Recognition of Blocking Categories for UWB Positioning in Complex Indoor Environment

Yaguang Kong, Chuang Li *, Zhangping Chen and Xiaodong Zhao

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China; ygkong@hdu.edu.cn (Y.K.); czp.apple@hdu.edu.cn (Z.C.); xdzhao@hdu.edu.cn (X.Z.)

* Correspondence: JasonLee1781@163.com; Tel.: +86-188-8892-9980

Abstract: The recognition of non-line-of-sight (NLOS) state is a prerequisite for alleviating NLOS errors and is crucial to ensure the accuracy of positioning. Recent studies only identify the line-of-sight (LOS) state and the NLOS state, but ignore the contribution of occlusion categories to spatial information perception. This paper proposes a bidirectional search algorithm based on maximum correlation, minimum redundancy, and minimum computational cost (BS-mRMRMC). The optimal channel impulse response (CIR) feature set, which can identify NLOS and LOS states well, as well as the blocking categories, are determined by setting the constraint thresholds of both the maximum evaluation index, and the computational cost. The identification of blocking categories provides more effective information for the indoor space perception of ultra-wide band (UWB). Based on the vector projection method, the hierarchical structure of decision tree support vector machine (DT-SVM) is designed to verify the recognition accuracy of each category. Experiments show that the proposed algorithm has an average recognition accuracy of 96.7% for each occlusion category, which is better than those of the other three algorithms based on the same number of CIR signal characteristics of UWB.

Keywords: UWB indoor positioning; NLOS status recognition; CIR signal characteristics; mRMR feature selection; DT-SVM

1. Introduction

1.1. Research Status

Outdoor positioning technologies such as the Global Positioning System (GPS) can achieve good results in outdoor positioning, but they cannot achieve good positioning results in indoor positioning. Note that indoor position perception plays an important role in many applications such as tracking. Thus it is important and necessary to do more investigation and exploration on indoor positioning technologies. Among the existing indoor-positioning systems, ultra-wide band (UWB) technology has become one of the most promising methods due to its high precision, good time-delay resolution, low power consumption, and high robustness in complex indoor environments. In particular, it has significant advantages over other indoor positioning technologies in terms of accuracy [1]. The UWB positioning system shows the best performance under line-of-sight (LOS) propagation conditions, and the positioning accuracy can reach centimeter level. However, due to the complexity of the indoor environment, there are often barriers between fixed base station (FS) and mobile base station (MS), namely non-line-of-sight propagation (NLOS) conditions, which will lead to the blocking of the signal propagation path between FS and MS and introduce positive deviation [2]. The propagation of UWB under NLOS will cause the measured distance between FS and MS to be greater than the actual distance, then result in a sharp decline in the final positioning accuracy. Therefore, the status recognition of

NLOS in the UWB indoor positioning system is crucial to improving the performance of the UWB-based positioning system, thus it has become one of the research hotspots of many scholars [3–6].

Note that NLOS state recognition algorithms based on distance estimation [7–9] and channel impulse response (CIR) features of UWB [10–16] are most commonly used to identify NLOS propagation state. However, compared with the CIR feature method, the distance estimation will cause extra delay due to the collection and calculation of distance. Moreover, the real-time performance of these two methods is often unsatisfactory, and the detection accuracy often fails in meeting actual demands. Furthermore, six features (energy, maximum amplitude, rise time, average excess delay, root mean square of delayed spread, and kurtosis) were extracted by collecting a large amount of experimental data under LOS and NLOS conditions and analyzing the CIR waveforms [10]. Least square support vector machine (LS-SVM) was utilised to classify LOS and NLOS states, but the relevance of features was not considered, so the recognition accuracy was limited. A method was proposed to identify LOS and NLOS states by identifying the signal characteristics under LOS/NLOS using the maximum likelihood ratio [11,12]. Kurtosis, mean additional delay and root mean square of delayed propagation were extracted, and maximum likelihood ratio of features was used to identify LOS and NLOS conditions [12]. However, this method is more suitable for the identification of multipath effect under NLOS, but not fully applicable to an indoor complex environment. The NLOS condition was identified by extracting kurtosis from CIR [13] and by obtaining four different features (skewness, kurtosis, RMS delay and mean excess delay) from CIR [14]. The energy of the first path in CIR was obtained as the recognition condition of NLOS [15]. Eight features were acquired from CIR [16] (standard deviation and skewness were added on the basis of literature [10]), and the convolution method was used to determine that 30 points in CIR could achieve good detection results. However, the above results did not consider the redundancy between the features, and few features would have a limitation on the detection accuracy. In addition, there is also the use of an inertial measurement unit (IMU) combined with UWB for NLOS identification [17,18], but it would utilize a second sensor.

Among the 8 CIR signal features selected in the existing research, there are often correlations between different features, and there are redundant features. Therefore, feature selection is required when identifying the propagation status of LOS and NLOS in the UWB indoor environment. Feature selection algorithms are mainly divided into two categories: one is a classifier-dependent mode (Wrapper), and the other is a classifier-independent mode (Filter) [19–26]. Compared with Wrapper mode with poor real-time performance, Filter mode has received extensive attention because it does not rely on classifiers, and it has high efficiency and scalability in reducing feature dimensions. The mutual information method in Filter mode mainly relies on the size of the mutual information value between the features and the target object to sort the features, while the classic information gain (IG) algorithm is to sort the features according to the correlation between the features and the tags [19]. A feature selection method of dynamic mutual information was proposed, which combined more information measures to construct a general criterion function [20]. However, as the feature dimension increases, the computational complexity also increases. A feature selection algorithm combining information gain and divergence was proposed for classification, but the selection index of divergence was increased, which increased the computational complexity [21]. The mutual information feature selection algorithm with uniform information distribution (MIFS-U) was proposed, which improved the penalty factor of the evaluation function in the MIFS method and used the uncertainty coefficient to describe the degree of redundancy between features, and the degree of relevance of the selected feature set and category was introduced into the penalty factor [22]. A minimum-redundancy maximum-relevance feature search algorithm and criterion (mRMR) were proposed [23]. A non-linear feature selection algorithm based on forward search was proposed [24], which used mutual information and interactive information. However, neither MIFS-U nor mRMR considers whether the first feature selected is the feature that contributes the most to the whole. An algorithm combing the mutual information criterion with the greedy search strategy was proposed, which used the nearest neighbor estimation method to estimate the mutual information and select the feature subset [25]. This algorithm has a good effect

on the selection of high-dimensional features. In addition, some scholars also studied other methods for feature selection [26]. For example, a genetic algorithm was used for the CIR features of UWB selection [27]. A classical sorting algorithm ReliefF based on feature distance was introduced [28]. A backward recursive feature selection algorithm SVM recursive feature elimination (SVM-RFE) was also proposed [29]. However, the computational time complexity of these algorithms is not as good as that of feature selection methods based on mutual information. Therefore, the performance of feature selection methods based on mutual information is better for higher dimensions.

All the above methods only study the recognition of the propagation of LOS and NLOS in the indoor environment, and the influence of NLOS state on the positioning accuracy of UWB. However, the contribution of blocking category information to the perception of indoor space information is ignored, and the recognition of different blocking categories is not studied in depth. In addition, in the aforementioned CIR signal feature selection method applied to UWB, the calculation cost of different features and the number of CIR signal features are often not comprehensively considered. Some algorithms cannot achieve good feature selection results when the feature dimension is low. Meanwhile, the first feature selected by these algorithms is not the feature that contributes the most to the whole, and the computational cost of some features is often higher than other features.

1.2. Contributions

In order to address these problems of the above methods, this paper not only studies the recognition of LOS and NLOS states, but also considers the contribution of the effective recognition of each blocking category to indoor spatial information perception, and realizes the recognition of different blocking categories, which can be used for the indoor positioning and indoor information perception by providing more useful information. In the experiment, the blocking categories of NLOS state are expanded, including water, human body, metal, wall and wooden board, to reflect the complex indoor environment. In addition, 30 effective signal points are extracted from the CIR to improve the calculation efficiency of eight features, and the influence of different features on various shielding is analyzed.

In addition, considering the detection speed in practical application, the calculation cost of different features is introduced into the feature evaluation index. A bidirectional search algorithm based on the feature evaluation index of maximum relevance, minimum redundancy and minimum calculation cost is proposed to determine the optimal feature subset. In addition, based on the method of vector projection, the separability between the classes of measurement is realized, the decision tree support vector machine DT-SVM hierarchy is designed to complete the identification accuracy of each blocking category verification. More importantly, compared with the research of other scholars, this paper not only studies the LOS and NLOS state recognition, but also studies the recognition of different blocking categories in the NLOS state. Experiments have verified that the method proposed in this paper is superior to other algorithms in the recognition of LOS and NLOS states, but has achieved good recognition results for all blocking categories, with an average accuracy of 96.7%. This paper provides a new research method for UWB subsequent indoor positioning, indoor information perception construction and target tracking.

