Publisher: State and Provincial Joint Engineering Lab. of Advanced Network Monitoring and Control (ANMC)

Cooperate:

Xi'an Technological University (CHINA)
West Virginia University (USA)
Huddersfield University of UK (UK)
Missouri Western State University (USA)
James Cook University of Australia
National University of Singapore (Singapore)

Approval:

Library of Congress of the United States

Shaanxi provincial Bureau of press, Publication, Radio and Television

Address:

4525 Downs Drive, St. Joseph, MO64507, USA

No. 2 XueFu Road, WeiYang District, Xi’an, 710021, China

Telephone: +1-816-2715618 (USA) +86-29-86173290 (CHINA)

Website: www.ijanmc.org

E-mail: ijanmc@ijanmc.org

xxwlcn@163.com

ISSN: 2470-8038

Print No. (China): 61-94101

Publication Date: June 28, 2022

Associate Editors

Table of Contents

Research on Driving Conditions and Fuel Consumption of Improved K-means Clustering Algorithm

Shuping Xu
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China
E-mail: 563937848@qq.com

Xuanlv Wei
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China

Leyi Wang
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China
E-mail: 634877232@qq.com

Xiaodun Xiong
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China

***Abstract*—In order to solve the problem that the initial center of traditional clustering algorithm is easy to fall into local optimum and time-consuming. An improved combination optimization algorithm of principal component analysis and weighted K-means clustering is proposed. The algorithm introduces the maximum and minimum distance, weighted Euclidean distance, starting from the mean sum of the distances of the remaining clustering points, avoiding the influence of outliers and edge data. The proportion method is used to improve the principal component, and the characteristic influence factor obtained is used as the initial characteristic weight to construct a weighted Euclidean distance metric. According to the influence factors of feature contribution rate on clustering, a clustering method of feature weight influence factors is proposed. The representative feature factors are selected to highlight the clustering effect. Finally, the driving cycle of automobile is synthesized and the instantaneous fuel consumption is analyzed. The results show that: the difference value of speed acceleration joint distribution of the proposed method is only 1.05%, which saves 44.2% of the time compared with the traditional K-means clustering, and the driving cycle fitting degree is high, which can reflect the actual vehicle operation characteristics and fuel consumption.**

Keywords-Driving Cycle; Influence Factors; Feature Weight; Weighted K-Means Clustering

I. INTRODUCTION

The driving condition of a car is also called the operating cycle, which is the speed-time variation law of a vehicle in a specific environment. It is mainly used to evaluate vehicle pollutant emissions and energy consumption, and is of great value to the research and development of new vehicle models and risk assessment of traffic control [1]. Many scholars have conducted research on it, and Nguyen et al. [2] proposed a driving cycle construction process based on

DOI: 10.2478/ijanmc-2022-0011

Markov chain theory. Ding Yifeng et al. [3] used multivariate statistical methods such as principal component and cluster analysis to construct automobile road conditions. Liu Yingji et al. [4] used the characteristics of kinematics segment connection fuzzy to construct working conditions by combining principal components and fuzzy C-means clustering. Most scholars' research on driving cycle mainly focuses on the selection of K-means clustering initial center and single improved k-means clustering algorithm, but lack of research on principal component analysis and clustering combination optimization and execution time consumption. In order to achieve the ideal clustering effect and time consumption, it is still necessary to focus on the improvement of K-means clustering. Zhang Rui et al. [5] proposed OICCK-means algorithm in order to make up for the deficiency that the clustering effect of traditional K-means algorithm depends heavily on the initial clustering center. Zhang Lin et al. [6] adopted the idea of density to overcome the sensitive defect of traditional initial center. Luo Junfeng et al. [7] introduced information entropy and weighted distance to remove outliers. Zhang Yan [8] proposed an improved rough K-means clustering algorithm based on density weighting, which not only improves the clustering accuracy and reduces the number of iterations, but also weakens the interference of noise data and outliers on the results. However, the algorithm improves the clustering accuracy at the expense of efficiency cost. The algorithm puts most of the time consumption on the density of data objects, and the time complexity is too high.

Through the above analysis, this paper proposes an improved principal component analysis and improved K-means clustering combination optimization method, introduces the maximum and minimum clustering method and weighted Euclidean distance, and increases the weight of clustering eigenvalues according to the contribution factor. The results show that the clustering effect is stable, the time consumption is low, and the driving cycle constructed has strong applicability and meets the characteristics of traffic conditions.

II. ANALYSIS OF DRIVING CYCLE DATA

A. Data preprocessing

The data collected in this paper are the actual road driving conditions of a city light vehicle in September 2019 (sampling frequency is 1Hz), among which, the data information includes time, GPS speed measurement, longitude and latitude, instantaneous fuel consumption, etc. Using fitting interpolation method to interpolate and fit the disturbed discontinuous data, wavelet decomposition and reconstruction method to smooth the contaminated data [9] the original data was reduced from 194511 to 164039 by Matlab preprocessing

B. Feature parameter extraction and kinematic segmentation

Based on the analysis of relevant data and related research, 12 characteristic parameters are defined to describe the kinematic segments [10]. In this paper, 12 characteristic parameters including segment duration/T, travel distance/S, average speed/V_a, average driving speed/V_d, idle time ratio/ T_i, acceleration time ratio/ T_a, deceleration time ratio/T_d, cruise time ratio/T_c, speed standard deviation/V_{std}, average acceleration /a_a, average standard deviation of acceleration / a_{std}, average deceleration/a_d etc.

The interval from the start of one idling speed to the beginning of the next idling speed is called the kinematic segment [11]. This paper uses Python to develop related programs, uses stack and loop traversal data for processing, and divides

2445 kinematics segments from 164039 preprocessed data.

III. IMPROVED PRINCIPAL COMPONENT ANALYSIS

The traditional principal component uses linear technology to reduce the dimension of data, which eliminates the influence of order of magnitude and the difference information of each characteristic factor. In real life, the relationship between data is often nonlinear.

The comprehensive evaluation method with variance contribution rate as the weight can not reasonably explain the analysis results, and even the evaluation results deviate greatly from the facts [12]. Therefore, using the specific gravity method proposed in reference [13], the improved principal component can not only eliminate the dimension noise, but also can represent more feature parameter information and realize dimension reduction. The formula is as follows:

$$ZX_i = x_i / \sqrt{\sum_{i=1}^{n} x_{ij}^{\ 2}} \qquad (1)$$

In the case of dimension reduction, the improved principal component forms a matrix with the obtained number of data samples $(n)\times$ characteristic parameters (p), and select the principal component whose cumulative contribution rate reaches more than 80% for reduction and de-correlation. It can be seen from Figure 1 that the cumulative contribution rate of the first four principal components has reached 82.76%, which basically represents all the information of the 12 characteristic parameters of the fragment.

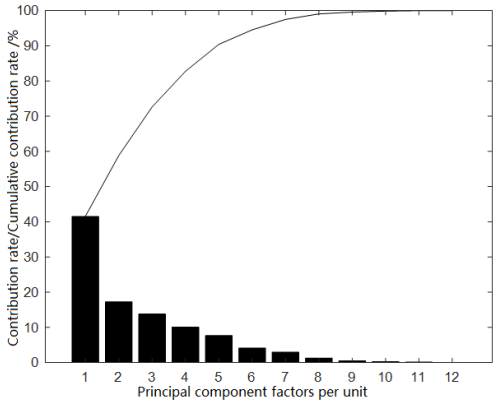


Figure 1. Contribution rate and cumulative contribution rate

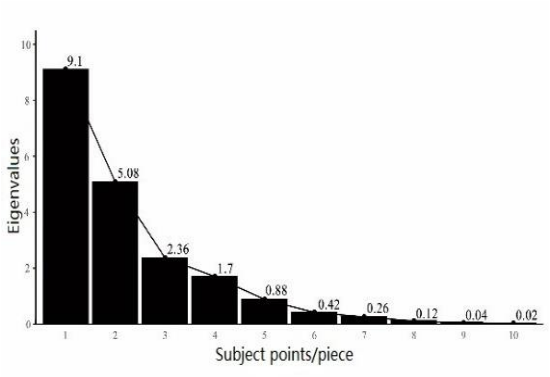


Figure 2. Gravel map

It can be seen from Figure 2 that each principal component is gradually decreasing, and there is an obvious inflection point in the change curve. It can be seen from Figure 1 that the first principal component contains 41.5% information in the improved principal component analysis results, so it meets the requirement that less principal components represent more information.

TABLE I. PRINCIPAL COMPONENT LOADING MATRIX

Characteristic parameter	M_1	M_2	M_3	M_4
Deceleration time ratio T_d	0.423	0.341	-0.723	0.248
Distance traveled S	0.893	0.134	0.045	0.432
Fragment duration T	0.432	0.231	-0.142	0.768
Acceleration time ratio T_a	0.394	-0.156	0.060	0.491
Cruise time ratio T_c	0.341	0.835	-0.045	-0.138

Average velocity V_a	0.499	0.763	0.025	0.255
Average driving speed V_d	0.778	0.315	0.112	0.358
Speed standard deviation V_{std}	0.198	0.033	0.034	0.189
Accelerate standard deviation a_{std}	0.145	0.267	-0.067	-0.121
Average acceleration a_a	0.014	0.223	0.033	0.024
Average deceleration a_d	0.566	-0.433	-0.052	0.315
Idle Time Ratio T_i	0.125	-0.351	0.843	0.467

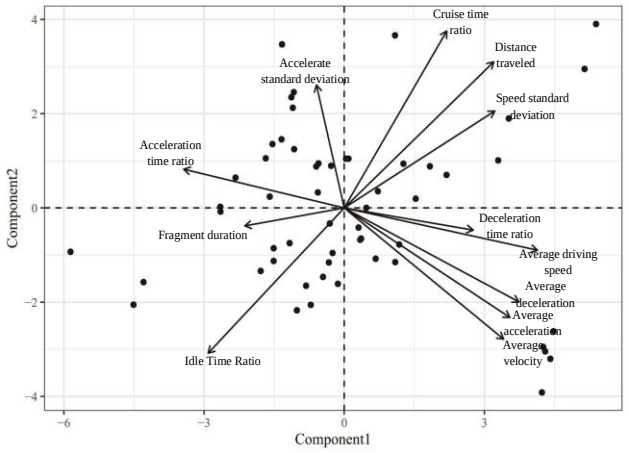


Figure 3. Principal component analysis scatter plot

When the absolute value of the principal component load factor of the selected parameter is larger, the correlation coefficient between a parameter and a principal component is higher [14]. From Figure 3, we can see the correlation of each eigenvalue directly. According to the above table 1, the first principal component eigenvalues are driving distance, average deceleration and average driving speed, and the correlation coefficients are 3.15, 2.08 and 3.69, respectively, so they have great correlation with driving distance and average driving speed; The second principal component eigenvalues have the average speed and cruise time ratio, and the correlation coefficients are 2.75 and 3.84 respectively, so they have a greater correlation with the cruise time ratio; the third principal component eigenvalues have the idle time ratio and deceleration time ratio, and the correlation coefficients are 3.06 and 2.85 respectively, so they have a greater correlation with the idle time ratio; The fourth principal component eigenvalue has fragment duration, and the correlation coefficient is 2.43, which indicates that it has a strong correlation with fragment duration. Through the analysis of IPCA, the first four principal components can reflect the characteristics of the original segment, and the 12 characteristic parameter matrices of the population sample are compressed into one eight characteristic parameter matrix which can represent the vast majority of sample information.

IV. IMPROVED K-MEANS CLUSTERING ANALYSIS

A. Outlier processing

The actual test species will have more or less interference, which often produces outliers or noises, which will affect the clustering effect. Here, we construct a residual point distance mean sum method to eliminate the influence of noise and outliers [15]. For the i point in the data, the sum of distances between each point and other points is S_i, and the sum of distances is H .When $S_i > H$, point i is regarded as an isolated point. Among them, the sample data is, the data dimension is, and the calculation is as follows:

$$S_i = \sum_{j-1}^{n} \sqrt{\sum_{h-1}^{d} (x_{ih} - x_{jh})^2} \tag{2}$$

$$H = \sum_{i-1}^{n} \frac{S_i}{n} \tag{3}$$

B. Maximum and minimum distance

1) The maximum and minimum distance of the remaining data in the cluster and dataset is defined as:

$$D_{\max} = \mathrm{Max}(d) \tag{4}$$

Among them, d is a set consisting of the minimum value of the distance between each cluster and the remaining data in the data set.

2) d_k is the minimum value of the distance between each cluster and the remaining data in the data set,

$$d_k = \mathrm{Min}(\sum_{k=1}^{m}(X_{ik} - X_{jk})^2) \tag{5}$$

Among them, X_i is the cluster center, X_j is the remaining data in the data set, and m is the dimension of the data.

3) Determine whether to select the initial candidate center as the optimized candidate center.

$$\mathrm{Max}(\mathrm{Min}(\mathrm{D}_t)) > \theta \left\| v_1 - v_2 \right\| \tag{6}$$

Among them, v_1, v_2 are the points that first become the candidate centers after optimization, θ is the parameter, which can be 0.5.

4) The criterion function of the $K-means$ algorithm for clustering is the error sum of square criterion function.

$$J_c = \sum_{i=1}^{k}\sum_{p\in C_i}(\left\| P - M_i \right\|)^2 \tag{7}$$

Among them, M_i is the mean value of all data in class C_i, p is each data in class C_i, and J_c is the function of sample and cluster center.

C. Weighted Euclidean distance

$\omega = [\omega_1, \omega_2, \cdots, \omega_n]^T \in R^{n\times d}$, The weight ω is introduced to distinguish the relationship between the sample data and the cluster center,

$$\sigma d_{\omega}(x_j, c_i) = \sqrt{\sum_{m=1}^{d}\omega_{jm}(x_{jm} - c_{im})^2} \tag{8}$$

$$\omega_{jm} = \frac{x_{jm}}{\frac{1}{n}\sum_{j=1}^{n}x_{jm}} \tag{9}$$

The initial new weight is as follows:

$$\mathrm{W_{iNew}} = \mathrm{W_i}(1 + \frac{\mathrm{A_{init}} - \mathrm{A_i}}{\mathrm{A_{init}}}) \tag{10}$$

Among them, the clustering accuracy is

$$\mathrm{A_i} = \frac{N_{cor}}{N}\% \tag{11}$$

Among them, $\omega_j = (\omega_{j1}, \omega_{j2}, \cdots, \omega_{jd})^T$ is d dimensional vector, x_{jm} is the m component of the j sample, $\frac{1}{n}\sum_{j=1}^{n}x_{jm}$ the average of the sum of the m component of each data object in the sample data set. It can be seen that ω is a weight that can reflect the overall distribution characteristics of the sample Value [5].

D. Feature weighted K-means clustering algorithm

1) By processing the noise and outliers, a new data set is obtained, and the related feature list is obtained.

2) The improved principal component analysis calculates the contribution factor of each feature to obtain the initial weight.

$$W=(W_{1X_1},W_{2X_2},\cdots,W_{nXn}) \quad (12)$$

3) The maximum minimum distance multi center clustering algorithm iteratively implements the proposed clustering center selection method to determine k *initial clustering centers.*

4) Based on the weighted features and the initial clustering center, K-means is executed to obtain K *clusters.*

5) Calculate the initial clustering accuracy A_{init}

6) For each feature in ω*, K-means clustering without the feature is performed, and the clustering accuracy is calculated;If* $A_i<A_{init}$ *, increase its weight* W_{iNew} *; otherwise remove the feature.*

7) Normalize the weights, perform K-Means clustering based on the new weights, and calculate the clustering accuracy A_{init}*;*

If $A_{final}>A_{init}$, accept the new weight and set $A_{init}=A_{final}$; otherwise, keep the old weight unchanged.

According to the above working condition data, the improved K-means algorithm is used for processing. First, edge data and outliers are detected, and abnormal points are eliminated. As shown in Figure 4 below, cluster 1 is a normal clustered point. Cluster 2 is the outlier of edge data. As can be seen in Figure 5, the edge data is relatively distant from most normal points, and most of the edge data are outliers, which can be eliminated.

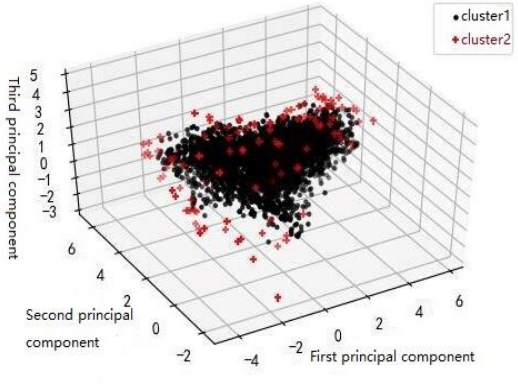


Figure 4. Scatter plot of edge data points of working conditions

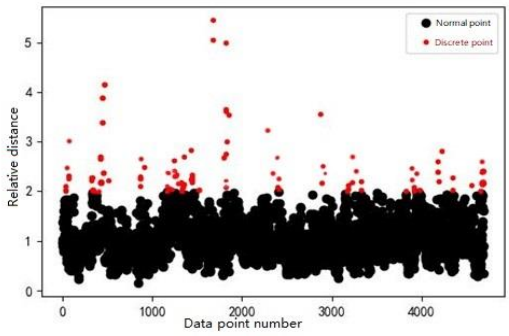


Figure 5. Relative distance comparison of outliers

According to the above-mentioned improved principal component analysis, the contribution factor and the characteristic value with high correlation are used to draw the three-dimensional graph, as shown in Figure 6. In this paper, the average speed, driving distance and cruise time ratio are selected to represent each point of clustering.

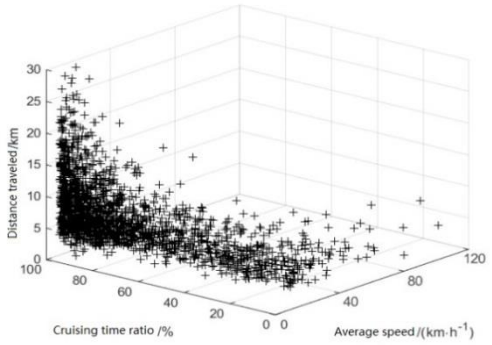


Figure 6. Three-dimensional scatter plot of working conditions

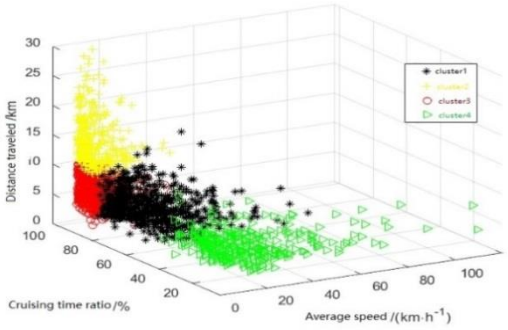


Figure 7. Working condition cluster analysis scatter plot

The improved K-Means clustering algorithm divides the kinematic segments into four categories, which are represented by cluster 1, cluster 2, cluster 3 and Cluster 4. It can be seen from Figure 7 that the first type is downtown area, where the vehicles start and stop frequently and the speed is low, and the average speed, cruise time ratio and driving distance are low; the second type is the living area, which is congested, with more start and stop times, and lower average speed, cruise time ratio and driving distance; the third type is suburban area, with smooth road conditions, less starting and stopping times, average speed, cruise time ratio and driving distance The fourth type is high-speed area, with smooth traffic, less start and stop times, high average speed, cruise time ratio and driving distance.

V. DRIVING CYCLE CONSTRUCTION AND FUEL CONSUMPTION ANALYSIS

A. Construction and verification of working conditions

According to the proportion of the total time of various time segments in the driving cycle of all data sets, the time taken by each driving cycle in the final construction cycle can be calculated [16]. This paper takes 1400s to construct vehicle driving cycle, as shown in Figure 8 below. The first type of low speed segment, the second type of medium speed segment, and the third type of medium high speed segment. The fourth type of high-speed video.

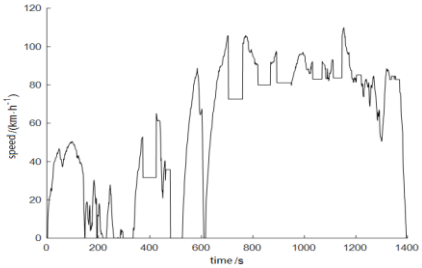


Figure 8. Synthetic driving conditions

From the speed and acceleration to verify the difference between the constructed driving cycle and the experimental data [11], this is a relatively standard verification method. Matlab software is used to calculate the speed acceleration joint distribution matrix of the vehicle driving cycle data, as shown in Figure 9.

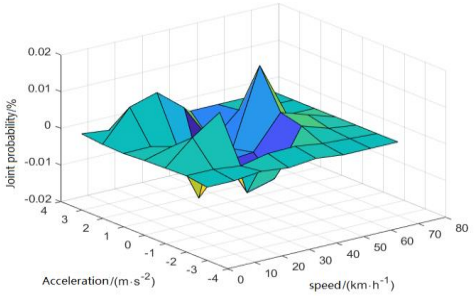


Figure 9. SAFD difference between experimental data and synthetic conditions

As can be seen from Figure 9 above, the joint velocity acceleration difference distribution of the experimental data and the improved clustering algorithm in this paper is within the ±1.2% range, and the calculated distribution difference value ($SAFD_{diff}$) is 1.05%，while the difference value ($SAFD_{diff}$) of the speed acceleration joint distribution between the experimental data and TKM is 0.97%. Therefore, the driving cycle constructed in this paper meets the driving characteristics of light vehicles, meets the development requirements of vehicle driving cycle construction, and has strong applicability.

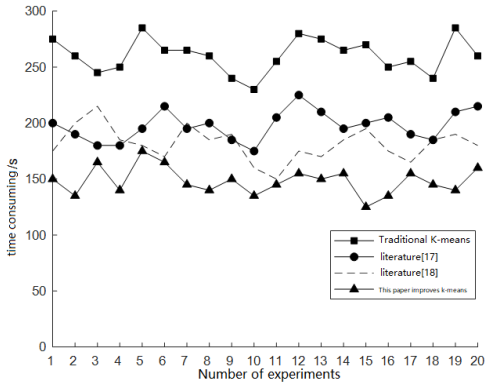


Figure 10. The results of the running time of the four methods

This paper uses the working condition construction method of literature [17] and literature [18]. According to the data of this paper, the improved principal component algorithm of this paper is combined with the four algorithms respectively, and 20 experiments are performed, as shown in Figure 10. The results show that, In this paper, the improved K-means clustering algorithm can not only weaken the influence of noise points on the initial center, but also greatly shorten the clustering time based on the stable clustering effect.

TABLE II. FOUR METHODS TO COMPARE THE RESULTS OF THE EXPERIMENT

Clustering method	The number of wrong samples	Average running time / s	Average accuracy /%	$SAFD_{diff}$ /%
k-means	184	260.5	89	1.98
Literature[17]	121	202.75	97	1.54
Literature[18]	98	181.5	99	1.25
The algorithm in this paper	101	145.25	98	1.05

The results of programming using Matlab are shown in Table 2 above. The comparison of the four algorithms in terms of the number of error-clustering samples, average running time, average correct rate and $SAFD_{diff}$, the improved K-means algorithm in this paper performs better in clustering performance and time consumption. The average running time is 44.2% less than that of traditional K-means clustering.

B. Fuel consumption analysis

As shown in figures 11 and 12, the instantaneous fuel consumption is large at low speed, medium and low speed, the torque fluctuation in the region is larger than that in the high speed region, the instantaneous fuel consumption rate in the high speed region is relatively stable, and the instantaneous fuel consumption rate in the low speed region and medium speed region is obviously increased. It can be observed in Figure 13 that the instantaneous fuel consumption increases briefly at low speed, and then the fluctuation trend is roughly consistent with the driving speed. As can be seen from Figure 14, the engine speed is mainly distributed in 1500-2500r / min under driving condition, and the opening of accelerator pedal is concentrated in 0.12-0.18, indicating that the driving condition is in medium high speed state.

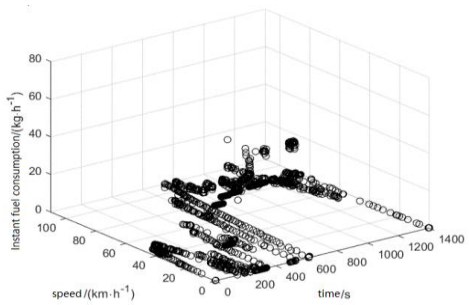


Figure 11. The relationship between driving time and speed instant fuel consumption

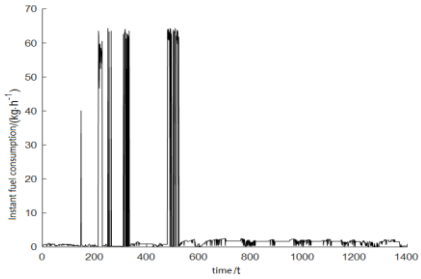


Figure 12. Relationship between driving time and instantaneous fuel consumption

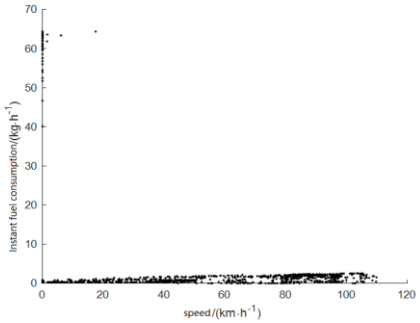


Figure 13. The relationship between driving speed and instantaneous fuel consumption

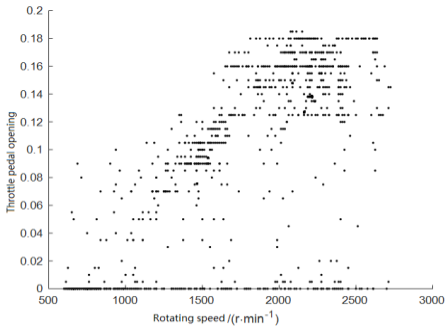


Figure 14. The relationship between driving speed and accelerator pedal opening

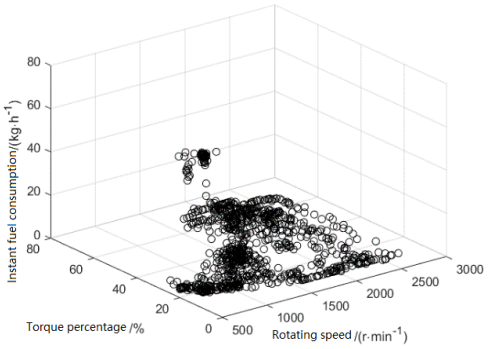


Figure 15. Instantaneous fuel consumption off for driving time and speed

It can be observed from Figure 15 above that the instantaneous fuel consumption is mostly concentrated in the speed of 1000-1500r / min, and the percentage of torque is 10% - 30%, which indicates that this part is composed of high-speed, medium speed and low-speed driving conditions. There are a few relatively concentrated areas in the speed range of 1500-2500r / min. it can be observed that this part is the instantaneous fuel consumption generated under the condition of high engine speed and low torque percentage, which may be due to driving It is caused by the extreme operation of the driver.

VI. CONCLUDING REMARKS

This paper proposes an improved optimization algorithm for the combination of principal components and feature-weighted K-means clustering, and introduces the residual point clustering mean method to eliminate outliers and reduce clustering time. The maximum minimum distance method can optimize the candidate initial centers, so that K-means avoids falling into the local optimal solution, so as to achieve a good clustering effect. According to the contribution rate of the eigenvalue contribution factor to the cluster, the initial feature weight is obtained, and a weighted Euclidean distance metric is proposed. Select characteristic values such as cruise time ratio, travel distance, average speed and so on with larger contribution factors, and then increase the weight to perform cluster analysis to construct vehicle driving conditions. The improved clustering algorithm proposed in

this paper still has room for improvement. The weighted density K-means clustering algorithm can be proposed on the basis of the algorithm in this paper. You can also consider directly removing outliers in the data preprocessing part of this paper to reduce the running time of subsequent clustering. You can also add more dimensional feature information.

ACKNOWLEDGMENT

The authors wish to thank the cooperators. This research is partially funded by the Project funds in Shaanxi province University Student Innovation and Entrepreneurship Fund Project (S S202010702115X) and the Project funds in engineering laboratory project (GSYSJ2018013).

REFERENCE

[1] Yuan Su-fen. Research on driving conditions of urban vehicles and optimal matching of transmission system[D]. Wuhan University of Technology, 2013.

[2] Nguyen, Nghiem, Le, et al. Development of the typical driving cycle for buses in Hanoi, Vietnam. 2019, 69(4):423-437.

[3] Ding Yi-feng, Li Jun, Liu Yu. Experimental study on actual road driving conditions of heavy diesel vehicles[J]. Automotive Engineering, 2017, 39(12): 1438-1443.

[4] Liu Ying-ji, Xia Hong-wen, Yao Yu, et al. Vehicle driving condition formulation method combining principal component analysis and fuzzy c-means clustering [J]. Highway and Transportation Science and Technology, 2018, 35(03): 79-85.

[5] Zhang Rui, Wang Yi-wu, Zhu Xiao-long, et al. Research on K-means algorithm for optimizing initial center based on UPGMA [J]. Computer Technology and Development,2018,28(02):50-53+58.

[6] Zhang Lin, Chen Yan, Ji Ye, et al. Research on a density-based K-means algorithm [J]. Application Research of Computers, 2011, 28(11): 4071-4073+4085.

[7] Luo Jun-feng, Suo Zhi-hai. A density-based k-means clustering algorithm [J]. Microelectronics and Computer, 2014, 31(10): 28-31.

[8] Zhang Yan. Clustering algorithm based on rough set and genetic algorithm [D]. Shaanxi Normal University, 2010.

[9] Ding Yi-feng, Li Jun, Gai Hong-chao, et al. Application of wavelet transform in vehicle speed data processing for construction of driving conditions [J]. Science Technology and Engineering, 2017, 17(28): 274-279.

[10] Li A-wu, Zhang Cui-ping, Wang Yang, et al. Research on the construction of driving conditions and emission values of light vehicles in Taiyuan City [J]. Chinese Science and Technology Papers, 2017, 12(22): 2537-2542.

[11] Peng Yu-hui, Zhuang Yuan. Combinatorial optimization clustering and Markov chain construction method of urban sanitation vehicle driving conditions[J]. Journal of Fuzhou University (Natural Science Edition), 2019, 47(04): 502-508.

[12] Chen Zhao-ming, Wang Wei, Zhao Ying, et al. Improved Principal Component Analysis and Multiple Regression Integration of Hanfeng Lake Water Quality Assessment and Prediction [J]. Environmental Monitoring Management and Technology, 2020, 32(04): 15-19 .

[13] Liu Qing-yuan, Li Yong, Pu Xun-chi, et al. Application research of improved principal component analysis method in reservoir water quality evaluation[J]. Sichuan Environment, 2017, 36(06): 116-122.

[14] Yuan Su-fen. Research on driving conditions of urban vehicles and optimal matching of transmission system[D]. Wuhan University of Technology, 2013.

[15] Zhang Jie, Zhuo Ling, Zhu Yun-you. Improvement and application of a K-means clustering algorithm [J]. Application of Electronic Technology, 2015, 41(01): 125-128+131.

[16] Song Yi-fan. Construction of urban road vehicle driving conditions in Shenzhen based on clustering and Python language [D]. Chang'an University, 2018.

[17] Gao Jian-ping, Gao Xiao-jie. Construction of actual driving conditions of vehicles based on improved fuzzy C-means clustering method [J]. Journal of Henan University of Science and Technology (Natural Science Edition), 2017, 38(06): 21-27+4-5.

[18] Liu Bing-jiao, Shi Qin, Qiu Duo-yang, et al. Construction of driving conditions and accuracy analysis based on improved ant colony algorithm [J]. Journal of Hefei University of Technology (Natural Science Edition), 2017, 40(10): 1297- 1302.

DNS is the Internet Pivotal Basics and Fundamental

Chengjin Mou
[1.] Senior Researcher of Kunlun Strategy Research Institute and Information Security;
[2.] Director of International Strategic Research Center of CMCA;
[3.] National Conditions Deputy Director of Development Strategic Security Research Center;
[4.] Zhejiang Province Chief Researcher of Beidou Future Networks Space Research Institute
E-mail: mcjzp139@139.com

***Abstract*—DNS provides name resolution services for Internet applications and is an important infrastructure of the Internet. The domain name system is the core infrastructure of the Internet, which is responsible for the composition of irregular digital sequences Internet protocol address (IP) and highly readable domain names are converted to each other, which is an important prerequisite for maintaining the normal operation of the Internet. The domain name system provides domain name to IP address translation. DNS system was designed to run in a trusted environment at the beginning, but now the complex Internet environment makes the vulnerability of DNS protocol appear.**

This paper briefly describes the current situation of the Internet and the domain name system. By analyzing the current situation of the domain name system, the structure of the IPv6 domain name system and the development of DNS related technologies, it concludes that DNS Security issues are not limited to "vulnerability" or "harassment", but have a clear strategic and systematic nature, and have become one of the focuses of unprecedented struggle and competition, that is, "whoever controls DNS will own the Internet". At the same time, this paper also summarizes the latest research achievements in DNS protocol design and system implementation, and prospects the possible hot research directions in the future.

Keyword-DNS; Internet; Domain Name Resolution; IPV6

I. DNS IS THE INTERNET

DNS is the abbreviation of English domain name system which refers to the domain name system of the Internet (the same below) When DNS is referenced in Chinese; it is usually understood directly as "domain name resolution system".

In September, 2021, Geoff Huston, chief scientist of the Asia Pacific Network Information Center (APNIC), pointed out at the Symposium on "DNS openness" held by the European electronic communication regulatory authority (beret), "Every Internet user connected to the Internet must first access DNS without selectivity. In fact, this attribute essentially defines what the Internet is, that is, DNS is the Internet."

In this sense, today's Internet at least includes the DNS system resolved by the bind software of the United States the DNS system[1] resolved by the NSD software independently developed by the Netherlands, and the "Russian sovereign Internet"(RuNet) that independently adjusts the autonomous domain and legislates to regulate DNS resolution.

DOI: 10.2478/ijanmc-2022-0012

While the Russian Ukrainian cyber war, which is called "300000 global hackers" by the US media, is under way, the US internet technology and capital leaders are working together to "sever" Russia, the US will expand its cyber forces, the US has signed the Declaration on the future of the Internet with more than 50 countries and Taiwan, and the US led NATO has accepted South Korea to join the Cyber Defense Center, which has repeatedly stated that, The United States has never given up its hegemonic proposition of "one world, one Internet".

Facts have constantly proved that the view in the Research Report on "China and the domain name system" of the London Institute of economic and Political Sciences (LSE) on March 19, 2009 is groundless, that is, the domain name system DNS is a typical "inherently political" technology; The attempt to change the politicization of DNS lacks binding force; Getting rid of the inherent political nature of DNS technology depends on the change of new standards and architecture.

Facts have further proved that DNS constitutes the key core foundation of the global network addressing mechanism and virtualization services (such as "content push network" CDN) on which the basic functions of the Internet depend. With the rapid development of Internet technology and application, the role of domain name system DNS not only lies in its importance and security, but also highlights the ownership of its command and control.

As of March 23, 2022, the United States has revoked the authorization of "Article 214"[2] of five state-owned telecom operators in China, all of which have taken effect. This means that the Chinese public network, which was fully functional connected to the U.S. Internet in 1994, will be disconnected at any time, regardless of the IPv4 or IPv6 protocol, or other virtual overlay networks attached to the Internet, that is, the most basic DNS link of the Internet will be disconnected, and the addressing mechanism of the Internet root name server system and the basic support for virtualization services will be disconnected.

II. DOMAIN NAME SYSTEM

DNS includes an ecosystem composed of domain name registration, domain name application protocol, domain name resolution hierarchical service, domain name server software, and communication networks, which constitutes an information and communication technology and service (ICTs) supply chain. Therefore, DNS is the "system of systems" of the Internet, and the key foundation and foundation for the "multi stakeholders" of the Internet to attach great importance to (and seize).

A. Domain name resolution

DNS is actually a distributed database. Its hierarchical structure is similar to the file system structure of UNIX (Figure 1), presenting an inverted "tree" with roots at the top.

