

ICS 35.140

L60/69

团 体 标 准

T/CVIA-64-2018

室内机器人视觉同步定位及地图构建的 技术规范

Specification of visual simultaneous localization and mapping for indoor robot

The technical specification

2018-7-24 发布

2018-7-25 实施

中国电子视像行业协会 发布

前 言

本标准是关于在室内机器人上使用的基于摄像头视频数据流等多传感器的视觉同步定位与地图构建的技术规范，该技术规范阐述定义了以下内容：

- 1、视觉室内定位技术的原理概述；
- 2、视觉定位模块传感器输入信号标准及输出接口标准；
- 3、视觉室内定位系统中坐标系选取的规范；
- 4、视觉室内定位技术应具备的功能；
- 5、视觉室内定位技术的技术指标；
- 6、视觉室内定位技术的检测标准；

本标准由【中国电子视像行业协会】归口。

本标准主要起草单位：珊口(上海)智能科技有限公司，哈尔滨工业大学，浙江大学，乐生（东莞）智能科技有限公司，深圳市中彩联科技有限公司。

本标准主要起草人：崔戡玮、李磊、温任华、陈浩耀、章国锋、白为民、郝亚斌、冯晓曦、彭健锋、杨勇、张利利。

引 言

随着机器人技术、计算机视觉技术的发展和嵌入式设备计算能力的提高，基于各种传感器所提供的移动信息进行同步定位与地图构建的技术在移动机器人上逐步普及。这类移动机器人可用于家庭或工业场景中，具有自主移动、环境建模、地图构建、定位导航等功能。本标准提出了关于在室内机器人上使用的基于摄像头视频数据流、陀螺仪、里程计等多传感器的视觉同步定位和制图的实现技术规范以及测试标准。

视觉同步定位与地图构建技术，简称 vSLAM，解决了同时建立机器人周围场景的三维模型，并定位机器人自身的空间位置还原出相机运动轨迹的问题。基于视觉的 vSLAM 技术采用面阵 CMOS 相机作为核心原件、融合惯性测量单元 (IMU)、里程计等传感器辅助定位。相比于激光雷达，室内无线载波等定位方案，具有成本低廉、功耗低、精度高、环境适应性好、使用范围广泛等优点。

根据目前室内视觉定位技术在机器人产品上使用情况，本标准提出了基于摄像头视频数据流等多传感器的视觉同步定位和制图的实现技术规范。技术框架内包含视觉特征提取模块、定位跟踪模块、地图构建模块、闭环矫正模块四个处理模块；在功能上体现为视觉跟踪定位、视觉场景识别及全局重定位、视觉坐标系与房屋坐标系对齐等功能。

作为一种快速发展的新兴的定位方式，视觉定位技术的误差特性与路径积分等传统的定位方式有很大不同。为了规范视觉定位技术的技术指标，本标准提出了测量室内视觉定位技术性能及精度的一系列相关技术指标，并给出了标准的测量方法。

目 录

1、范围.....	1
2、术语与缩略语.....	1
3、视觉同步定位及地图构建的原理概述.....	2
4、传感器采集层输入标准.....	4
5、视觉定位模块输出标准.....	5
6、视觉室内定位系统中坐标系选取的规范.....	6
7、室内定位技术应具备的功能.....	7
8、视觉室内定位技术的技术指标.....	7
9、室内定位技术的检测标准.....	9

室内机器人视觉同步定位及地图构建的技术规范

1、范围

本标准规定了室内视觉定位技术的实现方式与通讯接口。针对传感器与机器人的连接方式和接口定义、定位构图信息的输出格式进行了规范化的设计。同时给出了室内视觉定位技术方案的关键技术指标定义和测量方法。

