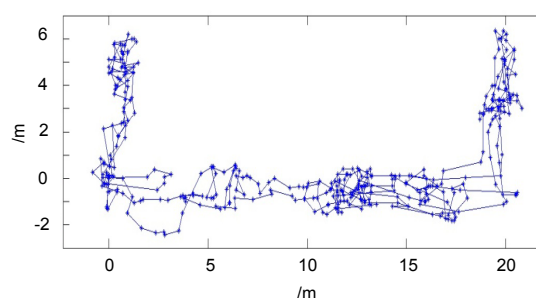




基于视觉内容匹配的机器人自主定位系统

曹天扬^{1,2}, 蔡浩原^{1*}, 方东明¹, 刘昶¹

¹中国科学院电子学研究所传感技术联合国家重点实验室, 北京 100190; ²中国科学院大学, 北京 100049



摘要: 为了使机器人准确使用全局地图、解决被绑架问题, 设计了基于视觉内容匹配的自主定位系统, 充分利用每个房间、每段走廊中的物体与布局信息进行机器人导航定位。视觉内容匹配包括提取图像重叠区、重叠区域的子块分解重建和相似度匹配。首先将待匹配图像中由机器人视觉和位移造成的畸变调整为一致、然后分析图像内容相似度, 并设计了天花板特征点筛选和误匹配子块剔除算法, 实现重叠区的准确提取与重建。实验中, 机器人视觉可与关键帧序列构建的全局地图准确匹配, 实时提取与之最相似的关键帧对机器人定位, 匹配准确率不低于 95%, 定位精度 RMSE<0.5 m, 被绑架时仍能准确自主定位。

关键词: 机器人视觉; 图像内容匹配; 图像畸变; 图像重叠区; 重建重叠区

中图分类号: TP242; TP391.41

文献标志码: A

Robot vision localization system based on image content matching

Tianyang Cao^{1,2}, Haoyuan Cai^{1*}, Dongming Fang¹ and Chang Liu¹

¹State Key Laboratory of Transducer Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China;

²University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: In order to take advantage of global navigation map for robot self-localization and solve kidnap problem, a robot vision localization system is presented based on graphic content matching. It can make good use of the different objects and their layout in different rooms or corridors to fix robot position, which cannot be disturbed by similar objects. This vision localization system is composed of frames overlap region extraction and overlap region rebuilding through sub-blocks matching, and the interference caused by points on the wall and mismatching sub-blocks can be deleted. The image distortion can be adjusted to the same before matching. In the experiment, this graphic matching method can match the real-time robot vision with keyframes global map effectively, and find out the most similar keyframe for each vision image and fix robot position exactly. More than 95% robot vision can be matched and position RMSE<0.5 m. Robot can also localize itself effectively when it is kidnapped.

Keywords: robot vision; graphic content matching; image distortion; image overlap region; overlap region rebuilding

DOI: 10.3969/j.issn.1003-501X.2017.05.008

Citation: *Opto-Elec Eng*, 2017, 44(5): 523–533

收稿日期: 2017-01-03; 收到修改稿日期: 2017-03-20

*E-mail: hycail@mail.ie.ac.cn

1 引言

视觉系统是机器人获取信息的主要来源,它可以像人眼一样为机器人提供导航定位服务,尤其是在室内等 GPS 无法有效应用的场景中^[1,2]。理想的视觉导航定位系统应能使机器人在包含多个房间、多条走廊的复杂环境中准确自主运动^[3]。

目前,在机器人视觉导航定位领域中最有效的方法是基于视觉的同步定位与地图构建技术(visual simultaneous localization and mapping, V-SLAM),包括 ORB SLAM、dense SLAM、semi-dense SLAM、LSD SLAM 和 CV-SLAM 等多种方法。这些方法都是以充分挖掘视觉图像中的导航特征信息为基础实现的,可分为稠密法、半稠密法和稀疏法三大类^[4]。稀疏法主要是以提取和匹配图像的特征点实施导航定位,稠密法是以图像中的全部像素点为研究对象实施匹配、提高了图像信息的利用率^[5]。dense SLAM 属于典型的稠密法,使用了 3D 特征进行匹配和地图构建^[6]。在此基础上发展出了半稠密法,如 semi-dense SLAM、LSD SLAM,主要利用视觉图像中物体的边界信息提升图像的运算速度^[7]。LSD SLAM 还引入了关键帧技术提高信息处理效率^[8]。ORB SLAM 是一种稀疏法,引入了 ORB 特征构建特征点,可兼顾解算精度和运算速度^[9]。CV-SLAM 也是一种稀疏法,它利用天花板上的特征点进行机器人导航^[10]。由于天花板与地面等高, CV-SLAM 使用单目摄像机即可快速对天花板上的空间特征点进行检测^[11,12]。

但这些 V-SLAM 方法也有一定的局限性,不论是稀疏法、半稠密法还是稠密法,都是仅使用了特征点、像素点等“图像点”信息。而由这些图像点按照一定规则排列构成的图像景物内容尚未得到有效利用,如图像中每个物体的外形、结构、位置分布等。如果机器人不使用这些视觉的内容信息、仅使用物体上的特征点信息,自主导航时极易被绑架^[13]。绑架发生时,机器人被突然移动了一段距离,如被人或其它机器人误碰撞移动,此时 SLAM 算法很容易失效。因为与前一时刻的位置过远超出视距,SLAM 很难通过先前的定位结果校准新增位移,SLAM 的里程计也会因车轮打滑而失效,甚至测量出的位移增量都不准确,这种情况下需要在全局地图中匹配导航才能对机器人重新准确定位^[14]。如 ORB SLAM 算法就设计了这一功能,一旦出现被绑架等无法成功匹配的情况,它会将当前视频帧中的特征点与所有关键帧重新匹配、在全局地图

中进行定位。但室内通常会存在大量的相似物体、这些相似物体通常还十分明显,如写字楼的不同房间都有相同的空调口和吸顶灯,它们的特征基本一致、极易导致机器人把自己误定位到错误的房间。

为了解决上述问题,本文以视觉内容匹配为基础设计了机器人自主定位系统。图像内容匹配可以充分利用室内不同房间、不同走廊中的各种物体的外形与布局信息,使机器人能实时确定自己是在哪间屋子及在屋中的具体位置。图像内容匹配最大的干扰来自于每张照片都具有不同的拍摄角度和拍摄位置,造成图像中的景物呈现不同程度的形状畸变,严重影响了图像之间的相似度分析。为了消除这些畸变的影响,本文采用先将待匹配的图像畸变调整为一致后再进行匹配的策略,实现机器人视觉图像与由关键帧序列构成的全局导航地图的准确匹配和定位。首先,求取视觉图像与各个关键帧的拍摄景物重叠区。根据图像中的特征点对机器人视觉图像平移和旋转,求取它与全局地图中各个关键帧可能的拍摄重叠区。然后,对视觉图像重叠区重建。通过重叠区域子块分解重建算法对视觉图像与每个关键帧的重叠区进行匹配重建,可消除因拍摄视角和位置不同造成的图像畸变,能够准确评估机器人视觉与各个关键帧的相似度、挑选出与当前视觉内容最相似的关键帧。最后根据该关键帧对机器人当前的位置进行定位。