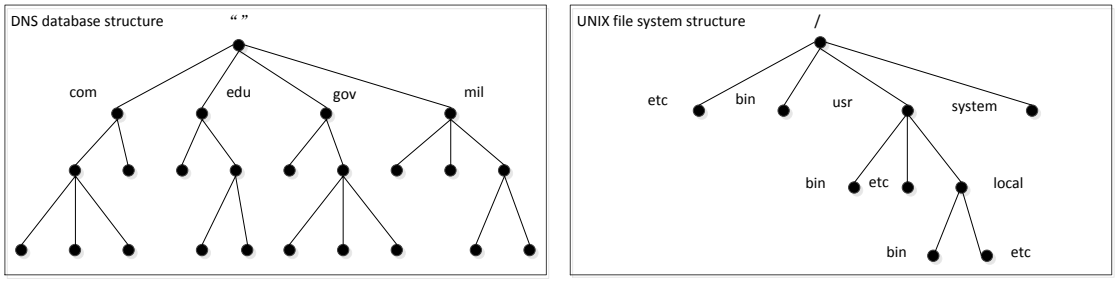


Figure 1. DNS structure of domain name system and UNIX file system structure

The composition and resolution of domain names follow the rule of "right to left" and flow in an orderly manner. They are the root domain name, the top-level (Level 1) domain name, the authoritative (Level 2) domain name, and the sub (Level 3, level 4, etc.) domain name, followed by "." to distinguish. For example, "www.icbc.com.cn" is a three-level domain name. The United States attaches great importance to and closely controls the DNS hierarchical tree system domain name resolution, and never allows any reversible, variable and movable situation to occur.

According to the definition of "DNS terminology" (RFC 84992019-1), DNS "instances" allow multiple DNS servers to have the same IP address in anycast routing, and each server (cluster) is called a "node"; Such DNS domain name server nodes are also called "anycast nodes".

Since 2008, all root domain name servers have applied the "anycast" technology to realize the transformation of the root domain name server into a "system of systems" and provide faster services. Therefore, the global Internet is no longer 13 root domain name servers, but the integration of 13 root domain name server systems [3].

TABLE I. DISTRIBUTION OF NODES OF 13 ROOT DOMAIN NAME SERVERS IN CHINA

Root domain name system(As of November 18, 2021)	**Chinese Mainland (22 nodes)**	**Hong Kong (11 nodes)**	**Taipei (7 nodes)**
A (Verisign)	-	1	-
E (NASA)	-	2	1
F (ISC)	BeiJing:1 HangZhou:1 ChongQing:1 XiNing:1	3	2
H (Army Research Laboratory,ARL)	-	1	-
I (Netnod,Switzerland)	BeiJing:1	1	1
J (Verisign)	BeiJing:1 HangZhou:1	3	-
K (RIPE NCC,Netherlands)	BeiJing:1 GuangZhou:1 GuiYang:1	-	1
L (ICAN N)	BeiJing:1 WuHan:1 ZhengZhou:1 XiNing:1 HaiKou:1 ShangHai:1	-	2

The total number of nodes in the 13 root domain name systems increases or decreases dynamically.

The above Tble1 shows that with the help of anycast technology, the 13 root domain name server systems of the Internet have set up 1469 nodes in the world (configured with established IPv4 and IPv6 addresses).

Please note that any one root domain name server and its corresponding anycast site or node will not be composed of one or two computers or several cabinets, but a server (data cabinet) cluster designed and constructed according to the needs of the service object and scale (target and target group) and sufficient computing support. Among them, the core supporting the computing power of the server cluster is the algorithm.

B. DNS protocol family

In November 1983, the concept, facility, implementation and specification of domain name system DNS (RFC 882, RFC 883) was formally proposed.

The Internet Engineering Task Force (IETF)[4], established on January 14th 1986, is responsible for formulating and promoting the voluntarily adopted Internet standards and specifications, especially the standards constituting the TCP/IP protocol family. At present, RFC involving DNS and related to TCP/IP four-layer architecture and protocols includes 299 protocols (standards) in 7 categories. As shown in Table 2.

TABLE II. RFC CATEGORY, QUANTITY AND DNS RELATED QUANTITY

Category of RFC	Number of RFCs	Where the number of RFCs associated with or related to DNS
Standard	122	5
Proposed Standard	3,819	142
Best Current Practice	301	25
Informational	2,791	87
Experimental	522	29
Historic	331	10
Uncategorized	887	1
Total	8,773	299

C. Top level domain name

The service Zone of the root domain name is the top-level domain name. In other words, the root domain name system is available and serviceable only when the top-level domain name is registered and enabled. The authorization and management of the Zone File of the root domain name and the services of the root domain name system are currently only available for 1588 top-level (or first level) domain names. The registration of top-level (or first level) domain names requires the approval and authorization of ICANN [5].

On January 1st, 1985, six top-level domains ".com,.net,.org,.edu,.gov,.mil" were first registered, marking the official operation of the root domain name system. Among them, "com,.net,.org" is called the general top-level domain name, and ".edu,.gov,.mil" is used as the special top-level domain name.

The US national top-level domain name ".us" was registered on February 15th, 1985; The

Chinese national top-level domain name ".cn" was registered on November 28th, 1990.

Currently, top-level domain names are divided into seven types (As shown in Table 3):

TABLE III. TYPE AND QUANTITY OF TOP-LEVEL DOMAIN NAMES

TLD type of top-level domain name	Abbreviation	Current quantity
Common top-level domain name	gTLD	3
New generic top-level domain name	ngTLD	1,240
Country / region code top level domain name	ccTLD	316
Qualified generic top-level domain name	grTLD	3
Infrastructure top level domain name	(arpa)	1
Private top-level domain name	sTLD	14
Test top level domain name	tTLD	11
Total	-	1,588

In practical applications, a large number of registered domain names are secondary (or authoritative) domain names, that is, the names of organizations, enterprises and websites, such as "ccb.com".

VeriSign reported that as of June 2021, 367.3 million secondary domain names had been registered worldwide; among them, 181million are general top-level domain names (GTLD), accounting for 49.3%.

D. Leading DNS software

In 1984, the system software running on the first root domain name server was called "JEEVES", which was designed and developed by Paul Mockapetris. At the same time, the first DNS software version, funded by the Defense Advanced Research Projects Agency (DARPA) of the US Department of defense, developed by Berkeley University (four graduate students) and released in May1984, is called "BIND"[6]. Later, Doug Kingston and Mike Muuss of the U.S. Army ballistic laboratory made major changes to the BIND software code, which was used in the H-Root domain name server of the U.S. Army ballistic laboratory in 1985. Perhaps the H-Root can be considered as the Taproot of DNS domain name resolution.

With the support of the U.S. Department of homeland security, BIND has been completely upgraded and changed from structural design modification to comprehensive code update. In September, 2000, the Internet Software Alliance Corporation (ISC) released the new main version of BIND (BIND 9), which has been used until now: from the release of version 9.0.0 on January 28, 2004 to the release of version 9.17.18 on September 15, 2021, 673 sub versions have been released in 18 years (the life cycle of the sub versions is generally one year), including software upgrade, defect modification and vulnerability patch.

With its first mover advantage and the "promotion" mode of providing free software and open source code, BIND has a market share (allegedly) of more than 90% and is also regarded as a "de facto standard" in the industry. It must be noted that the "free software" of BIND is not equal to "open source code" and should not be confused.

In May, 2002, NLnet Lab in the Netherlands released the DNS software independently

developed, called "NSD", which was enabled in the "K" root domain name server in February 2003, replacing "BIND". At present, four root domain name server systems (D, H, K, L) have adopted "NSD" [7].

In addition, in March 2021, the National Security Agency (NSA) of the United States released the new DNS software developed, called "Protective DNS" (PDNS), which is currently mainly used in military networks and defense infrastructure (DIB) networks. In the sense that "DNS is the Internet", it should also be a system that can be resolved and run independently.

E. DNS supply chain

The simplified relationship of the ICTs supply chain of the root domain name system DNS is shown in the following figure 2:

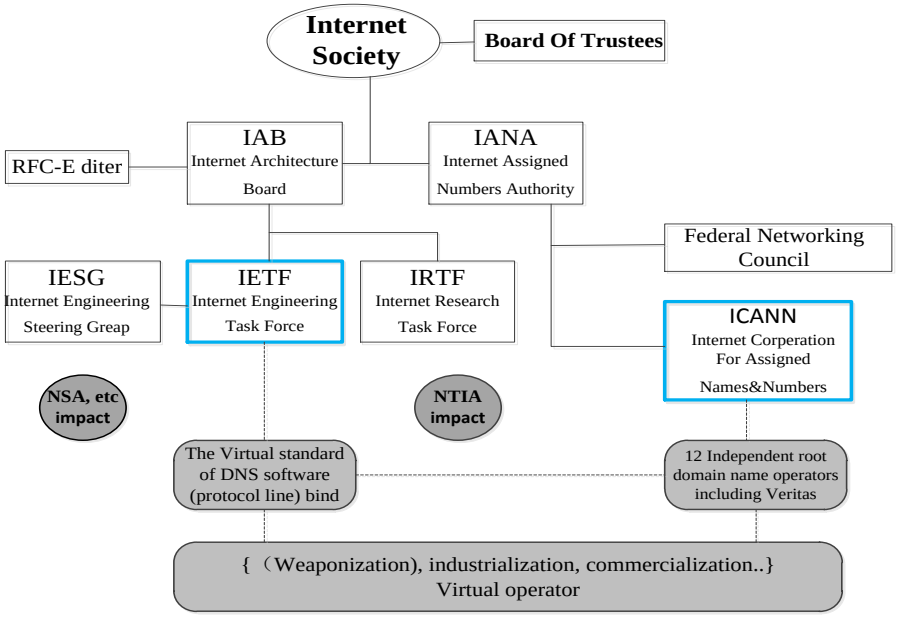


Figure 2. Simplified diagram of ICTs supply chain based on root domain name system DNS

1) ICANN and IETF are parallel, or IETF has no direct relationship between the R & D of technical specifications and the authorization and management of digital resources;

2) 12 root domain name system operating units such as Veritas are not fully subject to ICANN, including the influence exerted by NTIA on behalf of the U.S. government in fact;

3) The DNS standards and specifications of IETF are only the minimum set of codes in DNS software implementation (that is, the codes that implement DNS software protocol stack have enough "discretionary" space, which varies with different software protocol stack developers);

4) All kinds of virtual operators can adapt to the "vest" and change or subvert the application mode and mode of DNS. Such as DNS based content push network (CDN);

5) The National Security Agency[8] (NSA) and the national cyber security and Infrastructure Security Agency (CISA) of the United States have considerable (demand oriented) guidance and (weaponization) influence on the research, development and application of DNS software technology. For example, BIND software has a special and customized version, which is different from the free version open to the public.

In the ecological environment of the root domain name system, if ICANN has the decision-making power over digital resources (domain name, IP address and Asn autonomous system number) and root domain name management, then the domain name execution power in applications and services is another matter.

III. Regulation and governance of DNS

A. Root domain name server

In 1984, the first root domain name server of the Internet was established to serve only the ARPAnet network of the USD Epartment of defense.

In 1985, the number of root domain name servers increased to four, which were hosted by the American Academy of Information Sciences (ISI), Stanford Institute of International Studies (SRI) and the US Army ballistic Laboratory (BRL).

In 1987, the number of root domain name servers was increased to 7, and the services were extended to ARPAnet (Development and Testing network), MILnet (military network), NSFnet (National Science Foundation Network), SURAnet (Southeast University Network), BARRnet [9] (Western Silicon Valley Research Network) and NASA-Science (NASA research network).

In 1995, the year after China's full-featured access to the U.S. Internet, the number of root domain name servers increased to 9, the services were extended to NORDUnet, and the root domain name servers were renamed from "A" to "I".

1997 was an important year for the United States and China to develop the internet almost synchronously. The United States has achieved and improved the Internet in DNS technology, and China has provided a large market for Internet applications:

▲ In January, the United States added four new root domain name servers from "J" to "M". So far, 13 Internet root domain name servers have been built and renumbered from "A" to "M", completing the overall architecture deployment of the Internet domain name system. On January 1st, China's people's network was connected to the Internet.

▲ In May, the "K" root domain name server was transferred from the United States to the Internet Exchange Center (Linx) in London, England, and then to Amsterdam, the Netherlands (autonomous system AS 25152), which was managed and operated by the European Internet Coordination Center (RIPE NCC). On June 3th, China Internet Network Information Center (CNNIC) was established.

▲ In August, the United States transferred the "M" root domain name server to Tokyo, Japan (autonomous system AS 7500). In October, CHINANET was interconnected with CSTNET, CERNET and CHINAGBN.

▲ Please note that since the DNS data message adopts UDP transmission protocol, the data packet length is 512 bytes; The length of IPv4 address is 32 bytes. The IPv4 addresses of 13 root domain names occupy 416 bytes (embedded in DNS message), and only 96 bytes are DNS data and information. Therefore, it is the main reason for limiting the Internet to set up 13 root domain name servers at most.

However, this does not mean that if the IPv6 address length is 128 bytes, the root domain name server can be added. In fact, after more than 30 years of transitional tests, IPv6 is still only attached to and subject to the existing 13 root domain name server (system) architecture of IPv4,

and has not established a "pure IPv6" root domain name server (system).

B. Root domain image server

The root domain name mirror server is divided into two types: Global and Local. Two or more servers with identical online content and synchronized updates are all mirror servers except the host server.

Mirror servers are also called "Instances" in the industry. Instances are not interconnected and can be controlled or managed independently based on the Web or the command line.

In the node list of each root domain name server, some instances are marked as "Whole Network", while others are marked as "local". The instance mark indicates the application range of the image server. The application range of the image server is determined and limited by the routing method of the instance (the Border Gateway Routing Protocol BGP of the autonomous system running on TCP).

The whole network instance allows routing announcements to be broadcast on the global Internet, that is, any router on the Internet can know the routing path to (link to) the instance. For a specific source, the established route of the instance may not be the best route, and other instances can be selected as the destination (via). All root domain name server operators must have at least one network wide instance to provide services for the global Internet [10].

For local instances, route advertisements are limited to connected networks. For example, the instance might be visible to only one network operator (ISP) or to an ISP connected at a particular switching point. Other (or remote) domain name resolution requests cannot be viewed and queried. Some root domain name server operators may also choose to deploy local instances according to their own and partner needs.

The mirror server is a server that shares the load of the host. It synchronously maps the data information passing through the host server like a mirror. It can be seen, but it may not be "retained" or complete. Because the mapping of the mirror server is subject to the mirrored host, what the host has can be mirrored, and no change, change or modification is possible or allowed.

It must be noted that the "anycast node" of DNS cannot be considered or used as the "mirror point" of the domain name; hosting an anycast node is conditional, and related parties must sign a confidentiality agreement (NDA).

C. Root domain name server control

The national telecommunications and Information Administration (NTIA), established in 1978 under the Department of Commerce, is an administrative department of the federal government. NTIA's plans and decisions mainly focus on expanding broadband Internet access and adoption in the United States, promoting the use of spectrum for all users, and ensuring that the Internet continues to be an engine of continuous innovation and economic growth.

The "cooperation agreement with Veritas" published on NTIA website states that Veritas manages the authorized root zone documents in accordance with the cooperation agreement No. NCR 92-18742 signed with the U.S. government. Verizon's responsibilities include: editing the root zone file according to the suggested changes, publishing the file, and then distributing the file (through the a root domain name server) to other root domain name server operators. From this point of view, the status and role of root a domain name server [11] in the 13 root domain name

servers is equivalent to the parent root, or the "Female-child root" integrated into one.

Root A has been managed by Verizon for decades under the strong protection of the U.S. government and military. One of the main functions is to distribute and push authoritative updates of top-level domain names to other 12 secondary root domain name server systems every 24 hours to ensure the consistency and uniqueness of the real-time operation of global DNS domain name resolution. Any root set up by countries around the world (including China) according to the Internet layout (including all root domain name mirror servers around the world) must and can only follow root A for real-time synchronous update, coordination and domination, that is, subject to root A. Otherwise, the operation of the whole network (including China's local Internet) may be seriously disrupted, or even a large area of congestion, block and interruption may occur, and the normal service cannot (or cannot) be provided.

Based on the "cooperation agreement" between NTIA and Veritas, from october1, 1998 to october26, 2018, NTIA and Veritas signed 26 public amendment and supplementary agreements. According to the 2015 financial report submitted by Verizon to the Securities and Exchange Commission (SEC), it clearly states:

1) DNS is supervised by the Department of Commerce on behalf of the U.S. government. According to the letter of commitment (AOC) signed by the Ministry of Commerce and ICANN, which took effect on October 1st, 2009, the Ministry of commerce is one of the subjects of continuous review and accountability of ICANN's performance [12].

2) The role of ICANN is to serve as the coordination core among multi stakeholders. The above-mentioned letter of commitment is not binding on ICANN.

3) The role of the U.S. government is to coordinate the management of important aspects of DNS through NTIA, including the functions of the Internet digital distribution authority (IANA) and the DNS Management of the root domain name zone.

The key point is that all the above root domain names (regardless of the parent root, Taproot and secondary root), root domain name system (host and image distribution, establishment and operation control system), IPv4 and IPv6 protocols are the network terms, proprietary functions and specific elements of the Internet developed by the United States. The Network sovereignty of the Internet, including naming right, jurisdiction, design and planning right, rule determination right, operation dominant right, routing dominant right, data control right, as well as the distribution and lease right of domain name address, the distribution and coordination right of root domain name mirror server node, etc., can only be decided by the United States. Any country (including China, Japan, European countries, etc.) and organization (including the United Nations, ISO/IEC, ITU, etc.) outside the United States does not count.

Female root refers to the source root of the whole Internet. The U.S. military network evolved and reconstructed from ARPANET is the core network, origin network, network in network and leading network (main network) of the Internet. The "Female root" should be hidden in the "main network" that the U.S. military absolutely controls and provides a high degree of security.

Taproot is the direct root and the direct root of the Internet. Root A is considered to be the Taproot among the 13 Internet roots, and the Female 12 roots are Auxiliary roots. For example, root M in Japan is the Auxiliary root.

The U.S. government has carefully planned and constructed the jurisdiction (legal system) and control (governance mechanism) over the Female root, Taproot and the Auxiliary root of the root domain name, which can not be disobeyed and changed by any other countries, organizations and individuals.

IV. WHERE IS DNS GOING

A. DNS Security Extension

With the increasingly prominent security problems and risks of domain name resolution system DNS, IETF has continuously issued a set of standards and specifications "Domain Name System DNS Security Extension" (DNSSEC) since 2005.

At 0:00 on October 12th, 2018 (Beijing time), ICANN implemented the domain name root zone key reversal (KSK, key signing key) on the global Internet, replacing the single trust root used to verify the consistency of DNSSEC response, which is the first time in the history of the Internet.

But so far, the application of DNSSEC is far lower than expected. Among them, the technical reasons include: the digital signature of DNSSEC increases the number of bytes of DNS resolution response packets, making most DNS resolution response packets remain under the UDP transmission limit of 512 bytes, which is becoming more and more challenging. At the same time, in order to keep the DNS specification unchanged, packets with more than 512 bytes can be truncated and switched to TCP to obtain domain name resolution responses with more than 512 bytes, potentially reducing the efficiency of DNS (and increasing the delay and fragmentation of packets).

Therefore, in reality, the deployment and application of DNSSEC are "layered and segmented". For example, DNSSEC is adopted for the domain name resolution service of root domain name and top-level domain name, while DNSSEC is basically not adopted for the domain name resolution service of authoritative domain name, recursive domain name server and user terminal.

Geoff Huston, chief scientist of APNIC [13], believes that the current status shows that DNSSEC is one of the typical cases of application failure.

According to the measurement statistics of APNIC, as of May 6th, 2022, the average verification rate of DNSSEC in the world is 29.91%, of which (As shown in Table 4):

TABLE IV. AVERAGE VALIDATION RATE OF DNSSEC IN THE WORLD

Country(and Region)	DNSSEC Average validation rate
India	59.60%
Russia	52.23%
America	38.78%
China(mainland)	0.93%
(Hong Kong)	57.25%
(Taiwan)	6.41%
(Macao)	5.23%

The interoperability of IPv6 and IPv4 technologies ("mutual incompatibility") is one of the key foundations and fundamental cruxes of Internet security issues. Although both belong to the Internet Protocol (different versions) and are controlled and operated by 13 root domain name server systems that also use and rely on the Internet, there are still a large number of potential instability factors and unknown security risks that cannot be predicted and prevented, Necessary and necessary practical (parallel) operation experiments and verifications must be carried out to explore possible solutions, which requires incalculable economic costs and security costs, or the gains outweigh the losses.

B. Yeti DNS Project

In June 2015, ICANN launched the "Yeti DNS Project" proposed by American expert Paul Vixie and the Japan WIDE organization (Japanese translation “雪だるまプラン”,Translated into English as "Yeti DNS project")

The "Yeti DNS Project" is an experimental project for parallel testing of IPv4 and IPv6 domain names. ICANN uses the "Yeti DNS" DNSSEC key to test all restart and Reset settings in the M root domain. It does not provide a substitute domain name space, but only changes (adds) the delegation information of the M root domain name system resolution.

American professionals acknowledge that this preliminary technical test project, which is temporary and allowed to fail in the test environment, is neither the experiment of "technical prototype", nor the deployment of root domain name server system in the production environment, nor the "new pattern of IPv6 root domain name server hypothesis", and has been completed at the end of December 2017.

American professionals clearly pointed out that the tests and experiments of the "Yeti DNS Project" have proved that:

1) The "Taproot root server" of IPv6 is not a truly independent Taproot server, but a testing server under the Taproot root server of IPv4. Technically, the status of the 25 new IPv6 root servers in the "Yeti DNS Project" is actually lower than that of the 13 IPv4 root servers.

2) The so-called IPv6 "Taproot root server" in China is controlled and monitored by the IPv4 Taproot root server and f root server in the United States. "The root server of IPv4 still has the right to interpret the root server of IPv6." "Even if China has a root server for IPv6 in the future, it does not mean that China can play a leading role."

3) The security performance of IPv6 is inferior to that of IPv4. IPv4 addresses can be dynamically allocated, and each IPv6 website has only a certain static address, which is easy to be accurately located and attacked. Therefore, the U.S. government and the U.S. military do not use IPv6, and deliberately push China to spend a lot of human, material and financial resources to build a pilot IPv6 system.

4) The "Yeti DNS Project" does not attempt to "bifurcate the domain name space". All tests are based on the expansion under the IPv4 architecture, rather than "starting a new business" to re-establish a new domain name management system and root domain name server. In other words, whether IPv4 or IPv6, the global Internet's total hub, data exchange center, backbone network, Taproot root server and Web master station are still in the United States, which are built, controlled and managed by American enterprises. IPv6 and the "Yeti DNS Project" have not solved China's Internet security problems at all.

In short, the United States has full sovereignty over both IPv4 and IPv6 domain name space. The US government, US politicians and US politics do not allow any country, any organization or any individual to change or shake the US "one net dominating the world", which has become a well-known Internet common sense, scientific common sense and political common sense all over the world.

C. Development of DNS related technologies

1) Since January this year, ICANN has widely publicized a revised action measure: "Knowledge-Sharing and Instantiating Norms for DNS and Naming Security", referred to as "KINDNS".

On March 10th, at the 73rd annual video conference of ICANN (ICANN 73) [14], the Middle East representative of AFRINIC, the management organization of Africa, issued a "statement" on "digital signature and verification of DNSSEC":

It is emphasized that DNS is crucial to ensure the continuity of network services. Defective or invalid DNS services will have a negative impact on the experience of any institution and organization (including customers, partners or employees), affect e-commerce applications, cause loss of revenue and damage the brand image. It is disclosed that 63% of institutions and organizations will be offline and out of service due to DNS attacks in 2021; It is recommended to clarify the "return on investment" (ROI) in DNSSEC deployment and application and the "risk loss rate" (ROR) for delayed deployment and application of DNSSEC.

On May 19th, ICANN made an official reply, acknowledging that the importance of DNS Security operation to the overall stability and flexibility of the Internet was recognized, which is the core of ICANN's mission; Indicates that the series of recommendations put forward in the statement are consistent with ICANN's "five year strategic plan" and "five year plan for the Middle East and surrounding countries and regions", that is, they are related to ICANN's regional objectives; Define the regional objectives of ICANN, including:

—Through cooperation and support with multi stakeholders, to develop technical capacity and establish a regional network of technical experts;

—Identify and mitigate the security threats DNS faces by participating in the work of multiple stakeholders.

"Multi stakeholder", in fact, is to build a new "threshold" of exclusive competition through the cooperation between the government and private enterprises, and take their own interests or vested interests. One of the "principles" put forward in the "Declaration on the future of the Internet" signed and issued by the United States with the EU, more than 50 countries and Taiwan on April 28, namely "protecting and strengthening multi stakeholder governance methods".

2) The technology and services of DNS are undergoing fundamental changes, not only in the underlying protocols (and key technologies), but also in the alliance of governance models and the strengthening of management means.

The "QUIC" protocol [15], a new generation of network data transmission protocol based on the "UDP" protocol, is known as a subversive innovation comparable to the "TCP" protocol.

On May 11, IETF released the updated version RFC 9250 in the standardization process of "DNS based on private QUIC connection"; On May 20, IETF updated the second version of the standardization process based on the "UDP" protocol.

D. Attention tips

1) The global Internet digital resources allocated and managed by ICANN mainly include: top-level domain name, IP address and ASN autonomous system number. Although the average verification rate of global DNSSEC (29.91%) is close to the average application rate of IPv6 (31.22%), ICANN builds and promotes "DNS" (KINDNS framework) and does not try to build a global IPv6 application ecosystem. Is it consciously "favoring one over the other"?!

2) The non mandatory technical standards for DNS are formulated by IETF, but the digital resources of DNS (such as top-level domain name

authorization and port allocation) are managed by ICANN (and IANA). ICANN has built a "KINDNS framework" for the research and development of IETF standards and related technologies, building (seizing) a "first mover advantage" platform for the global Internet.

3) Although ICANN is still maintaining DNSSEC, when QUIC based DNS (DOQ) becomes the standard, it may replace DNSSEC. In other words, DNSSEC is only a transitional or "dead meat meal".

Combing the domain name system DNS, an intuitive and simple evolution is that DNS has already become the "hub" (command and control system) and "commanding point" (navigation system for positioning and redirection) of the Internet from its initial idea as the Internet "phone book" (file system).

E. Epilogue

DNS Security is not limited to "vulnerability" or "harassment", but has a clear strategic and systematic nature, and has become one of the focuses of unprecedented struggle and competition, that is, "whoever controls DNS will own the Internet" [16].

The Internet is man-made. It is the creative result of human collective wisdom. It is a systematic innovation that carries forward the past and advances with the times with the continuous sublimation of human knowledge, culture, science and technology. Especially in the evolution and development of DNS domain name system and its security technology, there is no shortcut, no "one move to beat the world" and no "eternal" technical know-how once and for all.

DNS is the key foundation and foundation of the Internet. The innovation of DNS domain name system is not only a technical programming based on experience, but also an ecosystem project, including dynamic supply chain and potential political, economic, diplomatic and military strategies and strategies.

Complete the manuscript May 24, 2022.

REFERENCES

[1] Huning. dengwenping. Yaosu Research status and challenges of Internet DNS Security [J] Journal of network and information security, 2017, 3 (03): 13-21.

[2] Miaochen. Analysis and Research on Internet DNS traffic [D] Beijing University of Posts and telecommunications, 2013.

[3] Salnikov, Andrii, Kónya, Balázs. DNS-embedded service endpoint registry for distributed e-Infrastructures [J]. Cluster Computing, 2021 (prepublish).

[4] Xie Chongfeng. Viewing the development trend of IPv6 from IETF dynamics [J] ICT and policy, 2020, (08): 12-17.

[5] Kouwenjun. Research on IPv6 domain name system [J] Scientific consulting (Science and technology. Management), 2016 (03): 42-44.

[6] Liuqing. Security issues of Internet domain name system in China [J] Modern telecommunication technology, 2010, 40 (04): 9-11+17.

[7] Liyanxing. Research on DNS Security Extension and scalable distributed DNS [D] University of Electronic Science and technology, 2021,Doi:10.27005/d.cnki gdzku 2021. 002068.

[8] Zhangwenjia. Research on Key Technologies of DNS root zone resolution self verification [D] Harbin Institute of technology, 2019, Doi:10.27061/d.cnki ghgdu 2019.000952.

[9] Chenghongbo. Research and deployment of DNSSEC security mechanism [D] Shanghai Jiaotong University, 2006.

[10] Luodexiang. Research on attack and defense technology of some Internet security protocols [D] Shanghai Jiaotong University, 2007.

[11] Lei He, Quan Ren, Bolin Ma, Weili Zhang,Jiangxing Wu.Anti-Attacking Modeling and Analysis of Cyberspace Mimic DNS [J]. China Communications, 2022, 19(05):218-230.

[12] Dengchengjun, LiuYing. Implementation of DNS service in next generation Internet [J] Journal of Chongqing Electric Power College, 2021, 26 (02): 35-37+48.

[13] Jiazhuosheng. Research on active defense architecture and key technologies based on domain name service log analysis [D] Beijing Jiaotong University, 2021, Doi:10.26944/d.cnki.gbfju.2021.000328.

[14] Chang Deliang, Hao Shanshan, Li Zhou, Liu Baojun, Li Xing. DNSWeight: Quantifying Country-Wise Importance of Domain Name System [J]. IEEE ACCESS, 2021, 9.

[15] ChenBo. Analysis on security protection technology of DNS [J] Electronic technology, 2020,49 (06): 80-81.

[16] Zhangjichuan. Research on enterprise DNS Security Scheme Based on blockchain technology [D] Harbin Institute of technology, 2020,Doi:10.27061/d.cnki ghgdu 2020.002902.

A Review of Virtual Surgical Object Modeling Technology Based on Force Feedback

Yu Liu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710032, China
E-mail: 1484347675@qq.com

Baolong Liu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710032, China
E-mail: liu.bao.long@hotmail.com

***Abstract*—The early virtual surgical system was only developed on the computer software, and users could only interact with the surgical model through the mouse. The disadvantage of this way was that users could not get tactile feelings. Therefore, with the development and progress of virtual reality technology, force feedback equipment has been introduced into virtual surgery, and has become a research hotspot in this field, in order to make efforts to feedback equipment in the research institute and higher and higher utilization rate. Researchers, teachers and students used force feedback devices to improve the immersion of virtual surgical system. This article mainly introduced the virtual modeling method in the operation, including geometric modeling and physical modeling, the geometric modeling of the two models are introduced, and summarizes their advantages and disadvantages, the physical modeling of soft tissue deformation modeling of three models are analyzed and compared, and the feedback force calculation model are introduced, finally, expounds some problems faced by modeling technology, virtual surgery The future development is also prospected.**

Keywords-Force Feedback; Virtual Surgery; Geometric Modeling; Physical Modeling

I. INTRODUCTION

With the development of science and technology in medical surgery, virtual surgery has become a hot topic and research direction, and a lot of research institute and hospital using computer simulation technology to reproduce the surgical scenario, it's a multidisciplinary cross area of research, including medical, computer graphics, math, mechanical, etc. [1]. In medical teaching, ever took the form of teaching material teaching theory teaching to students, but because of the limited resources, cannot guarantee that every student to practice the operation, cause the teacher cannot effectively guide students [2], as a result, virtual surgery have solved the problem of resource scarcity, provides students with a realistic operation environment, can undertake training over and over again, But because most of the virtual surgery can only provide the feeling on the vision [3], unable to bring in other senses, such as the sense of touch, leading to the doctor in the virtual environment to complete the operation effect of the operation and clinical surgery a greater difference between the effect of the environment, in order to allow users to get tactile perception, is gradually applied to the virtual force feedback devices [4] in the operation, Users by

DOI: 10.2478/ijanmc-2022-0013

force feedback device control virtual surgery tool model and object model of operation environment interact [5], when the force feedback device access to the user's operation, the operation information for the corresponding calculation and feedback to the user of an appropriate size, thereby enhancing the user immersion in virtual surgery [6], the user in the process of training again and again, Having a comprehensive understanding of the entire surgical process and the structural characteristics of the surgical object saves resources and reduces costs.

The modeling of surgical objects and surgical tools in virtual surgery is related to the degree of reality of the virtual surgery system, and also has a great influence on the collision detection, feedback force calculation and deformation calculation. Therefore, geometric modeling and physical modeling are the focus of virtual surgery research. In clinical surgery, in the case of soft tissue, if for press operation, not only with focus on production, and its surface will with corresponding deformation pressure strength, but for this rigid body such as teeth, because the surface of the teeth will not happen with the touch of needle deformation, therefore in virtual surgery, the software not only need to consider the feedback force calculation model, Deformation models are also needed. The establishment of a good geometric model and physical model has an important impact on the real-time and authenticity of virtual surgery.

II. GEOMETRIC MODELING

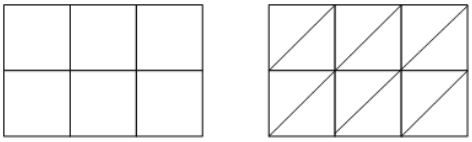
Geometric modeling refers to drawing the geometric shape of the model in the virtual three-dimensional scene, and establishing an accurate geometric model of the surgical object, which can bring a realistic visual experience to the user, which is of great significance for the visual rendering of virtual surgery, such as in clinical oral cavity. In surgery, drilling operations are particularly common. Teeth are the main object of oral surgery. When the doctor drills the teeth, tooth debris will be generated. In the virtual oral surgery, certain structural changes must be made to the geometric model, so that the Users have a clear visual experience. Therefore, in virtual surgery, a reasonable geometric modeling method is a prerequisite to ensure the authenticity of virtual surgery.

There are two methods of geometric modeling: surface model based on surface mesh and volume model based on voxel.

A. Surface model based on surface mesh

Surface model based on surface mesh refers to many discrete particle by line connected together by the model, the model of topological relations between particle and particle, according to the topological relationship model can be further calculated the stress or deformation, therefore, choose the appropriate topology relationship is the premise of guarantee authenticity model deformation. Common topologies of face models include triangles, quadrangles, or other polygons, as shown in Figure 1. In the process of grid division, triangle structure is the smallest topological structure that can be divided. Compared with quadrilateral structure, there is no need to judge whether four points are on the same plane. Therefore, triangle structure is simple, flexible, easy to describe and efficient, and has been adopted by most scholars [7]. Early in virtual surgery system, most scholars use the surface of the triangular structure model to simulate the surgery simulation, for example, in 1994, Massie is simulated by the method, the operation objects in the virtual environment, and the operation object gives stiffness characteristics, force analysis through the spring model can calculate the size of the force feedback, and feedback to the

user, The user can get a tactile feeling, but the model can not reflect the deformation characteristics. In 2017, Liao Denghong et al. used the grid deformation algorithm to realize the drilling operation of teeth, but the model could only reflect the changes of tooth surface, but could not reflect the structural changes inside the tooth model.



(a) Square structure (b) Triangular structure

Figure 1. General surface model surface topology.

Surface model based on surface mesh has certain advantages in displaying the surface structure of the model, and the modeling speed of the surface model is faster than that of the volume model, which can fully display the topological relationship between each particle on the surface of the surgical model. The model of single tooth surface based on triangular plane is shown in Figure 2.

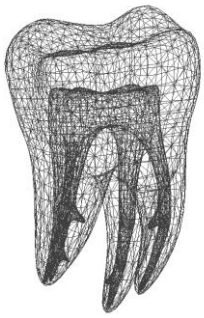


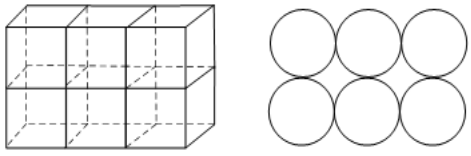
Figure 2. Single tooth surface model based on triangular plane.

