Research Article

Keyframes Global Map Establishing Method for Robot Localization through Content-Based Image Matching

Tianyang Cao,[1,2] Haoyuan Cai,[1] Dongming Fang,[1] Hui Huang,[1] and Chang Liu[1]

[1]*State Key Laboratory of Transducer Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*The University of Chinese Academy of Sciences, Beijing 100190, China*

Correspondence should be addressed to Haoyuan Cai; hycai@mail.ie.ac.cn and Chang Liu; tuengineer@qq.com

Self-localization and mapping are important for indoor mobile robot. We report a robust algorithm for map building and subsequent localization especially suited for indoor floor-cleaning robots. Common methods, for example, SLAM, can easily be kidnapped by colliding or disturbed by similar objects. Therefore, keyframes global map establishing method for robot localization in multiple rooms and corridors is needed. Content-based image matching is the core of this method. It is designed for the situation, by establishing keyframes containing both floor and distorted wall images. Image distortion, caused by robot view angle and movement, is analyzed and deduced. And an image matching solution is presented, consisting of extraction of overlap regions of keyframes extraction and overlap region rebuild through subblocks matching. For improving accuracy, ceiling points detecting and mismatching subblocks checking methods are incorporated. This matching method can process environment video effectively. In experiments, less than 5% frames are extracted as keyframes to build global map, which have large space distance and overlap each other. Through this method, robot can localize itself by matching its real-time vision frames with our keyframes map. Even with many similar objects/background in the environment or kidnapping robot, robot localization is achieved with position RMSE <0.5 m.

1. Introduction

The ideal indoor mobile robot should store a global map of the entire indoor space for self-localization, especially in a building including multiple rooms and corridors [1]. It would be much better if this global map can be built by the robot itself during indoor environment studying and without the help of human [2].

For this research field, SLAM (Simultaneous Localization and Mapping) is the commonly employed [3], especially the V-SLAM (Visual Simultaneous Localization and Mapping) [4, 5]. Many latest variants take advantage of new features or 3D information to build navigation map [6, 7], including ORB SLAM [8], dense SLAM [9, 10], semidense SLAM [11], LSD SLAM [12], and CV-SLAM [13, 14]. The CV-SLAM method is specially designed for indoor robot localization [15] and could make good use of ceiling features as navigation map through an upward-looking camera [16]. This technology has been used widely in the Roomba, a product of

iRobot Inc. Similar methods are used in other models made by Dyson, SAMSUNG, and LG.

However, the SLAM method is easy to be disturbed, and the most troublesome problems include kidnap problem and similar objects interference [17]. The kidnap problem occurs frequently when a robot is suddenly involved in a collision, kicked, or intentionally repositioned during operation. The robot cannot localize itself according to video information of previous moments [18]. The similar object interference refers to the fact that a robot is easy to be confused by similar features on similar objects in different places and error fixed itself to the wrong position. An indoor robot that looks mostly at ceilings of indoor space do not detect many features of significance. To make matters worse, there are usually many similar objects in the indoor environment (e.g., air condition outlets, ceiling lamps, and ceiling). Hence established methods are not sufficiently reliable for global indoor positioning.

In order to solve these problems, this paper presents keyframes global map establishing method for robot localization through content-based image matching, involving both ceiling and wall regions. The robot could establish global map by itself and fix its position at any time without the need of its localization information of previous moments and cannot be kidnapped.

Comparing with the common keyframes extraction method (e.g., according to the robot moving time or distance), our keyframes extracted according to image content would be more useful (image content mainly including the objects in the room and their layout and their color distribution). Problems related to losing keyframes or redundancy can be avoided, especially when robot is changing speed or turning suddenly [19, 20].

In this paper, the keyframes global map establishing method is also based on a specially designed content-based image matching method, which can calculate images similarity exactly. The keyframes for each room and corridor can be extracted from robot environment studying video by image content matching. This matching method is designed to avert image distortion interference caused by robot view angle and movement. It is designed based on a distortion model, which is deduced to analyze the distortion features of ceiling and wall region (e.g., wall, door, window, and furniture). According to these features, two image processing methods, including overlap region extraction method and overlap region rebuild method through subblocks matching, are presented. We first *adjust the distortion of vision frames nearly consistent and then calculate the similarity between frames*. In order to adjust distortion more exactly, ceiling points detecting method and mismatching subblocks checking method are incorporated. Through content-based image matching, this paper can take advantage of the information of different objects imaged in different room to extract keyframes for each room from robot vision. *These keyframes have very large space distance, and the whole indoor environment could be described only by 10 s of keyframes.* Keyframes overlap with each other, and their position relationship can be extracted. These keyframes and their global position would compose global map for indoor environment. This content-based image matching method can also be used for robot self-localization. When the robot is moving in rooms or corridors, for the real-time updated robot vision, the most similar keyframe for each robot vision frame can be chosen automatically from this global map and the robot can fix its position precisely.

In the test, the experiment site is composed of two rooms (20-meter-square) and two corridors (11 m). The robot can move automatically in this site to study environment features, and a video about 9.5 min (1710 frames) long had been taken by robot vision during indoor environment studying. Through the map establishing method presented in this paper, 72 keyframes (less than 5% frames) are extracted from this video to build the map of experiment site. The robot can make good use of this map to localize itself in the whole building. It is able to draw robot trajectory exactly, even in the event of kidnapping, with position RMSE less than 0.4 m.

The rest of this paper is organized as follows. Section 2 describes the design of robot vision system and the distortion features of common indoor objects. In Section 3, we discuss content-based image matching method and the keyframes global map establishing method. In Sections 4 and 5, the experiment result and conclusion are presented and discussed.

2. Image Distortion Model and Feature Analysis for Common Indoor Objects

Content-based image matching with both ceilings and walls is one of the best ways for detecting keyframes and building global map, which can extract keyframes from environment video as efficiently as possible to describe the indoor environment and keep the connection between these frames. However, the main problem for content-based image matching is the image distortion. If two frames are taken in different view angle and position, the objects distortions in them are different, and image content similarity between the two frames is difficult to be calculated.

Here we discuss the method for analyzing features of image content distortion through establishing a distortion model. An upward-looking camera (SY012HD wide angle camera, Weixinshijie Technology Co., Ltd, Shenzhen, China) is installed on the wheeled robot (Roomba, R1-L081A, Midea, Suzhou, China) as the robot vision sensor [21], as shown in Figure 1(a), the same as that carrying a CV-SLAM [22, 23].

2.1. The Features of Robot View Angle and Movement for Ceiling and Wall Region. The camera model, which can change the three-dimensional object in the real world to two-dimensional picture, is as follows [24, 25]:

$$
z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & 0 \\ 0 & \dfrac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (1)
$$

where

$$
\begin{aligned}
R &= \begin{bmatrix} \cos R & 0 & -\sin R \\ 0 & 1 & 0 \\ \sin R & 0 & \cos R \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos P & \sin P \\ 0 & -\sin P & \cos P \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} \cos H & \sin H & 0 \\ -\sin H & \cos H & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{2}
$$

is calculated through robot view angles of roll R, pitch P, and heading H. The term $T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$ is calculated through robot movements T_x and T_y and ceiling height T_z, (X, Y, Z) are the coordinates of an object point in three-dimensional real world, and (x, y) are its corresponding points in the robot vision. The terms f, dx, and dy are the constant of the
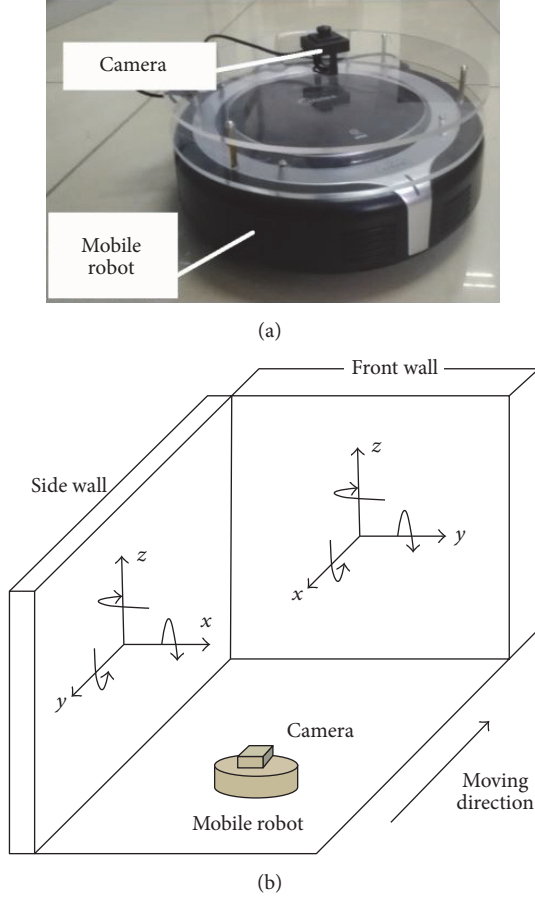
FIGURE 1: The relationship between wall and upward-looking camera. (a) Robot and camera. (b) Relationship between front wall and side wall and camera.

camera: f is the camera focus and dx and dy are the scale factors.

There are two kinds of main indoor objects: ceiling and wall regions. The ceiling is parallel to the floor, and wall regions (e.g., wall, window, door, and furniture) are perpendicular to the floor. For the upward-looking camera on the robot, the view angles for them are different.

When a robot is moving on the floor, its route parallels the ceiling, the heading (given H) is the only robot view angle can be changed, and both roll and pitch are equal to 0.

