# A Method of Indoor Space Layout for Home Stay Based on Binocular Vision SLAM

Liushi Qin[1](✉) and Zhengfeng Huang[2]

[1] College of Art and Design, Nanning University, Nanning 530000, China
315032245@qq.com

[2] Polytechnic Institute, Guangxi Agricultural Vocational and Technical University, Nanning 530000, China

**Abstract.** Reasonable layout method can ensure reasonable spatial layout and data accuracy. Traditional layout methods can not guarantee the accuracy of feature data in complex spatial environment, which leads to unreasonable layout. In order to solve this problem, a method of indoor space layout of home stay based on binocular vision SLAM is proposed. According to the RGB color image of a given indoor scene, the interior room layout boundary of the scene is detected, and the foreground and background are segmented. Use Kinect to obtain spatial point cloud data to avoid data loss. The geometric simulation model of spatial features is established to realize the projection from spatial points to image plane. The binocular vision SLAM point and line feature selection method is adopted to make the point and line features cluster evenly. Match features, analyze the constraint relationship between the projected line segment and the line segment to be matched. The visual area of each functional area is calculated, and the indoor space layout model is constructed by combining the anthropometric technology. According to the geometric characteristics of layout containers and the attributes of layout objects, the open layout of indoor space is realized. The experimental results show that the clustering effect of this method is good, and the maximum space utilization rate can reach 92%, which increases the spatial interaction and openness, and is consistent with the ideal layout effect.

**Keywords:** Binocular Vision SLAM · Home Stay Room · Spatial Layout

## 1 Introduction

The simulation of indoor space layout of home stay has played a huge role in improving people's quality of life, because people live indoors 87%–90% of the time, and the simulation of indoor space feature layout has been the focus of relevant scholars' research. The conventional indoor spatial feature layout method can be completed by 3D laser scanner or sensor, but due to its poor clustering effect, the overall deviation can not meet the ideal requirements, resulting in low data accuracy. With the introduction of Kinect sensor, it has a certain impact on the traditional multi-layer layout simulation of indoor


B. Wang et al. (Eds.): ICMTEL 2023, LNICST 532, pp. 377–393, 2024.
https://doi.org/10.1007/978-3-031-50571-3_28

space features. Kinect sensor can obtain the spatial information of indoor scene at a certain frequency, and output RCB image and infrared depth image with high resolution. In the research fields of indoor 3D map building and indoor space simulation, the use of Kinect devices can help the space to be built better and faster, which not only makes the Kinect sensor get in-depth research, but also makes space reconstruction and other technologies get better development. In the research at home and abroad, the simulation method of multi-level layout of indoor space features is more to improve the convergence and efficiency, but the accuracy is insufficient [1]. At present, literature [2] proposes a building space layout planning method based on the differential evolution method. By constructing a parametric snake curve, the internal energy and external energy utilization function of the building edge can be obtained, the variation probability and cross probability of the edge contour can be calculated, the maximum number of iterations can be obtained, and the external edge space layout of high-rise buildings can be reasonably planned. This method has strong convergence, but the overall accuracy is insufficient. Literature [3] proposed a spatial layout simulation method based on the output structure model. Through its semantic constrained multi label images, the unstructured point cloud was automatically divided into rooms. The horizontal slice of the point stone with a single room was projected to the plan to form a binary image. Line extraction and regularization were performed to generate the plan lines. Finally, a structured model is constructed by multi label graph cutting, and the indoor space layout is carried out using this model. This method has high efficiency in the indoor layout, but the accuracy is still insufficient. Although these two simulation methods have been relatively mature, it is difficult to ensure reasonable spatial layout and data accuracy in the face of complex spatial structure. Therefore, a method of indoor space layout of home stay based on binocular vision SLAM is proposed to solve the problem of poor accuracy in the above traditional simulation methods. The superpixel segmentation algorithm was used to initially segment the image, and binocular visual SLAM was used to extract feature points, and the dynamic region was detected and segmented. A depth sensor is used to collect indoor spatial position information. After removing noise from the data and completing filtering processing, the indoor spatial feature set model is established by using point cloud data. The multi-level spatial feature model was constructed, the residential interior space layout scheme was designed, and the binocular visual SLAM layout point and line features were selected to establish the open interior space layout.

## 2 Layout Scene Segmentation Based on Binocular Vision SLAM

The proposed spatial layout method uses binocular vision SLAM to extract feature points and detect and segment dynamic regions. However, due to the sparse dynamic feature points in the scene, the overall contour of the target cannot be well described, and the scene needs to be segmented. The purpose of scene segmentation is to obtain the approximate contour of multiple foreground objects in the field of vision. By combining image segmentation methods and slope smoothing algorithms, reducing the number of iterations and using semi global matching algorithms, the calculation time of scene segmentation is changed from the original 2 s/frame to the existing 80 ms/frame under the premise of acceptable error, which can meet the real-time computing requirements on the CPU [3].

