is less than the old one
    generation = generation + 1;
    Save the new minimum value and update Ax according to the new chromosomes;
  end
end
```

Algorithm 1: The distribution of anchors is optimized via simplified MPGA.

Table 1: GDOP value for different anchors.

Point (cm)	GDOP	
	Optimized distribution	Unoptimized distribution
(100, 100, 80)	1.6987	2.7863
(50, 60, 110)	1.8452	3.0972
(300, 360, 10)	1.6425	155.2138

It can be concluded that the TWR-TDOA-KF algorithm has better performance than the TWR-TDOA algorithm and the Chan-EKF. In addition, the optimized algorithms have better performance than the unoptimized algorithms. More accurately, for three different noise standard deviations, the average RMSE of the optimized TWR-TDOA-KF algorithm is lower than the Chan-EKF by 1.67 cm, lower than the TWR-TDOA-EKF by 1.03 cm. And the average RMSE of the TWR-TDOA-KF is lower than TWR-TDOA by 1.51 cm.

4. Hardware Platform and Measurement Scenario

The system measurement model is shown in Figure 5. The hardware part is composed of a single-chip microcomputer and a UWB communication module. In our system, only the primary anchor is connected to the upper PC through a USB data cable, and the secondary anchors are wireless communication. ADS-TWR ranging and wireless data transmission are applied between the tag and the anchors through the UWB communication module. After measuring the distance, the secondary anchors transmit the data to the anchor, and then the primary anchor transmits the data to the upper computer through a USB to calculate the target position. Here, we use the optimum anchors' distribution based on the proposed simplified MPGA algorithm. For the internal hardware module structure of the tag and anchors, we adopt the DWM1000 system. Time error is on the nanosecond level and the typical update rate is 3.5 Hz. This system has a feature of low cost, high ranging accuracy, high positioning accuracy, and fast ranging speed, and the serial port of hardware is very robust. Compared with the traditional ranging mode, the power consumption is much lower [21].

4.1. An Outdoor Scenario. The layout of our outdoor experiment scenario is shown in Figure 6(a), and it is implemented in cropland. Anchor 1 is connected to the PC as the primary device, and anchors 2, 3, and 4 are secondary devices. In this experiment, $h = 80$, Genmax = 10, and $X^{2\times12} = \begin{bmatrix} 0 & 0 & 0 & 0 & 470 & 0 & 0 & 470 & 0 & 0 & 470 & 0 \\ 0 & 0 & 0 & 495 & 470 & 370 & 495 & 470 & 370 & 495 & 470 & 370 \end{bmatrix}$.

After calculating by the simplified MPGA, anchor 1 is set as (0,0,0), and the other anchors are set as (0,470,370), (490, 0,370), and (495,470,0), respectively.

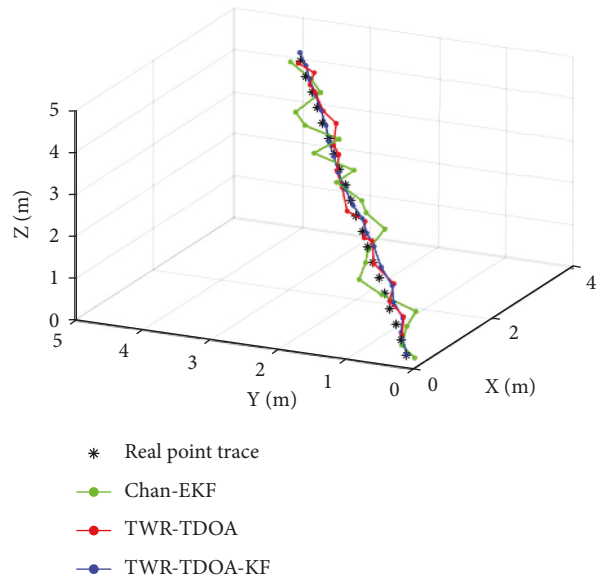


Figure 3: Experimental position estimation of three algorithms under 3D coordinates when tag moving on a straight line.

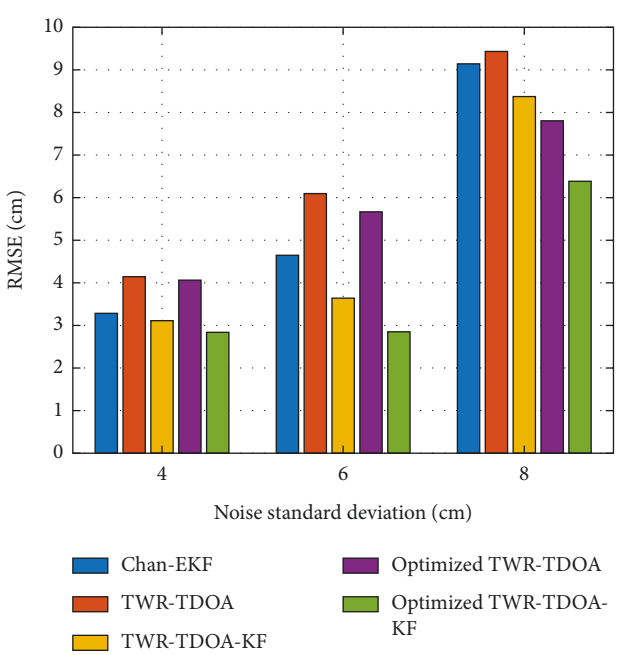


Figure 4: RMSE of localization by Chan-EKF, TWR-TDOA, TWR-TDOA-KF, Optimized TWR-TDOA, and Optimized TWR-TDOA-KF, when $\delta = 4$ cm, $\delta = 6$ cm, and $\delta = 8$ cm.

In this section, the experimenter debugs the positioning system from the midpoint between anchor 3 and anchor 4 to the midpoint of anchor 1 and anchor 2 by holding the tag at a constant velocity of 0.8 m/s. We use the Link PG system to get the mobile data and then process these data by our proposed method. The results are shown in Figure 6(b). The true trajectory of the target's motion is illustrated with the blue line at the top. The positioning result is depicted with the red dots. The result processed with the Kalman filter is plotted with a yellow line, which is the final location trajectory. It shows that the Kalman filter put the scattered anchor point trace into a smooth tracking curve, and the anchor point trace deviating from the real trajectory is corrected to a certain extent.

4.2. Indoor Scenario. The GPS signals cannot penetrate buildings and are not able to work indoors. In this part, we designed an indoor position experiment. The setup is shown in Figure 7(a), which is a room with a length of 640 cm, a width of 430 cm, and a height of 374 cm. The moving target tracks the blue rectangular on the desk, which has a height of 80 cm. Anchor 1 is the primary sensor and the others are the secondary sensors.

In the experiment, $h = 80$, Genmax = 10, and $X^{2\times 12} = \begin{bmatrix} 0 & 0 & 0 & 0 & 400 & 0 & 0 & 400 & 0 & 0 & 400 & 0 \\ 0 & 0 & 0 & 500 & 400 & 370 & 500 & 400 & 370 & 500 & 400 & 370 \end{bmatrix}$. After running the MPGA, anchor 1 is set as (0,0,0) and the other anchors are set as (500, 0,370), (500,400,0), (0,400,370), respectively. In this indoor experiment, by using the proposed algorithm, the 3D comparison between the positioning results and the real trajectory is shown in Figure 7(b). It should be mentioned that the electromagnetic interference from GPS and WiFi may affect the measurement accuracy.

5. Results and Analysis

For indoor and outdoor environments, effective measurement experiments are carried out. The RMSE between the actual position and the estimated location is evaluated. The RMSE of the outdoor scene is shown in Figure 8(a), and the error result for the indoor scene is calculated, as shown in Figure 8(b).

The average RMSE of the position error for the outdoor environment is 9.8 cm, and the average position error is about 11.21 cm for the indoor scene. The indoor error is

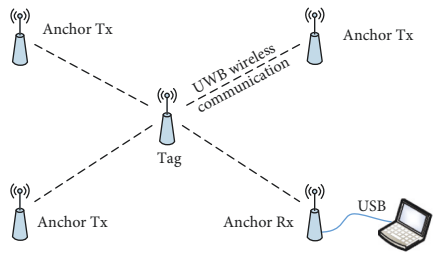
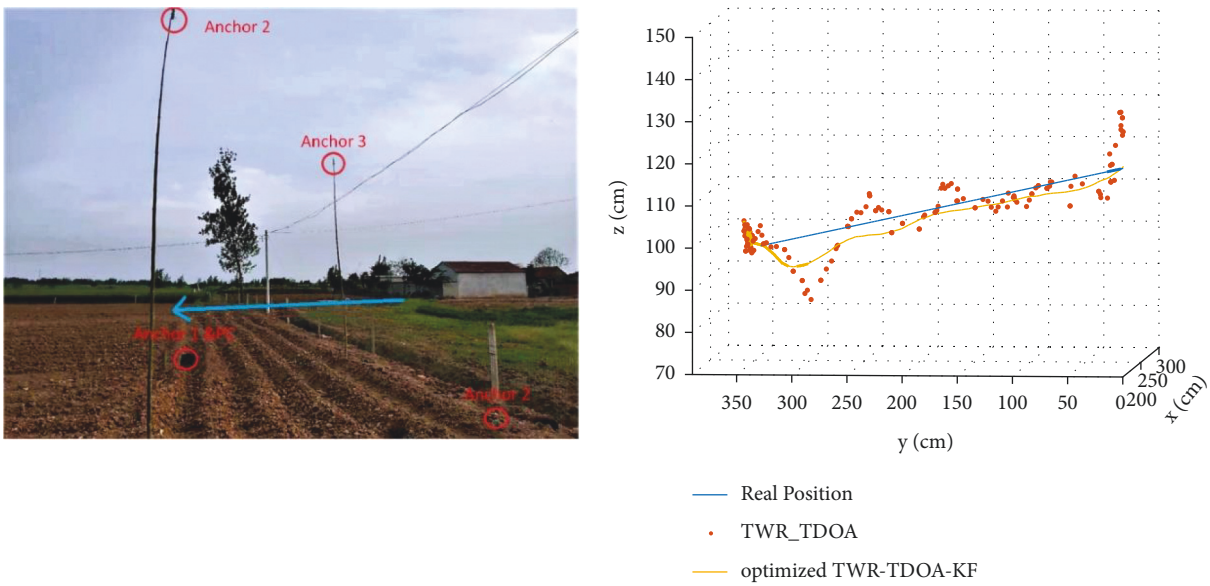
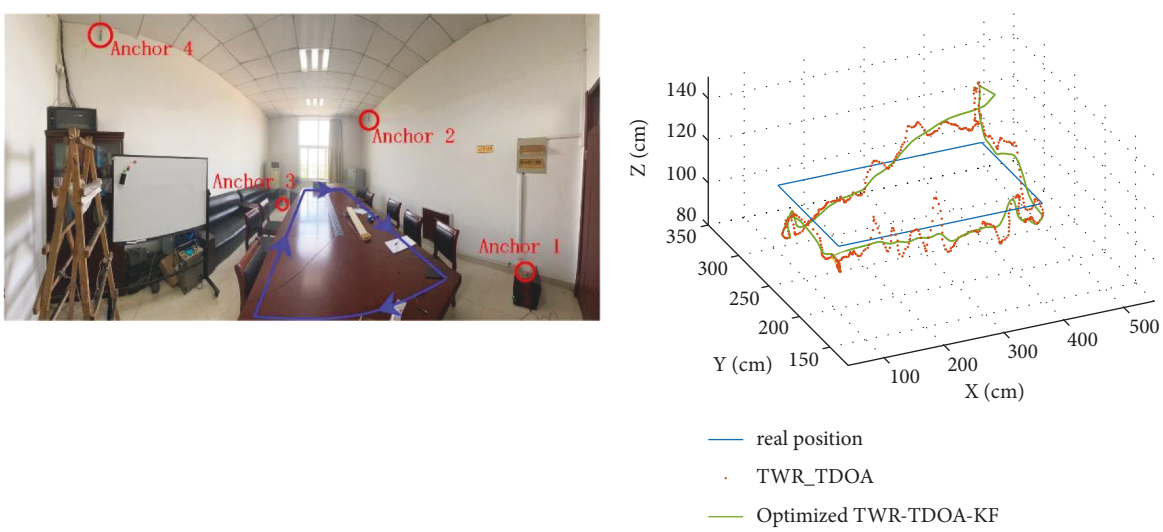


Figure 5: The system measurement model.



(a) (b)

Figure 6: The outdoor scene and test results. (a) Outdoor cropland experiment scene. (b) Comparison of location and tracking results with the real trajectory.



(a) (b)

Figure 7: Indoor testing environment and results. (a) Indoor moving target positioning scene. (b) Comparison of location and tracking results with real trajectory.

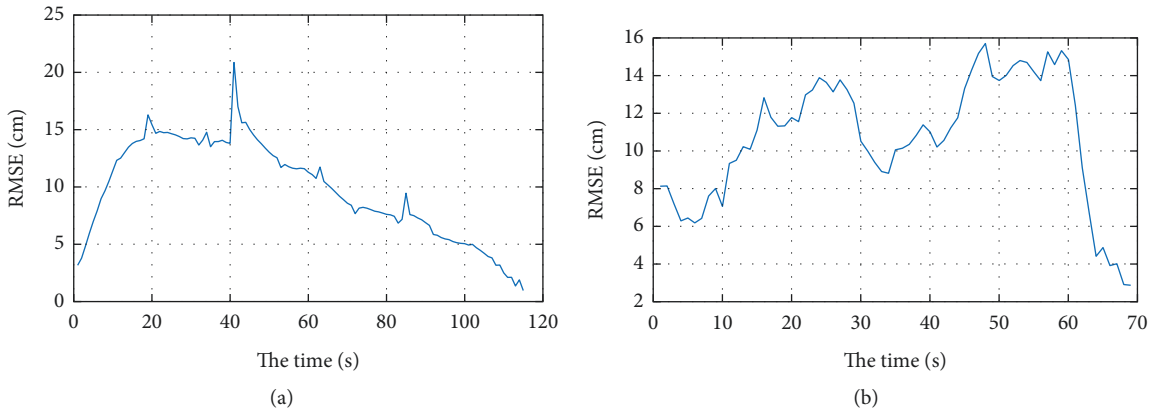


Figure 8: The RMSE curves for the outdoor and indoor systems. (a) The system positions RMSE for the outdoor scene. (b) Position measurement RMSE for the indoor scene.

Table 2: Comparisons among different UWB position systems.

System	Algorithm	Accuracy (cm)
Ubisense [22]	AOA, TDOA	15–20
Zebra [23]	TDOA	30
Sewio [24]	TDOA	30–50
Asynchronous [25]	TDOA	15.2
Sync-free [26]	TDOA	20.4
This article	TWR-TDOA-KF	10

larger than that in the outdoor environment, which may be caused by adjacent frequency electromagnetic interferences. Combining the outdoor and indoor scenarios, the whole-system RMSE error for 3D moving targets is around 10 cm.

Comparison based on localization accuracy and the implemented localization approach between the proposed technique and other reported algorithms is listed in Table 2. The technique proposed in this article has the best accuracy performance when the new TWR-TDOA solution and layout optimization are applied, and the advantage is that it can be adapted to different environments.

6. Conclusion and Discussion

The system presented in this article successfully demonstrated the target positioning and tracking process for a moving target in 3D UWB indoor and outdoor scenes. The idea of the system is to use ranging to realize positioning. After obtaining the distance between the target and the base station, the proposed TDOA algorithm is used for the positioning solution. We also discuss which factors affect the theory's accuracy. An optimum layout with low complexity based on the 3D MPGA is used to reduce the error in positioning accuracy. Single horizontal plane searching can greatly reduce computing and has similar precision to whole space searching. The hardware of the proposed system uses a UWB communication to realize the data transmission between the measured target and the base station and then realizes the distance measurement. This is an important step to getting the original data. Indoor and outdoor experiments are provided. If only the new TWR-TDOA solution is adopted, it has poor stability. Kalman filter algorithm is used to solve the problems related to fuzzy and missing subjects of TDOA positioning in the indoor and outdoor experiments, which has better location results. Our proposed system has better precision than mainstream algorithms. Due to the limited experimental conditions, the error test of the system is not very accurate. If the stepping motor is used to simulate and analyze the moving target in actual scenes, the system error can be further reduced. In this design, the Kalman filter is selected to filter the TDOA positioning results when tracking the target. In fact, this method has the problem of filter divergence, that is, when the target motion state does not meet the filter setting, the Kalman filter results will deviate and accumulate gradually. This problem can be improved by some specific algorithms for tracking maneuvering targets, such as the Singer algorithm, Interactive Multi-Model algorithm, and so on.

Data Availability

The MatLab and python code 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.

Acknowledgments

The project was supported by the National Natural Science Foundation of China (Grant nos. 61801231 and 62071238), the Natural Science Foundation of Higher Education of Jiangsu Province (Grant no. 17KJB510020).

References

[1] J. D. Ni, D. Arndt, and P. Ngo, "Ultra-wideband time-difference-of-arrival two-point-tracking system," in *Proceedings of the AIAA Houston Annual Technical Symposium 2009*, Houston, TX, USA, March 2021.

[2] J. Ni, D. Arndt, P. Ngo, C. Phan, K. Dekome, and J. Dusl, "Ultra-wideband time-difference-of-arrival high resolution 3D proximity tracking system," in *Proceedings of the IEEE/ION Position, Location and Navigation Symposium*, pp. 37–43, Indian Wells, CA, USA, May 2010.

[3] Z. Deng, X. Zheng, C. Zhang, H. Wang, and W. Liu, "A TDOA and PDR fusion method for 5G indoor localization based on virtual base stations in unknown areas," *IEEE Access*, vol. 8, pp. 225123–225133, 2020.

[4] K. C. Ho, "Unified near-field and far-field localization for AOA and hybrid AOA-TDOA positionings," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1242–1254, 2018.

[5] J. He and H. C. So, "A hybrid TDOA-fingerprinting-based localization system for LTE network," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13653–13665, 2020.

[6] Q. Li, B. Chen, and M. Yang, "Improved two-step constrained total least-squares TDOA localization algorithm based on the alternating direction method of multipliers," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13666–13673, 2020.

[7] Z. Su, G. Shao, and H. Liu, "Semidefinite programming for NLOS error mitigation in TDOA localization," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1430–1433, 2018.

[8] B. Xu, G. Sun, and Z. Yang, "High-accuracy TDOA-based localization without time synchronization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1567–1576, 2013.

[9] S. He and X. Dong, "High-accuracy localization platform using asynchronous time difference of arrival technology," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1728–1742, July 2017.

[10] J. Yin, Q. Wan, and K. C. Ho, "A simple and accurate TDOA-AOA localization method using two stations," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 144–148, Jan. 2016.

[11] R. Mazraani, M. Saez, L. Govoni, and D. Knobloch, "Experimental results of a combined TDOA/TOF technique for UWB based localization systems," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1043–1048, Paris, France, June 2017.

[12] C. Qian, W. Xia, W. Cui, Z. Lan, F. Yan, and L. Shen, "A 3-D indoor positioning system using asymmetry double-sided two-way ranging and chan assisted extended kalman filter," in *Proceedings of the 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Hangzhou, China, October 2018.

[13] Z. Cui, Y. Gao, J. Hu, and J. Cheng, "LOS/NLOS identification for indoor UWB positioning based on morlet wavelet transform and convolutional neural networks," *IEEE Communications Letters*, vol. 25, no. 3, pp. 879–882, March 2021.

[14] A. Niitsoo, T. Edelhäußer, and C. Mutschler, "Convolutional neural networks for position estimation in TDoA-based locating systems," in *Proceedings of the 2018 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–8, IPIN, Nantes, France, September 2018.

[15] C. Jiang, J. Shen, S. Chen, Y. Chen, and Y. Bo, "UWB NLOS/LOS classification using deep learning method," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2226–2230, 2020.

[16] A. Seyed, R. M. Buehrer, *Handbook of Position Location: Theory, Practice, and Advances*, Wiley-IEEE Press, Hoboken, NJ, USA, Second Edition, 2018.

[17] Y. Zhao, Z. Li, and J. Shi, "Sensor selection for TDOA-based localization in wireless sensor networks with non-line-of-sight condition," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 9935–9950, 2019.

[18] W. Li, "GDOP and the CRB for Positioning Systems," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E100, no. 2, pp. 733–737, 2017.

[19] W. Tong, B. Tao, X. Jin, and Z. Li, "Design optimization of multipole galatea trap coils by multiple population genetic algorithm," *IEEE Transactions on Plasma Science*, vol. 44, no. 6, pp. 1018–1024, June 2016.

[20] R. Olivera, R. Olivera, O. Vite, H. Gamboa, M. A. Navarrete, and C. A. Rivera, "Application of the three state Kalman filtering for moving vehicle tracking," *IEEE Latin America Transactions*, vol. 14, no. 5, pp. 2072–2076, May 2016.

[21] D. W. 1000 User Manual, "How to Use, conFig," *And Program the DW UWB Transceiver*, vol. 23, p. 231, 2017.

[22] A. R. Jiménez Ruiz and F. Seco Granja, "Comparing ubisense, BeSpoon, and DecaWave UWB location systems: indoor performance analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 2106–2117, Aug. 2017.

[23] Z. User Mannual, "Dart UWB Vision Reader," *UWD*, vol. 1000, p. 223, 2013.

[24] *Sewio RTLS UWB Kit Guide*, SEWIO, 2021.

[25] S. He and X. Dong, "High-accuracy localization platform using asynchronous time difference of arrival technology," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1728–1742, 2017, ISSN:0018-9456.