For the wall region, because they are fixed on the floor, the freedom of robot view angles is also very limited. We analyze each wall separately, by giving each wall an independent coordinate (front/back wall or side wall), as shown in Figure 1(b), with the X-axis perpendicular to the wall, Y-axis parallel to the wall and floor, and Z-axis perpendicular to the floor. The heading H between a wall and the robot is the only view angle which can be changed. As wall is perpendicular to floor and ceiling, it can be seen as a part of ceiling with roll or pitch of 90 deg. For the front wall (or back) of robot, the pitch is 90 deg and roll is 0, as shown in Figure 1(b). For the left wall (or right wall) on the robot, the roll is 90 deg and pitch is 0, as shown in Figure 1(b).

Because of the different view angles, ceiling and wall deform differently in the robot vision. In order to match precisely, this deformation should be revised before image content matching. According to (1), it can also be deduced that robot movement (T_x, T_y) and object position (X, Y, Z) can also contribute to image distortion. Therefore, bringing the robot view angle and movement and object position into (1), the distortion models of ceiling and wall region could be established, and their distortion features can be extracted.

2.2. Ceiling Distortion Model and Feature Extraction. Given the coordinate origin on the ceiling, T_x and T_y show robot movement, and T_z is ceiling height. The heading between ceiling and camera is H, and roll and pitch are 0. Therefore (1) can be transformed as follows:

$$
z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} \dfrac{1}{dx} & 0 & 0 \\ 0 & \dfrac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos H & \sin H & 0 & T_x \\ -\sin H & \cos H & 0 & T_y \\ 0 & 0 & 1 & T_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3}
$$

It can be simplified as follows:

$$
x = \frac{f}{dx \left(T_z + Z \right)} \left(X \cos H + Y \sin H + T_x \right),
$$

$$
y = \frac{f}{dy \left(T_z + Z \right)} \left(-X \sin H + Y \cos H + T_y \right). \tag{4}
$$

The distortion of ceiling only includes rotation and translation, and the shape of ceiling is unchanged. As rotation and translation are affine transform, they can be revised by affine transform too. Through SURF (Speed Up Robust Features), the matched feature points in frames A and B are extracted [26, 27]. Bringing these points coordinate into (5), the heading difference ΔH and movement differences Δt_x and Δt_y between these two frames could be resolved [28]:

$$
\begin{pmatrix} x_B \\ y_B \end{pmatrix} = \begin{pmatrix} \cos \Delta H & \sin \Delta H \\ -\sin \Delta H & \cos \Delta H \end{pmatrix} \begin{pmatrix} x_A \\ y_A \end{pmatrix} + \begin{pmatrix} \Delta t_x \\ \Delta t_y \end{pmatrix}. \tag{5}
$$

Two frames, A and B, can be adjusted to the same robot view angles and camera shooting position after rotation and translation according to ΔH, Δt_x, and Δt_y. Their overlap regions (containing the same objects in two frames) could be extracted. The image similarity would be obtained by comparing the similarity of their overlap regions.

2.3. Wall Distortion Model and Feature Extraction. For the front wall ($P = 90$, $R = 0$) and side wall ($P = 0$, $R = 90$), the heading difference is 90 deg (or −90 deg), but their image distortions are the same and can be deduced as follows.

Given that the robot heading is H when it is taking frame A, for a point (X, Y, Z) on the front wall, its corresponding

point (x_{Aw}, y_{Aw}) on frame A of robot vision is calculated as follows:

$$z \begin{bmatrix} x_{Aw} \\ y_{Aw} \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & 0 \\ 0 & \dfrac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \cos H & \sin H & 0 & T_x \\ -\sin H & \cos H & 0 & T_y \\ 0 & 0 & 1 & T_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{6}$$

And it can be simplified as follows:

$$x_{Aw} = \frac{f}{dx} \frac{X \cos H + Y \sin H + T_x}{X \sin H - Y \cos H + T_z},$$

$$y_{Bw} = \frac{f}{dy} \frac{Z + T_y}{X \sin H - Y \cos H + T_z}. \tag{7}$$

$$z \begin{bmatrix} x_{Aws} \\ y_{Aws} \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & 0 \\ 0 & \dfrac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos H & \sin H & 0 & T_x \\ -\sin H & \cos H & 0 & T_y \\ 0 & 0 & 1 & T_z \end{bmatrix} \begin{bmatrix} X_s \\ Y_s \\ Z_s \\ 1 \end{bmatrix}. \tag{8}$$

And it can be simplified as

$$x_{Aws} = \frac{f}{dx} \frac{-Z_s + T_x}{X_s \cos H + Y_s \sin H + T_z}, \tag{9}$$

$$y_{Aws} = \frac{f}{dy} \frac{-X_s \sin H + Y_s \cos H + T_y}{X_s \cos H + Y_s \sin H + T_z}. \tag{10}$$

Comparing with (7), it means that the x-axis and y-axis terms are exchanged in (9) and (10). Accordingly, x, y are exchanged, and T_x, T_y are exchanged, too. The distortion feature of front wall in the robot vision is the same as that of side wall, and the distortion revise method for the front wall would be the same as that of side wall.

Take the distortion of front wall for example. Given the robot heading and movement for frame A are H, t_x, t_y for frame B which would be marched with frame A, its parameters are $H + \Delta H$, $t_x + \Delta t_x$, and $t_y + \Delta t_y$. According to (8), the object point (X, Y, Z) on frame B is (x_{Bw}, y_{Bw}) and is calculated as follows:

$$x_{Bw} = \frac{f}{dx}$$
$$\cdot \frac{X \cos (H + \Delta H) + Y \sin (H + \Delta H) + T_x + \Delta t_x}{X \sin (H + \Delta H) - Y \cos (H + \Delta H) + T_z}, \tag{11}$$

For the point (X_s, Y_s, Z_s) on the side wall, its corresponding point (x_{Aws}, y_{Aws}) on the frame A of robot vision is calculated as follows:

$$y_{Bw} = \frac{f}{dy} \frac{Z + T_y + \Delta t_y}{X \sin (H + \Delta H) - Y \cos (H + \Delta H) + T_z}. \tag{12}$$

For the ceiling and wall region in frames A and B, the deformations of their ceiling images, including rotation and translation, can be adjusted similarly through (5). However, the wall region in frame A, after rotation and translation, would be transformed differently from that of frame B:

$$\begin{pmatrix} x_{AB} \\ y_{AB} \end{pmatrix}$$
$$= \begin{pmatrix} \cos \Delta H & \sin \Delta H \\ -\sin \Delta H & \cos \Delta H \end{pmatrix} \begin{pmatrix} \dfrac{f}{dx} \dfrac{X \cos H + Y \sin H + T_x}{X \sin H - Y \cos H + T_z} \\ \dfrac{f}{dy} \dfrac{Z + T_y}{X \sin H - Y \cos H + T_z} \end{pmatrix} \tag{13}$$
$$+ \begin{pmatrix} \Delta t_x \\ \Delta t_y \end{pmatrix},$$

where (x_{AB}, y_{AB}) is point in frame A processed by (5) and can be simplified as

$$x_{AB} = \frac{f}{dx} \frac{X \cos H \cos \Delta H + Y \sin H \cos \Delta H + T_x \cos \Delta H + Z \sin \Delta H + T_y \sin \Delta H}{X \sin H - Y \cos H + T_z} + \Delta t_x,$$

$$y_{AB} = \frac{f}{dy} \frac{-X \cos H \sin \Delta H - Y \sin H \sin \Delta H + Z \cos \Delta H - T_x \sin \Delta H + T_y \cos \Delta H}{X \sin H - Y \cos H + T_z} + \Delta t_y. \tag{14}$$

If this point is translated by an amount (S_x, S_y),

$$
\begin{aligned}
S_x = {}& \cos \Delta H \frac{f}{dx} \frac{-T_x}{X \sin H - Y \cos H + T_z} \\
& + \sin \Delta H \frac{f}{dx} \frac{-X \sin H + Y \cos H - Z - T_y}{X \sin H - Y \cos H + T_z} \\
& + \frac{f}{dx} \frac{\Delta t_x}{X \sin H - Y \cos H + T_z} - \Delta t_x, \\
S_y = {}& \sin \Delta H \frac{f}{dy} \frac{X \cos H + Y \sin H + T_x}{X \sin H - Y \cos H + T_z} \\
& + \cos \Delta H \frac{f}{dy} \frac{-Z - T_y}{X \sin H - Y \cos H + T_z} \\
& + \frac{f}{dy} \frac{\Delta t_y + Z + T_y}{X \sin H - Y \cos H + T_z} - \Delta t_y.
\end{aligned}
\tag{15}
$$

The transform result is

$$
\begin{aligned}
& x_{AB} + S_x \\
& = \frac{f}{dx} \frac{X \cos (H + \Delta H) + Y \sin (H + \Delta H) + T_x + \Delta t_x}{X \sin H - Y \cos H + T_z} \\
& y_{AB} + S_y = \frac{f}{dy} \frac{Z + T_y + \Delta t_y}{X \sin H - Y \cos H + T_z}.
\end{aligned}
\tag{16}
$$

Comparing with (10) and (11), the transformed frame A is much closer to frame B than the original frame A. The similarity between frames A and B can be calculated more exactly through matching with transformed frame A compared to original frame A.

But there is still some difference between transformed frame A and frame B. On one hand, their denominators are different, which are $X \sin H - Y \cos H + T_z$ and $X \sin(H + \Delta H) - Y \cos(H + \Delta H) + T_z$. On the other hand, some terms in (15) are unknown (e.g., point coordinate (X, Y, Z) in the room space), as the terms (S_x, S_y) cannot be calculated through the two equations. In order to neutralize these disadvantages, overlap region extraction method and subblock matching method would be presented, and (S_x, S_y) can be resolved through subblock matching. The term subblock is defined as follows. An image is divided into many small blocks which are equal sized, and these blocks are named as subblock in this paper. Equation (5) would also be accomplished through overlap region extraction method designed in this paper.