B. Volume model based on voxel

Volume model based on voxel refers to contain internal structure information of three-dimensional data model, the model consists of a number of individual element, voxel volume element, namely the voxel is through the body to a space according to certain size and manner of dividing, similar to the pixels, each individual element are closely linked and has its own location information, and fully displayed the model the internal organization of the information, Therefore, the selection of voxels is of great significance to the authenticity of model deformation. Common voxel structures include cubes, spheres or other shapes, as shown in Figure 3. Because cube is the easiest method for structure division, researchers mostly use cube structure to divide voxels and generate corresponding volume models [8, 9]. In 1997, Stijn Oomes applied Scale-space Theory to achieve voxelization of the model. In 1998, Roni Yagel et al. voxelized the model using a function of a normal vector. In 2000, Jones et al. used Distance Fields and Distance Transform to realize the voxelization of the model. Although these voxel methods can generate voxel models faster and better, they all need graphic workstations to participate in and have high requirements on hardware. Therefore, in 2004, Wu Xiaojun et al. proposed a method to voxelize the mesh model by using the structure of the octree. This method calculates the distance from the triangle to the center of the voxel. If the calculated distance is less than the preset threshold, it is considered that the voxel unit intersects with the triangle, and finally voxelization is realized. Although this method does not require the participation of a graphics workstation and only requires a computer, it consumes a lot of computing resources in the calculation process, and the threshold is difficult to determine, resulting in inefficient voxelization and incomplete voxel search. In 2010, Mu Bin et al. proposed a new voxelization method based on the octree structure. By calculating the projected volume, it is judged whether the voxel intersects the triangle, and the adjacency relationship is used to fill the interior of the model with voxels. , and finally realize the voxelization process. This

method projects the 3D to the 2D plane, and then returns the 2D data to the 3D space to realize the voxelization. Although the speed is improved, the processing process of this method is relatively Complicated, returning from 2D to 3D will cause errors to cause incomplete voxel search errors. In 2017, in order to improve this shortcoming, Duan Weiwei et al. proposed a new voxelization method based on the octree structure. First, the octree was used to subdivide the model, and then the model was scanned from multiple directions. This method Although it can quickly determine the surface voxels and internal voxels of the model, it is not suitable for the case where the model contains holes, and it will treat the internal holes as internal voxels. Therefore, the voxelization method still needs to be further improved.

Compared with the surface model, the volume model based on voxel can show the structure of the internal information, for example in the oral cavity in virtual surgery, voxel model can according to different physical properties of the internal structure of simulated teeth of layered internal structure, such as enamel, dentin and dental pulp, etc., thus brings the user in the process of the whole operation and clinical surgery similar visual and tactile sensation. Therefore, the body model is used by most scholars to build the virtual surgical object model. The single tooth volume model based on cube is shown in Figure 4.



(a) Cube structure (b) Sphere structure

Figure 3. General volume model topology.

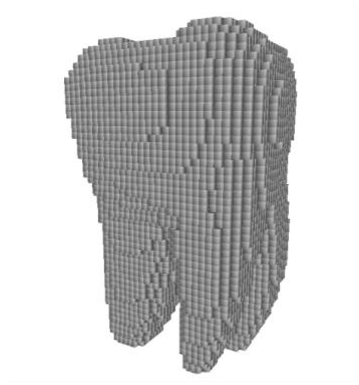


Figure 4. Single tooth volume model based on cube.

III. PHYSICAL MODELING

Physical modeling refers to adding some physical characteristics, such as weight, surface roughness, suction, adhesion and deformation characteristics, on the basis of the geometric model, so that the model in the virtual environment is more consistent with the object in the real world. Therefore, the establishment of physical model is closely related to the authenticity and real-time of the system.

For tissues and organs, soft tissues will be deformed when pressed. In order to achieve this visual deformation effect, a deformation model of soft tissues needs to be established [10]. When pressing, not only the visual deformation effect will be produced, but also the tactile force feedback will be accompanied. In order to realize the tactile feedback, the feedback force calculation model needs to be established.

A. Soft tissue deformation model

At present, common modeling methods for soft tissue deformation include Mass Spring Method (MSM), Finite Element Method (FEM) and Boundary Element Method (BEM) [11]. The analysis and comparison of soft tissue deformation modeling methods is shown in Table Ⅰ.

TABLE I. COMPARISON OF THREE MODELING METHODS

Modeling method	Instantaneity	Complexity	Computational accuracy	Robustness
Mass Spring Method	best	simple	general	general
Finite Element Method	general	complex	best	better
Boundary Element Method	general	more complex	better	better

1) Mass Spring Method

Mass spring method refers to the use of some discrete particles to describe the object, the particles and particles are connected through the spring, the force of the particle meets Newton's laws of motion, the spring meets Hooke's law. The mass spring model is shown in Figure 5.

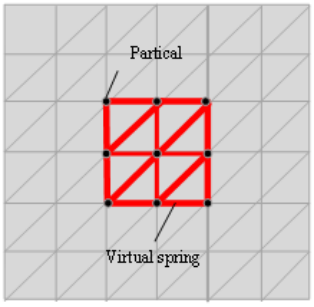


Figure 5. Mass Spring Model

2) Finite Element Method

Refers to a continuum of finite element method is used to solve the domain of discrete into several units, each unit through the certain way of interconnecting approximation instead of the original system, and then within each unit, with the assumption of approximate function to piecewise said the whole solution domain and the unknown variables, finally through with the original problem, the mathematical model of the equivalent variational principle or weighted method, The equations of ordinary differential equations are established to solve the basic unknowns, and then numerical analysis is used to solve the equations, and finally the solution of the original problem is obtained.

3）Boundary Element Method

Boundary element method refers to discretely solving the boundary of the solution domain by piecewise functions. This method is also a numerical analysis method, but it requires less computation than the finite element method. The boundary element model is shown in Figure 6.

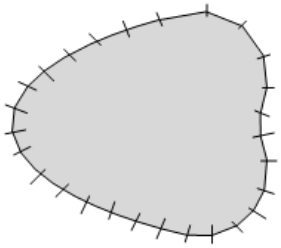


Figure 6. Boundary Element Model

B. Feedback force calculation model

In order to enable users to get tactile feelings in virtual surgery, that is, users can obviously feel the contour, stiffness and other characteristics of the model surface through force feedback equipment [13], so as to make the experience more real, it is necessary to add mechanical properties to the geometric model and establish an appropriate feedback force calculation model for

tissue and organ models. At present, the most commonly used force feedback model is the mass-spring-damper model. However, for rigid body models, such as teeth and human bones, their surfaces will not deform with the contact of the surgical tool model, so damping force is not needed to be considered.

The tactile feedback of the mass-spring-damper model is rendered by the relative position of the surgical tool model, In order to prevent the "piercing" phenomenon, that is, when the surgical object model is touched with the surgical tool, the surgical tool is prevented from being embedded inside the surgical object model. Therefore, for the situation where the visual and tactile positions of the surgical tool model are inconsistent, 1995 In 2008, Zilles and Salisbury proposed a point-based three-degree-of-freedom force feedback calculation method, also known as the God-Object method. This method saves two position information, one is the force contact point, which refers to the position where the surgical tool does not receive any resistance during the movement process, that is, the ideal position; the other is the God-Object point, when the surgical tool is in contact with the position of the collision point when the model object collides. When the surgical tool touches the surgical object in the process of moving, the force contact point will penetrate the surgical object, and the God-Object point will stay on the surface of the surgical object, and the feedback force can be obtained by calculating the distance between the two points. . However, this method has a disadvantage. It is easy to cause the God-Object point to fall into the gap due to the error, resulting in discontinuity of force. Therefore, in 1997, Ruspini proposed a Virtual-proxy method based on this method. The Object point is replaced by a small ball without weight, which improves the disadvantage of falling into the gap and realizes the feedback effect of friction [14].The idea of this method is the same as the God-Object method, require the force feedback Device to store two location information in the virtual surgical scene, one is the actual terminal location of the force feedback device, the other is the proxy location of the surgical tool model. In the initial state, the tool model has not collided with the object model. Therefore, the proxy and device overlap at this time. When the collision between the tool model and object model is detected, the proxy of the tool model will always be on the surface of the object model where the collision occurs. The device "pricks" into the object model as the force feedback device applies varying amounts of force. That is, visually, the tool model is on the surface of the object model, but tactile, the tool model is already deep inside the object model, as shown in Figure 7.

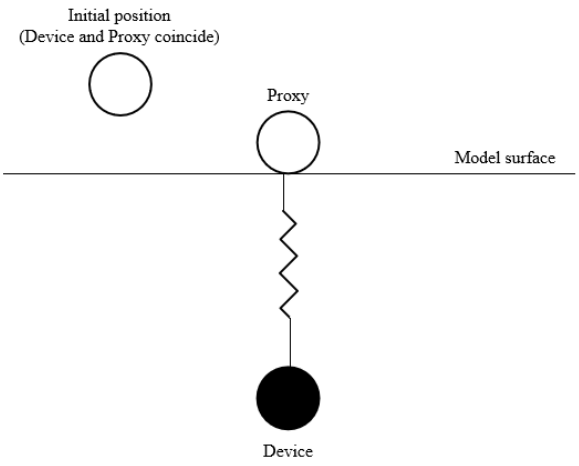


Figure 7. Mass-spring-damper Model

When there is no collision in the moving process of the tool model, the feedback force is 0. When a collision occurs, the spring will stretch with the force applied by the force feedback device. The coordinate of the collision point should be found according to the line between the target position of the tool model and the position of the previous moment and the intersection of the object model surface. Then hooke's Law is used to

calculate the size of the feedback force. Its feedback force can be described as formula (1).

$$F = (k\Delta x + d\Delta x)\vec{n} \quad (1)$$

Where k represents the elastic coefficient of the virtual spring, the distance between the device and the proxy of the surgical tool model, and d represents the damping coefficient, and $\vec{n}$ represents the direction vector of the normal vector of the action point [15]. By adjusting k and d values, the mechanical properties of the surgical object in the actual operation can be accurately simulated, so that doctors can feel the feedback force in the real operation.

After a good model is established, if you want to get the tactile feeling in the virtual surgical system, you need to introduce force feedback devices. Now there are many force feedback devices on the market. Examples include Phantom Omni (Geomagic Touch) from Sensable, 3D Touch from Novint Falcon, Omega.6 and Omega.7 from Force Dimension, as shown in Figure 8. Phantom Omni force feedback device supports 3-Dof position and 3-Dof force feedback, 3D Touch force feedback device supports 6-Dof position and 3-Dof force feedback, Omega.6 force feedback device supports 6-Dof position and 3-Dof force feedback, Omega.7 force feedback equipment supports seven degrees of freedom position and three degrees of freedom force feedback. User through the manipulation of the force feedback device control handle control surgical tools and operation model to interact, the force feedback device access to the user by controlling the control handle to the operation model of the force f a state information, the state information including the tool position or size of the force, and then through the establishment of good physical model, Calculate the size and direction of the reaction force generated by this force, so as to generate force control signal, and then calculate the feedback force through the actuator in the force feedback device, and transmit it to the user through the control handle, so that the user can get the tactile feeling in the virtual environment.

(a)Phantom Omni

(b)3D Touch

(c)Omega.6

Figure 8. Common force feedback devices

IV. CONCLUSIONS

The introduction of force feedback virtual surgery system, for users with visual perception at the same time, also gives the user more true feelings in the sense of touch, and establish a precise surgical object model is the basis and important part of virtual surgery training, to the visual rendering of virtual surgery and feedback force calculation has important influence. This

article mainly introduced the virtual modeling method in the operation, after introducing the force feedback include geometric modeling and physical modeling, the geometric modeling of the two models are introduced, including physical modeling of soft tissue deformation model and the feedback force calculation model, the soft tissue deformation model of the three common models are introduced and compared, the feedback force calculation model and force feedback algorithm are introduced. In this paper, by analyzing a variety of geometric model and physical model for the visual rendering of force sensing and rendering in virtual surgery training provides the theoretical basis, although the present modeling technology can provide users with visual and tactile feeling good, but its real time needs to be improved, the need to improve the model to improve the real-time performance of the system, bring a good sense of immersion.

ACKNOWLEDGMENT

This work is partially supported by Science & Technology Program of Shaanxi Province with project “2021GY-005”. In addition the authors would like to thank Prof. Liu Baolong for his contributions and suggestions.

REFERENCES

[1] David, Pinzon, Simon, et al. Prevailing Trends in Haptic Feedback Simulation for Minimally Invasive Surgery. [J]. Surgical Innovation, 2016, 23(4):415-421.

[2] Zhang Y, Luo D, Li J, et al. Study on Collision Detection and Force Feedback Algorithm in Virtual Surgery [J]. Journal of Healthcare Engineering, 2021, 2021(1):1-12.

[3] Verstreken K, Van Cleynenbreugel J, Martens K, Marchal G, van Steenberghe D, Suetens P. An image-guided planning system for endosseous oral implants. [J]. IEEE transactions on medical imaging,1998, 17(5).

[4] Kusumoto Naoki, Sohmura Taiji, Yamada Shinichi, Wakabayashi Kazumichi, Nakamura Takashi, Yatani Hirofumi. Application of virtual reality force feedback haptic device for oral implant surgery. [J]. Clinical oral implants research, 2006, 17(6).

[5] http://www.measurego.com/Force-feedback.

[6] Lin Y, Wang X, Wu F, et al. Development and validation of a surgical training simulator with haptic feedback for learning bone-sawing skill [J]. Journal of Biomedical Informatics, 2014, 48(2):122-129.

[7] Massie T H, Salisbury J K. The PHANTOM haptic interface: a device for probing virtual objects[C]// Proc. of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. 1994.

[8] Bogoni T, Scarparo R, Pinho M. A virtual reality simulator for training endodontics procedures using manual files. IEEE, 2015.

[9] Jones M W, Satherley R.Voxelisation: Modelling for volume graphics[C]// Proceedings of the 2000 Conference on Vision Modeling and Visualization (VMV-00), Saarbrücken, Germany, November 22-24, 2000. DBLP, 2000.

[10] K. Jayasudha, Mohan G. Kabadi. Soft tissues deformation and removal simulation modelling for virtual surgery [J]. International Journal of Intelligence and Sustainable Computing, 2020, 1(1).

[11] Dan K, Chandrasekaran S, Wang Y F. A New Framework for Behavior Modeling of Organs and Soft Tissue using the Boundary-Element Methods[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition Workshops. IEEE, 2008.

[12] ReinhardMnner, Grimm J W, ClemensWagner, et al. Interactive Real-Time Simulation of the Internal Limiting Membrane. [J].Lecture Notes in Computer Science.Spring-Verlag Herdelberg.2004.3078:153-160.

[13] Soon D, Chae M P, Pilgrim C, et al. 3D Haptic Modelling for Preoperative Planning of Hepatic Resection: A Systematic Review [J]. Annals of Medicine & Surgery, 2016, 10:1-7.

[14] Ruspini D C, Kolarov K, Khatib O. The haptic display of complex graphical environments[C]// Proceedings of the 24th annual conference on Computer graphics and interactive techniques. DBLP, 1997.

[15] Yanping L, Dedong Y , Xiaojun C , et al. Simulation and evaluation of a bone sawing procedure for orthognathic surgery based on an experimental force model. [J]. J Biomech Eng, 2014, 136(3):034501.

E-Commerce Middle Office Management System Based on Springboot

Hejing Wu
East University of Heilongjiang
Heilongjiang, 150086
E-mail: 499917928@qq.com

***Abstract*—This topic takes Vue framework as the front-end framework of e-commerce middle office management system, uses springboot framework as the construction mode of back-end framework, and uses Java to write function code. [1] The data of the system is stored in MySQL database, which can be provided to the employee end of registered employees and the management end of managers. Employees can complete registration and login through the e-commerce middle office management system, and then view department information, store information Warehouse information and view order information and add, view commodity file information. The system administrator manages employee information, department information, store type information, store information, warehouse information, commodity file information and order information through the management end. Through the realization of the different functions of the above employees and administrators, the normal operation of the system is ensured, and the process involving stores, commodities and orders is monitored, So as to make a complete set of e-commerce middle office management system that can be provided to all kinds of users.**

Keywords- E-Commerce Middle Office Management System; Order Management; Vue. Framework; Springboot Framework; Mysql Database

I. INTRODUCTION

At present, with the increase of the number of merchants engaged in e-commerce business and the rapid increase in the transaction volume of e-commerce platforms, the traditional business management methods have obviously been unable to adapt to the effective control of merchants and orders, and the concept of business middle office can realize the digital output and precipitation of the whole process from the store to the transaction into visible data, which can provide the e-commerce platform with omni-channel operation capabilities, so that all the data related to the order can be shared and common. [2]

The e-commerce middle office management system developed by the project is built on the SpringBoot framework, the system's functional code is written in Java language, the system's operation interface is designed and implemented through VUE technology, and all the data of the system is accessed and invoked using the MySQL database, which can be provided to employees and administrators for two different users to use. [3] The ultimate development purpose of this project is to provide users with an e-commerce middle office management system that can operate stably and meet the basic usage needs, so that users can realize the digitization of the whole process of

DOI: 10.2478/ijanmc-2022-0014

department information, store information, warehouse information, commodity file information and order information through the e-commerce middle office management system, and provide certain data support for the user's business decisions.

The SpringBoot Development Framework is a lightweight framework that can support the development of commercial application systems, which can be combined with a large number of development frameworks to be used as a development technology for application system development process. The SpringBoot development framework also supports aspect programming, which can be introduced to directly automate the configuration of the project, saving a lot of time in environment configuration; in addition, it can also complete the call to a variety of interfaces, saving the resource occupation of the system itself; the program developed by the framework can be used with the Tomcat server to make its deployment more convenient.

Tomcat Server is a background server that allows users to deploy programs with only a few very simple configurations, greatly simplifying the difficulty of their deployment. After the enterprise copies the developed web system program directly to the relevant location of the Tomcat server, it can be run and provided to all kinds of users after a simple modification of the web .xml or related files.

VUE technology is a lightweight programming technology developed by using JavaScript for front-end interface development, which is mainly through HTML technology to complete the presentation of the interface, and the code programming is done by using VUE technology. In VUE technology, the association between the view and the model can be completed by using the provided ViewModel, and the binding and access to the data can be completed directly by modifying the logical business rules, so as to realize the display on the HTML interface. [4]

MySQL is a commonly used database system for web system development, which can not only complete the above work, but also use the provided visual management environment to complete the database table creation and execution of data queries and other related actions, of course, can also meet the basic needs of data processing in different languages.

The Java language is a high-level language based on the C++ programming language and introduced in response to the needs of multi-platform applications. The Java language can support the development of applications with interfaces; it can also develop web pages of applets for complete stage coding; it can take advantage of existing components and can run on multiple platforms with only one compilation, which greatly improves the efficiency of development programs.

The overall functional design of the e-commerce middle office management system is shown in Figure 1 below.

II. SYSTEM DESIGN

The e-commerce middle office management system is a web order management system that can be applied to the order aggregation management scenario, which can be provided to employees to complete registration and login, after logging in, you can view personal information and modify passwords in the personal center interface, employees can also view department information, store information and warehouse information, and can also enter order information and commodity file information into the system; after the administrator logs in, you can also manage department information, store and type

information. Order information and product file information entered into the system by employees can also be maintained and processed. Through the design of the above employee and administrator functions, a complete and operational e-commerce middle office management system is formed. [5] The employee registration process of the e-commerce middle office management system developed in this topic is that if any employee wants to enter the e-commerce middle office management system to view the warehouse information and add relevant information such as orders and commodity files, he must first select the registration button to enter all the registration information into the Chinese side of the system. During registration, employees are required to enter all relevant information such as employee job number and employee name into the edit box to complete the registration in time, and carefully fill in the password, contact number and position information. If one item is omitted, it will be prompted that this item cannot be blank. Until all the information is filled in and confirmed to be correct, click registration to register successfully, Only when any registration information is not entered, it cannot be registered successfully. The employee login process of the e-commerce middle office management system developed in this topic is to enter the system and complete the filling of main information. This step can be carried out only after the registration is successful. Specifically, you can enter the login information into the edit box, including user name, password and other relevant information, and select the user role, including administrator and employee module, If the e-commerce login information is not successfully entered or the password is not correct, the employee will be prompted to log in to the middle console. Otherwise, if the login information is not successfully entered or the login information is not correct, the employee will not be able to log in to the middle console.

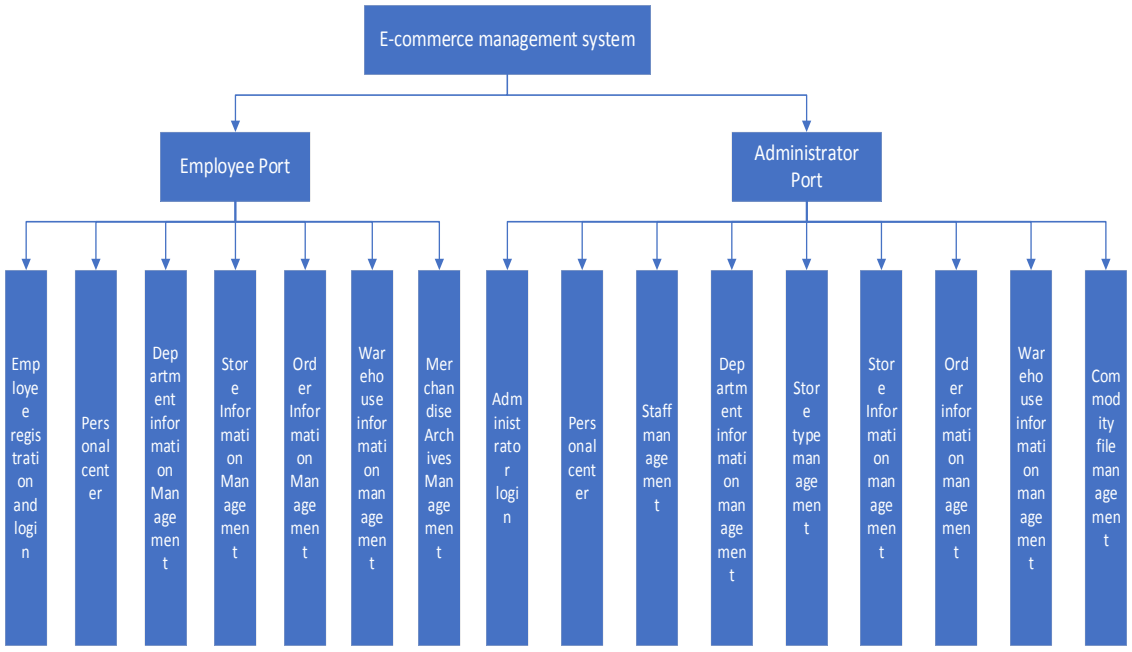


Figure 1. Overall system design structure diagram

The entry of the order information of the e-commerce middle office management system developed in this topic is that the employee first logs in to the e-commerce middle office management system, verifies that the order-related information is edited and entered into the edit box by selecting the order information management menu, and the successful submission of the order information cannot be completed without entering the order information into the system. After the employees enter the order information into the system, the administrator can complete the view of the order information entered into the system by all employees, and can also be maintained in the case that the order information such as the order quantity, the order price and other related information is entered incorrectly, and the order information can be cleared or modified from the list. When designing the database of an e-commerce middle office management system, the entity abstraction of the database is given through the proposed functional requirements. With the continuous and further improvement of the prototype diagram, the database can be partially modified and designed in time. It generally constructs the entities by using some functions, and these entities are generally the data structures corresponding to the functional interface, and the attributes connected with the entities are the data elements attached to these data structures. For the e-commerce middle office management system, it can include the following entities: employee entity, administrator entity, Department entity, warehouse entity and Order entity, store entity, store type entity, commodity file entity, etc. its design is as follows:Its order information management process is shown in Figure 2 below. [6]

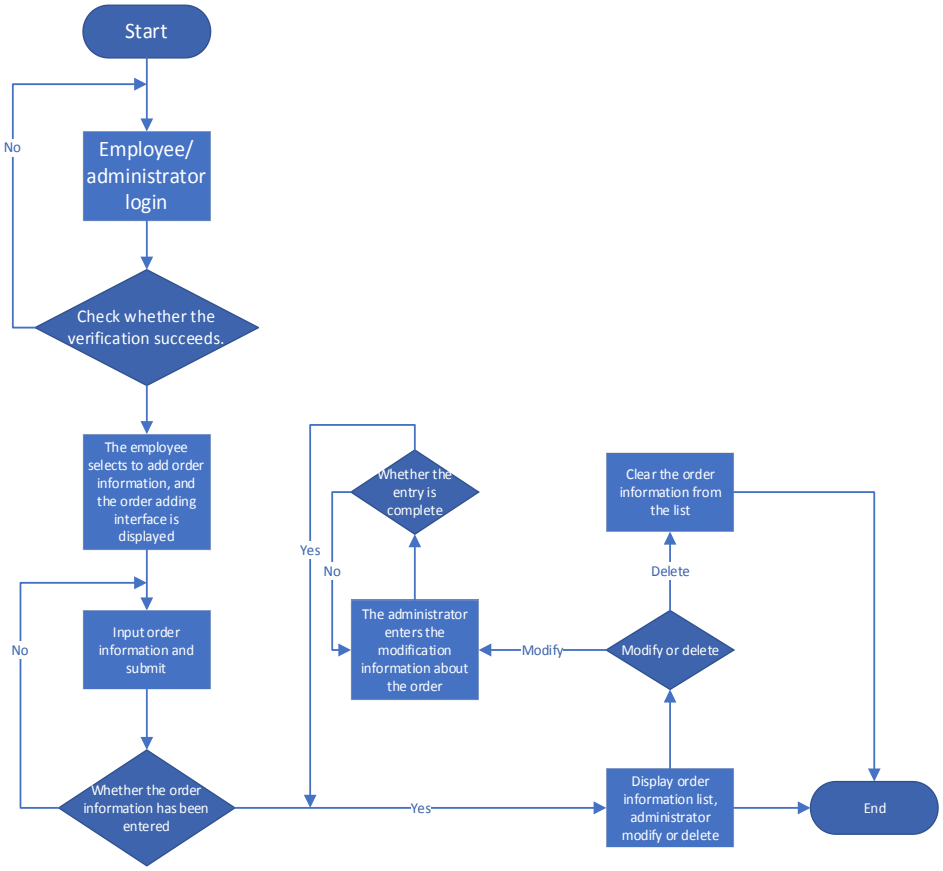


Figure 2. Order information management flowchart

III. System Implementation and Testing

The registration interface is to enter the employee number, employee name, password, contact number and position into the edit box, any of the above five kinds of information must be entered into the system, otherwise the employee is not allowed to register successfully. The administrator entity is specifically used to provide users with management authority with login verification information when they can manage and maintain a series of permissions such as employee, department information, store type, store information, order information, warehouse information and commodity file information when entering the e-commerce middle office management system. [7] The employee registration interface is shown in Figure 3 below.

Figure 3. Employee registration interface

Once an employee is registered, they can use the login screen to complete the verification, in which they can enter the login information into the edit box, select the role, and only if the login information matches the role can the login be successful. Department entity is an effective way to complete the ownership of the specific organizational structure of all employees using the system. It is also the most common and most basic organizational structure of an enterprise. Its department entity mainly includes department name, department director, department number, personnel structure and detailed establishment date. The employee login interface is shown in Figure 4 below.

Figure 4. Employee login screen

In order to provide employees with the safe use of the e-commerce middle office management system, provides a mechanism to modify the password to complete, here, to ensure that the original password is correctly entered into the system under the premise, and to ensure that the new password of the two consecutive entries are consistent in the case of the password can be successfully modified. Warehouse entity is a kind of address information used to record the goods sold by enterprises or stores for storage. Its attributes include: warehouse number, warehouse name, warehouse location, warehouse picture, warehouse area and stored goods.Its interface for modifying passwords is shown in Figure 5 below.

Figure 5. Modify the password interface

Information such as positions, contact numbers, etc., that may be entered by individual employees at the time of registration may be modified by the employees themselves. Its interface for modifying personal information is shown in Figure 6 below.

Figure 6. Modifying the personal information interface

Department information is a kind of information that can be released by the administrator through the system and provided to employees to view, at this time, employees can also quickly query department information and view the details of each department according to

the name of the department and the head of the department. [8] The store type entity is used to display the specific type information of the store where the employees using the system record the sold goods, whether it comes from JD store, Taobao store or other stores,Its interface is shown in Figure 7 below.

Figure 7. Departmental information management interface

Employees can view all the store information released by the administrator according to the actual store situation, in addition, employees can also view a certain store or a certain type of store by store name, store type, etc., and can also view the details of a store. The store entity is the entity information that can be used to sell goods through the e-commerce platform, and can also be used to provide services and order distribution for the sold goods, including the detailed number of the store, the name of the store, the type of the store, the exquisite photos of the store, the opening time of the store and the detailed business scope, The order is generated by the name of the store, the price of the order, the employee's contact number, and the order data recorded by the user, including the name of the store, the order number, and the order number of the order, The commodity file entity is used to record the detailed information of the commodity file provided to the user to purchase and the store can sell to the user, including: commodity name, commodity size, commodity storage location, commodity inventory, monthly commodity sales volume, commodity picture, registration time, employee account number, employee name and employee contact number, The store information list viewing interface is shown in Figure 8 below.

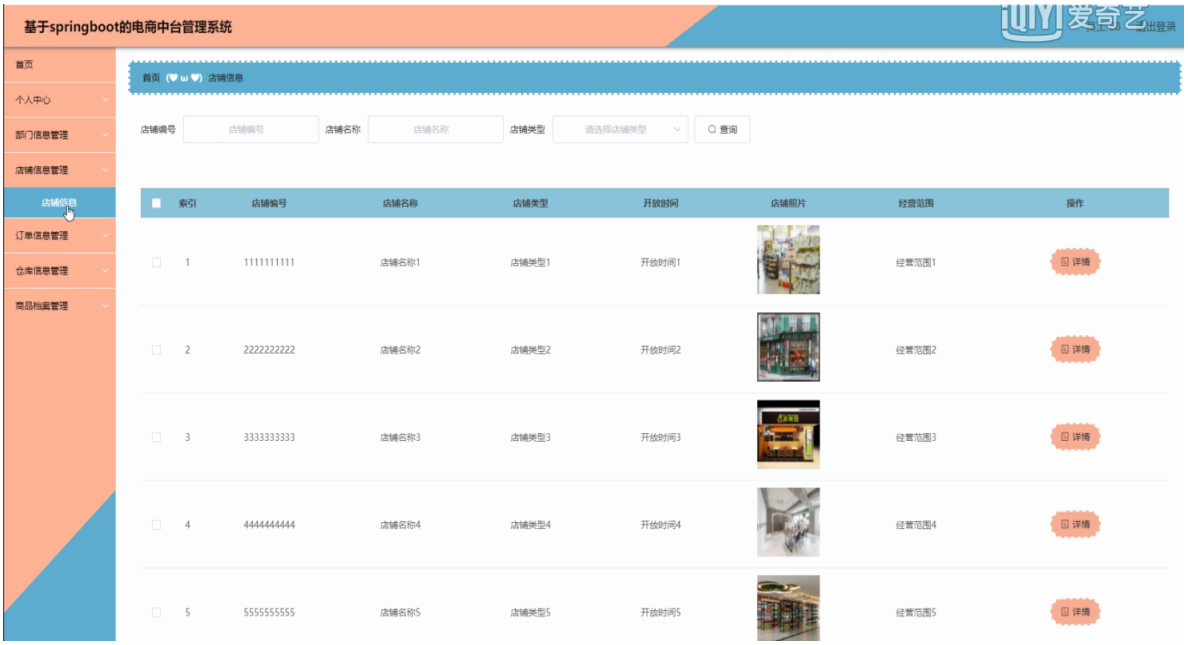


Figure 8. Store information list viewing interface

Order information management

Order information is an employee can be based on the actual order sales situation to enter the system of information, in the order information to add interface, the order quantity must be entered into the system, you can also directly select the store to complete the automatic appearance of the store data, in the order quantity is not entered into the system can not complete the successful submission of the order. Its order information addition interface is shown in Figure 9 below.

Figure 9. Order Information Addition Interface

After the employee enters the order information into the system according to the actual situation or the administrator maintains the order information according to the actual situation of the order, the order information can be viewed by the employee; in addition, the employee can also query the way to quickly find the order and view the details of an order. Its order information list interface is shown in Figure 10 below.

Figure 10. Order information list interface

Warehouse information is a kind of information related to the storage location of the goods that can be provided to employees in the system according to the actual storage of goods by the administrator to complete the information related to the storage location of the goods; employees can also query the details of a warehouse according to the name of the warehouse, location, etc. The commodity file entity is used to record the detailed employee entity of the commodity file provided to users to purchase and the store can sell to users. It is used to record users who use the e-commerce middle office management system to view department information, store information and warehouse information. It can also add order information and commodity file information, including employee job number, employee name, employee gender, employee avatar Employee contact number, employee position and other information, including: commodity name, commodity size, commodity storage location, commodity inventory, monthly commodity sales volume, commodity picture, registration time, employee account number, employee name and employee contact number, etc, The commodity file entity is used to record the detailed employee entity of the commodity file provided to users to purchase and the store can sell to users. It is used to record users who use the e-commerce middle office management system to view department information, store information and warehouse information. It can also add order information and commodity file information, including employee job number, employee name, employee gender, employee avatar Employee contact number, employee position and other information, including: commodity name, commodity size, commodity storage location, commodity inventory,

monthly commodity sales volume, commodity picture, registration time, employee account number, employee name and employee contact number. [9] After the above entities are designed, the data table can be designed according to the required attributes and the occupied space may be used, Provide necessary conditions for the creation of the actual database. E-commerce database management system Its warehouse information list viewing interface is shown in Figure 11 below.

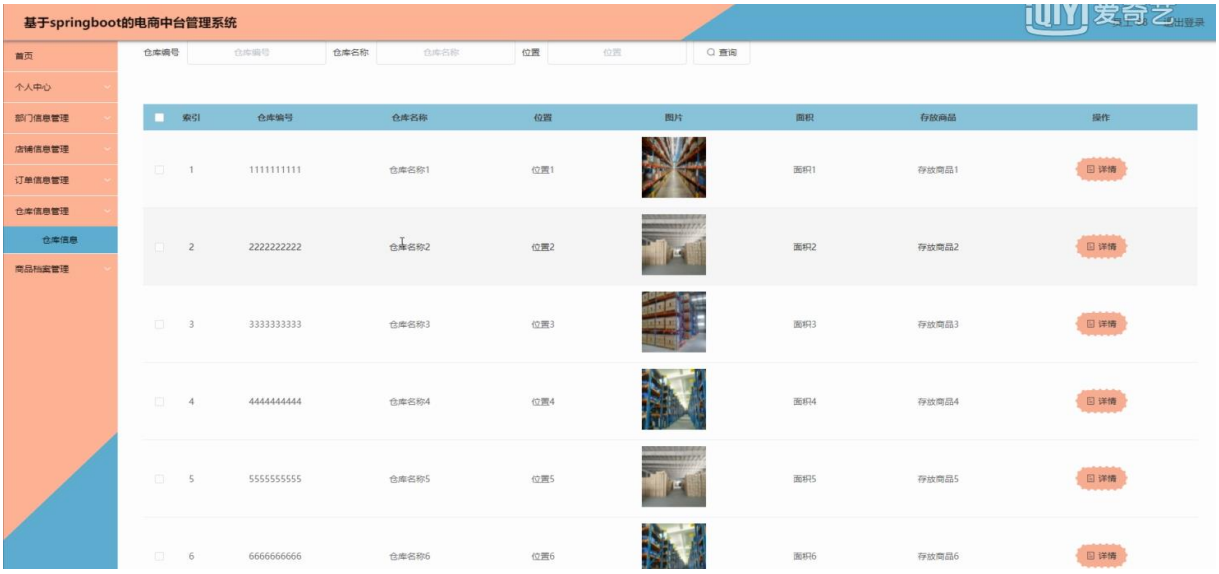


Figure 11. Repository Information List Viewing Interface

Commodity file is a staff can be based on the ability to sell the product information to complete the entry into the system, where the product name is necessary to enter the system, other product information can not be entered into the system temporarily, its product file information to add interface as shown in Figure 12 below.