2. Experimental Environment and Channel Impulse Response (CIR) Signal Characteristic Model

2.1. Hardware Equipment and Experimental Environment

This test uses two UWB devices, one is a fixed base station BS as a receiving device and the other is a mobile base station MS as a transmitting device. DW1000 chip is used for each UWB base station, and the time difference of arrival (TDOA) method is used to calculate the distance. The operating frequency range of DW1000 used by BS and MS is 3244 MHz to 6999 MHz. The maximum transmission power density is −35 dBm/MHz. The communication rate is 6.8 Mbps. This experiment was conducted in a large experimental center with a relatively complicated indoor environment (listed in Figure 1). The whole room is divided into data acquisition areas for training sets and data acquisition areas for

testing. In the complex indoor environment, the blocking categories usually include metal blocking (such as iron cabinets), wooden blocking (such as tables and chairs), human blocking, and water blocking (such as fish tanks and buckets of water). Therefore, in the experimental environment, the aforementioned several blocking materials were simulated to conduct experiments to verify the blocking effects of different materials on UWB. Data collection can be divided into six categories: LOS condition (no blocking) and water blocking, people blocking, metal blocking, wall blocking and wood board blocking. The physical conditions for each blocking category are as follows: In the experiment, a metal iron cabinet with 70 cm × 50 cm × 200 cm is placed between the MS and the BS, and the blocking angle of the iron cabinet is adjusted to complete the frontal blocking and the side blocking to achieve metal blocking; 1 to 2 people stand between the MS and BS at different angles to achieve human blocking. A plastic transparent water tank 80 cm × 40 cm × 80 cm, is filled with water and placed between the MS and BS to achieve water blocking. The BS is placed in the laboratory room, and the MS is placed in the corridor, separated by a concrete wall to block the wall. A wooden board 150 cm × 15 cm × 150 cm is placed between the MS and BS to block the board.
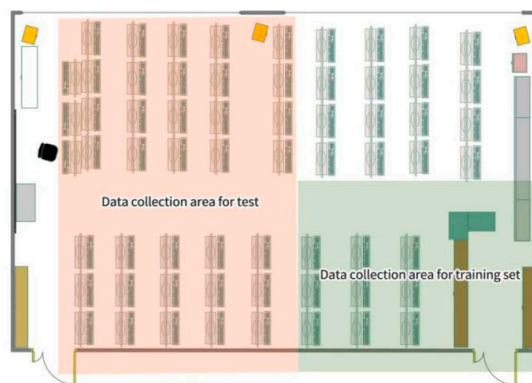

Figure 1. Experimental test environment.

With the same distance between MS and BS, in the data collection area of the training set, each blocking category was repeatedly tested 10,000 times, and a total of 60,000 CIR samples were collected as the training set. In the data collection area of the test set, each blocking category was repeatedly tested 1000 times and a total of 6000 CIR samples were collected as the test set. The true distance between BS and MS is measured by Leica DISTO D5 laser rangefinder.

2.2. CIR Description

Channel state information (CSI) includes the whole process of the signal from the transmitter to the receiver, including signal fading, multipath channel, transmission path and a series of information. It is usually used to describe the transmission process of a wireless channel. CSI can be measured by channel impulse response (CIR), which is expressed as:

$$h(t) = \sum_{i=1}^{N} a_i \delta(t - \tau_i) \tag{1}$$

where N is total of propagation paths, a_i and τ_i are the amplitudes and time delays of path i, and δ is the impulse function.

The CIR signals received by BS are shown in Figure 2. It can be seen that most of the 1015 signal points are noise (blue point area), while only a few are useful information (red point area). To improve the system efficiency, only 100 points of CIR data in the red area are collected.

The 100 CIR signal points of six class data (no blocking, water blocking, people blocking, metal blocking, concrete wall blocking and wood board blocking) are shown in Figure 3. However, it can be clearly seen from Figure 3a that the CIR signal differences of several types of features only concentrate

on 30 points, which is confirmed in literature [15], while the remaining points cannot contribute much difference information. Therefore, this paper extracts 30 useful signal points from the 100 points, as shown in Figure 3b.
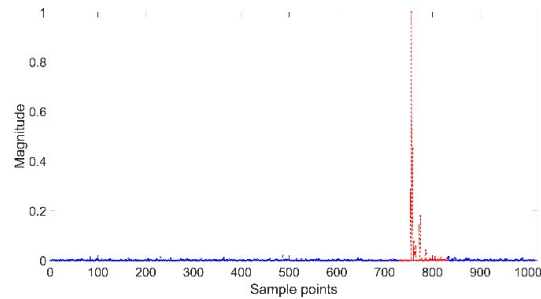

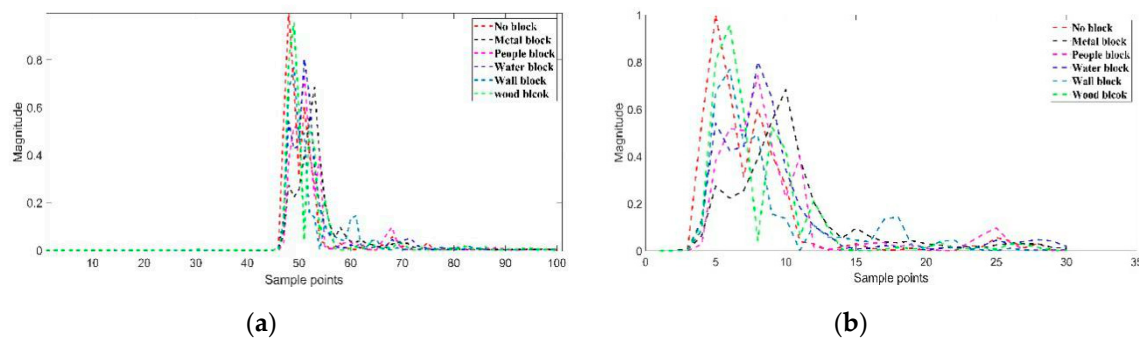**Figure 2.** 1015 CIR signal points.

| (**a**) | (**b**) |

Figure 3. The different channel impulse response (CIR) signal points of 6 categories: (**a**) 100 CIR signal points of 6 categories; (**b**) 30 CIR signal points of 6 categories.

2.3. Construct CIR Signal Features

By analyzing the CIR signals in the LOS and NLOS states under different blocking materials, the maximum amplitude of the signal in the NLOS state is usually smaller than that in the LOS state due to the blocking and attenuation of the signal during propagation. The energy of the first path is much less than the LOS state because of signal attenuation. In addition, the reflected, diffracted, or dispersed pulses after blocking are merged after the first path, resulting in the path amplitude after the first path in the condition of NLOS usually being larger than that in the condition of LOS, making the data below the state of NLOS closer to the mean and smaller standard deviation. At the same time, because blocking will cause delayed propagation in the NLOS state, the rise time, mean excess delay and RMS delay spread are selected as characteristics. Kurtosis and skewness are selected as characteristics due to the energy dispersion in the LOS and NLOS states. The discrete mathematical models of the above eight features are as follows:

(1)　The energy of the CIR signal:

$$E_f = \sum_{t=1}^{N_f} h(t)^2 \tag{2}$$

$$E = \sum_{t=1}^{N} h(t)^2 \tag{3}$$

where E and E_f represent the energy of all CIR signals and the energy of the CIR signal reaching the first peak respectively, and N and N_f represent all the points and the number of points reaching the first peak respectively.

(2)   Maximum amplitude:

$$h_{\max} = \max(h(t)) \tag{4}$$

(3)   The rise time to the maximum t_{rise}
(4)   The standard deviation σ
(5)   Mean excess delay: τ_m

$$\tau_m = \frac{\sum\limits_{t=1}^{N} t\left|h(t)\right|^2}{E} \tag{5}$$

(6)   RMS delay spread: τ_{RMS}

$$\tau_{RMS} = \sqrt{\frac{\sum\limits_{t=1}^{N} (t - \tau_m)^2 \left|h(t)\right|^2}{E}} \tag{6}$$

(7)   Kurtosis:

$$\kappa = \frac{E[(\left|h(t)\right| - u)^4]}{\sigma^4} \tag{7}$$

where u is the mean of $h(t)$.

