

Dynamic Optimization Data Association based on JCBB Algorithm

Huiheng Suo^{1,3,a}, Bosi Wei^{1, b*}, Qiang Hu^{1,c}, Jiingjia Pei^{3,d}, Xie Ma^{2,e}, Yujie Song^{2,f}, Bibo Yu^{4,g},
Xiushui Ma^{3,h}

¹Nanchang Hangkong University, Nanchang, China

²Ningbo University of Finance & Economics, Ningbo, China

³NingboTech University, Ningbo, China

⁴Huayuan New Materials Co., Ltd, Ningbo, China

^asuohuiheng@163.com, ^bwei_bosi@163.com, ^c1289762339@qq.com, ^d17815927609@163.com,

^emaxie@163.com, ^f1819377256@qq.com, ^g583771406@qq.com, ^hmxsh63@aliyun.com

*corresponding author

Keywords: JCBB, DOJCBB, Data Association, SLAM

Abstract: This paper presents the principles of data association algorithms and describes them using mathematical language. It then provides a detailed analysis of the application of the Joint Compatible Branch and Bound (JCBB) algorithm in the data association stage, proposing the DOJCBB data association algorithm. Firstly, to avoid excessive computational resource waste caused by numerous environmental features in complex environments, the data association is restricted to a localized association region. Secondly, to address the error accumulation resulting from uncertainties during robot operation, a threshold constant dynamic adaptation is introduced. Lastly, considering the issues of multiple hypotheses with the maximum number of associations and false associations in the data association process of the JCBB algorithm, corresponding optimization criteria are designed for improvement. The effectiveness of the proposed improved DOJCBB data association algorithm is validated through comparative experiments on a simulation platform.

1. Introduction

Data association initially originated from the field of target tracking, which aims to determine the source of each acquired data. In the context of SLAM, data association involves matching sensor observations with the constructed map of environmental features to determine if the observed data

corresponds to existing map features [1, 2]. Currently, there are three main research approaches to data association algorithms: probability-based methods, fuzzy methods, and optimization methods.

In 1971, Singer et al. [3] introduced the Nearest Neighbor (NN) algorithm, which selects the closest feature within a threshold range as the matching result. This method is only applicable in simple environments and fails in complex environments due to the occurrence of repetitive associations [4], leading to significant errors in SLAM.

In 1974, Bar-Shalom [5, 6] proposed the Probabilistic Data Association (PDA) algorithm. This method assumes independence between each feature and computes the prior information of the features through weighted summation. However, due to the neglect of inter-feature correlations, the performance of the PDA algorithm deteriorates in dense feature scenarios. To address this issue, Bar-Shalom incorporated feature correlations [7] and improved the PDA algorithm, introducing the Joint Probabilistic Data Association (JPDA) algorithm.

In 2001, Neira et al. [8] proposed the Joint Compatibility Branch and Bound (JCBB) data association algorithm, which addresses the problem of false associations that increase with the number of associations in the hypothesis during independent compatibility tests. Compared to the Nearest Neighbor data association algorithm, the JCBB algorithm yields more robust association hypotheses in complex environments. However, it shares the drawback of high computational complexity with the JPDA method, which hinders real-time implementation. In 2009, Zhou Wu, Zhao Chunxia, and others [9] optimized the JCBB algorithm to reduce computational complexity and improve matching accuracy. In 2015, Yao Cong [10] proposed a JCBB data association algorithm based on K-means clustering. It reduces the number of involved features by clustering and grouping the observations, thereby reducing computational complexity. In 2018, Wang et al. [11] introduced a SLAM data association algorithm for multiple hypothesis target tracking. They improved environment exploration and mapping accuracy by sorting and eliminating association hypotheses.

In 2009, Zeng Wenjing et al. [12] transformed the data association model into a combinatorial optimization problem and solved it using ant colony optimization. In 2011, Du Hangyuan et al. [13] proposed a fuzzy logic-based data association algorithm by establishing fuzzy rules. This method can parallelly process multiple association hypotheses, enhancing the robustness of the data association algorithm. In 2017, Li Yanju et al. [14] designed a data association algorithm that combines independent compatible nearest neighbor and JCBB algorithms for use in complex environments. This approach effectively improves association accuracy in complex environments.

2. Description and Principles of Data Association Problem

The environment perception of a robot is an incremental process where environmental landmarks need to be continuously incorporated into the constructed map as the robot explores the environment. Data association, in essence, involves matching sensor observations with the existing map environmental features to determine whether the observed data corresponds to the same environmental feature present in the map.