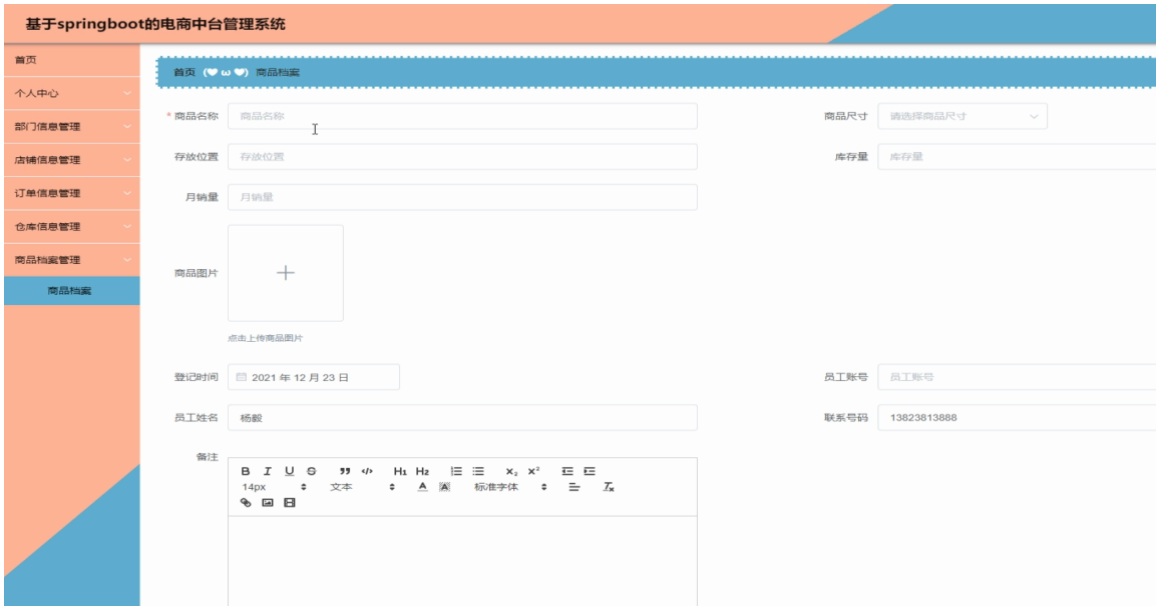


Figure 12. Product file addition interface

After the employee enters the product file information into the system or the administrator maintains the product file information, the employee can view the product file information list, and can also query and view the details of the product file. Its product file information viewing interface is shown in Figure 13 below.

Figure 13. Product file list viewing interface

Administrators can not only view the employee information of all employees who use the employee terminal of the e-commerce middle office management system to complete the registration, but also give employees a permission according to the actual situation of the employees and maintain and process them according to the errors in the employee's information such as contact numbers, positions, etc. Its employee management interface is shown in Figure 14 below.

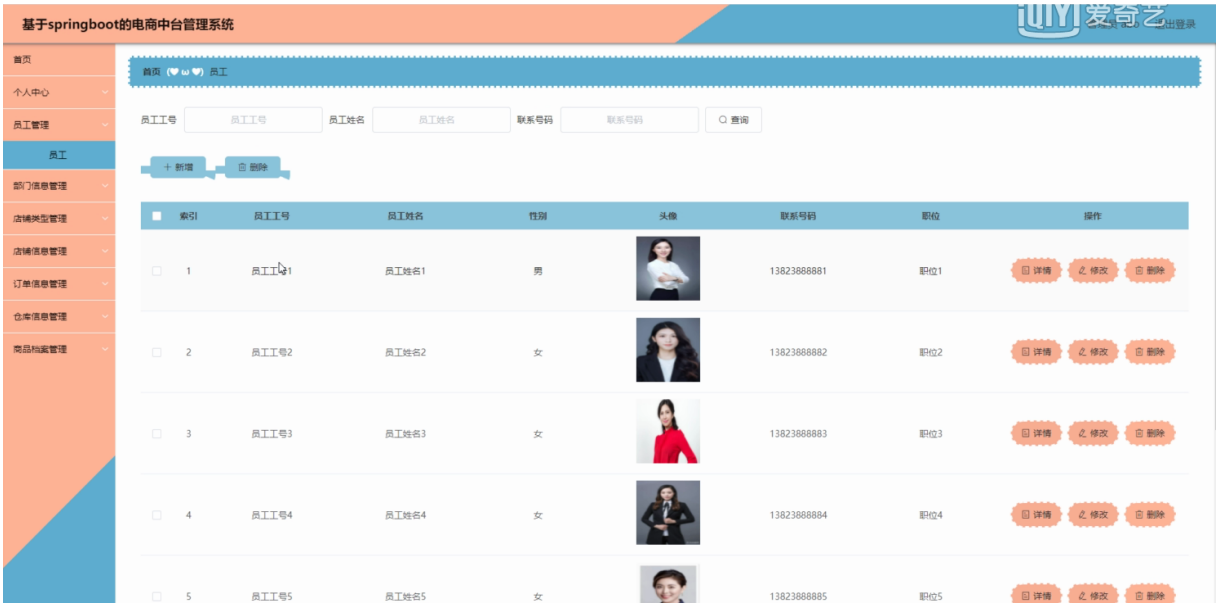


Figure 14. Employee management interface

Department information is based on the actual enterprise organization situation to enter the department name, department number and department head and other information into the system, and its department information addition interface is shown in Figure 15 below.

Figure 15. Department Information Add Interface

After the administrator has completed the most basic department information for the organizational structure of the enterprise, it can also be maintained in the case of structural adjustments and changes such as department name and department head. Its department management list interface is shown in Figure 16 below.

Figure 16. Department Management List Interface

Store information is based on the actual situation to the store information such as store name, business scope and other information to enter into the system, and then can be provided to employees to view, as long as the store information is not entered in the case can not publish store information.

After the store information is added to the system by the administrator, it can also be maintained according to the opening hours, business scope, and store photos of the store information. [10]

The administrator can view the list of order information entered into the system by all employees, and can also complete the maintenance of the order in the event that the order information is entered incorrectly, such as the order quantity, the order price, etc. [11]

The administrator can view the product file information entered by all employees into the e-commerce middle office management system; can also maintain the product file information in the case of incorrect entry such as product size, storage location, etc.; of course, can also view the detailed information of each product file. [12]

IV. CONCLUSION

The e-commerce middle office management system developed by this project is a web system for order management, which can be taken by means of collecting materials related to the subject through the library and the Internet, through the analysis and summary of these literature, to complete the analysis and database design of the functions and processes of the e-commerce middle office management system, and finally to use technology to complete the system used by employees and administrators.

The e-commerce middle office management system uses the Vue framework as the design technology of the front-end interface, and the background end uses the SpringBoot framework to build the system, uses the Java language to write the functional code used by employees and administrators, and stores the relevant data such as orders and commodity files using the MySQL database. The e-commerce middle office management system is a web system that can be used for the information associated with the commodity file and the order, which can be provided to employees and administrators to use, employees can view department information, store information, warehouse information, and can also add and query order information and commodity file information; can be provided to the administrator to manage the department information, store information and warehouse information, etc., and can also be added by the staff the order information and commodity file information to be maintained by the administrator in the case of changes in these information. Thus, it can be provided with a complete set of e-commerce middle office web management system.

REFERENCE

[1] Shufan Liu, Ximei Li, Peng Sun. Design and Implementation of Communication Base Station Survey System based on SpringBoot [J]. Proceedings of 2018 Academic Conference on intelligent education and artificial intelligence development, 2018:34-36.

[2] Zhang Shiying. Development and Implementation of College Students' Ideological and Political Practice Course Network Teaching Platform Based on Jsp Technology [J]. 2019 3rd International Conference on Advancement of the Theory and Practices in Education (ICATPE 2019), 2019:364-369.

[3] Big Data:Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce [J]. Sergio Ramírez-Gallego, Alberto Fernández, Salvador García, Min Chen, Francisco Herrera. Information Fusion. 2018:100-103.

[4] Analysis of user preference and expectation on shared economy platform: An examination of correlation between points of interest on Airbnb [J]. Moloud Abdar, Neil Y. Yen. Computers in Human Behavior. 2018:71-73.

[5] Liao Bin Analysis of MVC pattern in developing web application based on Java [J] Electronic technology and software engineering, 2020 (21): 49-50.

[6] Zhangsiqing Design and implementation of e-commerce adoption system for agricultural products based on J2EE [d] Anhui Agricultural University, 2021 (33): 113-115.

[7] Yang Jin Design and implementation of fresh e-commerce order fulfillment system [d] Beijing Jiaotong University, 2021 (20): 160-161.

[8] Wang Hao Design and implementation of e-commerce system for retail industry [d] University of Electronic Science and technology, 2021 (19): 37-38.

[9] Wu Tong Design and implementation of a commodity background management system for a small and medium-sized enterprise [d] University of Electronic Science and technology, 2021 (5): 87-89.

[10] Zhaoanxiue, hurui town Design and implementation of order management system based on JavaEE [j] Science and technology innovation and application, 2021 (04): 115-117.

[11] Lian Da, xiexiaoling, liupingping Design of e-commerce management system for rural economy [j] Automation technology and application, 2020 (12): 164-167.

[12] Songdapeng Development practice of large enterprise electronic mall system [j] China management informatization, 2020 (15): 179-182.

A Circuit Principle and Simulation Test for Negative Group Delay

Han Shen
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: sunny_shine_zj@163.com

Zhongsheng Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: wzhsh1681@163.com

***Abstract*—Group velocity superluminal phenomenon, also known as negative group delay or negative group velocity phenomenon, refers to a group of envelope signal at the output end of the medium before the input, in the time axis, is to leave the medium envelope appears before entering the medium, but this does not violate causality. Based on the waveguide theory of negative group delay, this paper uses the transfer function, amplitude response and phase shift response of the electronic circuit to control group velocity, and introduces three first-order bandpass amplifier RC circuits to control group velocity in low frequency band, which realize positive group delay, negative group delay and filtering functions respectively. Finally, the phenomenon of negative group delay is preliminarily realized by using circuit simulation software, which lays a foundation for subsequent research. As an envelope that can carry and transmit information, the study of negative group velocity is of great significance to the improvement of signal transmission efficiency.**

Keywords-Negative Group Delay; Band-pass Amplifier; Electronic Circuits; Transfer Function; Signal Delay Compensation

I. INTRODUCTION

In the mid-twentieth century, Brillouin and Sommerfeld showed that in irregular dispersion regions, group velocities can exceed the speed of light in a vacuum, and can even be negative [1]. For superluminal group velocities, the envelope takes less time to travel through the medium than it does for light to travel the same length in a vacuum. For the negative group velocity, what appears on the time axis is that the envelope as you leave the medium appears before you enter the medium.

T. Nakanishi, K. Sugiyama and M. Kitano [4] proposed a model to realize this phenomenon by using common electronic components, which mainly include pulse generator, bandpass amplifier, resistor, capacitor and LED. The circuits structure constituted by them is shown in Figure 1. From the perspective of pulse analysis, the time generator mainly generates a voltage signal, namely rectangular pulse, the LPF part of the low-pass filter will conduct pulse modulation to Gaussian pulse, and the ND part of the negative group delay will produce the phenomenon of phase advance and group velocity advance. On the time axis, the pulses at the C end appear before the pulses at the B end, but neither exceeds the actual pulse generator, the rectangular pulse at the A end, and thus does not violate causality. In addition, when the LEDs are connected at B and

DOI: 10.2478/ijanmc-2022-0015

C, the light on the C terminal is bright earlier than that on the B terminal.

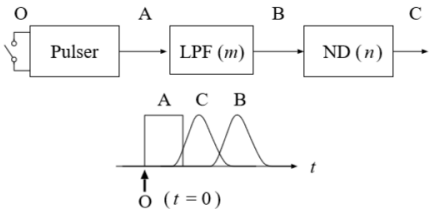


Figure 1. Experimental schematic diagram of negative group delay and the pulse relationship on the time axis

II. MATHEMATICAL PRINCIPLES

A. Principle of Negative Group Delay in Signal Transmission

Group velocity refers to the overall propagation velocity of wave train or the envelope propagation velocity of wave. In dispersive media, it is defined as:

$$v_g^{-1} = \frac{dk}{d\omega}|_{\omega_0} \tag{1}$$

Where $k = k(\omega)$ is the wave function related to frequency ω. In a dispersive medium, the phase shift $\phi(\omega)$ is related to the wave function as follows:

$$\phi(\omega) = -k(\omega)L \tag{2}$$

Where L represents the channel length of wave train signal transmission, and the delay of group velocity is defined as:

$$t_d = -\frac{d\phi}{d\omega}|_{\omega_0} \tag{3}$$

After substitution, it will be as:

$$t_d = -\frac{d(-kL)}{d\omega} = \frac{dk}{d\omega}L = v_g^{-1}L \tag{4}$$

Combined with signal propagation in vacuum, the total time of signal transmission is:

$$t_{total} = \frac{L}{c} + t_d \tag{5}$$

Combined with the definition of speed, it can be concluded that the actual speed of signal transmission in the channel is:

$$v_g = \frac{L}{t_{total}} \tag{6}$$

Then, the relationship between group velocity, light speed and propagation time can be as follows:

$$\frac{1}{v_g} = \frac{1}{c} + \frac{t_d}{L} \tag{7}$$

For the normal positive group delay state, $t_d > 0$, then $v_g < c$, the speed is less than the speed of light.

In an unconventional state, the output wave is required to arrive earlier than the input wave, that is, the group delay is required to be negative, that is, $t_d < 0$. Here we can also analyze two situations:

If $(-t_d) < L/c$, then $v_g > c$, which is superluminal in the narrow sense.

If $(-t_d) > L/c$, then $v_g < 0$. At this time, negative group velocity is realized, which is the main research objects of this paper.

B. Group Delay Principle in Electronic Circuits

The relationship between input and output waves in an electronic circuit is as follows:

$$v_{out}(t) = (h * v_{in})(t) = v_{in}(t - t_d) \quad (8)$$

Where $h(t) = \delta(t - t_d)$, is the impulse response, and t_d is the delay time of the output signal envelope.

Fourier expansion is used to obtain the following relations:

$$V_{out}(\omega) = H(\omega)V_{in}(\omega) \quad (9)$$

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-i\omega t_D} dt = \exp(-i\omega t_D) \quad (10)$$

When negative group delay (group velocity faster than light speed) is realized in electronic circuit, the fidelity of the original input signal should be as low as possible, and the phase shift should be easy to analyze. Therefore, the ideal amplitude should be equal to 1 and the phase shift should be linear. According to the characteristics of the electronic circuit, the amplitude and phase shift of the transfer function H(ω) of the circuit are defined respectively [5]:

$$A(\omega) \equiv |H(\omega)| = 1 \quad (11)$$

$$\phi(\omega) \equiv \arg H(\omega) = -\omega t_D \quad (12)$$

Combined with the definition of group delay, the group delay is:

$$t_d = -\frac{d\phi}{d\omega}|_{\omega_0} = t_D \quad (13)$$

C. Processing of Transfer Functions

In practical circuits, complex numbers are directly used to express circuit characteristics. Here, the mathematical relationship between transfer function, amplitude function, phase shift function and group delay is explained. Since the response function is an imaginary number, in combination with the way of deriving the transfer function in the actual electronic circuit, the expression can be transformed, then other parameters can also be transformed accordingly:

$$H(\omega) = a + ib \quad (14)$$

$$A(\omega) = |H(\omega)| = a^2 + \omega^2 b^2 \quad (15)$$

$$\phi(\omega) = \arg H(\omega) = \arctan(\frac{b\omega}{a}) \quad (16)$$

Under certain conditions, an approximation can be made as follows:

$$\lim_{b\omega/a \to 0} \phi(\omega) = \lim_{b\omega/a \to 0} \arctan(\frac{b\omega}{a}) = \frac{b\omega}{a} \quad (17)$$

$$t_d = -\frac{d\phi}{d\omega}|_{\omega_0} = -\frac{b}{a} \quad (18)$$

When both a and b are positive or negative, the case of negative group delay is realized.

III. THREE TYPICAL AMPLIFIER CIRCUITS

In order to design the electronic circuit for low frequency negative group delay phenomenon, the RC amplifier electronic circuit is mainly used, where the impedance function of capacitor C is related to imaginary number and frequency, and there is a relatively simple and suitable transfer function. This section describes the three simplest RC amplifier electronic circuits, and analyzes in detail how to calculate the circuit's transfer function, the associated amplitude and phase shift, and the effects of simulation tests. It should be noted that, in order to explain the mathematical principle of amplifier circuit, this section introduces the simplest circuit model, the simulation effect is not ideal, but there are obvious phenomena, the actual used circuit is expanded and applied on this basis.

A. A First-order Low-pass Filter Which Can Only Achieve Positive Group Delay

The simplest RC amplifier circuits, shown in Figure 3, is a first-order low-pass filter, but can only achieve the function of positive delay.

The bandpass amplifier itself has three interfaces, which are positive interface, negative interface and output connection interface. Amplifiers with five interfaces can also be found in practical applications. In analog software for electronic circuits and in practical use of electronic components, the amplifier has five interfaces, and two additional interfaces are used to connect the amplifier's driving power supply. In the actual experiment, the driving power of the amplifier needs to consider how to set up and how to set up in the test platform, but it does not affect the analysis of the circuit effect in the theoretical stage, and this part will not be discussed in the subsequent analysis. Generally, the positive interface of the amplifier is mainly connected with the input end and its related electronic components, and the negative interface is mainly connected with the output end and its related electronic components.

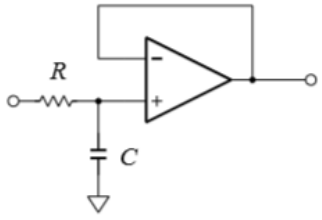


Figure 2. A first-order RC circuits with positive group delay

The transfer function of the electronic circuits shown in Figure 2 is:

$$H(\omega) = \frac{V_{out}}{V_{in}} = \frac{1}{1+i\omega RC} = \frac{1}{1+i\omega T} \tag{12}$$

As can be seen from the above analysis, the parameters related to frequency ω will finally be presented in the calculation results of group delay. Considering the characteristics of the circuit itself, let $T = RC$ be called the time constant of the circuit, that is, the relation between the transfer function and frequency and time constant is as follows:

$$H(\omega) = \frac{1}{1+i\omega T} = \frac{1-i\omega T}{1+\omega^2 T^2} \tag{13}$$

The amplitude and phase shift are:

$$A(\omega) = |H(\omega)| = \sqrt{\frac{1}{1+\omega^2 T^2}} \tag{14}$$

$$\phi(\omega) = \arctan(-\omega T) \tag{15}$$

For convenience of estimation, the amplitude response when $\omega T \sim 0$ is approximately 1 (slightly less than 1), that is, the shape of the input pulse does not undergo much deformation, and the approximate group delay thus obtained is:

$$t_d = -\frac{d\phi}{d\omega}|_{\omega_0} \sim T \tag{16}$$

Because $T = RC > 0$, it is positive group delay. Even without any restriction on frequency, the phase shift is always positive. Therefore, the circuit design can only get positive group delay in the whole frequency band. With the increase of input pulse frequency, the phase shift function can no longer be approximated as a linear function, and the delay time becomes smaller and smaller and approaches 0, but will not be negative.

The simulation effect of the circuit is shown in Figure3. Red is the input pulse and blue is the output pulse. It can be seen that the output signal does delay on the time axis.

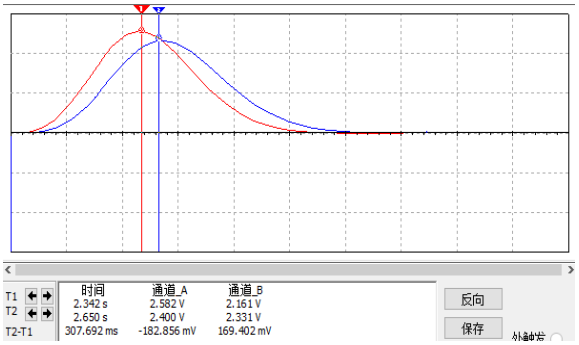


Figure 3. Simulation effect of first-order RC circuits with positive group delay

B. A First-order Low-pass Filter with Negative Group Delay

Figure 4 shows another simple first-order low-pass filter, which can realize the negative group delay phenomenon. The circuit characteristics and transfer function characteristics are analyzed in detail below.

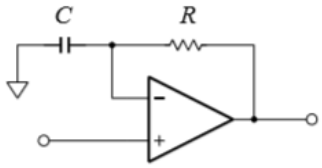


Figure 4. A first-order RC circuits with negative group delay

The transfer function, amplitude and phase shift of the circuit are:

$$H(\omega)=\frac{V_{out}}{V_{in}}=1+i\omega T \tag{17}$$

$$A(\omega)=|H(\omega)|=\sqrt{1+(\omega T)^2} \tag{18}$$

$$\phi(\omega)=\arctan(\omega T) \tag{19}$$

According to the calculation method of group delay, the group delay is negative, that is, the negative group delay is realized, $\omega \ll 1/T$, amplitude response approximately 1, no large deformation of the input signal is generated, and the phase shift can be approximated as follows:

$$\phi(\omega)\sim\omega T \tag{20}$$

$$t_d\sim -T \tag{21}$$

After taking the first derivative of frequency, the group delay is:

This is the simplest kind of negative group delay circuits. As the frequency of the input signal increases, the effect of negative group delay becomes smaller and smaller, and finally approaches 0.

Figure 5 shows its simulation effect under sine. Red is the input pulse, blue is the output pulse, from the time axis, the output pulse does arrive earlier than the input pulse effect. However, it should also be noted that the first-order filter has a large distortion problem, and it is found in the test that the distortion of the first-order circuit at high frequency is almost destructive. Therefore, in practical application, this circuit cannot be directly used, and further modification is needed.

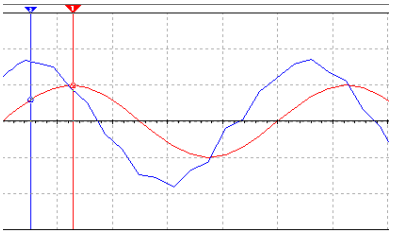


Figure 5. Simulation effect of first-order RC circuits with negative group delay

C. A First-order All-pass Filter that can be used to Modulate An Input Pulse Signal

As shown in Figure6, the simplest all-pass filter that can filter rectangular signals is introduced to generate input signals with different requirements in the circuit. It can only achieve positive group delay by itself, but the input pulse generated thereby will play a certain role in future

exploration and analysis of group favela phenomenon.

This circuit is a little more complicated than the first two, in that the input is connected not only to the ground terminal and the amplifier positive, but also to the amplifier negative, so the current direction in the circuit needs to be carefully analyzed in order to analyze the voltage situation. The signal entering from the input terminal is divided into two parts at the first node, one goes right to the resistor R, capacitor C and ground terminal, and the other goes up to the two series resistors R1 and the output terminal.

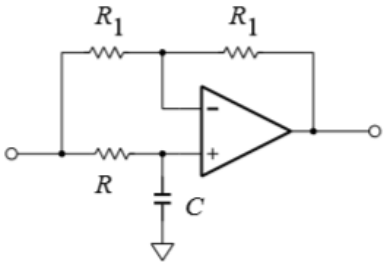


Figure 6. A first-order filter for modulating the input pulse

As shown in Figure 6, is the first-order all-pass filter, and its transfer function is:

$$H(\omega)=\frac{V_{out}}{V_{in}}=\frac{1-i\omega T}{1+i\omega T}=\frac{1-\omega^2T^2-2i\omega T}{1+\omega^2T^2} \quad (22)$$

Its amplitude response is exactly 1, which requires no approximation, but the phase shift is complicated:

$$A(\omega)=1 \quad (23)$$

$$\phi(\omega)=\arctan\left(\frac{-2\omega T}{1-\omega^2T^2}\right) \quad (24)$$

Computes the group delay is always greater than zero, so this is a positive group delay circuit, mainly used for filtering, can be simple to use stable voltage signal of rectangular pulse, pulse signal is converted to a certain, as shown in Figure7, red for the input of rectangular pulse, blue for the output pulse, visible had made a certain processing of signal, on the basis of optimizing circuit, Capable of generating a Gaussian pulse that can be used for subsequent testing.

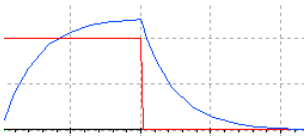


Figure 7. Simulation results of first-order filter modulating rectangular pulse

IV. Design and Analysis of Electronic Circuits

Referring to the model proposed by T. Nakanishi, K. Sugiyama and M. Kitano [4], as shown in Figure 8, a pulse generator, two second-order Bessel low-pass filters and two negative group delay circuits are successively connected in the circuit. Two LEDS are connected before and after the two negative group delay circuits to display the phenomenon that the output end is brighter than the input end. In order to analyze the signal condition, two LED parts are connected to the same oscilloscope in software simulation [3].

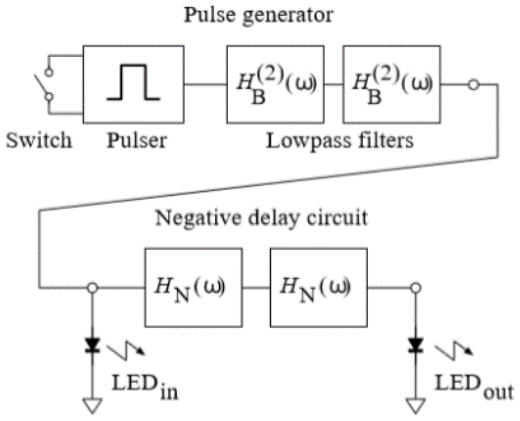


Figure 8. The electronic circuit realizing the negative group delay phenomenon

A. Pulse Generator

The function of the first part of the pulse generator is to generate a suitable Gaussian pulse, which is mainly composed of a rectangular pulse generator and two second-order Bessel filters.

After the switch is closed, a rectangular pulse is generated, and then a Gaussian pulse is generated by filtering through the filter.

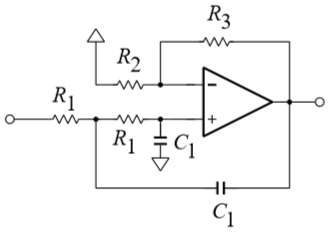


Figure 9. Second-order Bessel filter

The analysis focuses on the parts of two second-order Bessel filters, which filter the rectangular pulse into approximate Gaussian pulse in the circuit, which can carry the signal in application. As shown in Figure9, similar to the all-pass filter analysis method introduced in the previous section, its transfer function is:

$$H_B^{(2)} = \frac{1}{1+i\omega\alpha T+\frac{1}{3}(i\omega\alpha T)^2} \tag{25}$$

Particularly, $T = 1.272R_1C_1$, $\alpha = \frac{R_3}{R_2} = 0.268$

B. Negative Group Delay Circuits

Based on the second circuit in the previous section, C' and R 'are added to optimize, suppress amplitude gain in the transfer function and, more importantly, suppress the large distortion of the signal caused by fast gain at high frequencies. The negative group delay circuit used in this design method is shown in Figure 10 [2]. According to the above calculation method, the transfer function can be obtained as follows:

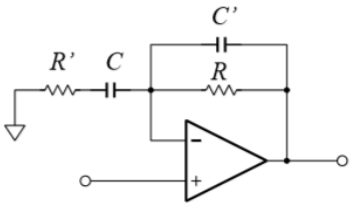


Figure 10. Negative group delay electronic circuit

$$H(\omega) = \frac{V_{out}}{V_{in}} = 1+\frac{i\omega T}{(1+i\omega aT)(1+i\omega bT)} \tag{26}$$

Particularly, $T = RC$, $a \equiv C'/C$, $b = R'/R$, note $C' \ll C$, $R' \ll R$ when selecting component parameters to facilitate approximate calculation.

$$H(\omega) \sim 1+i\omega T \tag{27}$$

$$A(\omega) = |H(\omega)| \sim 1 \tag{28}$$

$$\phi(\omega) = \arg H(\omega) \sim \omega T \tag{29}$$

In the lower frequency band, the amplitude response is close to 1, the input pulse basically does not produce deformation, and the negative group delay is close to T. In order to facilitate estimation, the approximate response function is equal to the simple circuit before modification, but in fact, due to the existence of a and b, the high-frequency part of the input pulse can be better suppressed, which makes the pulse fidelity better, and the phenomenon of negative group delay more stable.

V. SIMULATION

According to the above analysis results, choose the appropriate electronic simulation software, build the circuit in the simulation software, test its pulse advance phenomenon, and do a preliminary comparative experiment. Theoretically, a negative group delay circuits can get about 220 ms ahead, and two circuits in series can get about 440 ms ahead.

A. Circuit Structures

In accordance with the way of circuit construction, the effect of circuit is simulated in the circuit simulation software, and the input and output of the negative group velocity delay part are connected to the same oscilloscope, so as to analyze the sequence of input pulse and output pulse on the same time axis. Figure11 shows the electronic circuit built in the software simulation. All the electronic components are set in the ideal state. Two second-order Bessel filters are connected in series in part of the low-pass filter, and two negative group delay circuits are also connected in series. Figure12 shows the oscilloscope display, in which the red pulse is the rectangle pulse initially generated; The green pulse is the Gaussian pulse generated after the filter, that is, the input pulse of the negative group delay circuits. The blue pulse is the output pulse.

As can be seen by naked eyes, on the time axis, the output pulse arrives earlier than the input pulse. Using the ruler to test, it can be found that the peak value comparison between the input pulse and output pulse does produce about 0.5 seconds earlier, which is more consistent with the theoretical value. It can also be noted that neither the input pulse nor the output pulse precedes the rectangular pulse that originally generated the signal, so there is no violation of causality.

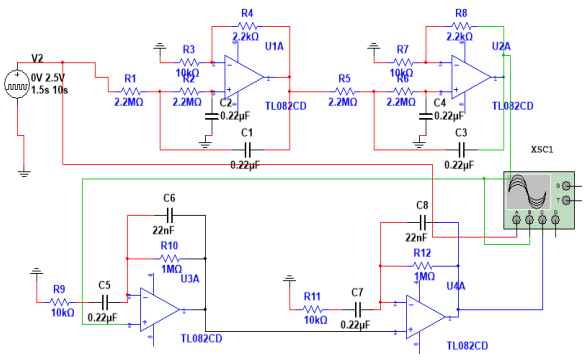


Figure 11. Electronic route built in simulation software

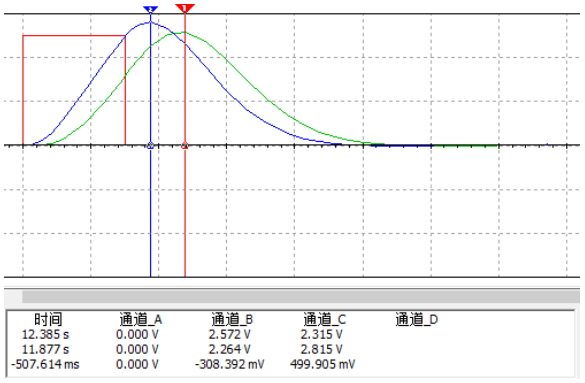


Figure 12. Simulation results of negative group delay phenomenon

B. Change the Number of Filter Ratio Test

On the basis of the above circuits, the number of Bessel filters and negative group delay electronic circuits were simply adjusted to simulate the experimental effect of the circuit, and the changes in advance and pulse peak values were compared. The experimental results are shown in Table 1. Where BF represents the number of Bessel filters, NGD represents the number of negative group delay circuits, NDT represents the time of negative delay, IPP represents the peak value of output pulse, OPP represents the peak value of output pulse, and AMP represents the amplitude response of the transfer function [7].

TABLE I. SIMULATION RESULTS OF ADJUSTING THE RATIO OF THE NUMBER OF BESSEL FILTERS AND NEGATIVE GROUP DELAY CIRCUITS

BF	NGD	NDT/ms	IPP/V	OPP/V	AMP
2	1	214	2.577	2.685	1.042
2	2	427	2.577	2.81	1.090
2	3	615	2.577	3.007	1.167
2	4	803	2.577	3.215	1.248
3	1	205	2.677	2.754	1.029
3	2	479	2.677	2.837	1.060
3	3	615	2.677	2.946	1.101
3	4	821	2.677	3.053	1.140
3	5	1027			
3	6	1164			

4	1	256	2.903	2.968	1.022
4	2	462	2.903	3.036	1.046
4	3	701	2.903	3.116	1.073
4	4	889	2.093	3.187	1.098
4	5	1068			
4	6	1099			
4	7	1294			
5	1	205	3.229	3.288	1.018
5	2	462	3.229	3.351	1.038
5	3	752	3.229	3.432	1.063
5	4	906	3.229	3.48	1.078
5	5	1050			
5	6	1205			
5	7	1378			
6	1	222	3.652	3.709	1.016
6	2	427	3.652	3.767	1.031
6	3	701	3.652	3.829	1.048
6	4	940	3.652	3.902	1.068
6	5	1022			
6	6	1265			
6	7	1425			
7	1	239	4.179	4.236	1.014
7	2	444	4.179	4.295	1.028
7	3	718	4.179	4.358	1.043
7	4	889	4.179	4.46	1.067
7	5	933			
7	6	1134			
7	7	1283			
8	1	205	4.827	4.886	1.012
8	2	444	4.827	4.946	1.027
8	3	684	4.827	5.008	1.037
8	4	855	4.827	5.064	1.049
8	5	1004			
8	6	1199			
8	7	1283			
9	1	239	5.618	5.678	1.011
9	2	444	5.618	5.74	1.022
9	3	615	5.618	5.803	1.033
9	4	889	5.618	5.872	1.045
9	5	1060			
9	6	1186			
9	7	1345			
10	1	251	6.503	6.565	1.010
10	2	474	6.503	6.63	1.020
10	3	669	6.503	6.697	1.030
10	4	865	6.503	6.762	1.040
10	5	1088			
10	6	1199			
10	7	1381			

The number of Bessel filters changed from 2 to 10, and the number of negative group delay circuits changed from 1 to 7. When there are more than 4 negative group delay circuits, pulse deformation and resonance interference are very serious, and it is difficult to measure the peak value of output pulse. Therefore, amplitude changes are only measured to 4 negative group delay circuits. In addition, when the Bessel filter is small, too many negative group delay circuits will also make it impossible to measure, so the data is not completely complete, but it does not affect the observation and analysis of the general trend.

The lead time and amplitude of the experiment were plotted to analyze the initial effect of the two parts of the circuit [6].

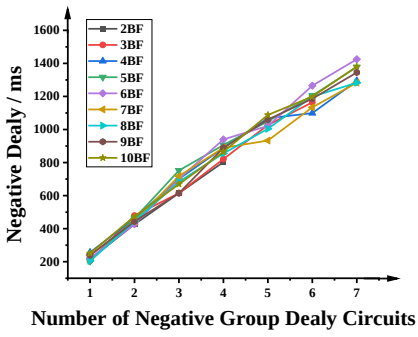


Figure 13. The variation of delay time with the number of circuits under different Bessel filters numbers

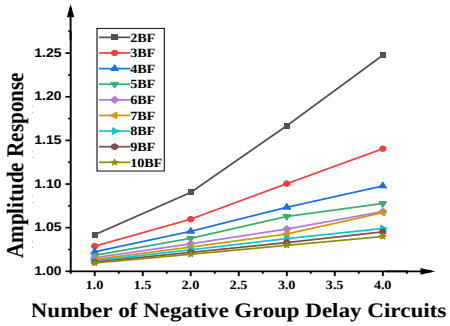


Figure 14. Amplitude response varies with the number of circuits under different Bessel filters numbers

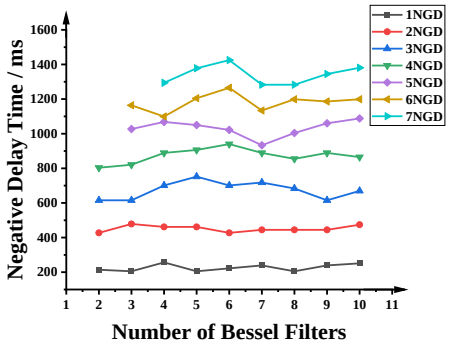


Figure 15. The change of negative delay time with the number of Bessel filters under different number of negative group delay circuits

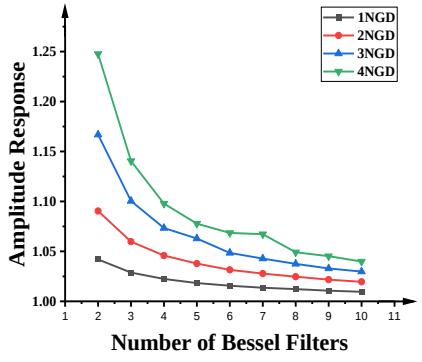


Figure 16. The amplitude response varies with the number of Bessel filters under different number of negative group delay circuits

Figure 13 and Figure14 show the changes of negative delay and amplitude response as the number of negative group delay circuits increases under different number of Bessel filters.