If frames A and B contain the same image content, their transform results would be very similar after being processed by (5) and (15) and very useful for image matching and similarity analysis. For the same ceiling regions in frames A and B, the correlation coefficient between them would be very large. For the same wall regions in two frames, as containing the same objects, the number of correct matching subblocks would be also very large. This paper would combine them together to analyze the similarity between frames A and B.

3. Match Method Design

Content-based image matching method would be designed firstly and then the keyframes global map can be established through this matching method.

3.1. Content-Based Image Matching through Overlap Region Extraction and Subblocks Rebuild. The content-based image matching method is designed according to the features of image distortion and consists of three parts: image overlap region extraction, overlap region rebuild through subblock matching, and, lastly, image content similarity calculating, as shown in Figure 2. For frames A and B, this method would adjust the distortion of frame A similar to that of frame B and then calculate their degree of similarity. In order to adjust more exactly, ceiling points detecting method is designed to improve the accuracy of overlap region extraction, and mismatched subblocks checking method is deigned to rebuilt the overlap region more exactly.

3.1.1. Image Overlap Region Extraction. For frames A and B, which are taken by robot at different positions, only parts of them might overlap. Content-based image matching would focus on analyzing the image similarity in this overlap region.

In order to extract this overlap region, frames A and B would be adjusted to the same view angle and camera shooting position according to their matched feature points extracted through the SURF method. If all the feature points are on the ceiling, the translation Δt_x, Δt_y, and rotation ΔH between the two frames can be calculated directly according to (5). Through rotation and translation, the same objects in frames A and B would be adjusted to the same position, and this image processing progress is shown in Figure 3. The overlap regions can be extracted effectively from these two adjusted frames.

The image procession result is shown in Figure 4. The rotation and translation of frames A and B are shown in Figures 4(c) and 4(d), and the same objects are overlapped, as shown in Figure 4(e). The overlap region mask is the pixels covered by the two image, as shown in Figure 4(f). The overlap regions in frames A and B can be extracted through this mask.

If the feature points were extracted from wall regions, they would interfere with the calculation result of Δt_x, Δt_y, and ΔH. In order to delete the points not on the ceiling (including wall, furniture, window, and door), this paper designs ceiling points detecting method according the feature of ceiling, which are of equal height.

For the ceiling points a and b, given that the origin of coordinate is on the ceiling and ceiling points $Z = 0$, according to (4), the length of their connect lines in frames A and B (for points "a" and "b" in Figures 5(a) and 5(b), their connect line is "line ab") can be calculated as follows:

$$
\begin{aligned}
d_{A,ab} = {}& \left(\left(x_{A,a} - x_{A,b} \right)^2 + \left(y_{A,a} - y_{A,b} \right)^2 \right)^{1/2} \\
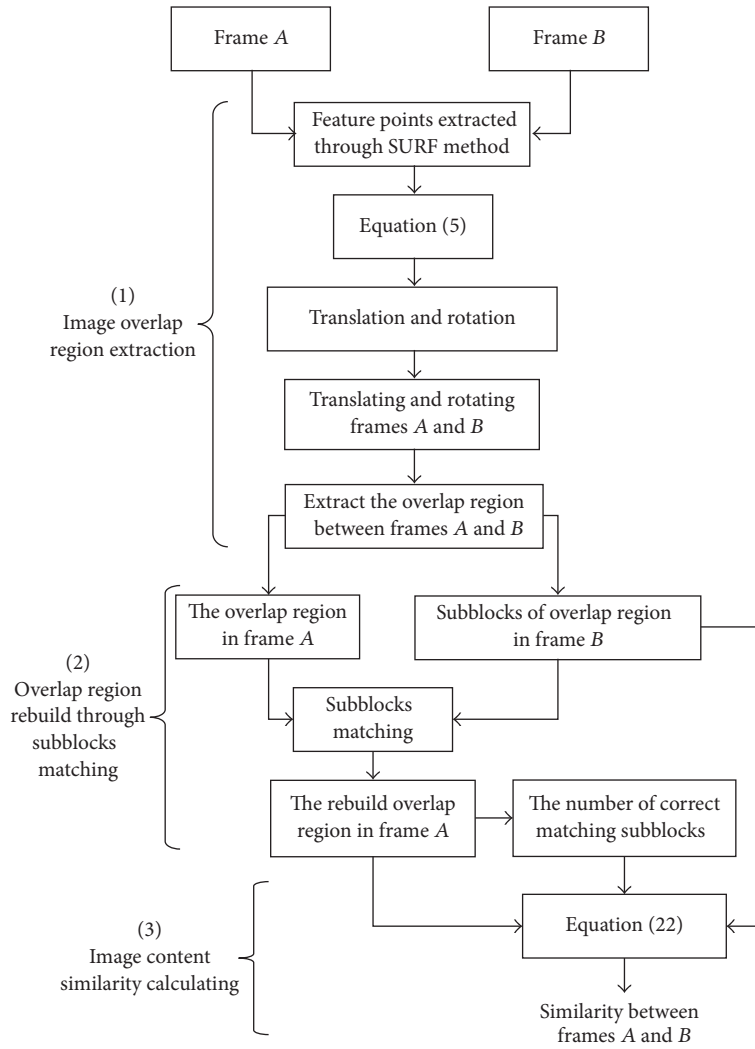= {}& \left(\left(\frac{f}{dx T_z} \right)^2 \left(\left(X_a \cos H + Y_a \sin H + T_{x,A} \right) \right. \right.
\end{aligned}
$$

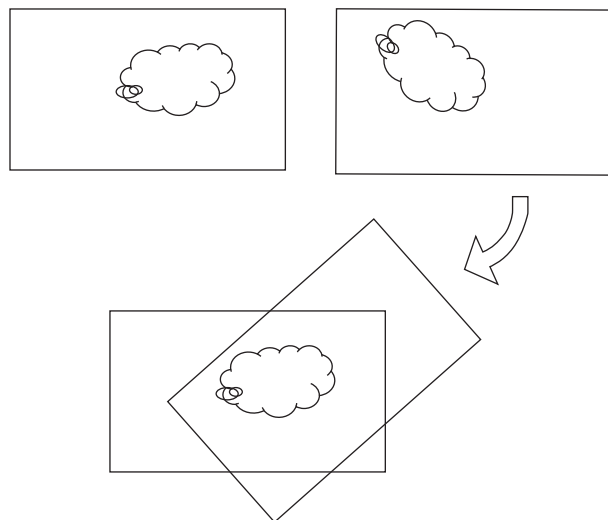FIGURE 2: The image processing of content-based image matching.

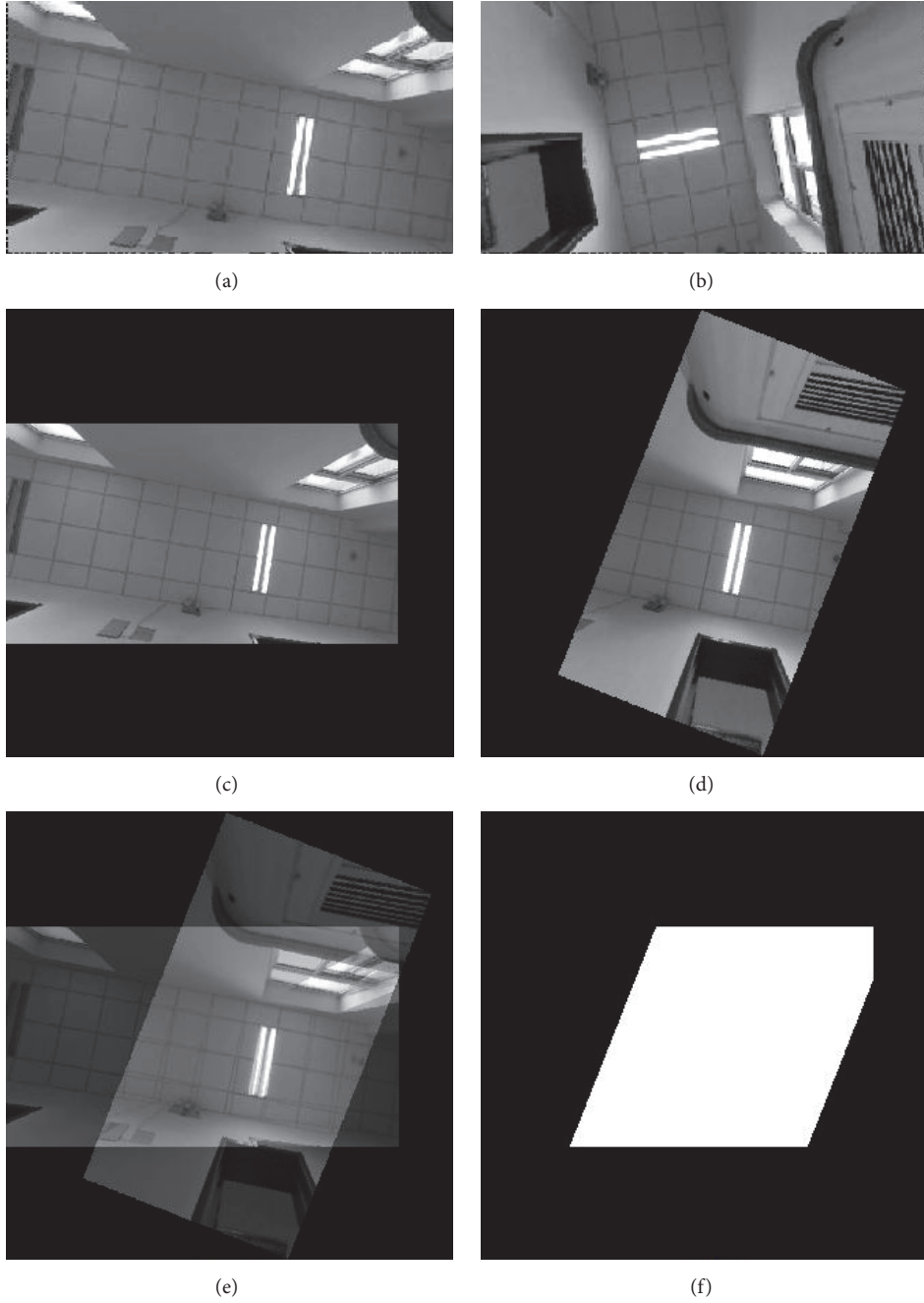FIGURE 3: The image processing of rotation and translation.