2.1 Description of the Data Association Problem

Data association refers to the matching of observation data with environmental features. Assuming that the state vector of the robot at different time instances is denoted as $x_k (k \in 0, 1, L, T)$, the collection of state vectors $X = \{x_k, k \in 0, 1, L, T\}$ represents the robot's trajectory. The robot's control inputs at time instance k are represented as $u_k (k \in 0, 1, L, T)$, and

$U = \{u_k, k \in 0, 1, L, T\}$ represents the set of control inputs. The robot's observations at time instance k are represented as $Z_k(z_i, i \in 1, 2, L, I)$, where a single observation at time instance k is denoted as $z_i (i \in 1, 2, L, I)$, and the collection of all robot observations is represented as $Z = \{Z_k, k \in 0, 1, L, T\}$. The environmental landmark features are denoted as $m_j (j \in 1, 2, L, J)$, and the set of landmark features is represented as $M \{m_j, j \in 1, 2, L, J\}$. The data association variable between an environmental landmark feature m_j and a single observation z_i is represented as $c_n (n \in 1, 2, L, N)$, and the data association variable of the robot's observation Z_k at time instance k is denoted as $C_k = \{c_n, n \in 1, 2, L, N\}$. The data association variable for all robot observations Z is represented as $C = \{C_k, k \in 0, 1, L, T\}$.

Based on the above assumptions, the data association problem can be formulated as an optimal estimation problem, as shown in Equation (1):

$$(X^*, M^*, C^*) = \arg \max_{X, M, C} (P(X, M | Z, U, C)) \quad (1)$$

By using Equation (1), it is possible to calculate the likelihood function $P(X, M | Z, U, C)$ for all possible data association variables in the solution space of C to find the optimal solution. However, due to the correlation between the data association variable C and the observation information z_i , as well as the accumulation of the number of observations and landmark features m_j as the robot moves, the solution space of the data association variable C becomes infinitely large. This ultimately leads to increased computational complexity and dimensionality. Therefore, brute-force solutions are not appropriate. To address this problem, an approach based on maximum likelihood estimation [15] can be used to obtain the optimal estimate C^* of the data association variable C , as shown in Equation (2):

$$C^* = \arg \max_C (P(X, M | Z, U, C)) \quad (2)$$

The solution for the optimal estimation of the robot trajectory X and landmark features M is obtained as shown in Equation (3):

$$(X^*, M^*) = \arg \max_{X, M} (P(X, M | Z, U, C^*)) \quad (3)$$

In Equation (3), it represents the optimal estimation of the robot trajectory X and landmark features M . In this study, the focus of the data association research is on the real-time robot position and landmark positions, which can be approximated as solved in Equation (4):

$$(X_k^*, M^*) = \arg \max_{x_k, M} (P(x_k, M | Z, U, C^*)) \quad (4)$$

2.2. Principles of Data Association

From the previous section's description, it is evident that solving the data association problem involves determining the optimal estimate of the data association variable C . Generally, the data association problem can be divided into three parts: threshold filtering, computing association metrics, determining association criteria.

2.2.1 Threshold Filtering

This part is able to remove association hypotheses with low likelihood and filter out undesired values in the observations, such as clutter and noise. In robot research, errors are often assumed to follow a Gaussian distribution. Therefore, each observation data from the robot can potentially originate from all environmental landmark features. This leads to a significant amount of data involved in association matching. To effectively reduce the computational load of data association, it is necessary to filter out unnecessary observations and only consider those within a certain range. A window, referred to as an association gate, is set based on this range. The principle behind setting the gate is to allow the observation data to have a higher probability of falling within the window. The setting of the threshold directly affects the quality of data association, and an appropriate threshold value helps improve the accuracy of data association and reduce the computational load. Common association gates [16] include elliptical gates and rectangular gates. The following sections will introduce these two types of association gates.

(1) Rectangular association gate

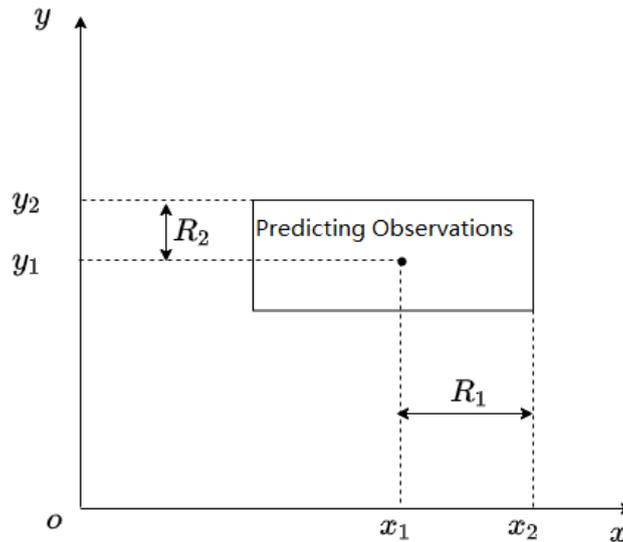


Figure 1. Rectangular association gates

The rectangular association gate, as shown in Figure 1, $R_1 = K_{G,1}\sigma_r$, $R_2 = K_{G,2}\sigma_r$. If the observed measurement z_k obtained by the robot and the predicted observation value \hat{z}_k of the existing landmark feature in the map satisfy Equation (5), then the association is considered valid.