(8)   Skewness:

$$\gamma = \frac{E[(\left|h(t)\right| - u)^3]}{\sigma^3} \tag{8}$$

This paper calculates the cumulative distribution of different occlusion categories under each feature. Figure 4 shows the cumulative distribution of each feature of different occlusion categories under a 30-point CIR signal, indicating that the first path energy and the maximum amplitude under LOS status are bigger than under the NLOS condition. Rise time, standard deviation, average excess delay and delay propagation RMS under LOS status are significantly less than the NLOS condition. However, it can be seen from Figure 4 that there is no significant difference between the features of LOS and wood blocking. Meanwhile, the difference between the measured distance and the actual distance when the 15 cm thick wood is blocked is similar to the difference under LOS, which verifies that wood has little effect on the attenuation of the incident electromagnetic field when the indoor wood moisture content is low. Therefore, the following work mainly focuses on the feature selection and the best feature set of no blocking, water blocking, people blocking, metal blocking and wall blocking.
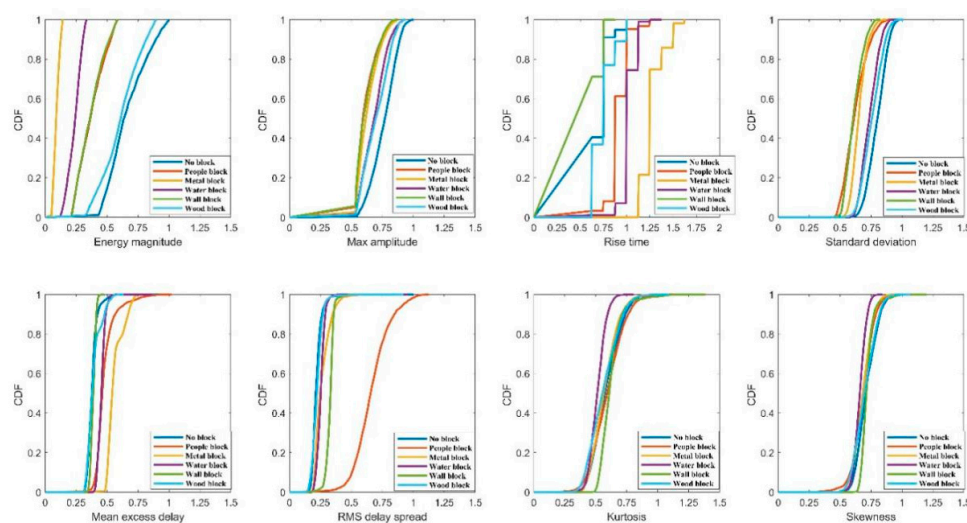

Figure 4. The cumulative distribution of features of different blocking classes at 30 CIR signal points.

3. Feature Combination Selection Algorithm and Multi-Classifier Design

3.1. System Model Description

Given a dataset $\mathbf{D}$ containing k samples, $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k\}$, where k represents the number of samples. In dataset $\mathbf{D}$, each kind of sample $\mathbf{d}_i = \{\mathbf{X}_i, \mathbf{L}_i\}$ has feature set $\mathbf{X}_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$, where $n = 1, 2, \ldots, 8$ represents the eight CIR signal: features of the first path energy, maximum amplitude, rise time, standard deviation, mean excess delay, RMS delay spread, kurtosis and skewness. $\mathbf{L}_i = \left\{l_i^1, l_i^2, \ldots, l_i^m\right\}$, where $m = 1, 2, \ldots, 5$ respectively represents no blocking, water blocking, people blocking, metal blocking and wall blocking. $l_i^m = \{1, -1\}$, if $l_i^m = 1$, it is labeled as a sample of class m; otherwise $l_i^m = -1$. The meanings of all symbols in the article are shown in Table A1.

Define $\mathbf{B} = \{b_1, b_2, \ldots, b_n\}$, $1 \leq n \leq 8$, n is the total number of selected feature; b_i, $i = 1, 2, \ldots, n$, is the index of the feature that meet the set criteria for evaluating features.

3.2. Mutual Information and Relevance Definitions

The relationship between features can be divided into relevance, redundancy, interaction and independence according to mutual information. The definition of a feature's relevance, redundancy, independence, irrelevance and interactivity is well defined in literature [25].

Mutual Information(MI) $I(\mathbf{X}; \mathbf{Y})$ is used to describe the degree of relevance between two random variables $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ and $\mathbf{Y} = \{y_1, y_2, \ldots, y_m\}$ [19]. When $I(\mathbf{X}; \mathbf{Y})$ is large, it means the relevance between X and Y is strong; otherwise, it means the relevance is small. In particular, when $I(\mathbf{X}; \mathbf{Y}) = 0$, it means that $\mathbf{X}$ and $\mathbf{Y}$ are independent of each other.

$$I(\mathbf{X}; \mathbf{Y}) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_i)} \tag{9}$$

where $p(x, y)$ is the probability density of x and y.

Conditional mutual information (CMI) can reflect the relevance between different features under the same label. Suppose there are three random variables set $\mathbf{X}, \mathbf{Y}$ and $\mathbf{L} = \{l_1, l_2, \ldots, l_d\}$, and their conditional probability density is $p(x|l), p(y|l), p(x, y|l)$, respectively. In the case of a given $\mathbf{L}$, then the mutual conditional information of $\mathbf{X}$ and $\mathbf{Y}$ about $\mathbf{L}$ is:

$$I(\mathbf{X}; \mathbf{Y}|\mathbf{L}) = \sum_{k=1}^{d} p(l_k) \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j|l_k) \log \frac{p(x_i, y_j|l_k)}{p(x_i|l_k)p(y_j|l_k)} \tag{10}$$

Based on the relevance between features and categories, a measurable multivariate model $\mathbf{F}$ can be obtained, and the specific form of $\mathbf{F}$ is as follows:

$$\mathbf{F} = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1m} \\ I_{21} & I_{22} & \cdots & I_{2m} \\ \vdots & \vdots & & \vdots \\ I_{n1} & I_{n2} & \cdots & I_{nm} \end{bmatrix} \tag{11}$$

where I represents the mutual information between each feature and the corresponding category, and the relevance between the 8 features and the label is shown in Figure 5.

In general, the matrix $\overline{\mathbf{F}} = [\bar{I}_{ij}]$ can be obtained by standardizing and centralizing the data. Calculate the feature's mutual information matrix $\mathbf{Q} = \overline{\mathbf{F}} \cdot \overline{\mathbf{F}}^T$. A $n \times n$ dimensional symmetric matrix is obtained, in which each element represents the mutual information between features under the same category, as shown in Figure 6.

$$\hat{\bar{I}}_{ij} = I_{ij} - \frac{\max I_{ij} - \min I_{ij}}{2} \tag{12}$$

$$\bar{I}_{ij} = \frac{\hat{I}_{ij}}{\sqrt{\sum\limits_{j=1}^{m} (\hat{I}_{ij})}} \tag{13}$$

where I_{ij} represents the mutual information between each feature and the corresponding category, and the relevance between the 8 features and the label is shown in Figure 6.
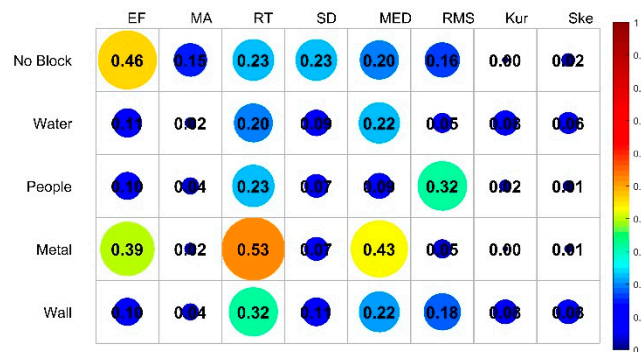

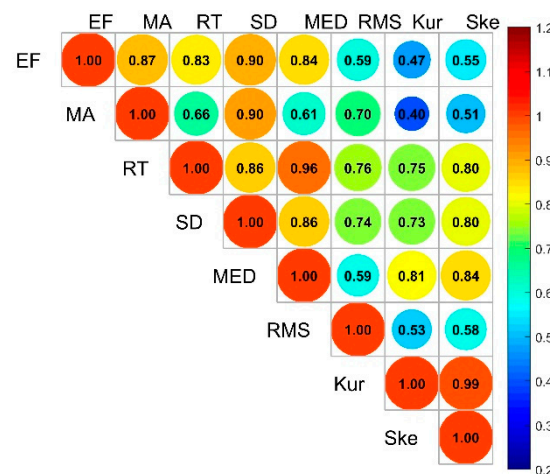**Figure 5.** The relevance between features and labels.

Figure 6. The relevance between features and features under the same label.

3.3. Construction of Bidirectional Search Algorithm Based on Maximum Correlation, Minimum Redundancy, and Minimum Computational Cost (mRMRMC) Feature Evaluation Criteria

The mRMR (max-relevance and min-redundancy) algorithm [23] is proposed on the basis of the mutual information-based feature selection algorithm (MIFS). The mRMR algorithm not only considers the relevance $D(x_i, \mathbf{L})$ between features and labels, but also considers the redundancy $R(x_i, x_j)$ between features under the same label.

$$D(x_i, \mathbf{L}) = \frac{1}{|\mathbf{L}|} \sum_{l_j \in \mathbf{L}} I(x_i; l_j) \tag{14}$$

$$R(x_i; x_j) = \frac{1}{|\mathbf{X}|} \sum_{x_j \in \mathbf{X}} I(x_i; x_j) \tag{15}$$

where $|\mathbf{L}|$ represents the number of labels in tag set $\mathbf{L}$, and $|\mathbf{X}|$ represents the number of features in feature set $\mathbf{X}$.