(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 4: The overlap region extraction progress. (a) Frame *A*, (b) Frame *B*, (c) the translation result, (d) the rotation result, (e) the overlap between two frames, and (f) the overlap region mask.

$$
\begin{aligned}
&- \left(X_b \cos H + Y_b \sin H + T_{x,A} \right) \right)^2 - \left(\frac{f}{dy T_y} \right)^2 \\
&\cdot \left(\left(-X_a \sin H + Y_b \cos H + T_{y,A} \right) - \left(-X_a \sin H \right. \right. \\
&\left. \left. + Y_b \cos H + T_{y,A} \right) \right)^2 \right)^{1/2} = \left(\left(\frac{f}{dx T_z} \right)^2 \left(X_a \cos H \right. \right. \\
&+ Y_a \sin H - X_b \cos H - Y_b \sin H \right)^2 + \left(\frac{f}{dy T_y} \right)^2
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \left(-X_a \sin H + Y_b \cos H + X_a \sin H - Y_b \right. \\
&\left. \cdot \cos H \right)^2 \right)^{1/2}, \\
&d_{B,ab} = \left(\left(x_{B,a} - x_{B,b} \right)^2 + \left(y_{B,a} - y_{B,b} \right)^2 \right)^{1/2} \\
&= \left(\left(\frac{f}{dx T_z} \right)^2 \left(\left(X_a \cos \left(H + \Delta H \right) \right. \right. \right.
\end{aligned}
$$
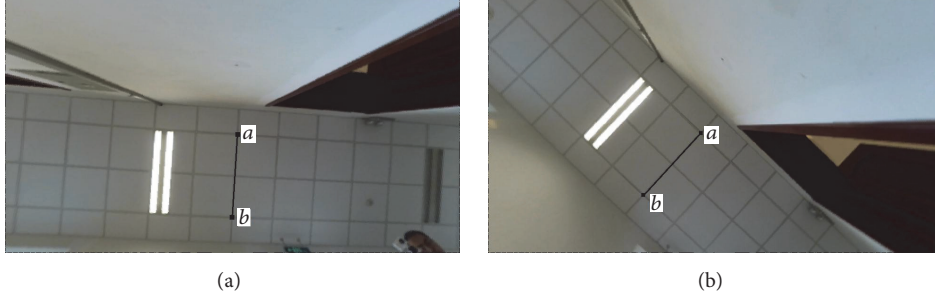
(a)　　　　　　　　　　　　　　　　　　　　　(b)

FIGURE 5: The connect lines on the ceiling. (a) The connect lines on the original ceiling. (b) The connect lines on the ceiling after rotation and translation.

$$+ Y_a \sin(H + \Delta H) + T_{x,B}) - (X_b \cos(H + \Delta H)$$

$$+ Y_b \sin(H + \Delta H) + T_{x,B}))^2 - \left(\frac{f}{dyT_y}\right)^2$$

$$\cdot \left((-X_a \sin(H + \Delta H) + Y_b \cos(H + \Delta H) + T_{y,B}\right)$$

$$- \left(-X_a \sin(H + \Delta H) + Y_b \cos(H + \Delta H)\right.$$

$$+ T_{y,B}))^2\right)^{1/2} = \left(\left(\frac{f}{dxT_z}\right)^2 (X_a \cos H + Y_a \sin H\right.$$

$$- X_b \cos H - Y_b \sin H)^2 + \left(\frac{f}{dyT_y}\right)^2 (-X_a \sin H$$

$$+ Y_b \cos H + X_a \sin H - Y_b \cos H)^2\right)^{1/2} = d_{A,ab},$$

$$(17)$$

where (X_a, Y_a, Z_a) and (X_b, Y_b, Z_b) are the coordinates of points a and b in the real world, $(x_{A,a}, y_{A,a})$ and $(x_{A,b}, y_{A,b})$ are the coordinates of points a and b in frame A, and $(x_{B,a}, y_{B,a})$ and $(x_{B,b}, y_{B,b})$ are the coordinates of points a and b in frame B. $(T_{x,A}, T_{y,A})$ and $(T_{x,B}, T_{y,B})$ are the robot movement for frames A and B. As points a and b are equal in height, it can be deduced by (17) that $d_{A,ab} = d_{B,ab}$ and the length of connect lines of the same ceiling points in the two images are unchanged, as shown in Figure 5.

However, if one of the feature points b_w is on the wall (given height is Z_{bw}) and not on the ceiling, its height would differ from that of ceiling (T_z). In this case, the length of the connect line is

$$d_{A,abw} = \left((x_{A,a} - x_{A,bw})^2 + (y_{A,a} - y_{A,bw})^2\right)^{1/2}$$

$$= \left(\left(\frac{f}{dxT_z}(X_a \cos H + Y_a \sin H + T_{x,A}) - \frac{f}{dx}\right.\right.$$

$$\cdot \frac{X_{bw} \cos H + Y_{bw} \sin H + T_{x,A}}{X_{bw} \sin H - Y_{bw} \cos H + T_z}\right)^2$$

$$+ \left(\frac{f}{dyT_z}(-X_a \sin H + Y_{bw} \cos H + T_{y,A}) - \frac{f}{dy}\right.$$

$$\cdot \frac{Z_{bw} + T_{y,A}}{X_a \sin H - Y_{bw} \cos H + T_z}\right)\right)^{1/2},$$

$$(18)$$

$$d_{B,abw} = \left((x_{B,a} - x_{B,bw})^2 + (y_{B,a} - y_{B,bw})^2\right)^{1/2}$$

$$= \left(\left(\frac{f}{dxT_z}(X_a \cos(H + \Delta H) + Y_a \sin(H + \Delta H)\right.\right.$$

$$+ T_{x,B}) - \frac{f}{dx}$$

$$\cdot \frac{X_{bw} \cos(H + \Delta H) + Y_{bw} \sin(H + \Delta H) + T_{x,B}}{X_{bw} \sin(H + \Delta H) - Y_{bw} \cos(H + \Delta H) + T_z}\right)^2$$

$$+ \left(\frac{f}{dyT_z}(-X_a \sin(H + \Delta H) + Y_{bw} \cos(H + \Delta H)\right.$$

$$+ T_{y,B}) - \frac{f}{dy}$$

$$\cdot \frac{Z_{bw} + T_{y,B}}{X_a \sin(H + \Delta H) - Y_{bw} \cos(H + \Delta H) + T_z}\right)\right)^{1/2}$$

$$\neq d_{A,abw},$$

$$(19)$$

where $(x_{A,bw}, y_{A,bw})$ and $(x_{B,bw}, y_{B,bw})$ are coordinates of wall point b_w in frames A and B and $d_{A,abw}$ and $d_{B,abw}$ are the length of connect lines ab_w in these two frames, as ΔH is in the denominator of (19), and (18) and (19) are not equal, and $d_{A,abw} \neq d_{B,abw}$. Therefore, the points on the wall can be deleted effectively by comparing the length of connect lines of feature points in the two frames, and the feature points on the ceiling can be reserved. The image process is shown in the Figure 6.

Therefore the points on the ceiling can be extracted by comparing the length of their connect lines in frames A and B. And these points can make the calculating result of (5) more exact and extracting overlap regions more effective. In the overlap regions, as rotation and translation have been
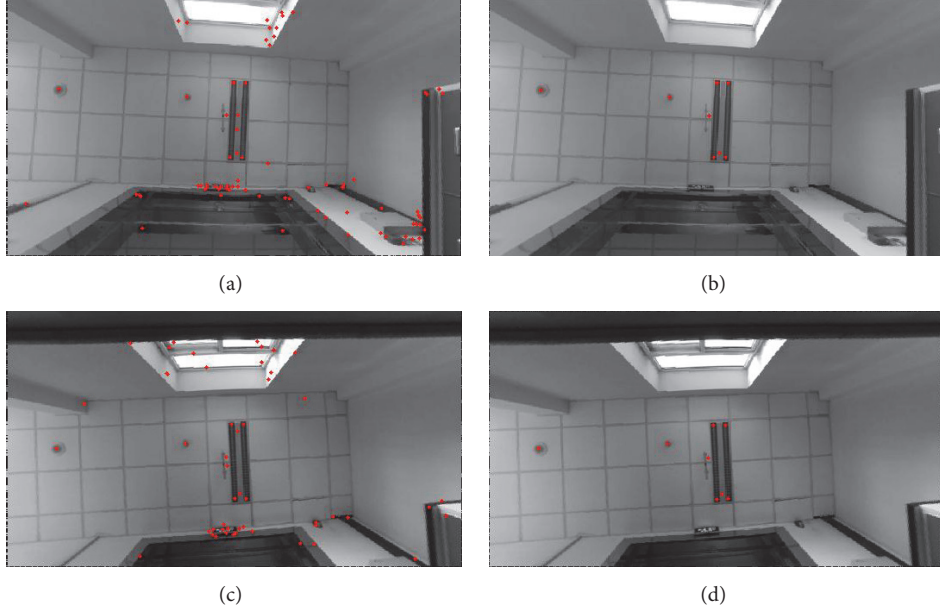
(a)

(b)

(c)

(d)

FIGURE 6: The image processing of ceiling feature point detection (red points are the feature points). (a) Original feature points in frame A, (b) ceiling points detecting result in frame A, (c) original feature points in frame B, and (d) ceiling points detecting result in frame B.

completed, the difference of ceiling image between frames A and B caused by robot rotation and translation can be deleted. However, for the wall image distortion difference between these two frames, the distortion adjustment would be completed in the next section.