在测试中,实验区域包括两间房间(15 m²)和一条 20 m 的走廊。该区域的全局地图包含 58 幅关键帧。对于每幅机器人实时视觉中的图像,通过与全局地图进行图像内容匹配,与之最相似的关键帧可以被准确提取出来用于对机器人定位。超过 95%的机器人视觉图像可以准确匹配,定位精度 RMSE<0.5 m。

2 基于视觉图像内容匹配的定位方法

2.1 算法的架构设计

基于视觉内容匹配的机器人自主定位系统可以充分利用不同房间、不同走廊区域中的不同物体对自己的位置实时定位。算法由三个环节构成:机器人视觉与关键帧的图像重叠区提取、机器人视觉重叠区的子块分解重建、关键帧重叠区与机器人视觉重建区匹配,它们之间的关系如图 1 所示。

该系统需要两项输入,由多幅关键帧组成的建筑物内部全局导航地图、机器人的实时视觉图像。地图

和实时视觉图像都由机器人顶部垂直天花板的广角摄像机拍摄^[15],如图2所示。

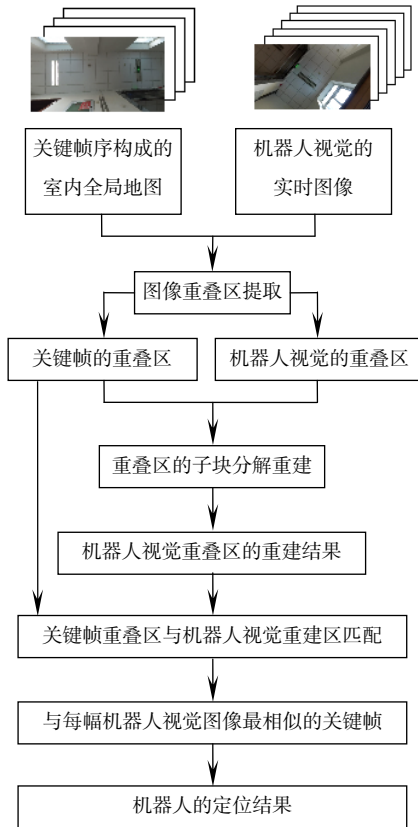


图1 系统的架构图。
Fig. 1 System structure.

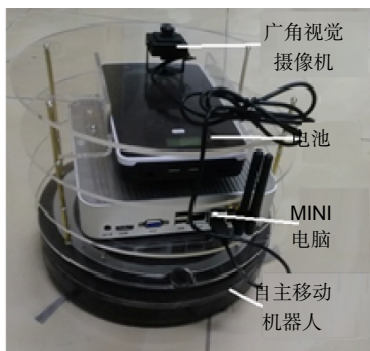


图2 机器人与它的视觉系统。
Fig. 2 Robot and its vision system.

本文中的实验区域由两个房间(各 15 m²)和一条走廊(长 20 m)组成。构建全局导航地图时,遥控机器人在屋子和走廊中运动,机器人根据里程计的输定时拍摄关键帧、并记录拍摄各帧的位置坐标。可以让机器人按照最短的路径行进,有效减少拍摄冗余。如

图3所示,实验区域的全局地图包含58幅关键帧,每幅关键帧拍摄的范围是6 m×3.4 m,可确保对试验区域的完全覆盖。机器人自主工作后将实时拍摄的图像与该地图中的关键帧进行匹配,找出与当前时刻最相似的关键帧,即可确定当前机器人所在的位置。这是一种全局地图匹配方法,可以有效消除SLAM算法无法解决的绑架问题,机器人无论被突然放置到室内的任何位置,都可以准确地自主定位。

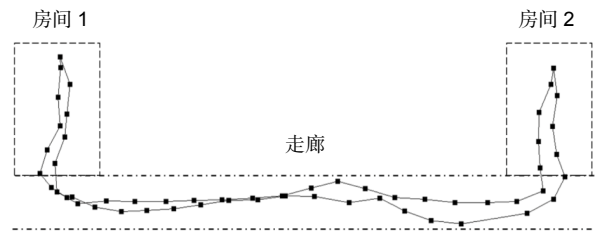


图3 关键帧序列构建的全局地图(图中小方块为拍摄关键帧的位置)。

Fig. 3 The global map composed by keyframes (the blocks in the figure is keyframes taken position).

2.2 图像重叠区提取

对于机器人拍摄的实时视觉图像,即便是与它最相似的关键帧,二者存在拍摄位置和拍摄角度的差异,也仅是有部分区域重叠。因此两帧图像内容的相似度分析与匹配就是针对它们的重叠区域进行。

为了实现这一目标,首先要将拍摄于不同位置、不同航向的机器人视觉图像和关键帧图像平移和旋转到相同的拍摄位置和视角。如图4所示,经过旋转和平移,帧(a)、(b)中的三角形物体和平行四边形物体被调整到了相同位置。

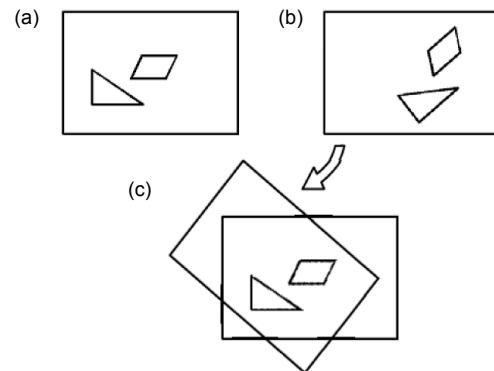


图4 图像的旋转和平移过程。

Fig. 4 The image rotation and translation.

平移量(T_x, T_y)和旋转量 H 可通过由 SURF 算法提

取的图像中的导航特征点解算出来^[16]。对于天花板上的点,由于与地面等高、摄像机的视角无俯仰和横滚,因此天花板三维空间点 (X,Y,Z) 和视觉图像点 (x,y) 的对应关系可以通过如下的相机数学模型描述^[17,18]:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & 0 \\ 0 & \frac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (1)$$

式中: H 为机器人航向角, f 、 dx 、 dy 是机器人视觉系统的参数。可简化为

$$x = \frac{f}{dxZ} (X \cos H + Y \sin H + T_x), \quad (2)$$

$$y = \frac{f}{dyZ} (-X \sin H + Y \cos H + T_y). \quad (3)$$

因此,对于天花板上的特征点,令它在关键帧中的坐标为 $(x_{1,n}, y_{1,n})$ 和机器人视觉中的坐标为 $(x_{2,n}, y_{2,n})$,可由式(2)、式(3)推导出它们之间具有以下的线性关系:

$$x_{2,n} = \cos H \cdot x_{1,n} + \sin H \cdot y_{1,n} + \frac{f}{dxZ_c} T_x, \quad (4)$$

$$y_{2,n} = -\sin H \cdot x_{1,n} + \cos H \cdot y_{1,n} + \frac{f}{dyZ_c} T_y, \quad (5)$$