The mRMR evaluation index is:

$$J_{mRMR}(x) = \frac{1}{|\mathbf{L}|}\sum_{l_i \in \mathbf{L}} I(l_i; x) - \frac{1}{|\mathbf{B}|}\sum_{x_i \in \mathbf{B}} I(x_i; x) \tag{16}$$

where $I(l; x)$ is the mutual information between feature x and category label l, $I(x_i; x)$ are the mutual information between feature x and the selected feature x_i, and $\mathbf{B}$ represents the selected feature subset. With the increase of the number of features in the selected subsets, the calculation of the system will increase accordingly, the time cost of feature computation is introduced. Big O notation is usually used to express time complexity. Since only the highest term is retained and the coefficients are ignored, the time complexity can only be roughly described. Therefore, the calculation time complexity of each feature is expressed by calculating the number of statement execution times of each feature in this paper. Through the calculation formula of different features, the number of execution statements for each feature is obtained, and the calculation cost of each feature is obtained after ignoring the constant term.

$$C(\mathbf{B}) = \sum_{x_i \in \mathbf{B}} \beta_i c(x_i) \tag{17}$$

$$\beta_i = \frac{c(x_i)}{\sum\limits_{x_i \in \mathbf{X}} c(x_i)} e^{-\left(\frac{D(x_i, \mathbf{L})}{\sum\limits_{x_i \in \mathbf{X}} D(x_i, \mathbf{L})}\right)} \tag{18}$$

where $C(\mathbf{B})$ represents time cost of the selected feature set $\mathbf{B}$, $c(x_i)$ represents the calculated cost of feature x_i, and β_i is the relative weight of the calculation cost.

Finally, the comprehensive evaluation criteria based on maximum relevance, minimum redundancy and minimum computational cost are determined as mRMRMC:

$$J_{mRMRMC}(x_i) = \left(\frac{1}{|\mathbf{L}|}\sum_{l_j \in \mathbf{L}} I(x_i; l_j) - \frac{1}{|\mathbf{B}|}\sum_{x_j \in \mathbf{B}, j \neq i} I(x_i; x_j) - \omega\beta_i c(x_i)\right) \tag{19}$$

where $x_i \in \mathbf{B}_s$ represents the features in the selection area; ω is an adjustable constant, indicating the weight of calculation cost in the evaluation index, and under this evaluation index, the selected set $\mathbf{B}$ shall be satisfied.

$$\mathbf{B} = \mathbf{B} \cup \text{argmax}_{x_i \in \mathbf{B}_s}(J_{mRMRMC}(x_i)) \tag{20}$$

3.4. The Optimal Feature Set Is Determined Based on Multiple Threshold Constraints

In the traditional feature selection algorithm based on mutual information, such as mRMR, MIFS, IG, etc., the feature with the greatest relevance to the category is put into the selected feature subset as the first feature. However, these methods often ignore the combination effect between features, and do not consider whether the selected first feature is the feature that contributes the most to the whole. In this paper, a bidirectional search strategy based on mRMRMC, namely BS-mRMRMC algorithm, is proposed. First, calculate the relevance $D_{all}(\mathbf{X}, \mathbf{L})$ of all features and labels, and then inversely delete feature x_i to obtain the relevance $D_{\text{miss}(x_i)}(x_i, \mathbf{L})$ of the feature set and labels after missing x_i. The feature that causes the greatest change in relevance is selected as the first feature vector of the selected combination. After that, the features in $\mathbf{B}_s$ to be selected are selected in a positive sequence, and the features with the largest evaluation index $J_{mRMRMC}(x_i)$ are added to the selected $\mathbf{B}$, and the best feature subset is finally determined.

$$D_{all}(\mathbf{X}, \mathbf{L}) = \frac{1}{|\mathbf{L}|}\sum_{x_i \in \mathbf{X}}\sum_{l_j \in \mathbf{L}} I(x_i; l_j) \tag{21}$$

$$D_{miss(x_i)} = \frac{1}{|\mathbf{L}| - 1} \sum_{x_j \in \mathbf{X}, x_j \neq x_i} \sum_{l_k \in \mathbf{L}} I(x_j; l_k) \tag{22}$$

If $x_i = \text{argmax}(D_{all}(\mathbf{X}, \mathbf{L}) - D_{miss(x_i)}(x_i, \mathbf{L}))$, then feature x_i provides the most information for the whole, so x_i is taken as the first feature of the selected set, $\mathbf{B}(1) = x_i$.

Reverse deletion ensures that the first feature selected must be the feature that provides the most information for the whole, avoids that the features which are most relevant to the tag are not the feature that provides the most information for the whole due to the combination effect.

This paper makes the system model more efficient and stable by adding three constraint conditions: minimum relevance constraint, computational cost constraint and maximum evaluation index constraint. The selection of the optimal feature subset based on BS-mRMRMC algorithm under three constraints is shown in Figure 7.

Input:	Sample data set **D**, features set **X**, category tag set **L**;		
Output:	Best subset of features $\mathbf{B}_{best}$;		
First feature of the selected feature subset by the reverse deletion strategy		**Best feature subset selection by the forward search strategy**	
(1)	Set $\mathbf{B}_{best} = \emptyset$, B = Ø, η_d, η_c, η_j, $k = 2$;	(12)	**for** $k \leq |\mathbf{X}|$
(2)	**for each** $x_i \in \mathbf{X}, l_j \in \mathbf{L}$	(13)	**if** $((C(\mathbf{B}) > \eta_c) \| (\sum_{x_i \in \mathbf{B}} J_{mRMRMC}(x_i) > \eta_j))$
(3)	The relevance between features and labels is calculated according to formula (9) and formula (14–15) $D(x_i, \mathbf{L})$, $R(x_i, x_j)$;	(14)	Break;
(4)	**if** $D(x_i, \mathbf{L}) < \eta_d$	(15)	**else for each** $x_i \in \mathbf{B}_s$
(5)	Remove x_i in the **X**;	(16)	Calculated evaluation index $J_{mRMRMC}(x_i)$ according to formula (19);
(6)	**end if**	(17)	**end for**
(7)	**end for**	(18)	find $\mathbf{B}(k) = \text{argmax}_{x_i \in \mathbf{B}_s}(J_{mRMRMC}(x_i))$; $\mathbf{B}_s = \mathbf{B}_s - x_i$;
(8)	**for each** $x_i \in \mathbf{X}$	(19)	**end for**
(9)	Calculate $D_{all}(\mathbf{X}, \mathbf{L})$ and $D_{miss(x_i)}$ according to formula (21–22);	(20)	Return $\mathbf{B}_{best} = \mathbf{B}$;
(10)	**end for**		
(11)	Determines the first feature of the selected set: $B(1) = \text{argmax}(D_{all}(\mathbf{X}, \mathbf{L}) - D_{miss(x_i)}(x_i, \mathbf{L}))$; $\mathbf{B}_s = \mathbf{X} - \mathbf{B}(1)$;		

Figure 7. The solution process of the proposed algorithm.

Constraint 1: minimum relevance constraint. Set the threshold of relevance between feature and label η_d. Once the feature is related to the label $D(x_i, \mathbf{L}) < \eta_d$, the feature x_i is directly discarded to reduce the complexity of later calculation and improve the system efficiency. Where η_d is selected according to the relevance between the global feature vector and the label:

$$\eta_d = \lambda_d \max\{D(x_i; \mathbf{L}), x_i \in \mathbf{X}\} \tag{23}$$

where $0 < \lambda_d \leq 1$ represents the parameter of minimum relevance threshold.

Constraint 2: computational cost constraint. Set the computational cost constraint threshold η_c, $x_j = \text{argmax}_{x_i \in \mathbf{B}_s}(J_{mRMRMC}(x_i))$, if $C(\mathbf{B}) + c(x_j) > \eta_c$, then the best subset of features is directly determined to be $\mathbf{B}_{best} = \mathbf{B}$. The computing cost constraint threshold η_c is determined according

to the complexity of computing system characteristics and the computing power of the system in practical application.

$$\eta_c = \lambda_c \sum_{x_i \in \mathbf{X}} c(x_i) \tag{24}$$

where $0 < \lambda_c \le 1$ represents the parameter of computing cost threshold.

Constraint condition 3: maximum evaluation index constraint. Set the maximum evaluation index constraint threshold η_j, Indicates that the recognition accuracy of the system can reach the actual requirement η_j after x_j is added, the best subset of features $\mathbf{B}_{best} = \mathbf{B} \cup x_j$.

3.5. Decision Tree Support Vector Machine (DT-SVM) Classifier Design

SVM, one of the most commonly used classifiers, was proposed by Vapnik et al. based on statistical learning theory and its learning method [30]. Due to its strong learning ability and generalization ability, SVM has been widely used in multi-classification problems with small samples. However, SVM was originally designed to handle two types of classification tasks. For multi-classification support vector machines, N-class tasks can be converted into multiple two-class tasks as in one-against-all SVM (OAA-SVM) [31], one-against-one SVM (OAO-SVM) [32], decision directed acyclic graph SVM (DDAG-SVM) [33], and the multi-classification method decision tree SVM (DT-SVM) [34]. For N-class classification problems, DDAG-SVM and DT-SVM only need to construct the decision surface, which greatly improves the training and detection speed. But the recognition accuracy of DDAG-SVM and DT-SVM is not ideal due to the accumulation of classification errors. In order to solve this problem, both [35] and [36] proposed the use of the particle swarm optimization algorithm DT-SVM. Although this method improves the classification accuracy, it also increases the corresponding computational complexity and has a good effect on more classification. The structure of DT-SVM classifier is designed by different methods [37–40], but there are always problems such as limited accuracy improvement and limited adaptability.