The input of this algorithm is a binocular image. First, a super pixel segmentation algorithm is used to segment the image initially, convert the RGB image into a Lab color space image, and combine the color values and coordinates of each pixel into a 5-dimensional vector. This vector is used as the standard for measuring the similarity between pixels, and the segmentation results are obtained through multiple iterations. It can be seen that the segmentation boundary in the image can better describe the contour of various objects. By defining a total cost function, and using the block coordinate descent method as an optimization algorithm, the cost is minimized. After iteration for a certain number of times, the bevel parallax model of each hyperpixel block in the scene and the boundary type between the hyperpixel blocks are obtained [4]. According to the parameters of two adjacent inclined planes, the boundary can be divided into three categories, namely mutual occlusion, hinge connection and coplanarity. The cost function mainly includes the following five parts:

Part 1: Lab color space cost. The closer the pixel color is to the average color of the segmentation block, the smaller the cost is. The formula is as follows:

$$f_1(x, z_s) = \|\tau(x) - z_s\|_2^2 \tag{1}$$

In Formula (1), $x$ represents the current pixel coordinate; $\tau$ represents the current pixel color value; $z_s$ represents the average color value of the partition block $s$ where the current pixel is located.

Part 2: pixel position cost. The closer the pixel position is to the average position of the segmentation block, the smaller the cost is. The formula is as follows:

$$f_2(x, o_s) = \|x - o_s\|_2^2 \tag{2}$$

In Formula (2), $o_s$ represents the center coordinate of the partition block where $x$ is located.

Part 3: Parallax cost of the inclined plane, that is, the error between the real parallax of the current pixel and the parallax calculated by the corresponding segmentation block using the inclined plane parameters. The smaller the error, the smaller the parallax cost. The formula is:

$$f_3(x, \alpha, g) = \begin{cases} \left(d(x) - \hat{d}(x, \alpha)\right)^2 & if\ g = 0 \\ \lambda & if\ g = 1 \end{cases} \tag{3}$$

In Formula (3), $\alpha$ is the slope parameter of the partition block corresponding to the pixel; $g$ represents the local point marker bit of the pixel; $\lambda$ is the cost constant $d, \hat{d}$ The true parallax of the current pixel point is calculated by using the slope parameters [5].

Part 4: boundary length cost, that is, the smaller the number of adjacent segmentation blocks and middle pixels, the smaller the cost. The formula is as follows:

$$f_4(s_1, s_2) = \begin{cases} 0 & if\ s_1 = s_2 \\ 1 & otherwise \end{cases} \tag{4}$$

In Formula (4), $s_1, s_2$ represents two partitions respectively.

Part 5: Boundary type cost. Set the penalty coefficients of the three boundary types (occlusion, hinge and coplanar) as $\eta_1$, $\eta_2$, 0 respectively. The formula is as follows:

$$f_5(k) = \begin{cases} \eta_1 & \textit{if}\ \text{Boundary occlusion} \\ \eta_2 & \textit{if}\ \text{Boundary hinge} \\ 0 & f\ \text{Coplanar boundary} \end{cases} \quad (5)$$

In Formula (5), $k$ represents the boundary type of adjacent blocks.

Home stay indoor scene image layout estimation is the initial unit as well as the key unit in image scene understanding research. Its basic goal is to detect the layout boundary of the room in the scene according to the given RGB color image of the indoor scene (or the frame captured in the video of the indoor scene) [6]. In order to intuitively express the concept of indoor scene layout estimation, the occluded boundary obtained after segmentation can be obtained according to the optimization results, that is, the contour boundary line of the foreground and background. Figure 1 provides an indoor scene layout annotation work.

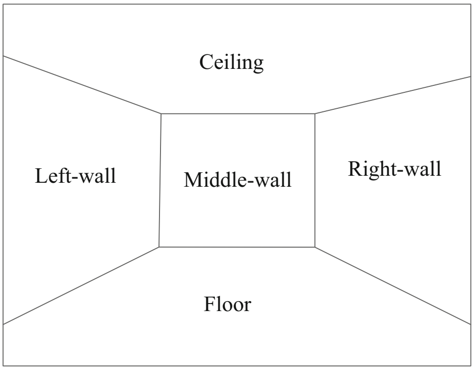



**Fig. 1.** Indoor scene layout annotation

The upper part of Fig. 1 lists five types of geometric semantic information that may exist in the datum mark, namely, the ground, the middle wall, the right wall, the left wall, and the ceiling, and lists the corresponding mark colors. In general, the benchmark annotation image is a subjective visual evaluation standard used to evaluate the semantic annotation results obtained by the semantic annotation method of indoor scenes [7]. The indoor scene image layout estimation is to obtain the accurate segmentation of each geometric area surface in the indoor scene image. With this segmentation image, we can further analyze and infer the internal relationship between each semantic category in the indoor scene image, and also continue to mine the deeper information contained in the semantic category itself, so as to achieve a higher level of image scene understanding.