[26] W. Wang, J. Huang, S. Cai, and J. Yang, "Design and implementation of synchronization-free TDOA localization system based on UWB," *Radio Engineering*, vol. 27, no. 1, pp. 320–330, 2019.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 7049377, 12 pages
https://doi.org/10.1155/2022/7049377

Research Article

Development of Wireless Transmission System for Microseismicity in Complex Mountainous Area

Qingming Xie,[1,2] **Chunling Wu**,[1] **Hongliang Liao**,[3] **Lichuan Chen**,[2] **Yunbin Hu**,[1] **Guilan He**,[1] **and Yueming Kang**[4]

[1]*Chongqing College of Electronic Engineering, Chongqing 401331, China*
[2]*Technology Innovation Center of Geohazards Automatic Monitoring, Ministry of Natural Resources, Chongqing Institute of Geology and Mineral Resources, Chongqing 401120, China*
[3]*Beijing Urban Construction Survey, Design and Research Institute Co. Ltd., Beijing 100101, China*
[4]*China Coal Science and Engineering Group Chongqing Research Institute Co. Ltd., Chongqing 400039, China*

Correspondence should be addressed to Qingming Xie; mingming8842@126.com

Received 1 March 2022; Revised 24 April 2022; Accepted 17 May 2022; Published 31 May 2022

Academic Editor: Wang Wenyong

A large amount of data would be generated during the process of microseismic monitoring, and data transmission with the cable has some shortcomings such as poor anti-interference, time-consuming, laborious, high cost, and low construction efficiency. On the other hand, the influence of signal diffraction and the multipath effect will greatly cut down the energy of the signal during the wireless transmission in the complex mountainous area. The wireless transmission system based on Wi-Fi is designed in this paper, which includes the base station and data transceiver. The maximum distance and transmission rate at point-to-point communication of the base station are up to 1000 m and 150 Mbps, respectively, and it provides the communication backbone line for the network. The data transceiver adopts the network protocol based on Ad Hoc, and the transmission distance is up to 200 m at the condition of complex topography and lush mountains. We have applied the system to in-site microseismic monitoring with 28 geophones of the coalbed methane fracturing in South Chongqing. It achieved a good performance with a transmission rate of 1.2 Mbps, a time delay of 1.95 ms, and a signal strength of up to −52.3 dBm for real-time data transmission in the field. The results show that the system has the advantages of low BER, fast transmission speed, long communication distance, and stable and safety hardware and has a great value for more applications of microseismicity in complex topography.

1. Introduction

Microseismicity is the earthquake with a small magnitude induced by the rock failure originated in the change of stress field of underground rock [1]. The events directly are related to the mechanism of rock fracturing at the action of internal and external driving forces. Source parameters, geometric size, and the extension direction of the crack can be predicted on temporal and spatial scale. It has been widely used in unconventional oil and gas fracture monitoring and evaluation [2–4], earthquakes induced by hydraulic fracturing [5–7], mining safety monitoring [8–10], and early warning and prediction of landslides [11–13].

With the improvement of the sampling rate, a large amount of data would be generated during the process of microseismic monitoring. The transmission with cable has some shortcomings such as poor anti-interference, time-consuming, laborious, high cost, and low construction efficiency. As the development of wireless local area network (WLAN) transmission technology, such as Wi-Fi, ZigBee, LoRa, and Bluetooth, it has brought the progress of microseismic data transmission technology in real-time.

Remote, wireless network of seismic sensor stations was achieved for microseismic monitoring. The seismic information is received by a control center station and displayed and analyzed [14]. Zigbee wireless network communication technique is used for wireless data collection of seismic wave detection sensor. Meanwhile, GPRS wireless packet switching technique is used to complete remote data transmission [15]. Remote wireless module, Ad Hoc

technology, and AODV (Ad hoc on-demand distance vector) wireless routing protocol are used to complete multihop data forwarding in the case of no other infrastructure [16]. WSNs (wireless sensor networks) based on compression perception is applied to the system design of the source location node, which includes acquisition, storage, and wireless transmission for microseismicity [17]. Weak signal acquisition, wireless communication, and database management are implemented by the optimal layout of a borehole-surface monitoring system [18]. A routine high-resolution microseismic monitoring system was installed in an opencast coal mine to investigate the impact of induced seismicity on the slope failures in real time [19]. The methodology related to data acquisition, analysis, and interpretation from microseismic monitoring was used to determine possible fault locations [20]. Combining a load-balancing scheme with a high-throughput polling mechanism in WLANs, the system allows all the seismographs to associate with the available APs and keeps load balance among the APs [21].

A set of high-precision distributed wireless microseismic acquisition stations was developed by Sun et al. [22], including the acquisition circuit, main control circuit, and other hardware circuits. Advanced acquisition systems are integrated into the ARMs (advanced RISC machines) of the main control board, and Wi-Fi technology was used to achieve wireless data communication [23]. Hardware was developed to address the wireless microseismic acquisition stations and deliver monitoring software based on the Android platform [24]. Microseismic sensors are integrated into the toe of the SMART geotechnical instruments with the MineHop mesh network from Mine Design Technologies (MDTs), which will allow mines to drill one hole to satisfy the requirements of microseismic and traditional geotechnical monitoring [25]. Seismic signals processing system (SSPS) is an embedded computer system that receives real-time waveform data from Sensor Interface & Signals Acquisition (SISA). The SSPS is processed real-time Short-Time-Average through Long-Time-Average (STA/LTA) [26]. A new wireless seismic sensor network system based on Wi-Fi and existing network resources, especially designed for seismic monitoring of buildings, allows remote control, and real-time monitoring of the recorded signals by any Internet browser [27]. A wireless seismic exploration system using a dual-layer network based on Wi-Fi and LTE is developed for long-distance high-rate seismic data transmission with a high reliability [28]. Internet of Things-based wireless technology is developed to a considerable amount of data created by complex seismic scenarios, with the advantages of long range, low power, and inherent compatibility to cloud storage and computing [29, 30].

Due to the complex topography and lush mountains, there are many difficulties in wireless transmission equipment layout, short transmission distance, fast energy scattering, and obvious multipath effect in Southwest China. In this paper, a wireless transmission system based on Wi-Fi mode is developed for the complex mountainous area, which has the advantages of flexible networking, fast transmission speed, long communication distance, and stable and safety hardware. We have applied the system to in-site microseismic monitoring of the coalbed methane fracturing with 28 geophones in South Chongqing, it achieved a good performance for real-time transmission.

2. Wireless Transmission System Design for Microseismicity

2.1. Microseismic Monitoring Network. The sensors for microseismicity are usually arranged in four ways [31]. Surface monitoring is to collect signals by laying geophones on the ground above the monitoring target area (Figure 1(a)). More than 1000 single-component (1C) geophones are usually inserted into the surface. In recent years, the layout of distributed stations is more and more widely used for microseismicity because of its convenient construction and low cost (Figure 1(b)). The three-component (3C) geophones are buried in several meters underground in order to reduce the interference of environmental noise. The advantages of both are the large observation aperture, accurate location on horizontal positioning, and the conditions of focal mechanism inversion. But it is difficult to distinguish from the ground noise. Microseismic monitoring in a shallow hole (Figure 1(c)) usually adopts the combination observation with multiple shallow holes underground 50 to 300 m. The 3C geophones are placed in each hole to detect the signal. This way can effectively avoid intense environmental noise and improve signal quality. However, it brings out the expensive cost because of multiple boreholes drilling. Microseismic monitoring in borehole (Figure 1(d)) is usually arranged the receivers in one well (commonly deep in several thousand meters) near the monitoring target area. The advantages are the good SNR and accurate location, and it is often used to fracturing monitor for unconventional oil and gas.

2.2. Wireless Transmission System in Complex Mountainous Area. The system adopts WLAN communication technology with the wireless topology of base stations and data transceivers based on Wi-Fi mode, as shown in Figure 2(a). The backbone network is composed of base stations, and the communication access point (AP) is consisted of transceivers. The system adopts an open "plug and play" networking mode, which is conducive to the addition of data acquisition nodes.

Single chain is composed of a sensor, data collector, and wireless transceiver, which are connected with each other through cables. The microseismic signal is collected by the sensors, and the analog signal is converted into the digital signal by the ADC in the data collector and converted it into the wireless signal by the transceiver to meet the network communication protocol. The wireless signal from microseismicity is remotely transmitted to the microseismic processing center to analysis and storage through different base stations, as shown in Figure 2(b).

2.3. Base Station. The wireless base station is the main networking equipment of the transmission system for microseismic data. It provides the communication trunk line

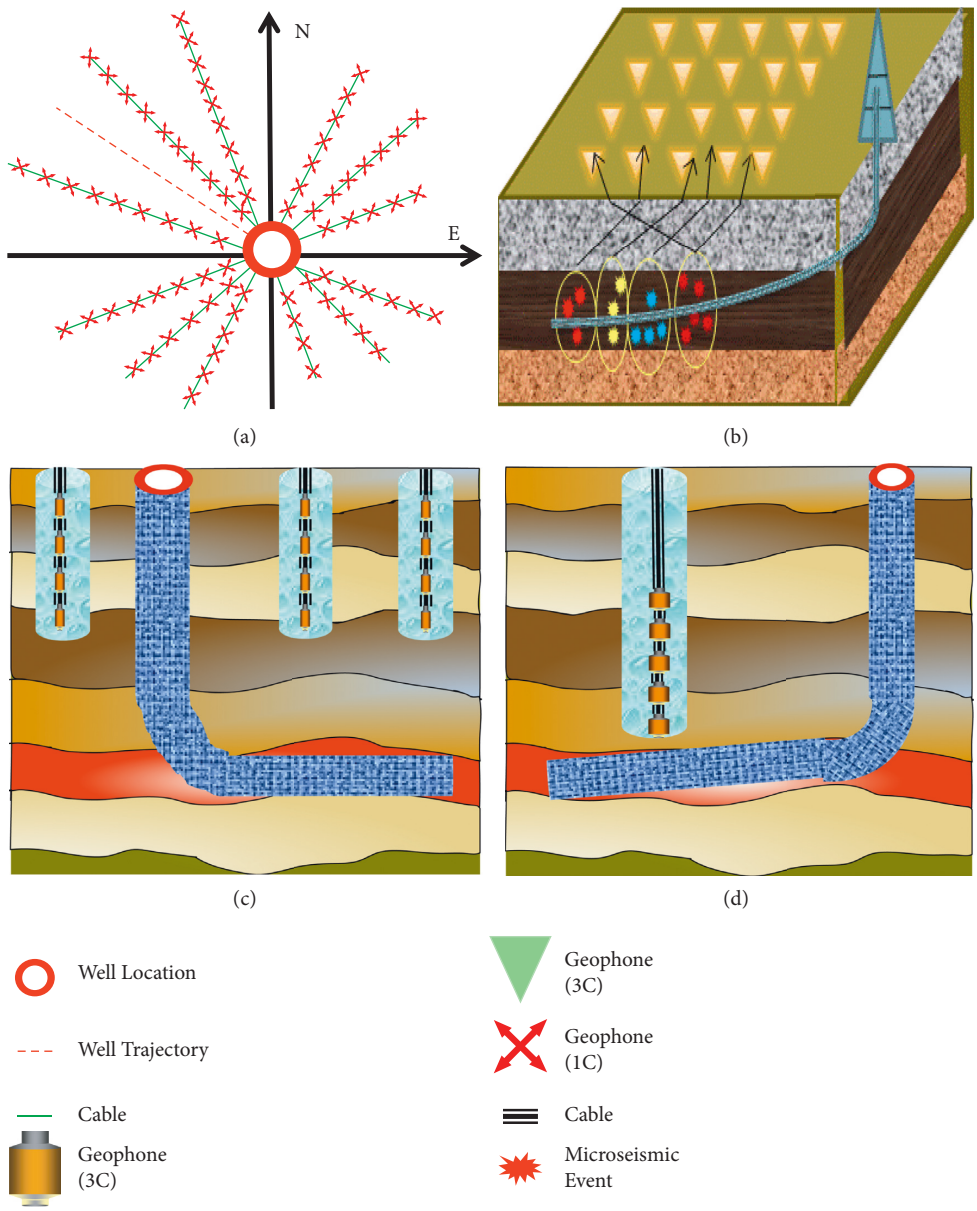


Figure 1: Microseismic monitoring network. (a) Surface array with 1C sensors; (b) distributed station network with 3C sensors; (c) 3C geophones deployed at the various shallow holes; (d) 3C geophones placed in deep boreholes near the fracturing well.

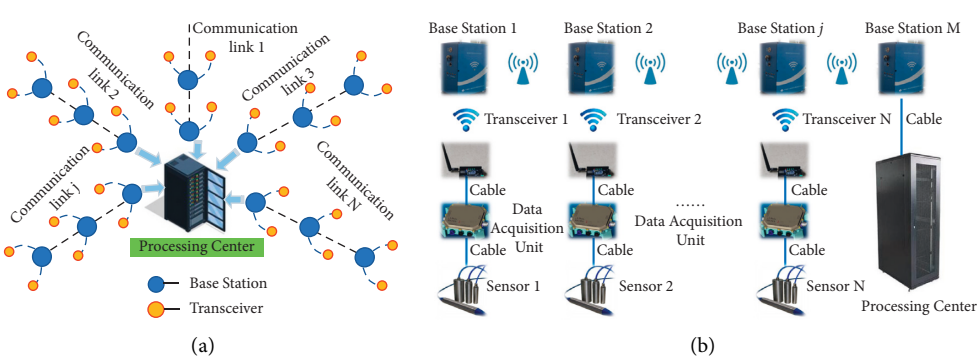


Figure 2: Multilink wireless transmission system architecture. (a) Microseismic data transmission system of multilink cascade; (b) connection diagram of a single link is composed of sensor, data acquisition unit, transceiver, and base station.

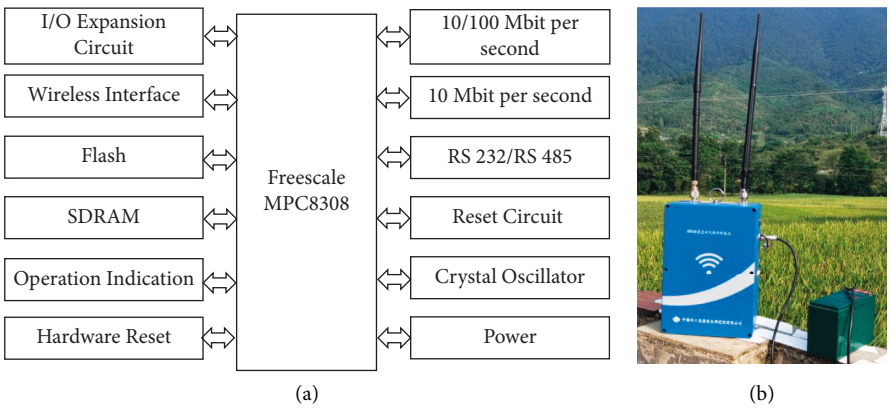


Figure 3: Internal structure diagram of base station (a) and its practicality picture (b).

Table 1: Main parameters of the base station.

Communication protocol	802.11 b/g/n	Serial port mode	RS232/RS485
Frequency range	2.32 GHz to 2.51 GHz	Maximum transmission rate	150 Mbps
Transmission distance	700m @6dBi antenna, flat and no shelter	Maximum access point	255
Encryption type	WEP64/WEP128/TKIP/AES	Security mechanism	WEP/WPA-PSK/WPA2-PSK
Working current	≤600 mA	Working voltage	24.0 V

for the whole network and cascade function in the single trunk link. When the transceiver enters the signal coverage range of the base station, the base station automatically searches for the access node in the same network and allows the security node to automatically join the network after communication handshake authentication.

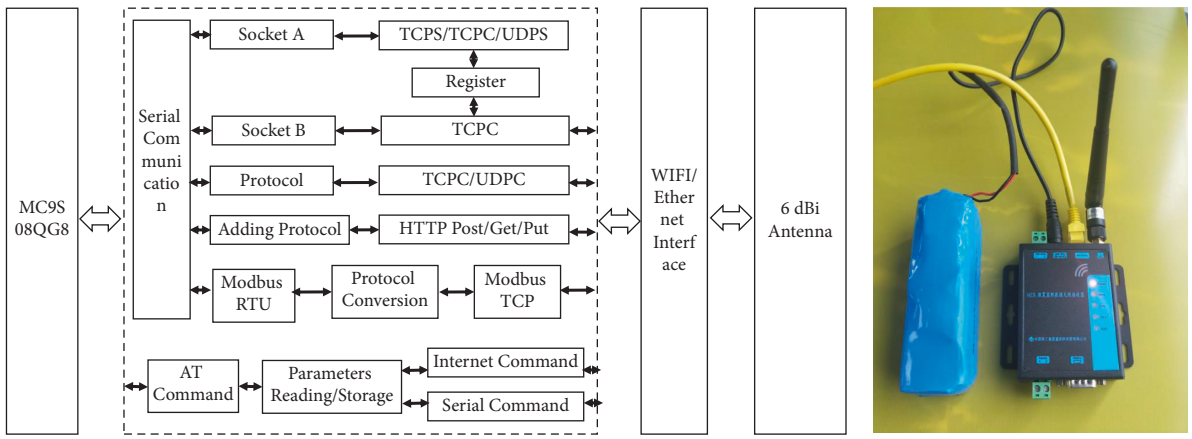
The base station adopts an embedded system composed of processor, memory, and peripheral chip. The main control chip is Freescale MPC8308 with low power consumption and high integration. The PowerPC e300 core in the MPC8308 is a superscalar processor with the 400 MHz maximum operating frequency and includes the independent on-chip 32K bytes physically addressed cache, on-chip L1 instructions, and memory management units (MMU). The memory includes a 128 MByte unbuffered DDR2 SDRAM discrete devices, 8 MByte NOR flash single-chip memory, 32 Mbit NAND flash memory, and 256 kbit M24256 serial EEPROM [32]. The chip supports dual three-speed (10, 100, 1000 Mbps) ethernet controllers and is mainly applied in wireless base stations, data concentrators, and wireless LAN access points. In this paper, the crystal oscillator with 150 MHz is connected to the processor. The structure diagram of the base station is shown in Figure 3.

In order to meet the transmission requirements of complex mountainous areas, the parameters of the base stations are shown in Table 1. The operating frequency is in the range of Wi-Fi (~2.4 GHz). The maximum distance and transmission rate at point-to-point communication are 1000 m and 150 Mbps, respectively. With the increase of the number of access points (APs) in the network, the transmission rate would be reduced seriously induced by channel congestion due to the limitation of bandwidth. For management consideration, the number of APs should be less than 255 to ensure the real-time transmission of raw data. The maximum power consumption is less than 1.23 W.

The routing protocol is Ad-Hoc on-demand distance vector (AODV) to realize multibroadcast routing. Combination of authentication algorithm and encryption algorithm, the security strategy is worked in WPA-PSK or WPA2-PSK protocol mode. The authentication algorithm adopts an authentication routing protocol based on IEEE 802.11x standard, and the encryption algorithm adopts AES (Advanced Encryption Standard) with 128 bits key and preshared key authentication mode.

2.4. Wireless Data Transceiver. The data transceiver supports three types of network topologies, namely, star, mesh, and cluster tree. In view of the complex environment in the mountainous area, the data transceiver adopts the network protocol based on Ad Hoc, which can minimize the power consumption and cost on the basis of general network layer functions and has the functions of self-organization and self-maintenance. In order to extend the communication distance of the base station as far as possible, any data transceiver will dynamically and automatically join the network as long as there is a network with corresponding ID.

The wireless data transceiver uses MC9S08QG8 as the main control chip, which is connected with the wireless communication module and a 6 dBi antenna to realize the transmission of microseismic monitoring data from the transceiver to the base station. MC9S08QG8 is an 8-bit microcontroller with low power consumption and high performance. It adopts enhanced kernel hcs08, has 8kbit flash and 512 bits ram, 8-bit ADC, and 15 MHz crystal oscillator [33]. The wireless communication module realizes 2.4 GHz Wi-Fi data sending and receiving. The data transmission layer includes two sockets and supports TCP server, TCP client, UDP server, HTTP, and other communication protocols. In addition, new communication



(a) (b)

Figure 4: Internal structure diagram of wireless transceiver (a) and its practicality picture (b).

Table 2: Main parameters of the transceiver.

Communication protocol	802.11 b/g/n	Serial port mode	RS232/RS485
Frequency range	2.412 GHz to 2.484 GHz	Baud rate	300 to 460.8 kbps
Transmit power	802.11 b: +19 dBm @11 M bps	Network protocol	TCP, UPD, ARP, DHCPC, DNS, PING
Receiving sensitivity	−89 dBm @11M bps	Working temperatur	−40 to 85°C
Transmission distance	300 m @6 dBi antenna, Flat and no shelter	Security mechanism	WEP/WPA-PSK/WPA2-PSK
Encryption type	WEP64/WEP128/TKIP/AES	Working voltage	5.0 V to 36.0 V

protocols can be added according to user needs. The transceiver also supports Modbus communication protocol, including RTU, ASCII, and TCP. The internal structure diagram of the wireless data transceiver is shown in Figure 4.

The digital signal from the microseismic data collector is transmitted to the wireless communication module through RS 232 or RS 485 serial port and converted it to Modbus TCP communication mode with protocol conversion. It would be sent to the corresponding base station by Wi-Fi interface and antenna and then transmitted to the microseismic data processing center.

The parameters of the data transceiver are shown in Table 2. The baud rate is 300 to 460.8 kbps, and the receiving sensitivity is −89 dBm operating at 11 Mbps. When it is connected with the antenna (transmission gain 6 dBi), the transmission distance is up to 300 m at the condition of the flat and no shelter environment.

3. System Performance Test

3.1. BER (Bit Error Rate) and Time Delay. The digital signal would be distorted by the influence of environmental noise in the transmission process, or the signal voltage is changed due to energy decay. In addition, the signal would be damaged when the hardware works abnormally. Bit error rate (BER) is an index to measure the accuracy of data transmission within a specified time. In this paper, BER is defined as the ratio of error bits to the total number of transmitted bits.

In order to test the stability of the system, the BER of the base station is tested in the field. The base station has been continuously sent the messages to the receiving port for eighteen hours, and the transmission bits is about 2.2×10^8. The error bits are 544, and the BER is 2.47×10^{-6}. The test result shows that the wireless system has the performance of high transmission efficiency and good stability. It is satisfied with the requirements of wireless transmission of microseismic monitoring data. The BER testing results are shown in Table 3.