Figure13 presents the result of analysis, with the increase of number of negative group delay circuit, negative time delay value is higher and higher, and its value will not because of the change in the number of Bessel filter showed significant difference between the two, that is, under the condition of the input pulse conform to the requirements of the circuit, the more negative group delay circuit, the negative effect of time delay, the better, The main variable of the specific delay value is the number of negative group delay circuits, and the relevant value of input pulse is not the main influencing factor.

Analysis of the relationship presented in Figure14 clearly shows that the amplitude response relationship has different performance. The number of negative group delay circuits and the number of Bessel filters are very important. With the increase of the number of negative group delay circuits, the impulse response increases obviously, but with the increase of the number of Bessel filters, the increase of impulse response will be inhibited. It can be seen that both the parameter characteristics of the negative group delay circuits and the important parameters of the input pulse have a great influence on the amplitude response of the circuit. In the specific simulation process, with the increase of negative group delay circuits, the distortion of the output pulse becomes more and more obvious. When there are more than four negative group delay circuits, the pulse has shown a large deformation, and more than seven circuits are connected in series, and the shape of the pulse can not be identified.

Figure 15 and 16 show the changes of negative delay time and amplitude response as the number of Bessel filters increases in the case of different number of negative group delay circuits. The results presented match those of the first two

figures. The number of Bessel filters reflects the difference of input pulse. Under the condition of the circuit, it has no direct influence on the negative delay time, but has a more obvious influence on the amplitude response. The number of negative group delay circuits represents the characteristics of negative group delay electronic circuits (namely the transfer function), which has a very obvious effect on the negative delay time and amplitude response.

The Bessel filter part is equivalent to the actual process of modulating a signal, generating an input pulse; The subsequent negative group delay circuits is the main research part of negative group delay and faster-than-light phenomenon. It is worthy of further analysis in subsequent experimental exploration and practical application, especially the effect of parameters and number of negative group delay circuits, which is the most important influencing factor.

C. LED Effect Simulation

The input and output ends of the analog circuit are respectively connected with LED, and the sequence of lighting is recorded by recording the screen time axis. The results are shown in Figure17. The input lamp represents the input end of the negative group delay circuits, the output1 lamp represents the output end of one negative group delay circuits, and the output2 lamp represents the output end of two negative group delay circuits. The four-digit time axis is hour, minute, second, and sixtieth of a second respectively. When converted into milliseconds, the data in Table 2 can be obtained. It can be seen that the simulation effect of LED has a relatively obvious advance, and it matches the theoretical calculation and oscilloscope measurement results.

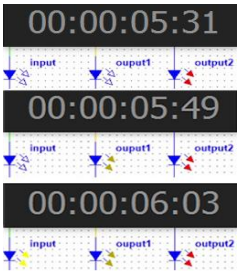


Figure 17. LED simulation results

TABLE II. LED SIMULATION RESULTS CONVERSION

Port	Time Code	Unit Conversion /ms	Time Delay/ms
Input	06:03	6050.000	
Output1	05:49	5816.667	-233.333
Output2	05:31	5516.667	-533.333

VI. CONCLUSION

Faster-than-light or negative group velocity pulse propagation is contrary to conventional wisdom and can be misleading [8]. Einstein's theory of relativity states that nothing can travel faster than light in a vacuum. But negative group velocities are the result of interference between waves of different frequencies, consistent with relativistic causality. Using the negative group delay circuits, the rising part of the input Gaussian pulse is increased, while the falling part of the Gaussian pulse is suppressed. The pulse shape can be maintained, but the output wave arrives earlier than the input wave in terms of time.

This paper mainly introduces the principle and simulation test of a RC negative group delay circuits with good demonstration effect, which can realize the phenomenon of signal advance of second magnitude in low frequency band. Among them, the first order positive delay, the first order negative delay and the transfer function calculation of the first order filter are analyzed in detail, in order to facilitate the actual use of more complex circuit analysis. Finally, the electronic circuit is tested by software simulation, and a good phenomenon is obtained. This paper mainly

focuses on the analysis of the circuit principle. After simulating the signal modulation under the actual situation, negative group delay is used to modulate the phase advance effect, etc. For other principles contained in it, further research and testing will be carried out in the future.

The phenomenon of negative group delay can be understood as superluminal in a broad sense, but its main principle is the advance of signal phase and the advance of group velocity of envelop signal. Because envelop can carry certain information, this research is of great significance to the compensation of time delay in signal transmission.

REFERENCES

[1] Brillouin L. Wave propagation and group velocity. NewYork: Academic Press, 1960.

[2] M.Kitano, T.Nakanishi, and K, Sugiyama. Negative Group Delay and Superluminal Propagation:An Electronic Circuit Approach. IEEE Journal of Selected Topics in Quantum Electronics, 2003, 9 (1) :43-51.

[3] Morgan W. Mitchell, Raymond Y. Chiao. Negative group delay and "front" in a causal system: An experiment with very low frequency bandpass amplifiers. Physics Letters A,230(1997):133-138.

[4] T.Nakanishi, K.Sugiyama, and M.Kitano. Demonstration of negative group delays in a simple electronic circuit. Am. J. Phys. 70(11), November, 2002:1117-1121.

[5] Hua Cao, Arthur Dogariu, and L.J.Wang, Negative Group Delay and Pulse Compression in Superluminal Pulse Propagation. IEEE Journal of Selected Topics in Quantum Electronics, 2003, 9 (1):52-58.

[6] Huiling Mao. Studies on Group Delay and Signal Fidelity in Negative Group Delay Circuit [D]. Zhejiang University.

[7] Huiling Mao, Linhua Ye, Li-Gang Wang. High fidelity of electric pulses in normal and anomalous cascaded electronic circuit systems [J]. Results in Physics, 2019, 13.

[8] HUANG Zhi-xun. The Achievements and Problems of the Superluminal Light Physics [J]. Journal of Communication University of China (Natural Science), 2013, 20(06):1-19.

3D Reconstruction System Based on Multi Sensor

Fan Yu
School of Computer Science and Engineering
Xi'an Technological University
No.2 Xuefu Middle Road, Weiyang district,
Xi'an, Shaanxi, China
E-mai: yffshun@163.com

Xue Fan
School of Computer Science and Engineering
Xi'an Technological University
No.2 Xuefu Middle Road, Weiyang district,
Xi'an, Shaanxi, China
E-mai: startfanxue@163.com

***Abstract*—In the 3d dense map construction system of indoor scene by mobile robot, the existing single sensor method cannot improve the positioning accuracy and reconstruction accuracy of robot, as well as the requirement of rapidity. Therefore, it is applied to THE ORB-SLAM with three parallel threads of track tracking, map reconstruction and loopback detection. Through depth camera pose to splice point cloud of building three-dimensional dense point cloud, in the 3 d reconstruction, a computer can not rely on GPU parallel computing, using only the CPU recovery environment three-dimensional dense scene map method, further reduce the time of the map construction, improve the efficiency of the reconstruction, thus improve the overall performance. Since only the ORB features were retained in the map during the construction of ORB-SLAM2, sparse point cloud map was established. Fortunately, the framework structure of ORB-SlAM2 was relatively clear. Only one thread needed to be added for the maintenance of point cloud map, and the key frames generated by ORB-SLAM2 were passed into the point cloud map construction thread. Use incoming keyframes to generate a map with dense point clouds.**

Keywords-Mobilerobot; Three-Dimensional Reconstruction; The ORB-SLAM; Depth Sensor; Inertial Sensor

I. INTRODUCTION

3D Reconstruction refers to an important work in the field of 3D computer vision, which aims to restore and reconstruct some 3D objects or 3D scenes. The reconstructed models can be easily processed by computers and expressed digitally. 3d reconstruction technology is the key technology to establish and express the objective world by using computer, including object reconstruction in 3D scene, dynamic reconstruction of human body, reconstruction of large buildings and so on. 3D reconstruction technology through the acquisition of depth data, pre-processing, feature extraction and matching, point cloud fusion reconstruction and other processes. Is to depict real scenes as mathematical models that can be logically expressed by computers, The application fields cover cultural relic protection, game development, architectural design, aerospace, shipbuilding, archaeology, robotics, virtual reality (VR), augmented reality (AR), reverse engineering, computer animation, surveying and mapping industry, engineering measurement, clinical medicine and other research to play an auxiliary role. The 3D reconstruction algorithm based on SLAM takes the camera as environment sensing SLAM and turns it into visual SLAM. In this way, it not only focuses on scene reconstruction but also can get the motion track of the camera when acquiring image data. It is a 3D reconstruction algorithm compatible with positioning and mapping, and has a good application prospect.

II. THE ORB - SLAM WERE REVIEWED

The ORB-SLAM2 (Oriented Fast View Brief SLAM2) is a monocular visual SLAM system based on image features and nonlinear optimization, including ORB feature extraction, ORB image feature bag for location recognition and loopback detection, information association view, G2O image optimization general framework. Scale - aware loop detection is applied to large-scale map construction. This algorithm only uses ORB as feature detection and description in all the processing, which can improve the effect of location recognition and loopback detection. This paper summarizes three important steps of ORB-SLAM algorithm: tracking thread, local map

thread, closed loop detection thread and dense map construction thread. As shown in Figure 1.

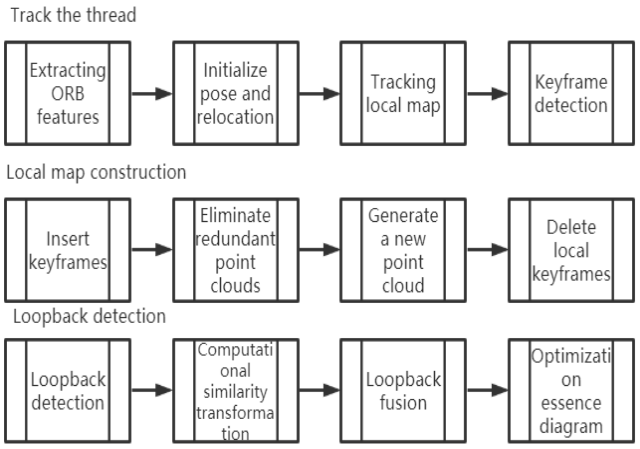


Figure 1. Key threads for ORB-SLAM

A. Track the thread

In tracking thread part mainly through the depth sensor RGB camera ORB feature extracting and feature matching, and triangulation to establish a link between characteristics of map point depth and 3 d coordinate, follow-up images by tracking motion model, reference key frames and relocation camera pose, with a minimum weight projection error to optimize the current position, Then determine whether to generate a new key frame according to the preset conditions.

B. Local map thread

Local map thread part is mainly on camera RGB camera get key frames is dealing with the key frames tracking thread creation, update the map point and the corresponding relationship between the key frames, delete maps the newly added observations are less map point, then to the high degree of common view of key frames by triangulation restores map, Repeat map key points of key frames and adjacent key frames are checked. When all key frames in the key frame queue are processed, local trapped set adjustment is performed for the current key frame, adjacent key frames and observed map points. Finally, the pose of key frames and map point accuracy are optimized by minimizing reprojection error. In the process of map building, the whole map building process is divided into three parts: the first part is the driving part, responsible for driving the sensor, here is replaced by the data set; The second part is pose estimation. ORB_SLAM2 obtains the camera data and outputs the camera pose. Finally, 3d scene recovery is realized by drawing part to provide environment map for robot navigation in complex scene.

C. Closed loop detection thread

Local map processing thread insert key frames, mainly includes three processes, respectively is closed loop testing, calculation of similarity transformation matrix, and the closed loop correction, the closed-loop detection is by calculating the word bag similar score candidate key frames, then for each key frame to calculate similarity transformation matrix, by random sampling to select the optimal consistency of key frames, Finally, the key frame pose was optimized by the Essential Graph, and the global trapped set adjustment was performed to obtain the global consistent environment map and camera running track.

D. Dense drawing threads

Dense built figure thread is to perform map point depth range search key frames, then match the price within the depth range, perform keyframe the initial depth, according to the principle of similar depth of adjacent pixels to obtain the initial depth map for inverse depth fusion adjacent pixels and filling vacant pixels, through the depth of the adjacent key frames optimal depth information fusion, Furthermore, the final depth map was obtained by intra-frame filling and external point elimination. Finally, the 3d reconstruction of indoor environment was obtained by point cloud splicing. The specific steps for dense reconstruction are as follows:

Step 1: Estimation of scene depth range: The image obtained by the RGB camera in the depth camera is input as the key frame, and every map point observed by the key frame at any time is projected into the key frame image. The depth value of the map point under the coordinates of the key frame is calculated, and the maximum or minimum depth is selected to set the inverse depth search range of the scene (ρmin, ρ Max).

Step 2: Feature point matching: extract special points in the process of camera movement.

Step 3: Fill and eliminate the depth map by smoothing inside the key frame and eliminating the outer points.

Step 4: Optimize the depth information of the key frame on the basis of the position and pose of the key frame calculated by the tracking thread.

Step 5: Smooth the key frame, remove the outer points, and fill and eliminate the depth map obtained from the depth data.

III. RGBD+IMU SENSOR INFORMATION FUSION

The dense 3D point cloud reconstruction based on the fusion of DEPTH sensor and inertial data in THE ORB-SLAM2 algorithm adds angular velocity and acceleration measurement information of IMU data in addition to input RGB data and DEPTH data. RGB data and IMU data are tightly coupled. And adopt the method of optimization for the ORB – SLAM2 information system of tracking, built figure and loopback detection of the three parallel threads, angular velocity and acceleration of the join of IMU data can improve the mobile devices pose estimation precision and real-time performance, at the same time the depth data to build dense point cloud, in under the action of both real-time indoor scene of 3 d reconstruction, Figure 2. shows the structure of 3D reconstruction by fusion of depth data and IMU data:

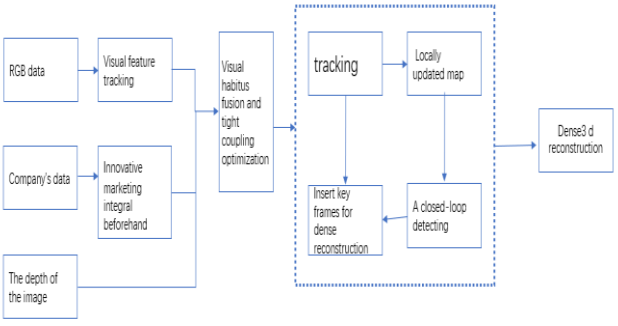


Figure 2. STRUCTURE diagram of 3D reconstruction

A. Innovative marketing integral beforehand

Innovative marketing module data including 3 axis gyroscope, 3 axis accelerometer and magnetometer sensor are independent of each other and not easy to influence each other between inertial sensor, the integral solution of displacement and rotation, with the passage of time accumulated error is bigger, data are available in a short period of time only, and magnetometers easily affected by the local environment, need correction in advance, Therefore generally do not adopt the method of integral calculation pose directly, instead of multi-sensor fusion method can get good posture calculation results, the current in order to achieve good posture combination of inertial sensor fusion method which USES the visual sensor, vision sensors in feature points clear, feature rich texture of effect is better, Met feature points such as glass, white walls, but fewer cases, recovery scenarios are defective or not working, and IMU sensor use for a long time has very big accumulated error, but the advantage in a short period of time when the displacement data of high accuracy, so adopt the way of visual and inertial data fusion compensate each other, can get good posture calculation results.

The acceleration of IMU coordinate system can be obtained by the accelerometer and gyroscope of IMU, as shown in the following formula (3-1). On the left is the direct output of IMU, $R_{WB(t)}$ is the rotation matrix from IMU coordinate system to the world coordinate system, while $R^{T}_{WB(t)}$ is the rotation matrix representing the world coordinate system to IMU coordinate system. $w^{a(t)}$ represents the acceleration in the world coordinate system, $w^{g(t)}$ represents the gravity vector in the world coordinate system, $b^{a}_{(t)}$ and $b^{a}_{(t)}$ represent the offset of the accelerometer and gyroscope, $\eta^{a}_{(t)}$ and $\eta^{g}_{(t)}$ represent the noise of accelerometer and gyroscope respectively, and the effect of earth rotation is ignored, thus assuming that the world coordinate system is an inertial system. The formula is as follows:

$$B^{\alpha(t)} = R^{T}_{WB}(t)(W^{\alpha(t)} - wg) + b^{a}(t) + \eta^{a}(t) \quad (1)$$

$$B^{\varpi}WB(t) = B^{\varpi}WB(t) + b^{g}(t) + \eta^{g}(t) \quad (2)$$

IV. VISUAL INERTIA IS TIGHTLY COUPLED

The tight coupling based on visual inertial sensor fusion uses RGB image for feature extraction and feature matching and combines IMU information for joint optimization. Tightly

coupled optimizations in trace threads include map updates and no map updates.

1) Map updates

The tracking thread can be divided into two situations: map update and no map update. At the beginning, there was no map update, and map update was completed in local map building and closed-loop thread. When the map is updated, the overall optimization state vector should be constructed first, including rotation, translation, accelerometer paranoia and gyroscope paranoia at the current moment J. The overall error equation includes visual reprojection error and IMU measurement error, and the last moment is expressed.

$$\theta = \left\{ R_{WB}^{j}, pp_{B}^{j}, w = v_{B}^{j}, b_{g}^{j}, b_{a}^{j} \right\} \tag{3}$$

$$\theta^{*} = \arg_{\theta} \min((\sum_{k} E_{proig}(\mathrm{k}, j) + E_{IUM}(i, j)) \tag{4}$$

2) When no map is updated

When there is no map update, the overall optimization state vector is constructed, including the rotation, translation velocity displacement, accelerometer bias and gyroscope bias of the current moment J +1 and the last moment J. The overall error equation includes visual reprojection error and IMU measurement error, as shown in the formula:

$$\theta = \left\{ R_{WB}^{j}, P_{W}^{j}, v_{W}^{j}, b_{a}^{j}, R_{WB}^{j+1}, v_{W}^{j+1}, b_{g}^{j+1}, b_{a}^{j+1} \right\} \tag{5}$$

V. IMPROVED ORB-SLAM BUILDS DENSE 3D MODELS

In order to be able to make no matter in a wide range of mobile robot and small scale and unknown environment online implementation of high precision positioning and reconstruction, need to dense, sparse map as sparse map based on feature points cannot be applied in practical fields such as robot navigation, path planning, so dense was carried out on the map is very necessary. Visual SLAM is the simultaneous localization and mapping, so who is the positioning? If the camera is on the robot, position the robot. If is holding a camera is a fixed camera, similarly built figure is also set up the camera or robot through the map, the two collections, can determine the robot in a certain location on the map as well as the continuous motion trajectory in the scene, although using SLAM built figure can very clearly see the camera or the trajectory of the robot, key frames, feature points, However, the visual information provided is less, and the established map is not clear and intuitive. Therefore, it is necessary to convert the sparse map into a dense THREE-DIMENSIONAL point cloud map, and splice the point cloud information of key frames. The generated point cloud map can clearly see the surrounding visual features and obtain a good visual experience. According to the internal parameters of the camera, the corresponding relationship between three-dimensional point cloud and two-position coordinates can be calculated, as shown in Formula.

By the location of the camera is run by generating each key frames of the point cloud, need to remove the depth value is too large or invalid points, and then use statistical filter to remove away from the point of "the masses", which is isolated points, each point by statistical distance and that point with other recent point distribution, keep a distance precision of points, Discarding the relatively far distance between the point, because the key frames overlap, so the points of the chapter, there are a lot of position close, will occupy a lot of memory and consumption space position, so it is necessary to drop the sample using voxel filtering, the advantages of speed filter is to ensure that each individual element within only a point, The space consumption of memory is reduced by down sampling of space, the pose of camera is calculated in the visual odometer and back-end optimization part, and the global point cloud can be obtained by splicing the point cloud data. The formula for calculating three-dimensional point cloud from two-dimensional color images and depth images is shown as below (6):

$$\begin{cases} z = \dfrac{d}{s} \\ x = (u - c_x) \cdot (\dfrac{z}{f_z}) \\ y = (u - c_y) \cdot (\dfrac{z}{f_z}) \end{cases} \quad (6)$$

Among them: f_x, f_y, f_x, f_y for camera inside, (u, v) as the image coordinates of (x, y, z) coordinate system for the image coordinates, d for depth camera measured pixel distance, s for the actual distance and the proportion of the measured distance d sparse coefficient, from the point cloud camera coordinate system to the global coordinate system transformation formula of point cloud is shown in the following (7):

$$x_{w,j} = T_{w,ci} X_{ci,j} \quad (7)$$

Where, $T_{w,ci}$ is the pose of the ith key frame, $T_{w,ci}$ is the point cloud in the coordinate system of the ith key frame, and $T_{w,ci}$ is the point cloud obtained in the global coordinate system after transformation.

Aiming at the deficiency of ORB SLAM algorithm, a method of rapidly constructing dense 3D map of unknown environment is added on this basis. The so-called dense map is to construct the corresponding point cloud for each key frame, and then assemble all the point clouds according to the key frame location information obtained from ORB SLAM2 to form a global point cloud map. The whole algorithm steps are as follows Figure 3.

Build a dense point cloud map by adding the PointCloudMapping.cc class to the SRC folder in the ORB-slam2 algorithm, which contains some member variables and member functions. This paper build a dense real-time 3 d reconstruction module is inserted into the key frames (mobile camera or robots get key frames) and the corresponding color and depth information, in according to the algorithm of tracking, BA optimization and key frames after the loopback detection and real-time update dense 3 d reconstruction module, complete 3 d point cloud splicing generates dense module.

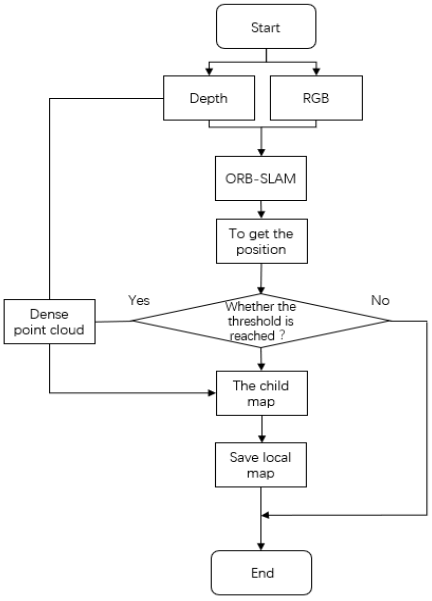


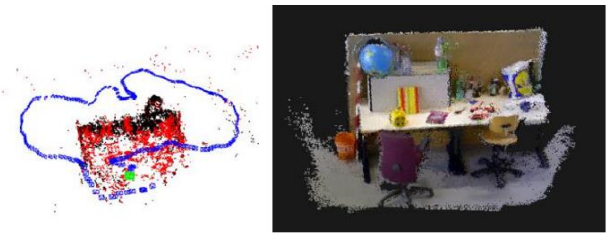
Figure 3. ORB-SLAM algorithm flow

In this paper, Realsense D435 is used as a sensor device. Firstly, the ORB-SLAM algorithm is used to extract key frames and obtain the pose of the robot. Then, whether the movement of the robot exceeds a certain threshold value is judged. The large scale map is decomposed into small maps of a certain size, thus reducing the operation time of saving the map and building the map, thus improving the overall performance of the algorithm.

According to the above theory, this chapter adopts the open dataset TUM of Technical University of Munich to conduct dense reconstruction test, and selects fr1_room, FR1_floor,andRGBD_DATASet_freiburg3_teddy. The results of sparse point cloud reconstruction and dense reconstruction are compared.

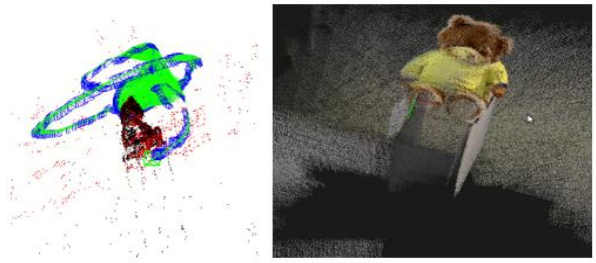
a Sparse map b Dense map

Figure 4. Dense reconstruction results of data set FR1_room



a Sparse map b Dense map

Figure 5. Dense reconstruction results of data set FR3_office



a Sparse map b Dense map

Figure 6. Reconstruction results of dataset RGBD_DATASet_freiburg3_TEDDY

The establishment of dense point cloud map plays an important role in the robot's perception and positioning of the surrounding environment. Compared with sparse point cloud map, it can make people understand the surrounding environment more intuitively. The establishment of THREE-DIMENSIONAL model can allow people to observe the real world from multiple angles and provide an immersive experience. It is of great significance to explore the unknown world with robots.

VI. EXPERIMENTAL PROCESS AND STEPS

All project need to use experiment to verify the overall performance of the system, this chapter will experiment link, the first to experiment environment is introduced and the hardware and software for laboratory use, through to the open source test data sets, and then carry on indoor mobile robot real scene assessment, this paper compares and analyzes so as to verify the reliability of this system.

A. Hardware experimental environment

This article USES the experimental equipment is independently developed by the indoor mobile robot, and above the robot of the original part of the reform so as to build the experiment platform of the topic, because this topic considering the hardware sensors and integration between software development is practical, need for robot original laser radar sensor will be demolished, Control system is a model for the small new Air15 lenovo workstation PC platform is used to control the robot on the surrounding environment scanning, the PC platform for Intel processor i5, run for 16 gb memory, graphics card memory is 4 gb, hardware platform is made of the robot include the robots, robot control, mobile robot perception of these three parts, Among them, Microsoft InteRealsense Depth Camera D435 is used as the main sensor, while the IMU sensor of MODEL MPU9250 is installed. The bottom of the robot is composed of chassis and quadrupedal wheels, which can realize 360 ° free movement of the robot.

TABLE I. HARDWARE CONFIGURATION PARAMETERS OF THE ROBOT

Hardware	Model
Upper-computer workstation	Lenovo Air15
IMU sensor	MPU9250
Depth sensor	Realsense D435

The Inter Realsense Depth Camera D435 is the preferred solution for applications such as phase robot navigation and object recognition. The Realsense D435 offers global shutter sensors and a larger lens for better low-light performance than cheaper D415 cameras. The D435 also features the more powerful RealSense module D430. The D435 camera has four circular holes on its front, running from left to right. The first and third are the IR Stereo Cameral. The second is an IR laser Projector, and the fourth is a color camera (color sensor). As shown below:

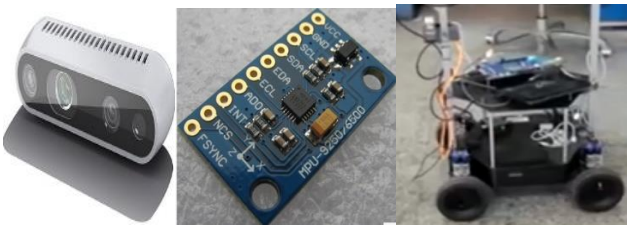


Figure 7. Depth camera schematic diagram

B. Software experiment environment

This paper uses lenovo Xiaoxin Air15 computer to build Windows 10 home edition and

Ubuntu18.04 mirror double operating system as the system software experimental platform. Compared with building virtual machines (Vmware Workstation Pro), Vmware workstation Pro has a faster loading speed, which is convenient, fast, safe and free, and has a convenient configuration environment. Ubuntu18.04 features simple and beautiful interface, convenient configuration environment, and a large number of third-party libraries and dependencies, including pangloin0.5, eigen3.4.5, OpenCV, etc. Eigen provides fast linear algebra operations on the matrix, but also includes the steps to solve the equation, many of the upper software library is also using Eigen matrix operations OpenCV in this paper is mainly on Realsense D435 depth image processing and optical flow tracking.

TABLE II. SOFTWARE SYSTEM CONFIGURATION

The name of the software	Model and Version
The operating system	Ubuntu18.04
Robot operating system	ROS-Melodic
Development programming language	c/c++
Visual library to use	opencv
Point cloud library	PCL

C. Sensor calibration experiment

In this experiment, the calibration of sensors is a link that cannot be ignored. For the calibration of cameras, a checkerboard calibration grid should be prepared in advance, as shown in the figure 8.

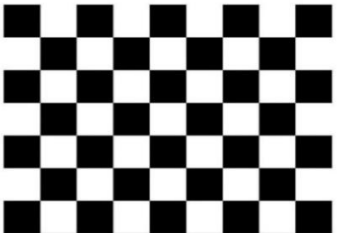


Figure 8. Schematic diagram of checkerboard calibration board

In this project, the calibration plate is fixed and images at different positions and angles of the calibration plate are collected by moving cameras. In this paper, 15 pictures from different angles are selected, as shown in Figure 9.

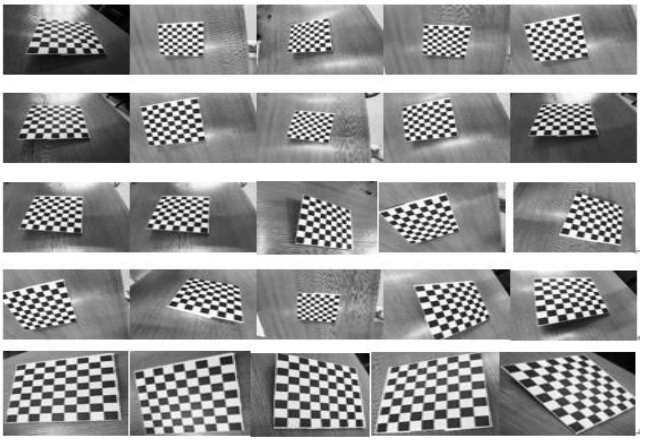


Figure 9. Schematic diagram of multi-angle calibration plate

Then, the calibration algorithm was used to extract corner information of each checkerboard calibration board, calculate the homography matrix of each image, and calibrate the external parameters and distortion parameters of the camera according to the size of the checkerboard, as shown in Figure 10. It can be seen that the corner information of the calibration board was extracted accurately.

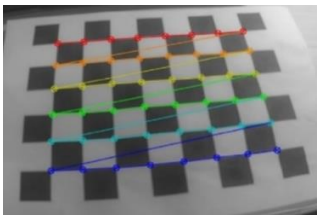


Figure 10. Corner extraction of the calibration board

According to the reprojection error test, the average error is 0.28213 pixels, which meets the requirements of this experiment. The experimental calibration results are shown in Table 3.

TABLE III. CALIBRATION RESULTS OF THE REALSENSE D435 CAMERA

The camera parameters	The calibration results		
Internal	605.894	0	317.887
	0	606.856	256.416
	0	0	1
Distortion parameter 1	0.15689		
Distortion parameter 2	-0.3425		
Mean reprojection difference	0.28213		

D. Multisensor calibration

Of multiple sensors is mainly on the camera and IMU sensor calibration, through access to the inside of the camera, in order to get more accurate

camera calibration need to get the camera and joint of IMU sensor calibration for the transformation matrix, transformation matrix can give 3 d reconstruction system provides a good initial value, higher accuracy, In this paper, Kalibr tool is used for joint calibration of the two. Just now, internal parameters of the camera have been obtained. Again, TOPIC and IMU data in the image obtained by the camera are extracted through ROS communication. Sufficient rotation and translation camera can collect enough data from different angles to make the calibration results between sensors more accurate. The calibration results are shown in Table 4.

TABLE IV. CALIBRATION RESULTS OF MULTIPLE SENSORS

The camera parameters	The calibration results		
Rotation matrix of camera and IMU	0.9999524 -0.0054268 -0.001426238	0.0042236 0.9995236 0.0017052	0.0012432 0.00158425 0.9999842
Translation matrix of camera and IMU	0.01238847	0.00842635	-0.02145263

Figure 11 shows the reprojection error diagram:

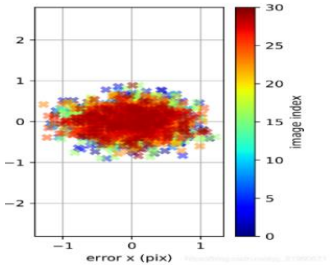


Figure 11. Reprojection error diagram

E. Real data experimental results

In this paper, the RGB-D sensor loaded by the robot is used instead of the IMU sensor module to carry out dense THREE-DIMENSIONAL reconstruction. The displacement curves of the robot in the X and Y directions and the trajectory of the robot are shown in Figure 12. After several experiments, it was found that dense reconstruction using a separate sensor would appear as follows

The effect shown in Figure 12, and the effect shown in Figure 13. is produced by using the multi-sensor fusion mode:

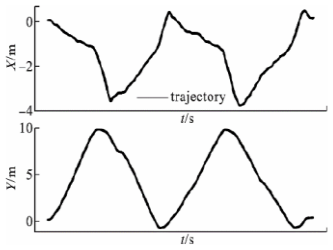


Figure 12. Displacement curve in XY direction

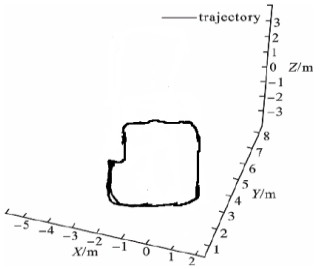


Figure 13. Trajectory diagram of the robot

As shown in Figure 13, the total length of the robot's moving track is 5m, and the end point positioning error is 0.12m in the X direction and 0.16m in the Y axis direction, indicating that the algorithm in this paper can work stably and effectively in the process of robot movement.

Figure 14. Top view of adding an IMU

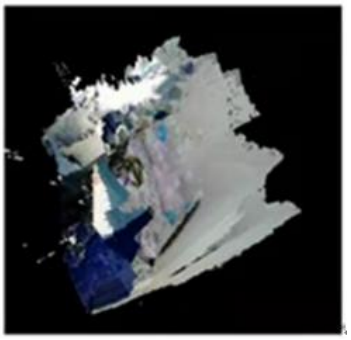


Figure 15. Top view of adding an IMU without an IMU

TABLE V. EXPERIMENTAL RESULTS OF 3D RECONSTRUCTION FOR DIFFERENT SENSORS

Sensor	Track the percentage of partial breaks caused by loss
Without IMU	35%
With IMU	10%

By Fig 12，13, 14 and table 5. shows that only rely on RGBD sensor vision camera, the colour image and depth map as the input for 3 d reconstruction, in the harsh conditions in complex scene and scene, camera is easy to track lost can't find between scanned, so will break, reconstruction of the map will not be able to use, However, when RGBD camera is constructed with IMU sensor for 3D reconstruction, the problem of image fracture and loss caused by the above single sensor can be alleviated in most implementations.

VII. SUMMARY

This paper mainly analyzes the SLAM scheme based on multiple visual sensors. Due to the defects of the pure visual scheme and its difficulty in stabilizing in various environments and practical scenes, the fusion of multiple sensors is adopted to construct 3D reconstruction. Finally, experiments are used to verify it. Pure vision will have fracture, while partial fracture caused by multi-sensor fusion tracking loss reaches 35%, and fracture caused by multi-sensor tracking loss reaches 10%. Compared with single sensor, the accuracy of multi-sensor fusion reconstruction is improved by 25%. It improves the accuracy and precision of indoor reconstruction.

REFERENCE

[1] Wang H, Wang H, Zhu X, et al. Three-Dimensional Reconstruction of Dilute Bubbly Flow Field With Light-Field Images Based on Deep Learning Method [J]. IEEE Sensors Journal, 2021, PP(99):1-1.

[2] Tokanai K, Kamei Y, Minokawa T. An easy and rapid staining method for confocal microscopic observation and reconstruction of three-dimensional images of echinoderm larvae and juveniles [J]. Development, Growth & Differentiation, 2021, 63.

[3] Kim J, Lee D, Doh G, et al. Three-dimensional tomographically reconstructed optical emission profiles of Hall thruster plasmas [J]. Plasma Sources Science and Technology, 2022, 31(1):015013 (11pp).