## 3 Building a Multi-level Model of Spatial Features

Use Kinect to obtain spatial point cloud data. When Kinect scans the scene, due to occlusion or light, there may be missing data. Generally, some positions or scanning angles that cannot be scanned are incorrect [8]. Therefore, in the scanning process, it is necessary to adjust the position and angle of Kinect equipment, establish a local coordinate system with Kinect as the center, and collect indoor spatial data with the support of this coordinate system. In the data acquisition process of Kinect equipment, the main device is the depth camera, which adopts optical coding technology and mainly uses the depth sensor to collect the indoor space location information, and its internal depth distance is represented by a 13 bit binary number. Suppose that the color information of a single pixel in the depth map collected from the Kinect device is a 16 bit data parameter [9]. According to the description of pixel information in the SDK document, in the pixel information of the depth image, the first 13 bits of the data contain the depth information, and the last 3 bits of the data contain the user index. After the original depth data of Kinect is obtained, the obtained color information and user ID are divided to ensure that the first 13 bits of data remain unchanged and the last 3 bits return to 0 [10–12]. After SDK internal processing, depth data between 0–4095 can be obtained through Kinect device. Generally, the depth information of the current pixel is obtained by shifting Kinect's original data, which reflects the real world distance of the pixel.

There are some areas to be repaired in the data collected by Kinect. These areas are hollow areas, areas that cannot be monitored, and shadow areas. For different areas, corresponding repair measures are taken to obtain more complete depth data. The data repaired can not be directly used in the simulation model building, because some noise points with continuous depth will appear at the edge of the image during the repair process, You need to use filters to process point cloud data. After the noise in the data is removed and the filtering process is completed, the indoor spatial feature set model is established using point cloud data.

The spatial data obtained through the above process are mainly displayed in the form of depth images. The depth images are two-dimensional maps. Considering that the spatial characteristics are mainly composed of two-dimensional data, when acquiring the depth images of indoor space, it is necessary to convert the depth maps into the relationships of three-dimensional spatial points in the real world. Kinect has an automatic correction function, which can calibrate the depth and color cameras in real time, Therefore, the process of establishing the geometric simulation model of spatial features is to complete the projection of spatial points to the image plane. In order to quantitatively describe the establishment process of the geometric model, the world coordinate system, camera coordinate system, image pixel coordinate system and physical coordinate system are respectively established, as shown in Fig. 2.

The $(O_1-X_1Y_1Z_1)$ shown in Fig. 2 represents the camera coordinate system, the $(O_2-X_2Y_2Z_1)$ represents the physical coordinate system, and the $(O_3-X_3Y_3)$ represents the world coordinate system. Suppose that the physical size of the image pixel obtained by Kinect in the x direction is $d_x$, and the physical size in the y direction is $d_y$. It can be seen that the conversion relationship between the world coordinate system and the camera coordinate system is as follows:

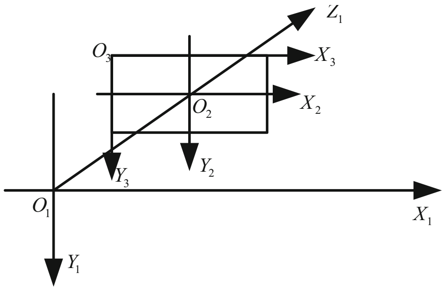



**Fig. 2.** Camera imaging model

$$\begin{cases} X_1 = (x - x_0) \times \dfrac{d_x}{\alpha} \\ Y_1 = (y - y_0) \times \dfrac{d_y}{\beta} \\ Z_1 = d_z \end{cases} \tag{6}$$

In Formula (6), $x_0, y_0$ represents the initial camera coordinate; $\alpha, \beta$ represents the internal parameters of the camera; $d_z$ is the physical dimension in the z direction.

The formula can be used to complete the conversion of pixel coordinates of depth image to three-dimensional coordinates of space points, and the Kinect sensor is used as the origin of the world coordinate system. Through this transformation, the simulation model corresponding to the indoor space depth image can be obtained.

# 4 Interior Space Layout Scheme Design of Home Stay

## 4.1 Point Line Feature Selection of Binocular Vision SLAM Layout

It is extended on the basis of SLAM and adopts the same principal structure. The system is mainly divided into three parallel threads: tracking thread, local composition thread, closed-loop detection thread, and additional global optimization thread. Global optimization is only created during closed-loop. The main contents of each thread are as follows:

(1) Tracking thread: the input is the image sequence collected by the binocular camera, which is divided into left image and right image. The left and right images at the same time are called one frame. Image preprocessing includes image distortion correction, detection and description of feature points and line segments, and binocular matching. Tracking is divided into two stages, one is to track adjacent frames, the other is to track local maps, and get the camera's position and attitude by minimizing the re projection error. Finally, the key frame of the current frame is judged.