$$z_{k,l} = z_{k,l} - \hat{z}_{k,l} \leq K_{G,l}\sigma_r \quad (5)$$

In Equation (5), $l \in M$, M represents the dimension of the association gate; $z_{k,l}$ represents the residual; σ_r represents the standard deviation of the residual; $K_{G,l}$ represents the threshold constant, which is determined by the detection probability, observation density, and dimensionality of the state vector. The value of σ_r is related to the observation error and the predicted covariance matrix in the Kalman filter, as follows:

$$\sigma_r = \sqrt{\sigma^2 + \sigma_p^2} \quad (6)$$

In Equation (6), σ represents the observation standard deviation; σ_p represents the predicted standard deviation of the Kalman filter. Here, we assume that the errors in \hat{z}_k and the Gaussian error model are independent of each other. Therefore, the probability of the observation falling within the threshold can be expressed as Equation (7):

$$P_G = [1 - P(|t_1| \geq K_{G,1})][1 - P(|t_2| \geq K_{G,2})] \dots [1 - P(|t_l| \geq K_{G,l})] \quad (7)$$

In Equation (7), $P(|t_l| \geq K_{G,l})$ represents the probability of a standard normal random variable exceeding the threshold $K_{G,l}$. In the data association research of this paper, all threshold sizes are the same, denoted as K_G . Therefore, Equation (7) can be simplified as:

$$P_G = [1 - P(|t| \geq K_G)]^M \approx 1 - MP(|t| \geq K_G) \quad (8)$$

We can determine the threshold value by looking up the standard normal distribution table based on the proportion of observations that fall within the threshold P_G .

(2) Elliptical association gates

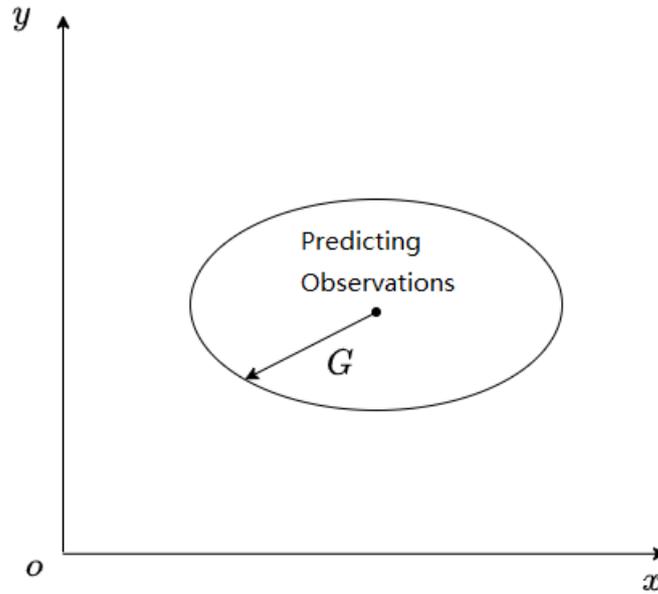


Figure 2. Elliptical gates

Figure 2 shows the elliptical association gate with G as the threshold constant. This association gate uses the norm of the residual vector to determine the correlation, as shown in equation (9). The threshold is set using the χ^2 distribution method.

$$d^2 = \left(z - \hat{z} \right)^T S^{-1} \left(z - \hat{z} \right) \leq G \quad (9)$$

Here, d^2 represents the Mahalanobis distance, S denotes the covariance matrix of observation

errors, and $\begin{pmatrix} z - \hat{z} \end{pmatrix}$ represents the innovation vector. If $\begin{pmatrix} z - \hat{z} \end{pmatrix}$ follows a Gaussian distribution, then d^2 follows a χ^2 distribution with $\dim\begin{pmatrix} z - \hat{z} \end{pmatrix}$ degrees of freedom. In the data association discussed in this paper, the observed data consists of distance and angle measurements between the robot and map features. Therefore, the Mahalanobis distance follows a Chi-square distribution with 2 degrees of freedom. By referring to the χ^2 distribution table, the threshold G for the test, as indicated by equation (9), can be determined. When this condition is met, it signifies that the observation falls within the association range of a particular landmark feature.

Table 1. Chi-square distribution table with 2 degrees of freedom

Confidence level	75%	90%	95%	97.5%	99%	99.5%	99.9%
Value	2.773	4.605	5.991	7.378	9.210	10.597	13.820

2.2.2 Computing Association Metrics

For the filtered observations after thresholding, it is necessary to calculate the similarity, namely the relative distance, between the observation and the existing landmark features. Typically, two distance metrics, Euclidean distance and Mahalanobis distance, are used as measurement methods for data association. The Euclidean distance only considers the relative distance between the observation and the landmark feature positions, while the Mahalanobis distance takes into account the correlation between variables as well.

In this study, the Mahalanobis distance is chosen as the measurement method between the observation and the landmark features. Let x, y be samples extracted from the population G , with mean μ and covariance V , the following equation holds:

$$D^2(x, y) = (x - y)^T V^{-1} (x - y) \quad (10)$$

Based on equation (10), it can be observed that when the covariance matrix representing the uncertainty between variables in the Mahalanobis distance is an identity matrix, the Mahalanobis distance reduces to the Euclidean distance. In SLAM algorithms, when the robot obtains the i -th observation at time k , the innovation $v_i(k)$ is obtained by subtracting the predicted observation value $z_i(\hat{k})$ at time k from the actual observation. At the same time, the predicted observation covariance matrix $S(k)$ is obtained. Therefore, the Mahalanobis distance can be represented as $d_k^2 = v_i(k)^T S(k)^{-1} v_i(k)$.