[4] Donati D M, Frisoni T. Implant Reconstruction of the Pelvis: IV: 3D-Printed Custom-Made Prosthesis. 2022.

AMM, BAFA, CMJGA, et al. Intraoperative three-dimensional bioprinting: A transformative technology for burn wound reconstruction [J]. Burns, 2022.

[5] AHK, ACL, ASK, et al. Three-dimensional volume reconstruction from multi-slice data using a shape transformation. 2022.

[6] XuY, NanL, ZhouL, et al. HRBF-Fusion: Accurate 3D reconstruction from RGB-D data using on-the-fly implicits [J]. 2022.

[7] Liu Y, Zhang S. Mediastinal basal pulmonary artery identification and classification by three-dimensional reconstruction [J]. Surgical and Radiologic Anatomy, 2022, 44(3):447-453.

[8] Guilloux Y L. THREE-DIMENSIONAL RECONSTRUCTION METHOD USING A PLENOPTIC CAMERA:, US20190156501A1[P]. 2019.

[9] K Zieliński, Staszak R, Nowaczyk M, et al. 3D Dense Mapping with the Graph of Keyframe-Based and View-Dependent Local Maps [J]. Journal of Intelligent & Robotic Systems, 2021, 103(2).

[10] Cong M, Fedkiw R, Lan L. OBTAINING HIGH RESOLUTION AND DENSE RECONSTRUCTION OF FACE FROM SPARSE FACIAL MARKERS:, US20210150810A1[P]. 2021.

[11] Weilharter R, F Fraundorfer. HighRes-MVSNet: A Fast Multi-View Stereo Network for Dense 3D Reconstruction from High-Resolution Images [J]. IEEE Access, 2021, PP(99):1-1.

[12] Lombardi M, Savardi M, Signoroni A. DenseMatch: a dataset for real-time 3D reconstruction [J]. 2021.

[13] Pan Z, Hou J, Yu L. Optimization algorithm for high precision RGB-D dense point cloud 3D reconstruction in indoor unbounded extension area [J]. Measurement Science and Technology, 2022, 33(5):055402 (15pp).

[14] Wang P, Shi L, Chen B, et al. Pursuing 3D Scene Structures with Optical Satellite Images from Affine Reconstruction to Euclidean Reconstruction [J]. 2022.

[15] ShakeriM, Loo Y, Zhang H . Polarimetric Monocular Dense Mapping Using Relative Deep Depth Prior [J]. 2021.

[16] Hller B, Mossel A, Kaufmann H. Automatic object annotation in streamed and remotely explored large 3D reconstructions [J]. Computational Visual Media, 2021.

[17] Theu L T, Tran Q H, Solanki V K, et al. Influence of the multi-resolution technique on tomographic reconstruction in ultrasound tomography [J]. International Journal of Parallel Emergent and Distributed Systems, 2021.

[18] Hermann M, Ruf B, Weinmann M. REAL-TIME DENSE 3D RECONSTRUCTION FROM MONOCULAR VIDEO DATA CAPTURED BY LOW-COST UAVS [J]. 2021.

[19] Farsangi S, Naiel M A, Lamm M, et al. Rectification Based Single-Shot Structured Light for Accurate and Dense 3D Reconstruction [J]. Journal of Computational Vision and Imaging Systems, 2021, 6(1):1-3.6932.

Face Mask Wearing Detection Based on YOLOv5

Yunshan Xie
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: 2398576240@qq.com

Zhiyi Hu
Engineering Design Institute
Army Research Loboratory
Beijing, 100042, China
E-mail: 763757335@qq.com

Jun Yu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: yujun@xatu.edu.cn

***Abstract*—In recent years, COVID-19 has swept the world, and people in crowded public places are usually large. In order to reduce the risk of virus transmission, stop the spread of the epidemic and reduce cross-infection, wearing masks correctly has become an important measure to prevent the virus. Aiming at the time-consuming and laborious situation of wearing masks manually, this paper proposes a mask wearing detection method based on yolov5. The input layer is mainly used for mosaic data enhancement, that is, adaptive anchor box and adaptive image scaling technology; Yolov5 in backbone mainly adopts focus and CSP (cross stage partial) structure; The neck layer adopts spp (spatial pyramid pooling) module and FPN (feature pyramid networks) + pan (pixel aggregation network) structure; The output mainly adopts ciou for the bounding box loss function_ Loss is the average index of NMS (non maximum suppression). This method uses 8000 preprocessed images as the data set and trains 200 epochs to get the final model. The algorithm visually displays the training and test results through tensor board, and inputs the pictures captured by the camera into the model to detect whether the face wears a mask. The accuracy, recall and mean accuracy (map) of the algorithm on the test set are 94.8%, 89.0% and 93.5% respectively, which are higher than the detection results of yolov3 and yolov4 algorithms.**

Keywords-Mask Detection; Yolov5; CIOU Loss Function

I. INTRODUCTION

By April, 2022, more than 500million cases of COVID-19 had been confirmed in the world, and more than 6.2 million cases had died. Under the normal epidemic situation, wearing masks in public places is an important means of effective epidemic prevention. Therefore, mask wearing detection has become a core work of epidemic prevention. At present, the main detection method is manual detection, which is not only time-consuming and laborious, but also increases the risk of virus infection. Therefore, it will be of great practical significance to build a mask detection and monitoring system to realize the automation and intelligence of epidemic prevention and control.

At present, some scholars have done research on mask wearing detection. Ren Yu [1] and others proposed a fast r-cnn mask wearing detection algorithm based on deep learning,

which accelerated the convergence of the model by migrating the weight of the pre training model on the Imagenet data set, and the accuracy of the final test set reached 89.41%. Dong Yanhua [2] and others proposed a SSD network based on residual structure. By adding residual structure before the positioning and classification of SSD network, the feature extraction network and classification and positioning layer are separated, which effectively solves the dual task of learning local information and high-level information at the same time. Finally, the average detection accuracy reaches 92.3%. Cao chengshuo [3] and others proposed a YOLO mask algorithm, which is based on YOLOv3, introduces the attention mechanism into the feature extraction network, and uses the feature pyramid and path aggregation strategy for feature fusion. The average accuracy of the algorithm is 93.33%. Based on YOLOv4, Cheng Tinghao [4] and others redesigned the anchor frame size with K-means clustering algorithm and improved the network structure. The recall rate and accuracy rate on the test set reached 88.20% and 92.30% respectively.

This paper adopts the YOLOv5 target detection network model. Compared with YOLOv3 and YOLOv4, the innovation of YOLOv5 is that it uses the focus structure in the backbone, which reduces flops and improves the training speed. Compared with the ordinary convolution operation in YOLOv4neck structure, YOLOv5 uses CSP_X structure, which strengthens the fusion ability of network features and improves the detection accuracy. Overall, the performance of YOLOv5 is better than that of YOLOv4 and YOLOv3.

II. YOLOV5 NETWORK MODEL

A. YOLOv5 algorithm

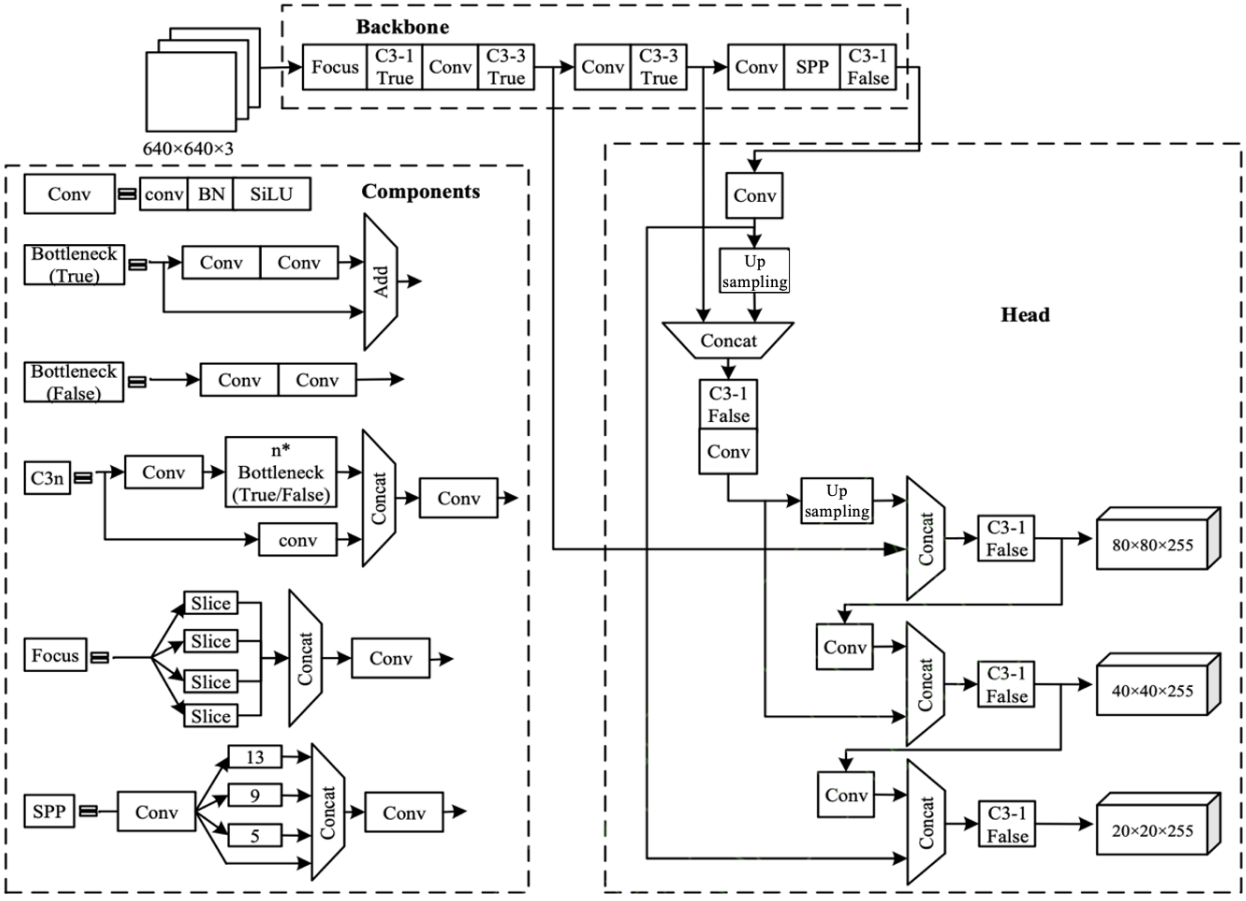


Figure 1. YOLOv5s network structure diagram

Yolov5 algorithm is developed on the basis of yolov4 and yolov3. Compared with yolov4, the architecture of yolov5 is about 90% smaller [9]. In terms of accuracy, the performance of yolov5 is better than that of yolov3 and yolov4 algorithms in the current market.YOLOv5 is a single-stage target detection algorithm. YOLOv5 [10] series can be divided into YOLOv5s [11], YOLOv5m, YOLOv5l [12] and YOLOv5x [13]. Among them, YOLOv5s has the smallest network. Although the accuracy is slightly worse than the other three, it has the fastest speed. Because of the speed requirement of mask detection algorithm, this paper selects YOLOv5s network model.

The network structure diagram of YOLOv5s is shown in Figure 1. It can be seen that the model is mainly divided into four parts: input, backbone, neck and prediction. Mosaic data enhancement is used at the input end to improve the detection effect of small targets [14]. In addition, using the adaptive anchor frame calculation method [15], different data sets will have anchor frames with initial length and width, which can get a larger intersection and union ratio, and greatly improve the efficiency of training and prediction [16]; The backbone part is a convolutional neural network that aggregates and forms image features on different image granularity [17], mainly including Focus structure and CSP structure. The neck part is between the backbone and prediction. It is a network layer of a series of mixed and combined image features, which transmits the image features to the prediction part [18]. The prediction part is the final detection part, which mainly predicts the image features, generates the boundary box and predicts the type of target [19].

B. Input

The input terminal includes mosaic data enhancement [20], image size processing and adaptive anchor box calculation. Masaic data enhancement adopts 4 pictures to be spliced in the way of random scaling, random clipping and random arrangement. Using mosaic data enhancement can greatly enrich the data set and add many small targets. YOLO algorithms need to change the input image size into a fixed size. The standard size of the image in this paper is 640 × six hundred and forty × 3。 It is necessary to set the initial knowledge anchor box before network training. The initial knowledge anchor box of YOLOv5 is [10, 13, 16, 30, 33, 23], [30, 61, 62, 45, 59119], [116, 90156198373326].

Adaptive image scaling refers to the unified scaling of images with different lengths and widths to a standard size, and then used as a data set for detection and processing. After the picture is scaled, if there are many black edges filled on both sides of the picture, there will be information redundancy and affect the speed. In this regard, yolov5 has been improved and the letterbox function has been modified, so that the original picture can adaptively reduce information redundancy, that is, reduce black edges. The steps of yolov5 algorithm for image adaptive scaling are to calculate the scaling scale, calculate the size after scaling, and calculate the black edge filling value. The calculation formula is as follows (1):

$$\begin{cases} l_3 = l_2 + np \cdot mod\left(\left(l_1 - w_1\right)\right) \cdot min\left\{\dfrac{l_2}{l_1}, \dfrac{w_2}{w_1}\right\}, 32\Big) \\ w_3 = w_1 \cdot min\left\{\dfrac{l_2}{l_1}, \dfrac{w_2}{w_1}\right\} + np \cdot mod\left(\left(l_1 - w_1\right) \cdot min\left\{\dfrac{l_2}{l_1}, \dfrac{w_2}{w_1}\right\}, 32\right) \end{cases} \tag{1}$$

Where: l_1, w_1 is the length and width of the original image; l_2, w_2 is the length and width of the original scaled size; $l_3, w_3$is the length and width of adaptive scaling size; $\frac{l_2}{l_1}, \frac{w_2}{w_1}$ is the scaling factor. Select a small scaling factor and multiply it by the length and width of the original image size to obtain l_2 and w_2 ; $(l_1 - w_1) \cdot min\left\{\frac{l_2}{l_1}, \frac{w_2}{w_1}\right\}$ is the height to be filled, and then in numpy use np.mod to remove the remainder, get 8 pixels, and divide 8 by 2 to get the values that need to be filled in at both ends of the image adaptation.

Adaptive anchor box calculation, setting anchor boxes with initial length and width for different data sets. In the network training of YOLOv5, the predicted anchor frame is output on the set initial anchor frame, the error between the predicted anchor frame and the ground truth of the actual calculation anchor frame is calculated, and the data is updated by backpropagation, and the network parameters are optimized through multiple iterations. The anchor box is calculated. In network training, the network outputs the prediction frame on the basis of the initial anchor frame, then compares it with the real frame, calculates the gap between the two, and then updates it in reverse to iterate the network parameters.

C. Backbone

YOLOv5 uses the focus and CSP structures in backbone [21]. The CSP[22] structure is shown in the network structure diagram in Figure 1, The Focus structure is the operation of the YOLOv5 algorithm that is different from YOLOv3 and YOLOv4 ,and the key step of focus structure is slicing operation, as shown in Figure 2. For example, the image of 608 * 608 * 3 is input into the focus structure, and the slicing operation is adopted. First, it becomes the feature map of 304 * 304 * 12, and then after a convolution operation of 32 convolution cores, it finally becomes the feature map of 304 * 304 * 32. The original intention of the design of the Cross Stage Partial structure is to reduce redundant computation and enhance gradients. The difference between YOLOv5 and YOLOv4 is that only the backbone network in YOLOv4 uses CSP structure. Two CSP structures are designed in YOLOv5. Take YOLOv5s network as an example, CSP1_ X structure is applied to backbone network, another CSP2_ X structure is applied to neck.

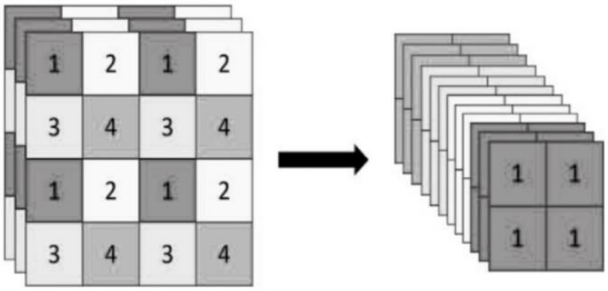


Figure 2. Slicing operation

D. Neck

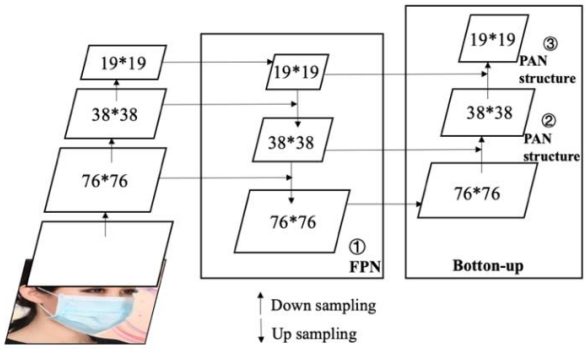


Figure 3. FPN + PAN structure

The network structure design of neck adopts the structure of FPN + PAN. FPN [23] uses a top-down side connection to build a high-level semantic feature map on all scales and construct the classic structure of the feature pyramid. PAN [24] uses a bottom-up feature pyramid. After passing through a multi-layer network in the middle of FPN, the target information at the bottom has been very blurred. Therefore, pan is added to make up for and strengthen the

positioning information. The specific structure is shown in Figure 3.

E. Prediction

Prediction includes bounding box loss function and non maximum suppression (NMS). YOLOv5 uses CIOU_Loss [25], which effectively solves the problem of overlapping bounding boxes. In the post-processing process of target detection, for the screening of many target frames, weighted NMS operation is adopted to obtain the optimal target frame.

Target detection algorithms often output multiple overlapping prediction frames for the same target, resulting in false detection. Therefore, non maximum suppression algorithm [26] is generally used as post-processing technology to suppress redundant prediction frames in order to obtain the final detection results. NMS [27] and soft NMS are commonly used non maximum suppression algorithms [28]. It mainly reduces the redundant prediction frames for the same target [29], reduces the number of false positive prediction frames and improves the accuracy of target detection by suppressing the confidence of the prediction frame with non maximum confidence [30] (hereinafter referred to as the non maximum box). However, NMS and soft NMS only use IOU as the judgment standard [31] to suppress the prediction box that highly overlaps with the confidence maximum box. This kind of algorithm can successfully suppress redundant prediction frames, but it will also suppress adjacent targets [32]. Therefore, this paper uses weighted NMS operation [33]. Compared with the traditional non maximum suppression, weighted NMS does not directly eliminate those frames with the same category as the current rectangular frame whose IOU is greater than the threshold in the process of rectangular frame elimination, but weights them according to the confidence of network prediction to obtain a new rectangular frame, Take the rectangular box as the rectangular box of the final prediction, and then eliminate those boxes [34].

III. EXPERIMENT AND RESULT ANALYSIS

A. Experimental preparation

There are 8000 pictures in this experimental data set, including two categories: mask and no mask. Mask indicates that the personnel have worn masks, and no mask indicates that the personnel have not worn masks. 80% of the images in the data set are used as training sets and 20% as test sets. Some dataset images are shown in Figure 4.

Figure 4. Dataset partial image

The experiment uses pytoch to build the network framework, the initial learning rate is set to 0.01, and 200 epochs are trained iteratively. Firstly, the obtained images are normalized and preprocessed, and then the processed images are input into the YOLOv5 network model training to obtain the best weight data, and then the images are tested and analyzed. The experimental process is shown in Figure 5.

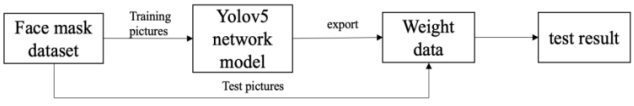


Figure 5. Experimental flow chart

B. Mosaic data enhancement

Mosaic splices four pictures by randomly selecting them and randomly cutting, arranging and scaling them. The effect picture is shown in

Figure 6. This method enriches the background and small targets of the detection object, and improves the robustness of the network to a certain extent. In addition, after mosaic data enhancement, it is equivalent to processing four pictures at a time, and the batch size increases implicitly, which makes the initially set batch size value do not need to be large, and a good model can be obtained, reducing the performance requirements of GPU.

Figure 6. Mosaic data enhancement

C. Experimental environment

The configuration of algorithm experiment platform is shown in Table 1.

TABLE I. EXPERIMENTAL ENVIRONMENT CONFIGURATION

Parameter	Configuration
CPU	Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz
GPU	RTX 3090
Memory	60G
Display memory	24G
System environment	Ubuntu18.04
Experimental platform	PyTorch1.6.0、Python3.8
Accelerated environment	CUDA10.1

D. YOLOv5 network training

In this paper, the training epoch is 200 and the batch size is 64. The official pre training weight of YOLOv5 is used to accelerate the convergence of the model. In the experiment, the learning rates of BN layer, weight layer and Bais layer are lr0, lr1 and lr2 respectively, in which lr0 and lr1 change the same. The learning change rate is shown in Figure 7.

In this paper, the loss function adopts CIOU_ Loss, as shown in formula (2). Where v is used to measure the consistency of the relative proportion of two rectangular boxes, α Is the weight coefficient, from α It can be seen from the definition of parameters that the loss function will be more inclined to optimize in the direction of increasing overlapping areas, especially when the IOU is zero. CIOU_Loss comprehensively considers the overlapping area, center distance and aspect ratio, and further considers the relative proportion of rectangular frame, which makes the detection effect further.

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2\left(\mathbf{b},\mathbf{b}^{gt}\right)}{c^2} + \alpha v \qquad (2)$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2$$

$$\alpha = \frac{v}{(1-\mathrm{IoU})+v}$$

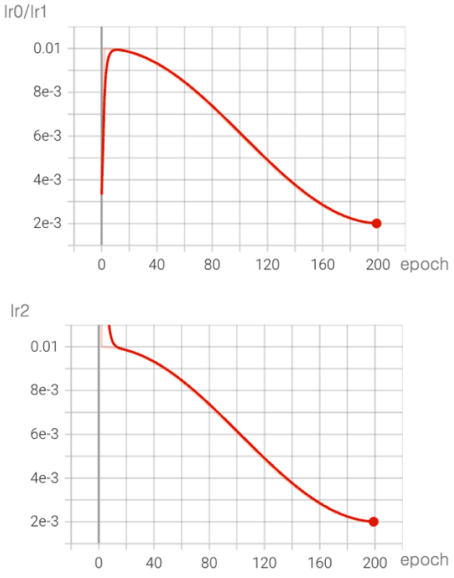


Figure 7. Learning rate curve

E. Experiment and result analysis

The results of this experiment will be from mAP@0.5, mAP@0.5: 0.95, Precision and Recall. The calculation formula is as follows.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$AP = \int_0^1 P \mathrm{d}R \tag{5}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{6}$$

TP, FP and FN in (3) and (4) above refer to the number of correct inspection frames, false inspection frames and missed inspection frames respectively. AP value is the area of P-R curve, and N in formula (6) represents the total number of detection categories, which is 2 in this paper. mAP@0.5 it refers to the average AP of all categories when IOU is set to 0.5, mAP@0.5: 0.95 refers to the average mAP on different IOU thresholds. The IOU value ranges from 0.5 to 0.95 in steps of 0.05.

While outputting the prediction box, this paper will also output the classification confidence score belonging to the box. Generally speaking, for a certain model, by adjusting the confidence threshold, the positive or negative of the predicted value can be changed, and the Precision and Recall also change accordingly. By observing the changes in the confidence thresholds of Precision and Recall followers, the quality of the model can be evaluated to a certain extent. If a model maintains a stable Precision at a high level while Recall grows, it proves that the model has better performance. If a model needs to lose a lot of Precision in exchange for the improvement of Rccall, the performance of the model is poor. Typically, researchers use the Precision-Recall curve to measure the model's trade-off between Precision and Recall. The mean precision of each category is calculated from the area under its corresponding Precision-Recall curve. Generally, the higher the AP of each category, the better. The average precision of the evaluation indicators commonly used in target detection is the average of each category of AP.

The experimental process and experimental test results are shown in Figure 8 and Table 2 respectively.

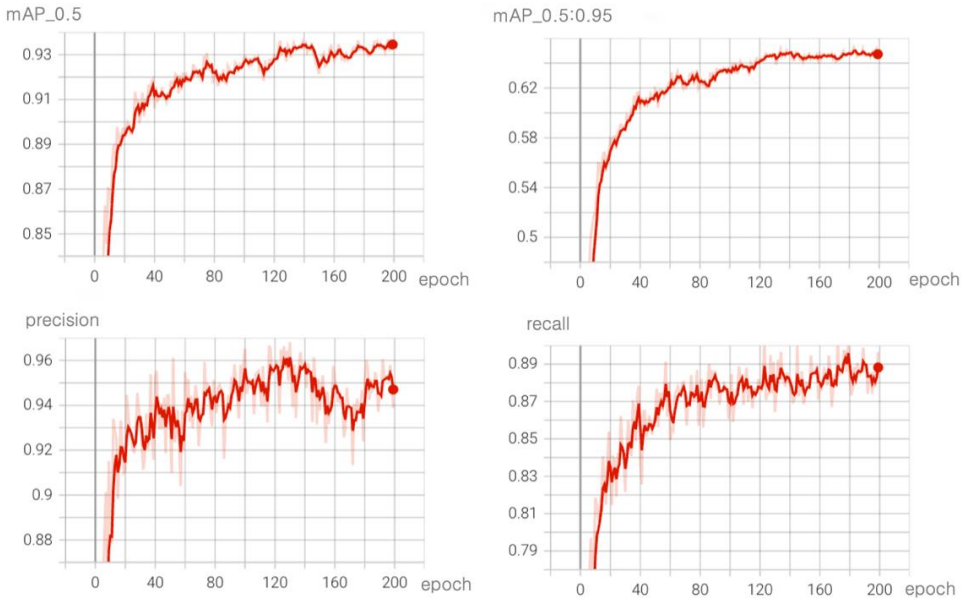


Figure 8. Experimental training process

TABLE II. TEST RESULTS OF EACH MODEL ON THE TEST SET

Model	Precision/%	Recall/%	mAP@0.5/%
YOLOv5s	94.8	89	93.5
YOLOv4	76.2	85.4	51.2
YOLOv3	73.6	82.3	48.9

It can be seen from table 2 that the Precision, Recall and mAP of YOLOv5s on the test set are 94.8%, 89% and 93.5% respectively. These three indicators are significantly higher than YOLOv3 and YOLOv4 algorithms.

After the algorithm training in this paper, input the data of the test set into the network model to get the test results, as shown in Figure 9. If the mask is worn, the value on the target box represents the confidence of classification. If the mask is not worn, no mask will be displayed on the target box. According to the performance analysis of various data, the training effect of YOLOv5 network model is ideal.

Figure 9. Experimental test effect

IV. CONCLUSION

To sum up, in order to achieve the accuracy and real-time of detection and liberate the manual from the complex detection work, this paper proposes a personnel mask wearing detection method based on yolov5, which can effectively solve the problem of time-consuming and laborious manual detection of mask wearing. If the mask is not worn in the public place, it will be detected in the mask confidence box. If the mask is not worn in the public place, it will be detected and displayed. Experiments show that the accuracy, recall and average accuracy (map) on the test set reach 94.8%, 89.0% and 93.5% respectively. The detection effect is better than that of yolov4 and yolov3, and the overall effect is ideal. Further improvements can be made in the follow-up, such as how to detect the non-standard wearing of personnel masks, how to lighten the model, and how to verify and improve the network model in practical application.

REFERENCES

[1] Ren Yu, Liu Quanjin, Huang Zhong, Hu langtao, Liu Guoming. Mask wearing detection algorithm based on fast r-cnn and transfer learning [J] Journal of Anqing Normal University (NATURAL SCIENCE EDITION), 2021, 27(04): 25-30. DOI: 10.13757/j.cnki.cn34-1328/n. 2021. 04. 005.

[2] Dong Yanhua, Zhang Shumei, Zhao Junli.SSD mask detection based on residual structure [J] Computer technology and development, 2021, 31(12):67-72.

[3] Cao chengshuo, Yuan Jie. Mask wearing detection method based on Yolo mask algorithm [J] Progress in laser and Optoelectronics, 2021, 58(08):211-218.

[4] Cheng Tinghao, Cui Yuchao, Wu Xinmiao. Research on mask wearing detection in public places based on yolov4 [J] Modern computer, 2021(16):134-140.

[5] https://zhuanlan.zhihu.com/p/172121380, 2021-04-2

[6] Wang hengtao, Zhang Shang, Zhang Chaoyang, Liu Zhanwei Lightweight PCB defect detection based on yolov5 [J / OL] Radio Engineering: 1-9 [2022-05-10] http://h-p.kns.cnki.net.lib-ycfw.xatu.edu.cn/kcms/detail/13.1097.TN.20220509.1015.004.html

[7] Zhang Xuan, Yan Mingzhong, Zhu Daqi, Guan Yang. Marine ship detection and classification based on YOLOv5 model [J]. Journal of Physics: Conference Series, 2022, 2181(1).

[8] Li Chao, Cao Yining, Peng Yakun. Research on Automatic Driving Target Detection Based on YOLOv5s [J]. Journal of Physics: Conference Series, 2022, 2171(1).

[9] Guo Xiaotong, Zuo Min, Yan Wenjing, Zhang Qingchuan, Xie Sijun, Zhong Iker. Behavior monitoring model of kitchen staff based on YOLOv5l and DeepSort techniques [J]. MATEC Web of Conferences, 2022, 355.

[10] Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Vasyl Melkin, Oleh Biloborodov, Mykola Samoilenko, Olena Kravchenko, Halyna Kalynychenko, Anton Rohovyi, Mykhaylo Soloshchuk. Improving the model of object detection on aerial photographs and video in unmanned aerial systems [J]. Eastern-European Journal of Enterprise Technologies, 2022,1(9).

[11] He Yi, Li Han Dong Mask wearing recognition in complex scenes based on improved yolov5 model [J] Microprocessor, 2022, 43 (02): 42-46

[12] Chen Jiping, Chen Yongping, Xie Yi, Zhu Jianqing, Zeng huanqiang Ghost Yolo: lightweight mask face detection algorithm [J / OL] Signal processing: 1-13 [2022-05-10] http://h-p.kns.cnki.net.lib-ycfw.xatu.edu.cn/kcms/detail/11.2406.TN.20220322.1023.002.html

[13] Wang Xinran, Tian Qichuan, Zhang Dong Overview of research on wearing detection of face masks [J / OL] Computer engineering and application: 1-15 [2022-05-10] http://h-p.kns.cnki.net.lib-ycfw.xatu.edu.cn/kcms/detail/11.2127.TP.20220124.1558.014.html

[14] Guo Lei, Wang Qiulong, Xue Wei, Guo Ji Mask wearing detection in dim light based on attention mechanism [J] Journal of University of Electronic Science and technology, 2022, 51 (01): 123-129

[15] Yu Shuo, Li Hui, GUI Fangjun, Yang Yanqi, LV Chenyang Research on real-time detection algorithm of mask wearing based on yolov5 in complex scenes [J] Computer measurement and control, 2021,29 (12): 188-194 DOI:10.16526/j.cnki. 11-4762/tp. 2021.12.035.

[16] Zhu Xinpeng, Li Dan Detection system for wearing masks in railway station based on yolo5face [J] Electronic testing, 2021 (24): 50-52 DOI:10.16520/j.cnki. 1000-8519.2021.24.017.

[17] Chen Zhaojun, Chu Jun, Zeng lunjie. Multi category mask wearing detection based on dynamic weighted category balance loss [J / OL] Journal of graphics:1-10[2022-04-26].

[18] Zhang Qian.Research on acceleration of SAR image target detection algorithm based on deep learning [D] Beijing Jiaotong University, 2021.DOI:10.26944/d.cnki.gbfju.2021.000566.

[19] Li Bao. Design and implementation of pedestrian detection algorithm based on deep learning [D] Beijing University of Posts and Telecommunications, 2021.DOI:10.26969/d.cnki.gbydu.2021.000482.

[20] Niu Ruixin, Zhao Zhengjian, Zhang Shiyuan, Dang Jie, Du Lu.Ship detection in optical remote sensing image based on FPN feature extraction [J] Radio Engineering, 2021,51(11):1296-1302.

[21] Yu Wenqing. Research and application of natural scene text detection algorithm based on deep learning [D] Hebei University of Geosciences, 2022.DOI:10.27752/d.cnki.gsjzj. 2022.000068.

[22] Li Yongshang, Ma Ronggui, Zhang Meiyue.Improve the monitoring video traffic flow statistics of yolov5s + deepsort [J] Computer engineering and Application,2022,58(05):271-279.

[23] Zhang Zhiyan, Li Dan Face detection with mask on campus based on yolo5face [J] Electronic test, 2021 (23): 40-42 + 99 DOI:10.16520/j.cnki. 1000-8519.2021.23.012.

[24] Fan qinrui, Li Dan Mask face detection based on yolo5face [J] Electronic production, 2021 (19): 61-63 + 8 DOI:10.16589/j.cnki. cn11-3571/tn.2021.19.019.

[25] Ye Xingyu Research on mask wearing detection algorithm based on deep learning [J] Information and computer (theoretical Edition), 2021, 33 (18): 72-76

[26] Zhang Luyao, Han Hua Face mask detection based on yolov5s [J] Intelligent computer and application, 2021, 11 (09): 196-199

[27] Yu Fangxu, Wang Shuai, Liu Shichao, Liang Peng, Yu Xi Campus mask wearing detection system based on convolutional neural network [J] Electronic components and information technology, 2021, 5 (08): 19-20 + 24 DOI: 10.19772/j.cnki. 2096-4455.2021.8.009.

[28] Peng Cheng, Zhang Qiaohong, Tang Zhaohui, GUI Weihua Research on mask wearing detection method based on yolov5 enhancement model [J] Computer engineering, 2022, 48 (04): 39-49 DOI: 10.19678/j.issn. 1000-3428.0061502.

[29] Liu Qiyuan Research on single-stage complex face detection method [D] People's Public Security University of China, 2021 DOI: 10.27634/d.cnki. gzrgu. 2021.000271.

[30] Tan Shilei, BIE xiongbo, Lu Gonglin, Tan Xiaohu Real time detection of personnel wearing masks based on yolov5 network model [J] Laser journal, 2021,42 (02): 147-150 DOI:10.14016/j.cnki. jgzz. 2021.02.147.

[31] Wang Feng Improved yolov5 artificial intelligence detection and recognition algorithm for wearing masks and helmets [J] Architecture and budget, 2020 (11): 67-69 DOI:10.13993/j.cnki. jzyys. 2020.11.021.

Research on Construction Method of Wavelet Telemetry Data with Improved Threshold

Yangyang Sun
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 2675379281@qq.com

Shuping Xu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 563937848 @qq.com

Haonan Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1160411807 @qq.com

Yueqiu Huang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 410666963 @qq.com

***Abstract*—In order to strengthen the applicability of data denoising algorithm, this thesis study the common telemetry data denoising algorithm based on the data of engine speed, flight space speed, cabin temperature and humidity, and establishes the evaluation model of error square sum and curve similarity to evaluate the denoising performance. Experiments show that the polynomial fitting has the greatest denoising error and slow convergence speed. The five-point cubic smoothing has the smallest overall denoising error, the median filtering algorithm can change the effect of smoothing effect by adjust it's moothing window, but ignores the authenticity of data. Therefore, the above three data denoising algorithms do not meet the requirements of telemetry data processing. In this thesis, an improved threshold function is proposed which effectively improves the data jump and excessive smoothing and reduce the denoising accuracy compared with the traditional thresholding function in order to makes the measured value closer to the true value. The algorithm is applied to the noise processing of four kinds of telemetry data, the results show that the denoising accuracy is improved significantly compared with the other three algorithms, which makes the measured value closer to the true value to reflect the changing trend of the original measurement data more truthfully, and the curve similarity is improved significantly, which are all above 80%.**

Keywords-Function Construction; Measurement Error; Telemetry Data; Curve Similarity

I. INTRODUCTION

In order to ensure the safety of aircraft flight, a lot of testing work needs to be done before the official flight. Due to the measurement of telemetry parameters and the vibration, electromagnetic interference, quantization error and propagation path during transmission, the original data will inevitably have measurement errors. At this time, the measured values usually have burrs, sharp corners, sudden changes, etc., which may cause some interference to the signal analysis. In order to improve the measurement accuracy of the telemetry system, reduce the fitting error, and ensure the safety of the aircraft, the research on the construction method of the telemetry data function is very important.