(2) Local composition thread: after the tracking thread inserts the key frame, optimize the points, lines and positions in the local map. At the same time, the spatial points and spatial lines in the map are eliminated according to the statistical information, the stable tracking part is retained, and the key frames with redundant information in the map are eliminated. After keyframe insertion, new map points and lines are created by combining another frame in the local map.
(3) Closed loop thread: the dictionary tree is used for closed loop detection. When a closed loop is detected, the transformation between the closed loop frame and the current frame is calculated, and the cumulative error and the position of map points and lines are corrected through the pose map optimization.
(4) Global optimization thread.

In addition, a scene recognition module is constructed based on point and line features for closed-loop detection. At the same time, the system maintains the elements in the environment map, including map points, map lines, key frames, and the connection relationship established between key frames, that is, common view and minimum generated numerical map. If there are commonly observed features between two frames, the two frames are taken as the vertices in the graph, and the number of commonly observed features is used to establish an undirected graph for the edge weight. Finally, a common view is formed. The minimum spanning tree is a sub graph with higher weight in the common view. By querying the common view, you can get a window connected to the current frame to form a local map.

Under this map, a method of point and line feature selection for binocular vision SLAM is provided, including:

Step 1: Based on the removal step of abnormal point and line features, first randomly extract the matching pairs of point and line features with preset proportions, solve the relative pose of the front and rear frames, and then obtain the interior point set under the pose, then repeat the above process for a fixed number of times to determine the maximum interior point set, and finally use the maximum interior point set to solve the final pose;

Randomly extract one or more point features and line features, and use the randomly extracted point and line features to estimate the inter frame pose to obtain the optimized pose. According to the optimized pose, the re projection error of all point line features is obtained, the re projection error threshold is set, and the number of interior points is counted. The re projection error of each point feature is calculated as follows:

$$e = x' - h(\varepsilon, c) \tag{7}$$

In Formula (7), $x'$ represents the coordinates of point features under the image coordinates of the next frame; $\varepsilon$ represents the algebra of the relative position and pose of the camera coordinate system of the previous frame and the camera coordinate system of the next frame; $c$ represents the 3D feature point under the camera coordinate system of the previous frame; $h$ represents the function of 3D-2D projection after the pose transformation of $c$.

Step 2: Construction of error transfer model for point and line features: First, assume the measurement error covariance matrix of the image according to the actual situation, and then, based on the nonlinear transformation relationship of Gaussian distribution random variables, finally obtain the error covariance matrix for feature introduction pose solution. The schematic diagram of the error transfer process for point and line features is shown in Fig. 3.

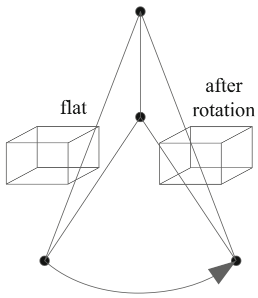



**Fig. 3.** Schematic diagram of the transmission process of point line characteristic error

Step 3: Selection of point and line features based on error transfer model Step: solve the problem with greedy algorithm, and retain the specified number of features by removing the contribution of features to pose solution;

Step 4: Select the point and line features based on the uniform distribution of feature space, respectively construct the artificial features of feature points and lines, including their own features and domain features, and then cluster the point features and line features in the artificial feature space, so that the point and line features are evenly distributed in each class.

### 4.2 Feature Matching

After binocular matching and triangulation in the first frame, the initial environment map is obtained. As the camera moves, the space points and lines of the environment map will increase. The movement of the camera will not change the spatial position of the points and lines in the environment map, but their positions in the image will change accordingly due to the movement of the camera. The pose estimation firstly finds the matching relationship between 3D features and 2D features through tracking matching, and then solves the pose by minimizing the cost function through nonlinear optimization.

The matching includes two parts: adjacent frame feature matching and local map feature matching. The feature matching of adjacent frames is to use the information of previous and subsequent frames and use fewer features to roughly estimate the pose of the camera at the current time; However, local map feature matching involves multi frame information, and more constraints are used to bring more accurate solutions. For feature matching of two adjacent frames, the space points and lines tracked in the previous frame can be projected into the image at the current time, and descriptive sub matching can be carried out by limiting the search range, which not only improves the efficiency but also reduces the error matching rate. It is known that there is a space point and a space line under the world coordinate system observed in the last frame (the information of the start point and end point of the space line is mainly used in projection). In order to project onto the current image, you need to know the pose of the camera at the current frame. However, the current pose is a quantity to be solved, and the uniform velocity model is used to predict the current pose.