2.2.3 Determining Association Criteria

The data association criterion refers to selecting the observation that is closest to the existing environmental features based on this criterion. Then, this observation is associated and matched with the corresponding environmental feature.

2.3 Jointly Compatible Branch and Bound Algorithm

The JCBB algorithm combines the jointly compatible and branch and bound methods to achieve optimal data association. It is based on the premise that false data associations in the independent compatibility test increase with the number of associations in the hypothesis. The JCBB data association algorithm, under the condition of joint compatibility test, examines the correlation between the measurements and landmark features, as well as the correlation between the robot's state and the features [], and searches the explanation tree to find the non-empty hypothesis with the maximum number of joint compatible associations.

Given the association hypothesis $H_m = \{j_1, j_2, \dots, j_m\}$, the joint observation equation is derived as follows:

$$Z_{H_m} = h_{H_m}(x_k) + v_{H_m} \quad (11)$$

$$h_{H_m}(x_k) = \begin{bmatrix} h_{j_1}(x_k) \\ \mathbf{M} \\ h_{j_m}(x_k) \end{bmatrix} \quad (12)$$

Given the joint state vector I_{H_m} and the covariance matrix S_{H_m} of the observation Jacobian S_{H_m} , we can obtain the equation as follows:

$$\begin{aligned} I_{H_m} &= z_k - h_{H_m}(\hat{x}_{k,k-1}) \\ S_{H_m} &= H_{H_m} P_k H_{H_m}^T + G_{H_m} V G_{H_m}^T \end{aligned} \quad (13)$$

In equation (13), $z_k = [z_{k_1}, z_{k_2}, \dots, z_{k_m}]^T$ represents the joint observation vector of existing environmental landmarks. $h_{H_m}(\hat{x}_{k,k-1})$ represents the joint compatibility predicted observation of the environmental landmarks. V represents the variance of observation noise. H_{H_m} and G_{H_m} can be obtained using equation (14):

$$\begin{aligned} H_{H_m} &= \frac{\partial h_{H_m}}{\partial x} \Big|_{\hat{x}_{k,k-1}, h_{H_m}(\hat{x}_{k,k-1})} \\ G_{H_m} &= \frac{\partial h_{H_m}}{\partial z} \Big|_{\hat{x}_{k,k-1}, h_{H_m}(\hat{x}_{k,k-1})} \end{aligned} \quad (14)$$

The JCBB data association algorithm uses the criterion of finding the association hypothesis with the maximum non-empty joint compatibility pairings as a monotonically non-decreasing standard to traverse the explanation tree. In other words, the association hypothesis H_m with the maximum number of association pairings is sought. Then, equation (15) is used as a method to determine whether the association hypothesis H_m satisfies the joint compatibility criterion. Equation (15) is as follows:

$$D_{H_m} = I_{H_m}^T S_{H_m}^{-1} I_{H_m} \leq \chi^2_{d,1-\alpha} \quad (15)$$

In equation (15), D_{H_m} represents the Mahalanobis distance of the association hypothesis H_m ; $d = \dim(z_k)$. The chi-square distribution (χ^2) is a d-dimensional matrix that satisfies the desired confidence level $1-\alpha$. When equation (3-15) holds true, the association hypothesis is considered as the data association result.

3. Design of the DOJCBB Data Association Algorithm

3.1. Local Feature Association Strategy

When a robot operates in a complex environment with multiple features, the sensors collect a large amount of observation data, while the number of environmental landmarks continues to increase over time. Traditional JCBB algorithm matches all stored environmental landmarks with the observation data, resulting in significant computational resources being consumed for data association, which severely affects the overall real-time performance of the system. The Local Feature Association strategy refers to performing data association only within the effective range of the robot's sensors. By adopting this strategy, the computational load of the data association algorithm remains stable and the number of map features involved in data association at any given time is reduced. This reduces the impact on the overall real-time performance of the robot. The selection of the local region is determined by the sensor's observation range and the robot's current pose at the moment.