式中: f 是相机的焦距, dx 和 dy 是像素尺寸,它们都是常数。只要能够从机器人视觉图像和关键帧之间提取多组特征点,带入式(4)、式(5)后通过最小二乘法即可解算出机器人在拍摄这两帧图像时的平移量 (T_x, T_y) 和旋转量 H 。

根据 (T_x, T_y) 和 H 进行反向旋转和平移,两幅图像的重叠区非常明显,它们的重叠区域可以通过旋转平移后两幅图像重合点制成的掩模提取。平移和旋转的结果如图5(a)和(b)所示。旋转和平移之后,重叠区域很明显,如图5(c)所示。提取重叠区可以通过两幅图像的掩模实现,如图5(d)所示。

如果特征点在墙上,由于与天花板不等高,则式(2)~式(5)会受到很大干扰。因此本文以特征点之间的几何关系为基础,设计了一种通过特征点连线和连线夹角来识别和剔除墙壁特征点的方法。

对于帧 A 中天花板上的特征点 a 和 b ,如果机器人发生旋转和平移之后拍摄到的帧 B 中仍然包含特征点 a 和 b 所在的区域,则 ab 的长度在两帧中不变:

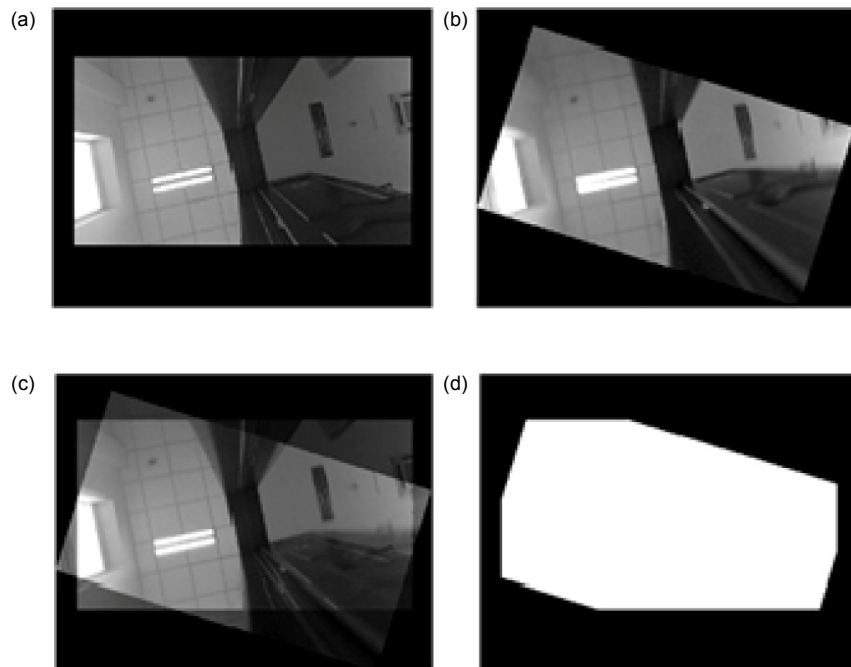


图5 旋转和平移的结果。(a) 关键帧平移。(b) 机器人视觉旋转。(c) 两帧重叠区域。(d) 重叠区域的掩模。

Fig. 5 The rotation and translation result. (a) Keyframe rotation. (b) Robot vision rotation. (c) Overlap region. (d) Overlap region mask.

$$d_{A,ab} = [(x_{A,a} - x_{A,b})^2 + (y_{A,a} - y_{A,b})^2]^{1/2}, \quad (6)$$

$$\begin{aligned} d_{B,ab} = & [(x_{B,a} - x_{B,b})^2 + (y_{B,a} - y_{B,b})^2]^{1/2} = \\ & \{[(C_1(X) + S_1(Y) + \frac{fT_x}{dZ_c}) - \\ & (C_2(X) + S_2(Y) + \frac{fT_x}{dZ_c})]^2 - \\ & [(-S_1(X) + C_1(Y) + \frac{fT_y}{dZ_c}) - \\ & (-S_2(X) + C_2(Y) + \frac{fT_y}{dZ_c})]^2\}^{1/2} = \\ & [(C_1(X) + S_1(Y) - C_2(X) - S_2(Y))^2 + \\ & (-S_1(X) + C_1(Y) + S_2(X) - C_2(Y))^2]^{1/2} = \\ & d_{A,ab}, \end{aligned} \quad (7)$$

式中:

$$\begin{aligned} C_1(X) &= X_{A,a} \cos H, \quad S_1(Y) = Y_{A,a} \sin H, \\ C_2(X) &= X_{A,b} \cos H, \quad S_2(Y) = Y_{A,b} \sin H, \\ S_1(X) &= X_{A,a} \sin H, \quad C_2(Y) = Y_{A,b} \cos H, \\ S_2(X) &= X_{A,b} \sin H, \quad C_1(Y) = Y_{A,a} \cos H, \end{aligned}$$

$(x_{A,a}, y_{A,a})$ 、 $(x_{A,b}, y_{A,b})$ 是图像帧 A 中天花板上 a 点、 b 点的坐标, $(x_{B,a}, y_{B,a})$ 、 $(x_{B,b}, y_{B,b})$ 是 a 点、 b 点在图像 B 中的坐标, $d_{A,ab}$ 是关键帧中 a 点、 b 点连线的长度, $d_{B,ab}$ 是视觉图像中 a 点、 b 点连线的长度。若两个特征点都在天花板上, 则它们的连线长度在不同的照片中均保持不变。

而如果特征点 b_w 在墙壁上, 该特征点与 a 点连线的长度在不同的照片中会发生变化, 不再保持不变。连线长度的推导过程为

$$d_{A,ab_w} = [(x_{A,a} - x_{A,b_w})^2 + (y_{A,a} - y_{A,b_w})^2]^{1/2}, \quad (8)$$

$$\begin{aligned} d_{B,ab_w} = & [(x_{B,a} - x_{B,b_w})^2 + (y_{B,a} - y_{B,b_w})^2]^{1/2} = \\ & \{[(C_1(X) + S_1(Y) + \frac{fT_x}{dZ_c}) - \\ & (C_3(X) + S_3(Y) + \frac{fT_x}{dZ_w})]^2 + \\ & [(-S_1(X) + C_1(Y) + \frac{fT_y}{dZ_c}) - \\ & (-S_3(X) + C_3(Y) + \frac{fT_y}{dZ_w})]^2\}^{1/2} \neq \\ & d_{A,ab}, \end{aligned} \quad (9)$$