If the camera is moving at a uniform speed, the position and attitude of the previous frame can be used to estimate the position and attitude of the current frame. This model is only applicable to the situation where the speed and direction of motion are relatively consistent or the speed of motion is relatively slow, which has certain limitations. The uniform motion model of the camera in Lie group form is as follows:

$$U_k' = \exp(\varsigma)U_k \tag{8}$$

In Formula (8), $\varsigma$ represents the motion parameter; $U_k$ represents the current frame pose. First, determine whether the feature is within the camera's field of view at the current time. There are two judging conditions: one is that when the current frame is used as the reference system, the feature is in front of the camera, and the other is that the re projected coordinates must be within the image range. The projected point cannot coincide with the matching point, so it is necessary to search in the circular area with the projection point as the center radius.

The projection of the space line segment is similar to that of the space point, but the space line segment is partially observed. With the movement of the camera, a part of the space line segment will be observed by the camera. There are only two cases of the space point, namely, in the camera's field of view and not in the camera's field of view. This paper adopts the following strategies to deal with this problem:

(1) Take the current camera coordinate system as the reference coordinate system of the space segment, and calculate the two endpoints relative to the current camera coordinate system;
(2) If both endpoints are behind the camera, projection matching is not performed;
(3) If there is an endpoint behind the camera, the intersection point between the space line and the image plane is calculated, and then the intersection point coordinates are obtained by substituting;
(4) The two endpoints in front of the camera are projected to obtain the coordinates in the image pixel coordinate system. Generally, the re projection coordinates of the intersection point in the previous step are not within the image, and the projected line segments may be outside the image range. Therefore, Liang Basky segment clipping algorithm is adopted for processing. Segment clipping means that the specified window is used as the graphic boundary to retain the segments inside the window, while the segments outside the window are discarded. Liang Barsky line clipping algorithm has high computational efficiency, and can retain the starting and ending information of the original line end points. The space line re projection result shows the space line re projection result. The red line represents the projected position of the space line in the previous frame, the black line represents the line segment detected by feature extraction in the current frame, and the starting point of each line segment is represented by a dot. It can be found that the projected line segment and the line segment to be matched have certain constraints.

### 4.3 Open Indoor Space Layout

Effective layout of indoor space to meet the needs of users of different identities is the main factor considered in indoor space modeling layout. The design and layout of indoor space environment based on openness can objectively reflect the different behavior needs of users in different functional areas of different comfort indoor spaces. The indoor space layout method based on the openness of indoor space includes:

Obtain the indoor space parameter data to be arranged, and obtain the parameter data classification information. According to the parameter data and classification information, the three-dimensional model of indoor space is built. According to the 3D model, obtain the openness parameters of each functional partition of the indoor space to be laid out, the range of openness standard parameters corresponding to users of each functional partition of the indoor space, and the openness parameters of each functional partition of the indoor space to be laid out, and adjust the layout of each functional partition. Based on the indoor space after layout adjustment, the openness parameters of each functional partition are obtained again until the openness parameters of each functional partition meet the range of the openness standard parameters. The view area of each functional partition of the open parameter indoor space and the view volume of each functional partition, and the view area of each functional partition, expressed as:

$$A_v = \sum_{0 \le i \le 1} a_i A_v^{1i} + \sum_{0 \le j \le m} b_j A_v^{2j} + \chi A_v^3 \tag{9}$$

In Formula (9), $a_i$, $b_j$, $\chi$ respectively represent the area coefficient of the $i$ transparent area, the area coefficient of the $j$ transparent area, and the area coefficient of the opaque area $A_v^{1i}$, $A_v^{2j}$, $A_v^3$ represents the area of the $i$ transparent area, the area of the $j$ transparent area, and the area of the opaque area in the viewing area corresponding to the viewpoint, respectively.

For the selection of viewpoints, it is necessary to first set the viewpoint height and the minimum area corresponding to the viewpoint, and divide them within the viewpoint range of the indoor space according to the viewpoint height and the minimum area corresponding to the viewpoint, so as to obtain the number of optional viewpoints in the indoor space. The indoor space openness parameter also includes the standard deviation of the view area and view volume corresponding to the optional view point. The parameter data is used as all entity elements existing in the occlusion. The parameter data is classified according to the transparency, translucency and opacity of the entity element materials to obtain the classification information. The viewpoint range is determined in the indoor space according to the average height of users using the indoor space to be laid out, the minimum walking width of users, and the indoor space parameter data. The layout adjustment of each functional partition, including obtaining the layout rules and constraints of each functional partition of the indoor space before the layout adjustment of the functional partition; The layout of each functional partition is adjusted according to the layout rules and constraints of each functional partition. The openness parameters of each functional partition are obtained again based on the indoor space after layout adjustment until the openness parameters of each functional partition meet the range of the openness standard parameters; When there are multiple optional layout schemes that meet the range of the openness standard parameters, the scheme corresponding to the minimum sum of the apparent area standard deviation and the apparent volume standard deviation in all the optional layout schemes is the optimal scheme.