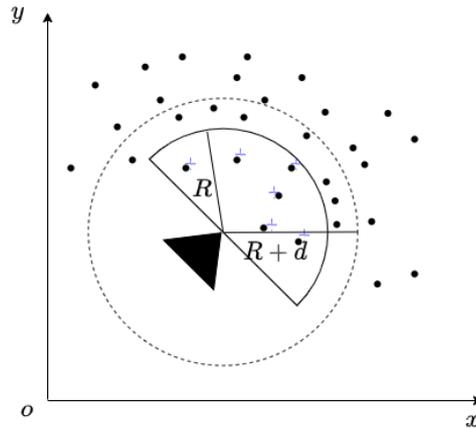


Figure 3. Illustration of Local Region

As shown in Figure 3, the black triangle represents the robot, the solid semicircle represents the effective sensing distance of the sensor, and the dashed circle represents the local association region with the added compensation distance. The compensation distance is introduced to prevent the omission of observed features corresponding to the true map. The black dots represent the existing environmental landmarks, and the asterisks represent the observed data. This helps to reduce the number of environmental landmark features involved in data association at the same time, thus reducing computational complexity. The local association region is determined by Equation (16):

$$\sqrt{(x_v - x_i)^2 + (y_v - y_i)^2} \leq R + d \quad (16)$$

In Equation (16), (x_v, y_v) represents the current pose of the robot at the moment; (x_i, y_i) represents the position of the i th environmental landmark feature point; R represents the

effective distance of the observation sensor; d represents the compensation distance.

3.2. Adaptive Threshold Value

The robot's sensors have errors that accumulate over time as the robot operates in the environment, leading to dynamic changes in the robot's uncertainty. The data association step is also affected by the overall system uncertainty. If the data association is performed using a fixed association threshold constant $d_{threshold}$, it can deviate from reality and increase the errors in robot localization and mapping.

To address this issue, this section improves the association threshold constant by dynamically adjusting it based on the robot's estimated uncertainty. The approach involves using the system error covariance matrix to represent the uncertainty of the robot's feature estimation. Then, the confidence level of the chi-square distribution is determined based on the error covariance matrix at different time steps, which in turn determines the association threshold constant. This allows the threshold constant to dynamically adapt.

In the EKF-SLAM algorithm, $P(k)$ represents the uncertainty of the robot's state. Since the robot's pose is represented as (x, y, θ) , the uncertainty of the robot's pose state can be quantified using the first three dimensions of the covariance matrix $P(k)$, as shown in Equation (17):

$$P_v(k) = P(k)_{11} + P(k)_{22} + P(k)_{33} \quad (17)$$

Where $P_v(k)$ represents the uncertainty of the robot's pose state, which is numerically equal to the trace (the sum of the diagonal elements) of the covariance matrix of the pose state.

The environment landmark features are represented as $(r, \beta)^T$. Therefore, the uncertainty of the estimation for the i -th environment landmark feature can be quantified using the $P(k)$, $(2i+2)$ and $(2i+3)$ elements on the diagonal of the covariance matrix, as shown in the following equation:

$$P_i(k) = P(k)_{(2i+2)(2i+2)} + P(k)_{(2i+3)(2i+3)} \quad (18)$$

Where, $P_i(k)$ represents the uncertainty estimation of the environment landmark feature.

When using the JCBB algorithm for data association, the choice of confidence level can have different effects on the robot's pose estimation error. In the initial stage of robot motion, where the errors have not yet accumulated, the robot's state uncertainty is low, and the confidence level has little impact on pose estimation. However, as the robot moves over time, the estimation errors of the environment landmark features accumulate due to system uncertainty, ultimately affecting the accuracy of the data association algorithm. The confidence level is directly proportional to the threshold constant used for data association. When the robot's state uncertainty increases, the threshold constant becomes larger, leading to a higher possibility of misjudgment in data association, such as considering an existing environment landmark feature as a new one, which affects the robot's localization and mapping. To address the issue of misjudgment in data association and its impact on localization and mapping, an improvement can be made as follows: in the initial stage of robot operation when the system uncertainty is minimal, a larger threshold value can be used to associate more information. As the robot operates for a period of time and the uncertainty increases, the possibility of misjudgment also increases. At this point, it is necessary to appropriately reduce the confidence level to lower the association threshold. Formula (19) represents a method for

dynamically calculating the confidence level based on system changes:

$$(1 - \sigma)^* = 99.5\% \times \frac{P_0}{P_i} \quad (19)$$

Where, P_0 represents the initial uncertainty of the robot system state, which is the initial value of the state error covariance. P_i is the quantified value of the uncertainty of the environment landmark feature estimation obtained from Formula (18). By using the confidence level calculated based on Formula (19) and referring to the chi-square distribution table shown in Table 1, the threshold constant for data association can be dynamically and adaptively selected.

3.3. Optimization of Misassociations

Although the presence of joint compatibility test in the JCBB data association algorithm reduces false associations, there still exists a certain degree of mismatched data. For instance, when multiple association hypotheses with the same and maximum number of joint compatible pairs exist, the JCBB data association algorithm selects the first discovered association hypothesis as the optimal association, even though the first search result may not necessarily be the best. Furthermore, there should be uniqueness between observations and environment landmark features, meaning that an observation can only come from a single landmark feature and cannot be associated with multiple features. However, in the data association solution, there are cases where an already associated landmark feature may participate in association again. To address these two types of mismatching in the JCBB data association, this section proposes two optimization methods: joint compatibility cost minimization and mutual exclusion optimization. Joint compatibility cost minimization refers to the situation where multiple association hypotheses with the maximum number of associations exist in the JCBB data association algorithm. By comparing the joint Mahalanobis distances of these n association hypotheses, the one with the minimum joint Mahalanobis distance is selected as the optimal association hypothesis. The definition of joint compatibility cost minimization is provided below:

$$\begin{cases} D = (D_{H_1}, D_{H_2}, \dots, D_{H_n}) \\ D_{H_i} = \min(D), i = 1, 2, \dots, n \\ H = H_i \end{cases} \quad (20)$$

Where, D represents the set of joint Mahalanobis distances for these n association hypotheses. D_{H_i} represents the minimum joint Mahalanobis distance, and H corresponds to the optimal hypothesis among these association hypotheses.