3.1.2. Overlap Region Rebuild through Subblocks Matching. For the points on the ceiling region, it is difficult to calculate the translation result (S_x, S_y) of each point in the wall directly through (15), as robot cannot measure each point coordinates (X, Y, Z) in the room through its monocular camera. This paper presents subblocks matching method to calculate the translation value.

As the distorted object in the frame is still an entity in the image, the translation values (S_x, S_y) of distorted object points in a small image region are nearly equal. Therefore this region can be translated as smaller units according to the average translation value $(\overline{S_x}, \overline{S_y})$ of this region. The overlap region in frame A can be divided into many small subblocks firstly and then match the overlap region in frame B through SAD method (Sum of Absolute Difference) to get the average translation value $(\overline{S_x}, \overline{S_y})$ of each block. Therefore the distortion of overlap region in frame A can be adjusted similar to that of frame B, as shown in Figure 7.

The SAD matching method is as follows:

$$
\begin{aligned}
&v(i, j) \\
&= \arg\min_{\substack{i \in [1,L] \\ j \in [1,L]}} \sum_{s=1}^{M} \sum_{t=1}^{N} |B_L(i+s-1, j+t-1) - A_T(s,t)|,
\end{aligned}
\quad (20)
$$

where A_T is one of the subblocks in the overlap region in frame A, and size is $M \times N$, and B_L is the overlap region in frame B, and size is $L \times D$. Traversal B_L, the terms (i, j),

which can make (20) smallest, is the most suitable rebuild position of A_T in B_L. This (i, j) is the $(\overline{S_x}, \overline{S_y})$ of subblock A_T. Through this method, every subblock in frame A can find its rebuild position and get its translation $(\overline{S_x}, \overline{S_y})$, and the rebuild overlap region of frame A is similar to that of frame B.

Considering that there might be some mismatched subblocks, which are caused by similar objects and mismatched to wrong position, this paper presents mismatched subblocks checking method to check the wrong $(\overline{S_x}, \overline{S_y})$.

As $(\overline{S_x}, \overline{S_y})$ is the average of (S_x, S_y) of all the points in a subblock; the performance of $(\overline{S_x}, \overline{S_y})$ is very similar to that of (S_x, S_y). For (S_x, S_y), (15) can changed into matrix from:

$$
\begin{aligned}
\begin{pmatrix} S_x \\ S_y \end{pmatrix} &= \frac{f}{X \sin H - Y \cos H + T_z} \\
&\cdot \begin{pmatrix} \dfrac{1}{dx} & 0 \\ 0 & \dfrac{1}{dy} \end{pmatrix} \begin{pmatrix} \cos(-H) & \sin(-H) \\ -\sin(-H) & \cos(-H) \end{pmatrix} \begin{pmatrix} Y \\ X \end{pmatrix} \\
&\cdot \sin \Delta H - \frac{f}{X \sin H - Y \cos H + T_z} \\
&\cdot \begin{pmatrix} \dfrac{1}{dx} & 0 \\ 0 & \dfrac{1}{dy} \end{pmatrix} \begin{pmatrix} \cos \Delta H & \sin \Delta H \\ -\sin \Delta H & \cos \Delta H \end{pmatrix} \begin{pmatrix} T_x \\ Z + T_y \end{pmatrix} \\
&+ \frac{f}{X \sin H - Y \cos H + T_z} \\
&\cdot \begin{pmatrix} \dfrac{1}{dx} & 0 \\ 0 & \dfrac{1}{dy} \end{pmatrix} \begin{pmatrix} \Delta t_x \\ \Delta t_y + T_y + Z \end{pmatrix} - \begin{pmatrix} \Delta t_x \\ \Delta t_y \end{pmatrix}.
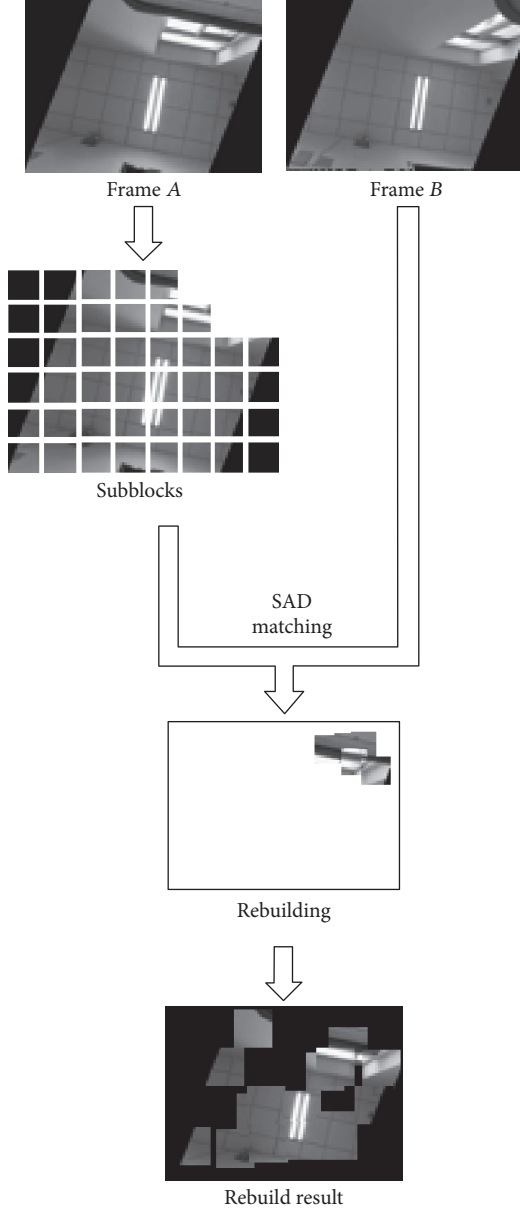\end{aligned}
\quad (21)
$$

Figure 7: The overlap region rebuilding process.

It can be seen from (21) that the translation (S_x, S_y) is constituted of two parts: the projection of heading difference ΔH produced by the first and second term and the translation produced by the third and fourth term. The core of first term is $\left(\begin{smallmatrix} \cos(-H) & \sin(-H) \\ -\sin(-H) & \cos(-H) \end{smallmatrix} \right) \left(\begin{smallmatrix} Y \\ X \end{smallmatrix} \right) \sin \Delta H$, and it is the function of $\sin \Delta H$, which is the projection of ΔH. The core of second term is $\left(\begin{smallmatrix} \cos \Delta H & \sin \Delta H \\ -\sin \Delta H & \cos \Delta H \end{smallmatrix} \right) \left(\begin{smallmatrix} T_x \\ Z+T_y \end{smallmatrix} \right)$, the same as the rotation part in (5); it shows frame A rotating ΔH during overlap region extraction. So the subtraction result of the first two terms in (21) can extract the projection of heading difference ΔH between frame A and B and excludes the rotation caused by overlap region extraction. The core of third term is $\left(\begin{smallmatrix} \Delta t_x \\ \Delta t_y + T_y + Z \end{smallmatrix} \right)$; it is the translation difference between frames A and B caused by wall object height and robot movement difference between frames A and B. For the fourth

term, the same as the translation part in (5), it shows frame A translation result during overlap region extraction. So the subtraction result of the last two terms in (21) can extract the translation difference between frames A and B and exclude the translation of overlap region extraction.

Therefore the wrong $(\overline{S_x}, \overline{S_y})$ of mismatched subblocks can be picked out according to the rotation difference and translation difference between frames A and B after overlap region extraction, as $(f/(X \sin H - Y \cos H + T_z)) \left(\begin{smallmatrix} 1/dx & 0 \\ 0 & 1/dy \end{smallmatrix} \right)$ is the projection ratio between real world and camera image, and (X, Y, Z) can be substituted by its corresponding subblock coordinate (x, y). The thresholds $(\Delta P_x, \Delta P_y)$ for $(\overline{S_x}, \overline{S_y})$ of subblock in (x, y) can be calculated by affine model:

$$\begin{pmatrix} \Delta P_x \\ \Delta P_y \end{pmatrix} = \begin{pmatrix} \cos \Delta H_T & \sin \Delta H_T \\ -\sin \Delta H_T & \cos \Delta H_T \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \Delta x_T \\ \Delta y_T \end{pmatrix}, \quad (22)$$

where ΔH_T, Δx_T, and Δy_T are the thresholds of rotation difference and translation difference between frames A and B after overlap region extraction. In order to delete the wrong mismatched subblocks effectively, ΔH_T, Δx_T, and Δy_T are smaller than the maximums of rotation difference and translation difference.

If $(\overline{S_x}, \overline{S_y})$ of a subblock is larger than its threshold $(\Delta P_x, \Delta P_y)$, it can be deduced that this block is mismatched and should be deleted. Through mismatched block detection, the correct content-based image matching result can be gotten. The matching result for two similar frames is shown in Figure 8, and the matching result for the dissimilar frames is shown in Figure 9. This subblock matching method can match similar frames and evaluate their similarity effectively.