Therefore, considering the actual number of categories in this paper, the vector projection method to realize the separability measurement between classes is adopted. Accordingly, the hierarchical structure of the DT-SVM classifier is designed to realize the recognition of each blocking category. The method based on vector projection was proposed by Li et al. [41], by which the number of intersecting samples between two types of samples could be better distinguished. The method based on vector projection was proposed by Li et al. [41], by which the number of intersecting samples between two types of samples could be better distinguished.

Set the sample set $\mathbf{D}_l = \{\mathbf{d}_{l1}, \mathbf{d}_{l2}, \ldots, \mathbf{d}_{li}\}$ of class l, indicating that there are k_l samples of class l. $\mathbf{d}_{li} = \{x_{i1}, x_{i2}, \ldots, x_{ij}\}$, $j = 1, 2, \ldots, n$, indicating that each sample contains n features. Let:

$$\mathbf{m}_l = \frac{1}{k_l} \sum_{i=1}^{k_l} \mathbf{d}_{li} \tag{25}$$

where $\mathbf{d}'_{li}$ is the projection of sample $\mathbf{d}_{li}$ to the feature direction of this class, Euclidean distance from $\mathbf{d}'_{li}$ to $\mathbf{m}_l$ is $\|\mathbf{m}_l \mathbf{d}'_{li}\|_2 = \|\mathbf{m}_l - \mathbf{d}'_{li}\|_2$.

Let: $d = \|\mathbf{m}_1 - \mathbf{m}_2\|_2$ represents the distance between the two classes $\mathbf{D}_1 = \{\mathbf{d}_{11}, \mathbf{d}_{12}, \ldots, \mathbf{d}_{1i}\}$ $\mathbf{D}_2 = \{\mathbf{d}_{21}, \mathbf{d}_{22}, \ldots, \mathbf{d}_{2i}\}$. And

$$\begin{cases} r_1 = \max_{\mathbf{d}_{1i} \in \mathbf{D}_1} (\|\mathbf{m}_1 \mathbf{d}'_{1i}\|_2) \\ r_2 = \max_{\mathbf{d}_{2i} \in \mathbf{D}_2} (\|\mathbf{m}_2 \mathbf{d}'_{2i}\|_2) \end{cases} \tag{26}$$

In order to measure the separability between classes, the separability measure value se_{12} is calculated. If $r_1 + r_2 \ge d$, the number of samples in $\mathbf{X}_1$ that satisfy $d - r_2 \le \|\mathbf{m}_1 \mathbf{d}'_{1i}\|_2 \le r_1$ and

$\|\mathbf{m}_1\mathbf{d}'_{1i}\|_2 \leq d$ is n_1. In the same way, the number of samples in $\mathbf{X}_2$ that satisfy $d - r_1 \leq \|\mathbf{m}_2\mathbf{d}'_{2i}\|_2 \leq r_2$ and $\|\mathbf{m}_2\mathbf{d}'_{2i}\|_2 \leq d$ is n_2. Then, the separability measure between classes is defined as se_{12}.

$$se_{12} = \frac{n_1 + n_2}{k_1 + k_2} \tag{27}$$

If $r_1 + r_2 < d$, that means the two classes don't intersect, $se_{12} = 0$.

Step1: Let $d = 1$, and the separable measure se_{ij} between classes was obtained from the class samples, $i, j = 1, 2, \ldots, l, i \neq j$, and the separable measure matrix **SE** was constructed.

$$\mathbf{SE} = \begin{bmatrix} 0 & se_{12} & \cdots & se_{1,l-1} & se_{1,l} \\ se_{21} & 0 & \cdots & se_{2,l-1} & se_{2,l} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ se_{l-1,1} & se_{l-1,2} & \cdots & 0 & se_{l-1,c} \\ se_{l,1} & se_{l,2} & \cdots & se_{l,l-1} & 0 \end{bmatrix} \tag{28}$$

Setp2: according to the sum of each row, find the trip and the smallest row, and record the number of rows at this time $L(d) = m + d - 1$, then delete the elements of that row and column.

Step3: $d = d + 1$, and repeat step2 to guide the sorting of inter-class relationships of all categories.

Step4: Initialize $d = 1$, and take the sample set of class $d = 1$ as the positive sample of subclassifier, and the remaining sample of classes as the negative sample of the classifier. Train the SVM classifier and record the node information $node(d)$. Remove the class $L(d)$ sample set from the sample set.

Step5: $d = d + 1$, and repeat step4 until all sub-classifiers are trained. The final DT-SVM multi-classifier is shown in Figure 8.
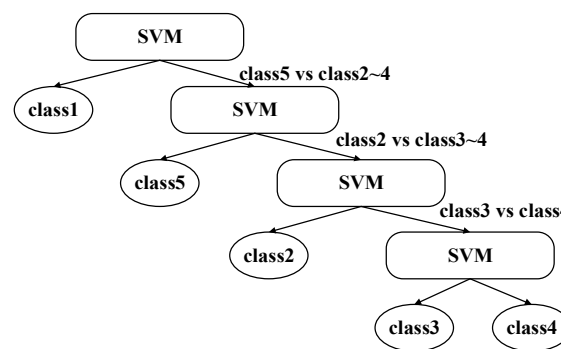

Figure 8. Hierarchical structure of decision tree support vector machine (DT-SVM) multi-classifier.

The flow chart of the whole paper is addressed in Figure 9. Firstly, data collection and data preprocessing were carried out, and then the best feature subset was obtained based on the BS-mRMRMC algorithm. Finally, the classification hierarchy of multiple classifiers was determined according to the vector projection method, and the performance of the classifier was evaluated with four indexes.
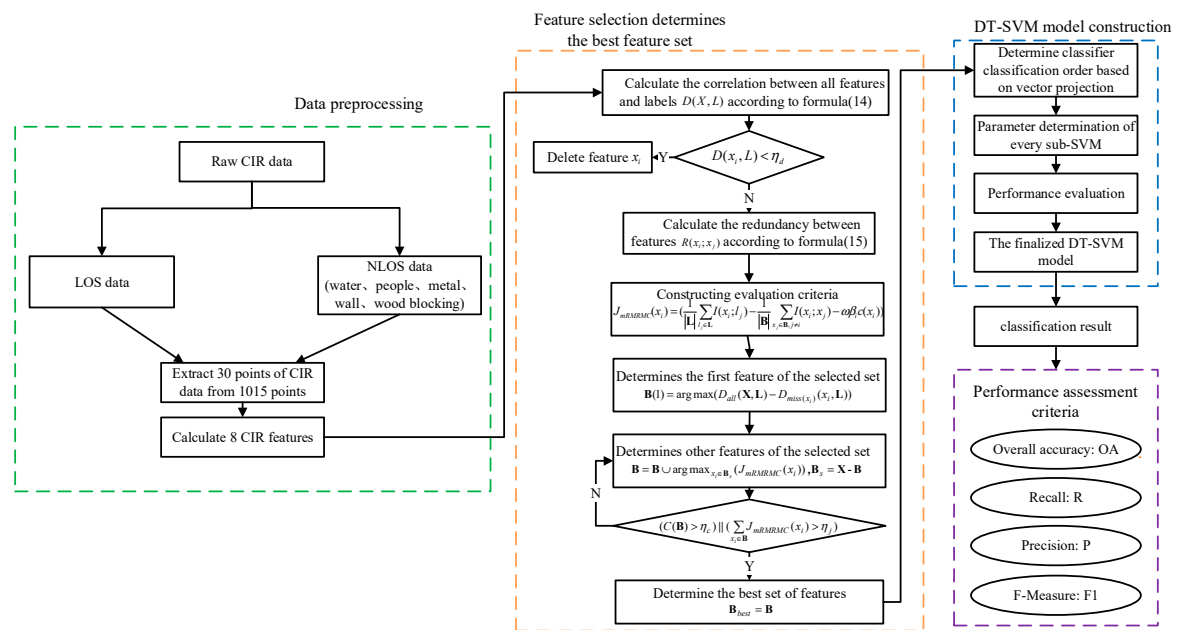
Figure 9. General diagram of algorithm flow.

4. Results and Discussion

In this section, the effectiveness of the proposed method is verified by the selection effect of features in the best subset and the evaluation indexes of recognition results. As shown in Figure 10, three feature selection algorithms IG [20], ReliefF [27] and SVM-RFE [29] are selected to compare the results with the BS-mRMRMC algorithm proposed in this paper.
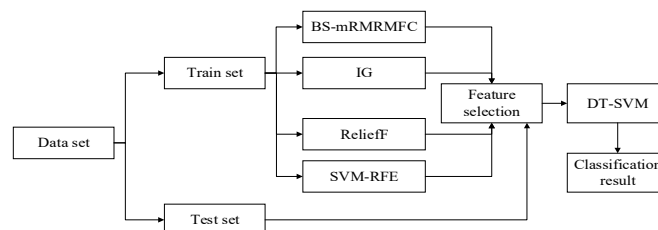

Figure 10. Experimental framework.