Mutual exclusion optimization can be understood as the rejection of matching when another observation attempts to associate with an environment landmark feature that has already been associated with a measurement.

3.4. DOJCBB

Based on the previously proposed strategies of local feature association, adaptive gating, and mismatch optimization, this paper presents the Dynamic Optimized Joint Compatibility Branch and Bound (DOJCBB) data association algorithm. The specific procedure of the DOJCBB algorithm is as follows:

(1) Determine the local feature association region based on the robot's system parameters and the currently stored environment landmark features using Equation (16).

(2) Calculate the Mahalanobis distance between each environment landmark feature within the determined local region and the current observation, as shown in the following equation:

$$\begin{cases} v_i(k) = z_j(k) - z_i(\hat{k}), j = 1, 2, L, N \\ d_k^2 = v_i(k)^T S(k) v_i(k) \end{cases} \quad (21)$$

Where $v_i(k)$ represents the observation residual, and $S(k)$ represents the predicted covariance of the current observation. Combining the robot's current system state error covariance matrix $P(k)$, the uncertainty of the environment landmark feature estimation $P_i(k)$ from Equation (18), and the initial value of the state error covariance P_0 , the dynamic confidence $\mu_i(k)$ at time k is calculated using Equation (19):

$$\mu_i(k) = 99.5\% \times \frac{P_0}{P_i(k)} \quad (22)$$

Based on the dynamic confidence at different time steps $\mu_i(k)$, different chi-square distribution values are selected from Table 1 as the dynamic adaptive threshold G for data association.

(3) Combining the explanation tree model with the Joint Compatibility Branch and Bound (JCBB) algorithm, search for association solutions that satisfy both the joint compatibility criterion and the mutual exclusion optimization condition.

(4) If multiple association solutions with the maximum number of non-empty associations exist, select the solution with the minimum joint Mahalanobis distance as the optimal association solution using Equation (20).

4. Simulation and Analysis

To validate the effectiveness of the improved data association algorithm in this paper, this section establishes a simulation environment for data association algorithms based on the simulation platform developed by Tim Bailey et al. at the University of Sydney, as shown in Figure 5. The size of the simulation environment is 10×10 , and the green line represents the true path of the moving robot. The robot moves counterclockwise along this path starting from the coordinate origin. The blue asterisk (*) points represent the environmental landmark features. The simulation parameters are shown in Table 2.

Table 2. Simulation Environment Parameter Settings

Movement Speed	0.5m/s	Control Frequency	40Hz
Maximum Observation Distance	2m	Observation Frequency	5Hz
Maximum Steering Angle	20°/s	Control Noise	(0.05m/s, 3°)
Observation Range	180°	Observation Noise	(0.1m/s, 1°)