本标准适用于使用摄像头及其他辅助传感器定位的室内机器人，也可供其他类型的机器人应用，以及视觉定位技术在增强现实，虚拟现实等应用参考。

2、术语和缩略语

2.1 术语和定义

下列术语和定义适用于本标准。

2.1.1

视觉跟踪 Visual Tracking

指基于视频流数据计算摄像头的实时位置。主要通过视频流中前后帧之间的图像匹配计算位置。

2.1.2

视觉特征 Visual Features

指图像中可识别，匹配的图像特征。典型的视觉特征包括角点(Corners)，线段 (Edges)，平面(Planes)等。

2.1.3

闭环检测 Loop Closure Detection

指基于摄像头视频流数据检测移动机器人是否回到之前去过的场景。当发生闭环时，系统会通过比较当前的坐标和在相同场景下的历史坐标对地图和坐标同时进行修正。

2.1.4

视觉重定位 Visual Relocalization

指在具有地图且视觉跟踪丢失的情况下，根据摄像头视频数据和地图数据进行分析匹配，重新计算出机器人在地图中的坐标。

2.1.5

I/O设备 Input/Output equipment

管理和控制计算机的所有输入/输出(I/O)设备是操作系统的主要功能之一，主要分为字符设备和块设备。一般由机械和电子两个部分组成。

2.2 缩略语

下列缩略语适用于本标准：

CPU Central Processing Unit 中央处理器

IMU Inertial Measurement Unit 惯性测量单元

V SLAM visual simultaneous localization and mapping 视觉同步定位与地图构建技术

RMSE root mean squared error 均方根误差

CMOS Complementary metal - oxide - semiconductor 互补金属氧化物半导体

3、视觉同步定位及地图构建的原理概述



图 1 视觉同步定位及地图构建系统应用方案图

3.1 视觉同步定位及地图构建系统利用摄像头采集到的视频流数据及其他类型的辅助传感器完成机器人的定位及地图构建。

3.2 视觉同步定位及地图构建技术中应包含实时传感器采集层。传感器采集层负责实时采集多种传感器的数据流并获取每个传感器数据点的时间戳。时间戳可以通过硬件同步的方式获取（同步触发多个传感器使其同时采集数据），也可以通过软件层的方式获取（基于获取到传感器的系统时间来制订时间戳）。

3.3 视觉同步定位及地图构建系统中应包含一个或多个摄像头以采集视频数据。也可以包含基于结构光或飞行时间法的深度镜头模组同时获取场景深度信息。

3.4 视觉同步定位及地图构建系统中还可以包括其他辅助传感器，例如惯性测量单元 (IMU)，里程计 (odometer, wheel encoder)，光流地面跟踪传感器 (floor tracking sensor) 等。这些辅助传感器的作用包括帮助视觉定位模块获取地图尺度信息，以及通过提供初始位置估计减少视觉定位模块的计算量及定位误差。

3.5 在视觉定位模块内部应包含利用图像视频流和其他传感器的信息跟踪定位机器人的位置（视觉跟踪定位）的功能并提供输出接口实时输出机器人的位置。视觉跟踪定位模块通过提取，分析，及匹配视频流中前后图像帧的视觉特征，并和地图中的视觉地标进行比对，推算出机器人的实时位置。

3.6 在视觉定位模块内部应包含地图构建及维护的功能。地图构建及维护模块根据机器人的位置，创建机器人周围环境的三维模型。该三维模型以稀疏或半稠密，或稠密的三维点云形式存在。定位模块可以通过串行通讯接口输出视觉地标的位置及描述。

3.7 为了修正长时间运行导致的累积误差，视觉定位模块内部还应具有环境识别及闭环检测与矫正的功能。该模块周期性地检测机器人是否位于之前去过的坐标附近，如果发现匹配，则根据历史记录的同场景下坐标和当前坐标的差别对定位和地图进行修正，以消除累积误差。在发生闭环的时候，机器人的定位可能出现跳变，外部程序可以通过API接口获取到闭环矫正的信号以进行相应的处理。

3.8 视觉定位系统在摄像头出现大面积遮挡，光线条件恶劣，或环境视觉纹理稀少的情况下会出现跟踪丢失的情况。当视觉定位丢失，且辅助传感器数据存在的情况下，定位模块应使用辅助传感器进行定位，避免出现没有定位输出的情况。

3.9 定位系统应具有视觉重定位的能力，通过分析比对摄像头采集到的图片和地图中存储的信息，重新计算出机器人的坐标并开始视觉跟踪。在发生重定位的时候，机器人的定位可能出现跳变，定位模块应输出视觉重定位信号以通知外部程序进行相应的处理。