Table 3: The BER test results.

No.	Sending (bits)	Receiving (bits)	BIT
1	220343392	220342848	$2.47e-6$
2	27200	27200	0
3	26651892	26651883	$3.38e-7$

The wireless transmission system has to meet the requirements of real-time performance and reliability for microseismic data transmission. The system is a hybrid communication mode of multilevel networking, and the tests are carried out under the shelter of trees in the field to check out the time delay. Six base stations with the interval of 400 m are used in the test to form 5 hops (HOP) backbone link with the distance of 2.1 km. The results show that all the time delays of base stations in three HOPs are less than 1 ms, and the delay of the fourth hop increases to 1.85 ms. However, when the number of Hops increases to six, the communication rate decreases significantly and the time delay increases linearly, and the maximum time delay reaches 15 ms, as shown in Figure 5. Therefore, in the complex mountainous environment, the number of base station HOPs is designed to four for the reason of convenience, economy, and effectiveness.

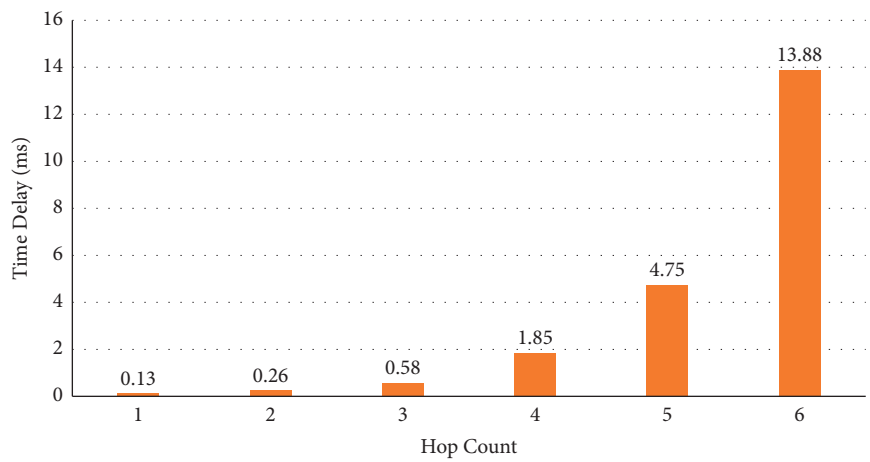


Figure 5: Time delay test for the multilevel cascade of the base station.

Table 4: Signal strength test in LOS and complex mountain environments.

Transmitting power (mW)	Antenna gain (dBi)	Surface topography	Communication distance (m)	Signal strength (dBm)
100	3	Line of sight	10	−32.4
			200	−72.5
			300	−84.5
500		Line of sight	10	−26.3
			200	−65.7
			300	−77.5
1000		Line of sight	10	−22.3
			200	−60.5
			300	−71.2
100	6	Line of sight	10	−31.2
			200	−70.3
			300	−83.7
500		Line of sight	10	−25.9
			200	−65.5
			300	−77.2
1000		Line of sight	10	−21.9
			200	−60.8
			300	−71.1
100	6	Complex mountain region	10	−31.6
			200	−80.2
			300	−94.6
500		Complex mountain region	10	−26.8
			200	−76.8
			300	−86.2
1000		Complex mountain region	10	−22.7
			200	−70.3
			300	−80.2

3.2. Signal Strength. Due to the complex terrain and lush forests in complex mountainous areas, the influence of signal diffraction and multipath effect will greatly cut down the energy of the signal. The transmission frequency is set in the Wi-Fi working frequency range (2.4 GHz), and various communication parameters (transmission power, reception distance, and antenna gain) are changed to provide the optimal communication parameters for the application of an in-site microseismic monitoring system in a complex mountainous environment. The mobile portable wireless monitoring receiver PR100 (R&S company, Germany) is used for testing of signal power strength. It has high sensitivity and scanning speed up to 2.0 GHz/s. The frequency range is 9 kHz to 7.5 GHz. In the scanning process, it can store the intercepted signal at the action of 10 MHz real-time bandwidth, which is suitable for radio reconnaissance, locating interference sources, frequency monitoring, etc.

The test results are shown in Table 4 and Figure 6. Despite the enhancement from the antenna gain, the signal would be greatly attenuated by the shelter of trees and the topographical relief in a complex mountainous area. On the other hand, the signal strength would be improved with

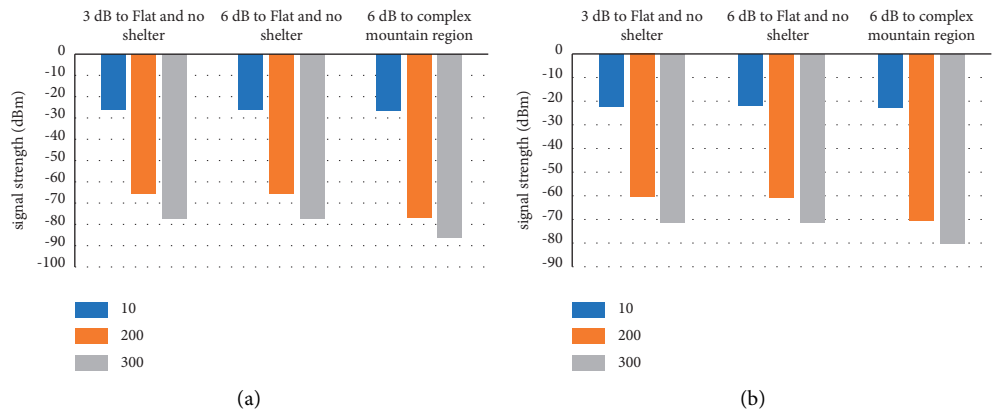


Figure 6: Test for the signal strength with various transmission parameters of the data transceiver, the transmitting power is 0.5 watt (a) and 1 watt (b), respectively.

the transmission power of the source, which is directly related to the effectiveness of data transmission. With the communication distance of 10 m and 200 m, the maximum signal strength is, respectively, −22.7 dBm and −70.3 dBm at the condition of transmission power 1 watt, 6 dBi gain antenna. At the distance of 300 m away from the source, the maximum signal strength is −80.2 dBm, which basically reaches the limit of the signal receiver. Considering the performance and economy of the data transceiver comprehensively, the main communication parameters are adopted 6 dbi gain antenna, transmission power 1 watt, and the transmission distance to the base station is less than 200 m.

3.3. Optimal Throughput Rate of Base Station. The maximum transmission rate of the base station in the backbone link is tested by the method of optimal throughput rate (OTR). The relationship between the number of data transceivers and the OTR of the base station at different distances are achieved at the condition of the complex mountain environment in the field. In general, the OTR test should be operated in a shielded darkroom to reduce the impact from environmental noise. In the process of the field test, the result of the OTR test is affected by the different terrain conditions and noise sources. In order to avoid the impact of multirate transmission on the OTR test of the base stations, the testing environment should not be changed greatly with the increase of the number of base station cascades.

The tests are carried in the field with the shelter from the trees. According to the previous test results, the number of data transceivers is determined to be four, and the distances of the base station are gradually changed as 100 m, 200 m, 400 m, 600 m, and 800 m, respectively. The results are shown in Figure 7. The OTR of the base station increases with the acquisition nodes (data transceivers) when the number of data transceivers in the single link is less than four. The OTR achieves the highest transmission speed when the number of acquisition nodes is four. However, when the number of data transceiver nodes is greater than 4, the OTR begins to decrease with the increase of the acquisition node. The transmission rate of the base station only reaches 1.5 MHz at the condition of the distance of 800 meters and six HOPs. It cannot meet the data transmission requirements when the number is more than 30 microseismic sensors. As a result, in the complex mountainous environment, the number of base station HOPs is designed to four, and the number of the data transceiver is four too.

4. Case Study

4.1. Introduction of the Coalbed Methane (CBM) Well. In order to explore the resources and productivity of coalbed methane in southern Chongqing, the wireless transmission system is applied to microseismic monitoring of two coalbed methane horizontal wells fracturing (Q-H1 and Q-H2) in real-time. The spatial distribution of cracks and the stimulated reservoir volume (SRV) would be predicted to improve the productivity of CBM.

The monitoring area is located in the south of Chongqing, belonging to the passive edge fold thrust belt and Jinfoshan dome fold belt at the upper Yangtze block. It mainly develops multiple fold structures with NE-NNE strike, which is the typical complex geological structure of mountainous areas in Southwest China. The average gas content of the coal seam is 26.14 to 28.18 m^3/t in the monitoring area, and the layer named M8 is highest up to 29.45 m^3/t. The chemical components of the gas are mainly methane, nitrogen, carbon dioxide, and a small amount of heavy hydrocarbon gas. The methane concentration is 89.69~99.36% (the average of 95.74%), which indicates a good industrial value. The wells in the monitoring area are divided into two stages of fracturing. Quartz sand is mainly used as a fracturing proppant, and the fracturing fluid consumption is 1920 m^3 (Q-H1) and 1880 m^3 (Q-H2), respectively. The fracturing belongs to a medium-sized scale compared with the stimulation of coalbed methane reservoirs in China.

4.2. Deployment of the Microseismic Wireless Transmission System

4.2.1. Deployment of the Monitoring Network. According to the topography, distributed microseismic monitoring

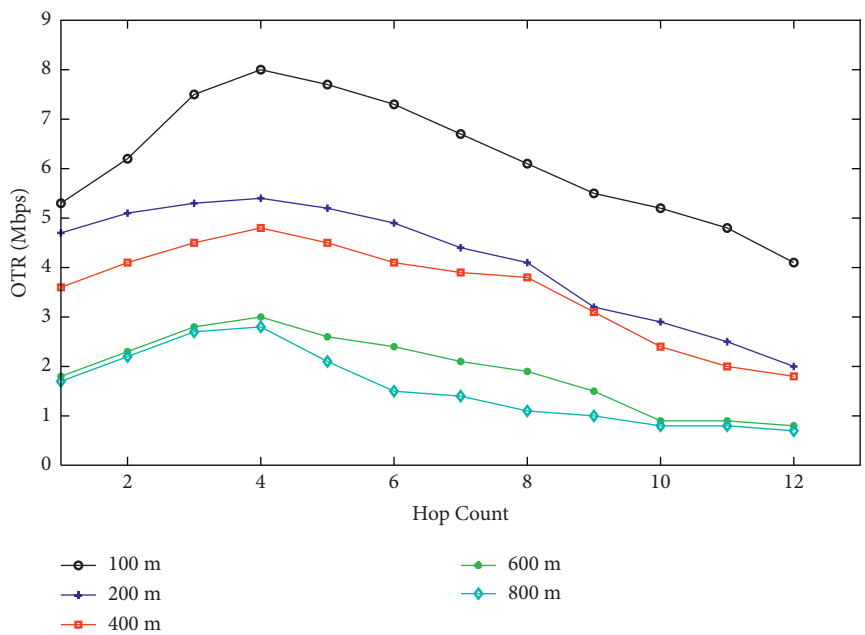
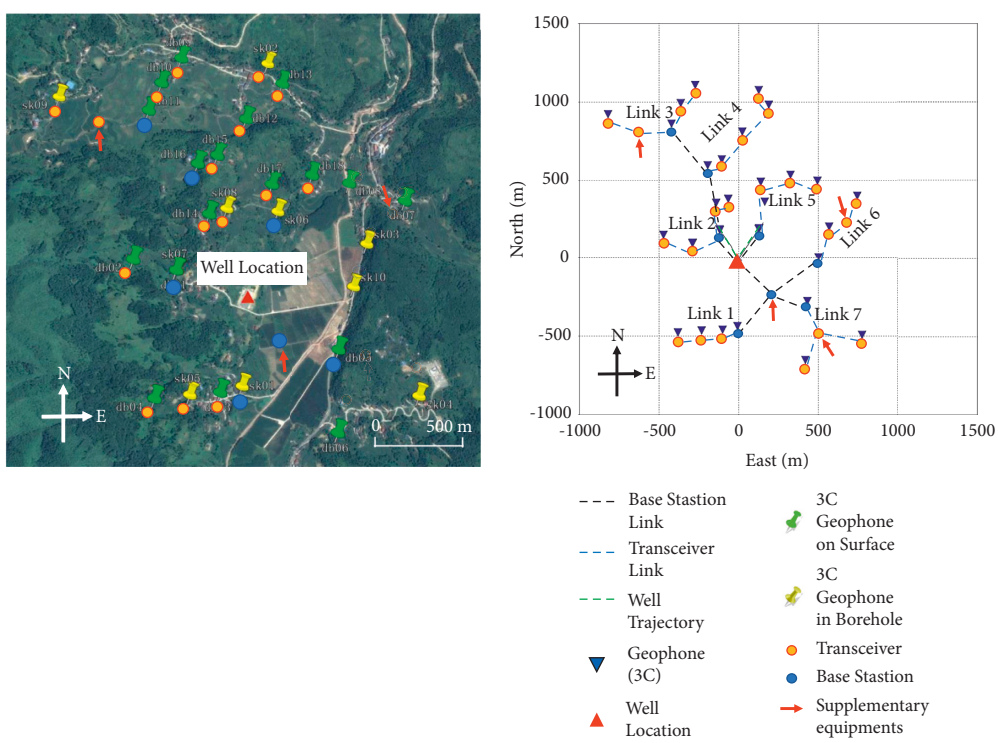


Figure 7: Optimal throughput rate (OTR) test of the base station.



(a) (b)

Figure 8: Deployment of geophones and wireless transmission networks. (a) Microseismic monitoring network; (b) the wireless networks for microseismic data transmission.

system is carried out by a 3C geophone, deployed at the surface, and shallow hole around the monitoring area (Figure 8). Firstly, the geophone in the shallow hole is less affected by environmental noise, and the SNR of raw data is better than the surface. Secondly, the weak signal received from the surface geophone can be calibrated by the high SNR signal detected in the borehole, improving the correct recognition of microseismic events. Consequently, the combination network has the advantages of flexible layout, wide azimuth range and reliable data quality [34].

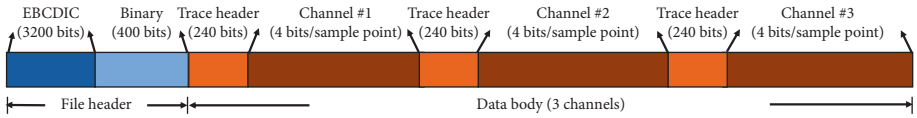


Figure 9: Data body of SEGY format proposed by SEG.

The acquisition system is composed of seismograph, geophones, solar panel, battery, router, and data transceiver, which forms a single transmission link. Nineteen 3C accelerometers (VAS-200) are placed out of the fracturing well with an interval of about 160 m, which is represented by a green pin shape in Figure 8(a). The depth is 0.5 to 1 m, covered by the protection box to decrease the environmental noise. The dynamic frequency range of the sensor is 3 Hz to 1200 Hz, and the sensitivity is better than 220 mV/g. Nine 3C sensors are deployed at the shallow hole with an interval of about 200 m, which is represented by a yellow pin shape in Figure 8(a). The performance is the same as the accelerometers deployed at the surface. The sensor's depth is 15 to 30 m, with cement slurry above to ensure it coupling with the bedrock. The data collector is Sigma 3 plus with 3 channels, high-precision 32-bit AD converter, 12 V voltage power supply, 256 G memory, and the sampling period is from 0.25 to 8 milliseconds triggered by GPS. It supports the transmission model of 4G and Wi-Fi.

4.2.2. Wireless Transmission Network. The microseismic wireless transmission system adopts the hybrid networking mode of star cascade, and a total of 8 base stations and 24 wireless transceivers are arranged to form 7 backbone links, as shown in Figure 8(b). The analog signals collected by the microseismic sensor are converted into digital signals by the data collector and transmitted to the base station through the wireless data transceiver. Then, the signal would be transmitted to the data processing center with the multiple cascaded base stations. According to the testing results, in the complex mountainous environment, the distance from the data transceiver to the base station should be less than 200 m, and the distance of the base station cascade should be less than 400 m. In order to improve the data transmission rate and reduce the time delay, two transceivers are supplemented in link 3 and link 6, respectively, and one transceiver and base station are added in link 7, as indicated by the red arrow. The main communication parameters of the transceiver are adopted 6 dBi gain antenna, the transmission power of 1 watt, and the transmission distance to the base station is less than 200 m. The number of base stations HOPs in a single link are less than 5 data transceivers. The maximum interval between base stations is 400 m, and each trunk line can cascade up to 4 base stations.

The data from the geophone is encapsulated in the standard SEGY format proposed by SEG (Society of exploration geophysicists), which is composed of a file header and data body (Figure 9). The file header includes an EBCDIC header (3200 bits) and binary header (400 bits), which are used to store microseismic description and sampling rate, equipment status, measurement parameters, etc. The data body is composed of trace header (240 bits) and sampling data, the latter is generally floating-point format (4 bits) in IEEE or IBM standard. In the case of the 1000 sps of the sampling rate, each three-component sensor will generate a SEGY file of about 16 kbits packet per second, and 448 kbits for 28 sensors.

On the condition of the sampling rate of 1000 SPS, the data volume generated by 28 sensors is 448 kbps. Besides the routing message, the data transmission throughput of the system in one second needs to reach at least 1 Mbps. Finally, the system has played a good performance with a transmission rate of 1.2 Mbps, a time delay of 1.95 ms, and a signal strength is −52.3 dBm.

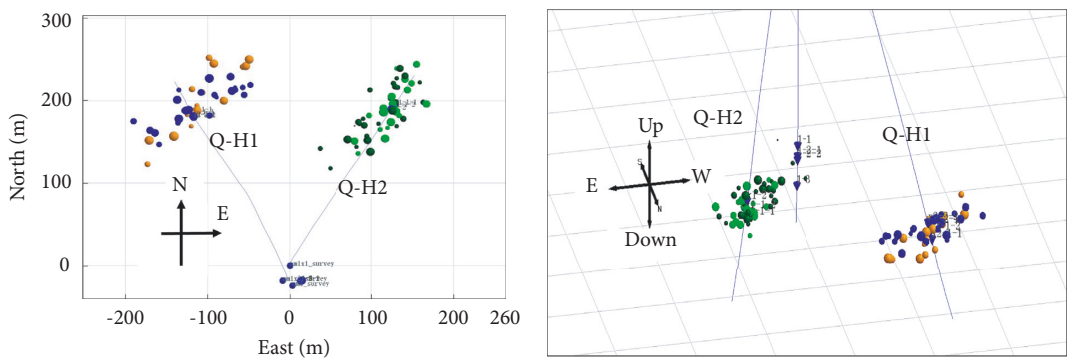
4.3. Performance Comparison with Various Systems. Savazzi and Spagnolini [35] discussed the feasibility of employing Wi-Fi, Wi-Max, Bluetooth, ZigBee, and Ultra-Wideband (UWB) technologies in wireless geophone network. General cable-free land seismic data acquisition and current state-of-the-art wireless seismic data acquisition systems are overviewed by Makama et al. [36] and Yi et al. [37]. Compared with other transmission modes, Wi-Fi has the advantages of wide bandwidth, networking flexibility, reliability, and security in complex application environments. At present, based on IEEE 802.11 protocol, there are several typical wireless systems used in microseismic monitoring, such as UNITE, RT3, and Sigma™.

UNITE system is developed by "Sercel Inc." (Carquefou, Pays de la Loire, France) and consists of a central control station (CCS), remote acquisition units (RAU), and cell access nodes (CAN) [38]. The antennas for Wi-Fi and GPS are set on the top of RAU to improve the data quality through real-time GPS time synchronization. In order to avoid transmission interference, RAU and CAN are operated in the range of 2.4 GHz and 5.8 GHz, respectively. The actual short-distance transmission rate of RAU, based on 802.11b, is normally less than 1 Mbps in outdoor line-of-sight (LOS) environments [28].

RT3 (RealTIME 3) is from the "Wireless Seismic Inc." (Stafford, TX, USA), designed with two-tier radio telemetry architecture and supports over 250,000 cable-less channels [39]. It consists of recording units named Mote, ground Relay Units (GRU) operated in the 2.4 GHz (ISM band), and a central recording system (CRS) worked in the range of 5.6 GHz to 5.8 GHz. The data from the Mote to CRS are transported via GRU communication in real time. The GRU is a full duplex transceiver supported in a line segment, the throughput rate is up to 55 Mbps with the "burst" type. CRS provides three independent views of the spread, including continuous seismic energy and ambient noise levels.

Table 5: Performance comparison with the advanced systems.

Device	Protocol	Technology employed	Dynamic range (dB)	Power consumption	Communication range (m)	Data rate	Manufacturer
RT3	802.11 b/g/n	2.4 GHz ISM band	143	—	<400	<55 Mbps (LOS)	Wireless Seismic, USA
UNITE	802.11a/b/g/n	2.405~2.4835 GHz	128	0.085 w/channel	<1500 (LOS)	<11 Mbps (LOS)	Sercel, France
Sigma TM	802.11 b/g/n	2.4 GHz ISM band	126	0.48 w/channel	<400	—	International Seismic, USA
In the paper	802.11 b/g/n	2.4 GHz	128	0.41w/channel	<400	<11 Mbps	—



(a) (b)

Figure 10: Microseismic events in inclined well Q-H1 and Q-H2. (a) Vertical view of the events; (b) lateral view of the events. The colors of the sphere represent different fracturing sections, and the size of the sphere represents the local magnitude (M_L) of the events.

Sigma™ is a wireless continuous seismic data acquisition system developed by "International seismic Inc." (Ponca City, Oklahoma, USA). It offers multiple data acquisition and retrieval modes and can be extended to hundreds of channels, including blind data acquisition, control & status nodes, and real-time data transmission within the range of the 2.4 GHz ISM band [40]. When facing different observation conditions and objects, observation strategies can be formed with the corresponding software and observation strategy. Flexible and self-configurable features of Mesh Radio Network (MRN) are allowing Sigma Acquisition Units (SAU) to be deployed in aggressive environments. SAU communicates to maintain the MRN network naturally, providing a redundant wireless connectivity point to reach all units.