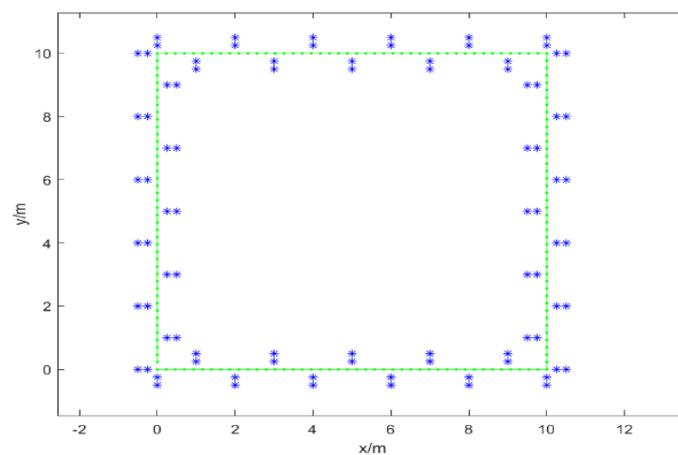


Figure 4 Simulation environment

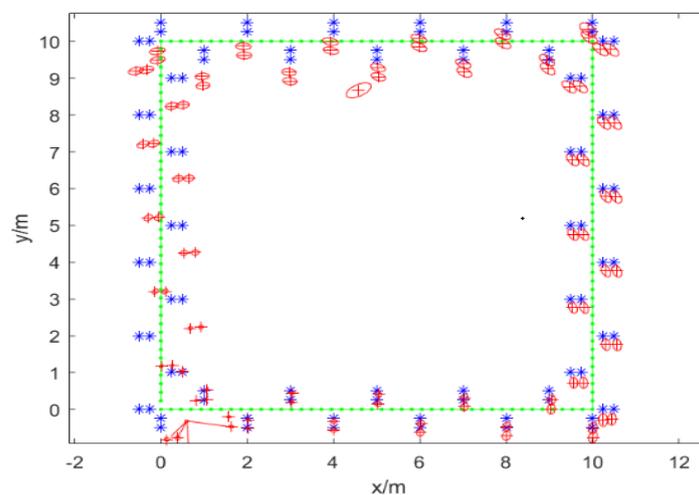


Figure 5 ICNN Simulation Results of Data Association

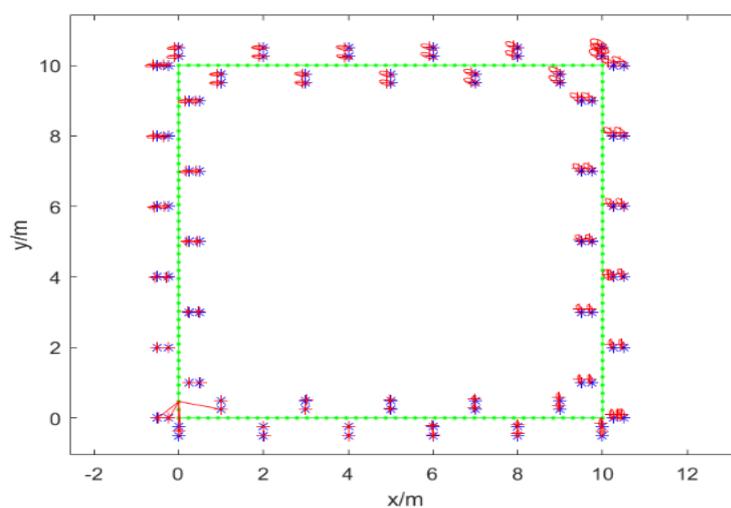


Figure 6 JCBB Simulation Results of Data Association

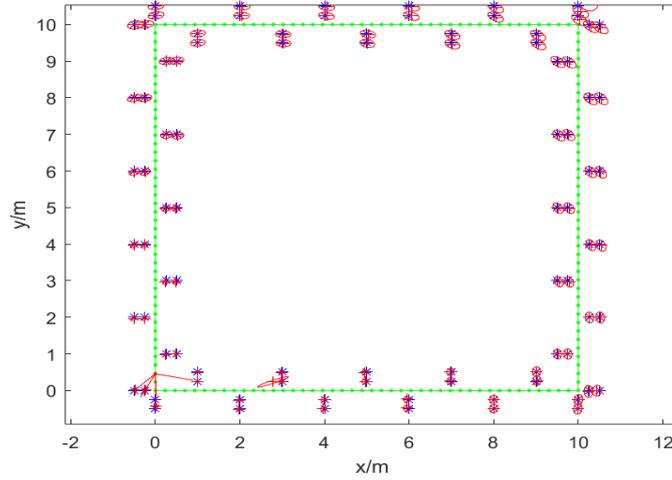


Figure 7. DOJCBB Simulation Results of Data Association

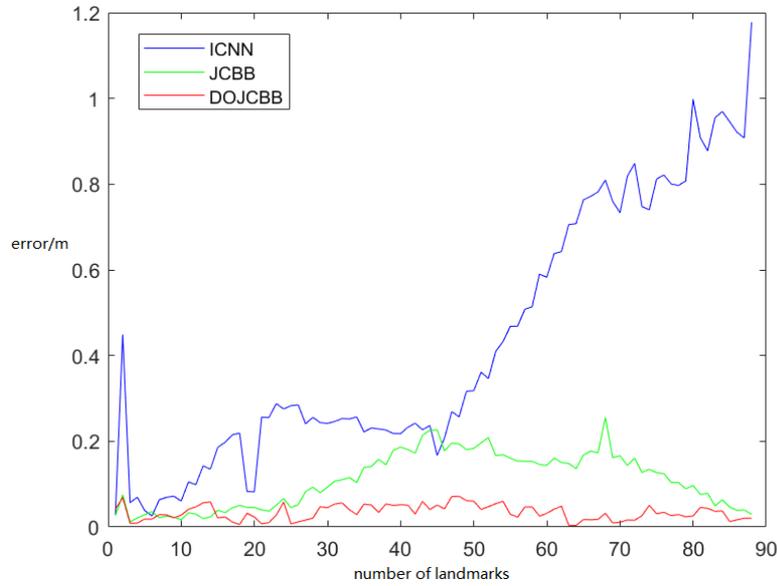


Figure 8. Data Association Error Analysis

Based on the simulated environment and parameter settings in Table 2, Figures 5, 6, and 7 show the data association results of the ICNN data association algorithm, JCBB data association algorithm, and the improved DOJCBB data association algorithm proposed in this paper, respectively. The red "+" symbols in the figures represent the estimated positions of the environmental landmarks, and the blue "*" symbols represent the actual positions of the environmental landmarks. Figure 8 shows the error variation curves of the average estimated landmark positions and actual positions for the three algorithms in 20 experiments, as described by Equation (23).

$$e = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m \left[\left(x_m^{(i)}(j) - x_t^{(i)}(j) \right)^2 + \left(y_m^{(i)}(j) - y_t^{(i)}(j) \right)^2 \right]}{n}} \quad (23)$$

Where, $(x_m^{(i)}(j), y_m^{(i)}(j))$ represents the estimated position of the j-th environmental landmark in the i-th experiment; $(x_t^{(i)}(j), y_t^{(i)}(j))$ represents the actual position of the environmental landmark. Through the analysis of the association results obtained by the three algorithms and the relative distances between landmark estimates and actual positions, the following observations are made:

(1)When using the Independent Compatible Nearest Neighbor (ICNN) algorithm, the uncertainty of the robot system increases with the motion time, leading to an increasing estimation error in the positions of the environmental landmarks.