4、传感器采集层输入标准

4.1 摄像头数据

4.1.1 摄像头数据的每个数据点包括图片数据及获取该图片数据的时间戳。

4.1.2 图片数据可以为彩色图像、黑白图像或深度图像。其中彩色图像及黑白图像的分辨率应至少为 640 像素x 360 像素，视场角大小应至少为 120 度，曝光方式可为卷帘相机曝光(rolling shutter)或全局快门曝光(global shutter)。深度图像可使用结构光或飞行时间(TOF)方式获得，分辨率应不低于 640 像素x360 像素。

4.1.3 摄像头数据应以DVP接口或MIPI接口的方式连接到视觉定位模组上。

4.1.4 图片数据的采集帧率应不低于 10 帧每秒。

4.1.5 出厂时摄像头应经过去畸变校准标定处理，畸变参数应以配置文件的形式存在模组内部。

4.2 惯性测量单元(IMU)数据

4.2.1 惯性测量单元(IMU)数据的每个数据点包括时间戳、加速度或位移增量、角速度或角度增量。

4.2.2 IMU数据应通过UART串口接入到视觉定位模组上。

4.2.3 IMU的协方差以配置文件的形式存储在模组内部。

4.3 里程计数据

4.3.1 里程计的每个数据点应包括时间戳、左轮、右轮的位增量。

4.3.2 里程计数据应通过UART串口接入到视觉定位模组上。

4.3.3 里程计数据的单位为脉冲值，每个脉冲值对应的距离长度应在出厂时进行校准标定，并以配置文件的形式存储在模组内部。

5、视觉定位模块输出标准

5.1 视觉定位模块的输出信号中应至少包含实时定位数据，每个数据点包括时间戳（64 位长度无符号整数，单位为微秒），三维坐标（每个坐标值为 32 位长度有符号整数，单位为 0.1 毫米），机器人在二维平面内的航向角（16 位长度有符号整数，单位为 0.01 度）。实时定位数据应以串行通讯接口输出。

5.2 视觉定位模块的输出信号中可包含视觉定位事件，每个数据点包括时间戳（64 位长度无符号整数，单位为微秒），视觉定位事件类型（32 位长度无符号整数）。视觉定位事件类型至少包括视觉重定位信号。在具有闭环检测功能的模块上，还应包括视觉闭环信号。在具有机器打滑检测功能的模块上，还可以包括打滑警告信号。

5.3 视觉定位模块的输出还可以包括视觉点云地图数据，每个数据点对应一个三维视觉地标点信号，至少包括视觉地标点在三维空间中的坐标（每个坐标值为 32 位长度有符号整

数，单位为 0.1 毫米）。点云中的每个数据点还可以包括地标点对应的视觉特征描述信息，和以及观测到该地标点时机器人所在的方向。

5.4 三维航向角信息，在具有三维导航功能的模块上，机器人的航向角还应包括俯仰角，横滚角信息。角度信息应以 16 位长度有符号整数表示，最小单位为 0.01 度。

6、视觉室内定位系统中坐标系选取的规范

6.1 室内定位系统应使用右手坐标系，X轴方向对应的航向角为 0 度，Y轴方向对应的航向角方向为 90 度。逆时针方向为航向角增加的方向，航向角的输出范围在-179.99 度到 180.00 度之间，最小精度不应低于 0.01 度。坐标输出范围应至少为-25 米至+25 米，坐标输出的单位为毫米。使用较小的单位可以避免整数和浮点数之间的转换。

6.2 在具有方形边界的室内环境中，视觉定位系统应当保证其坐标系的X轴及Y轴分别平行于一组边界。如图二所示。这样的坐标系定义方式使得室内定位系统的坐标系自然的与世界坐标系对齐。在具有多个房间的室内环境中，视觉定位系统的坐标系应当与初次初始化时的房间方向平行。视觉定位系统的默认原点位置应以机器人上电开机的位置为准。

