

文章编号 1004-924X(2017)08-2221-12

结合图像内容匹配的机器人视觉 导航定位与全局地图构建系统

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 100190)

摘要:为了解决机器人室内定位时的绑架问题和相似物体的干扰,设计了一种具有图像内容匹配功能的视觉系统,从而使机器人能有效提取关键帧序列构建室内全局地图并实现自主定位。考虑影响图像内容匹配的主要干扰是机器人视角和位移造成的图像畸变,本文通过对室内物体的图像畸变建模与特征分析,设计了一种图像内容匹配方法。该方法以图像重叠区提取、基于子块分解匹配的重叠区重建两部分为核心,可将待匹配的两帧图像畸变调整为一致后再进行内容匹配并准确解算它们的相似度。其能有效利用各个房间内不同的景物和布局信息来消除相似物体的影响,从机器人学习环境时采集的视频中提取空间间距大且重叠相连的关键帧序列建立整栋建筑内部的全局导航地图。机器人工作时,实时视觉的图像内容与地图关键帧序列匹配,提取出与每个时刻视觉图像最相似的关键帧对机器人实施定位。在由 3 个房间和 2 条走廊组成的实验区进行了实验测试,结果表明:机器人可有效消除相似物体的干扰,绑架发生时仍可通过与全局地图匹配实施准确自主定位,匹配准确率 $\geq 93\%$,定位精度误差(RMSE) < 0.5 m。

关键词:机器人视觉;机器人自主定位;关键帧全局地图;图像内容匹配;图像畸变

中图分类号:TP242.6 **文献标识码:**A **doi:**10.3788/OPE.20172508.2221

Robot vision system for keyframe global map establishment and robot localization based on graphic content matching

CAO Tian-yang^{1,2}, CAI Hao-yuan^{1*}, FANG Dong-ming¹, LIU Chang¹

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

2. *University of Chinese Academy of Sciences, Beijing 100190, China)*

* *Corresponding author, E-mail: hycail@mail.ie.ac.cn*

Abstract: To solve kidnapping problem and similar object interference at the time of indoor localization of robots, a visual system with graphic contents matching function was designed to make robots extract constructed indoor global map of key frame sequence effectively and to realize self-localization. Since the main interferences influencing graphic content matching was image distortion caused by visual angels of robots and displacement, a graphic content matching method was designed by image distortion modeling and feature analysis of indoor objects. With the method, both parts of image overlap-

收稿日期:2017-01-03;修订日期:2017-03-24.

基金项目:国家自然科学基金资助项目(No. 61372052, No. 61271147)

ping area extraction and overlapping area reconstruction based on sub-block decomposition and matching were taken as a core. The content matching could be implemented after two image distortions waiting for matching were adjusted to be conformity and their similarity could be calculated accurately. The method could make use of different sceneries and arrangement information in various rooms effectively to eliminate the influence of similar objects and to extract the overlapped and connected key frame sequence with great space distance from collected video when the robots learn environment to construct global navigation map in the whole building. When robots work, graphic contents of real-time vision and key frame sequence of map was matching, and key frame that is the most similar with visual image in all moments was extracted to implement localization on robots. Experimental test was performed in an experimental area composed of 3 rooms and 2 corridors. Experimental result indicates that robots can eliminate interference of similar objects effectively and can implement accurate self-localization by matching with global map when kidnapping happens. The matching accuracy is $\geq 93\%$ and localization accuracy error (RMSE) is $< 0.5\text{m}$.

Key words: robot vision; robot self-localization; keyframes global map; graphic content matching; image distortion

1 引言

机器人视觉的一个重要研究方向是为自主移动机器人提供导航定位服务。这一功能在 GPS (General Position System) 无法有效使用的室内环境中非常有用,视觉系统可首先自主学习室内环境、自动构建室内地图,然后依据该地图自主定位^[1-2]。

在这一研究中,基于视觉的同时定位与建图技术 (Visual Simultaneous Localization And Mapping, VSLAM) 最受关注^[3]。并以 3D 信息^[4]、新型特征为基础^[5],发展了 ORB SLAM^[6]、dense SLAM^[7]、semi-dense SLAM^[8]、LSD SLAM^[9]和 CV-SLAM^[10]等多种新方法。其中, CV-SLAM 采用等高的天花板作为导航参照物,十分简洁方便且适于室内应用^[11]。SLAM 的各个核心环节也是国内目前研究的热点。针对导航特征提取,张涛等^[12]将 RGB-D 摄像机采集的点云划分为数个平面,在平面中实现高精度、高稳定性的特征点提取与匹配。徐文立等^[13]定义了对新特征多级排序的误匹配损失函数与风险函数,降低了特征点的误匹配。邹丹平等^[14]从建筑结构边界线中提取导航特征。针对回环检测,张涛等^[15]采用深度神经网络对图像按照所包含的景物结构特征进行编码识别和构建回环。针对复杂的使用环境,杜航原等^[3]提出基于概率假设密度

滤波的 SLAM 算法,在观测噪声强度较大的情况下仍能有效地导航定位。

但 SLAM 也存在许多局限性,最大的问题是易“被绑架”和“相似物体”干扰。发生绑架时,机器人被碰撞或意外滑动到较远的新位置,脱离了先前定位信息的覆盖范围,无法利用这些信息估计和修正自己被绑架到的新位置,导致机器人迷失^[16-17]。为了解决这种问题,一些算法(如 ORB SLAM)通过特征匹配实施在全局地图中的机器人定位,但室内不同屋子中包含了很多相似物体,常用的匹配方法大多只利用了图像的特征点和直方图信息^[18-19](如像素法、全局直方图、局部直方图、BoW 等),它们无法有效区分相似物体上的特征,极易让机器人混淆,从而很容易造成机器人误匹配定位。

因此,本文设计了一种能够消除图像畸变影响的图像内容匹配方法,对传统的 SLAM 进行改进、实现基于图像内容的 SLAM (Graphic Content SLAM, GC-SLAM)。图像内容匹配可以代替 SLAM 中常用的特征匹配,使机器人的视觉系统能充分利用各个房间和走廊中的不同景物和布局信息以实现整栋建筑内部的全局地图构建与机器人自主导航定位。首先,针对由机器人视角和位移造成的图像畸变干扰问题,建立了天花板和墙体(包括墙、家具、门、窗)的图像畸变模型、提取畸变特征,构建了由重叠区提取、子块分解匹配的重叠区重建为核心的图像内容匹配算法,将待匹

配图像中的畸变调整为近似一致后匹配,准确实现了不同视角、不同位置的任意两帧图像的内容相似度比较,并设计了误匹配子块检测方法以消除相似物体的干扰、提高相似度的解算精度。然后,机器人视觉系统使用该图像内容匹配方法可以从室内环境的视频中提取出空间间距大、图像内容差异大的关键帧序列建立室内全局地图。在“多个房间+走廊”这种大型的复杂室内环境中只需提取几十幅关键帧即可完成全局地图的构建。工作中的机器人将实时视觉图像内容与关键帧序列地图一一匹配,找出与每一时刻视觉图像最相似的关键帧,即可准确地自主导航定位。