The performance of the microseismic wireless transmission system in this paper is compared with the advanced system based on Wi-Fi, as shown in Table 5. For the complex mountainous environment, the communication range and data rate would be significantly attenuated. Due to the different test environments, it is difficult to compare the performance. However, generally, RT3 has the widely dynamic range and the ability of channel expansion (over 250,000 channels) with the flexible layout. UNITE system can be connected with a digital or analog sensor, it is suitable for large-area monitoring, and the power consumption of a single channel is small. Although the data rate of the system in this paper is only up to 11 Mbps at LOS conditions, it can realize the stable transmission of 1.2 Mbps after multihop connection in the complex mountainous environment in Southwest China, which is satisfied with the stable transmission of less than 40 access points. In addition, the maximum communication range can be up to 400 m, and the UNITE and RT3 systems are less than 300 m in the harsh environments.

4.4. The Effect Evaluation of Well Fracturing. With the processes of noise reduction, signal recognition, first break pickup, velocity modeling, and location inversion, 88 effective microseismic events were identified in the fracturing of the two coalbed gas wells (47 events in Q-H1 and 41 events in Q-H2). The local magnitude (M_L) was between −2.13 and 1.24; it indicated that the rock fracturing belongs to a weak event. According to the distribution of the events, the stimulated reservoir volume (SRV) are $65.5 \times 10^4\ m^3$ and $59.2 \times 10^4\ m^3$, respectively. The spatial distribution of the events is shown in Figure 10.

Based on the fracturing data and the spatial distribution of the microseismic events, the cracks of the two wells did not connect with each other. However, due to the vertical positioning error caused by low signal-to-noise ratio raw data, some events have the problem of superposition in the vertical direction. In further processing, it is necessary to suppress the influence of noise and establish a more accurately acoustic velocity model of the formation.

5. Conclusions

We have presented a design for the wireless transmission system of microseismic monitoring in a complex mountainous area. It consists of a base station and a data transceiver based on Wi-Fi communication mode. The base station adopts an embedded system composed of processor, memory, and peripheral chip. It provides the communication backbone line for the network and cascade function with the trunk link. The maximum distance and transmission rate at point-to-point communication are 1000 m and 150 Mbps, respectively. The data transceiver adopts the network protocol based on Ad Hoc, which can minimize the power consumption and cost on the basis of general network layer functions. The baud rate is 300 to 460.8 kbps, the receiving sensitivity is −89 dBm operating at 11 Mbps. When it is connected with the antenna (transmission gain 6 dBi), the transmission distance is up to 300 m at the condition of the flat and no shelter environment.

The test results show that the system has the advantages of low BER (2.47×10^{-6}), fast transmission speed (up to 4.8 MHz at the distance of 400 m), long communication distance (about 2000 m), and stable and safety hardware. We have applied the system to in-site microseismic monitoring of the coalbed methane fracturing with 28 geophones in South Chongqing. It achieved a good performance with a transmission rate of 1.2 Mbps, a time delay of 1.95 ms, and a signal strength of up to −52.3 dBm for real-time data transmission in the field. These results indicate that the system has a great value for more applications of microseismicity in complex topography, such as the real time monitoring of landslide, unconventional oil and gas fracturing, and mining safety.

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 supported by the following funds: (1) Chongqing Natural Science Foundation Project: Study on dynamic evaluation method of four-dimensional stability of rock bank slope based on microseismic monitoring (cstc2020jcyj-msxmX1068); (2) Performance Incentive and Guidance Project from the Chongqing Science and Technology Bureau: "Research and application of dynamic monitoring technology of bank slope disaster in Three Gorges Reservoir area based on weak earthquake and seismic noise" (cstc2021jxjl20008); and (3) Performance Incentive and Guidance Project from the Chongqing Science and Technology Bureau: "Key technologies research on landslide early warning and dynamic visualization driven by monitoring data" (cstc2021jxjl0168). The authors are grateful to be supported by the Chongqing Science and Technology Bureau (Grant nos. cstc2020jcyj-msxmX1068, cstc2021jxjl20008, and cstc2021jxjl0168).

References

[1] J.-M. Kendall, A. Butcher, A. L. Stork, J. P. Verdon, R. Luckett, and B. J. Baptie, "How big is a small earthquake? challenges in determining microseismic magnitudes," *First Break*, vol. 37, no. 2, pp. 51–56, 2019.

[2] S. Maxwell, "Microseismic hydraulic fracture imaging: the path toward optimizing shale gas production," *The Leading Edge*, vol. 30, no. 3, pp. 340–346, 2011.

[3] C. L. Cipolla, S. C. Maxwell, and M. G. Mack, "Engineering guide to the application of microseismic interpretations," in *SPE Hydraulic Fracturing Technology Conference*, Society of Petroleum Engineers, Texas, USA, 2012.

[4] D. Fa, J. Qiu, M. Zhou et al., "Sichuan shale gas microseismic monitoring: acquisition, processing, and integrated analyses," in *International Petroleum Technology Conference*, Beijing, China, 2013.

[5] X. L. Lei, G. Z. Yu, S. L. Ma, X. Z. Wen, and Q. Wang, "Earthquakes induced by water injection at ~3 km depth within the Rongchang gas field, Chongqing, China," *Journal of Geophysical Research: Solid Earth*, vol. 113, Article ID B10310, 2008.

[6] D. Elsworth, C. J. Spiers, and A. R. Niemeijer, "Understanding induced seismicity," *Science*, vol. 354, no. 6318, pp. 1380-1381, 2016.

[7] M. Shirzaei, W. L. Ellsworth, K. F. Tiampo, P. J. Gonzalez, and M. Manga, "Surface uplift and time-dependent seismic hazard due to fluid injection in eastern Texas," *Science*, vol. 353, no. 6306, pp. 1416–1419, 2016.

[8] X. Luo and P. Hatherly, "Application of microseismic monitoring to characterise geomechanical conditions in longwall mining," *Exploration Geophysics*, vol. 29, no. 4, pp. 489–493, 1998.

[9] Y. Cao, L. M. Dou, C. B. Wang, X. X. Yao, and Y. Gu, "Microseismic precursory characteristics of rock burst hazard in mining areas near a large residual coal pillar: a case study from Xuzhuang coal mine, Xuzhou, China," *Rock Mechanics and Rock Engineering*, vol. 49, no. 11, pp. 1–16, 2016.

[10] Y. Li, T.-h. Yang, H.-l. Liu, X.-g. Hou, and H. Wang, "Effect of mining rate on the working face with high-intensity mining based on microseismic monitoring: a case study," *Journal of Geophysics and Engineering*, vol. 14, no. 2, pp. 350–358, 2017.

[11] G. Senfaute, A. Duperret, and J. A. Lawrence, "Micro-seismic precursory cracks prior to rock-fall on coastal chalk cliffs: a case study at Mesnil-Val, Normandie, NW France," *Natural Hazards and Earth System Sciences*, vol. 9, no. 5, pp. 1625–1641, 2009.

[12] N. W. Xu, C. A. Tang, L. C. Li et al., "Microseismic monitoring and stability analysis of the left bank slope in Jinping first stage hydropower station in southwestern China," *International Journal of Rock Mechanics and Mining Sciences*, vol. 48, no. 6, pp. 950–963, 2011.

[13] C. Colombero, C. Comina, S. Vinciguerra, and P. M. Benson, "Microseismicity of an unstable rock mass: from field monitoring to laboratory testing," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 2, pp. 1673–1693, 2018.

[14] C. B. Henry, R. S. Robert, B. B. Malcolm, V. G. Eric, and T. Jeff, "A remote, wireless microeismic monitoring system," *CREWES Research Report*, vol. 15, pp. 1–12, 2003.

[15] H. H. Bian, Y. D. Wang, Y. A. Cao, and W. Hao, "Wireless remote data acquisition and network transmission for microseismic signals," *Applied Mechanics and Materials*, vol. 419, pp. 555–562, 2013.

[16] F. Q. Zhang, T. L. Wang, J. C. Ye, G. Q. Huang, and G. F. Wang, "Design of microseismic monitoring data transmission system based on ad hoc," *Modern Electronics Technique*, vol. 39, no. 24, pp. 128–131, 2016.

[17] Y. F. Shao, Y. Han, J. Li, C. Shi, and M. Zhang, "WSN microseismic source location node design based on compression perception theory," *Microcontrollers & Embedded Systems*, vol. 10, pp. 19–22, 2017.

[18] Y. D. Y. Zhu, J. L. Wang, F. Sun, H. Lu, J. Lin, and Z. B. Chen, "Micro-seismic monitoring and instrument for hydraulic fracturing in the low-permeability oilfield," *Chinese Journal of Geophysics*, vol. 60, no. 11, pp. 4282–4293, 2017.

[19] S. Vinoth, L. Ajay Kumar, and E. Kumar, "Slope stability monitoring by quantification and behavior of microseismic events in an opencast coal mine," *Journal of the Geological Society of India*, vol. 85, no. 4, pp. 450–456, 2015.

[20] K. Harris and R. Bacon, "Utilizing source mechanism and microseismic event location to identify faults in real-time using wireless seismic recording systems - an eagle ford case study," *First Break*, vol. 33, no. 7, pp. 57–61, 2015.

[21] H. Y. Chen, R. W. Ouyang, and C. Wang, "Distributed source localization in wireless underground sensor networks," Eprint Arxiv, 2011, https://arxiv.org/abs/1112.4035.

[22] F. Sun, X. Wen, Z. Chen, Y. Zhu, and H. Lv, "Improvement of a microseismic monitoring data-transmission system based on a load-balancing scheme and a high-throughput polling mechanism," *IET Communications*, vol. 13, no. 20, pp. 3595–3600, 2019.

[23] S. Qiao, H. Duan, Q. Zhang et al., "Development of high-precision distributed wireless microseismic acquisition stations," *Geoscientific Instrumentation, Methods and Data Systems*, vol. 7, no. 3, pp. 253–263, 2018.

[24] Q. Zhang, S. Qiao, Q. Zhang, and S. Liu, "Design and implementation of the detection software of a wireless microseismic acquisition station based on the android platform," *Geoscientific Instrumentation, Methods and Data Systems*, vol. 10, no. 1, pp. 91–100, 2021.

[25] A. Dulmage and N. Ruddell, "Battery-powered wireless monitoring system for geotechnical, hydrology and microseismic sensors using the MineHop mesh network," in *Proceedings of the Ninth International Symposium on Field Measurements in Geomechanics*, pp. 613–620, Sydney, Australia, 2015.

[26] W. Rungshawang and Y. Suppakhun, "Development of wireless seismic sensor network for seismic activity sensing," in *Proceedings of the The 3rd International Conference on Engineering Science and Innovative Technology (ESIT2018)*, Phang-Nga, Thailand, 2018.

[27] J. J. Monteverde, J. S. Merino, and J. L. Llorens, "Design and implementation of a wireless sensor network for seismic monitoring of buildings," *Sensors*, vol. 2021, no. 11, Article ID 3875, 2021.

[28] Z. Y. Yin, Y. Zhou, and Y. X. Li Y, "Seismic exploration wireless Ssensor system based on Wi-Fi and LTE," *Sensors*, vol. 20, no. 4, Article ID 1018, 2020.

[29] H. Jamali-Rad and X. Campman, "Internet of Things-based wireless networking for seismic applications," *Geophysical Prospecting*, vol. 66, no. 4, pp. 833–853, 2018.

[30] H. Jamali-Rad, X. Campman, I. Mackay et al., "IoT-based wireless seismic quality control," *The Leading Edge*, vol. 37, no. 3, pp. 214–221, 2018.

[31] S. František, E. Leo, and J. M. Tijmen, "Stability of source mechanisms inverted from P-wave amplitude microseismic monitoring data acquired at the surface," *Geophysical Prospecting*, vol. 62, pp. 475–490, 2014.

[32] Nxp Semiconductor, "MPC8308: low-power PowerQUICC® II pro processor with DDR2, eSDHC, PCI express, eTSEC, USB, IEEE® 1588, Mpc8308Ec Datasheet," 2014, https://www.nxp.com.cn/docs/en/data-sheet/MPC8308EC.pdf.

[33] Nxp Semiconductor, "MC9S08QG8, MC9S08QG4 data sheet," 2014, https://www.nxp.com.cn/docs/en/data-sheet/MC9S08QG8.pdf.

[34] Q. Xie, D. Li, L. Cheng, F. Wang, Z. Huang, and D. Wang, "Noise suppression for micro-seismic on gas shale," *Acta Geologica Sinica - English Edition*, vol. 89, no. s1, pp. 367-368, 2015.

[35] S. Savazzi and U. Spagnolini, "Wireless geophone networks for high-density land acquisition: technologies and future potential," *The Leading Edge*, vol. 27, no. 7, pp. 882–886, 2008.

[36] A. Makama, K. Kuladinithi, and A. Timm-Giel, "Wireless geophone networks for land seismic data acquisition: a survey, tutorial and performance evaluation," *Sensors*, vol. 21, no. 15, p. 5171, Article ID 5171, 2021.

[37] B. J. Yi, T. Zhao, S. H. Yu, D. L. Lin, and J. H. Duan, "The present situation, development direction and existing probIems of seismic data acquisition system technology," *EGP*, vol. 28, no. 6, pp. 35l–358, 2018.

[38] Sercel, "UNITE, cable-free seismic acquisition," 2018, https://www.sercel.com/products/Lists/ProductSpecification/Unite_brochure_Sercel_EN.pdf.

[39] Wireless Seismic, "Wireless Seismic, Inc. Launches RT3-Ultra-High channel count seismic recording system featuring next generation radio technology," 2017, https://wirelessseismic.com/wireless-seismic-inc-launches-rt3-ultra-high-channel-count-seismic-recording-system-featuring-next-generation-radio-technology-2.

[40] International Seismic and Sigma, "The sum of all cableless experience, continuous seismic recording, system," 2021, http://www.iseis.com/documents/2.4GHz_issues_V2.pdf.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 3058499, 10 pages
https://doi.org/10.1155/2022/3058499

Research Article

A High-Performance Energy-Balanced Forwarding Strategy for Wireless Sensor Networks

Zhangxiang Hu, Xiaodan Jiang, Xiajun Ding, Kai Fang, and Xiaolong Zhou

College of Electrical and Information Engineering, Quzhou University, Quzhou 324000, China

Correspondence should be addressed to Xiajun Ding; 37050@qzc.edu.cn and Xiaolong Zhou; xiaolong@ieee.org

Received 15 March 2022; Accepted 7 April 2022; Published 5 May 2022

Academic Editor: Sai Zou

The wireless sensor network (WSN) is composed of several sensor nodes organized by multi-hop self-organization, which is a typical network for the industrial internet in industrial application. However, the energy using and processing capacity of each node are greatly limited. Therefore, it is of great significance to study energy-saving and efficient communication protocols for WSN. To prolong the lifetime of WSN and improve network throughput, a high throughput routing protocol with balanced energy consumption is proposed. The designed protocol first employs the *K*-means clustering algorithm to cluster the nodes, then calculates the weights based on the residual energy of and distance between the nodes, and finally selects the best node as the cluster head. Moreover, the optimal size of the package is determined by the parameters of the wireless transceiver and the channel conditions. In the data transmission stage, the Dijkstra algorithm is used to calculate the multi-objective weight function as the link cost. Experimental results demonstrate the superior performance of the proposed protocol over the CERP and TEEN routing protocols in terms of energy saving of network nodes, so as to improve the throughput and survival time of the entire system.

1. Introduction

Wireless sensor networks (WSNs) are broadly applied in the Industrial Internet of Things to enhance the productivity and efficiency of existing and prospective manufacturing industries. WSN is composed of a large number of low-power, micro-smart sensor nodes that are randomly deployed to perform sensing tasks in the monitoring area. In recent years, WSNs have played an important role in the production of life, environmental monitoring, national security, and other fields [1–5]. Because sensor nodes are usually powered passively with limited energy and weak computing power, it is important to build an energy-efficient WSN. However, how to construct an excellent WSN remains a challenging problem. Network topology control provides an effective way to solve this problem. Generally speaking, topology control refers to an underlying network topology conversion technology that can enhance system performance or minimize routing costs [6]. Clustering is an effective and widely used network topology scheme among topology control technologies. In addition, the clustering-based routing protocol helps to save the energy of sensor nodes, thereby prolonging the network lifetime [7, 8].

A typical clustering topology structure is shown in Figure 1. Each cluster contains a cluster member node (CM) and a cluster head node (CH). The CM is used to collect data and transmit the results to the CH, and then the CH merges the data within the cluster and forwards it to the remote base station (BS) through multi-hops transmission afterwards. The process is mainly divided into two stages: (1) the network topology construction stage, which carries out network clustering and cluster head node screening; and (2) the steady-state stage, which carries out data communication, data fusion, and data transmission.

Currently, many researchers have focused on research of clustering routing protocols. For example, Fan and Song [9] proposed an improved low-energy adaptive clustering

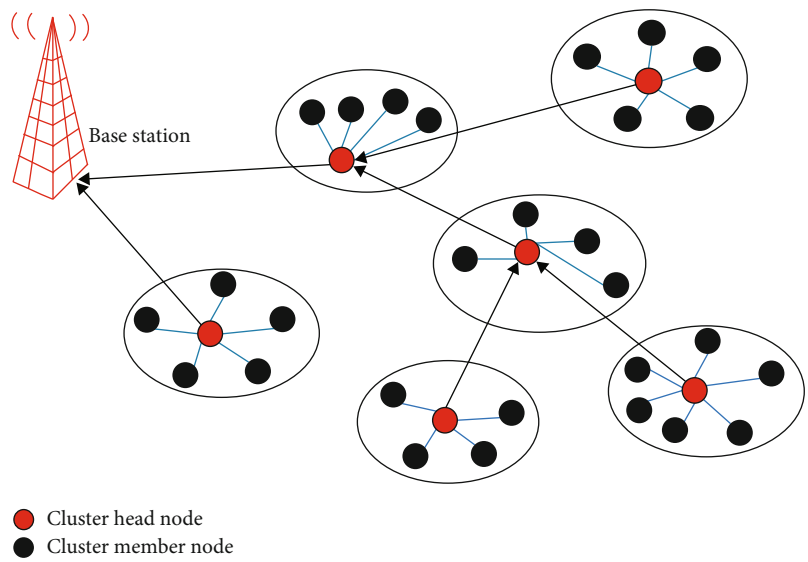


Figure 1: The structure of WSN clustering topology.

hierarchy (IM-LEACH), which was the most popular active routing protocol based on clustering. This protocol was composed of self-configuration and random adaptive methods, and provided local and low-energy access control for data transmission. Razzaq et al. [10] put forward an optimal K-means clustering scheme for data packets, which was another K-means clustering protocol based on high-energy efficiency. The protocol used a weighted function for cluster head selection to minimize node energy consumption and enhance system performance. Additionally, Liu et al. [11] proposed a multi-channel AODV routing protocol based on the Dijkstra algorithm, which both the energy consumed by nodes and the energy consumed during the data transmission were premeditated. Furthermore, the Dijkstra algorithm was used to select the path with the least energy consumption from the multi-data transmission channels. A clustering routing protocol based on the Dijkstra algorithm was suggested hereafter by Abderrahim et al. [12]. It was a centralized routing protocol in which the BS assigned a weight matrix to the network and then applied the Dijkstra algorithm to calculate the optimal data path from the source to the destination node. This approach was suitable for schemes that require periodic or query-based data reporting. Chen et al. [13] improved the traditional LEACH protocol, which considered load balance in the clusters. The main disadvantage of this scheme was that the residual energy of the node was neglected when selecting the relay nodes in the data transmission stage.

To ensure the high throughput of the WSN cluster communication protocol, a high throughput routing protocol with balanced energy consumption is proposed in this paper. The K-means algorithm is employed to achieve WSN clustering when exchanging data between nodes. For the CH node selection, we calculated two weight functions and opted the node not only containing the largest residual energy but also distancing the least to the initial cluster center accordingly. Since CH node is the only node that needs to aggregate data from CM nodes and relay it to BS, the proposed method guarantees a balanced energy consumption in the network. Moreover, the traditional Dijkstra algorithm is employed and the multi-objective weight function is used as link cost, which helps to refine the energy efficiency of the cluster data communication and eliminate the premature death of some nodes. The main contributions of this paper are those given here.

(1) The best cluster head is selected according to the remaining energy, and the size of the cluster is reasonably controlled according to the channel state. It is helpful to reduce the communication energy consumption of the nodes in the cluster and balance the energy consumption between the WSN clusters

(2) The distance between cluster heads and the remaining energy of cluster heads are fully taken into consideration when routing data. Compared with the conventional methods, the energy between cluster heads is well counterpoised and the network throughput is greatly ameliorated

The rest of this paper is organized as follows. In Section 2, related models include WSN topology model and energy consumption model of wireless communication are introduced. In Section 3, the proposed high throughput routing protocol for nodes with balanced energy consumption is presented in detail. The experimental results and analysis are discussed in Section 4, followed by concluding remarks in Section 5.