6.3 不具有规则方形边界的室内环境中，视觉定位系统的坐标方向应以机器人上电开机时的位置为坐标原点，机器人初始朝向，即前进方向为X轴，机器的左侧为Y轴方向。

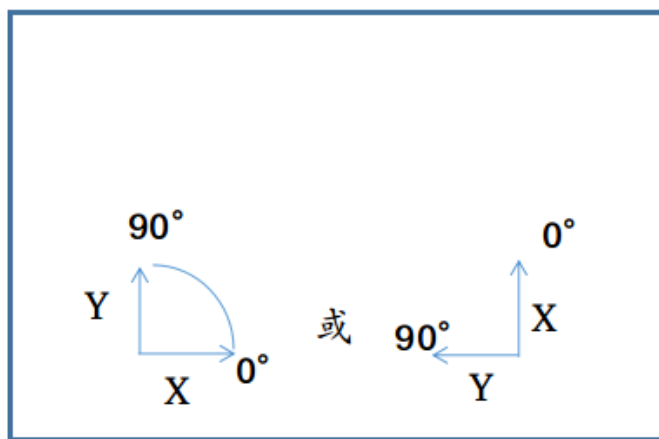


图 2 视觉定位模块的坐标系规范

7、室内定位技术应具备的功能

视觉室内定位技术应至少具备以下的功能。

7.1. 基于图像视频流作为输入数据，准确、实时的输出机器人的位置及航向角。

定位模块应能在不依赖其他传感器数据的情况下，实时输出机器人在空间中的定位及航向角。其中实时处理图片的帧率应高于每秒 10 帧图片以上，在标准室内测试环境中，平均绝对定位误差应小于 5 厘米，相对定位误差应小于 1%。

7.2. 基于图像视频流，具备跟踪丢失后快速重定位的能力。

在视觉跟踪丢失的情况下（例如镜头被遮挡或视觉条件恶劣的情况下）、或在机器人被人为移动的情况下，机器人在重新回到同一个场景中之前去过的位置时，应能基于视频数据重新快速计算出在世界坐标系中的位置，其中重定位成功率应高于 80%，重定位的响应速度应快于 1 秒钟。

7.3. 基于图像视频流，给出平行或垂直于房间方向的坐标系。

在初次开机时，机器人应能通过分析图片信息对房间进行快速建模，将自身视觉定位坐标系调整为平行或垂直于墙方向。在标准测试环境中，该功能的响应时间应小于 20 秒，坐标系与墙方向的夹角应小于 3 度，成功率应高于 90%。

7.4. 在同一环境中多次使用，具有地图重复使用的功能。

机器在同一环境中多次重复使用，机器人应具有地图保存并重新加载的能力。多次使用过程中视觉定位坐标系应保持一致。

8、视觉室内定位技术的技术指标

室内定位技术方案应提供以下技术参数以评估其性能。

8.1 图像处理帧率

图像处理帧率代表每秒钟系统可以处理的图像数量。在视觉定位系统中，图像处理的速度很大程度上影响了定位系统可以支持的最大运动速度及旋转速度。处理一帧图像应至少包括视觉特征提取及匹配，图像位姿计算等步骤。

8.1.1 图像处理的帧率可以通过运行标准测试数据集的速度测量。处理帧率为计算单元在标准测试数据集上的整体耗时，除以标准测试数据集所包含的图像数量。

8.1.2 图像处理帧率也可以通过在标准测试间实时测量。实时测量时，计算单元需要输出每一张所处理的图片所对应的时间戳及位姿，以及图片处理的中间结果以证明图片确实被处理。可以被接受的图片处理结果包括图片中所包含的视觉特征数量及描述，图片中所包含的视觉地标匹配数量及每个匹配地标的信息。

8.2 实时坐标输出频率

视觉定位系统可以通过融合惯性测量单元，里程计等其他传感器的方式以更高的频率输出定位坐标。实时坐标输出频率为计算单元在实际使用过程中每秒钟可输出的定位数据点的数量。

8.3 跟踪定位精度

定位精度反应了室内定位技术输出坐标的精确程度。定位精度包括位置精度和角度精度两部分。位置精度的单位为毫米，反应了在标准测试数据集上或标准测试间内运行时，定位模块给出的坐标和真实坐标的RMSE偏差（通常以单位长度内的偏差进行衡量，例如每 1 米平均偏差多少毫米）。角度误差的单位为度，反应了在标准测试数据集上或标准测试间内运行时，定位模块给出的角度和真实坐标角度的RMSE偏差（通常以每米的平均角度偏差进行衡量）。在测量精度时，由于定位模块的坐标系和真实值采集的坐标系可能存在差异，允许将所有的原始的定位数据进行一次整体的坐标平移及旋转变换以将原始坐标系和真实坐标对齐。