测试中,在实验区域(3个 15 m² 房间、1条 13 m 走廊、1条 7 m 走廊)内机器人拍摄了 882 帧环境视频,视觉系统中筛选出 72 幅关键帧来构建全局导航地图,有效地为机器人提供了导航定位服务和解算运动路线。同时,该系统具有很强的抗绑架能力,机器人被移动到任意位置都能通过全局地图准确地自主定位。测试中,关键帧匹配准确率≥93%,定位的均方根误差<0.5 m (Root Mean Square Error, RMSE)。

2 基于视觉内容的图像匹配

2.1 图像畸变建模与特征分析

本文的机器人和它的视觉系统如图 1 所示。采用一台朝上拍摄的 150° 广角摄像机^[20-21],可以使机器人视觉系统具有非常大的视场,能够拍摄包括天花板和大部分墙体区域(门窗、墙、家具)在内的室内景物。机器人以不同房间的不同景物为参照物,与自己的实时视觉进行匹配即可准确自

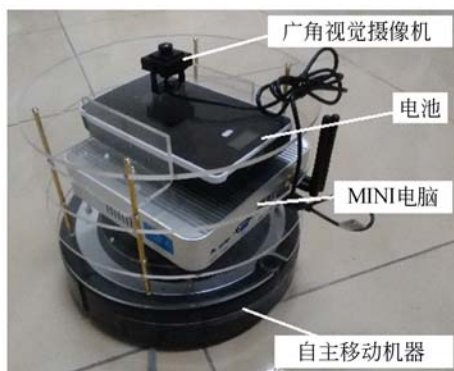


图 1 机器人和它的视觉系统

Fig. 1 Robot and its vision navigation system

主定位。

但这些景物在机器人视觉和地图中通常具有不同的拍摄视角,会出现不同程度的畸变,严重干扰了匹配的准确度。按照摄像机的视角,可将室内物体分为两类:正对摄像机的天花板区域与摄像机侧视的墙体区域(包括墙、家具、门、窗等),如图 2 所示。受不同视角的影响,这些景物将呈现不同的图像畸变。而且随着机器人的移动,这些视角造成的畸变程度也会持续改变。景物的这种图像畸变与机器人视角和位移之间的关系可通过摄像机的内外参数模型描述^[22-23]:

$$z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & 0 \\ 0 & \frac{1}{dy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{A} \quad \mathbf{T}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (1)$$

式中,(X,Y,Z)是景物中的空间中坐标,(x,y)是它在视觉图像中的坐标、与机器人的视角和位移相关,

$$\mathbf{A} = \begin{bmatrix} \cos R & 0 & -\sin R \\ 0 & 1 & 0 \\ \sin R & 0 & \cos R \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos P & \sin P \\ 0 & -\sin P & \cos P \end{bmatrix}$$

$$\begin{bmatrix} \cos H & \sin H & 0 \\ -\sin H & \cos H & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{T} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix},$$

由摄像机视角(横滚 R、俯仰 P、航向 H)和机器人的位移 T_x、T_y、T_z 构成(T_z 为天花板高度),f、d_x、d_y 是摄像机的参数^[21]。

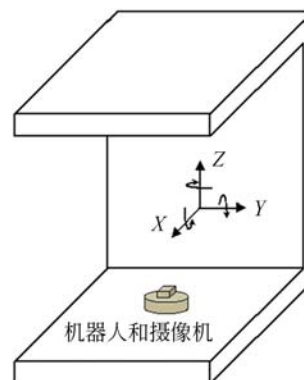


图 2 墙体区域和机器人摄像机的关系

Fig. 2 Relationship between wall and robot camera

天花板平行于机器人运动的地面,机器人的视觉系统只有 3 个参数可以自由改变:拍摄时的航向

角 H 和机器人的位移 T_x 、 T_y ，而俯仰和横滚都是 0。将航向角和位移带入公式(1)可以进一步推导出两帧图像中相同景物上的点具有以下关系：

$$\begin{aligned} x_B &= \cos H \cdot x_A + \sin H \cdot y_A + \frac{f}{dx} T_x, \quad (2) \\ y_B &= -\sin H \cdot x_A + \cos H \cdot y_A + \frac{f}{dy} T_y, \quad (3) \end{aligned}$$

式中， (x_A, y_A) 是帧 A 中天花板上的点，它在帧 B 中的对应点坐标是 (x_B, y_B) ，这两个坐标之间的数学关系为图像仿射变换^[24]。因此对于两幅照片中的天花板图像，只需进行逆向平移 T_x 、 T_y 和旋转 H ，即可将它们调整至相同的拍摄位置和相同的方向进行图像匹配。两帧天花板图像是否匹配可以通过计算它们的相关系数进行分析。

但对于墙体区域，相当于将天花板绕墙壁与天花板的交线转动 90° 后进行拍摄，如图 1 所示，在航向变化和位移的基础上叠加了 90° 的拍摄姿态变化。以机器人前方的墙体为例，墙体相当于

$$\begin{pmatrix} x_{AB} \\ y_{AB} \end{pmatrix} = \begin{pmatrix} \cos \Delta H & \sin \Delta H \\ -\sin \Delta H & \cos \Delta H \end{pmatrix} \begin{pmatrix} \frac{f}{dx} \frac{X \cos H + Y \sin H + T_x}{X \sin H - Y \cos H + T_z} \\ \frac{f}{dy} \frac{Z + T_y}{X \sin H - Y \cos H + T_z} \end{pmatrix} + \begin{pmatrix} \Delta t_x \\ \Delta t_y \end{pmatrix}, \quad (8)$$

式中， (x_{AB}, y_{AB}) 为帧 A 中变换后的点坐标。但公式(8)尚未实现将图 A 墙壁中的点平移到与图 B 对应的位置、还有一定的差距。因此还需要进行第二次平移 S_x 、 S_y ：

$$x_{AB} + S_x = \frac{f}{dx} \frac{X \cos(H + \Delta H) + Y \sin(H + \Delta H) + T_x + \Delta t_x}{X \sin H - Y \cos H + T_z}, \quad (9)$$

$$y_{AB} + S_y = \frac{f}{dy} \frac{Z + T_y + \Delta t_y}{X \sin H - Y \cos H + T_z}. \quad (10)$$