2. Related Models

2.1. WSN Topology Model. The WSN consists of N wireless communication nodes, numbered as S_i (i=1,2, ….., N). The nodes are randomly distributed in a rectangular area of $L * L$; the position of BS is fixed and far away from the sensing area. Nodes are then grouped into clusters, within which contains a CH node and several CM nodes. Due to

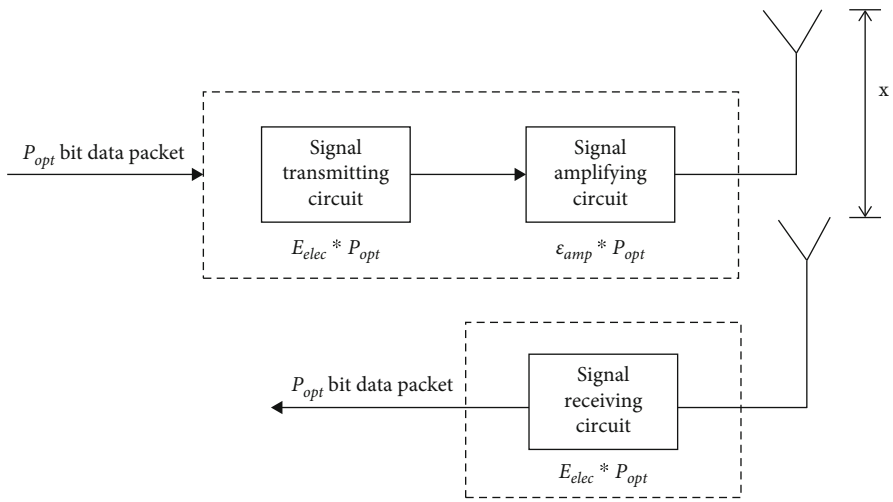


Figure 2: Energy consumption model of wireless communication.

the infinite energy and high computational capacity of the BS, the network clustering and the selection of CH nodes are carried out in the BS, and the selected CH information is notified to the CM nodes in the cluster. This process can reduce the energy consumption of the network topology construction. When constructing the network clustering topology, the nodes wait until a sensing event occurs, and the CM node that is close to the occurring position in the cluster will sense the environment and transmit data to the CH node. The CH node then fuses collected information and uses the Dijkstra's shortest path algorithm to efficiently transmit data to the BS in multi-hops.

2.2. Energy Consumption Model of Wireless Communication. Wireless communication between sensor nodes involves many aspects of energy consumption, such as signal amplification energy consumption, signal transmission energy consumption, and data processing energy consumption. In this paper, a classical energy consumption model [14] is utilized to evaluate the energy consumption of nodes. As shown in Figure 2, a large amount of energy is consumed in the data transmission stage in WSN. Theoretically, when sending the same data packet, the total energy consumption depends on the size of the data packet. To minimize the energy consumption of sending packets, the optimal packet size P_{opt} is considered to diminish energy consumption during data transmission. P_{opt} has been proved in [10], and its calculation can be formulated in Eq. (1).

$$P_{\text{opt}} = \frac{\sqrt{C_0^2 - 4C_0/\ln(1-p)} - C_0}{2}, \tag{1}$$

where $C_0 = +K_1$, α represents the header size (unit: bit), K_1 represents the energy consumed by the payload in communication, K_2 represents the energy consumed by the node to start up, and p represents the bit error rate (BER) of the channel. Thus, the energy E_{TX} and E_{RX} exploited by nodes to send and receive P_{opt} bit data are formulated in Eq. (2) and Eq. (3), respectively.

$$E_{\text{TX}}(P_{\text{opt}}, x) = P_{\text{opt}} \cdot E_{\text{elec}} + P_{\text{opt}} \cdot \varepsilon_{\text{amp}}, \tag{2}$$

$$E_{\text{RX}}(P_{\text{opt}}) = P_{\text{opt}} \cdot E_{\text{elec}}, \tag{3}$$

where x represents the distance between sensor nodes, E_{elec} represents the energy consumed by sending or receiving unit bit, and ε_{amp} represents the energy consumed by signal amplification of sending nodes as in Eq. (4).

$$\varepsilon_{\text{amp}} = \begin{cases} \varepsilon_{\text{fs}} \cdot x^2 \text{if}(x < x_{\text{th}}) \\ \varepsilon_{\text{mp}} \cdot x^4 \text{if}(x \geq x_{\text{th}}) \end{cases}, \tag{4}$$

where x_{th} denotes the distance threshold, and ε_{fs} and ε_{mp} denote the amplification energy and multipath fading parameters of the free space signal, respectively. If the distance x between the sending node and the receiving node is greater than or equal to x_{th}, the multipath fading channel model will be used. Otherwise, the free space propagation model will be used. Thus, a P_{opt} bit size packet is sent to and received, and the total energy consumed is E_{total} as in Eq. (5), which can be expanded to Eq. (6).

$$E_{\text{total}} = E_{\text{TX}} + E_{\text{RX}} + E_{\text{DA}}, \tag{5}$$

$$E_{\text{total}} = P_{\text{opt}} \cdot (2 \cdot E_{\text{elec}} + \varepsilon_{\text{amp}}) + E_{\text{DA}}, \tag{6}$$

where E_{DA} represents the energy consumed by data in the process of CH node fusion. Suppose that x_{CH} denotes the distance between the CM node and the CH node, and x_{BS} denotes the distance between the CH node and the BS, then the residual energy of CM node after CM transmits P_{opt} bit data to CH can be formulated in Eq. (7).

$$E_{\text{CM}} = E_{\text{init}} - E_{\text{TX}}(P_{\text{opt}}, x_{\text{CH}}), \tag{7}$$

K-means WSNs clustering algorithm
Step 1: Using Eq. (9) to calculate the value of the optimal cluster number *K*. Then, randomly select K nodes as the initial cluster head nodes.
Step 2: Calculating the Euclidean distance from each node to the clustering center and assigning the node to the nearest clustering center.
Step 3: Calculating the centers of all nodes within a particular cluster and updating the centers of the clusters.
Step 4: Repeating Step 2 for the new clustering center. If the cluster to which the node belongs changes, repeat Step 3 or stop the algorithm.

Algorithm 1: *K*-means WSNs clustering algorithm.

After receiving the data from the CM, the CH performs data fusion, and finally transmits the fused data to BS. The residual energy of CH node is then can be calculated in Eq. (8).

$$E_{\mathrm{CH}} = E_{\mathrm{init}} - E_{\mathrm{RX}}\left(P_{\mathrm{opt}}\right) - E_{\mathrm{DA}} - E_{\mathrm{TX}}\left(P_{\mathrm{opt}}, x_{\mathrm{BS}}\right). \quad (8)$$

3. Routing Protocol

3.1. Network Deployment Stage. The sensor nodes are randomly deployed in the monitoring area. Considering the nodes' inability to form a routing table, data transmission cannot be realized. Therefore, the sensor nodes in the network broadcast beacon to their neighbors in the communication range. The beacon frame includes the node number and node position. After all nodes in the network broadcasting beacon frame, all nodes can set up neighbor list, and routing tables can be easily generated in accordance with the list. Given to the high power of the BS, it can broadcast an initialization request data frame to all nodes in the network in a single hop. The nodes in the network will reply to the data immediately after receiving the message frame. According to the routing table inside the node, the response data frame is sent back to the BS along the shortest path provided by the Dijkstra algorithm in the form of multi-hops. The response data frame contains the residual information of the node residual energy and node position, assuming that the node position in the network has been obtained by GPS or positioning technology [15, 16]. So far, the deployment of WSN has been completed.

3.2. Network Clustering Stage. In this stage, the *K*-means clustering algorithm is employed for clustering the whole sensor network, which is an unsupervised learning algorithm that collects data into *K* clusters. Sunil et al. [17] have made a detailed study on the clustering problem which was most suitable for WSNs, and the value of *K* was calculated in Eq. (9). When the network is clustered into *K* clusters, the cluster similarity between clusters is low, while the intra-class similarity is high.

$$K = \sqrt{\frac{N}{2\pi} \cdot \frac{\varepsilon_{\mathrm{fs}}}{\varepsilon_{\mathrm{mp}}}} \cdot \frac{F}{x_{\mathrm{BS}}^2} \quad (9)$$

where N represents the number of sensor nodes in the network, F represents the dimension (two-dimensional plane) of a given network, x_{BS} represents the average distance between sensor nodes and BSs in the network, and K represents the most suitable class number divided from the original clusters.

The pseudo-code of *K*-means WSNs clustering algorithm is detailed in Algorithm 1. After the WSN has run for a fixed number of rounds, Algorithm 1 is re-executed to update the network clustering.

3.3. CH Node Selection Stage. Once a WSN is divided into *K* clusters, one CH node is selected from each clustering. The main function of the CH node is to aggregate the data of the CM nodes in the clustering, and forwards the data to the server by multi-hop after data fusion. Because the CH node needs to bear greater load, from the perspective of network energy balance, nodes with more residual energy and closer to the average distance of the CM node should be selected as the CH node. By doing so, the premature death of the CH node can be avoided and the energy consumption of the data transmission of the CM nodes can be decreased. In this paper, a weight function is designed to help select CH nodes, as in Eq. (10).

$$W_i = C_1 \cdot E_i + C_2 \cdot D_{C2i} \quad (10)$$

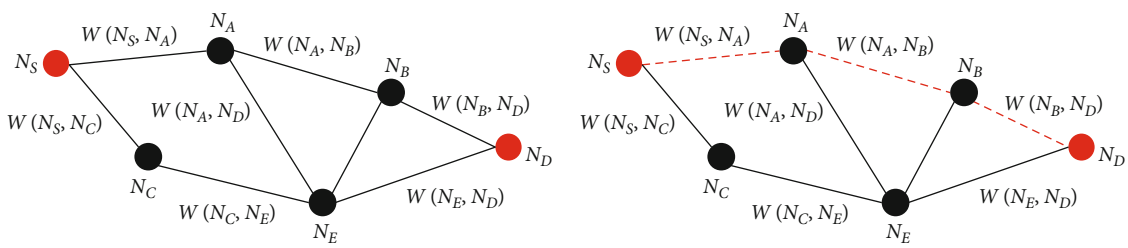
where $i = 1, 2, 3, 4 \ldots N$, C_1 and C_2 are constants, W_i represents the weight value of each node in the cluster, D_{C2i} represents the distance between the ith node and the center of the cluster, and E_i represents the residual energy of the ith node.

After obtaining the weight of all the sensor nodes in the cluster, a criterion is needed to determine which nodes are more suitable as CHs. In this paper, the evaluation function is designed based on energy consumption balance and energy-saving goals, as in Eq. (11).

$$W_{\mathrm{std}} = C_1 \cdot E_{\max} + C_2 \cdot \mathrm{avg}(D_{C2i}) \quad (11)$$

where W_{std} represents the standard weight, C_1 and C_2 are constants, $avg(D_{C2i})$ represents the average distance from all nodes in the cluster to the center, and $E_{\max}$ represents the highest residual energy of nodes in the cluster.

The weight W_i of each node in the cluster is compared with the standard weight value W_{std}, and node i corresponding to the minimum value of $|W_i - W_{\mathrm{std}}|$ is used as the CH. To ensure that the energy consumption of intra-cluster



(a) Cluster head neighbor topology

(b) Minimum energy consumption transmission path

Figure 3: Minimum energy consumption path selection.

Table 1: Experimental parameter settings.

Parameter	Value
E_{elec}	50 nJ/bit
E_{init}	0.5 J
E_{DA}	5 nJ/bit/signal
$L \times L$	100 m ×100 m
Number of nodes N	100
Coordinate of BS	(50 m,0 m)
Optimal grouping P_{opt}	1701bits
ε_{fs}	10pJ/bit/m^2
ε_{mp}	0.0013pJ/bit/m^4

nodes is balanced, the CH nodes are re-selected as described above for each fixed number of data acquisitions.

3.4. Data Transmission Stage. So far, the clustering and CH node selection of WSNs have been completed, at which time sensor nodes can collect and transmit data. Since most of the data collected by sensor nodes are basically meaningless, only when the sensing value exceeds a certain threshold value can the sensing data be meaningful. Therefore, this paper proposes an event-driven routing protocol. When the sensed data is larger than the threshold value T, the data is transmitted to the CH node. The CH node fuses together the data into a plurality of member nodes in the cluster, and then sends the data to the BS. The CH node is also selected from the same kind of nodes, and their communication distance is limited as well. So multi-hop transmission between cluster heads is crucial to transmit data to the BS. Therefore, the Dijkstra algorithm is employed in this paper to find the best path from the source CH node to the target BS according to the link cost. The main steps are detailed as follows.

Step 1: Create a set S, which contains only the source node N_S and the link cost $W(N_S, N_A)$ from N_S to the adjacent node N_A at the initial time.

$$S = \{N_s\}. \tag{12}$$

The weight function proposed in this paper, as link cost $W(N_S, N_A)$, contains two factors: the residual energy of the next hop node (the node to receive data) and the distance from the source node to the next hop node, as in Eq. (13),

$$W(N_S, N_A) = d1 \cdot E_{NA} + d2 \cdot D_{S2A}, \tag{13}$$

where $W(N_S, N_A)$ represents the link cost between the source node N_S and the neighboring node N_A, and D_{S2A} represents the distance between the source node N_S and the neighboring node N_A. d_1 and d_2 are adjustable parameters and satisfy $0 \le d_1 < d2$ and $d_1 < d_2 \le 1$.

Step 2: If the source node N_S and its neighboring node N_A are not in the set S, then adding the node to S, and subsequently continue to find the node N_C adjacent to the source node N_S. If the node is not in the set S either, then adding the N_C to S, and the link cost is $W(N_S, N_C)$.

Step 3: Compare the link cost $W(N_S, N_A)$ with $W(N_S, N_C)$, and then select a neighboring node with the least link cost as the next hop node of the source node, as in Eq. (14).

$$W(N_S, N_N) = \min\,[W(N_S, N_A), W(N_S, N_C)]. \tag{14}$$

Step 4: Use the neighboring node selected in Step 3 as the source node, and then repeat Steps 2 and 3 until the minimum cost of data transferring from the CH to the BS is found, that is, finding a transmission path that minimizes the Eq. (15). When data is transmitted to the BS, the algorithm ends.

$$C_{\min} = \sum_{i=1}^{m-1} W(Ns, N_D), \tag{15}$$

where N_D represents the BS and m represents the number of nodes included in the transmission path.

The minimum cost path selection process for transmitting data to the BS by multi-hop between CHs is shown in Figure 3. Finally, the path with the minimum energy consumption is selected to transmit data from CH to BS.

4. Experimental Results and Analysis

In this paper, the experiment is simulated on MATLAB platform. N =50 nodes are randomly deployed in a square area with L =100 m. These nodes are divided into K clusters, and each cluster contains N/K nodes. The performance of the data protocol for the WSN is evaluated according to three parameters: residual energy, energy consumption

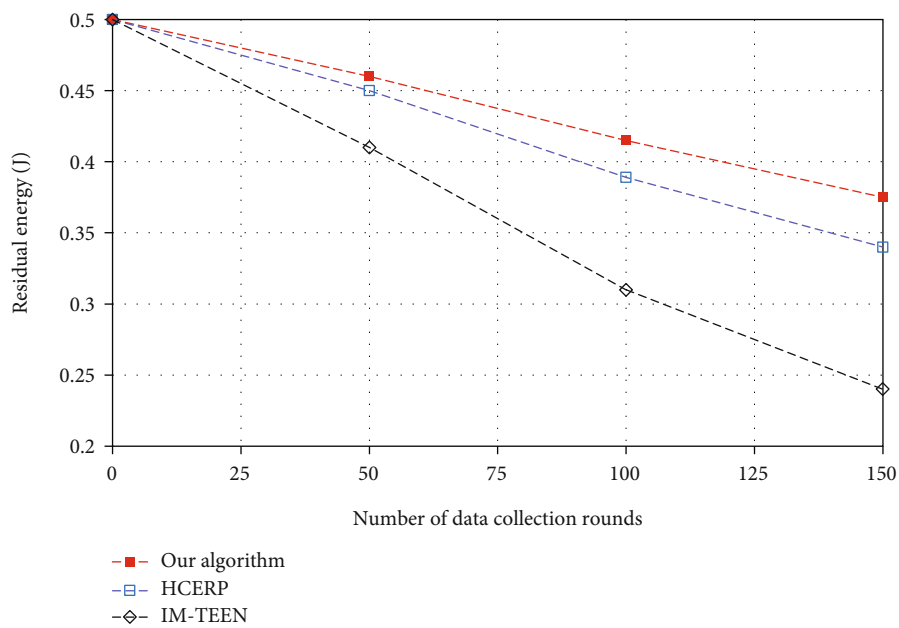


Figure 4: Average residual energy of nodes.

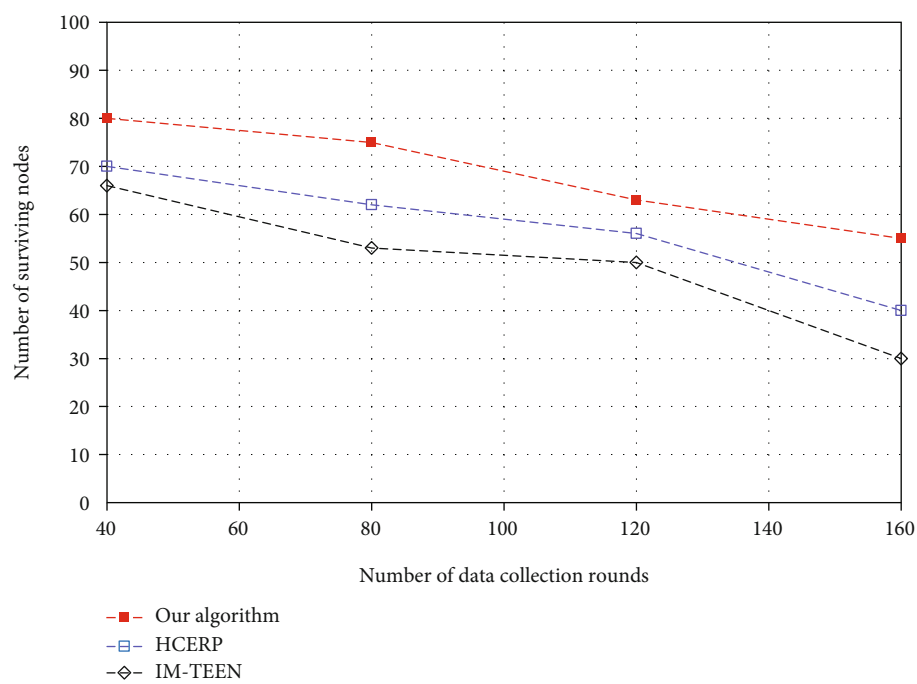


Figure 5: Number of surviving nodes.

balance, and network throughput. The residual energy is evaluated to illustrate the energy saving of the data routing protocol, the energy consumption balance is evaluated to demonstrate the energy consumption balance within the WSN, and the network throughput is evaluated to indicate the amount of data successfully transmitted to the BS per unit time. In the experiments, the concept of "round" is defined, and each "round" carries out 100 data transmissions. Each node in the WSN transmits P_{opt} data at each time step. Due to the mobility of the nodes in the network, their distribution is not fixed. We assumed that for every 10 "rounds" of data communication, the location of the nodes will change randomly, thus forming a new distribution of the node cluster. Therefore, for every 10 "rounds"

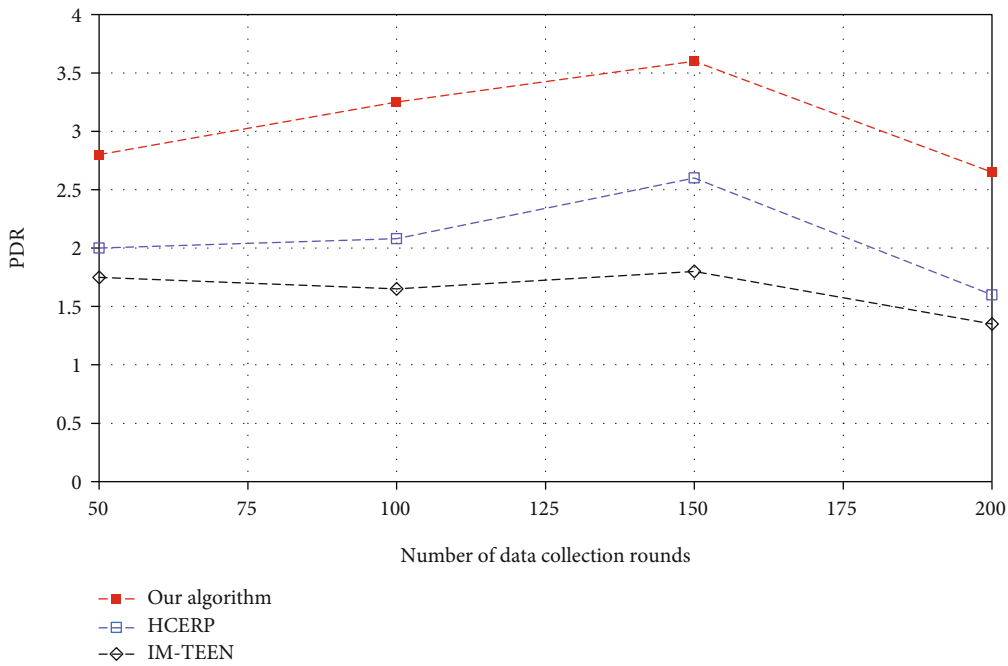


Figure 6: Packet transfer rate (K/s) and node density.

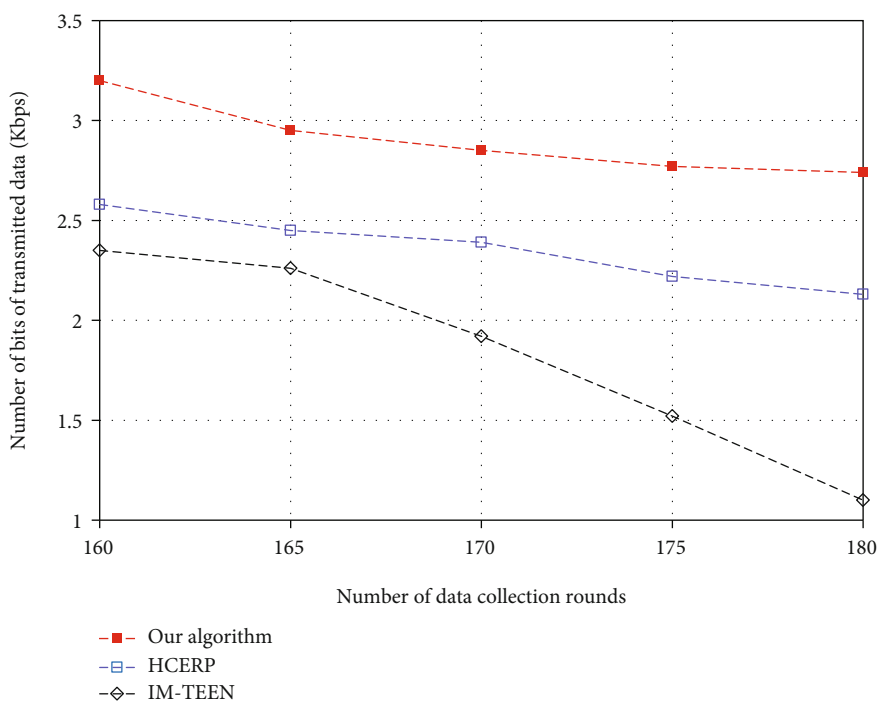


Figure 7: Network throughput and rounds.