式中: $C_1(X)$ 、 $S_1(X)$ 、 $C_1(Y)$ 、 $S_1(Y)$ 同式(7)定义一样, $S_3(X) = X_{A,b_w} \sin H$, $S_3(Y) = Y_{A,b_w} \sin H$, $C_3(X) = X_{A,b_w} \cos H$, $C_3(Y) = Y_{A,b_w} \cos H$; (x_{A,b_w}, y_{A,b_w}) 、 (x_{B,b_w}, y_{B,b_w}) 是墙壁点 b_w 在图像 A 、 B 中的坐标, d_{A,ab_w}

是关键帧中 a 、 b_w 点连线的长度, d_{B,ab_w} 是视觉图像中 a 、 b_w 点连线的长度。由式(8)、式(9)可知, 由于墙体不与天花板等高, $d_{A,ab} \neq d_{B,ab_w}$ 。

因此通过对比不同图像中由 SURF 算法提取的相同特征点之间的连线长度, 可以准确判断它们是否是在天花板上, 例如图 6(a)、图 6(b)两帧图像中的线段长度 mn 是否相等。为了进一步提高天花板特征点的提取精度, 考虑到天花板上的特征点连线相等, 则特征点集构成的三角形是全等三角形, 如图 6 中的三角形 Δopq 所示, 因此还可以根据两条特征点连线的夹角判断它们是否属于天花板上的点, 如图 6(a)、6(b)中的角 $\angle q$ (或 $\angle o$ 、 $\angle p$) 是否相等。

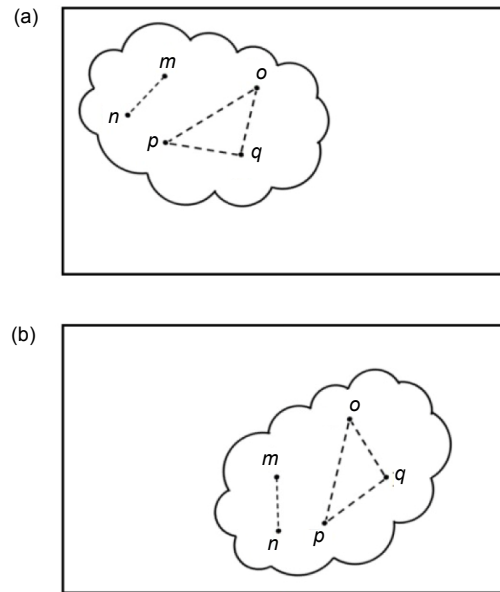


图 6 天花板上的特征点所具备的连线与夹角特性。

(a) 机器视觉图像。(b) 关键帧。

Fig. 6 The features of connect lines and angle of feature points. (a) Robot vision. (b) Keyframe.

因此本文的墙壁特征点滤除包括两步:

第一步: 连线筛选。在一定误差范围(图像尺寸的 3%)内比较 SURF 匹配特征点的连线长度、筛选天花板上的特征点。

第二步: 连接三角形筛选。在一定误差范围($\leq 3^\circ$)内比较 SURF 匹配特征点的连线夹角、再次筛选天花板上的特征点。

处理效果如图 7 所示, 特征点筛选的效果如图 7(c)、图 7(d)所示, 墙壁区域(包括家具)中的干扰点被有效滤除。

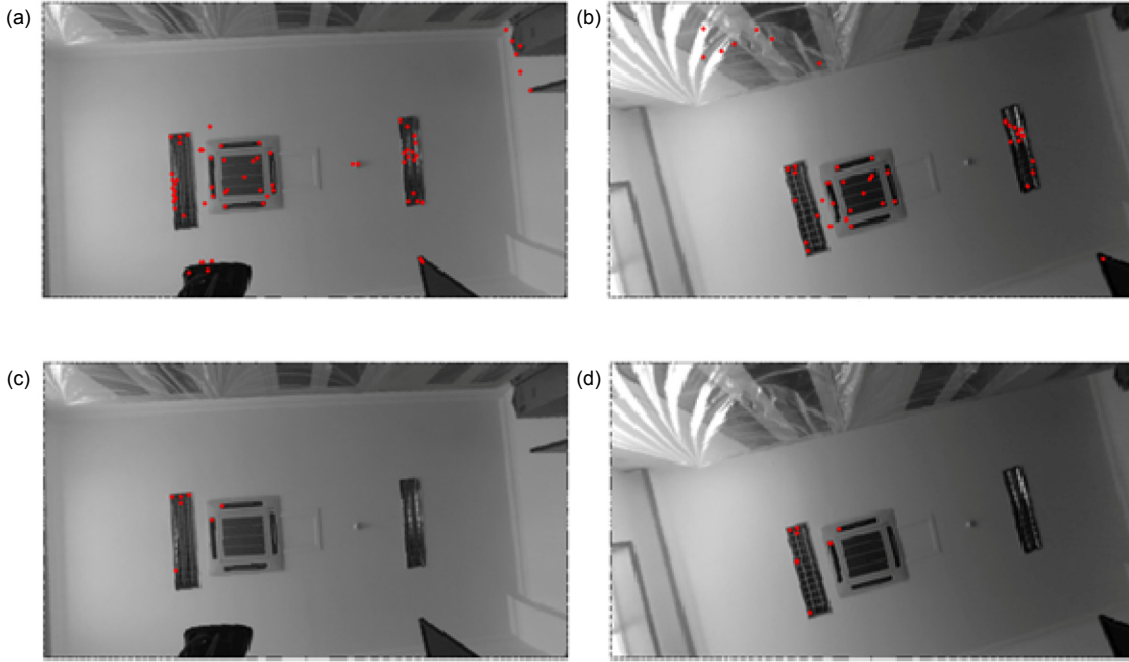


图7 图像特征点筛选结果(红点为特征点). (a) 关键帧中的特征点. (b) 视频帧中的特征点. (c) 筛选后的关键帧中的特征点. (d) 筛选后的视频帧中的特征点.

Fig. 7 The screen result of feature points (red points). (a) The feature points in keyframe. (b) The feature points in robot vision. (c) The screen result of feature points in keyframe. (d) The screen result of feature points in robot vision.