相比于原始的帧 A，第二次平移后的帧 A 与帧 B 非常接近。 S_x 、 S_y 的解算公式为：

$$\begin{aligned} S_x &= \cos \Delta H \frac{f}{dx} \frac{-T_x}{X \sin H - Y \cos H + T_z} + \\ &\sin \Delta H \frac{f}{dx} \frac{-X \sin H + Y \cos H - Z - T_y}{X \sin H - Y \cos H + T_z} + \\ &\frac{f}{dx} \frac{\Delta t_x}{X \sin H - Y \cos H + T_z} - \Delta t_x, \quad (11) \end{aligned}$$

$$\begin{aligned} S_y &= \sin \Delta H \frac{f}{dy} \frac{X \cos H + Y \sin H + T_x}{X \sin H - Y \cos H + T_z} + \\ &\cos \Delta H \frac{f}{dy} \frac{-Z - T_y}{X \sin H - Y \cos H + T_z} + \\ &\frac{f}{dy} \frac{\Delta t_y + Z + T_y}{X \sin H - Y \cos H + T_z} - \Delta t_y. \quad (12) \end{aligned}$$

俯仰旋转 90° ，若机器人拍摄帧 A 时航向角为 H° 、位移 T_x 、 T_y ，则根据公式(1)可推导出墙体上的点 (X, Y, Z) 在图像中的位置 (x_{Aw}, y_{Aw}) 的计算公式为：

$$x = \frac{f}{dx} \frac{X \cos H + Y \sin H + T_x}{X \sin H - Y \cos H + T_z}, \quad (4)$$

$$y = \frac{f}{dy} \frac{Z + T_y}{X \sin H - Y \cos H + T_z}. \quad (5)$$

令机器人拍摄帧 B 时的参数为 $H + \Delta H$ 、 $T_x + \Delta t_x$ 、 $T_y + \Delta t_y$ ，则帧 B 表达式为：

$$x = \frac{f}{dx} \frac{X \cos(H + \Delta H) + Y \sin(H + \Delta H) + T_x + \Delta t_x}{X \sin(H + \Delta H) - Y \cos(H + \Delta H) + T_z}, \quad (6)$$

$$y = \frac{f}{dy} \frac{Z + T_y + \Delta t_y}{X \sin(H + \Delta H) - Y \cos(H + \Delta H) + T_z}. \quad (7)$$

若对帧 A 旋转 ΔH 和平移 Δt_x 、 Δt_y ，它与帧 B 中相同的天花板图像会被调整至同一位置，但帧 A 中的墙体图像在处理后会发生如下变化：

为了减小公式(11)、(12)与公式(9)、(10)中由分母差异造成的影响，同时考虑到公式(11)、(12)中有许多参数未知(如 X 、 Y 、 Z)，本文将采用子块匹配重建的方法将图像分为数个小区域分别求取每个区域的平均 (S_x, S_y) 。对于机器人侧边的墙体，可以认为是墙体从天花板位置绕墙壁与天花板的交线横滚旋转 90° ，仍具有上述的性质，可以采取相同的方法分析和匹配。

若待匹配的图像包含相同的内容，则能够正确匹配的子块数量会非常多，匹配后的图像相关系数会非常大。对于天花板区域，因为无俯仰和横滚视角变化、无畸变，采用上述的平移方法，子块会匹配得更加准确，匹配的子块数量也会非常多。如果待匹配图像的内容不相同，则可以匹配的子块数量非常少，匹配后的图像中会有大量空洞且相关系数会非常低。因此将相关系数与匹配子块数量相结合，可以准确地评估两帧图像的相似度。

2.2 图像畸变校正与图像内容匹配

如果帧 A、B 都拍到一些相同的图像内容，则它们具有拍摄重叠区，图像内容匹配将在该区域

内进行。

因此对于帧 A、B,图像内容匹配方法如图 3 所示,主要包括两部分:(1)提取帧 A、B 之间的重叠区;(2)基于子块分解匹配的重叠区重建,将帧 A 重叠区划分为多个子块与帧 B 进行匹配,利用帧 B 的信息对帧 A 重叠区重建。比较帧 A 重建重叠区和帧 B 重叠区即可评估两帧的相似度。

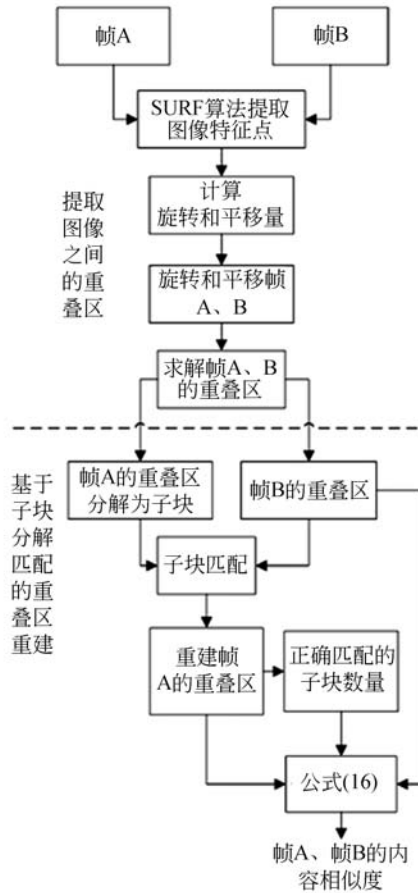


图 3 图像内容的匹配过程

Fig. 3 Process of graphic content matching

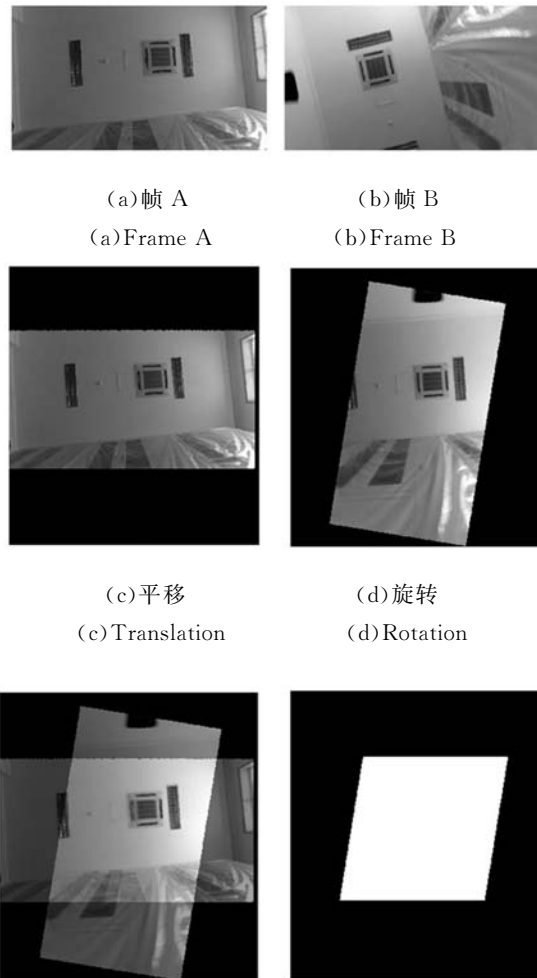
2.2.1 提取图像之间的重叠区

本文首先根据帧 A、B 中的特征点将它们调整到相同的拍摄位置和航向角,获取二者可能的重叠区,然后分析二者在重叠区中的图像内容相似度。

首先提取匹配帧 A、B 的特征点(SIFT 或 SURF 方法)^[25-26],带入公式(2)(3)通过最小二乘法求取航向和平移量^[27],然后旋转和平移帧 A、B。天花板上的特征点因为高度相同,求取航向和平移量会十分准确。但如果墙壁上的点也被提取为特征点,因为它们高度不同,求取的航向和平