In this paper, four system performance evaluation indexes are selected to evaluate the results of each index under different feature subsets selected by four methods. The meanings of several parameters of performance evaluation are shown in Table 1.

Table 1. Definition table of performance evaluation parameters.

		Predicted Class		
		Yes	No	Total
Actual class	Yes	*TP*	*FN*	*P*
	No	*FP*	*TN*	*N*
	Total	*P′*	*N′*	*P* + *N* = *P′* + *N′*

Four evaluation indexes are overall accuracy, precision, recall rate and the comprehensive evaluation index. The overall accuracy can evaluate the positive and negative samples detection accuracy, precision can evaluate the misjudgment of positive samples of the system, and the recall rate

can evaluate the missed sanction of positive samples of the system, and the comprehensive evaluation index can comprehensively evaluate the identification performance of the system.

(1)   Overall accuracy:

$$OA = \frac{TP + TN}{TP + FN + TN + FP} \tag{29}$$

(2)   Precision:

$$P = \frac{TP}{TP + FP} \tag{30}$$

(3)   Recall:

$$R = \frac{TP}{TP + FN} \tag{31}$$

(4)   Comprehensive evaluation index F-Score:

$$F = \frac{(\alpha^2 + 1)P * R}{\alpha^2 (P + R)} \tag{32}$$

Usually take $\alpha = 1$, $F1 = \frac{2*P*R}{P+R}$.

The feature ordering of the four methods adopted in this paper is shown in Table 2. It can be analyzed from the feature selection in the table that the feature with the greatest relevance with the label is not necessarily the feature that contributes the most to the overall information, which also verifies the effectiveness of the paper's reverse deletion search strategy used to determine the first feature.

Table 2. Feature selection ranking of four algorithms.

Algorithms	Feature Selection Sorting
BS-mRMRMC	$t_{rise} > \tau_{RMS} > E_f > \tau_m > \sigma > \kappa > h_{max} > \gamma$
IG	$E_f > t_{rise} > \tau_m > \tau_{RMS} > \sigma > h_{max} > \gamma > \kappa$
ReliefF	$\tau_{RMS} > E_f > \tau_m > \sigma > t_{rise} > \gamma > \kappa > h_{max}$
SVM-RFE:	$t_{rise} > E_f > \tau_m > \gamma > \kappa > \tau_{RMS} > h_{max} > \sigma$

4.1. Best Subset Selection Based on Different Constraints

First, the paper verifies the number of features in the optimal feature subset determined by each method under different constraints of computational cost and maximum detection accuracy. Different η_c and η_j are obtained by setting different threshold parameters of calculation cost and maximum evaluation index $\lambda_c = [0.1, 0.2, \ldots, 1]$, $\lambda_j = [0.5, 0.55, \ldots, 0.95]$. From Figure 11, the number of features in the best feature subset selected by the BS-mRMRMC method is less than that of other methods under higher detection accuracy requirements and lower computational cost requirements. That is to say, the method in this paper can achieve the same or higher detection accuracy with fewer features. For example, when the constraints are $\eta_c \leq 0.5C(\mathbf{X})$ and $\eta_j \geq 0.9$, the number of best sub-sets selected by the proposed algorithm is only 2, when the constraints are changed by $\eta_c \leq 0.5C(\mathbf{X})$ and $\eta_j \geq 0.95$, the number is only 3. However, the number of features selected by other algorithms under the same constraint conditions are more than the algorithm of this paper. It can be concluded that the number of features determined by the best subset of the proposed algorithm is better than that of the other three algorithms when the computational cost is lower and the demand for higher precision is higher.
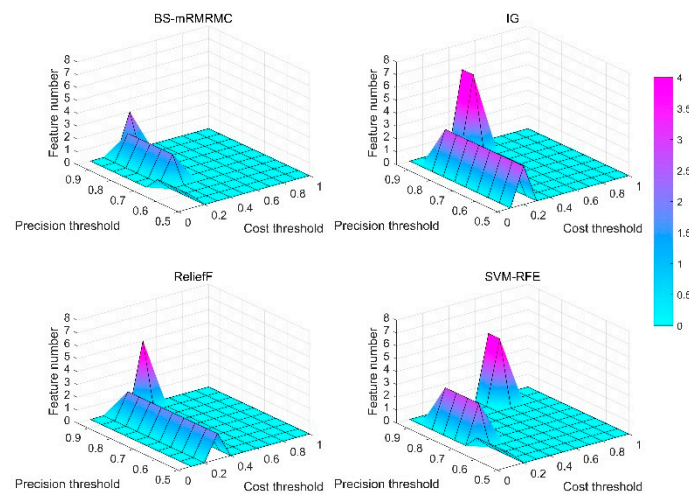
Figure 11. Surface graphs of the number of feature of best feature subset for the four methods under different computational cost constraints η_c and maximum evaluation index constraints η_j.

4.2. The Results Were Compared and Analyzed with Different Feature Number

In this section, the paper studies the comparison of the recognition results of different categories at first according to the sorting results of features in Table 2. The four pictures of Figure 12 show the comparison surface diagram of four evaluation indexes for identifying each category by four methods, when one to eight different features are selected successively. Figure 12a shows that our algorithm can achieve higher detection accuracy with fewer features. When the number of features selected is less than 4, the identification accuracy of each category is significantly better than IG and SVM-RFE, and slightly better than ReliefF. The Figure 12b–d show that our method has better recognition accuracy, recall rate and comprehensive evaluation index for most categories than the other three algorithms when the number of selected features is less than 4. This result is also consistent with that in the previous section, under the constraint of $\eta_c \leq 0.5C(\mathbf{X})$ and $\eta_j \geq 0.95$, the algorithm in this paper selects only 3 optimal subsets, which can achieve higher detection accuracy at a lower computational cost with a smaller number of features.

Figure 13 shows average accuracy rate, average precision, average recall rate and average comprehensive evaluation index contrast line chart of four methods when selecting a different number of features. If the number of features selected is less than 4, the indicators of the BS-mRMRMC algorithm are better than other algorithms. When there are more features, the selection of feature sets of different algorithms for the final indicators tend to be consistent. Due to the redundancy between features not considered by the IG algorithm, the detection accuracy decreases when the third feature is added in Figure 13, which also verifies the importance of considering the redundancy between features. In addition, the first feature selected can be regarded as the feature with the greatest contribution to overall performance. The accuracy of the recognition under the first feature is better than other algorithms, which also verifies the effectiveness of the bidirectional search strategy. However, the average recognition accuracy of the first feature selected by the BS-mRMRMC method is high, but the average recall rate is poor, indicating that this feature has a great correlation with other categories, which can be verified from Figure 5. As a result, many positive samples are missed and identified as negative samples, and there is a certain contradiction between the recall rate and accuracy. Therefore, it can be seen from the average comprehensive evaluation index F1 obtained from the comprehensive accuracy and recall rate that the first feature selected by BS-mRMRMC is better.
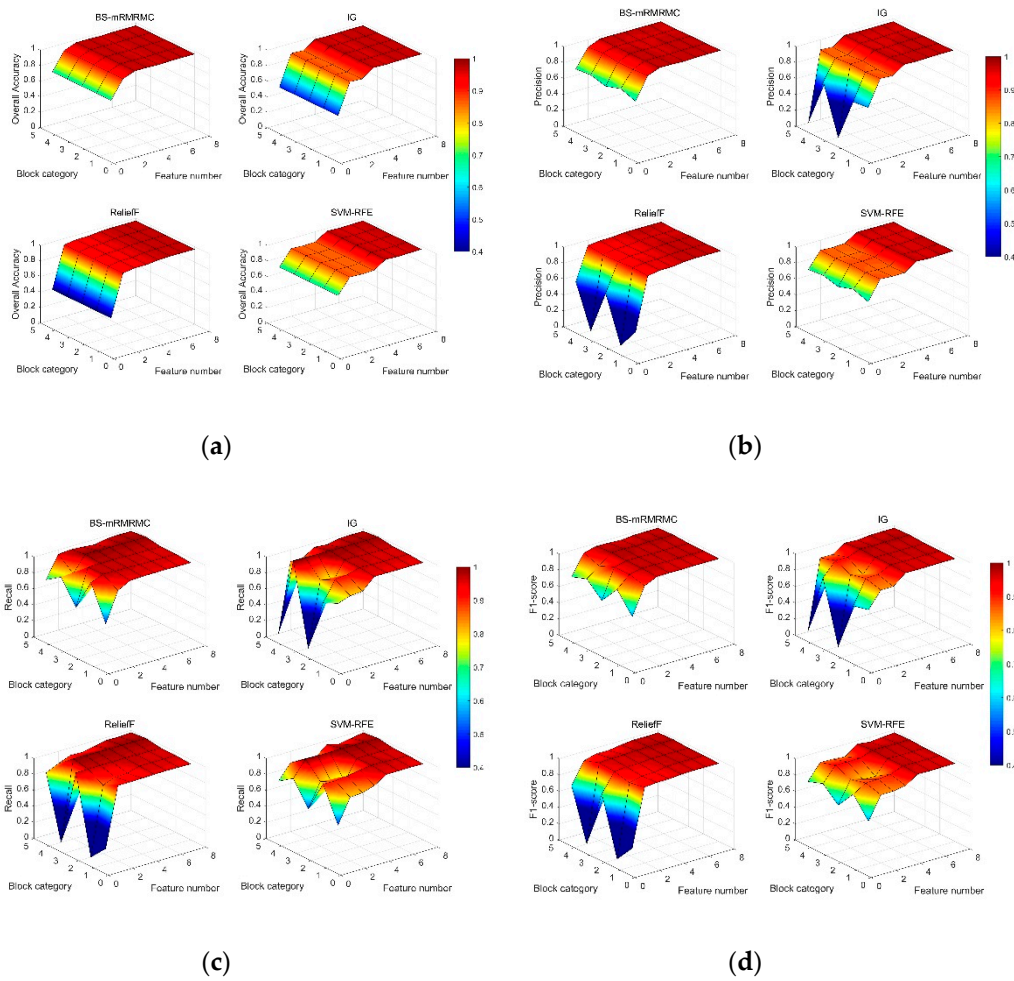
(**a**)　　　　　　　　　　　　　　　　　(**b**)