3.1.3. Image Content Similarity Calculating. If the content of frames A and B is similar, the number of correct matched subblocks is very great, and the correlation coefficients between rebuild overlap region in frame A and overlap region in frame B are also very large [29, 30], and the similarity S_k between two frames is the product of subblock number and correlation coefficients:

$$S_{AB} = N_k$$
$$\cdot \frac{\sum_{x,y} \left(C_A(x, y) - \overline{C}_A \right) \left(C_B(x, y) - \overline{C}_B \right)}{\sqrt{\sum_{x,y} \left(C_A(x, y) - \overline{C}_A \right)^2 \cdot \sum_{x,y} \left(C_B(x, y) - \overline{C}_B \right)^2}}, \quad (23)$$

where N_k is the number of matched subblock, $C_A(x, y)$ is the pixel value in rebuild overlap region in frame A, $C_B(x, y)$ is the pixel value in overlap region in frame B, and $\overline{C}_A$ and $\overline{C}_B$ are the average image pixel value.

3.2. Keyframes Global Map Establishing through Content-Based Image Matching. Before establishing map, robot would move in the building automatically to take a video of indoor environment through its camera firstly. And then, through this content-based image matching method, robot can extract keyframes sequence from vision video to build global map of indoor environment by itself.

The first keyframe is the first frame of vision video. And the other keyframes extraction progress is as follows:
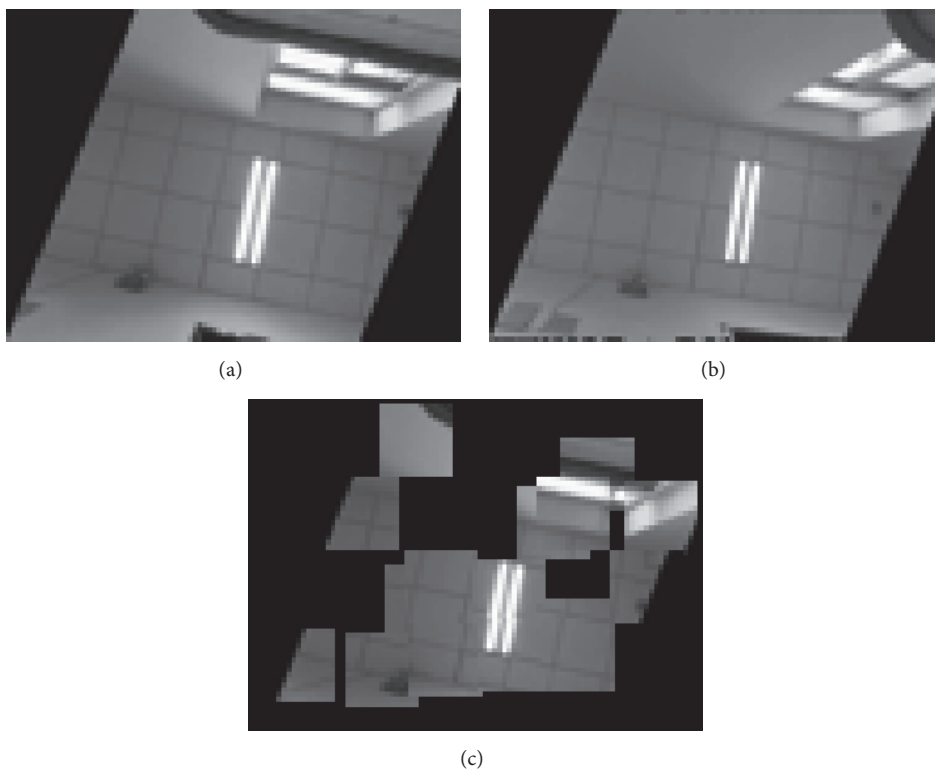
(a)

(b)

(c)

FIGURE 8: The rebuild overlap region by two similar frames. (a) The overlap region in frame *A*, (b) the overlap region in frame *B*, and (c) the rebuild overlap region in frame *A*.
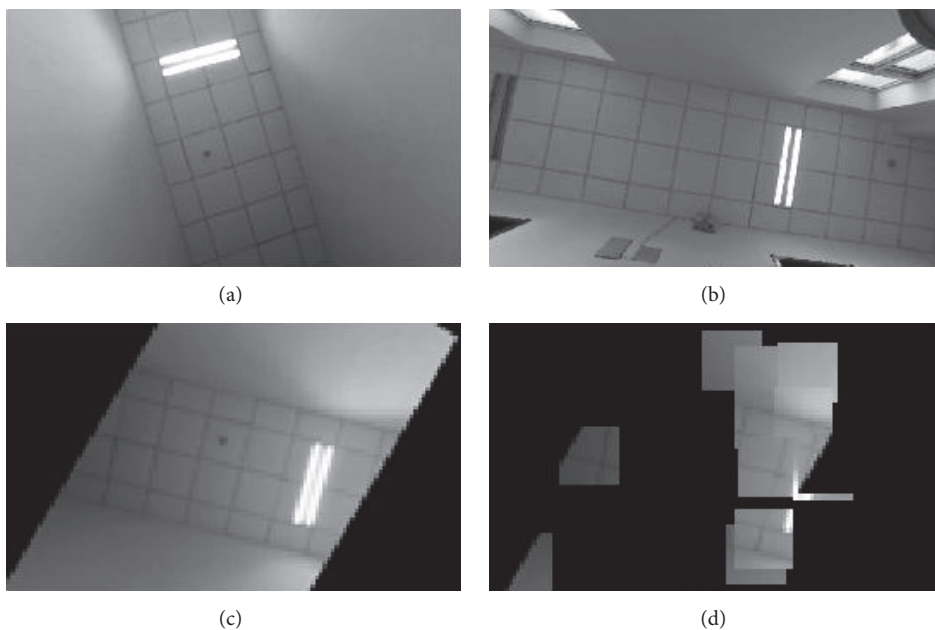

(a)

(b)

(c)

(d)

FIGURE 9: The rebuild overlap region by two dissimilar frames. (a) Frame *A*, (b) frame *B*, (c) the overlap region in frame *A*, and (d) the rebuild overlap region in frame *A*.
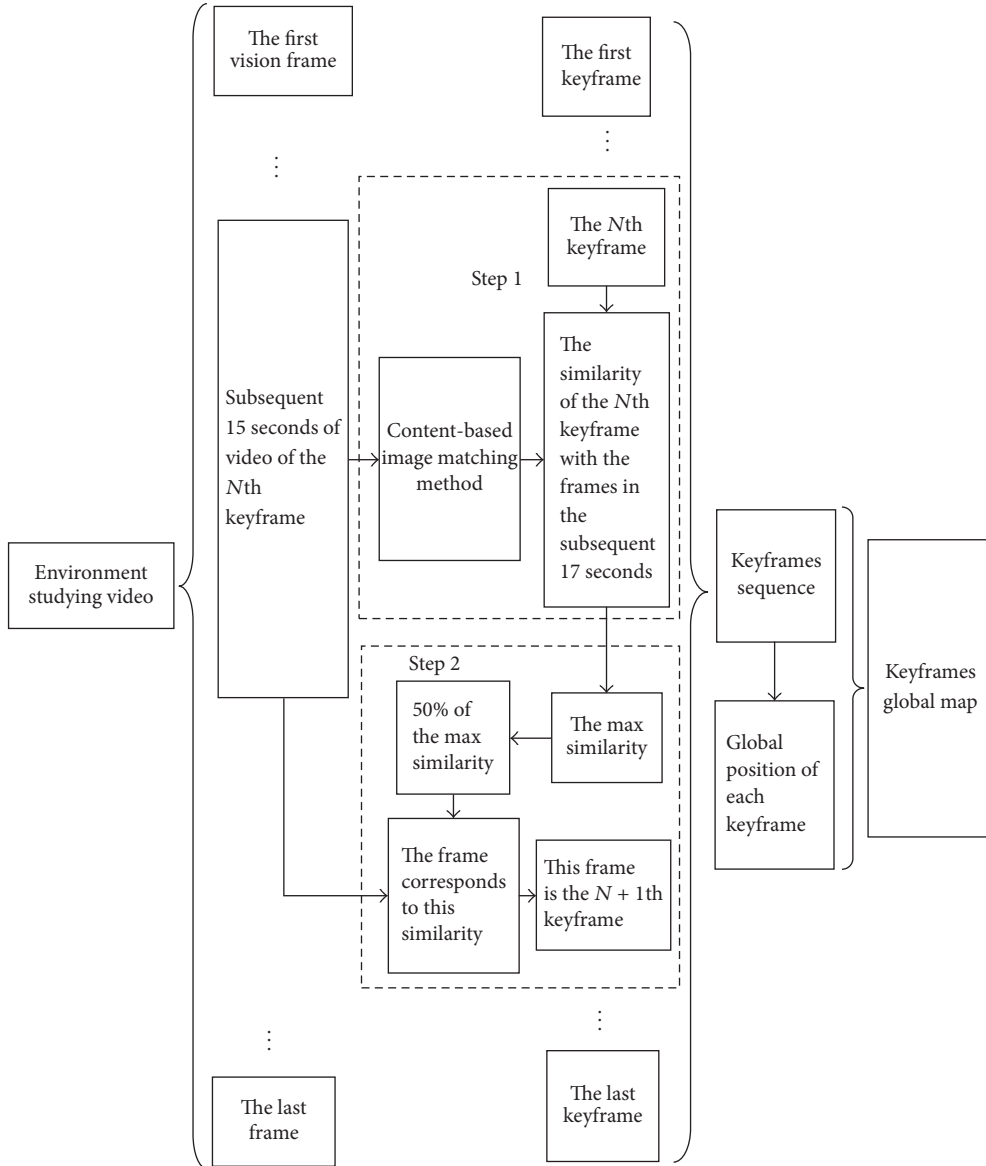
FIGURE 10: The image processing of keyframes global establishing.

Step 1. For the Nth keyframe, its similarity with subsequent 50 frames video (about 17 seconds) is calculated by the robot.