2.3 重叠区域的子块分解重建

如果机器人实时视觉和关键帧的重叠区仅是天花板,则经过平移和旋转校正之后,提取的重叠区近似一致,直接求取相关系数即可以判断出两帧图像的相似度。

如果重叠区包含了墙体部分(如门窗、家具、墙),因为机器人观察它们的视角为侧视、具有俯仰和横滚角,则会造成这些景物的形状畸变。这种畸变过程可以通过包含俯仰、横滚和航向信息的相机内外参数模型来描述。在相机模型中图像点与真实空间点的关系变为^[17]

$$z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & 0 \\ 0 & \frac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (10)$$

式中:

$$R = \begin{bmatrix} \cos R & 0 & -\sin R \\ 0 & 1 & 0 \\ \sin R & 0 & \cos R \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos P & \sin P \\ 0 & -\sin P & \cos P \end{bmatrix} \cdot \begin{bmatrix} \cos H & \sin H & 0 \\ -\sin H & \cos H & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

R 是横滚角, P 是俯仰角, H 是航向角, $T = [T_x, T_y, T_z]^T$ 是机器人的位移矩阵。化简后:

$$x = \frac{f}{dx} \frac{U_1 + T_x}{U + T_z}, \quad (11)$$

$$y = \frac{f}{dy} \frac{U_2 + T_y}{U + T_z}. \quad (12)$$

式中:

$$U = (\sin R \cos H + \cos R \sin P \sin H)X + (\sin R \sin H - \cos R \sin P \cos H)Y + (\cos R \cos P)Z$$

$$U_1 = (\cos R \cos H - \sin R \sin P \sin H)X + (\cos R \sin H + \sin R \sin P \cos H)Y + (-\sin R \cos P)Z$$

$$U_2 = (-\cos P \sin H)X + (\cos P \cos H)Y + (\sin P)Z ;$$

x 、 y 与 X 、 Y 、 Z 呈现非线性的关系，物体会出现远小近大的形状畸变。

对于待匹配的两帧图像中各个小区域(子块)，尽管存在畸变，但包含同一物体的子块还是有很大的相似性，如家具上的把手、门窗上的图案，尽管在不同的照片中畸变不同，但形状和像素的分布还是保持一定的相似度，只是畸变后包含相同内容的子块位置发生了比较大的变化，离照片中心越远的物体形状畸变越大，畸变造成的子块位移也越大。为了准确找到机器人实时视觉中每个子块在关键帧中的对应位置，消除畸变造成的位移，设计了一种图像重建方法，机器人视觉中的重叠区域将根据关键帧的信息进行重建。

每个子块在关键帧中的精确位置采用绝对误差和方法(sum of absolute difference, SAD)匹配求得：

$$D(i, j) = \arg \min_{\substack{i \in [1, L] \\ j \in [1, D]}} \sum_{s=1}^M \sum_{t=1}^N |I_L(i+s-1, j+t-1) - I_T(s, t)|, \quad (13)$$

式中： I_T 是机器人视觉的一个子块， M 和 N 是子块的尺寸， I_L 是关键帧序列中的一幅， L 和 D 是关键帧的尺寸。遍历该关键帧，找到坐标可以使式(13)输出最小的 (i, j) 。 (i, j) 就是子块 I_T 在关键帧 I_L 中的重建位置。通过这种方法，所有机器人视觉中的子块都将与

关键帧 I_L 进行匹配，重建重叠区。重建过程和结果如图 8 所示，机器视觉的重叠区(图 8(a))被分为很多子块(图 8(b))，这些子块可以与关键帧的重叠区通过 SAD 方法进行匹配(图 8(c))，找到它们在关键帧中最适合的位置(图 8(d))，重建的机器人视觉区域如图 8(e)所示。如果一幅关键帧与机器人视觉包含大量相同的图像内容，重建后的视觉重叠区将会与关键帧的重叠区非常相似。否则只会有少数几个子块能与关键帧进行匹配。

对于相似物体造成的子块误匹配问题，本文设计了一种动态阈值筛选方法。考虑到即使墙壁门窗等景物的照片发生了很严重的形状畸变，但景物依然保持为一个完整的连接体、不会被撕裂为几个分布于不同位置，即机器视觉重叠区和与之最相似的关键帧重叠区中的同一物体所在的子块位置会位于一个邻域内。因此可以设定一个邻域的阈值范围，超过该阈值的子块通常是相似物体造成的误匹配干扰，予以剔除。

本文为了更加精确地滤除误匹配，考虑到重叠区域提取过程中会出现旋转和平移误差，以式(4)、式(5)为基础将两种误差结合起来设计了一种动态邻域阈值 $[\Delta x, \Delta y]$ ：

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \cos \Delta H & \sin \Delta H & \Delta P_x \\ -\sin \Delta H & \cos \Delta H & \Delta P_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (14)$$

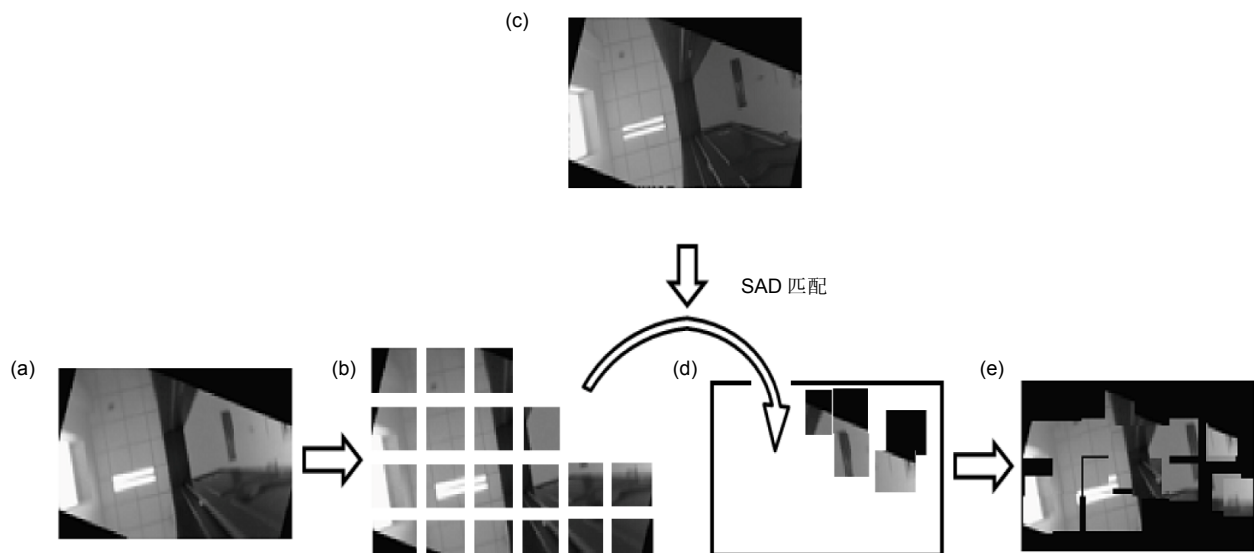


图 8 重建过程。(a) 机器人视觉的重叠区。(b) 图(a)的子块分割。(c) 关键帧的重叠区。(d) 子块的匹配重建。(e) 重建视觉图像重叠区。

Fig. 8 The process of rebuilding. (a) The overlap region in robot vision. (b) The sub-block division of fig. (a). (c) The overlap region in keyframe. (d) The sub-block rebuilding. (e) The rebuilding result of overlap region of robot vision.