Based on the optimal organizational structure of the indoor scene obtained above, the indoor space layout model is constructed by using anthropometric technology. The specific construction process is as follows: The indoor space layout model mainly distributes the attributes of the layout objects according to the characteristics of the layout containers. First, define the characteristics of the layout containers and the attributes of the layout objects. Second, select multiple indoor scene samples with reasonable layout, extract their relevant information, and build the indoor space layout model. For the indoor space layout of home stay, the closed space obtained by projecting the indoor space outline on the plane is called the layout container. The score of indoor space layout is calculated according to the geometric characteristics of the layout container and the attributes of the layout object. Through the above process, the indoor space layout model is built to provide tool support for the adaptation of indoor scene elements.

Based on the indoor space layout model constructed above, the comfort index is defined by using fuzzy theory. The specific definition process is as follows: According to the research, comfort is mainly determined by five indicators, namely floor area ratio, lighting ratio, noise, ventilation and thermal insulation. The obtained openness index is expressed as:

$$\psi = \frac{\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 + \varphi_5}{5} \tag{10}$$

In Formula (10), $\varphi_1$, $\varphi_2$, $\varphi_3$, $\varphi_4$, $\varphi_5$ represents plot ratio, lighting ratio, noise, ventilation and thermal insulation respectively. Based on the comfort evaluation index obtained from Formula (10), the indoor scene elements are adapted. The specific adaptation process is as follows: For the indoor space layout of home stay, the complete indoor space layout scene of home stay can be obtained only by adapting specific scene elements in the functional area. In order to complete the furniture combination of a certain function, semantic files are used to describe the attributes of functional areas. Through the above process, the adaptation of indoor scene elements is completed to achieve the comfortable layout of the indoor space of the home stay and provide a more comfortable living environment for residents.

## 5 Experiment

In order to verify the rationality of the research on the indoor space layout method of home stay based on binocular vision SLAM, the simulation environment is set as follows: Core quad core 4.0G processor, 16 GB memory, and Windows10 operating system. 1200 samples of 3D interior design images are collected, and the size of visual block area is 256 × two hundred and fifty-six × 224. For 3D design, the normalized feature matching coefficient is 0.10. The virtual form visual simulation experiment is used to verify the design effect of the system.

### 5.1 Experiment Description

In the simulation experiment, it is necessary to verify the correctness and advantages of introducing linear features into back-end optimization. The design of the simulation scene comes from Joan Sola's paper, including a bottom edge of 12× the 15 m house has 25 sides, and the number of points will be adjusted in the experiment, as shown in Fig. 4.

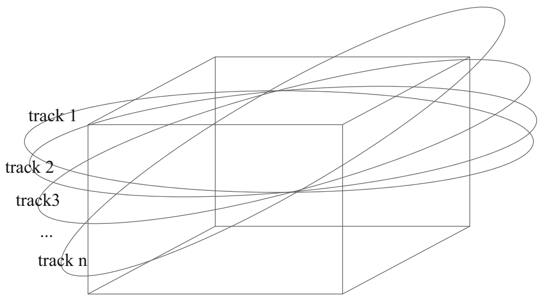


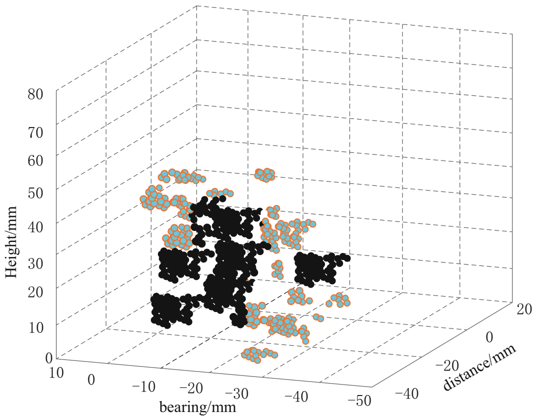

**Fig. 4.** Description of experiment track

The baseline of the binocular camera is 0.5m, and the pixel is 640 × 480, the focal length of both horizontal and vertical directions is 500 pixels, and its running track is shown by the curve in the figure.

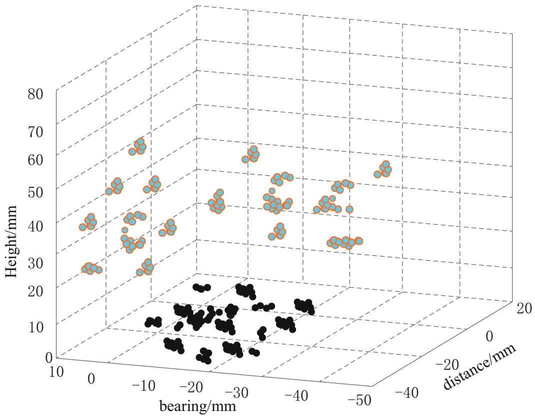
### 5.2 Clustering Analysis of Spatial Feature Data

In the spatial feature data clustering experiment, it is mainly to verify the accuracy of data from different methods. The better the data clustering effect is, the higher the data accuracy is, the less the burden on the simulation process is, and the more reliable the simulation results are; On the contrary, the worse the data clustering effect and the lower the data precision are, the less reliable the simulation results are. The building spatial layout planning method based on differential evolution method, the spatial layout simulation method based on output structure model and the layout method based on binocular vision SLAM are used to output simulation data clustering results, and different simulation methods are compared and analyzed according to the results. The specific results are shown in Fig. 5.