Step 2. The max similarity in the subsequent 50 frames is found out firstly, and it is correspondent to the frame whose space position is nearest to the Nth keyframe. And, then, the frame whose similarity is 50% of the max similarity can be extracted as the $N + 1$th keyframe. If all the subsequent 50 frames are larger than 50% of the max similarity, the 50th frame is the $N + 1$th keyframe. As there is 50% similarity between the N and $N + 1$ keyframes, they can both interval long space distance and overlap each other.

Repeating Steps 1 and 2 and processing all the frames of indoor environment video, keyframes sequence can be extracted, as shown in Figure 10.

This global map would be established by these keyframes, and consists of two parts: keyframes sequence and the global position of each keyframe. As the keyframes can partially overlap each other, extracting the feature points in the overlap region between neighboring keyframes and bringing them into (5), the relative position relationship of these keyframes, including heading difference $H_{N+1,N}$ and position difference $x_{N+1,N}$, $y_{N+1,N}$, would be resolved. Then, bringing these relation positions into (24), the global position of each keyframe can be gotten:

$$
\begin{pmatrix} x_{N+1} \\ y_{N+1} \end{pmatrix} = \begin{pmatrix} \cos H_{N+1,N} & \sin H_{N+1,N} \\ -\sin H_{N+1,N} & \cos H_{N+1,N} \end{pmatrix} \begin{pmatrix} x_{N+1,N} \\ y_{N+1,N} \end{pmatrix} + \begin{pmatrix} x_N \\ y_N \end{pmatrix},
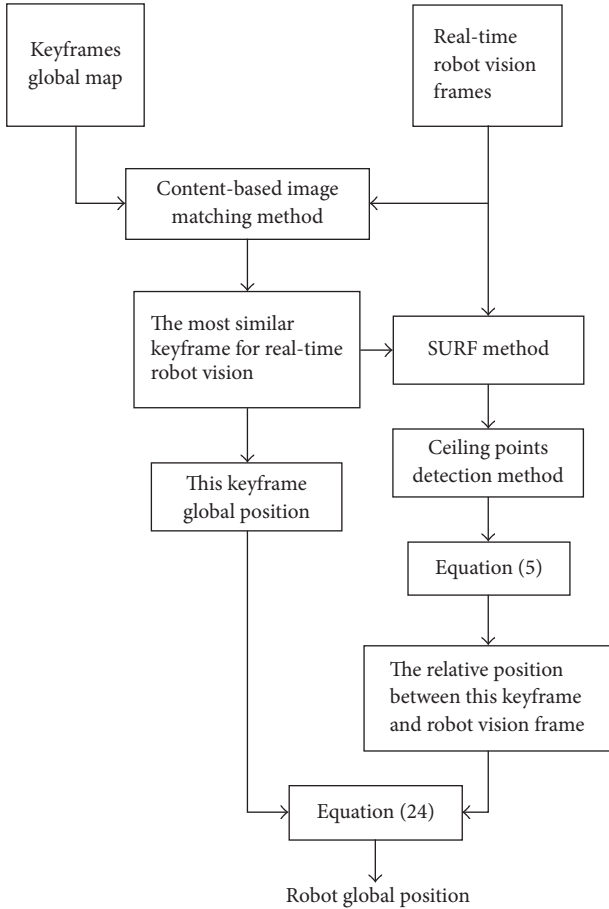\tag{24}
$$

Figure 11: The image processing of robot localization.

Table 1: Position RMSE comparison (m).

	Method in this paper	ORB SLAM	CV-SLAM
Corridor 1 (left)	0.144	0.086	0.123
Corridor 2 (right)	0.159	0.117	0.245
Room 1 (left)	0.468	2.026	0.358
Room 2 (right)	0.483	2.665	0.546

4. Results of Experiments

This keyframes global map establishing method for robot localization through content-based image matching had been tested, including two parts: map building and localization in large indoor environment and the test for robot localization under kidnap condition.

4.1. The Experiment of Map Building and Localization in Large Indoor Environment. The experiment site is composed of two rooms (20-meter-square) and two corridors (11 m). The experiment progress is shown in Figure 12. During robot indoor environment studying, a video about 9.5-minute long (1710 frames) had been taken by robot vision during robot indoor environment studying.

Total 72 keyframes (less than 5%) are extracted by our method from these video frames to build map of experiment site. Their global position resolving result is shown in Figure 13. The environment of the experiment site, including ceiling and wall region (wall, window, door, and furniture), can also be described by this keyframe sequence and their global position, and the position RMSE is less than 0.3 m, as shown in Figure 14.

The robot can make good use of this map to localize itself in the whole building and draw its route in the different room and corridor effectively, as shown in Figure 15.

In order to evaluate the localization precision of this map building method, the corners of tile floor are taken as ground signs, and the air conditioning port and ceiling lamp on the ceiling are taken as ceiling signs. And the localization RMSE between robot localization result and these signs is less than 0.5 m, as shown in Table 1.

Table 1 also shows the comparison between our method and ORB SLAM. The algorithm architecture of our method is similar to that of ORB SLAM, but this paper uses image content matching taking place of feature points matching. Through image content matching, the robot can pick up the keyframe from global map more precisely, which is most similar to its real-time robot vision, and is seldom disturbed by similar objects in the indoor environment. As the experiment site (Figure 14) includes four parts, two corridors and two rooms, the comparison between our method and ORB SLAM is also divided into four parts, as shown in Table 1.

It can be seen from Figure 14 that there are many similar objects in the experiment site, and this would be a serious test for our method and ORB SLAM.

When robot is moving the in the two rooms, the experiment result of our method is better than that of ORB SLAM. This is due to the fact that there are many similar objects in the two rooms, such as the air condition outlet and ceiling

where (x_N, y_N) and (x_{N+1}, y_{N+1}) are the global position of the Nth and $N + 1$th keyframe.

Through this content-based image matching method, the space distance between each keyframe could be very large, and the number of keyframes is very less. Robot would resolve its position quickly by matching with these keyframes. In the experiment, total 1710 frames are taken by the robot during indoor environment studying, and 72 keyframes are extracted by this method.

Through this global map, the robot can localize itself in real-time, and the image processing of robot localization is shown in Figure 11. When robot moving in the indoor environment to render service to human, this content-based image matching method could be used to match real-time robot vision frames with the keyframes sequence in the map and find out the most similar keyframe for each robot vision frame. The same as the global map establishing progress, feature points between robot vision and this keyframe can be extracted, and their relative position can be resolved through (5). And the global position of each robot vision frame, which is also the robot global position, can be resolved through (24), the same as resolving progress of the $N + 1$th keyframe global position.

(a)   (b)

FIGURE 12: The experiment site. (a) The robot moving in the experiment site. (b) The ceiling in the robot vision.
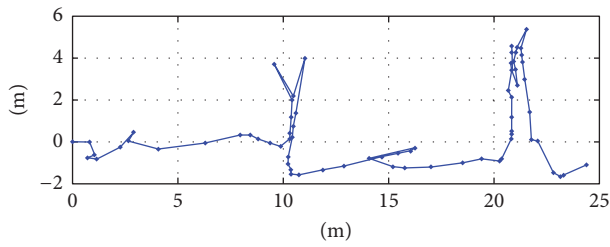

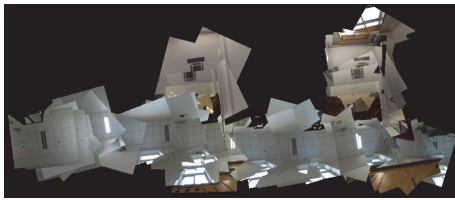FIGURE 13: The global position relationship of keyframes sequence.

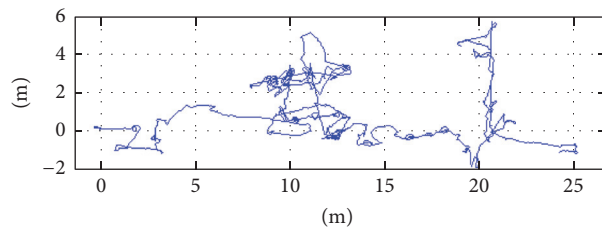FIGURE 14: Indoor environment described by keyframes sequence.

FIGURE 15: The robot route resolved through keyframes global map and content-based image matching.

FIGURE 16: The global position relationship of keyframes sequence; "∗" is the position of each keyframe.

lamps. The feature points on these similar objects in different rooms are easy to be error matched as the same points. If ORB SLAM needs to match all keyframes in the map, the robot can easily localize itself to wrong room (the distance between the two rooms is 12 m). But our method is able to take advantage of the image of dissimilar objects in different rooms, the interference caused by similar objects can be suppressed, and robot can fix its position precisely in these rooms and less error localized.

While the robot is moving in the two corridors, as when there are fewer similar objects than that of rooms and ORB, the method can extract feature points more effectively, the experiment result of ORB SLAM is better than that of our method.
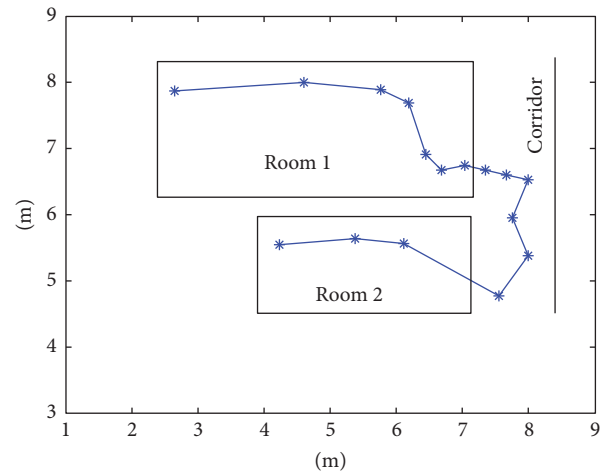
Table 1 also shows the comparison between our method and CV-SLAM. The difference is not significant under normal condition. However, compared with CV-SLAM method on commercial equipment, our method is immune to kidnapping events, because our method can build global map of indoor environment and make good use of this map to fix robot position at any time.