移量将有误差,这种误差将在下边的“基于子块分解匹配的重叠区重建”环节消除。

如图 4 所示,经过图 4(c)、(d)的旋转和平移,两帧的重叠效果在图 4(e)中十分明显。将同时包含帧 A、B 图像点的区域提取出来做成掩膜(图 4(f)),即可提取两幅图像可能的重叠区。



(e)两帧的重叠区域 (f)重叠区的掩膜
(e)Overlap region (f)Mask of overlap region
between two frames

图 4 图像之间的重叠区过程

Fig. 4 Extraction progress of overlap region

对于图像中的天花板区域,若包含相同的景物,经过旋转平移后,它们在帧 A、B 重叠区中的位置相同(如图 4(e)所示),求取重叠区的相关系数即可评估帧 A、B 的图像内容相似度。但对于重叠区中的墙体图像,考虑到景物形状发生了畸变,本文设计了基于子块分解匹配的重叠区重建方法对其处理,进行第二次平移修正畸变。

2.2.2 基于子块分解匹配的重叠区重建

该方法可代替含有未知参数的公式(11)和(12)求取平移量 S_x, S_y 。如图 5 所示,分割帧 A 重叠区为多个子块,分别与帧 B 重叠区匹配。平移后,尽管两帧图像的子块中仍存在细微的景物畸变,但只要两帧的重叠区域包含了相同的图像内容,帧 A 的每个子块都可以在帧 B 中匹配到适合的位置,帧 A 重叠区重建后将与帧 B 重叠区非常相似。

采用绝对误差求和的方法 SAD(Sum of Absolute Difference)解算帧 A 的各个子块在帧 B 中的匹配位置:

$$V(i, f) = \arg \min_{\substack{i \in [a, b] \\ j \in [a, p]}} \sum_{s=1}^M \sum_{t=1}^N | \mathbf{B}_L(i+s-1, j+t-1) - \mathbf{A}_T(s, t) |, \quad (13)$$

式中, \mathbf{A}_T 是帧 A 重叠区中的一个子块(尺寸 $M \times N$), \mathbf{B}_L 是帧 B 的重叠区(尺寸 $L \times D$)。遍历 \mathbf{B}_L , 使公式(13)的值最小的 (i, j) 就是子块 \mathbf{A}_T 的平移量 S_x, S_y 。帧 A 重叠区的每个子块都采用这种方法进行处理后,即可获得一幅帧 A 重叠区的重建图,如图 5(e)所示。虽然“提取图像之间的重叠区”环节会出现由旋转和平移误差造成的重叠区未对准问题,但经过子块匹配平移,可以基本消除这些误差的影响。

尽管平移后,两帧图像的子块中仍存在细微的景物畸变,但子块重建的视频帧重叠区与关键帧重叠区已经十分相似了。

考虑到可能由相似物体造成的误匹配,本文根据子块平移量 (S_x, S_y) 的特点,设计了一种检测误匹配子块的方法。

将 (S_x, S_y) 的解算公式(11)(12)变换为矩阵形式后,可以清晰显示出对图像畸变的修正作用:

$$\begin{aligned} \begin{bmatrix} S_x \\ S_y \end{bmatrix} &= -\frac{f}{X \sin H - Y \cos H + T_z} \\ \begin{bmatrix} dx & 0 \\ 0 & dy \end{bmatrix} \begin{bmatrix} \cos \Delta H & \sin \Delta H \\ -\sin \Delta H & \cos \Delta H \end{bmatrix} \begin{bmatrix} T_x \\ Z + T_y \end{bmatrix} &+ \\ \frac{f}{X \sin H - Y \cos H + T_z} \begin{bmatrix} dx & 0 \\ 0 & dy \end{bmatrix} & \\ \begin{bmatrix} \cos(-H) & \sin(-H) \\ -\sin(-H) & \cos(-H) \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \sin \Delta H &+ \\ \frac{f}{X \sin H - Y \cos H + T_z} \begin{bmatrix} dx & 0 \\ 0 & dy \end{bmatrix} & \\ \begin{bmatrix} \Delta t_x \\ \Delta t_y + T_y + Z \end{bmatrix} - \begin{bmatrix} \Delta t_x \\ \Delta t_y \end{bmatrix}. & \quad (14) \end{aligned}$$

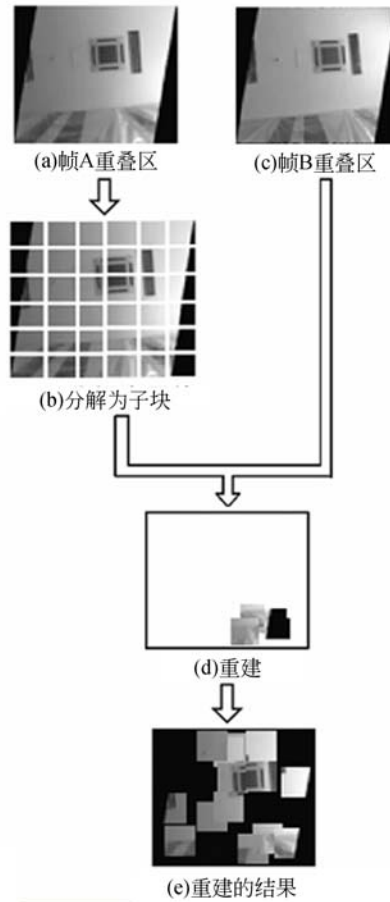


图 5 重叠区重建的图像处理过程

Fig. 5 Progress of overlap region rebuilding

公式(14)表明, (S_x, S_y) 可以修正墙体景物(包括墙、家具、门窗等)的图像畸变:

(1) 机器人旋转造成的畸变。

机器人视觉中,墙体上高度不同的区域与旋转中心距离不同,旋转相同的 ΔH 会导致这些区域在照片中发生不同的位移、造成景物变形,公式(14)第一项的高度参数 Z 与 ΔH 表征这种变形。同时由于该项还包含了水平面坐标 (X, Y) , 因此它表征了 (X, Y, Z) 三个参数在机器人旋转时对图像畸变的影响。公式(14)第二项的核心是 $\begin{bmatrix} \cos(-H) & \sin(-H) \\ -\sin(-H) & \cos(-H) \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix} \sin \Delta H$, 是重叠区提取时由旋转 ΔH 的投影 $\sin \Delta H$ 造成的位移,它不包含高度 Z 。因此二者相减的关系表明,子块匹配能够修正重叠区提取后残留在图像中由旋转 ΔH 和高度造成的墙体畸变。