By comparing the results in the diagram, we can see that the spatial layout policy method based on differential evolution method is mixed with different characteristic data, and there is no obvious segmentation boundary; The spatial layout simulation method based on the output structure model has obvious dispersion of data with different characteristics, but the data is scattered and not centralized; In the layout method based on binocular vision SLAM, the clustering of data with different characteristics is obvious. There is no complex situation. Combined with the overall deviation experiment results of the simulation model, we can see that the proposed method has small overall deviation of the simulation model, good data clustering effect, and high data accuracy. This method is superior to traditional simulation methods.




(a) A method of architectural space layout planning based on differential evolution



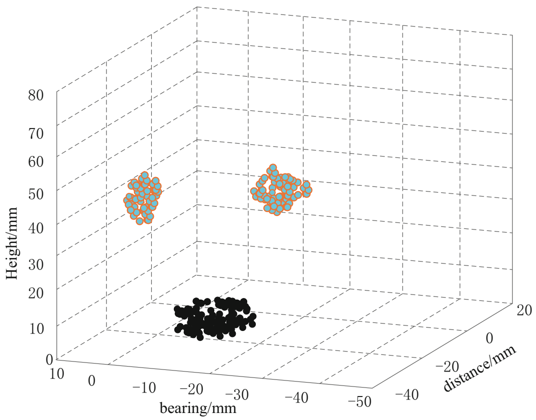

(b) Spatial layout simulation method based on output structure model

**Fig. 5.** Comparison and analysis of spatial layout feature data clustering results of three methods

### 5.3 Space Utilization Analysis

Use these three different methods to compare and analyze space utilization, and the comparison results are shown in Table 1.

As shown in Table 1, there is a large difference between the layout method based on binocular vision SLAM, the building space layout planning method based on differential evolution method, and the space utilization value of the space layout simulation method based on output structure model, which indicates that the experimental data is valid. Through comparison, it is found that the space utilization ratio of the proposed method is higher than that of the existing methods, and the maximum value can reach 92%.




(c) Layout method based on binocular vision SLAM

**Fig. 5.** (*continued*)

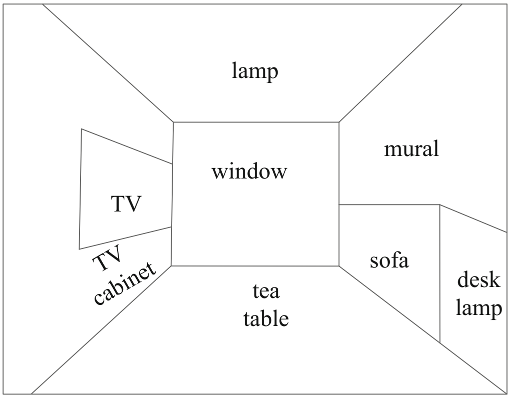
**Table 1.** Comparative analysis of space utilization rate of different methods/%

| Number of experiments/time | Differential evolution method | Output structure model | Binocular vision SLAM |
|---|---|---|---|
| 10 | 60 | 75 | 90 |
| 20 | 62 | 77 | 91 |
| 30 | 59 | 80 | 90 |
| 40 | 58 | 76 | 92 |
| 50 | 55 | 74 | 92 |
| 60 | 57 | 70 | 91 |

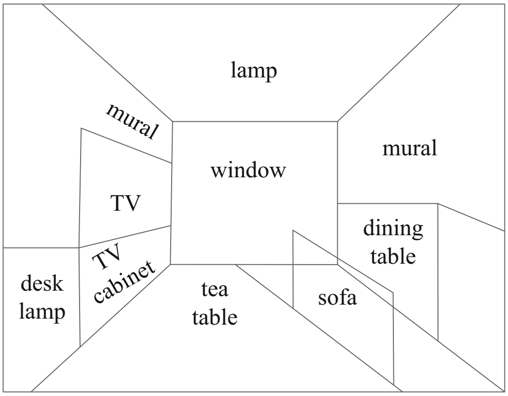
## 5.4 Analysis of Visual Effect of Space Layout

Taking the living room as an example, the space layout visual effects of the three methods are shown in Fig. 6.