8.4 视觉跟踪成功率

视觉跟踪成功率指在标准测试数据集或标准测试间内机器人运行时，视觉跟踪正常的时间和总体测试时间的比例。在视觉跟踪率测试数据上，存在一定比例的恶劣视觉条件，例如强光，弱光，弱纹理，会导致一定比例的视觉跟踪丢失。视觉跟踪成功率反应了视觉定位系统对恶劣视觉条件的鲁棒性。

8.5 重定位成功率、速度及精度

重定位成功率反应了定位模块的视觉重定位的反应速度。在重定位标准测试数据集或在标准测试间的重定位测试样例中，机器人在固定场景下在视觉无遮挡的情况下预先运行二十分钟时间。而后的二十分钟内，摄像头每隔一分钟会被完全遮盖长达一分钟的时间，总计被遮盖/打开十次。机器的重定位成功率指在每次摄像头重新打开的 20 秒时间内机器成功完成重定位的比例。

重定位的速度指在成功完成重定位的情况下，从摄像头打开到完成重定位的时间长度。

重定位精度指在成功完成重定位的情况下，在完成重定位的第一个定位数据点上的定位误差。

技术指标名称	标准测试间测量标准	测试数据集测量标准
图像处理帧率	≥ 5 帧每秒	≥ 5 帧每秒
实时坐标输出频率	≥ 50 Hz	≥ 50 Hz
跟踪定位精度	$\leq 1\%$	不适用
视觉跟踪成功率	$\geq 80\%$	$\geq 70\%$
重定位成功率	$\geq 80\%$	$\geq 80\%$
重定位反应时间	≤ 2 秒	≤ 2 秒

表 1：视觉定位系统的技术参数标准。

9. 室内定位技术的检测标准

室内定位技术相关技术指标的测量标准由标准测试数据集离线运行和在《GB/T 34454-2017 家用干式清洁机器人 性能测试方法》所规定的标准测试间内实时运行两部分组成。

9.1 使用标准测试间检测室内视觉定位及地图构建技术。

9.1.1 标准测试间可用于测量跟踪定位精度，坐标输出频率，图像处理帧率，视觉跟踪成功率，重定位成功率及重定位速度。

9.1.2 标准测试间的布置与《GB / T 34454-2017 家用干式清洁机器人 性能测试方法》所规定的标准测试间的要求相同

9.1.3 在测试间内屋顶布置有两个高速广角摄像头用于实时跟踪捕捉机器人的真实运动轨迹。测试间地面上铺设有棋盘格视觉标定板，作为动作捕捉系统的参考坐标系。在机器人运行过程中，机器人上的视觉定位模块给出的坐标信息和标准测试间内动作捕捉系统给出的坐标信息会被分别存放在服务器上。两个轨迹之间的同步由机器人开始及停止的事件进行对齐。在机器人运动结束后，视觉定位系统计算得到的定位轨迹和动作捕捉系统获得的轨迹可进行一次旋转平移变换进行对齐，然后计算二者间的RMSE作为定位系统的精度。

9.2 使用标准测试数据集检测室内视觉定位及地图构建技术。

9.2.1 标准测试数据集是在 20 个不同的家庭环境中录制，可用于测量图像处理帧率，视觉跟踪成功率，重定位成功率及重定位速度四个技术指标。

9.2.2 该数据集包含相机视频数据，陀螺仪航向角数据，里程计数据三种传感器数据。录制时的机器典型直线运动速度为 300mm/s，旋转时典型角速度为 90° /s，相机录制帧率为 10 帧/秒，分辨率为 640 x 360，陀螺仪采样率为 100Hz，里程计采样率为 50Hz。数据集录制时扫地机执行典型的弓字形全覆盖清扫的路径规划清扫策略。在家庭环境的数据集包含有多种恶劣的条件，例如强光，弱光，弱纹理，摄像头遮挡，地面打滑等。可以系统地评估一个室内定位技术的标准。

9.2.3 由于在普通家庭环境中架设动作捕捉系统难度大，该数据集不具备真实运动轨迹参考，因而不适合用于评估定位精度。