(2) 机器人平移造成的畸变。

平移导致景物发生相对移动,当机器人侧视

墙体时,这种移动还与景物各部分距离地面的高度相关,越低、移动越大,导致景物的不同位置发生不同程度的变形,公式(14)第三项中的高度参数 Z 表征这种变形畸变。公式(14)第四项是重叠区域提取时的平移量。因此二者的相减关系表明,子块匹配能够修正重叠区提取后残留在图像中由机器人移动和高度造成的墙体畸变。

因此,针对 (S_x, S_y) 的这种特点,本文设计了一种能够自动调整的动态阈值,滤除由相似物体造成的子块误匹配。由于帧 A、B 重叠区提取后残余的旋转差异和平移差异都很小,相比之下相似物体位于两帧的不同位置、所需的子块移动量会很大,因此通过为旋转和平移设定阈值即可识别和滤除相似物体造成的误匹配。考虑到景物的高度也在机器人视觉中变为与照片中心的距离(高度越低、距离越远),而旋转会因景物离旋转中心的不同而造成平移量不同、因此通过仿射变换将每个子块与照片中心的相对位置 (x, y) 、平移和旋转阈值 $\Delta x_T, \Delta y_T, \Delta H_T$ 结合在一起,可以构建识别误匹配子块的动态阈值 $\Delta P_x, \Delta P_y$:

$$\begin{bmatrix} \Delta P_x \\ \Delta P_y \end{bmatrix} = \begin{bmatrix} \cos \Delta H_T & \sin \Delta H_T \\ -\sin \Delta H_T & \cos \Delta H_T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x_T \\ \Delta y_T \end{bmatrix}. \quad (15)$$

若位于 (x, y) 的子块平移量 S_x, S_y 的绝对值大于阈值 $\Delta P_x, \Delta P_y$, 则判定为是误匹配、予以删除。公式(15)可对阈值自动调整,在中心小、在边界大,可以把相似物体误匹配的子块滤除、并确保不遗漏每个匹配正确的子块。删除误匹配子块之后,帧 A 重叠区的重建结果与帧 B 重叠区十分相似,如图 6 所示。



(a)帧 A 重叠区 (b)帧 B 重叠区 (c)子块分解重建的帧 A 重叠区
(a)Overlap region of frame A (b)Overlap region of frame B (c)Sub-blocks rebuild overlap region of frame A

图 6 重叠区重建

Fig. 6 Overlap region rebuilding

若两帧图像内容不相同,则帧 A 重建重叠区

与帧 B 会相差甚远,如图 7 所示。因此基于子块分解匹配的重叠区重建的方法可以很容易评估两幅图像是否包含相同的内容。

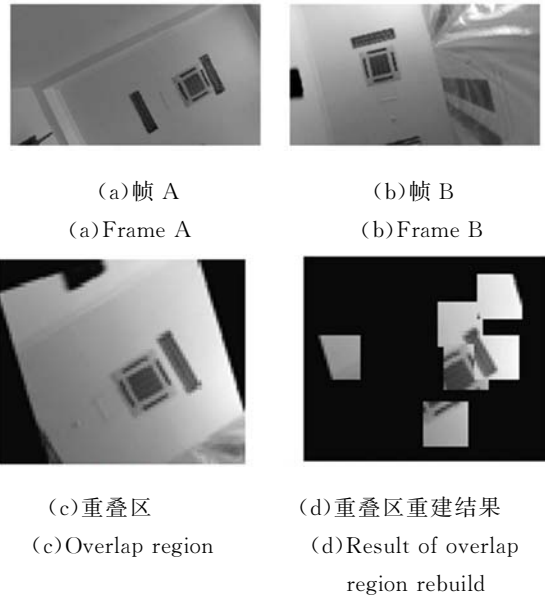


图 7 不相似帧的从重叠区重建结果

Fig. 7 Overlap region rebuild of dissimilar frames

2.2.3 待匹配图像的相似度解算

解算帧 A、B 的相似度可以通过比较帧 A 重建重叠区和帧 B 重叠区的相似度 S_{AB} 实现。 S_{AB} 由表征图像之间相似程度的相关系数、及能够正确匹配子块数量 N_k 组成,是二者的乘积^[28-29]:

$$S_{AB} = N_k \cdot \frac{\sum_{xy} (A(x, y) - \bar{A})(B(x, y) - \bar{B})}{\sqrt{\sum_{xy} (A(x, y) - \bar{A})^2 \sum_{xy} (B(x, y) - \bar{B})^2}}, \quad (16)$$

式中, $A(x, y)$ 是帧 A 重建重合区中的像素值, $B(x, y)$ 是帧 B 重合区中的像素值, \bar{A} 和 \bar{B} 是两帧的像素均值。

3 基于图像内容的关键帧全局地图建立与机器人自主导航

机器人自主导航需具备两项功能:自主建立地图和根据地图自主导航。首先,通过本文的视觉内容匹配方法,机器人自主拍摄和学习室内环境,分析室内环境视频中各帧之间的图像内容相似度、从中挑选出图像内容差异大、数量少且仍有部分区域重叠相连的关键帧序列构建室内全局地图^[30]。然后,在后续的自主运动工作中,机器人

视觉系统把实时采集的视觉图像与关键帧序列进行内容匹配,通过检测出与每一时刻机器视觉最相似的关键帧、实现机器人自主定位。

3.1 基于视觉内容的机器人全局导航地图构建

构建关键帧序列的室内全局地图的算法流程如图 8 所示。

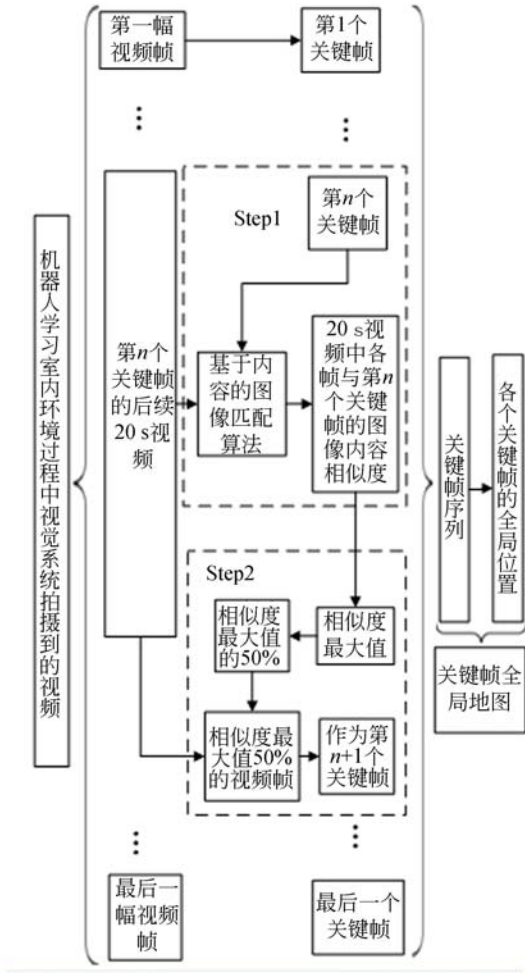


图 8 基于图像内容的关键帧全局地图构建过程

Fig. 8 Process of keyframe global map establishing according to graphic content

以室内环境视频的第一帧视频作为第一幅关键帧,关键帧序列中后续的关键帧提取过程如下:

Step1,对第 n 个关键帧,与后续 20 s 内视频(60 帧)逐帧匹配,解算它们与该关键帧的内容相似度;

Step2,找出相似度的最大值,最大值对应的视频帧与第 n 个关键帧的距离最近。取相似度降到 50% 的视频帧作为第 $n+1$ 个关键帧。若这 60

帧的相似度都大于 50%,则取最后一帧作为第 $n+1$ 个关键帧。50% 的相似度可以确保第 n 和 $n+1$ 个关键帧之间仍有部分区域重叠相连;

重复 Step1~2,提取关键帧序列。建立的关键帧全局地图包括两部分:关键帧组成的序列、及各个关键帧在室内的全局位置。

全局位置的解算过程为:对于第 $n+1$ 个关键帧,它与第 n 个关键帧重叠相连,用 SURF 算法求取两帧中的多对特征点,利用公式(2)、(3)解算它与第 n 个关键帧之间的相对位置(位置差 $x_{n+1,n}$, $y_{n+1,n}$ 、航向差 $H_{n+1,n}$),然后按照公式(17)进行累加即可得到第 $n+1$ 个关键帧的全局位置:

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} \cos H_{n+1,n} & \sin H_{n+1,n} \\ -\sin H_{n+1,n} & \cos H_{n+1,n} \end{pmatrix} \begin{pmatrix} x_{n+1,n} \\ y_{n+1,n} \end{pmatrix} + \begin{pmatrix} x_n \\ y_n \end{pmatrix}, \quad (17)$$

式中, (x_n, y_n) 、 (x_{n+1}, y_{n+1}) 是第 n 、 $n+1$ 帧的全局位置。按公式(17)逐帧迭代,即可求取关键帧序列中各帧在室内的全局位置。

该方法提取的关键帧数量少、间距大、内容独立性大,为机器人的快速匹配和实时定位奠定了基础。在实验中,机器人在环境学习时拍摄了 5 分钟约 882 帧图像,视觉系统从中提取了关键帧 72 幅。

3.2 全局地图中的机器人自主定位

随后机器人的自主工作过程中,实时视图

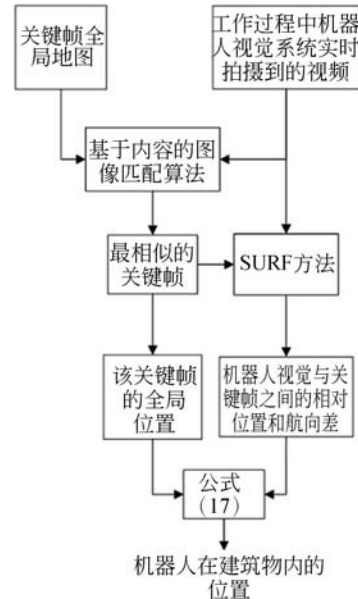


图 9 依据全局地图的机器人定位过程

Fig. 9 Robot localization through global map

像与地图中的关键帧序列通过图像内容进行匹配,找出和机器人当前视觉图像最相似的关键帧,通过该帧确定当前时刻自己在哪间屋子(或走廊)、及在屋中(或走廊)的精确位置。精确位置的解算过程如图 9 所示,首先,提取该视频帧和与它最相似关键帧之间的特征点,带入公式(2)和(3),求取它们的相对位置。然后带入公式(17),通过该关键帧的全局位置及它与机器人实时视觉图像之间的相对位置,求取机器人在当前时刻的全局位置。

4 实验和讨论

本文的机器人视觉导航定位与全局地图构建系统已经在大型复杂的建筑物内进行了测试。实验区域包括 3 个 15 m^2 房间及连接它们的 2 条走廊(1 条 13 m 、1 条 7 m),如图 10(a)、(b)所示。为了更好地测试绑架过程,如图 10(c),采用具有 SLAM 导航的塔米扫地机器人进行绑架实验对比。



(a)实验中的机器人 (b)机器人视觉中的图像 (c)绑架对比实验中的 SLAM 扫地机器人
(a)Robot in the experiment (b)Frame taken by robot (c)SLAM sweeping robot for comparison experiment

图 10 实验区域

Fig. 10 Experiment sites

机器人在环境学习时拍摄了 882 帧图像。本文通过图像内容匹配提取了 72 幅关键帧构建导航地图。这些关键帧按照它们的全局位置拼接(如图 11),可比较全面地描述室内的环境状况(每间屋子、每段走廊中的各种家具、门窗及它们的布局),如图 12 所示。选择一些地面瓷砖格的交叉点作为路标,建立地图时遥控机器人穿过这些路标,运动轨迹点的解算结果与这些路标真实位置之间的偏差作为地图的精度误差,实验中地图误差(RMSE) $<0.3\text{ m}$ 。

地图使用时,关键帧序列逐帧与机器人视觉

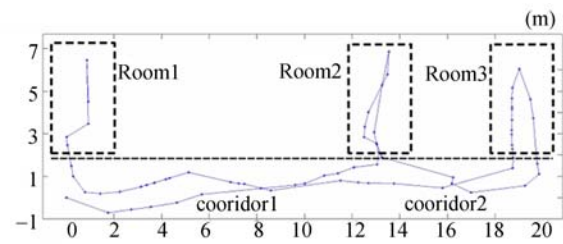


图 11 关键帧序列的全局位置关系

Fig. 11 Global position relationship of keyframe sequences



图 12 关键帧序列对实验区域的环境描述

Fig. 12 Indoor environment described by keyframes

匹配、对机器人定位,同时可实时解算出机器人的行驶轨迹。因为预先自主建立了全局导航地图,因此可以把定位误差控制在一定范围内,定位误差不会像 SLAM 算法一样随着运动距离及运动时间的增加而大幅增大。实验中,通过逐帧检查和统计匹配结果得到关键帧匹配的准确率 $\geq 93\%$ 。定位精度误差则是利用地面瓷砖格的交叉点作为路标,遥控机器人穿过这些路标,将解算出的运动轨迹点与这些路标真实位置之间的偏差作为定位精度误差,实验中机器人定位误差(RMSE) $<0.5\text{ m}$ 。