式中: $(\Delta P_x, \Delta P_y)$ 是重叠区域提取时平移误差阈值, (x, y) 是机器人视觉中每个子块距离重叠区中心的位置, ΔH 是重叠区域的旋转误差阈值。针对靠近重叠区边界时旋转误差影响大这一问题, 将 ΔH 与子块距离图像中心的距离结合到一起, 可以使式(14)在靠近边界时可适当增大阈值、有利于保留在图像边界处的正确匹配子块。经过式(14)处理后, 重建的视觉图像重叠区与关键帧重叠区十分相似, 如图 9 所示。

为了对比最相似关键帧和其它关键帧在与机器人视觉(图 9(b))匹配时的区别, 图 10 中显示了图 9(b)的视觉图像与另一关键帧的匹配结果。图 10 中视觉的重建区与关键帧的重叠区相差十分明显, 因此可以通过本文方法准确判断待匹配的两帧图像是否相似。

2.4 机器人视觉重建的重叠区和关键帧重叠区的相似度匹配

对于关键帧的重叠区和机器人视觉重建的重叠

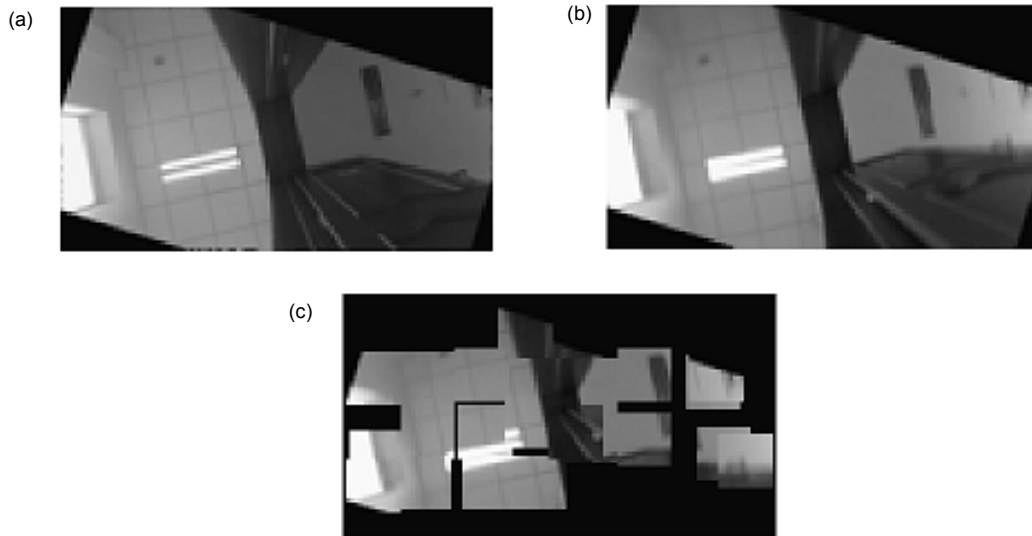


图 9 重叠区重建. (a) 关键帧重叠区. (b) 机器人视觉重叠区. (c) 子块分解重建的视觉图像重叠区.

Fig. 9 The rebuilding of overlap region. (a) The overlap region in keyframe. (b) The overlap region in robot vision. (c) The rebuilding result of overlap region of robot vision.

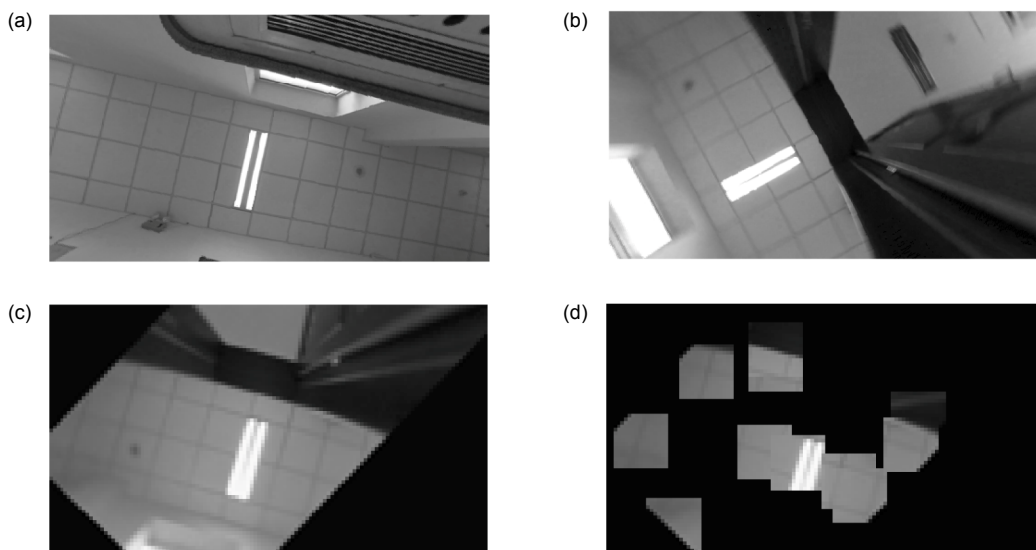


图 10 错误的重建重叠区. (a) 关键帧. (b) 机器人视觉. (c) 视觉重叠区. (d) 重叠区重建结果.

Fig. 10 The wrong rebuilding overlap region. (a) Keyframe. (b) Robot vision. (b) The overlap region in robot vision. (c) The rebuilding result of overlap region of robot vision.

区, 它们的相似度可以通过正确匹配的子块数量和两个区域的相关系数进行比较和评估。通过这种方法遍历全局地图中所有的关键帧, 与当前机器人视觉相似度最大的关键帧就是与之最相似的关键帧。

对于机器人视觉和与它最相似的关键帧, 成功匹配的子块数量和相关系数都是最大的。为了使相似度更加清晰, 使用它们的乘积作为相似度的结果输出。对于第 k 幅关键帧, 它的相似度 S_k 计算过程如下:

$$S_k = N_k \cdot \frac{C}{\sqrt{C_1 \cdot C_2}}, \quad (15)$$

式中:

$$C = \sum_{x,y} (C_{L,k}(x,y) - \bar{C}_{L,k})(C_{b,k}(x,y) - \bar{C}_{b,k}),$$

$$C_1 = \sum_{x,y} (C_{L,k}(x,y) - \bar{C}_{L,k})^2,$$

$$C_2 = \sum_{x,y} (C_{b,k}(x,y) - \bar{C}_{b,k})^2,$$

N_k 是可以匹配子块的数量, $C_{L,k}$ 是第 k 幅关键帧重叠区内的像素值, $C_{b,k}$ 是机器人视觉重建的重叠区内的像素值, $\bar{C}_{L,k}$ 和 $\bar{C}_{b,k}$ 是图像像素均值。

通过式(15), 可以精确求取各个关键帧与机器人视觉的相似度, 找出与之最相似的关键帧。机器人此刻的位置可以通过与最相似关键帧进行匹配, 将该视觉与该关键帧重叠区提取时求出的天花板特征点带入式(4)、式(5), 即可求出机器人相对于该关键帧的位置。

3 实验结果

本文的机器人视觉自主定位系统已经在大型复杂的室内环境中进行了测试。实验区域由一条走廊和两个房间组成。走廊 20 m 长, 每间屋子平均约 15 m²。实验区域的全局地图包含 58 幅关键帧, 如图 3 所示。