(2)Compared to the ICNN algorithm, the Joint Compatibility Branch and Bound (JCBB) algorithm considers the correlation between the environmental landmarks and the observation data using the joint compatibility test. It also takes into account the correlation between the system state and the environmental landmarks. The data association results of JCBB algorithm show a significant improvement in the estimation error of landmark positions compared to the ICNN algorithm.

(3)The proposed Dynamic Optimization of Joint Compatibility Branch and Bound (DOJCBB) algorithm incorporates dynamically adaptive changes in the association threshold and optimization criteria. It enhances the accuracy of the data association algorithm, and the error during the algorithm execution only fluctuates within a small range.

5. Conclusion

This paper first provides a mathematical description of data association and introduces the Joint Compatibility Branch and Bound (JCBB) algorithm. Then, addressing the issues of computational resource waste in complex environments and the presence of multiple maximum matching hypotheses in JCBB, the Dynamic Optimization JCBB (DOJCBB) data association algorithm is proposed and simulated. Experimental results validate the improved accuracy of the DOJCBB algorithm compared to ICNN and JCBB algorithms.

Funding

This paper is supported by Projects of major scientific and technological research of Ningbo City(2020Z065, 2021Z059,2022Z056,2023Z050(the second batch)), Major instrument special projects of the ministry of science and technology of China(2018YFF01013200), Projects of major scientific and technological research of Beilun District, Ningbo City(2021BLG002, 2022G009), Projects of engineering research center of Ningbo City (Yinzhou District Development and Reform Bureau [2022] 23), Projects of scientific and technological research of colleges student's of China(202213001008).

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

- [1] Atanasov N, Bowman S L, Daniilidis K, et al. A Unifying View of Geometry, Semantics, and Data Association in SLAM . *Proceedings of the IJCAI*. 2018: 5204-5208.
- [2] Choi J, Choi M, Chung W K, et al. Data association using relative compatibility of multiple observations for EKF-SLAM . *Intelligent Service Robotics*, 2016, 9: 177-185.
- [3] Singer R A, Stein J. An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems . *Proceedings of the IEEE Conference on Decision and Control*. 1971: 171-175.
- [4] Han Z, Wang F, Li Z. Research on nearest neighbor data association algorithm based on target “dynamic” monitoring model . *Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2020: 665-668.
- [5] Bar-Shalom Y. Extension of the probabilistic data association filter to multi-target tracking . *Proceedings of the Proc of the 5th Symp Nonlinear Estimation, San Diego, Sept 1974*. 1974.
- [6] Y. B-S, T. K, X. L. Probabilistic data association techniques for target tracking with applications to sonar, radar and EO sensors . *IEEE Aerospace and Electronic Systems Magazine*, 2005, 20(8): 37-56.
- [7] Fortmann T E, Bar-Shalom Y, Scheffe M. Multi-target tracking using joint probabilistic data association . *Proceedings of the 1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*. IEEE, 1980: 807-812.
- [8] Neira J, Tardós J D. Data association in stochastic mapping using the joint compatibility test . *IEEE Transactions on robotics and automation*, 2001, 17(6): 890-897.
- [9] Zhou and Zhao. An optimized data association algorithm for SLAM problem . *Robot*, 2009, 31 (3) : 217-223.
- [10] Research on mobile robot localization method based on Bayesian theory; chongqing University, 2015.
- [11] Wang J, Englot B. Robust exploration with multiple hypothesis data association . *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018: 3537-3544.
- [12] Zeng Wenjing, Zhang Tiedong, Xu Yuru, et al. A SLAM data association method based on ant colony algorithm . *Computer Applications*, 2009, 29 (1) : 136-138,148.
- [13] Du Hangyuan, Hao Yanling, Zhao Yuxin. SLAM data association method based on fuzzy logic . *System engineering and electronic technology*, 2011, 33 (11) : 2468-2473.
- [14] Li Yanju, Xiao Yufeng, Gu Song, et al. Research on SLAM data association algorithm based on laser sensor . *Microcomputer and application*, 2017, 36 (2) : 78-82.
- [15] Myung I J. Tutorial on maximum likelihood estimation . *Journal of mathematical Psychology*, 2003, 47(1): 90-100.
- [16] Wang Yaoqiang. Data association problem in simultaneous localization and mapping technology ; harbin Institute of Technology, 2008.