Common function construction methods mainly include simple polynomial fitting, differential evolution [1-3], ant colony optimization [4], moving average, weighted local polynomial

DOI: 10.2478/ijanmc-2022-0018

regression, etc [5]. (Zhang B, J. et al, 2015). polynomial fitting is simpler to implement, but it has disadvantages such as large fitting error and slow speed. Differential evolution and ant colony optimization algorithms have been widely used in numerical optimization, but only for combinatorial optimization problems on continuous regions, and are not suitable for numerical optimization on discrete regions [6-7]. The real-time cost of the sliding average algorithm data is exchanged for stability, and the larger the amount of data selected, the larger the delay; When the number of smoothing points is small, although the sum of the squares of the residuals is small, the smoothness is very poor and still has a broken line shape; The weighted local polynomial regression method has a poor smoothing effect when the error obeys the normal distribution; Zhao B. [8] presents a data processing method based on wavelet analysis for the problem of smoothing noisy data, the noise intensity is estimated and the data filtering is effective. Liang J, [9] derived from the literature to use Kalman filter for data smoothing, with small estimation bias and good real-time performance. However, for aircraft display systems, such methods are often poorly evaluated for parameters such as maximum delay and maximum error. Wavelet transform can simultaneously analyze signals in time domain and frequency domain. The denoising can be summarized into three methods: the wavelet coefficient modulus maxima denoising method proposed by Mallat; the spatial correlation noise reduction based on wavelet coefficients proposed by Sui W T, et al. [10] wavelet threshold denoising shrinkage method based on traditional hard and soft threshold ideas proposed by Donoho and Johnstone [11-13]. By studying the traditional soft threshold and hard threshold function, it can be concluded that the hard threshold method can preserve the edge features of the data and has the advantage of effectively retaining the authenticity of the data. However, this method does not have continuity, and it is easy to cause visual distortion such as vibration and pseudo-Gibbs effect at discontinuous points during signal reconstruction [14]. The advantage of soft threshold method is that the image is smoother after denoising, but it will make a fixed error between the estimated coefficient and the original coefficient, and make the edge information of the image blurred after denoising, resulting in image distortion after denoising [15]. In order to improve the denoising performance of the threshold function of wavelet method, Breiman proposes a compromise algorithm garrote threshold function, which is based on the traditional soft and hard threshold functions, and retains the advantages of the soft threshold function and the hard threshold function. However, the method ignores the feature that the noise will gradually decrease with the increase of decomposition scale under wavelet transform [16].

In order to effectively eliminate the noise of telemetry data and improve the accuracy and applicability of data measurement, this paper studies the construction method of aircraft telemetry data function. This method makes the measured value of the system closer to the true value, which makes it easier for the ground monitoring personnel to test the flight state of the aircraft. Let them perform accurate, real-time fault analysis.

II. Telemetry Data Processing System

With the requirement for ensuring flight safety of aircraft is becoming higher and higher. It is necessary to do quantity testing work before formal flight to real-time monitoring and maintenance of flight status to ensure formal flight safety. So, it is very important to develop a efficient telemetry data monitoring system and to study an efficient data denoising algorithm. In order to improve the flexibility and portability of the telemetry system, this thesis uses C++—QT technology to complete the design of data monitoring software which is realizes the dynamic configuration of telemetry engineering task files

based on IRIG 106 standard PCM data frame format, frame integrity judgment, data shunting, character conversion and engineering data conversion, to divides and restores the original engineering quantity signal by integrating various parameters along the way. At the same time, the thesis use the double buffer pool technology to receive and store data, which improves the efficiency of data analysis and meets the real-time requirements of system monitoring and maintenance.

Based on the telemetry system, this paper studies the software design of the portable data monitoring system and the construction method of telemetry data function. The telemetry system usually comprises two parts: an aircraft transmitting end and a ground receiving end, a telemetry transmitting end and a ground receiving end, and the transmitting end comprises a device such as a sensor, a converter, an encoder and a transmitter; the receiving end includes a receiving module, a data processing module, a recording module, a display module, and the like. The general structure of telemetry system is shown in Fig 1.

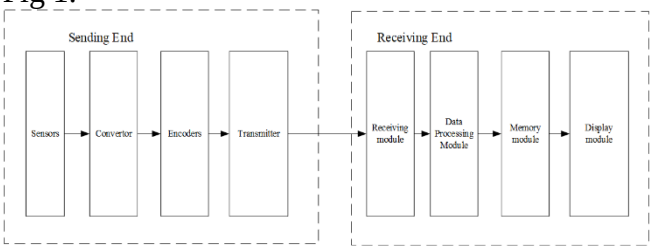


Figure 1. Overview of telemetry data processing structure.

As can be seen from Fig.1, at the sending end, the system collects and encodes the measured information, and edits the multi-channel parameter information into group signals suitable for single channel transmission according to a certain system. This signal is modulated by a transmitter carrier to form an electrical signal which is then transmitted into space.

At the receiving end, after the telemetry signal is transmitted via the radio link, it is first sent to the receiver by the receiving antenna for carrier demodulation to obtain the group signal. Then the data processing module performs signal processing, branching, storage, conversion and other analytical processing to recover the engineering quantity signal, and performs function construction on the parsed data to effectively reduce the measurement error, improve the telemetry data analysis efficiency and fault analysis accuracy. Finally, the measurement information is displayed through the visual interface.

III. Wavelet Function Construction Method With Improved Threshold Function

The wavelet threshold includes global threshold and hierarchical threshold. In order to remove data noise to the greatest extent and avoid local jitter caused by noise removal, the paper uses hierarchical threshold wavelet method to perform telemetry parameter denoising processing, ie for each group of data. The threshold processing is used for multiple times, and the noise of different frequencies is removed layer by layer from low frequency to high frequency, which not only can preserve the details of the data, but also can better smooth the data. Avoid excessive noise smoothing, or incomplete noise removal [17]. Wavelet Hierarchical Threshold Denoising Principle One-dimension wavelet delamination threshold denoising includes four modules: selecting wavelet basis function, determining the number of decomposition layers, carrying out threshold processing and signal reconstruction. The specific threshold noise reduction process is shown in Figure2.

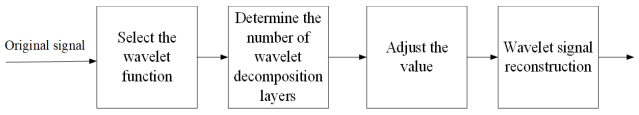


Figure 2. Wavelet threshold method signal reconstruction flowchart.

As shown in Fig 2, firstly, it is necessary to select a base wavelet function and an appropriate decomposition layer according to the data type to decompose the noise signal, so that the noise

signal is distributed in the high frequency part [18]. Then, a suitable threshold and threshold function are selected to quantize the high-frequency coefficients generated by the wavelet decomposition, and the wavelet coefficients of the noise suppression are generally expressed as the larger amplitude of the wavelet coefficients is retained, and the smaller amplitude is set to zero; Finally, the quantized wavelet coefficients are reconstructed to obtain an estimated value of the original signal, and data denoising is completed. In the construction of wavelet threshold denoising function, the selection of threshold and threshold function has a great influence on the denoising effect. Since the threshold and threshold functions are not unique, different threshold functions have different denoising effects and decomposition layers. Therefore, the paper will study the threshold function and threshold determination method and optimize and improve the defects of the existing threshold function to improve the denoising accuracy and applicability of the algorithm.

IV. Wavelet Algorithm Design for Improving Threshold Function

In order to effectively measure the data noise and suppress the influence of noise, according to the characteristics that the amplitude of the wavelet coefficient of the effective signal is large and the amplitude of the wavelet coefficient of the noise signal is small, the article uses a multi-layer threshold method for threshold estimation, which is for each group of data. Use more than one threshold processing. Removing noise at different frequencies layer by layer from low frequency to high frequency not only preserves the details of the data, but also smoothed the data better and minimizes the fitting error. Let the number of decomposition layers be j. When it is a sublayer not less than j, all coefficients are retained. When the number of decomposition layers is i ($0<i<$j), just keep the coefficient with the largest absolute value. There are K such coefficients. Let the general signal be α and the detailed signal is d, and the specific formula is defined as:

$$K = \frac{M}{(j+2-i)^{\alpha}}. \quad (1)$$

In formula (1), both M and α are empirical coefficients, and the default value is L(1), which is the length of wavelet coefficients after the first layer decomposition of M, namely M=L(1); As a rule of thumb, if you use the Brige-Massart strategy for signal compression, α = 1.5; and when denoising, α = 3. Based on the determination of the threshold, the paper uses MATLAB to carry out the simulation experiment. It is concluded that the optimal decomposition layer of the four parameters of the engine speed, airspeed, cabin temperature and humidity of the telemetry is 2, ie j = 2.

According to the principle of wavelet layered threshold, when the threshold of noise limited wavelet coefficients is determined, an optimal threshold function is needed to filter the noisy wavelet coefficients effectively to remove the noise coefficients [19-20]. In order to improve the signal distortion caused by data hopping and excessive smoothing of traditional soft and hard threshold function denoising, the article improves and optimizes on the basis of the traditional threshold function, so that the function has continuity, the threshold can change with the change of the decomposition scale, reduce the deviation between the wavelet coefficient and the original coefficient, and improve the smoothness of the curve. In this way, the measurement accuracy and applicability of the wavelet function construction method can be maximized.

First, introduce a contractible equation based on the hard threshold function so that there is continuity at point $\pm\lambda$. Since the amplitude and density of noise increase with the increase of the decomposition scale, it is not appropriate to use

the same threshold to process the wavelet coefficients at each scale. Therefore, in order to solve the problem of excessive killing of wavelet coefficients using global thresholds, the paper uses the multi-layer threshold method generated by Brige-Massart strategy to estimate the threshold, that is, using different thresholds for estimation at each layer. Let the threshold $\lambda_j = \sigma\sqrt{2lnN}/\log(j+1)$, where j represents the decomposition scale, $\sigma = median(|\omega_{j,k}|/0.6745)$. The improved threshold function is defined as follows:

$$\bar{\omega}_{j,k} = \begin{cases} \omega_{j,k} - \frac{\lambda_j^n}{\omega_{j,k}} & |\omega_{j,k} \geq \lambda| \\ 0 & |\omega_{j,k} < \lambda| \end{cases} \quad (2)$$

In equation (2), $\omega_{j,k}$ is the wavelet coefficient, $\bar{\omega}_{j,k}$ is the estimated wavelet coefficient value (k is a positive integer), λ_j is the threshold on the scale j, and λ_j is reduced because the scale j is increased, so that $\bar{\omega}_{j,k}$ is close to $\bar{\omega}_{j,k}$. Can effectively overcome the effects of constant deviation. When $\omega_{j,k} \to \pm\lambda$, this is the case in higher order power functions. $\frac{\lambda_j^n}{\omega_{j,k}} \to 1, \left(\omega_{j,k} - \frac{\lambda_j^n}{\omega_{j,k}}\right) \to 0$, This shows that the improved threshold function is continuous at λ; in addition, due to the situation of $\omega_{j,k} \to \infty$, and then $\left(\omega_{j,k} - \frac{\lambda_j^n}{\omega_{j,k}}\right) \to \omega_{j,k}$, it is known that the advantages in the hard threshold function are preserved; it can be seen from the function definition of equation (2) that when n = 1 and $\omega_{j,k} \geq \lambda$ occurs, the following situation is obtained: $\bar{\omega}_{j,k} = \omega_{j,k} - \lambda_j$. When $\omega_{j,k} < -\lambda$ occurs, the following situation is obtained: $\bar{\omega}_{j,k} = \omega_{j,k} + \lambda_j$, indicating the improved threshold function and the soft threshold function is the same; when$n \to \infty$,the improved threshold function is similar to the hard threshold function; indicating that the smaller the value of the higher-order factor n is, the smoother the curve of the improved threshold function is.

V. Algorithm Experiment Research And Result Analysis

In order to verify the superiority of the improved threshold function in the denoising of telemetry data, the paper selects 200 airspeed experimental parameters to verify the algorithm based on matlab. First select the appropriate wavelet basis function; then determine the optimal number of decomposition layers of the signal; on this basis, different threshold functions are applied to reconstruct the wavelet signal to remove the noise signal; finally, the paper uses the error square sum (SSE) and curve similarity (NCC) two quantitative indicators to quantitatively analyze the denoising effect of different threshold functions. The smaller the sum of error squares and the larger the similarity value of the curve, the better the denoising effect of the function is, and it is used to smooth the parameters of the actual telemetry system.

The paper conducts an experimental study based on the telemetry airspeed parameter signal of the telemetry. The curve changes before and after the airspeed data of the aircraft are shown in Fig 3.

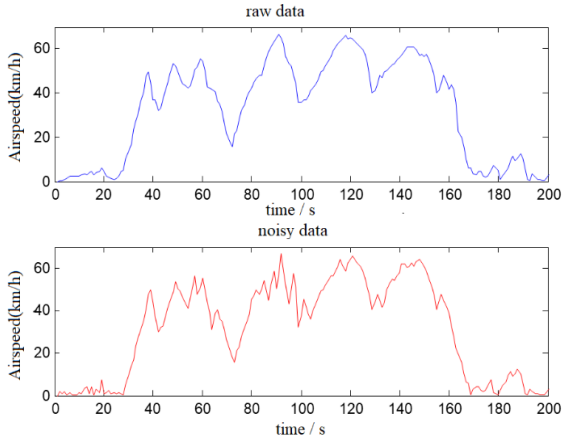


Figure 3. Airspeed raw data, noise data curve change diagram.

Based on the MATLAB platform, this paper uses two different wavelet bases, biorthogonal and symlets, to perform two-layer denoising on the noisy space velocity data. Taking the denoising

error value as the criterion, it is concluded that the noise-free airspeed signal is decomposed by the bior5.5 basis function, and the effect is the best.

In order to analyze the denoising superiority of the improved threshold function wavelet method, the paper uses three different threshold functions to perform two-layer wavelet decomposition on the space velocity data of noisy aircraft. The denoising minimum error squared sum value and the curve similarity value of different threshold functions are obtained by experiment. See Table Ⅰ, and the denoising curve display effect diagram is shown in Fig 4, 5, and 6, respectively.

TABLE I. THREE THRESHOLD FUNCTION AIRSPEED DATA DENOISING EVALUATION INDEX VALUE

Denoising Evaluation Index	Soft threshold	Hard Threshold	Improved Threshold Function
SSE	3046.4	2721.5	2477.4
NCC	0.985	0.988	0.990

It can be seen from the evaluation index values of the denoising of the noisy space velocity data by using three different threshold functions as shown in Table Ⅰ. The denoising error of the soft threshold function is the largest, indicating that the method will excessively denoise; the hard threshold function denoising relative soft threshold function denoising error value is significantly reduced, and the curve similarity is also improved; the improved threshold function has the best denoising performance, and the denoising accuracy and curve similarity are significantly better than the other two threshold functions.

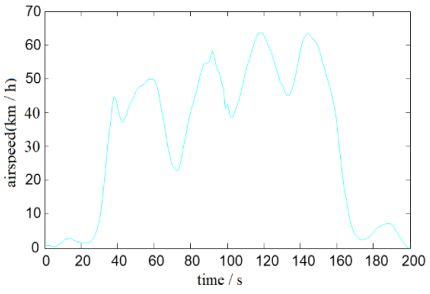


Figure 4. Soft threshold method airspeed denoising curve diagram.

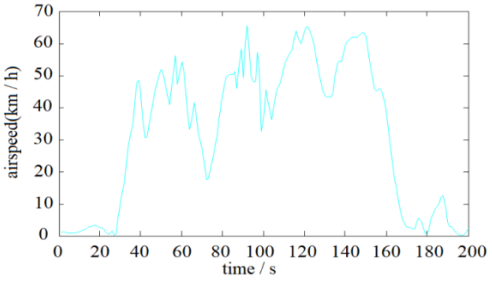


Figure 5. Hard threshold method airspeed denoising curve diagram.

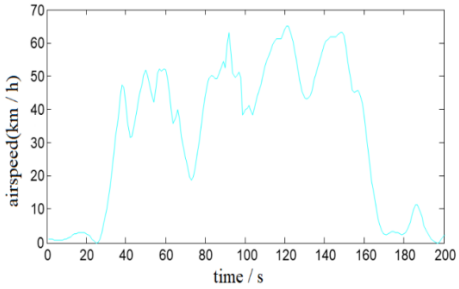


Figure 6. Hard threshold method airspeed denoising curve diagram.

Fig. 4 and Fig. 5 are graphs showing the denoising process of the noisy space velocity data shown in Figure 3 using the traditional soft and hard threshold functions, respectively. Combined with the denoising effect index value in Table Ⅰ, the soft threshold method is used to denoise the airspeed data as shown in Figure 6. The advantage is that it has the highest signal-to-noise ratio. The overall trend of the data curve after denoising is relatively smooth; However, the disadvantage is that the fitting error is the largest, especially for the data points between 40 and 60, the smoothness is too strong, some useful information is lost, and the data curve has the lowest similarity with the original data curve after denoising. The analysis of the airspeed data denoising effect of the hard threshold function shown in Figure 4 shows that although the denoising SNR is slightly lower than the soft threshold function, it has a lower error and the curve similarity is increased to 0.988. The denoising performance is generally superior to the soft threshold function; however, this method does not have continuous processing characteristics, and the smoothness is poor. It can be intuitively seen through the denoising curve diagram that the

data still has a glitch point, causing data oscillation, especially between data points 40-70 and 80-100. The data value mutation phenomenon is more serious.

Figure 6 is a graph showing the effect of the wave function on the denoising of the airspeed data after the threshold is improved. Comparing Figures 4, 5 and 6, it can be seen that the improved threshold method is used to denoise the airspeed data, and the denoising curve is smoother than the hard threshold function, which significantly improves the data glitch compared to the hard threshold function, especially at 80-100 data points between. The soft threshold function is oversmoothed between 40 and 60, while the improved threshold rule improves the problem. Combined with Table Ⅰ, it can be seen that the data denoising with the improved threshold wavelet method can not only effectively preserve the variation characteristics of the original data, and the similarity with the original data curve is as high as 0.990, and the influence of the error is also reduced. It also reduces the sum of squared errors after denoising to 2477.4. The denoising performance of this method is significantly better than the other two wavelet threshold methods.

In order to verify the applicability of the improved threshold function wavelet method in the data processing of telemetry system, the method is applied to the denoising processing of data such as engine speed, cabin temperature and humidity of telemetry aircraft. It is known that the signal-to-noise ratios of the three parameters after noise addition are -3.11989db, -0.7004db, 9.8033db, and the value of the variable n in the threshold function is still 3.5, and the interference signal is decomposed, and the number of layers is 2 layers. It can be found that the sum of the squares of the denoising errors for the four kinds of telemetry data and the value of the curve similarity are shown in Table Ⅱ.

TABLE II. IMPROVED THRESHOLD FUNCTION WAVELET CONSTRUCTION METHOD EXPERIMENTAL EFFECT EVALUATION FORM

Algorithm evaluation index	Engine Speed	Airspeed	Temperature Data	Humidity Data
SSE	462060	24774	24.682	21.835
NCC	0.998	0.989	0.978	0.810

From the value of the denoising effect evaluation index in Table Ⅱ. It can be seen that the wavelet method with improved threshold can effectively deal with the data noise, and the similarity of the denoising fitting curve to have strong applicability.

In order to improve the denoising accuracy of telemetry parameters, this paper studies the wavelet threshold method. The matlab simulation shows that the improved Garrote threshold wavelet is superior to the traditional threshold function in denoising performance, and compares the processing results of the airspeed data with the three algorithms mentioned above. The result shows that the denoising effect of the improved algorithm is effectively improved. Finally, in order to verify the applicability of the improved Garrote threshold wavelet method in the denoising of telemetry data, this paper applies the algorithm to the processing of many kinds of telemetry parameters. Through the denoising effect diagram and the denoising performance evaluation index, it is concluded that the algorithm can be better applied to the data processing of telemetry system. In order to effectively remove the data noise, make the measured value closer to the true value, and better maintain the aviation flight safety, the article deeply studies the effectiveness of wavelet transform in the denoising processing of telemetry data. The application characteristics of different thresholds and threshold functions in data denoising are mainly studied. This paper improves the defects of traditional soft and hard threshold functions in data denoising. A wavelet function construction method with improved threshold is proposed. The results show that the improved

threshold function has better denoising effect than the commonly used soft and hard threshold functions, and it can improve the data hopping phenomenon of hard threshold function and the excessive smoothing and denoising error of soft threshold function; and it can effectively remove multi-class telemetry data noise, with the highest measurement accuracy, the overall curve similarity is up to 80%, with good measurement accuracy and applicability.

VI. Conclusion

In order to ensure the flight safety of the aircraft, it is necessary to collect all kinds of parameters on the aircraft during flight test and transmit them to the ground acceptance equipment in real time for fault analysis by ground inspectors to ensure the flight safety of the aircraft. Because the remote sensing parameters of aircraft will be disturbed by various kinds in the process of acquisition and transmission, the measurement data will inevitably be affected by noise, such as spikes, burrs and other phenomena, which reduces the efficiency of fault analysis of ground detection. In order to avoid the interference caused by data noise, make the data display closer to the real value, and ensure the flight test safety of aircraft, this paper studies the denoising algorithm of telemetry parameter data, reduces the influence of error, and makes the measured value closer to the real value.

Firstly, the least squares method, five-point cubic algorithm and median filtering algorithm are studied. Through matlab simulation technology, the denoising of telemetry parameter signals is processed. It is concluded that the denoising of telemetry parameter signals has great limitations, it's fitting speed is slow, and the denoising effect is poor. Five-point cubic algorithm has certain effect on burr and cusp smoothing, and it can be used to a certain extent. Removal of data noise, but because of the discontinuity of smoothing function, there will be jumping phenomenon between different groups of data points, and the data curve can not be displayed smoothly. When the smoothing window is too small, the trend of data curve changes is quite different from the original data, and almost can not achieve the effect of drying. The bigger the window value, the smoother the curve will be, but it will be destroyed. The authenticity of the data and the denoising effect of the median filtering algorithm are general for the space velocity data with drastic changes.

Acknowledgment

This research is partially funded by the Project funds in Shaanxi province University Student Innovation and Entrepreneurship Fund Project (S S202110702080) and the Project funds in engineering laboratory project (GSYSJ2018013).

References

[1] Y. Jingfeng, "An Improved Adaptive Differential Evolution Algorithm," Journal of XuChang university, vol. 33, no. 2, pp.74-77, Mar 2014.

[2] M. Guanjun, D. Haibin, L. Senqi, Y. Yaxiang, "UCAV Path Planning Based on MAX-MIN Self-adaptive Ant Colony Optimization," Acta Aeronautica et Astronautica Sinica, vol. 29, pp.243-248, May 2008.

[3] Z. Bo, S. Lanxiang, Y. Haibin, X. yong, and C. Zhibo, "A method for improving wavelet threshold denoising in laser-induced breakdown spectroscopy," Spectrochimica Acta Part B: Atomic Spectroscopy, vol.107, pp.32-44, May 2015.

[4] L. Hui, W. Weida, X. Changle, H. Lijin, and N. Haizhao, "A de-noising method using the improved wavelet threshold function based on noise variance estimation," Mechanical Systems and Signal Processing, vol.99, pp. 30-46, Jan 2018.

[5] Z. Bing, and N. Shihong, "A Method of Pretreatment of Flight Data Based on Wavelet Analysis," Journal of missile and guidance, pp. 457-459, Aug 2015.

[6] L. Jian, W. Jingjing, "Path Smoothness Algorithm Based on Improved Kalman Filter," Radio Engineering, vol. 45, no. 5, pp.20-23, Apr 2015.

[7] S. Wentao and Z.Dan, "Noise reduction of ECG using spatial correlation filtering and stationary wavelet transform," 2010 5th International Conference on Computer Science & Education, pp. 1085-1088, Sep 2010.

[8] Z. Ruizhao, and C. Huimin, "Improved Threshold Denoising Method Based on Wavelet Transform," Physics Procedia, vol.33, no.1, pp.1354-1359, Jun 2012.

[9] R. Chao, S. Lei, L. Xianjian, "An Adaptive Wavelet Thresholding De-noising for Deformation Analysis,"

Geomatics And Information Science Of Wuhan University, ,vol.37, no. 7. Pp. 873-875. Jul 2012.

[10] S. Ruixia, L. da, and W. Xiaochun, "Low Illumination Image Enhancement Algorithm Based on HSI Color Space," Journal of Grahics, vol.38, no. 02, pp. 217-223, Apr 2017.

[11] Z. Fengbo, Z. Hongqiu, and L. Changgeng, "A pretreatment method based on wavelet transform for quantitative analysis of UV-vis spectroscopy," Optik - International Journal for Light and Electron Optics, vol.182, pp.786-792, Apr 2019.

[12] SHANG Li, ZHOU Yan, CHEN Jie, SUN Zhan-li. Image Denoising Using a Modified LNMF Algorithm [J], International Conference on Computer Science & Service System (CSSS), 2012: 1840-1843.

[13] FENG De-shan, DAI Qian-wei1, YU Kai1, GPR signal processing under low SNR based on empirical mode decomposition [J], Journal of Central South University (Science and Technology), 2012(02): 596-603.

[14] Omitaomu Olufemi A, Protopopescu Vladimir A, Ganguly Auroop R.Empirical Mode Decomposition Technique With Conditional Mutual Information for Denoising Operational Sensor Data [J], Sensors Journal, 2011(11): 2565-2575.

[15] Mouna Samaan, Stephen Cook.The Generation of Telemetry Frame Formats in a User-Friendly Environment, 1997, 337-341.

[16] Nicholas J. Murray, David A.The role of satellite remote sensing in structured ecosystem risk assessments.Science of the Total Environment, 2017.

[17] Ake Rosenqvist, Anthony Milne,Richard Lucas.A review of remote sensing technology in support of the Kyoto Protocol,Environmental Science&Policy 6(2003)441-455.

[18] Adam C. Watts, Vincent G, Ambrosia, Everett A. Hinkley. Unmanned Aircraft Systems in Remote Sensing and Scientific Research:Classification and Considerations of Use,Remote Sens. 2012, 4, 1671-1692.

[19] Esther Salami,Cristina Barrado, Enric Pastor.UAV Flight Experiments Applied to the Remote Sensing of Vegetated Areas,Remote Sens. 2014, 6, 11051-11081.

[20] Ken Whitehead,Chris H. Hugenholtz.Remote Sensing of the Environment with Small Unmanned Aircraft Systems(UASs), Part1: A review of progress and challenges.

Improved Random Forest Fault Diagnosis Model Based on Fault Ratio

Ziwei Ding
[1.] School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
[2.] State and Provincial Joint Engineering Lab of Advanced Network, Monitoring and Control
Xi'an, 710021, China
E-mail: 146377997@qq.com

Shunyuan Huang
[1.] School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
[2.] State and Provincial Joint Engineering Lab of Advanced Network, Monitoring and Control
Xi'an, 710021, China
E-mail: sylvianoemie@sina.com

***Abstract*—With the rapid development of information technology, the informatization, integration and complexity of more and more large equipment are increasing day by day, so it is very important to carry out fault diagnosis for such complex equipment. In the traditional way, expert system technology is usually used for fault diagnosis of complex equipment. However, with the increasing of equipment data information, traditional methods cannot solve the fault diagnosis requirements in the case of a large amount of data. Therefore, data-driven fault diagnosis method can solve this problem, The carrier of data-driven fault diagnosis is a large amount of engineering data, and its focus is to explore new methods of fault diagnosis from a large amount of historical data. In this paper, the classical random forest algorithm is selected as the basic model, and aiming at the imbalance of complex equipment data, the improved random forest voting mechanism based on the fault ratio is proposed to optimize the model, which makes the final model diagnosis accuracy more than 95%, and has good application value.**

Keywords-Complex Equipment; Fault Diagnosis; Random Forest; Unbalanced Data

I. INTRODUCTION

Along with the rapid development of information technology era, large equipment is more and more in different industries tend to electronic and complication, integration, and summarizes it is different in the field of large equipment increasingly tend to be intelligent, this development trend will largely increase the probability of equipment failure and the difficulty of the late breakdown maintenance, The traditional mode of "periodic maintenance" and "post-repair", such as manual scheduled maintenance and fault reprocessing, is no longer applicable to the current large and complex equipment. At the same time, a series of chain reaction caused by equipment failure will also cause serious safety accidents and bring huge economic losses. After long-term experience and practice, in order to ensure efficient, safe and reliable operation of equipment, reasonable use and in-depth study of fault diagnosis technology is particularly important [1].

The research on fault diagnosis was first carried out by NASA in the late 1960s. Since the research started, this technology has crossed many disciplines in other fields, and then derived many new fault diagnosis methods, attracting the attention of a large number of European developed countries. This technology has been applied to aviation, navigation, large-scale industrial projects, chemical industry and military fields in many countries. Westinghouse electric Company has been committed to the research of artificial intelligence expert system of power station since 1980s, and has achieved good results. Boeing, a civil aviation giant, has also developed IMA systems that combine artificial intelligence technology with fault diagnosis. Our country's Xiong Fanlun applied expert system in the field of agriculture, to achieve a more reasonable and convenient agricultural production [2].

Fault diagnosis refers to the technology of identifying and classifying the device status by collecting the current and historical status

information of the device. The purpose of this technology is to ensure the smooth and normal running of the system devices and avoid unnecessary emergencies. However, traditional fault diagnosis requires a high level of technical personnel in operation, and is not suitable for deeper diagnosis scenarios [3]. Therefore, with the continuous development of artificial intelligence and its derivatives, fault diagnosis technology has gradually realized the transformation to intelligent fault diagnosis. The core of intelligent fault diagnosis is to create an entity that can diagnose faults on devices as an "expert" and provide the same diagnosis results as traditional expert detection. At the same time, with the continuous development of machine learning, its performance in the field of fault diagnosis is becoming more and more excellent. Relevant research data of scholars show that the application of this technology to large and complex equipment can quickly identify faults, significantly improve the durability and reliability of equipment, and have universality and research ability [4]. In view of this equipment data imbalance, this paper choose machine learning in the classical model of random forest model based on random forest model in the history of unbalanced data sets, generally due to other machine learning model, and further introduce the basic principle of random forest algorithm, and then to the imbalance of a large number of complex equipment failure data based on the fault than improved random forest model, Experimental results show that the improved model has higher accuracy than the single random forest model.

II. BASIC PRINCIPLE OF RANDOM FOREST ALGORITHM

A. Random forest algorithm concept

Random Forest is a classification prediction model based on ensemble learning proposed by Leo Briman, academician of American Academy of Sciences in 2001. The smallest unit of the model is the decision tree. Intuitively speaking, every decision tree is a simple classifier. For an input sample, N decision trees will have N classification results. Random forest algorithm is a collection of all the classification results, the proportion of all results is the final decision results [5].

Generally speaking, each decision tree can be as a focus on a particular aspect of the referee, if only to listen to a referee rhetoric so there must be some deviation, but there are a number of referee each referee different perspective to deal with problems, eventually all vote the way the referee to determine the results, although there will be individual differences, But the overall prediction variance must be decreasing [6]. Compared with traditional classification algorithms, random forest algorithm has fast classification speed, large capacity of data processing, strong ability of error balance and difficult to over-fit. In addition, it is worth mentioning that the robustness of the random forest model depends on the number of decision trees. The more decision trees there are, the more precision of the model will increase. The random forest model is shown in the figure below.

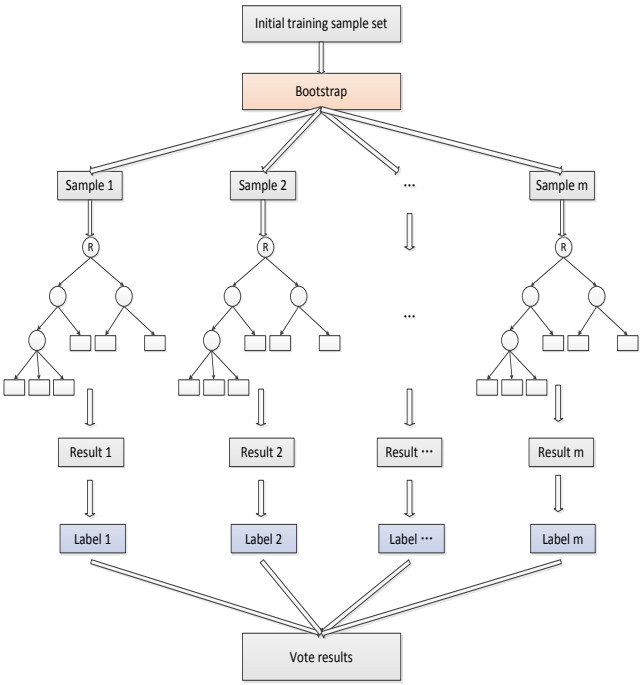


Figure 1. Random forest model

B. Random forest model construction process

The construction of random forest can be divided into four steps as shown in the following figure:

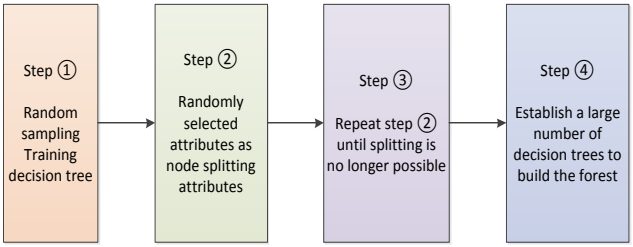


Figure 2. Random forest model construction process

1) Assume that there are currently N samples and select N samples by random extraction with put back. Random extraction with put back means that a sample is randomly selected from the data set every time. After the selection, the sample is put back into the data set and then returned to select new samples. The selected samples are used to train a decision tree;

2) When each sample has M attributes and each node in the decision tree needs to be split, M attributes are randomly selected from M attributes, and the condition to be met is M < Then, other indexes such as information gain and GINI value are used as splitting strategy to select the splitting attribute of this node;

3) A decision tree is established based on Step (2). Every node in the process of forming the decision tree should be split according to step (2). If the attribute selected by the node next happens to be the attribute used by its parent node when splitting, there is no need to split again. Repeat until it can no longer split;

4) Repeat all the above steps, and the completed decision trees constitute a complete random forest.

C. Advantages and disadvantages of random forest model

1) RF model is based on decision tree, so the model can also realize classification and regression functions [7]. However, the model is mostly applied to classification, and its advantages are as follows:

a) For features, the importance of each feature can be judged by tree structure;

b) In the case of feature missing or outliers, the model can also show good performance and still maintain accuracy;

c) The mutual influence of different features can be judged;

2) Compared with the advantages, the disadvantages are slightly insignificant. The model has the following disadvantages:

a) In dealing with regression problems, the performance of the model is not as good as that of dealing with classification problems;

b) When the external noise in the training sample set is relatively large, the over-fitting problem is more likely to occur;

c) In the case of imbalanced samples, it is impossible to accurately measure the contribution ratio of certain features to RF model.

III. IMPROVED RANDOM FOREST ALGORITHM DIAGNOSIS MODEL BASED ON FAULT RATIO

The imbalance of complex equipment failure data, the author of this paper chose random forests can better deal with unbalanced data model, but in practical engineering applications, the complex equipment failure probability is very low, so will lead to failure data and under normal circumstances the ratio of the equipment condition data samples is very small, the normal data sample size is far greater than the fault data sample size, We also refer to such data as highly unbalanced data sets [8].

A core part of the random forest model is the integrated voting process. In the basic RF model, the mode voting mechanism of "minority follows majority" is usually used to determine the final result. However, considering the extremely unbalanced data studied in this paper, the degree of imbalance will seriously affect the result of the mode voting. Random forest integrates multiple decision trees to achieve integrated learning. Although integrated learning can reduce the comprehensive error to a certain extent, the imbalance of the research object in this paper is too high, so it is particularly important to take corresponding measures. Therefore, this paper uses fault ratio to improve the voting decision [9].