实验中, 机器人定时拍摄图像、实时对自己进行定位。通过基于子块分解重建的图像内容匹配, 机器人每次都可以准确提取与每幅视觉图像最相似的关键帧、进而对自己进行准确定位。在走廊中的图像内容匹配结果如图 11 所示, 在房间内的重建和匹配结果如图 12 所示。测试中, 机器人每隔 10 s 拍摄一帧图像, 总计 384 帧, 超过 95% 的机器人视觉图像可以实现与其最相似的关键帧匹配。误匹配率非常低, 定位结果如图 13 所示, 因为是与全局地图进行匹配, 误差不会像通常的 CV-SLAM 方法一样累积增加, 本文实验中误差 RMSE < 0.5 m。机器人通过定位结果可准确获知自己的运动轨迹。

而在绑架发生时, 本文的图像内容匹配定位方法比 SLAM 方法更具优势。因为本文方法是与全局地图通过图像内容匹配定位, 在这种情况下机器人依然能够准确自主定位。绑架状态下的测试结果如图 14 所示, 当试验人员每次移动机器人 3 m 至 7 m 后, 机器人仍能准确确定自己的位置, 优于易受绑架问题困扰的 CV-SLAM 算法。

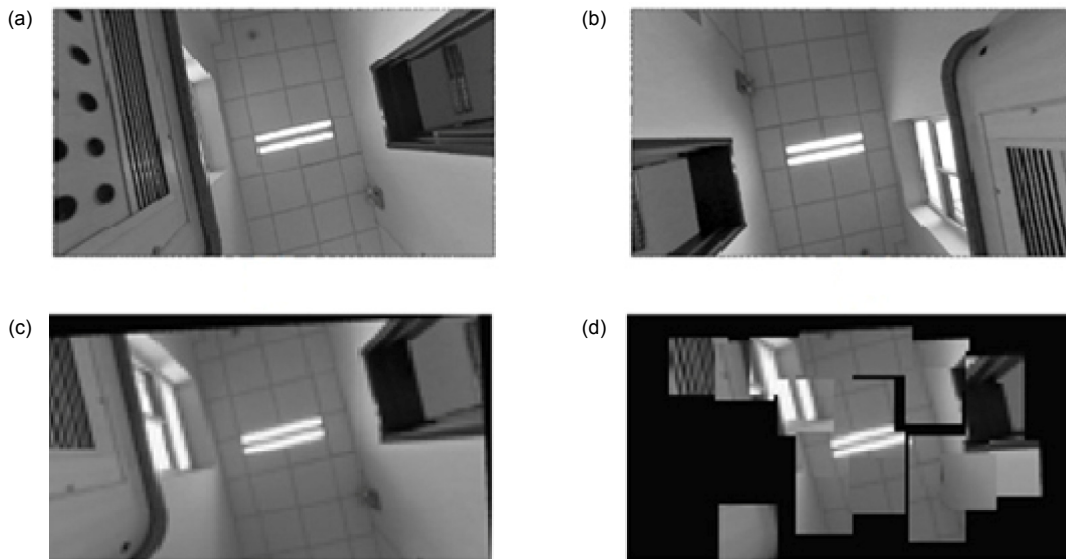


图 11 走廊里的实验结果. (a) 关键帧. (b) 机器人视觉. (c) 重叠区. (d) 重叠区重建结果.

Fig. 11 The experiment result in the corridor. (a) Keyframe. (b) Robot vision. (c) The overlap region in robot vision. (d) The rebuilding result of overlap region of robot vision.

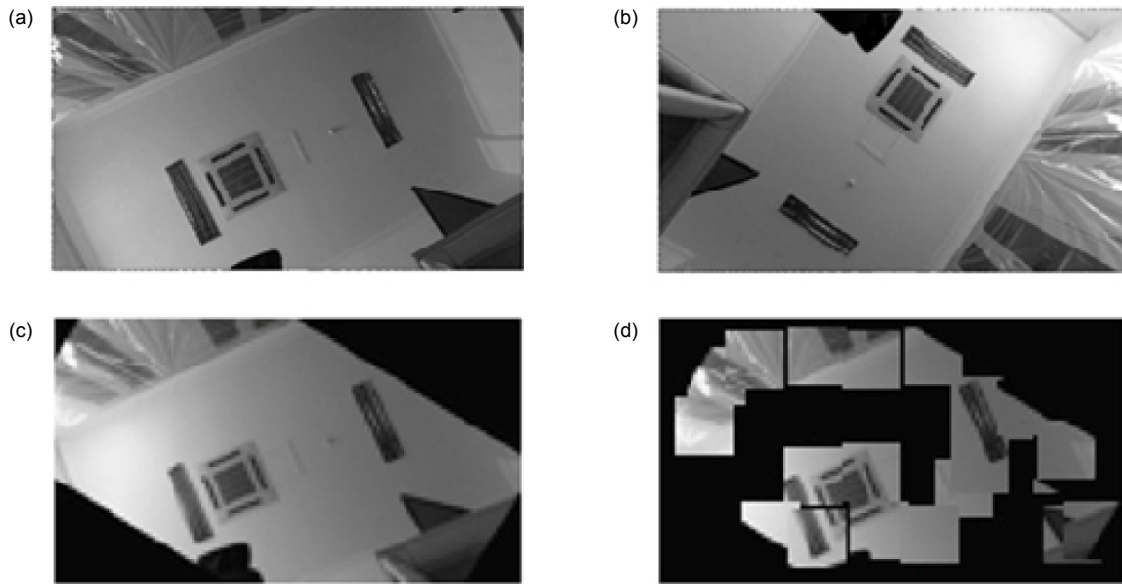


图 12 房间内的实验结果. (a) 关键帧. (b) 机器人视觉. (c) 重叠区. (d) 重叠区重建结果.

Fig. 12 The experiment result in the room. (a) Keyframe. (b) Robot vision. (c) The overlap region in robot vision. (d) The rebuilding result of overlap region of robot vision.

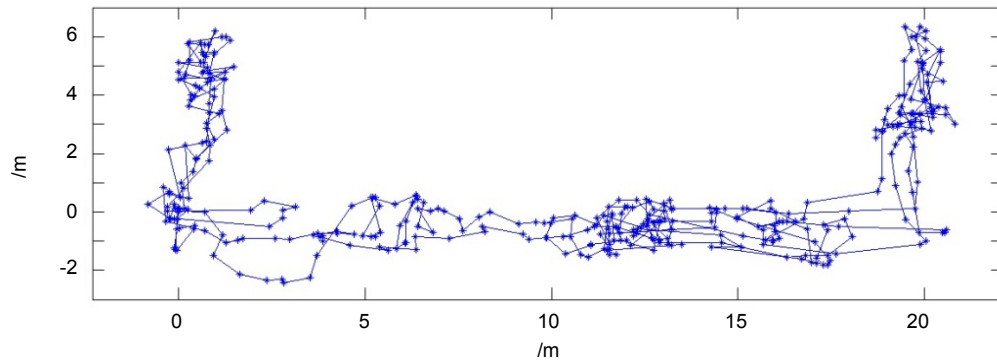


图 13 机器人匹配定位和运动路径结算结果(圆点为机器人的自主定位结果).