4.2. The Test for Robot Localization under Kidnap Condition. The kidnap problem for robot self-localization can be solved effectively by our method. In order to test robot in more complex kidnap condition, two adjoining rooms (20-meter-square and 10-meter-square) and the corridor (5 m) are chosen as experiment site. Through our method, 16 frames were extracted as keyframes to build the global map of these rooms and the corridor, with the position relationship of keyframes sequence shown in Figure 16. The keyframes mosaic result is shown in Figure 17. The rooms and small corridor could restrict the field of robot camera, such that the robot cannot fix its position by watching the distant landmarks. It was kidnapped and suddenly put into a far-away place. In the test, the robot was frequently taken from one place to another place (2 or 3 m apart) by experimenters to achieve kidnap. Through matching with global map, the robot was still able to fix its position effectively, especially
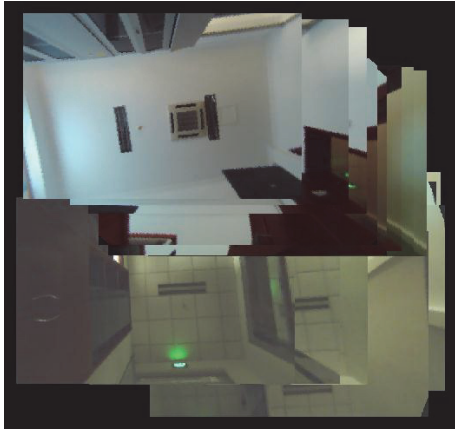
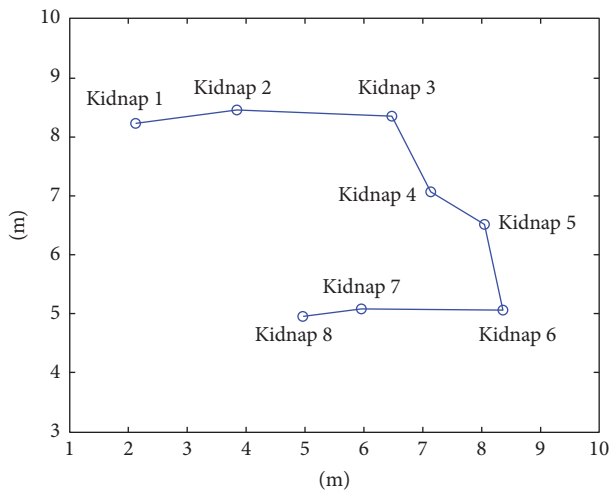FIGURE 17: The global map for kidnap test described by keyframes sequence.

FIGURE 18: The kidnap experiment; eight consecutive kidnaps were executed by experimenters, and sign "o" is the localization result for each kidnap position fix by the robot after landing.

when the robot is suddenly moved from room to the corridor, as shown in Figure 17, and position RMSE is less than 0.4 m.

5. Conclusion

For the issue of robot localization and mapping in the indoor environment, this paper presents keyframes global map establishing method for robot localization through content-based image matching with ability to analyze distortion and overlapping of keyframes. Results show that common problems, such as kidnapping or disturbances by similar objects, can be resolved through the content-based image matching method presented in this paper, which is specially designed for indoor environment. In the test, the keyframes global map can be established by this method and describe the indoor environment effectively. Although there are many similar objects in the experiment site, the robot cannot be kidnapped and can localize itself accurately (Figure 18), with the position RMSE being less than 0.5 m.

Competing Interests

Acknowledgments

References

[1] F. Y. Zhou, X. F. Yuan, Y. Yang et al., "A high precision visual localization sensor and its working methodology for an indoor mobile robot," *Frontiers of Information Technology & Electronic Engineering*, vol. 17, no. 4, pp. 365–374, 2016.

[2] E. Garcia-Fidalgo and A. Ortiz, "Vision-based topological mapping and localization by means of local invariant features and map refinement," *Robotica*, vol. 33, no. 7, pp. 1446–1470, 2015.

[3] J. M. Santos, M. S. Couceiro, D. Portugal, and R. P. Rocha, "A sensor fusion layer to cope with reduced visibility in SLAM," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 3, pp. 401–422, 2015.

[4] S. Jia, K. Wang, and X. Li, "Mobile robot simultaneous localization and mapping based on a monocular camera," *Journal of Robotics*, vol. 2016, Article ID 7630340, 11 pages, 2016.

[5] W. Y. Jeong and K. M. Lee, "CV-SLAM: a new ceiling vision-based slam technique," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '05)*, pp. 3195–3200, IEEE, Edmonton, Canada, 2005.

[6] I. Dryanovski, R. G. Valenti, and J. Xiao, "Fast visual odometry and mapping from RGB-D data," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '13)*, pp. 2305–2310, IEEE, Karlsruhe, Germany, May 2013.

[7] Y. Ji, A. Yamashita, and H. Asama, "RGB-D SLAM using vanishing point and door plate information in corridor environment," *Intelligent Service Robotics*, vol. 8, no. 2, pp. 105–114, 2015.

[8] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[9] K. Tateno, F. Tombari, and N. Navab, "Real-time and scalable incremental segmentation on dense SLAM," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '15)*, pp. 4465–4472, IEEE, Hamburg, Germany, October 2015.

[10] T. Schöps, J. Enge, and D. Cremers, "Dense planar SLAM," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR '14)*, pp. 157–164, IEEE, Munich, Germany, September 2014.

[11] T. Schöps, J. Enge, and D. Cremers, "Semi-dense visual odometry for AR on a smartphone," in *Proceedings of the 13th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '14)*, pp. 145–150, IEEE, Munich, Germany, September 2014.

[12] J. Engel, J. Stückler, and D. Cremers, "Large-scale direct SLAM with stereo cameras," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '15)*, pp. 1935–1942, Hamburg, Germany, October 2015.

[13] W. Y. Jeong and K. M. Lee, "CV-SLAM: a new ceiling vision-based SLAM technique," in *Proceedings of the IEEE IRS/RSJ*

International Conference on Intelligent Robots and Systems (IROS '05), pp. 3195–3200, IEEE, Alberta, Canada, August 2005.

[14] S. Jo, H. Choi, and E. Kim, "Ceiling vision based SLAM approach using sensor fusion of sonar sensor and monocular camera," in *Proceedings of the 12th International Conference on Control, Automation and Systems (ICCAS '12)*, pp. 1461–1464, Jeju Island, Korea, October 2012.

[15] H. Choi, D. Y. Kim, J. P. Hwang, C.-W. Park, and E. Kim, "Efficient simultaneous localization and mapping based on ceiling-view: ceiling boundary feature map approach," *Advanced Robotics*, vol. 26, no. 5-6, pp. 653–671, 2012.

[16] D. Y. Kim, H. Choi, H. Lee, and E. Kim, "A new cvSLAM exploiting a partially known landmark association," *Advanced Robotics*, vol. 27, no. 14, pp. 1073–1086, 2013.

[17] S. Lee, S. Lee, and S. Baek, "Vision-based kidnap recovery with SLAM for home cleaning robots," *Journal of Intelligent & Robotic Systems*, vol. 67, no. 1, pp. 7–24, 2012.

[18] S. T. Pfister and J. W. Burdick, "Multi-scale point and line range data algorithms for mapping and localization," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06)*, pp. 1159–1166, IEEE, Orlando, Fla, USA, May 2006.

[19] R. L. Stewart and H. Zhang, "Image similarity from feature-flow for keyframe detection in appearance-based SLAM," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '11)*, pp. 305–312, Phuket, Thailand, December 2011.

[20] J. Kim and I.-S. Kweon, "Robust feature matching for loop closing and localization," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, pp. 3905–3910, IEEE, San Diego, Calif, USA, November 2007.

[21] P. W. Smith, K. B. Johnson, and M. A. Abidi, "Efficient techniques for wide-angle stereo vision using surface projection mode," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 113–118, IEEE, Fort Collins, Colo, USA, 1999.

[22] H. Choi, Y. K. Dong, J. P. Hwang et al., "CV-SLAM using ceiling boundary," in *Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA '10)*, pp. 228–233, IEEE, Taichung, Taiwan, 2010.

[23] P. Chen, Z. Gu, G. Zhang, and H. Liu, "Ceiling vision localization with feature pairs for home service robots," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '14)*, pp. 2274–2279, IEEE, Bali, Indonesia, December 2014.

[24] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[25] K. H. Strobl and M. Lingenauber, "Stepwise calibration of focused plenoptic cameras," *Computer Vision and Image Understanding*, vol. 145, pp. 140–147, 2016.

[26] J. Yuan, Y. L. Huang, F. C. Sun, and T. Tao, "Active exploration using a scheme for autonomous allocation of landmarks," *Robotica*, vol. 32, no. 5, pp. 757–782, 2014.

[27] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[28] Z. Li, C. Shao, and Y. M. Liu, "Motion estimation based on axis affine model," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '10)*, pp. 572–576, IEEE, Xi'an, China, August 2010.

[29] D. Caballero, T. Antequera, A. Caro, M. L. Duran, and T. Perez-Palacios, "Data mining on MRI-computational texture features to predict sensory characteristics in ham," *Food and Bioprocess Technology*, vol. 9, no. 4, pp. 699–708, 2016.

[30] J.-S. Cho, H.-J. Lee, J.-H. Park, J.-H. Sung, J.-Y. Choi, and K.-D. Moon, "Image analysis to evaluate the browning degree of banana (*Musa* spp.) peel," *Food Chemistry*, vol. 194, pp. 1028–1033, 2016.