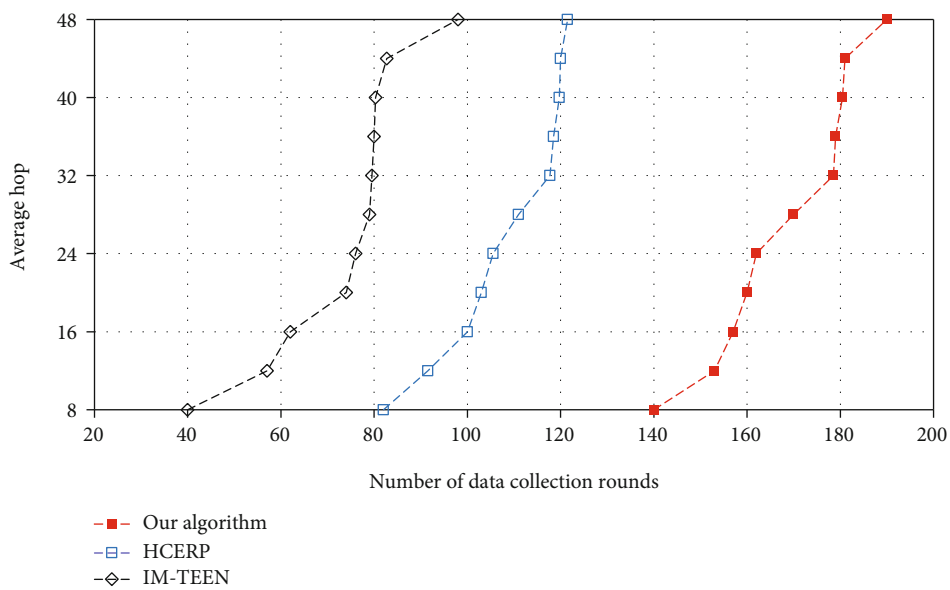


Figure 8: Node survival and rounds.

of data communication, the CH and cluster need reselecting. The specific parameter settings of the data communication model are listed in Table 1.

The optimal grouping size in Table 1 has been validated in reference [10]. Experiments compare the CERP improved protocol HCERP [18] and the TEEN improved IM-TEEN [19] routing protocol. Sensor networks have the following properties.

Property 1. Being homogeneous in nature, sensor clusters equip with the same initial energy, and are randomly deployed in the monitoring area.

Property 2. The whole network contains only one BS.

Property 3. After deployment, the sensor nodes and BS locations are fixed and known.

Property 4. CH collects the information of nodes in the cluster, and then forwards it to the BS in the form of multi-hop.

Property 5. The cluster uses the Rayleigh fading channel model [18, 20].

4.1. Analysis of Average Residual Energy of Nodes. With the operation of the sensor network, the residual energy of the sensor nodes will gradually decrease. The residual energy of the node can effectively reflect the energy consumption of the routing protocol. This experiment verifies the average residual energy of all nodes in the network when node communicates data in different rounds. As shown in Figure 4, the horizontal coordinate is the number of rounds of data communication, and the vertical coordinate is the average residual energy of nodes in the WSN.

The experimental results show that with the increase of data communication rounds, the average residual energy of nodes corresponding to the three routing protocols gradually decreases. After conducting the same number of routes of data communication, the average residual energy of the proposed routing protocol is higher than that of GCERP and IM-TEEN protocols, meaning that its performance is better than that of HCERP and IM-TEEN protocols. Because nodes in the IM-TEEN protocol keep sensing situated environment to reach the desired threshold, energy loss would be inevitable. The HCERP protocol increases energy consumption at the beginning of each round because of uneven clustering. The scheme proposed in this paper, however, takes energy and distance into account, shaping the whole network more energy-efficient and the entire network less energy-consuming.

4.2. Analysis of Node Lifetime. In the light of the different positions of nodes in the network, each node consumes different amount of energy in data communication and transmission, which will result in unbalanced energy consumption of the whole WSN. Theoretically, the early existence of multiple nodes in the network could risk paralyzing the whole network and subsequently lose its original function. Experiments verify the number of dead nodes in the network when three routing protocols are used to run data communication in different rounds. The experimental results are shown in Figure 5.

Figure 5 depicts the relationship between the number of surviving sensor nodes and the data communication rounds of the sensor network. As far as the results are considered, as the data communication rounds in sensor networks increases, the number of surviving sensor nodes decreases slowly. With the increase of communication rounds, the energy consumed by the nodes increases. The farther away some nodes are from the CH, the faster the energy is consumed, resulting in the death of these nodes with the increase of communication times. Meanwhile, when communicating for the same number of rounds in the network, the number of surviving nodes in our network is higher than

that of the HCERP and IM-TEEN protocols. Owning to the fact that the proposed routing protocol fully considers the residual energy of nodes and the distance between nodes and the center of the cluster, it can equally process the two factors, making the energy consumption of nodes inside the cluster more balanced.

4.3. Analysis of Network Transmission Rate and Throughput. Network transmission rate and network throughput are key indicators for WSNs. Network transmission rate PDR indicates the amount of data received by the BS in a unit time. The experiment verifies the network transmission rate and network throughput of the three routing protocols under the same conditions. The experimental results are shown in Figures 6 and 7, respectively.

Figure 6 shows that with the increase of the number of nodes, the PDR of IM-TEEN and HCERP ascends to a certain extent, and then begins to descend. When the number of nodes in the network is 150, the network transmission rates of the three routing protocols are the highest. Even though the number of nodes is dropping, the data transmission rate of the proposed routing protocol is still much higher than that of HCERP and IM-TEEN protocols, which indicates that the proposed routing protocol has the highest scalability out of the three.

Figure 7 shows the relationship between data communication rounds and overall throughput for WSNs. The throughput of WSN is calculated as the total number of data packets successfully received by the BS after each round of data communication. The experimental results show that the proposed algorithm has higher throughput than other algorithms. The main reason for the increase in network throughput is that the number of surviving nodes in HCERP and IM-TEEN protocols is less than that proposed in this paper for the same number of data communication rounds. In addition, the main reason for the significant decrease of TEEN throughput is that the node will not send data to the BS if the specific threshold is not reached.

4.4. Analysis of Surviving Nodes. As the number of collection rounds in the system increases, the energy of the nodes in the sensor network will also be consumed, and part of the energy in the nodes will be exhausted. This experiment uses the Average hop as an indicator of the energy consumption balance of the routing protocol. The experimental results shown in Figure 8 suggest that as the number of rounds collected by the system increases, the average hops for packet transmission also increase. As a result of the climbing number of dead nodes, packet transmission is put through many unnecessary communication links. On top of that, the experimental results reveal that our algorithm has the shortest average hops under the same conditions.

5. Conclusion

The routing protocol in WSN has problems such as uneven energy consumption and low throughput. To remedy these, this paper proposed a high throughput routing protocol for WSN with balanced energy consumption. In the cluster head data transmission stage, this protocol used the Dijkstra algorithm to take the weight function as the communication overhead of the links between adjacent nodes. The weight function considered the energy of nodes and the distance to neighboring nodes, which increased the lifetime of nodes and the throughput of the whole network. Moreover, the improved K-means clustering algorithm has been employed to divide the sensor nodes into K clusters, so that the energy consumption balance can be fully considered when selecting the CH. Experimental results have demonstrated that the proposed scheme has longer network life and higher network throughput. Furthermore, we can consider other indicators (such as network delay or reliability) to study the routing protocols in the future work.

Data Availability

The data used to support the findings of this study have not been made available because we use the MATLAB to simulate routing algorithms and mainly verify the performance metrics of several algorithms, so there is no need for a public dataset. The paper provides a comparative analysis through the data generated during the execution of the algorithms.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The Acknowledgements part is corrected to "This work was supported in part by National Natural Science Foundation of China (61876168), Zhejiang Provincial Natural Science Foundation of China (LGN21C130001, LGF21F010002), Quzhou Science and Technology Projects (2019K17, 2020K19), and National College Students' Innovation and Entrepreneurship Training Program (202111488012).

References

[1] X. Xuan, J. He, P. Zhai, A. Ebrahimi Basabi, and G. Liu, "Kalman filter algorithm for security of network clock synchronization in wireless sensor networks," *Mobile Information Systems*, vol. 2022, Article ID 2766796, 11 pages, 2022.

[2] P. J. Zhao, G. B. Hu, and L. T. Wan, "A novel sparse array configuration with low coarray redundancy for DOA estimation in mobile wireless sensor network," *Mobile Information Systems*, vol. 2021, Article ID 1362640, 8 pages, 2021.

[3] J. Wang, Y. Gao, W. Liu, A. K. Sangaiah, and H. J. Kim, "Energy efficient routing algorithm with mobile sink support for wireless senor networks," *Sensors*, vol. 19, no. 7, p. 1494, 2019.

[4] J. Wang, H. Han, H. Li, S. He, P. K. Sharma, and L. Chen, "Multiple strategies differential privacy on sparse tensor factorization for network traffic analysis in 5G," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1939–1948, 2022.

[5] K. Vijayalakshmi and P. Anandan, "Global levy flight of cuckoo search with particle swarm optimization for effective

cluster head selection in wireless sensor network," *Intelligent Automation & Soft Computing*, vol. 26, no. 2, pp. 303–311, 2019.

[6] W. H. Ren, K. Hao, C. Li, X. du, Y. Liu, and L. Wang, "Fuzzy probabilistic topology control algorithm for underwater wireless sensor networks," in *Proceedings of International Conference on Artificial Intelligence for Communications and Networks*, pp. 435–444, Harbin, China, 2019.

[7] S. Balaji, E. G. Julie, and H. Y. Robinson, "Development of fuzzy based energy efficient cluster routing protocol to increase the lifetime of wireless sensor networks," *Mobile Networks and Applications*, vol. 24, no. 2, pp. 394–406, 2019.

[8] S. E. Pour and R. Javidan, "A new energy aware cluster head selection for LEACH in wireless sensor networks," *IET Wireless Sensor Systems*, vol. 11, no. 1, pp. 45–53, 2021.

[9] X. N. Fan and Y. L. Song, "Improvement on LEACH protocol of wireless sensor network," in *Proceedings of International Conference on Sensor Technologies and Applications*, pp. 260–264, Valencia, Spain, 2007.

[10] M. Razzaq, D. D. Ningombam, and S. Shin, "Energy efficient K-means clustering-based routing protocol for WSN using optimal packet size," in *Proceedings of International Conference on Information Networking*, pp. 632–635, Chiang Mai, Thailand, 2018.

[11] C. X. Liu, Y. Li, W. Cheng, and G. Shi, "An improved multichannel AODV routing protocol based on dijkstra algorithm," in *Proceedings of IEEE Conference on Industrial Electronics and Applications*, pp. 547–551, Xi'an, China, 2019.

[12] M. Abderrahim, H. Hakim, H. Boujemaa, and F. Touati, "A clustering routing based on dijkstra algorithm for WSNs," in *Proceedings of International Conference on Sciences and Techniques of Automatic Control and Computer Engineering*, pp. 605–610, Sousse, Tunisia, 2019.

[13] Y. L. Chen, L. Y. Jiang, and Y. R. Mu, "A LEACH-based WSN energy balance routing algorithm," in *Proceedings of the World Symposium on Software Engineering*, pp. 37–41, Wuhan, China, 2019.

[14] P. Bakaraniya and S. Mehta, "K-LEACH: an improved LEACH protocol for lifetime improvement in WSN," *International Journal of Engineering Trends and Technology*, vol. 4, no. 5, pp. 1521–1526, 2013.

[15] P. Poonkuzhlai and R. Aarthi, "Child monitoring and safety system using WSN and IoT technology," *Annals of the Romanian Society for Cell Biology*, pp. 10839–10847, 2021.

[16] K. Fang, T. Wang, X. Zhou, Y. Ren, H. Guo, and J. Li, "A TOPSIS-based relocalization algorithm in wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1322–1332, 2022.

[17] S. Sunil, K. Prabhat, and S. Jyoti, "A survey on successors of LEACH protocol," *IEEE Access*, vol. 5, pp. 4298–4328, 2017.

[18] C. Sudhamani and M. S. S. Ram, "Cooperative spectrum sensing over Rayleigh fading channel," in *Innovations in Electronics and Communication Engineering*, H. Saini, R. Singh, V. Patel, K. Santhi, and S. Ranganayakulu, Eds., vol. 33 of Lecture Notes in Networks and Systems, Springer, Singapore, 2019.

[19] G. P. Maheshwari and A. K. Sharma, "Modified TEEN for handling inconsistent cluster size problem in WSN," in *2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1–6, Chennai, India, 2018.

[20] G. Anwar, N. Husnain, U. I. Muhammad, M. K. Khan, and A. Hassan, "Energy efficiency in multipath Rayleigh faded wireless sensor networks using collaborative communication," *IEEE Access*, vol. 7, pp. 26558–26570, 2019.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 2594000, 14 pages
https://doi.org/10.1155/2022/2594000

Research Article

Prediction of Industrial Network Security Situation Based on Noise Reduction Using EMD

Guanling Zhao,[1] **Lisheng Huang**,[1] **Lu Li**,[1] **and Yongfeng Zhang**[2]

[1]*University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China*
[2]*Sichuan University, Chengdu 610065, China*

Correspondence should be addressed to Lisheng Huang; lsh@uestc.edu.cn

Received 7 January 2022; Revised 2 March 2022; Accepted 7 March 2022; Published 23 March 2022

Academic Editor: Sai Zou

Industrial Internet security is a prerequisite to ensure the high-quality development of the Industrial Internet. The significant way to curb Industrial Internet security accidents and prevent cyber threats proactively is effectively controlling the changes in network situations. In this paper, we propose a new prediction model based on Long Short-Term Memory (LSTM), minimum mean square variance criterion (MMSVC), and empirical mode decomposition (EMD), with the aim of effective noise reduction and high prediction accuracy. To minimize the disturbance of random noise, we firstly deleted several outliers in high-frequency and noisy Intrinsic Mode Functions (IMFs) decomposed by EMD. MMSVC performs well in identifying noisy IMFs without using thresholds. For the blank places, we refilled them by a certain weight with relevant figures. After that, the LSTM model was applied to predict the denoised signal. The preliminary experimental analysis illustrated that noise reduction with the EMD method could provide a significant boost in forecasting performance.

1. Introduction

With the rapid development of the Industrial Internet, a growing number of production services are integrated with the Internet. It means that many industrial components such as R and D, production, and management are exposed to the Internet. The data covered by the Industrial Internet is diverse and widely distributed. Once the network is attacked, the production and the development of enterprises will be seriously affected. Therefore, effective network security protection is extremely important to ensure the high-quality development of the Industrial Internet.

Network Security Situation Awareness (NSSA) [1] is one of the most popular technologies in cybersecurity. Compared with the traditional methods, the essential components to the NSSA are evaluating the network security situation and predicting the trend of network characteristics. Faced with cyberthreats, it can help the network administrators make decisions efficiently and nip the matter in the bud. The network security situation can be abstracted as a multidimension time series like Equation (1), where x, y, and m denote different network characteristics. So, the network security situation prediction is a forecast of this multidimensional time series actually. It applies statistical models or other models to analyze the sets of historical network characteristics.

Nowadays, Artificial Neural Network (ANN) [2] is the most common method used in network security situation prediction, which has the distinguished advantages of self-learning and computing. To find the optimal parameters and construct the training model, the historical data sets need to be trained repeatedly. The accuracy of prediction is heavily reliant upon whether the collected data is effective. Nonetheless, a large number of noise points will inevitably appear in the data extraction process, leading to random errors. The more noise points exist, the more unreliable

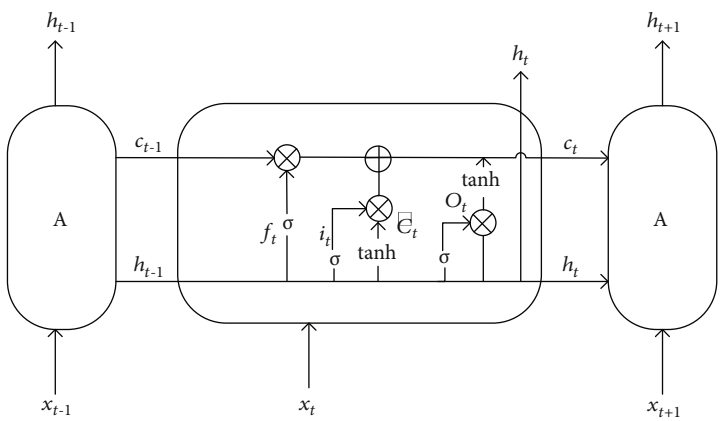


Figure 1: Architecture of the LSTM model.

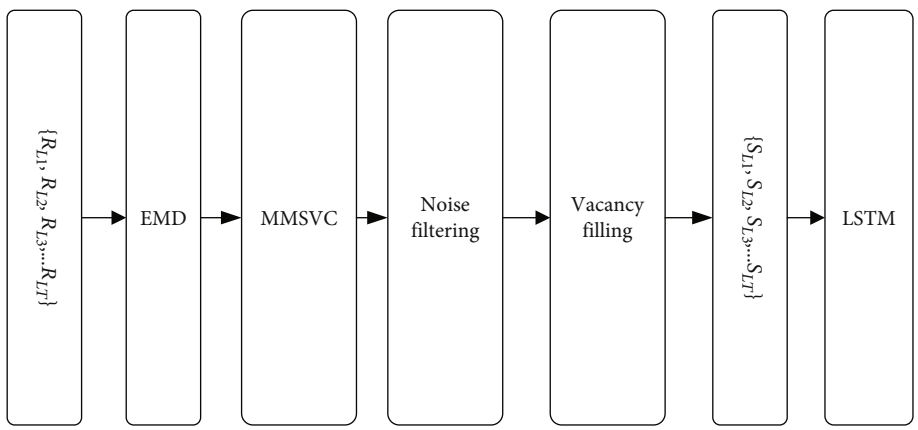


Figure 2: EMD-MMSVC-LSTM model.

the prediction is. EMD filtering is a new adaptive technology applied to minimize the risk of random noise. For conventional EMD filtering, the high-frequency IMFs will be completely discarded due to the presence of noise points, which not only consist of noise points but also include valid signals. This will result in severe signal distortion.

Because of the above problems, we propose a prediction algorithm combining EMD and MMSVC with LSTM. Firstly, we changed the signal from time domain to frequency domain in order to separate high-frequency and noisy IMFs. Secondly, we expurgated noise points under the condition that kept the valid signal. Eventually, we utilized LSTM to predict denoised data.

While most conventional EMD-based prediction methods predict individual IMFs directly, the proposed approach performs noise reduction on IMFs, which can significantly reduce the impact of noise on prediction. Noisy IMF identification using MMSVC avoids selecting thresholds of different permutation entropies. In order to locate and delete noise points, the signal is divided into groups. Outliers in each group are regarded as noise points. Then, refill them by a certain weight with relevant figures. Compared with other methods, this approach has the advantages of no special requirements for the signal itself and easy to operate.

$$\begin{pmatrix} x_1, x_2, x_3 \cdots x_n \\ y_1, y_2, y_3 \cdots y_n \\ \cdots \\ m_1, m_2, m_3 \cdots m_n \end{pmatrix}. \quad (1)$$

2. Related Work

The typical methods for network security situation prediction include Regression Analysis [3], Grey Theory prediction model [4], and Artificial Neural Network. These algorithms exhibit high performance in some particular applications, but they also have their limitations. For instance, the Regression Analysis lacks real time because it describes the regular by mathematical formulas. Various contingencies occurring in the network frequently lead Regression Analysis to become more inefficient. The Grey Theory prediction model has remarkable effects on small data sets, whereas, the accuracy of forecasting is considerably lower than ANN. The

```
noise_filter(data, n): /*parameter n refers to the packet size*/
    length <- length of data
    /*group data*/
    group[ ] <- each group contains n consecutive signals
    /*filter noise*/
    noise_filter_data = []
    for goup_data in the group do
        low_num <- the factor with the ranking percentage of B%
        high_num <- the factor with the ranking percentage of A%
         temp_data = noise_filter_and_replace(goup_data, low_num, high_num)
         noise_filter_data.append(tem_data) /*Merge all denoised groups*/
return noise_filter_data

/*delete outliers and fill the blank places*/
noise_filter_and_replace(data, low_num, high_num):
    /*If the noise points exist in the first position or the last position, they need to be processed separately*/
    for index <-1 to len(data-1) do
        if data[index] is noise point then
            Find the nearest two non noise points data[first_weight] and data[second_weight] to the left and right
            Calculate the distance x, y to the point
            first_weight <- y / (x + y)
            second_weight <- x / (x + y)
            data[index] <- data[first_weight] * first_weight + data[second_weight] * second_weight
return data
```

Algorithm 1: Noise reduction.