(**c**)　　　　　　　　　　　　　　　　　(**d**)

Figure 12. Under different number of features, the evaluate indicator surface graph of the four methods identify each category. (**a**) The overall accuracy, (**b**) the accuracy, (**c**) the recall rate, and (**d**) the comprehensive evaluation indicator f1-measure; where block category 0, . . . ,4 respectively represents no blocking, water blocking, people blocking, metal blocking and wall blocking).
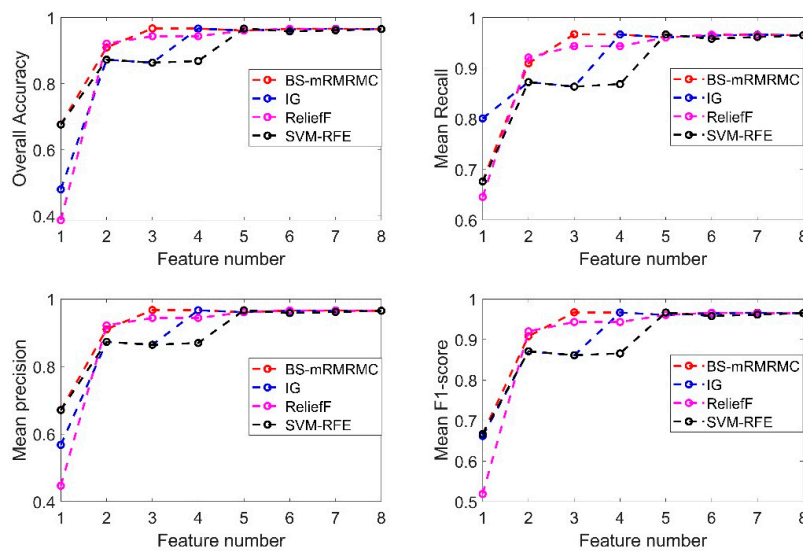

Figure 13. Comparison diagram of average indexes of each category identified by four methods under different feature quantities.

The purpose of feature selection is to obtain the best and smallest feature set, and the system performance meets the constraint requirements. Compared with the ReliefF method, our algorithm can comprehensively consider the correlation between features and categories and the redundancy between features. The limitation of ReliefF is that it cannot effectively remove redundant features. It can also be seen from Figure 13 that our proposed method selects three features, but ReliefF needs to select five features to achieve the same index requirements. Under the same constraints, our algorithm can select the feature set with the largest correlation, the smallest redundancy, and a smaller number of features. When the number of features in the selected best feature subset is 3, comparison of identification indicators of each category under the four methods is shown in Table 3. Except for the fact that the recognition index of some categories is slightly lower than other algorithms, the recognition index of BS-mRMRMC algorithm of most categories is better than other algorithms, and the average index of each category is more than 95%.

Table 3. The identification indexes of various categories are compared under the four methods.

Method	Precision				Recall				F1-Score				Overall Accuracy			
	BS-mRMRMC	IG	ReliefF	SVM-RFE	BS-mRMRMC	IG	ReliefF	SVM-RFE	BS-mRMRMC	IG	ReliefF	SVM-RFE	BS-mRMRMC	IG	ReliefF	SVM-RFE
No block	97.28%	86.85%	94.41%	86.85%	94.00%	80.50%	94.00%	80.50%	95.61%	83.55%	94.21%	83.55%	96.70%	86.35%	94.35%	86.35%
Water block	96.69%	85.90%	94.98%	85.90%	96.75%	92.50%	91.00%	92.50%	96.72%	89.08%	92.95%	89.08%	96.70%	86.35%	94.35%	86.35%
People block	96.59%	87.57%	93.97%	87.57%	97.25%	73.50%	96.50%	73.50%	96.92%	79.92%	95.22%	79.92%	96.70%	86.35%	94.35%	86.35%
Metal block	96.14%	85.45%	93.48%	85.45%	99.50%	99.50%	99.50%	99.50%	97.79%	91.94%	96.40%	91.94%	96.70%	86.35%	94.35%	86.35%
Wall block	96.85%	86.40%	95.03%	86.40%	96.00%	85.75%	90.75%	85.75%	96.42%	86.07%	92.84%	86.07%	96.70%	86.35%	94.35%	86.35%
Average	96.71%	86.43%	94.37%	86.43%	96.70%	86.35%	94.35%	86.35%	96.69%	86.11%	94.32%	86.11%	96.70%	86.35%	94.35%	86.35%

5. Conclusions

In this paper, a large amount of CIR data of a UWB device with five different blocking materials at LOS and NLOS are obtained in a large experimental center, and eight features are extracted from CIR to analyze the influence of different features on different occlusion categories. Selecting 30 effective CIR signal points for feature extraction can effectively improve the calculation efficiency of feature extraction. The proposed BS-mRMRMC method introduces the computational cost into the mRMRMC feature evaluation criteria, and uses a bidirectional search strategy for feature selection. Through multiple experiments, it has been proved that under the same constraints, BS-mRMRMC can better select the optimal feature subset with maximum correlation, minimum redundancy, minimum computational cost and fewer features. Meanwhile, when the same number of features are selected, the BS-mRMRMC method has better recognition accuracy for each category than the other three methods. When only three features are selected, the average accuracy of the recognition for each blocking category can reach 96.7%. More importantly, this paper not only identifies the LOS and NLOS categories, but also identifies the different blocking categories under NLOS, which provides a new research method for UWB subsequent indoor positioning, indoor information perception construction and target tracking.

Author Contributions: Under the supervision of Y.K., Y.K. proposed the method, established and verified the system model, and finished writing the first draft of the article. C.L. proposed and optimized methods, finished data processing and simulation, finished results verification and analysis, and modified article. Z.C. designed the UWB hardware equipment and software, and extracted the characteristics of the CIR signal. X.Z. unified the concept of the article and revised the article. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Nomenclature.

$\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k\}$:	Data set containing k samples	η_d:	Minimum relevance constraint threshold	Q:	Feature and feature mutual information under the same class	FN:	prediction is negative, reality is positive
$\mathbf{d}_i = \{\mathbf{X}_i, \mathbf{L}_i\}$:	Samples containing features and labels	η_j:	Maximum evaluation index constraint threshold	F:	Feature and label mutual information matrix	FP:	prediction is positive, reality is negative
$\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$:	Feature set containing n features	η_c:	Computational cost constraint threshold	$D(\mathbf{X}, \mathbf{Y})$:	Relevance calculation	FN:	prediction is negative, reality is positive
$\mathbf{L} = \{l^1, l^2, \ldots, l^m\}$:	Tag set containing m tags	λ_d:	Parameter of minimum relevance threshold	$R(x_i, x_j)$:	Redundancy calculation	OA:	Overall Accuracy
$\mathbf{B}$:	Selected feature set	λ_c:	Parameter of computing cost threshold	$J_{mRMRMC}(x_i)$:	mRMRMC characteristic evaluation index	P:	Precision
$\mathbf{B}s$:	Unselected feature set	$\mathbf{m}_l$:	Center point of the sample	$D_{all}(\mathbf{X}, \mathbf{L})$:	Relevance of all features and labels	R:	Recall
$C(\mathbf{B})$:	Feature computing time cost	se:	The separability measure between classes	$D_{miss}(x_i)$:	Relevance of all features and labels after missing x_i	F:	Comprehensive evaluation index F-Score
β_i:	Weight parameters	$\mathbf{SE}$:	the separable measure matrix	β_i:	relative weight of the calculate cost.	ω:	weight of calculation cost in the evaluation index
$I(\mathbf{X}; \mathbf{Y})$:	Mutual Information	FP:	prediction is positive, reality is negative				