实验中的机器人行驶速度为 0.2 m/s ,算法平均耗时为 2.5 s/F ,每次定位完成时机器人运动距离不超过 0.5 m ,对于以整个楼层为运动区域的机器人这种延迟造成的 0.5 m 偏差相对还是比较小。当机器人采用小于 0.2 m/s 的速度运动时,定位算法延时造成的定位偏差会更小、实时性会比较好。

相比于传统的 SLAM 方法,本文应用图像内容匹配设计的全局地图视觉导航系统在消除绑架问题方面具有很大的优势。机器人在发生绑架时的自主定位结果如图 14 所示,由实验人员通过突然移动机器人实施绑架实验,每次绑架时的移动

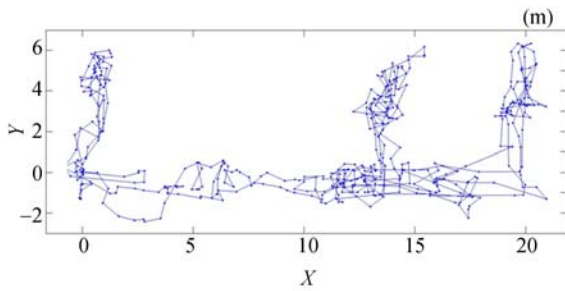


图 13 基于关键帧全局地图和图像内容匹配定位解算机器人的运动轨迹

Fig. 13 Robot route resolved through keyframes global map and content-based graphic matching

距离为 3~5 m, 机器人被绑架到新的位置后仍能够通过全局地图准确解算出自己的位置。

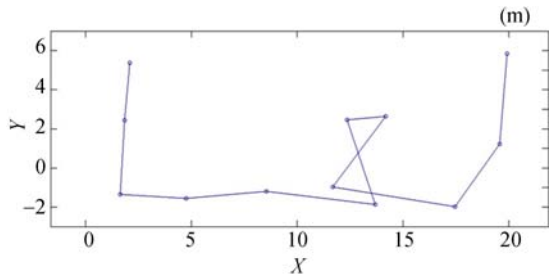


图 14 机器人被绑架时的自主定位结果(“o”为对绑架的定位结果)

Fig. 14 Localization result when kidnapped robot landing(“o” is localization for each kidnap position)

采用塔米扫地机器人实现了与传统 SLAM 绑架的对比试验。正常工作时, 具有 SLAM 导航功能的塔米扫地机器人以充电桩为起点开始运动, 完成扫地后能够自动导航返回起点处。发生绑架时, 扫地机器人将无法返回起点, 如图 15 所示。为了清晰显示绑架过程, 遥控本文的视觉全局定位机器人跟随塔米机器人同步运动、解算塔米机器人被绑架时的运动轨迹, 图 15 的粗、细线分别表示前进和返回时的轨迹。绑架过程为: 在机器人运动到 5 m 处时, 实验人员突然将它移动到了 8 m 处。尽管该机器人通过陀螺等传感器仍能保持航向继续前进、继续清扫房间, 但前几个时刻观测到的特征点路标都突然从机器人视觉图像中消失了、SLAM 不能在绑架结束时对机器人重新全局定位, 不能解算出绑架造成的这段位移, 机器人无法返回起点, 最后的终点如图 15 所示。

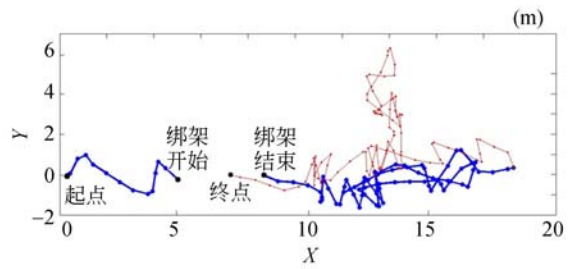


图 15 SLAM 方法遇到的绑架问题(粗、细线分别表示前进和返回时的轨迹)

Fig. 15 Kidnap problem for SLAM (Thick line is forward trajectory, and thin line is return trajectory)

本文方法与目前研究热点 ORB-SLAM 算法的性能对比如表 1 所示。在 20 m 的房间处发生了绑架, 尽管 ORB-SLAM 在这种情况下可以实施全局匹配定位, 但因为匹配基于 ORB 特征点实现, 很容易受到不同屋子中相似物体的干扰, 将机器人误匹配定位到与之相似的 13m 处的房间, 造成表 1 中出现比较大的误差。而本文的方法可以通过图像内容匹配消除这种问题, 仍能在 20 m 处的房间内对机器人准确定位。

表 1 本文方法与 ORB-SLAM 的定位误差 RMSE 对比

Tab. 1 RMSE comparison between our method and ORB-SLAM (m)

实验子区域	本文方法	ORB-SLAM
0 点处的房间	0.375	0.276
0-13 m 处的走廊	0.411	0.337
13 m 处的房间	0.459	0.42
13-20 m 处的走廊	0.435	0.487
20 m 处的房间	0.472	7.23

5 结 论

针对目前室内视觉导航与地图建立技术中存在诸如绑架、相似物体干扰等问题, 本文研制了一种结合图像内容匹配的机器人视觉导航定位与全局地图构建系统。它的核心是基于视觉内容的图像匹配, 通过利用室内各种物体的形状、布局, 可有效对待匹配图像之间的相似度进行分析。这种匹配方法能够从室内环境的视频中提取数量少且准确的关键帧序列并构建室内全局地图, 从而可

使机器人实现准确的实时全局定位。实验中机器人可以较准确自主定位,误差(RMSE) $<0.5\text{ m}$,

不会随着机器人运动范围及运动时间的增加而累积增大,并且能够使机器人有效消除被绑架问题。

参考文献:

- [1] BRAND C, SCHUSTER M J, HIRSCHM? LLER H, et al. Submap matching for stereo-vision based indoor/outdoor SLAM[C]. *Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2015: 5670-5677.
- [2] 温丰, 柴晓杰, 朱智平, 等. 基于单目视觉的SLAM算法研究[J]. *系统科学与数学*, 2010, 30(6): 827-839.
WEN F, CHAI X J, ZHU ZH P, et al. A visual slam algorithm based on monocular vision [J]. *Journal of Systems Science and Mathematical Sciences*, 2010, 30(6): 827-839. (in Chinese)
- [3] 杜航原, 郝燕玲, 赵玉新, 等. 用概率假设密度滤波实现同步定位与地图创建[J]. *光学精密工程*, 2011, 19(12): 3064-3073.
DU H Y, HAO Y L, ZHAO Y X, et al. Implementation of SLAM by probability hypothesis density filter[J]. *Opt. Precision Eng.*, 2011, 19(12): 3064-3073. (in Chinese)
- [4] 付梦印, 吕宪伟, 刘彤, 等. 基于RGB-D数据的实时SLAM算法[J]. *机器人*, 2015, 37(6): 683-692.
FU M Y, LV X W, LIU T, et al. Real-time SLAM algorithm based on RGB-D data[J]. *Robot*, 2015, 37(6): 683-692. (in Chinese)
- [5] GARCIA-FIDALGO E, ORTIZ A. Vision-based topological mapping and localization by means of local invariant features and map refinement [J]. *Robotica*, 2015, 33(7): 1446-1470.
- [6] 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.
- [7] 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)*, IEEE, 2015: 4465-4472.
- [8] KIM J H, CADENA C, REID I. Direct semi-dense SLAM for rolling shutter cameras[C]. *Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2016: 1308-1315.
- [9] BOIKOS K, BOUGANIS C S. Semi-dense SLAM on an FPGA SoC[C]. *Proceedings of the 26th International Conference on Field Programmable Logic and Applications (FPL)*, IEEE, 2015: 4465-4472.
- [10] CHEN P J, GU Z P, ZHANG G D, et al. Ceiling vision localization with feature pairs for home service robots[C]. *Proceedings of 2014 IEEE International Conference on Robotics and Biomimetics*, IEEE, 2015: 2274-2279.
- [11] JO S, CHOI H, KIM E. Ceiling vision based SLAM approach using sensor fusion of sonar sensor and monocular camera[C]. *Proceedings of the 12th International Conference on Control, Automation and Systems (ICCAS)*, IEEE, 2012: 1461-1464.
- [12] GAO X, ZHANG T. Robust RGB-D simultaneous localization and mapping using planar point features [J]. *Robotics and Autonomous Systems*, 2015, 72: 1-14.
- [13] 刘志斌, 吴显亮, 徐文立, 等. 视觉SLAM中基于误匹配风险预测的特征选择[J]. *机器人*, 2010, 32(5): 635-641.
LIU ZH B, WU X L, XU W L, et al. Erroneous matching risk prediction based feature selection for Visual SLAM[J]. *Robot*, 2015, 32(5): 635-641. (in Chinese)
- [14] ZHOU H Z, ZOU D P, PEI L, et al. StructSLAM: visual SLAM with building structure lines[J]. *IEEE Transactions on Vehicular Technology*, 2015, 64(4): 1364-1375.
- [15] GAO X, ZHANG T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. *Autonomous Robots*, 2017, 41(1): 1-18.
- [16] 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.
- [17] YI C, CHOI B U. Detection and recovery for kidnapped-robot problem using measurement entropy[C]. *Communications in Computer and In-*

- formation Science, Springer, 2011, 261: 293-299.
- [18] STEWART R L, ZHANG H. Image similarity from feature-flow for keyframe detection in appearance-based SLAM[C]. *Proceedings of 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, 2011: 305-312.
- [19] KIM J, KWEON I S. Robust feature matching for loop closing and localization[C]. *Proceedings of 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2007: 3905-3910.
- [20] KIM D Y, CHOI H, LEE H, *et al.*. A new cvSLAM exploiting a partially known landmark association [J]. *Advanced Robotics*, 2013, 27(14): 1073-1086.
- [21] 张宝龙, 李丹, 郭艳艳, 等. 基于 OV7962 的车载全景摄像头设计[J]. *液晶与显示*, 2015, 30(4): 634-640.
- ZHANG B L, LI D, GUO Y Y, *et al.*. Panoramic camera design for automotive applications based on OV7962[J]. *Chinese Journal of Liquid Crystals and Displays*, 2015, 30(4): 634-640. (in Chinese)
- [22] ZHUO J J, SUN L Q, SHI J J, *et al.*. Research on a type of camera calibration method based on high precision detection of X corners[C]. *Proceedings of the 8th International Symposium on Computational Intelligence and Design (ISCID)*, IEEE, 2015: 193-196.
- [23] STROBL K H, LINGENAUER M. Stepwise calibration of focused plenoptic cameras[J]. *Computer Vision and Image Understanding*, 2016, 145: 140-147.
- [24] 陈杰春, 丁振良, 袁峰. 三维重构不确定度的两阶段估计方法[J]. *光学精密工程*, 2008, 16(6): 1110-1116.
- CHEN J C, DING ZH L, YUAN F. Two-stage uncertainty evaluation of 3D reconstruction [J]. *Opt. Precision Eng.*, 2008, 16(6): 1110-1116. (in Chinese)
- [25] 翟优, 曾雷, 熊伟. 不同局部邻域划分 SURF 描述符的性能分析[J]. *光学精密工程*, 2013, 21(9): 2395-2404.
- ZHAI Y, ZENG L, XIONG W. Performance analysis of SURF descriptor with different local region partition[J]. *Opt. Precision Eng.*, 2013, 21(9): 2395-2404. (in Chinese)
- [26] 刘洋, 韩广良, 史春蕾. 基于 SIFT 算法的多表情人脸识别[J]. *液晶与显示*, 2016, 31(12): 1156-1160.
- LIU Y, HAN G L, SHI CH L. Recognition of expression-variant faces based on SIFT method[J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(12): 1156-1160. (in Chinese)
- [27] LI Z H, SHAO C Y, LIU Y M. Motion estimation based on axis affine model[C]. *Proceedings of 2010 International Conference on Mechatronics and Automation (ICMA)*, IEEE, 2010: 572-576.
- [28] CABALLERO D, ANTEQUERA T, CARO A, *et al.*. Data mining on MRI-computational texture features to predict sensory characteristics in ham [J]. *Food and Bioprocess Technology*, 2016, 9(4): 699-708.
- [29] CHO J S, LEE H J, PARK J H, *et al.*. Image analysis to evaluate the browning degree of banana (*Musa spp.*) peel [J]. *Food Chemistry*, 2016, 194: 1028-1033.
- [30] MOTTEKI A, YAMAGUCHI N, KARASUDANI A, *et al.*. Fast and accurate relocalization for keyframe-based SLAM using geometric model selection[C]. *Proceedings of 2016 IEEE Virtual Reality*, IEEE, 2016.

作者简介:



曹天扬(1984—),男,北京人,博士后,2006年于北京科技大学获得学士学位,2012年于北京科技大学获得博士学位,主要从事机器人视觉图像处理、传感器信号处理、多传感器信息融合的研究。E-mail: cao_tian_yang@sina.cn

导师简介:



刘昶(1969—),男,美籍人,研究员,1990年于清华大学获得学士学位,1996年于加州理工学院获得博士学位,主要从事微机电系统、传感技术及系统的研究。E-mail: tuengineer@qq.com