availability for network security situation prediction using ANN has been verified by many researches. For example, Liu et al. [5] proposed a method that applies GM (1,1) model and BP neural network model. In article [6], a network security situation prediction method based on BP neural network optimized by Seeker Optimization Algorithm (SOA) was proposed. The effectiveness of Artificial Neutral Network for network situation prediction can be effectively verified. LSTM is applied in this experiment, which has better performance than Recurrent Neural Network (RNN) in a long sequence [7]. LSTM can avoid the disturbance of the information attenuation due to a long time. However, data acquisition will produce random noise inevitably, which would make the learning of parameters deviate and give rise to decreased accuracy [8–10].

Therefore, it is vital to provide a significant boost in forecasting performance by filtering noise. In recent years, there has been an increasing interest in exploring methods for noise reduction, such as using filters, wavelet transform (WT) [11], EMD [12], and Fourier transform [13].

In article [14], a signal-filtering method based on empirical mode decomposition is proposed. Compared to well-known filtering methods, this method is a fully data-driven approach without too much human intervention.

In article [15], the authors present a comprehensive study of high-G calibration denoising method. They proposed a denoising method based on the combination of empirical mode decomposition (EMD) and wavelet threshold. They utilized wavelet threshold to processed IMFs, in which the filtering depends on the selection of decomposition number and basis function.

In article [16], a novel noise reduction technique for underwater acoustic signals based on complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), minimum mean square variance criterion (MMSVC), and least mean square adaptive filter (LMSAF) is proposed by Li and Wang. In their scheme, LMSAF was utilized for high-frequency IMF noise reduction, which addresses the problem of selection of decomposition number and basis function for wavelet noise reduction. However, LMSAF requires linear independence of input vectors at different times. The correlation of input signal will cause repeated error propagation, slow convergence speed, and poor tracking performance.

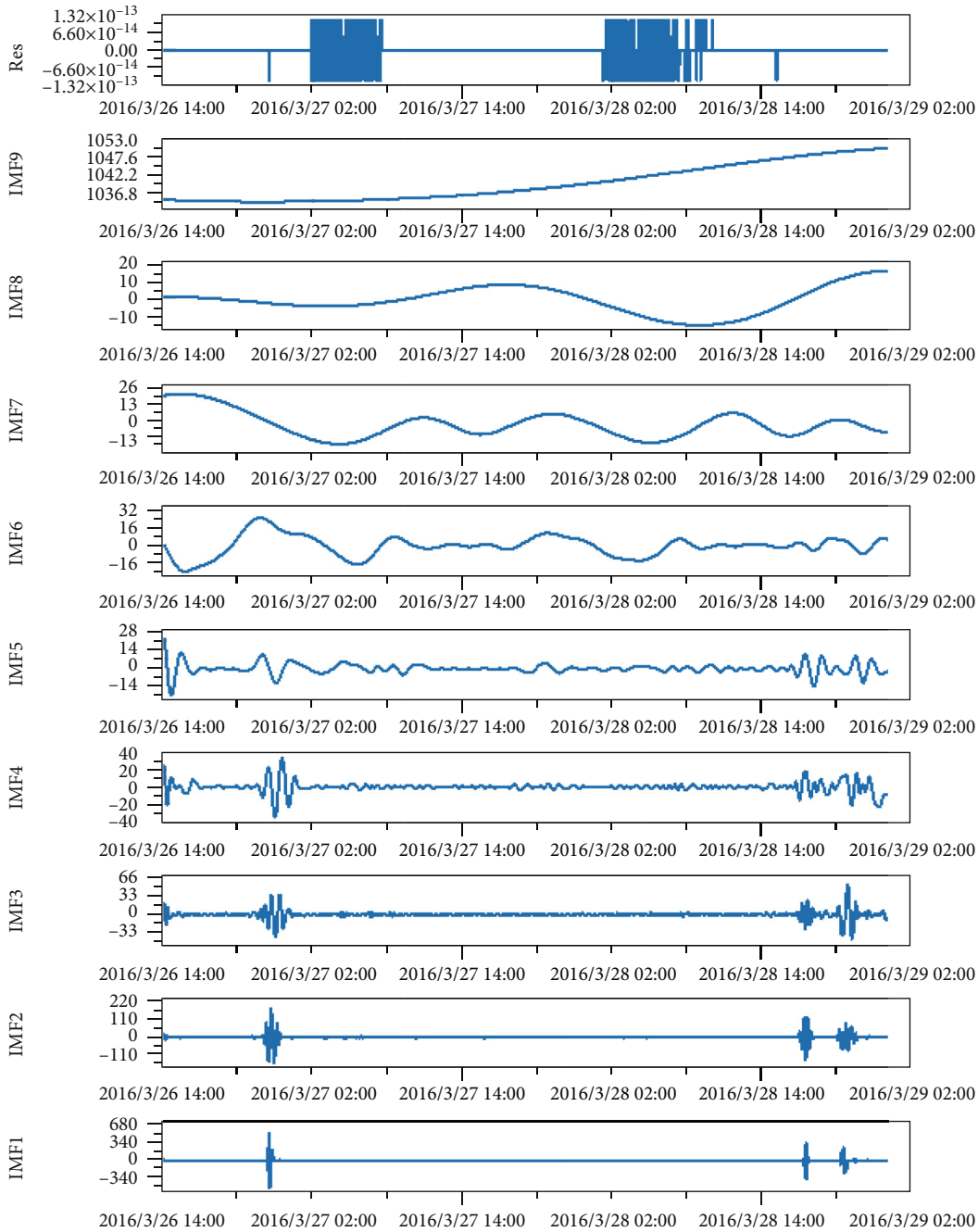
Here, we aim to address the issue of noise by using a technique, which adapted from combining EMD with LSTM.

3. Proposed Model

3.1. Principle of EMD. The greatest merit of EMD is it does not require too much human intervention. EMD is very simple and convenient for signal adaptive decomposition [17, 18].

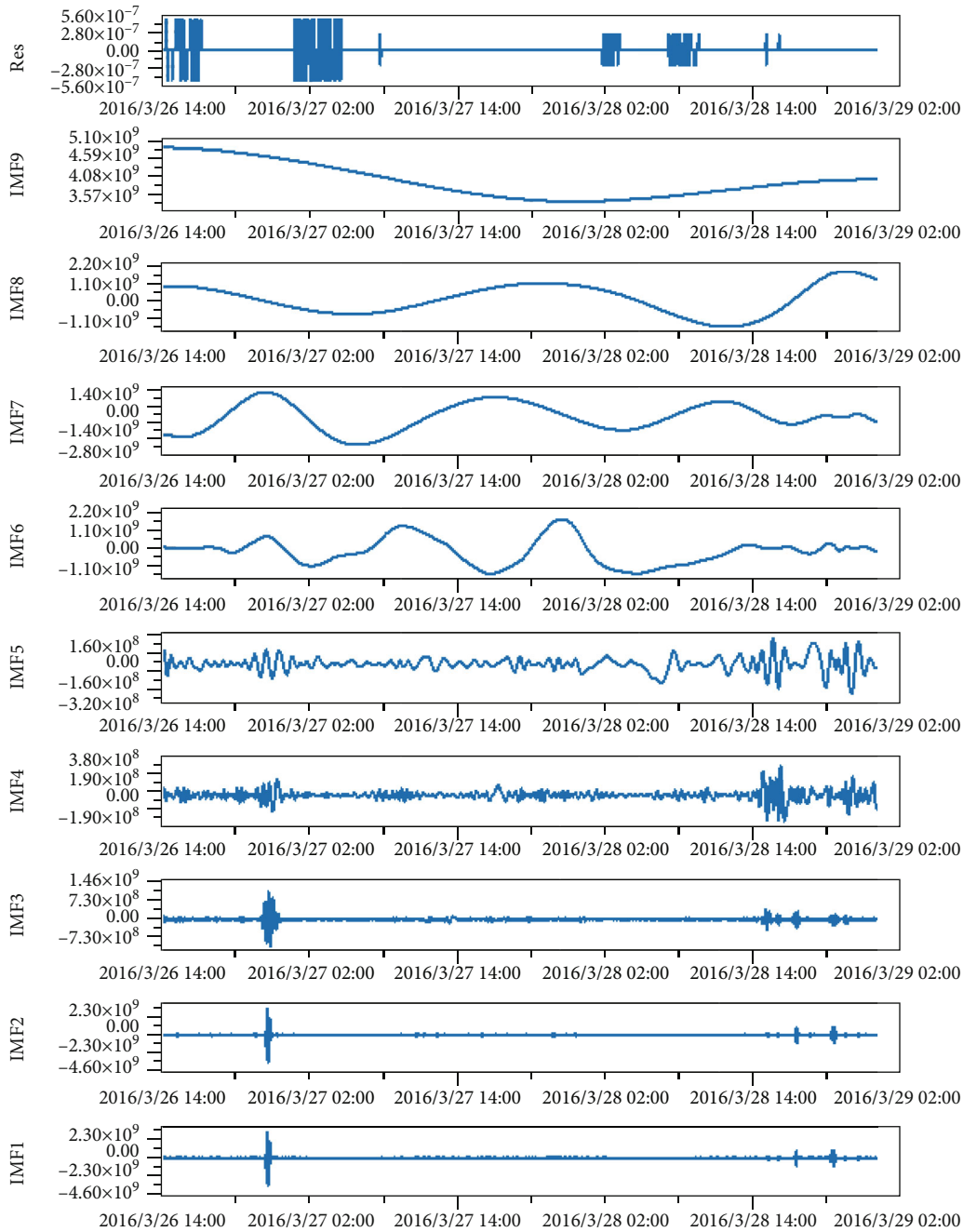
Any signals can be decomposed into IMFs in different frequency domains. Different features of time scales in the historical sequence sets are involved in different IMFs. The detailed procedure of EMD consists of six steps, which are described as follows:

Step 1. All the local maximums and minimums of the original signal $x(t)$ will be determined and joined as upper



(a)

Figure 3: Continued.

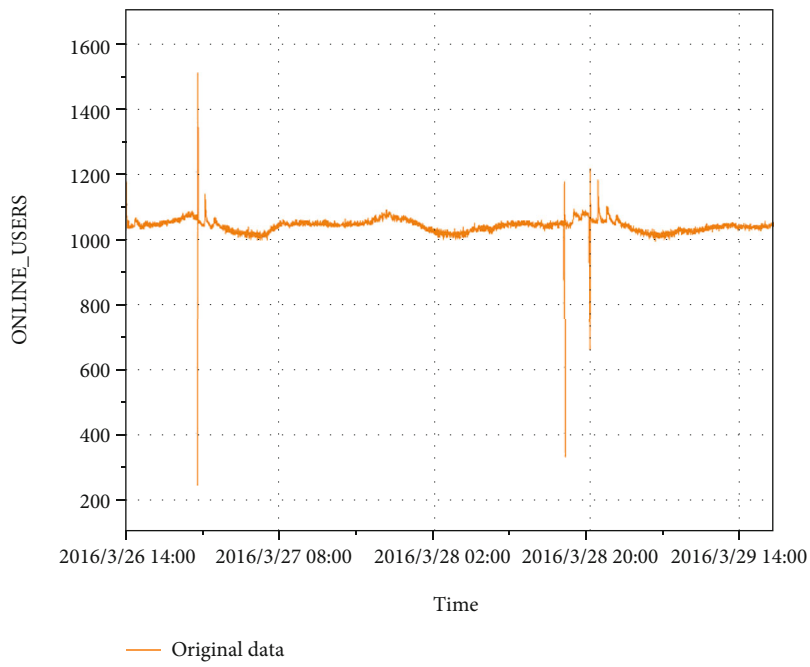


(b)

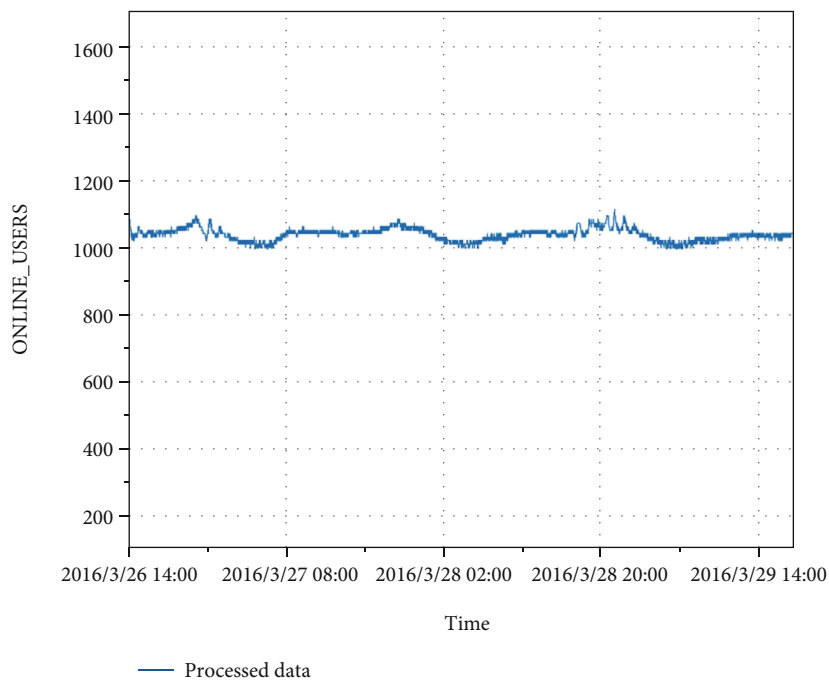
Figure 3: Decomposition of signal into IMFs: (a) ONLINE_ USERS IMFs and residual; (b) IP_INBPS IMFs and residual.

Table 1: Mean square variances of two adjacent IMFs.

	$M(1)$	$M(2)$	$M(3)$	$M(4)$	$M(5)$	$M(6)$	$M(7)$	$M(8)$
ONLINE_USERS	1970.1	560.5	181.4	52.8	56.5	34.8	117.6	159.8
IP_INBPS	$6.6E+16$	$2.7E+16$	$9.8+15$	$4.9E+15$	$4.8E+17$	$2.8E+18$	$1.6E+18$	$1.4E9$

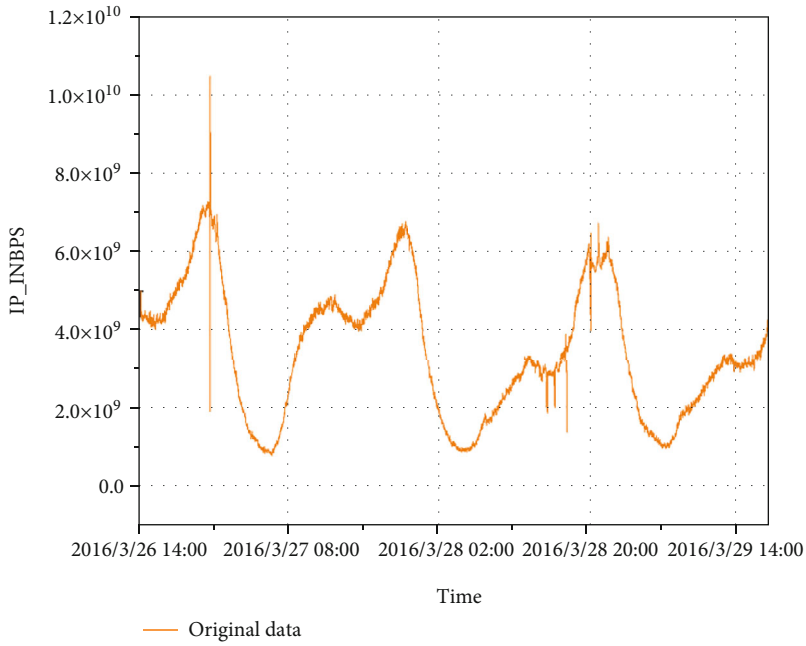


(a)

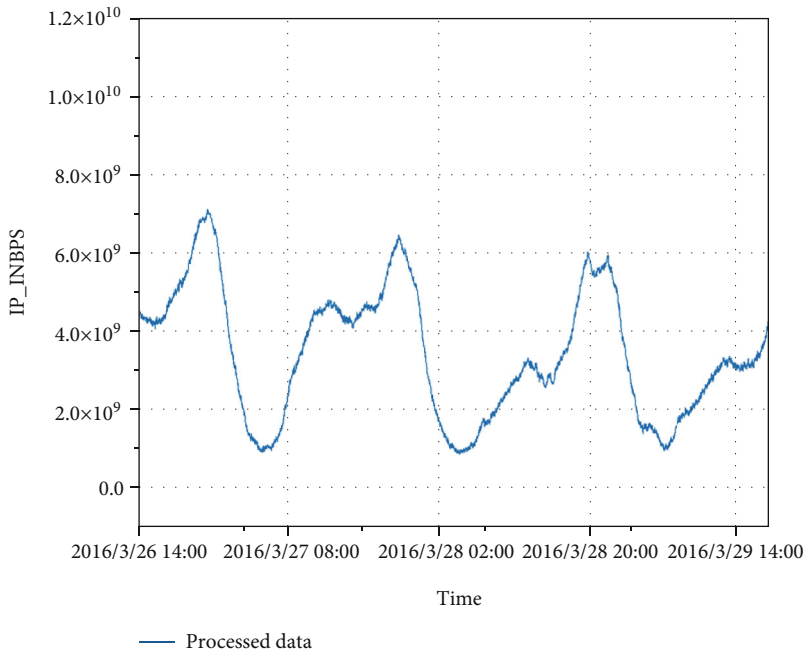


(b)

Figure 4: Continued.

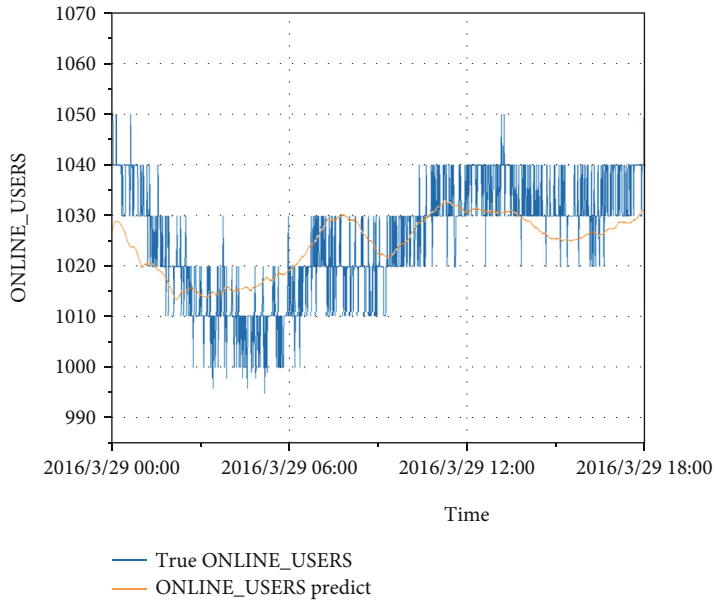


(c)

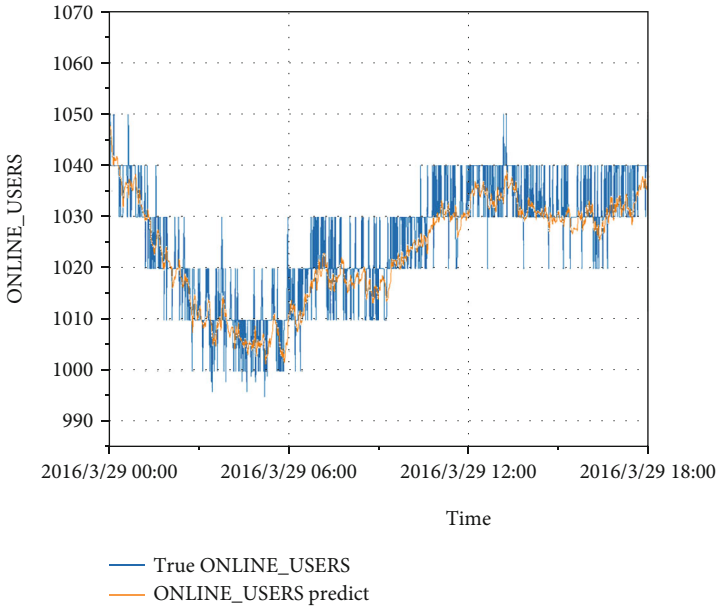


(d)

Figure 4: Comparison of noise reduction: (a) raw data of ONLINE_USERS; (b) denoised signal of ONLINE_USERS using the EMD method; (c) raw data of IP_INBPS; (d) denoised signal of IP_INBPS using the EMD method.

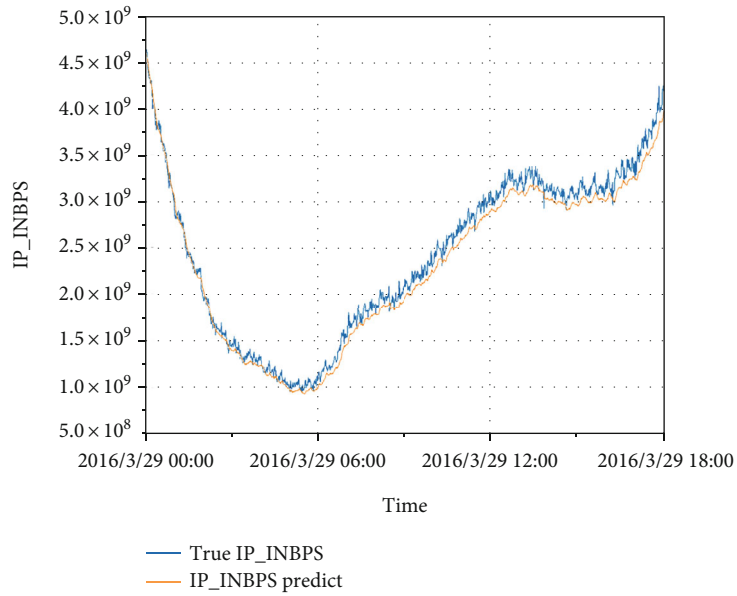


(a)

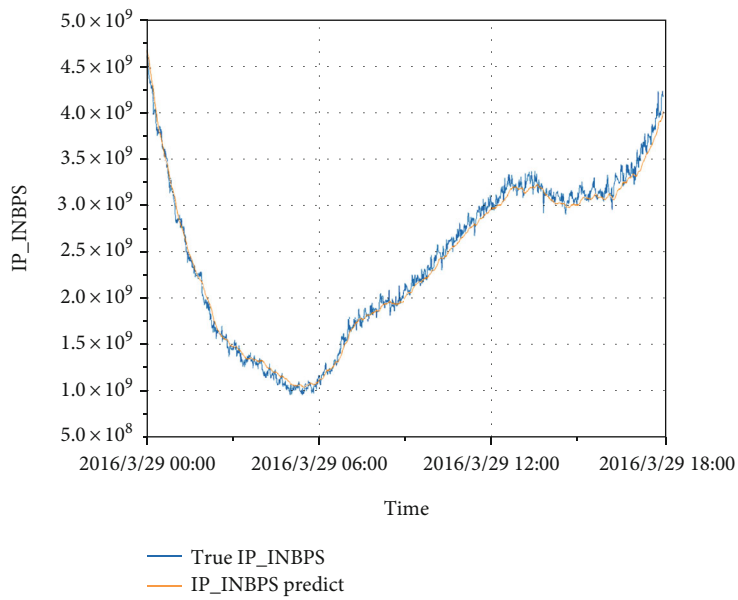


(b)

Figure 5: Continued.