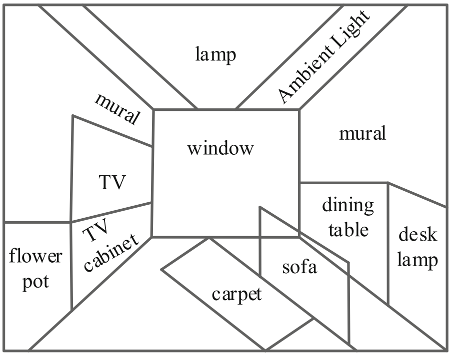
It can be seen from Fig. 6 that in order to adapt to the sustainable development of housing, the layout method based on binocular vision SLAM divides the space vertically, making each module connected and interconnected, and objectively weakening the functional space. The design effect formed by cutting in the overall space but mutually expanding increases the interactivity and openness of the space, which is consistent with the ideal layout effect. However, the other two methods differ greatly from the ideal layout effect, which indicates the possibility of using the research methods to achieve more functions. Compared with the two traditional layout methods, the layout method based on binocular visual SLAM has higher space utilization rate, small overall deviation, good data clustering effect and high data accuracy, which is closer to the ideal layout effect.




(a) Ideal layout

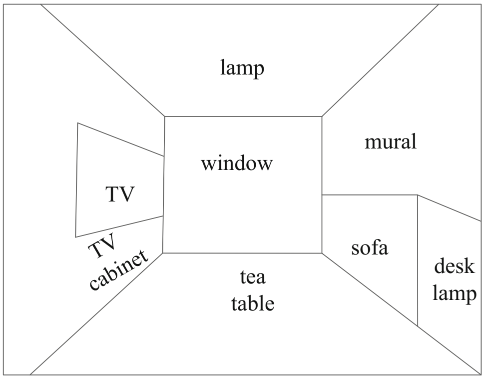



(b) A method of architectural space layout planning based on differential evolution




(c) Spatial layout simulation method based on output structure model

**Fig. 6.** Comparison and analysis of visual effects of three methods of spatial layout




(d) Layout method based on binocular vision SLAM

**Fig. 6.** (*continued*)

## 6 Conclusion

The indoor space layout method of home stay based on binocular vision SLAM is proposed. In complex environments, the distance between feature points of front and back frames and the polar line is used as the judgment standard of feature points, and sparse dynamic feature points in the image are obtained through threshold segmentation. Carry out in-depth research on the multi-level layout of indoor space characteristics. With the support of the original literature, carry out multi-level layout simulation on the indoor space characteristics. After the completion of the overall simulation, verify that the proposed simulation method has higher data accuracy through a number of comparative experiments, and provide a more comfortable living environment for residents. However, the space utilization rate and complexity of this method still have a large room for improvement, and further research and optimization of the comfortable layout method of small indoor space are needed. In the future development, we will further study the interior space layout method of stereo vision, increase its space utilization rate and reduce the computational complexity.

## References

1. Tao, J., Fan, X., Zhou, F.: Point-line feature fusion in stereo visual SLAM algorithm. J. Chin. Comput. Syst. **43**(06), 1191–1196 (2022)
2. Ren, J., Zheng, X., Li, Y., et al.: Mobile visual-inertial SLAM loopclosure algorithm based on novel multi-sensor fusion strategy. Chin. High Technol. Lett. **31**(07), 681–691 (2021)
3. Li, Y., Zhang, L.: Kinect-based simulation of multi-level layout of interior spatial features. Comput. Simulat. **38**(05), 357–361 (2021)
4. Zhou, J.: Aesthetic design of indoor space environment under the penetration of symbiosis theory consciousness. Indust. Construct. **51**(02), 224 (2021)
5. Lulu. Interior space design of traditional settlement homestay. Shanghai Text. Sci. Technol. **49**(07), 79 (2021)

6. Zhang, A., Wang, X.: Design of indoor environment spatial layout optimization system based on Bayesian network. Mod. Electron. Techniq. **44**(18), 114–118 (2021)
7. Cai, Y., Wu, C.: Construction of 3D layout planning model of indoor scene based on laser technology hologram. Laser J. **42**(06), 83–87 (2021)
8. Wang, Y.: Indoor optimal design of building renovation environment space layout based on genetic algorithm. Procedia Comput. Sci. **208**, 539–545 (2022)
9. Thomas, J.J., Mcomber, S., Ning, A.: Wake expansion continuation: multi-modality reduction in the wind farm layout optimization problem. Wind Energy **4**, 25 (2022)
10. Uddin, M.N., Chi, H.L., Wei, H.H., et al.: Influence of interior layouts on occupant energy-saving behaviour in buildings: an integrated approach using Agent-Based Modelling, System Dynamics and Building Information Modelling. Renew. Sustain. Energy Rev. **161**, 112382 (2022)
11. Pei, L., Chen, W., Zhang, Q., et al.: The design and optimization of ship cabin space layout based on crowd simulation. J. Comput.-Aided Design Comput. Graph. **33**(9), 1337–1348 (2022)
12. Wardhani, D.K., Susan, S.: Strategy to reduce the covid-19 transmission through adaptation of Greenship interior space (IS) criteria. IOP Conf. Ser. Earth Environ. Sci. **738**(1), 012072 (2021) (13pp)