References

1. Alsaleh, M.; Alnafessah, A.; Alhadhrami, S.A.; Al-Khalifa, H.; Al-Ammar, M.A. Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances. *Sensors* **2016**, *16*, 707.
2. Aditya, S.; Molisch, A.F.; Behairy, H.M. A Survey on the Impact of Multipath on Wideband Time-of-Arrival-Based Localization. *Proc. IEEE* **2018**, *99*, 1–21. [CrossRef]
3. Wen, K.; Yu, K.; Li, Y. NLOS identification and compensation for UWB ranging based on blocking classification. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 20 August–2 September 2017; pp. 2704–2708.
4. Wang, F.; Xu, Z.; Zhi, R.; Chen, J.; Zhang, P. LOS/NLOS Channel Identification Technology Based on CNN. In Proceedings of the 6th NAFOSTED Conference on Information and Computer Science, Hanoi, Vietnam, 12–13 December 2019; pp. 200–203.
5. Dardari, D.; Conti, A.; Ferner, U.; Giorgetti, A.; Win, M.Z. Ranging with Ultrawide Bandwidth Signals in Multipath Environments. *Proc. IEEE* **2009**, *97*, 404–426. [CrossRef]
6. Decarli, N.; Decarli, D.; Gezici, S.; D'Amico, A.A. LOS/NLOS Detection for UWB Signals: A Comparative Study Using Experimental Data. In Proceedings of the IEEE 5th International Symposium on Wireless Pervasive Computing 2010, Modena, Italy, 5–7 May 2010; pp. 169–173.
7. Almazrouei, E.; Sindi, N.A.; Al-Araji, S.R.; Ali, N.; Aweya, J. Measurement and analysis of NLOS identification metrics for WLAN systems. In Proceedings of the IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), Washington, DC, USA, 2–5 September 2014; pp. 280–284.
8. Wu, S.; Xu, D.; Wang, H. Adaptive NLOS Mitigation Location Algorithm in Wireless Cellular Network. *Wirel. Pers. Commun.* **2015**, *84*, 3143–3156. [CrossRef]
9. Gururaj, K.; Rajendra, A.K.; Song, Y.; Law, C.L.; Cai, G. Real-time identification of NLOS range measurements for enhanced UWB localization. In Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sapporo, Japan, 18–21 September 2017; pp. 1–7.
10. Marano, S.; Gifford, W.M.; Wymeersch, H.; Win, M.Z. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE J. Sel. Areas Commun.* **2010**, *28*, 1026–1035. [CrossRef]
11. Guvenc, I.; Chong, C.; Watanabe, F. NLOS Identification and Mitigation for UWB Localization Systems. In Proceedings of the 2007 IEEE Wireless Communications and Networking Conference, Kowloon, China, 11–15 May 2007; pp. 1571–1576.
12. Guvenc, I.; Chong, C.C.; Watanabe, F.; Inamura, H. NLOS Identification and Weighted Least-Squares Localization for UWB Systems Using Multipath Channel Statistics. *EURASIP J. Adv. Signal Process.* **2008**, 1–14.
13. Zhang, J.; Salmi, J.; Lohan, E.-S. Analysis of kurtosis-based LOS/NLOS identification using indoor MIMO channel measurement. *IEEE Trans. Veh. Technol.* **2013**, *62*, 2871–2874. [CrossRef]
14. Landolsi, M.A.; Almutairi, A.F.; Kourah, M.A. LOS/NLOS channel identification for improved localization in wireless ultra-wideband networks. *Telecommun. Syst.* **2019**, *72*, 441–456. [CrossRef]
15. Schroeer, G.; Haefner, B. Predictive NLOS Detection for UWB Indoor Positioning Systems Based on the CIR. In Proceedings of the 15th Workshop on Positioning, Navigation and Communications (WPNC), Bremen, Germany, 25–26 October 2018; pp. 1–5.
16. Zeng, Z.; Liu, S.; Wang, L. NLOS Identification for UWB Based on Channel Impulse Response. In Proceedings of the 12th International Conference on Signal Processing and Communication Systems, Cairns, Australia, 17–19 December 2018; pp. 1–6.
17. Zeng, Z.; Liu, S.; Wang, L. UWB/IMU integration approach with NLOS identification and mitigation. In Proceedings of the 52nd Annual Conference on Information Sciences and Systems, Princeton, NJ, USA, 21–23 March 2018; pp. 1–6.
18. Zeng, Z.; Liu, S.; Wang, L. NLOS Detection and Mitigation for UWB/IMU Fusion System Based on EKF and CIR. In Proceedings of the IEEE 18th International Conference on Communication Technology (ICCT), Chongqing, China, 8–11 October 2018; pp. 376–381.
19. Zhenhai, Z.; Shining, L.; Zhigang, L.; Hao, C. Multi-Label Feature Selection Algorithm Based on Information. *J. Comput. Res. Dev.* **2013**, *50*, 1177–1184.
20. Liu, H.; Sun, J.; Liu, L.; Zhang, H. Feature selection with dynamic mutual information. *Pattern Recognit.* **2009**, *42*, 1330–1339. [CrossRef]

21. Zhang, Y.; Yang, A.; Xiong, C.; Wang, T.; Zhang, Z. Feature Selection Using Data Envelopment Analysis. *Knowl. Based Syst.* **2014**, *64*, 70–80. [CrossRef]

22. Kwak, N.; Choi, C.-H. Input feature selection for classification problems. *IEEE Trans. Neural Netw.* **2002**, *13*, 143–159. [CrossRef] [PubMed]

23. Muhammad, A.S.; Labadin, J. Feature Selection based on Mutual Information. In Proceedings of the International Conference on It in Asia (CITA), Kota Samarahan, Malaysia, 4–5 August 2015; pp. 1–6.

24. Zhang, L.; Wang, C. Multi-label Feature Selection Algorithm Based on Maximum relevance and Minimal Redundant Joint Mutual Information. *J. Commun.* **2018**, *39*, 115–126.

25. Doquire, G.; Verleysen, M. Mutual information-based feature selection for multilabel classification. *Neurocomputing* **2013**, *122*, 148–155. [CrossRef]

26. Wang, Y.; Feng, L. A new hybrid feature selection based on multi-filter weights and multi-feature weights. *Appl. Intell.* **2019**, *6*, 4033–4057. [CrossRef]

27. Zeng, Z.; Liu, S.; Wang, L. UWB NLOS identification with feature combination selection based on genetic algorithm. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics, Las Vegas, NV, USA, 11–13 January 2019; pp. 1–5.

28. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]

29. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [CrossRef]

30. Sain, S.R. The Nature of Statistical Learning Theory. *Technometrics* **1997**, *38*, 409. [CrossRef]

31. Bottou, L.; Cortes, C.; Denker, J.S.; Drucker, H.; Guyon, I.; Jackel, L.D.; LeCun, Y.; Muller, U.A.; Sackinger, E.; Simard, E.; et al. Comparison of classifier methods: A case study in handwritten digit recognition. In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994; pp. 77–82.

32. Bernhard, S. *Advances in Kernel Methods—Support Vector Learning*; MIT Press: Cambridge, MS, USA, 1999.

33. Platt, J.C.; Cristianini, N.; Shawe-Taylor, J. Large Margin DAGs for Multiclass Classification. *Adv. Neural Inf. Process. Syst.* **2000**, *12*, 547–553.

34. Wang, X.; Shi, Z.; Wu, C.; Wang, W. An Improved Algorithm for Decision-Tree-Based SVM. In Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006; pp. 4234–4238.

35. Wang, D.M.; Lu, C.H.; Jiang, W.W.; Xiao, M.X.; Li, B. Study on PSO-based decision-tree SVM multi-class classification method. *J. Electron. Meas. Instrum.* **2015**, *4*, 611–615. [CrossRef]

36. Sha, H.; Zhang, C.; Shi, M.; Zheng, J. Voltage Sag Classification Based on PDT-SVM. In Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), Kyoto, Japan, 27 February–2 March 2019; pp. 1–6.

37. Li, X.R.; Zhao, G.Z.; Zhao, L.Y. Design of decision-tree-based support vector machines multi-class classifier based on vector projection. *Control Decis.* **2008**, *7*, 27–32. [CrossRef]

38. Kim, H.H.; Choi, J.Y. Hierarchical multi-class LAD based on OvA-binary tree using genetic algorithm. *Expert Syst. Appl.* **2015**, *42*, 8134–8145. [CrossRef]

39. Song, B.; Li, S.L.; Tan, M.; Ren, Q.-H. A Fast Imbalanced Binary Classification Approach to NLOS Identification in UWB Positioning. *Math. Probl. Eng.* **2018**, *2018*, 1–8. [CrossRef]

40. Zhang, L.; Suganthan, P.N. Oblique Decision Tree Ensemble via Multisurface Proximal Support Vector Machine. *IEEE Trans. Cybern.* **2015**, *45*, 2165–2176. [CrossRef] [PubMed]

41. Li, Q.; Jiao, L.C.; Zhou, W.D. Pre-extracting Support Vector for Support Vector Machine Based on Vector Projection. *Chin. J. Comput.* **2005**, *28*, 154–196.