A. Improved voting rules based on fault ratio

For complex equipment system, the number of fault classes is small, that is, the number of

positive classes is small. The number of normal classes is larger, that is, there are more negative classes. In this paper, the original mode voting mechanism is abandoned and the voting decision rules in RF model are improved by combining the ratio between the sample size of faulty data and the sample size of normal data, namely the fault ratio. The improved rule is described as follows: Check the category labels output by each decision tree in the forest. If the ratio of positive category labels to negative category labels is greater than the fault ratio, the final result is classified as positive category (small sample category). In this way, the original mechanism can be optimized, and the weight of positive and negative label decision trees in the forest is no longer fixed, which can better overcome the imbalance of positive and negative samples [10].

B. Improved random forest algorithm based on fault ratio

The implementation process of the improved random forest algorithm is shown in the following table:

Improved random forest algorithm based on fault ratio
Step1: The training samples were randomly put back from the data set, and were extracted for n times in total to obtain n independent training sets with repeated elements. Step2: The n decision trees are trained on different training sets. Step3: The sample category labels corresponding to N decision trees were analyzed, and the final voting induction was carried out by combining the improved voting decision method based on fault ratio.

IV. EXPERIMENT AND RESULT ANALYSIS

A. Introduction to complex equipment fault diagnosis data set

In this paper, the fault data set of a complex equipment is used, but considering the confidentiality of the relevant data of a complex equipment, the relevant feature attributes of the data set are expressed in the form of coding, and the specific situation of the data set is as follows:

1) Large amount of sample data. The data set provided is composed of training set and test set. The sample size of training set is 60000 (1000 positive classes and 59,000 negative classes). The sample size of the test set is 16000 (375 positive categories and 15625 negative categories), so the data set is rich in information and also in line with reality. Although complex equipment may have failures in real life, the occurrence of failures must be rare, so the proportion of normal and abnormal data in the original data set is 59:1, which is reasonable.

2) Rich sample attributes. Can be in the data set, each sample data are of the fire control system contains 171 features, it also means that each of the sample data is composed of 171 properties, research on the fault diagnosis of complicated equipment concerned about most is the word "complex", through the analysis of the data set can be obtained after data set features more, experimental complexity increase, It has reference value for the future study of fault diagnosis of complex equipment.

3) In essence, the fault diagnosis set is a typical binary data set with positive and negative imbalance.

B. Data preprocessing

1) Data missing value processing

Due to the high integration of complex equipment, some state data in the data set are missing. Therefore, the missing ratio of each feature is firstly calculated in feature engineering operation, and the results are shown in the form of bar graph as shown in the figure below.

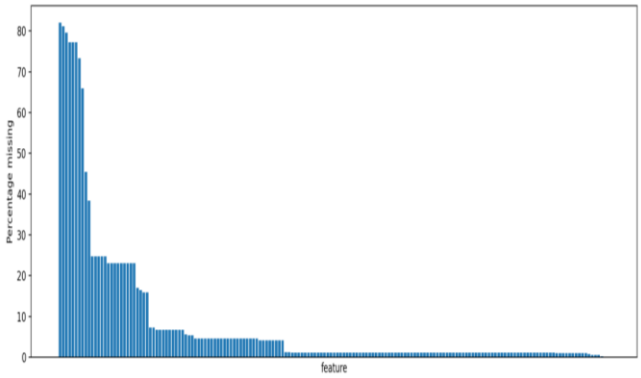


Figure 3. .Characteristic exact ratio

According to the figure above, there are a large number of missing attributes in the data set. In the operation of this paper, the attribute features with missing values greater than 75% were deleted, that is, six attribute columns were deleted, namely br_000, BP_000, BP_000, BO_000, AB_000 and CR_000.

Then, for the data with the missing percentage less than 75%, the missing value is completed. The main methods of missing value completion are median completion and mean completion. Here, the two methods are used to complete the data respectively, and the final method is measured by comparing the standard deviation of the data set after the two completions.

The standard deviation of each attribute after filling is shown in the following table:

TABLE I. DATA VARIANCE COMPARISON

Attribute name	Standard deviation after median supplement	The standard deviation of the mean
aa_000	1.454301e+05	1.454301e+05
ac_000	7.767625e+08	7.724678e+08
ad_000	3.504525e+07	3.504515e+07
ae_000	1.581479e+02	1.581420e+02
af_000	2.053871e+02	2.053753e+02
...	...	
ee_007	1.718666e+06	1.718366e+06
ee_008	4.472145e+05	4.469894e+05
ee_009	4.721249e+04	4.720424e+04
ef_000	4.268570e+00	4.268529e+00
eg_000	8.628043e+00	8.627929e+00

As can be seen from Table 1, for the attribute features with missing values, the standard deviation of the data after using the mean value is less than that after using the median value. Therefore, this paper adopts the mean value supplement method for the missing data.

2) Unbalanced data processing

Random forest has a good performance on unbalanced data sets, but the corresponding unbalanced processing of data sets is also very important.

Exist unbalanced data sets is to complex equipment fault diagnosis based on machine learning research a well-known problem in the process, there is no doubt that the small sample data in the process of research has played a more important role, and attract the attention of the researchers, in real life applications such small sample is a part of the researchers are more interested in.

The processing of unbalanced data sets can be divided into data level and algorithm level in general direction, as shown in the figure below:

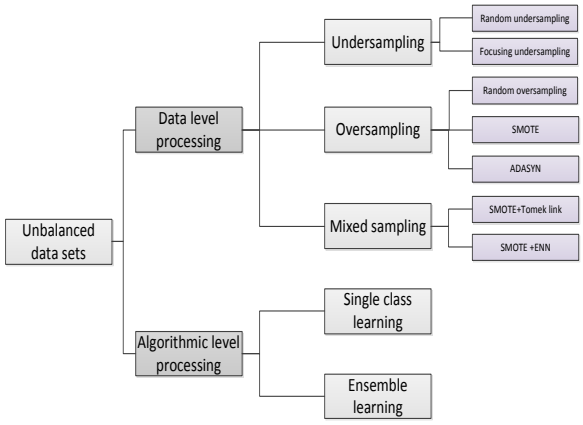


Figure 4. .Unbalanced data processing method

In this paper, ensemble learning is used at the algorithm level. Therefore, after comparing different sampling methods, we finally choose the improved method of smote borderline-smote1 as the solution for unbalanced data set processing. This method further divides a small number of samples into "safety", "danger" and "attention", and randomly selects a small number of samples of k-nearest neighbor in the "danger" attribute.

C. Introduction to complex equipment fault diagnosis data set

There are many super parameters in the random forest model. In this paper, the simple grid search method combined with 50% cross verification is used. On the premise of determining the node splitting index, the optimal parameters in the verification set are selected by changing the tree and the largest characteristic tree in the forest, and drawn into a broken line diagram, as shown below.

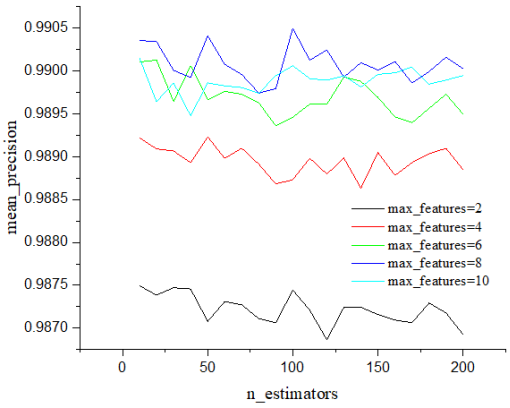


Figure 5. .Unbalanced data processing method

It can be intuitively observed from the figure that when the maximum feature number is selected differently, the corresponding average accuracy rate is also different. The average accuracy rate of the model here can be basically maintained above 0.99. To sum up, when the maximum feature number is 8, the number of decision trees in the forest is 100, and the splitting index is "Gini", the model achieves the best effect on the verification set.

D. Experimental results and comparison

After model fitting, the confusion matrix obtained on the test set is shown in the table below. According to the information in the table, the model correctly divides the normal data into 15611 and the fault data into 361. However, at the same time, 14 normal data are mistakenly divided into fault data and 14 fault data are mistakenly divided into normal data.

TABLE II. MODEL RESULTS

Forecast	Physical truth	
	Normal data	Fault data
Normal data	15611	14
Fault data	14	361

Combined with the model evaluation criteria, the evaluation results obtained according to the confusion matrix are as follows:

TABLE III. MODEL METRICS

Evaluation criteria	Result
ACC	96.94%
Precision	96.27%
Recall	96.27%
F1 score	0.9627
AUC	0.9701

In order to highlight the feasibility of the improved model based on fault ratio, the unmodified random forest model is used as the control experiment. The comparison of the experimental results is shown in the figure below.

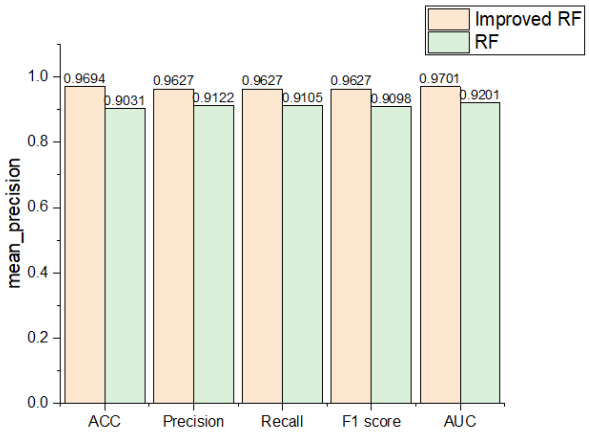


Figure 6. .Comparison of experimental results

Through the comparison of experimental results and indicators, it can be concluded that all indicators of the improved random forest model are better than the basic random forest model. The experiment shows that the random forest algorithm can better diagnose and process the unbalanced data, and the precision can reach more than 90%, but the accuracy rate of the model based on fault is better than that after improving the decision rules, and the precision can reach more than 96%. The improvement is feasible.

V. CONCLUSION

In view of the extremely unbalanced fault data of complex equipment, this paper proposes an improved random forest model based on fault ratio to realize the diagnosis task. The experimental results show that the performance of the model after using the improved method is better. The

importance of fault diagnosis in real life is self-evident. At present, there are many mainstream diagnosis methods, but it is very important to realize the analysis of "adjusting measures to local conditions" for different types of fields. Different machine learning algorithms may show different abilities in different data. This paper provides a new idea for the establishment of extremely unbalanced data model, In the follow-up research, We can study the following points: ①If the amount of data is large enough, fault diagnosis models can be built with the help of more emerging deep learning methods in recent years. ②In the creation of machine models, there are many choices for the selection of model parameters. In future research, intelligent optimization algorithms can be applied to the optimization of model hyperparameters. ③More sampling methods can be tried to solve the data imbalance problem.

REFERENCES

[1] Liu Zhantao Research on integrated method of condition monitoring and fault diagnosis of large equipment system [D] Beijing University of chemical technology, 2009.

[2] Xiong Fanlun Architecture and implementation of intelligent system technology for agricultural field [J] Pattern recognition and artificial intelligence, 2012, 25 (05): 729-736.

[3] Guo Zhi Research on fault diagnosis method of chemical machinery and equipment based on big data [J] Information recording materials, 2021,22 (09): 233-235.

[4] Zhang Peilin, Cao Jianjun, Ren Guoquan Research on condition monitoring system of large mobile complex equipment [J] Journal of Gun Launch and control, 2006 (03): 15-18.

[5] Xu Dongpo,Liu Yunqing,Wang Qian. Random forest-based human pose detection system for through-the-wall radar [J]. Journal of Physics: Conference Series, 2021, 1966(1).

[6] Sherif Ahmed Abu El-Magd, Sk Ajim Ali, Quoc Bao Pham. Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain [J]. Earth Science Informatics, 2021 (prepublish).

[7] Wu Weijie Research on application and optimization method of random forest algorithm [D] Jiangnan University, 2021.

[8] Dong Hongyao, Wang Yidan, Li Lihong. Overview of Random Forest Optimization Algorithms [J]. Information and Computer (Theoretical Edition), 2021, 33(17): 34-37.

[9] Sun Mingzhe, Bi Yaojia, Sun Chi. Overview of Improved Random Forest Algorithm [J]. Modern Information Technology, 2019,3(20):28-30.

[10] Wang Yisen, Xia Shutao. Overview of Random Forest Algorithm for Ensemble Learning [J]. Information and Communication Technology, 2018, 12(01): 49-55.

Air Attack Target Threat Assessment Based on Combination Weighting

Hong Li
[1.] School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
[2.] State and Provincial Joint Engineering Lab of Advanced Network, Monitoring and Control
Xi'an, 710021, China
E-mail: 3290910434@qq.com

Ruiqi Song
[1.] School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
[2.] State and Provincial Joint Engineering Lab of Advanced Network, Monitoring and Control
Xi'an, 710021, China
E-mail: 11118138@qq.com

Bailin Liu
[1.] School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
[2.] State and Provincial Joint Engineering Lab of Advanced Network, Monitoring and Control
Xi'an, 710021, China
E-mail: 498194312@qq.com

***Abstract*—Threat assessment is an important process of quantifying the threat of enemy attacking targets. It is also one of the main basis for commanders to make control decisions in air defense operations. Target threat assessment needs to obtain a large amount of air attack target information from various reconnaissance equipment and battlefield sensors, fuse these information, and get the ranking of the threat degree of air attack targets to our side. In view of the unbalanced distribution of index weight in threat assessment in air defense operations, a target threat assessment model based on combined weight is proposed in this paper. Firstly, according to the index system of air raid target threat assessment, the subjective and objective weights of the indexes are determined by analytic hierarchy process and critical method respectively, and the combined weights are calculated by multiplication synthesis method; Then the threat ranking of targets is obtained by TOPSIS method; Finally, the model is verified by an example. The simulation results show that the air target threat assessment model is reasonable.**

Keywords- Threat Assessment; Air Raid Targets; Combination Weighting; CRITIC; Multiplication Synthesis Method; TOPSIS

I. INTRODUCTION

In modern war air defense operations, with the wide application of various new technologies in the military field, the performance of air attack targets is higher and higher, and the mode of air attack has undergone qualitative changes, which makes modern air defense face more threats. Therefore, in the process of air defense battle command, the commander must analyze and evaluate the battlefield situation and threat of the enemy and ours according to air intelligence, and make decisions quickly and accurately in order to obtain the initiative. For the possible complex war environment and different threat factors of multiple incoming targets, it is very difficult to evaluate and judge them and establish a scientific and accurate evaluation model. Research on more accurate and credible threat assessment methods has become an important development trend in this field.

DOI: 10.2478/ijanmc-2022-0020

At present, there is no accurate definition of threat assessment. In the mid-1980s, the C3 Technical Committee under the joint directors of laboratories (JDL) established an information fusion expert group and developed a general information fusion processing model - JDL model. The model is mainly divided into five levels: information preprocessing, object refinement, situation assessment, threat assessment and excellent process. Its structure is shown in figure 1.

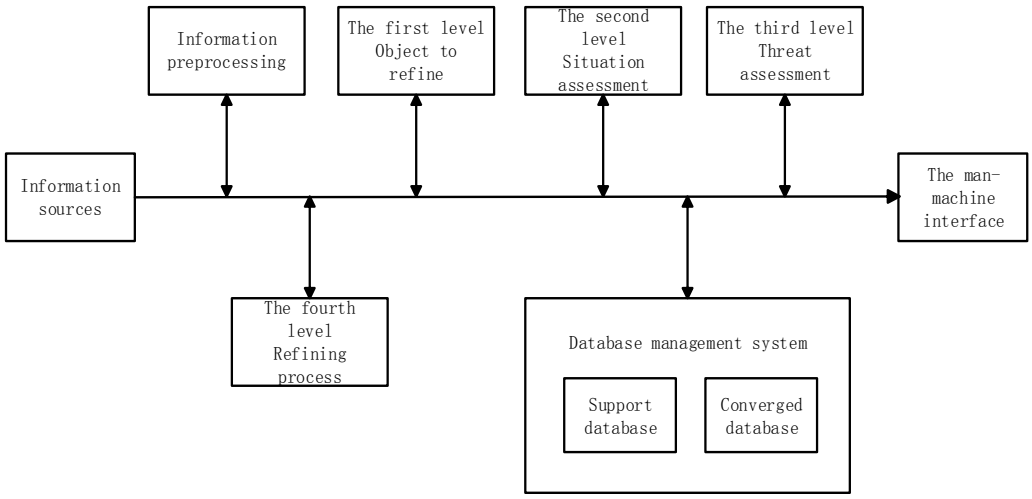


Figure 1. JDL model

Threat estimation is an important content of information fusion decision-making level. It is located in the third level in the model proposed by JDL. It is an important process to quantify the threat of enemy attacking targets. It is also one of the main basis for commanders to make control decisions in air defense operations. In the rapidly changing battlefield environment of information war, it is a very important work to quickly identify the original data of enemy targets obtained from a variety of complex sensors, obtain key information such as target type, position and speed through data preprocessing, judge the threat degree of incoming targets to our side, and provide data support for battlefield commanders to take corresponding combat deployment decisions. Many exploratory studies have been carried out on threat assessment at home and abroad. The main theories and methods are: multi-attribute decision-making method, fuzzy comprehensive evaluation method, grey correlation method and so on. Using these methods to evaluate the threat of air raid targets and determine the attribute weight of targets is a very important work, which is related to the reliability and correctness of target threat assessment results. The attribute weight of the target can be divided into subjective weighting method and objective weighting method according to its source. Subjective weighting method, such as analytic hierarchy process [2], whose index weight is flexibly determined by experts or commanders according to their own experience and battlefield situation, has great subjective randomness, and is also vulnerable to the lack of expert knowledge; Objective weighting method, such as entropy weight method, determines the weight according to the amount of information and correlation degree of indicators, which has a strong mathematical theoretical basis, but often ignores the subjective intention of decision-makers, and both of them have certain limitations.

For the threat assessment of air raid targets, there have been many assessment methods, but there are some problems in determining the index weight, such as over reliance on expert experience, unreasonable index weight distribution and too one-sided assessment results. Therefore, this paper proposes an air raid target threat assessment model based on combined weighting and TOPSIS method [14], which uses analytic hierarchy process and critical method to determine the subjective and objective weights of indicators respectively, taking into account both the subjective factors of experts and the correlation between indicators; The combination weight is

obtained by multiplication synthesis method; Using TOPSIS method to rank the threat of air raid targets [15]; Finally, an example of target threat assessment in air defense operation shows that the method is feasible and effective.

II. DETERMINE THE INDICATORS OF THREAT ASSESSMENT

In air defense operations, the threat target is an air attack target with the intention of attacking, threatening or even destroying our army's position. The threat assessment indicators affecting the target mainly include: target type, flight speed, flight altitude, arrival time, route shortcut, weapon type, penetration capability, jamming capability, etc. These indicators are not independent of each other, and they are more or less related to each other [1]. For example, the penetration ability, jamming ability and the number and type of weapons carried by the target are often determined by the target type. Moreover, since the threat assessment model is based on the fusion of target information collected by sensors, the assessment indicators should be operable in quantification. For the same batch of air raid targets, the fixed defense location will not affect the judgment of target threat [9]. As shown in Figure.2, the threat evaluation model can be established under the premise of fully identifying the target type of the paper.

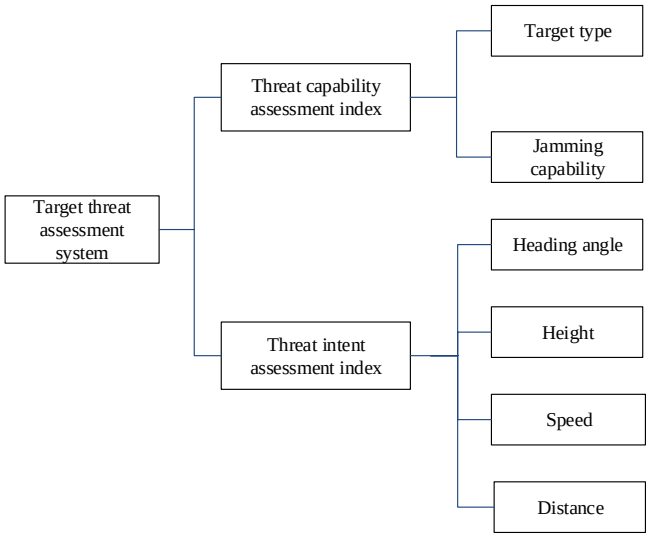


Figure 2. target threat assessment index system

Threat assessment indicators are divided into qualitative indicators and quantitative indicators. From the established threat assessment indicator system, it can be seen that threat capability assessment indicators belong to qualitative indicators, and the target height in threat intention assessment indicators is regarded as qualitative indicators. Here, G.A. Miller's 9-level quantitative theory is used to quantify three qualitative indicators: target type, target interference ability and target height. The incoming weapons for air defense operations in important places are mainly divided into missiles and aircraft. Among them, tactical ballistic missiles are a special kind of missiles, which will not be considered here. According to the actual air defense operations in the current naval key areas, the incoming targets are divided into large targets, small targets and armed helicopters in combination with the operational capability and reflection area of the incoming weapons, which are assigned as 8, 5 and 3 in turn. Among them, large targets can be divided into bombers, fighter bombers, assault aircraft; Small targets can be divided into cruise missiles, air to ground missiles, airborne missiles, stealth aircraft, etc. generally, the speed of such targets is relatively fast, and the maximum flight speed of airborne missiles and stealth aircraft can usually reach Mach 4 ~ 7 [3]; According to the target interference ability, it is divided into four levels: strong, medium, weak and none, with values of 9, 7, 5 and 3 respectively. According to the height of the incoming target, it is divided into high altitude, hollow altitude, low altitude and ultra-low altitude, which are quantified as 3, 5, 7 and 9 in turn.

III. AIR ATTACK TARGET THREAT ASSESSMENT MODEL BASED ON COMBINATION WEIGHTING

A. Determine subjective weight

Among the determination of subjective weight, the most common is analytic hierarchy process. Analytic hierarchy process (AHP) introduces Saaty's 9-level scoring system [13], uses human experience and judgment to quantify the influencing factors of the system hierarchically, constructs the judgment matrix through pairwise comparison, and calculates the relative weight of

the lower level elements of the adjacent level to the upper level elements according to the weight solution method. The specific calculation steps are as follows:

1) Consult experts to get the expert's judgment on the importance of each index and the expert judgment matrix u. The expert's judgment on the importance of each index is shown in Table I.

TABLE I. INDEX IMPORTANCE JUDGMENT

Equally important	Slightly important	Strong importance	Strongly important	Extremely important
1	3	5	7	9

2) Conduct consistency inspection. Test the consistency of U's thinking.

3) The weight of judgment matrix U is solved by analytic hierarchy process and normalized.

Punctuate equations with commas or periods when they are part of a sentence, as in

$$W_j = \sqrt[n]{\prod_{i=1}^{n} u_{ji}} \quad j = 1, 2, \cdots, n \tag{1}$$

$$\alpha = W \Big/ \sum_{j=1}^{n} W_j \quad j = 1, 2, \cdots, n \tag{2}$$

B. Determine objective weight

The common method to determine the objective weight is the information entropy method [6], but this method only considers the amount of information of the index value and ignores the correlation between the indexes. Therefore, this paper introduces the critical method to determine the objective weight of the index [12]. Critical method comprehensively determines the index weight according to the contrast strength and conflict between the evaluation indexes. At the same time, considering the difference and correlation between the indexes, critical method has the advantages of high reliability and independent of expert knowledge background. It is a more scientific objective weighting method. The calculation steps are as follows:

1) Determine the contrast strength S_i is

$$S_j = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(d_{ij} - \bar{d}_j \right)^2} \quad j = 1, 2, \cdots, n \tag{3}$$

Conflict *correlation coefficient* r_{ki} *is*

$$r_{kj} = \operatorname{cov}\left(D_k, D_j \right) \Big/ \left(s_k, s_j \right) \quad \text{k,j=1,2,}\cdots,n \tag{4}$$

2) Determine the comprehensive information content of each index

$$G_j = s_j \sum_{k=1}^{n} \left(1 - r_{kj} \right) \quad \text{j=1,2,}\cdots,\text{n} \tag{5}$$

3) The weight coefficient between indicators is determined by the comprehensive information of each indicator, Represented by β

$$\begin{cases} \omega = \left(\omega_1, \omega_2, \cdots, \omega_n \right) \\ \beta = \omega_j = \dfrac{G_j}{\sum_{j=1}^{n} G_j} \quad \text{j=1,2,}\cdots,\text{n} \end{cases} \tag{6}$$

C. Determine the combination weigh

Many experts usually use subjective weighting method or objective weighting method to determine the weight of the index system of the main attack direction of the enemy's incoming target. The subjective method mainly relies on the battlefield experience of battlefield experts to give a certain weight to the relevant battlefield indicators. This method relies too much on expert experience, which will lead to errors due to the lack of knowledge in the field of experts; Objective method refers to collecting relevant battlefield data according to various types of sensors to determine the weight of relevant battlefield indicators, ignoring expert experience and violating the principle of people-oriented in the battlefield.

In order to simultaneously consider the objective information obtained by the sensor and the experience judgment ability of the commander,

and make up for the defect of single subjective and objective weighting, this paper uses the multiplication synthesis method in literature [4] to determine the comprehensive coefficient of subjective and objective weighting, so as to ensure that under the premise of data analysis and excavation, combined weighting can be carried out according to expert criteria and the specific actual situation of air combat. The specific calculation is as follows.

$$\varepsilon_j = \frac{\alpha_j \times \beta_j}{\sum_{j=1}^{n} \alpha_j \times \beta_j}, j = 1, 2, \cdots, n \tag{7}$$

Inside, and β_j are the subjective and objective weights of air raid targets respectively, and ε_j are the weights of the ε_j index.

D. Threat assessment based on TOPSIS

The multi-attribute decision-making theory comprehensively considers multiple factors in the target threat, and can comprehensively reflect the impact of multiple factors on the evaluation. TOPSIS is a relatively mature multi-attribute decision-making method. TOPSIS theory normalizes the original data matrix, sorts and compares the decision-making schemes by calculating the weighted standardization matrix, finds out the optimal scheme (positive ideal scheme) and the worst scheme (negative ideal scheme) among the alternatives, and then calculates the distance between a scheme and the optimal scheme and the worst scheme, so as to obtain the proximity between the scheme and the optimal scheme, And take it as the basis for evaluating the advantages and disadvantages of each scheme [5]. The specific steps are as follows:

1) Construct objective decision matrix $D_{m\times n}$.

2) Determine the combination weight of each threat assessment index W.

3) Determine weighted decision matrix $\bar{A} = \left(\bar{a}_{ij}\right)_{m\times n}, \bar{a}_{ij} = D_{ij} W_j$.

4) Determine the ideal solution $\bar{a}_j^+$ *and negative ideal solution* $\bar{a}_j^-$.

$$\begin{cases} \bar{a}_j^+ = \max\limits_{1\le i\le n}\left(\bar{a}_{ij}\right) \\ \bar{a}_j^- = \min\limits_{1\le i\le n}\left(\bar{a}_{ij}\right) \end{cases} \tag{8}$$

5) Calculate the distance from each target to the ideal solution and negative ideal Solution d_i^+、d_i^-.

$$\begin{cases} d_i^+ = \sum\limits_{j=1}^{m} \left\| \bar{a}_{ij} - \bar{a}_j^+ \right\| \\ d_i^- = \sum\limits_{j=1}^{m} \left\| \bar{a}_{ij} - \bar{a}_j^- \right\| \end{cases} \tag{9}$$

6) Calculate the closeness between each target and the ideal solution T_i^+ *and rank the threats.*

$$T_i^+ = \frac{d_i^-}{d_i^+ + d_i^-} \tag{10}$$

Judge the target threat according to the calculated proximity T_i^+ .The greater the comprehensive evaluation Index T_i^+ , the greater the target threat; The smaller the comprehensive evaluation index T_i^+ ,the smaller the target threat [16].

IV. EXAMPLE SIMULATION

In terms of data selection, on the one hand, it should conform to the reality of air defense operations in important places, on the other hand, the data selection should not lose generality, and the importance of evaluation index factors should be highlighted, so as to fully verify the effectiveness of the method. Therefore, the initial simulation data of literature [8] is used for simulation analysis. Suppose that in an air defense battle, there are the following 6 groups of air attack targets. The target threat degree constitutes the evaluation index system according to the target type, speed, heading angle, jamming ability, air

raid altitude and distance. We obtained the threat evaluation index parameters of these six batches of targets through various sensors, as shown in Table II.

According to the calculation method in Section 1, the attribute values of the six incoming target threats are quantified, and the threat values are shown in Table III.

1) Determine subjective weights. According to AHP subjective weighting method, the calculated subjective weight is:

α=(0.2594, 0.2227, 0.0238, 0.1225, 0.0909, 0.0663)

2) Determine objective weights. First, normalize the threat attribute values, as shown in Table IV

TABLE II. THREAT INFORMATION OF AIR ATTACK TARGET

	Target type	**Speed(m/s)**	**Heading angle（° ）**	**Jamming capability**	**Air raid altitude**	**Distance（km）**
target 1	large	400	5	strong	hollow altitude	100
target 2	large	720	8	strong	hollow altitude	150
target 3	small-scale	1600	3	none	low altitude	300
target 4	small-scale	1200	5	none	low altitude	260
target 5	large	280	10	weak	ultra-low altitude	140
target 6	helicopter	100	15	medium	ultra-low altitude	120

TABLE III. TARGET ATTRIBUTE THREAT VALUE

	Target type	**Speed(m/s)**	**Heading angle（° ）**	**Jamming capability**	**Air raid altitude**	**Distance（km）**
target 1	5	400	5	8	4	100
target 2	5	720	8	8	4	150
target 3	8	1600	3	2	6	300
target 4	8	1200	5	2	6	260
target 5	5	280	10	4	8	140
target 6	3	100	15	6	8	120

TABLE IV. STANDARDIZATION DECISION TABLE

	Target type	**Speed(m/s)**	**Heading angle（° ）**	**Jamming capability**	**Air raid altitude**	**Distance（km）**
target 1	0.4	0.2	0.8333	1	0	1
target 2	0.4	0.4133	0.5833	1	0	0.75
target 3	1	1	1	0	0.5	0
target 4	1	0.7333	0.8333	0	0.5	0.2
target 5	0.4	0.12	0.4167	0.3333	1	0.8
target 6	0	0	0	0.6667	1	0.9

According to equation (3) ~ equation (6), the objective weight is calculated as:

$$\beta = (0.1836, 0.1294, 0.2706, 0.1753, 0.1144, 0.1267)$$

3) Determine the combination weight. According to equation (7), the combination weight is calculated as:

$$W = [0.1813, 0.2100, 0.0993, 0.1561, 0.2400, 0.1133]$$

4) The weighted normalized decision matrix can be obtained from the combination weight as follows:

$$\bar{A} = \begin{Bmatrix} 0.0725 & 0.042 & 0.0827 & 0.1561 & 0 & 0.1133 \\ 0.0725 & 0.0868 & 0.0579 & 0.1561 & 0 & 0.0850 \\ 0.1813 & 0.2100 & 0.0993 & 0 & 0.1200 & 0 \\ 0.1813 & 0.1540 & 0.0827 & 0 & 0.1200 & 0.0227 \\ 0.0725 & 0.0252 & 0.0414 & 0.0520 & 0.2400 & 0.0904 \\ 0 & 0 & 0 & 0.1041 & 0.2400 & 0.1020 \end{Bmatrix}$$

5) According to equation (8) and equation (10) of TOPSIS method, the relative closeness of target threat is , It can be seen from this that the ranking results of the threat size of the six groups of air raid targets are as follows: target3> target 4> target 2> target 1> target 5> target 6.

Figure 3. Comparison of simulation decision and sample decision results

It can be seen from Figure.3 that the ranking results are basically consistent with the decision results of the original samples in document [10], which shows the feasibility and effectiveness of this threat assessment method. It can be seen intuitively from the figure that when using this method for threat assessment, there is a large gap between the threat degree of each target, so the decision-maker can get the threat ranking of air raid targets more quickly.

V. USING THE TEMPLATE

For the threat assessment of air raid targets, in order to make good use of the objective information obtained by sensors in the threat assessment model and integrate the subjective experience and command preference of command decision-makers, this paper uses the method of combined weighted TOPSIS to establish the threat assessment model. Due to the impact of different weighting methods on threat assessment indicators, the target threat degree is also different. In the process of threat assessment, commanders flexibly use the combined weighting method combining analytic hierarchy process and critic method according to the battlefield situation to reasonably determine the weight of assessment indicators, use the ranking method approaching the ideal solution to quickly and accurately assess the threat of air raid targets, and quickly implement fire attack on targets with high threat, which has a certain auxiliary decision-making function. When the sensor detection ability is limited, the commander's experience judgment ability should be appropriately added in order to further improve the threat assessment method. The proposed method of determining the subjective weight has strong operability, and the commander's decision-making opinions are easy to quantify. However, when calculating the combination weight, the multiplication formula is easy to amplify the difference, which makes the evaluation result not objective enough. In the future research, we should further improve the above deficiencies.

REFERENCES

[1] Hanyu Li Research on target threat estimation based on iterative decision tree and BP neural network [D] China Ship Research Institute, 2018.

[2] Kun Zhang, Deyun Zhou TOPSIS method combined with entropy weight and group AHP for multi-objective threat assessment [J] Journal of system simulation, 2008 (07): 1661-1664.

[3] Xuesong Tang, Lihong Guo, Chen Changxi Research on threat assessment and ranking model based on AHP [J] Microcomputer information, 2006 (27): 35-38.

[4] Hongbo Zhou, Jincheng Zhang Grey target threat assessment based on combination weight [J] Firepower and command and control, 2018,43 (10): 143-147.

[5] Yinghao Hao, Yongli Zhang, Chuan Lei, Caihui Chen Simulation of air target threat assessment based on combined weighting TOPSIS method [J] Tactical

missile technology, 2015 (05): 103-108 DOI:10.16358/j.issn. 1009-1300.2015.05.18.

[6] Su Zhang Air target threat assessment technology [J] Intelligence Command and control system and simulation technology, 2005 (01): 41-45.

[7] Tao Zhang, Zhongliang Zhou, Xinyu Gou Target threat assessment and ranking based on information entropy and TOPSIS [J] Electro optic and control, 2012,19 (11): 35-38.

[8] Changjin Wang, Yonghui Zhang, Bin Huang Air Defense Threat Assessment of important places based on grey fuzzy matter-element analysis [J] Firepower and command and control, 2013,38 (08): 47-50 + 54.

[9] Haiyong Sun, Yangye Chen, Huigang Han Threat assessment and ranking of air raid targets based on grey clustering [J] Journal of Air Force Radar Academy, 2011,25 (05): 355-357 + 361.

[10] Yuan Zhou, Jun Yan, yuan sun, Jihui Xu, Huajie Lu Threat assessment model of important air defense targets based on Bayesian network [J] Journal of Naval Aeronautical Engineering College, 2015,30 (05): 467-472.

[11] Jia Guo Research on air target threat assessment method based on multi-attribute decision-making [D] Beijing University of technology, 2017 DOI:10.26948/d.cnki. gbjlu. 2017.000236.

[12] Chengzhe Fang, Yingxin Kou, an Xu, Shijie Deng, Mingyu Peng VIKOR air combat threat assessment based on ahp-critical combination weighting [J] Electro optic and control, 2021,28 (02): 24-28.

[13] Mingshuang Zhang, kehu Xu, Lingzhi Li Multi target threat assessment based on intuitionistic fuzzy set and VIKOR method [J] Journal of ordnance and equipment engineering, 2019,40 (06): 62-67.

[14] Jiawei Wu, Lin Zhou, Yong Jin, Junwei Li, Huanyu Liu Air target threat assessment based on subjective and objective combination [J] Command information system and technology, 2022,13 (01): 22-29 DOI:10.15908/j.cnki. cist. 2022.01.004.

[15] Ruijie Yin, Zuobin Yang, Cuixia Wu, Junjia Yang TOPSIS method for dynamic threat assessment of air raid weapons based on grey theory [J] Naval Electronic Engineering, 2021,41 (11): 107-110.

[16] Yang Gao, Dongsheng Li, Aixia Yong Threat assessment of target recognition system based on combined weight [J] Firepower and command and control, 2016,41 (05): 39-42 + 46.