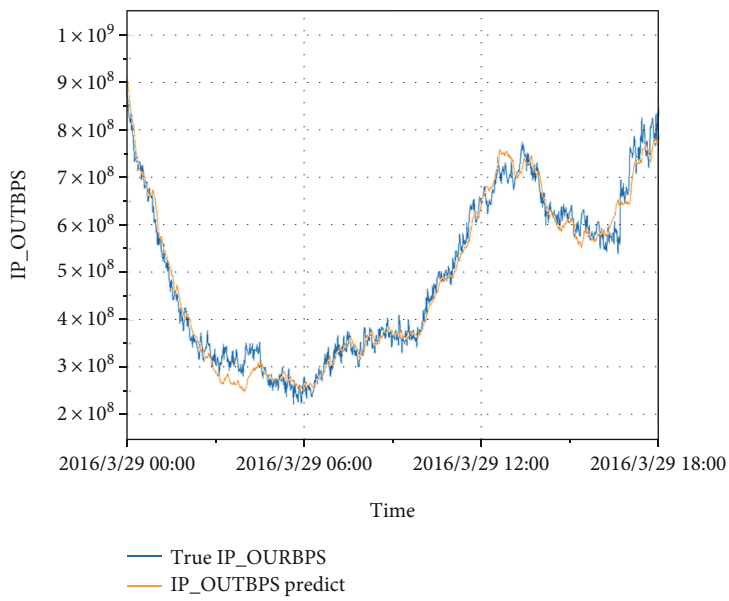


(c)

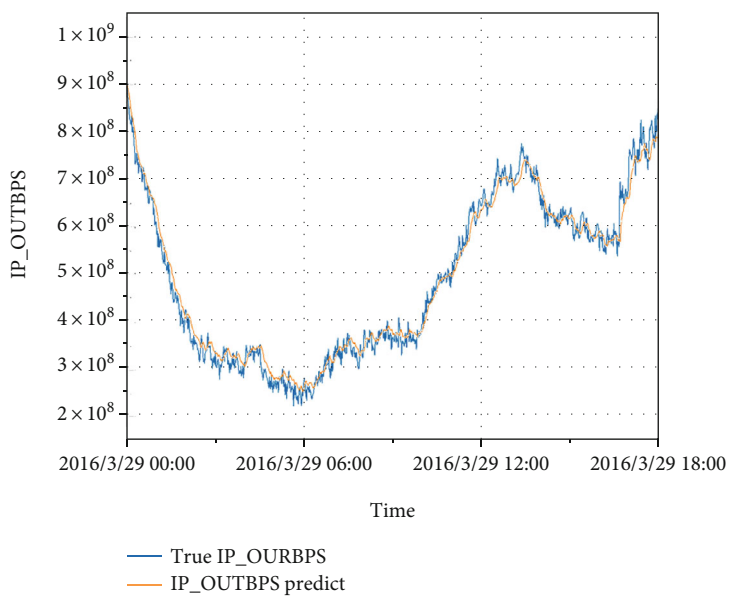


(d)

Figure 5: Continued.

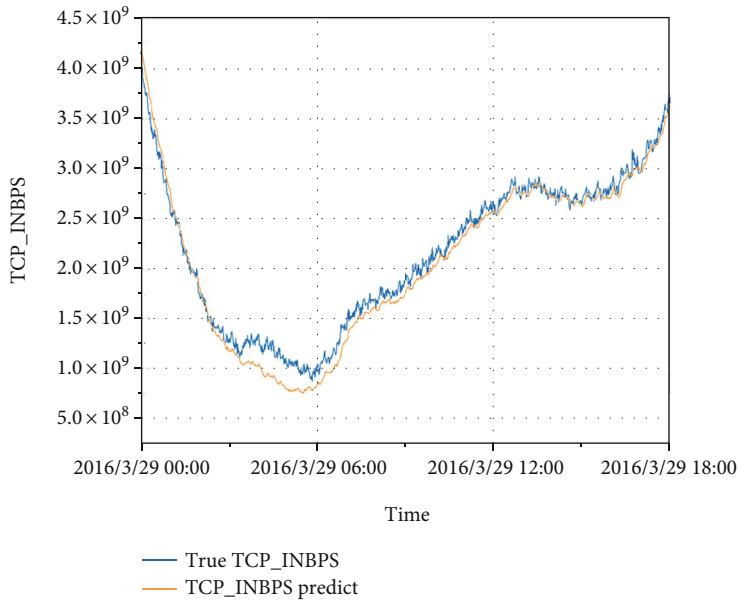


(e)

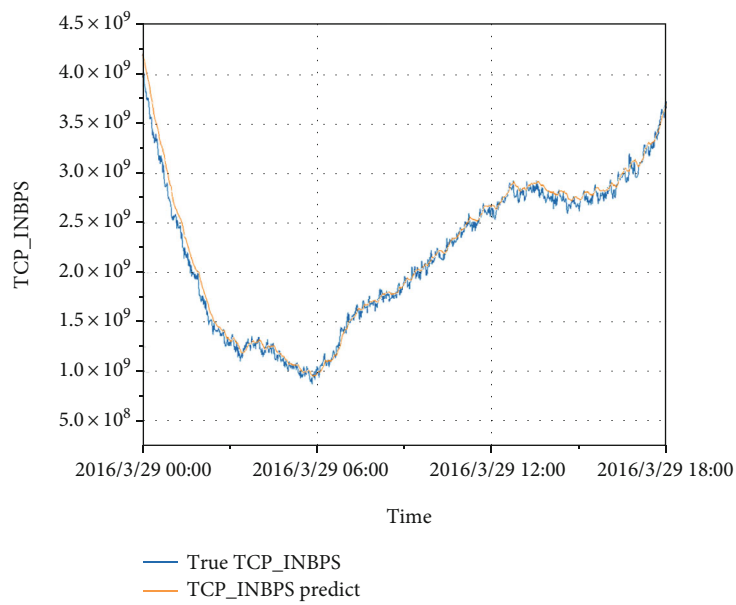


(f)

Figure 5: Continued.



(g)



(h)

Figure 5: Comparisons of prediction between raw data and denoised data: (a) prediction of raw ONLINE_USERS; (b) prediction of denoised ONLINE_USERS; (c) prediction of raw IP_INBPS; (d) prediction of denoised IP_INBPS; (e) prediction of raw IP_OUTBPS; (f) prediction of denoised IP_OUTBPS; (g) prediction of raw TCP_INBPS; (h) prediction of denoised TCP_INBPS.

Table 2: Performance comparison based on EMD against traditional LSTM models for forecasting.

	ONLINE_USERS		IP_INBPS		IP_OUTBPS		TCP_INBPS	
	LSTM	EMD-MMSVC-LSTM	LSTM	EMD-MMSVC-LSTM	LSTM	EMD-MMSVC-LSTM	LSTM	EMD-MMSVC-LSTM
MSE	79.35	39.68	$1.68E+16$	$8.63E+15$	$7.57E+14$	$4.50E+14$	$1.70E+16$	$6.92E+15$
RMSE	8.91	6.30	$1.30E+08$	$9.29E+07$	$2.75E+07$	$2.22E+07$	$1.30E+08$	$8.32E+07$
MAE	7.32	5.02	$1.09E+08$	$7.28E+07$	$2.13E+07$	$1.72E+07$	$1.04E+08$	$6.19E+07$
Average accuracy	+37%		+37%		+28%		+45%	

envelope $S_+(t)$ and lower envelope $S_-(t)$ by using the cubic spline difference.

Step 2. Calculate the average of $S_+(t)$ and $S_-(t)$ which is $m(t)$.

Step 3. The mean value $m(t)$ is subtracted from the original sequence to obtain the intermediate signal $h(t)$, $h(t) = x(t) - m(t)$.

Step 4. Judge whether the two conditions for becoming the IMF are met: the difference between the number of zero points and the number of extreme points cannot be more than one; meanwhile, the mean of the upper envelope and lower envelope must be zero at any time. If it satisfies, $h(t)$ will be considered as an IMF and perform Step 5; otherwise, repeat Step 1 to Step 4.

Step 5. Calculate the residual $r(t) = x(t) - h(t)$ and regard $r(t)$ as new $x(t)$. Then, repeat Step 1 to Step 4 until the signal will not be decomposed.

Last but not least, the decomposition of the original signal can be expressed as Equation (2), where n is determined by $x(t)$ and r is the residual signal.

$$x(t) = \sum_{i=1}^{n}(\mathrm{imf}_n) + r. \tag{2}$$

3.2. Principle of MMSVC. MMSVC is utilized to identify noisy IMFs decomposed by EMD [16]. The detailed procedure of MMSVC consists of three steps, which are described as follows:

Step 1. Original signal $x(t)$ removes the first n IMFs to construct $\mathrm{new}(n)$, $\mathrm{new}(n) = x(t) - \sum_{i=1}^{n}\mathrm{imf}_i$, $n = 1, 2, 3, \cdots, K$, where K is the total number of IMFs.

Step 2. Calculate square variance $M(n)$ of $\mathrm{new}(n)$ and $\mathrm{new}(n+1)$: $M(n) = \mathrm{MSE}(\mathrm{new}(n), \mathrm{new}(n+1))$.

Step 3. Repeat Step 2 to calculate any two adjacent IMFs, and find minimum mean square variance, $\min_{1<n<K}[\mathrm{MSE}(\mathrm{new}(n), \mathrm{new}(n+1))]$.

The first $n-1$ IMFs are considered as noisy IMFs while $\mathrm{MSE}(\mathrm{new}(n), \mathrm{new}(n+1))$ is the least value.

3.3. Principle of LSTM. LSTM was first proposed by Hochreiter and Schmidhuber [19] in 1997. From the internal structure, LSTM is adept at dealing with the time-series question [20, 21]. The specific structure of LSTM is depicted in Figure 1, where is composed of the updated state value $\widetilde{C_t}$ at time t, output signal (hidden state) h_t, sigmoid neural network layer σ, tanh function, and tanh neural network layer.

The forget gate is used to simulate the forgetting process by controlling the forgetting degree of the last memory cell C_{t-1} with a weight matrix w_f. Input signal x_t and h_{t-1} into sigmoid function, afterwards, output f_t with the value ranging 0~1. If the result is 0, it indicates that information is vanished completely, whereas, the whole data will be stored with the consequence of 1. The specific calculation procedure is as follows:

$$f_t = \sigma\left(w_f.[h_{t-1}, x_t] + b_f\right). \tag{3}$$

The input gate acts on determining how much information will be saved to the cell state C_t, which can effectively avoid the memory of irrelevant content. Sigmoid function and tanh function, both known as active function, are expressed by Equations (4)–(6) in detail, where w_i, and w_c are weight matrixes and b_i and b_c are bias terms. Add the information that needs to be forgotten and the information that needs to be remembered to update the cell state.

$$i_t = \sigma(w_i.[h_{t-1}, x_t] + b_i), \tag{4}$$

$$\widetilde{C_t} = \tanh\,(w_c.[h_{t-1}, x_t] + b_c), \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t}. \tag{6}$$

The output gate operates on controlling how much information from the cell state C_t will be outputted to h_t, which is the input signal of the next time. Summarizing, Equations (7) and (8) briefly describe the operations performed by an output gate.

$$o_t = \sigma(w_o.[h_{t-1}, x_t] + b_o), \tag{7}$$

$$h_t = o_t * \tanh\,(C_t). \tag{8}$$

3.4. EMD-MMSVC-LSTM Model. This algorithm fully exploits the properties of EMD, MMSVC, and LSTM. We firstly identify the noisy IMFs using MMSVC and delete several outliers in noisy IMFs. Secondly, for the blanks, we fill them by a certain weight, which is determined by the distance between the blank and adjacent reference data. Finally, the LSTM model is applied to predict denoised signals. The specific process of the algorithm is shown in Figure 2, where $\{R_{L1}, R_{L2}, R_{L3}, \cdots R_{LT}\}$ is regarded as the original time series and $\{S_{L1}, S_{L2}, S_{L3}, \cdots S_{LT}\}$ is identified as a new sequence.

The signal is decomposed into IMFs with diverse frequencies. The randomness of the signal decreases gradually. By deleting outliers in the high-frequency and noisy IMFs, the purpose of filtering noise and prediction with precision can be achieved. The specific procedures are described as follows:

Step 1. Signal decomposition. Apply EMD to decompose the network characteristic $R(t)$ to obtain n IMFs in different frequency domains and one residual.

Step 2. Utilize MMSVC to obtain noisy IMFs.

Step 3. Filter outliers in noisy IMFs. Take IMF1 for an example; to begin with, divide IMF1 into m groups according to period t. After that, sort the t sample points of each group

according to ascending order. Then, find the factor $a_i(i = 1, 2, 3 \cdots \cdots m)$ with the ranking percentage of A% and the factor $(b_i(i = 1, 2, 3 \cdots \cdots m))$ with the ranking percentage of B%. Ultimately, filter the data less than b_i and greater than a_i in each group. (The selection of A and B is determined by the data itself. Through continuous experiments, it is found that the effect of this group of data is the best when A is 90 and B is 10.)

Step 4. Fill in the blank places. The vacant parts are filled by a certain weight with relevant figures, and only nonnoise points can be used as reference data. For the noise point t_i, $t_{i\text{-}1}$ and t_{i+1} must be valid data; otherwise, we need to search for valid data nearby. If the distance between t_i and t_{i+1} is x and the distance between t_i and $t_{i\text{-}1}$ is y, then $t_i = xt_{i-1}/(x + y) + yt_{i+1}/(x + y)$. The pseudocode of Step 2 and Step 3 is represented in Algorithm 1 below.

Step 5. Reconstruct signal. Add IMFs and residual R_n item by item to be a new time series $S(t)$.

Step 6. Signal prediction. LSTM is utilized to predict the new time series.

4. Experiment and Verification

4.1. Data Set. The data in this paper is extracted from the traffic statistics logs collected from the CERNET campus network, which is composed of various indexes known as ONLINE_USERS, IP_INBPS, IP_OUTBPS, TCP_INBPS, and TCP_OUTBPS. The time range is from 14:00 on March 26, 2016, to 18:00 on March 29, 2016.

4.2. Noise Filtering Based on EMD. It is acknowledged that the noise in the original data will inevitably reduce the accuracy of the prediction model. As Figure 3 shows, ONLINE_ USERS and IP_INBPS are both decomposed into nine IMFs and one residual by EMD. Through the spectrum analysis of IMFs, the frequency of IMFs decreases step by step.

MMSVC is used to identify noisy IMFs of two indexes. The result is depicted in Table 1. As Table 1 shows, $M(6)$ is the minimum value for ONLINE_ USERS and $M(4)$ is the minimum value for IP_INBPS. Therefore, for ONLINE_USERS, the first five IMFs are noisy IMFs. And for IP_INBPS, the first three IMFs require noise reduction.

Process noisy IMFs. First, group the data according to the period. Second, filter outliers with the ranking percentile higher than 90% and the ranking percentile lower than 10% in each IMF. Third, fill the blanks by a certain weight with relevant figures. As Figure 4 shows, the noise is suppressed to a certain extent and the main image features of the original data are retained.

4.3. Experimental Results of EMD-MMSV-LSTM Model. Predict the two characteristics using the LSTM model. The data from 26th to 28th were regarded as training data, and the remaining data were used as the test data. For the conventional LSTM model, to a certain degree, the prediction of the current time depends on the value of the previous time, which leads to inefficiency readily. By converting the forecast results to rely on real data, the prediction will be effectively improved.

To prove the validity of the experiment, features named IP_OUTBPS and TCP_INBPS were involved. Figure 5 presents the comparisons of prediction between raw data described and denoised data described, in which the blue polyline is the observed values and the orange polyline is the predicted values. MSE, RMSE, and MAE were used as evaluation indexes in this experiment. See Table 2 for specific data.

The results of this experiment are summarized that the noise can be suppressed to a certain extent by EMD filtering; in the meantime, the main image features have almost remained. The utilization of filtering noise by EMD could provide a significant boost in LSTM forecasting performance.

5. Conclusion

In this study, we propose a new forecasting model known as EMD-MMSVC-LSTM. Compared with the most EMD-based prediction model, the proposed method can significantly reduce the impact of noise on prediction. MMSVC is employed to identify noisy IMFs without the selection of thresholds. After that, delete the outliers in each group. This approach of noise reduction has the advantages of no special requirements for the signal itself and easy to operate. Eventually, make prediction for the denoise signal.

Although the method of noise reduction has been realized in this paper, the filtering conditions are relatively unitary and the filtering process is nonadaptive, which would give rise to insufficient filtering. How to improve the effectiveness and adaptability of noise filtering is the direction of further research.

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 is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China, No. 2020YFB1711000 and No. 2018YFB08040505.

References

[1] Z. Haofang, K. Chunying, and X. Yao, "Research on network security situation awareness based on the LSTM-DT model," *Sensors*, vol. 21, no. 14, p. 4788, 2021.

[2] G. Wang, "Comparative study on different neural networks for network security situation prediction," *Security and Privacy*, vol. 4, no. 1, 2021.

[3] X. Wei-wei and W. Hai-feng, "Prediction model of network security situation based on regression analysis," in *2010 IEEE International Conference on Wireless Communications, Networking and Information Security*, pp. 616–619, Beijing, China, 2010.

[4] Q. Yu and Y. Shen, "Research of information security risk prediction based on grey theory and ANP," in *IEEE advanced information management, communicates, Electronic and Automation Control Conference (IMCEC)*, vol. 2016, pp. 107–113, Xi'an, China, 2016.

[5] L. Nian, L. Geng, and L. Yong, "A method of network security situation prediction based on gray neural network model," *Applied Mechanics and Materials*, vol. 63, pp. 936–939, 2011.

[6] R. Zhang, M. Liu, Q. Zhang, and Z. Cai, "A network security situation prediction algorithm based on BP neural network optimized by SOA," *In International Conference on Artificial Intelligence and Security*, vol. 12240, 2020.

[7] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting-a novel pooling deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.

[8] S. Xiangbo, Z. Liyan, S. Yunlian, and J. Tang, "Host-parasite: graph LSTM-in-LSTM for group activity recognition," *IEEE Transaction on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 663–674, 2021.

[9] Y. Kyuhwan, M. Kyunghan, S. Jaewook, M. Sunwoo, and M. Han, "Ego-vehicle speed prediction using a long short-term memory based recurrent neural network," *International Journal of Automotive Technology*, vol. 20, no. 4, pp. 713–722, 2019.

[10] G. Klaus, K. Srivastava Rupesh, K. Jan, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE Transaction on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[11] P. M. Ramos and I. Ruisánchez, "Noise and background removal in Raman spectra of ancient pigments using wavelet transform," *Journal of Raman Spectroscopy*, vol. 36, no. 9, pp. 848–856, 2005.

[12] L. Wenqing, Z. Laibin, L. Wei, and X. Yu, "Research on a small-noise reduction method based on EMD and its application in pipeline leakage detection," *Journal of Loss Prevention in the Process Industries*, vol. 41, pp. 282–293, 2016.

[13] M. Grédiac, F. Sur, and B. Blaysat, "Removing quasi-periodic noise in strain maps by filtering in the Fourier domain," *Experimental Techniques*, vol. 40, no. 3, pp. 959–971, 2016.

[14] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 6, pp. 2196–2202, 2007.

[15] Q. Lu, L. Pang, H. Huang et al., "High-G calibration denoising method for high-G MEMS accelerometer based on EMD and wavelet threshold," *Micromachines*, vol. 10, no. 2, p. 134, 2019.

[16] Y.-x. Li and L. Wang, "A novel noise reduction technique for underwater acoustic signals based on complete ensemble empirical mode decomposition with adaptive noise, minimum mean square variance criterion and least mean square adaptive filter," *Defence Technology*, vol. 16, no. 3, pp. 543–554, 2020.

[17] Y.-X. Bai, T.-T. Lin, and Z.-C. Zhong, "Noise reduction method of Φ-OTDR system based on EMD-TFPF algorithm," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24084–24089, 2021.

[18] Z. Jin, F. Fan, M.-P. Pere et al., "Serial-EMD: fast empirical mode decomposition method for multi-dimensional signals based on serialization," *Information Sciences*, vol. 581, pp. 215–232, 2021.

[19] H. Sepp and S. Jürgen, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[20] Y. Baek and H. Y. Kim, "ModAugNet: a new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Expert Systems with Applications*, vol. 113, pp. 457–480, 2018.

[21] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, and A. Masoudi-Nejad, "DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks," *Bioinformatics*, vol. 36, no. 17, pp. 4633–4642, 2020.