Fig. 13 The localization result and robot (The points are robot localization results).

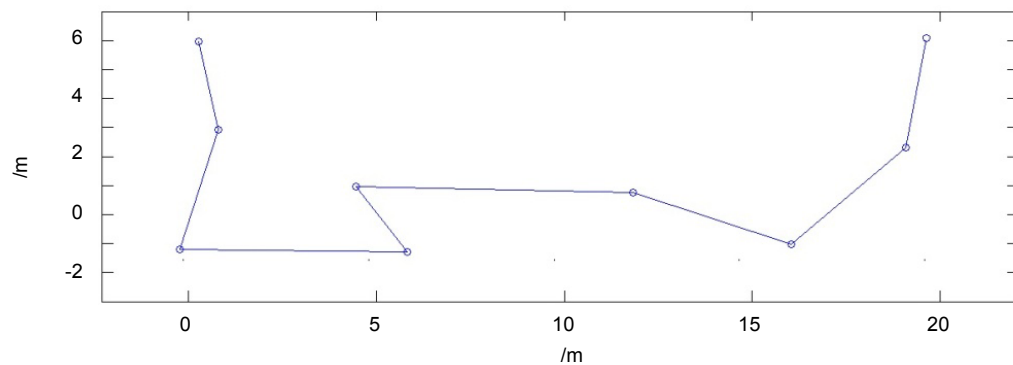


图 14 绑架环境下的机器人自主定位实验(圆圈处为机器人对绑架后位置的自主定位结果).

Fig. 14 The localization result when robot is kidnapped("o" is kidnap position when robot is landed).

4 结 论

针对机器人室内导航常出现的被绑架问题, 本文设计了基于视觉内容匹配的机器人自主定位系统, 可以使用全局地图对机器人匹配导航。该视觉定位系统的核心是视觉内容匹配, 包括图像重叠区提取、重叠区域的子块分解重建、相似度匹配三个图像处理环节, 可以有效校正重叠区中由拍摄位置和拍摄角度造成的图像内容畸变, 可以准确从全局导航地图的关键帧序列中挑选出与每幅机器人视觉图像最相似的关键帧、依据该帧对机器人进行准确定位。因为是全局地图匹配, 即使绑架中的机器人被误碰或搬移到很远的位置时, 它仍能找到此时与之最相似的关键帧, 对自己所在的绑架地点准确定位。在测试中, 机器人可以有效地在由多个屋子和走廊组成的环境中准确定位, 实验中图像匹配准确率 $\geq 95\%$, 定位精度 RMSE $< 0.5\text{m}$, 在机器人在被绑架后也可以对机器人准确定位。

基金项目

国家自然科学基金资助项目(61271147, 61372052)。

参考文献

- 1 Fabian J R, Clayton G M. Adaptive visual odometry using RGB-D cameras[C]// Proceedings of 2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Besacon, France, 2014: 1533–1538.
- 2 Xu Yunxi, Chen Fang. Real-time stereo visual localization based on multi-frame sequence motion estimation[J]. *Opto-Electronic Engineering*, 2016, **43**(2): 89–94.
许允喜, 陈方. 基于多帧序列运动估计的实时立体视觉定位[J]. *光电工程*, 2016, **43**(2): 89–94.
- 3 Wang Han, Mou Wei, Suratno H, et al. Visual odometry using RGB-D camera on ceiling Vision[C]// Proceedings of 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), Guangzhou, China, 2012: 710–714.
- 4 Lin Huican, Lü Qiang, Zhang Yang, et al. The sparse and dense VSLAM: a survey[J]. *Robot*, 2016, **38**(5): 621–631.
林辉灿, 吕强, 张洋, 等. 稀疏和稠密的 VSLAM 的研究进展[J]. *机器人*, 2016, **38**(5): 621–631.
- 5 Tateno K, Tombari F, Navab N. Real-time and scalable incremental segmentation on dense SLAM[C]// Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015: 4465–4472.
- 6 Schöps T, Engel J, Cremers D. Dense planar SLAM[C]// Proceedings of 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 2014: 157–164.
- 7 Schöps T, Engel J, Cremers D. Semi-dense visual odometry for AR on a smartphone[C]// Proceedings of 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 2014: 140–150.
- 8 Engel J, Stuckler J, Cremers D. Large-scale direct SLAM with stereo cameras[C]// Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015: 1935–1942.
- 9 Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. *IEEE Transactions on Robotics*, 2015, **31**(5): 1147–1163.
- 10 Chen Pengjin, Gu Zhaopeng, Zhang Guodong, et al. Ceiling vision localization with feature pairs for home service robots[C]// Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics, Bali, Indonesia, 2014: 2274–2279.
- 11 Choi H, Kim D Y, Hwang J P, et al. CV-SLAM using ceiling boundary[C]// Proceedings of 2010 the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA), Taichung, China, 2010: 228–233.
- 12 Choi H, Kim R, Kim E. An efficient ceiling-view SLAM using relational constraints between landmarks[J]. *International Journal of Advanced Robotic Systems*, 2014, **11**: 4.
- 13 Lee S, Lee S, Baek S. Vision-based kidnap recovery with SLAM for home cleaning robots[J]. *Journal of Intelligent & Robotic Systems*, 2012, **67**(1): 7–24.
- 14 Pfister S T, Burdick J W. Multi-scale point and line range data algorithms for mapping and localization[C]// Proceedings of 2006 IEEE International Conference on Robotics and Automation (ICRA), Orlando, America, 2006: 1159–1166.
- 15 Jeong W, Lee K M. CV-SLAM: a new ceiling vision-based SLAM technique[C]// Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada, 2005: 3195–3200.
- 16 Deng Jihong, Wei Yuxing. Location of object based on local feature descriptor[J]. *Opto-Electronic Engineering*, 2015, **42**(1): 58–64.
邓集洪, 魏宇星. 基于局部特征描述的目标定位[J]. *光电工程*, 2015, **42**(1): 58–64.
- 17 Huang Wenyong, Xu Xiangmin, Wu Fengqi, et al. Research of underwater binocular vision stereo positioning technology in nuclear condition[J]. *Opto-Electronic Engineering*, 2016, **43**(12): 28–33.
黄文有, 徐向民, 吴凤岐, 等. 核环境水下双目视觉立体定位技术研究[J]. *光电工程*, 2016, **43**(12): 28–33.
- 18 J Zhuo, L Sun, J Shi, et al. Research on a Type of Camera Calibration Method Based on High Precision Detection of X Corners[C]// Proceedings of 2015 8th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2015: 192–196.