**第2章 真实世界坐标系下的地图生成与定位方法设计**

2.1 引言

基于SLAM方法的即时建图和定位方法目前已经发展到的相当成熟，尤其是在无人机、无人车领域，基于视觉SLAM方法进行定位有着积极地应用，对于实际落地的场景，该系统输出的相机位姿和构建的地图必须具备真实尺度才能够进行导航与定位，此外依靠视觉构建的地图所处坐标系往往取决于初始化成功后的第一帧多建立的坐标系，这些问题的存在都使得传统SLAM系统难以有实际的使用条件。

对于当前传统的单目视觉SLAM算法，设备简易，处理数据较少，满足实时性的要求，但是也存在只能获取相机坐标系下得到相机位姿以及无法获取地图实际尺度的问题，无法展开实际应用；对于基于双目视觉的SLAM算法，可以解决地图尺度的问题，但面临成本较高，相机标定难度大且精度不高，鲁棒性较差等问题；对于融合视觉和IMU传感器的SLAM系统，可以获取真实尺度，以及得到真实世界坐标系下的相机和地图位姿，但是该系统对于IMU的精度要求较高，且在引入IMU后容易产生累计误差，难以初始化成功等问题。

针对上述问题，本章提出一种融合视觉传感器和二维码标签的单目视觉SLAM方法，在传统SLAM功能的基础上，可以得到相机和地图的真实尺度，并且根据一定的坐标转换，可以得到真实世界坐标系下的相机位姿和地图。在真实的应用场景中，布置的二维码和普通的自然关键点相比，更加容易捕捉到关键点，此外，比较固定，在重定位的流程中，往往会有更好的效果。该系统能够具备以下优良特性：能够对地图进行保存，复用和更新，通过不断完善的先验地图提高系统的鲁棒性；添加二维码信息增强传统SLAM中的重定位问题，此外通过二维码的尺度在仅适用单目相机的情况下估算出真实地图的尺度；依靠二维码中坐标系和真实世界的坐标系的转化，获取待估计物体在真实世界下的位姿。本章所提出来的结合二维码的视觉SLAM算法流程如图2.1所示，主要过程包括二维码。

2.2 二维码检测与识别方法

2.2.1 ArUco二维码的构成特点

二维码和一般的自然关键点相比较，具备比周围环境亮度更低得到的显著特点，如图2.2所示；且二维码本身是一个四边形的区域，可以凭借该特点约束出后续相机的位姿的绝对尺度；并且每一个二维码通过解码都可以获取到一个独一无二的对应ID号序号，在SLAM重定位的过程中，可以避免高重合区域的误检，在通过视觉检测二维码得到的过程中，我们希望尽可能多的检出场景中的所有二维码，随后再通过编码对误检值进行剔除。

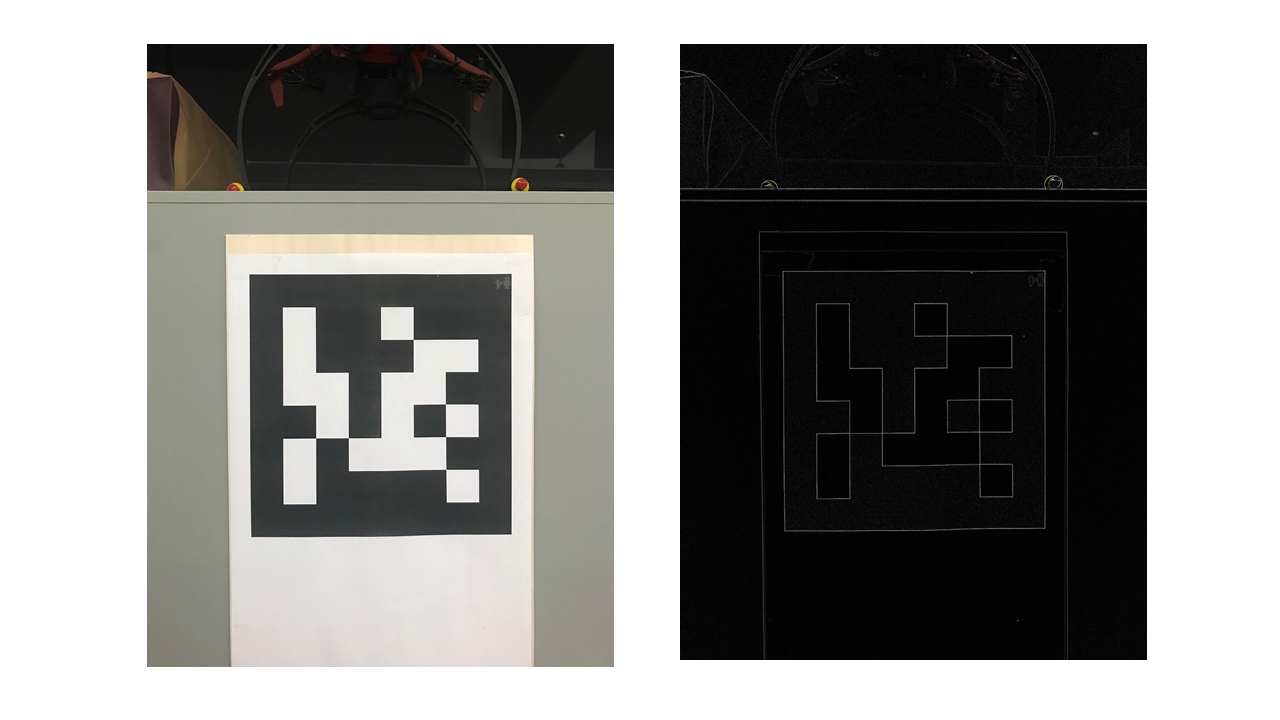


图2.2 ArUco二维码原始图像（左）和梯度图（右）

2.2.2 ArUco二维码的检测和识别过程

ArUco的检测和识别过程主要包括检测出二维码4个角点在图像中的位置，以及被检出的二维码的ID序号，在检测4个角点的位置时，需要检测图像中的线段和构成二维码的四边形。

在线段检测阶段，首先会计算整个图片中每一个像素的梯度强度大小和方向，随后对计算所得的梯度进行聚类，对所有满足聚类条件的像素点进行合并，可以的得到一组连续点，即检测出线段。

检测完线段后，进一步的需要检测构成二维码边缘的四边形，针对上一步中获取到的所有线段，对线段进行分组，若满足，上一线段的末端点和下一线段的起始点之间的距离小于某一阈值，即可首尾进行按照逆时针进行连接，若所有连接的线段数量达到4时，即认定生成的闭环可能为一个二维码的边缘四边形。

在检测二维码ID值之前，会先设定好一个包含所有二维码的字典，字典的大小即二维码的数量，字典中元素的大小即是每一个二维码的位数量。检测出构成二维码的四边形后，需要对图像进行透视变化规范图像，随后通过设定阈值分离出二维码上的黑色位和白色位，通过位数情况即可判断该二维码是否是字典外的不合规值，以及字典中的特定ID值，识别结果如图2.3所示。通过这样的方式检测和识别二维码具备非常好的鲁棒性，并且可以对错误值进行效验。

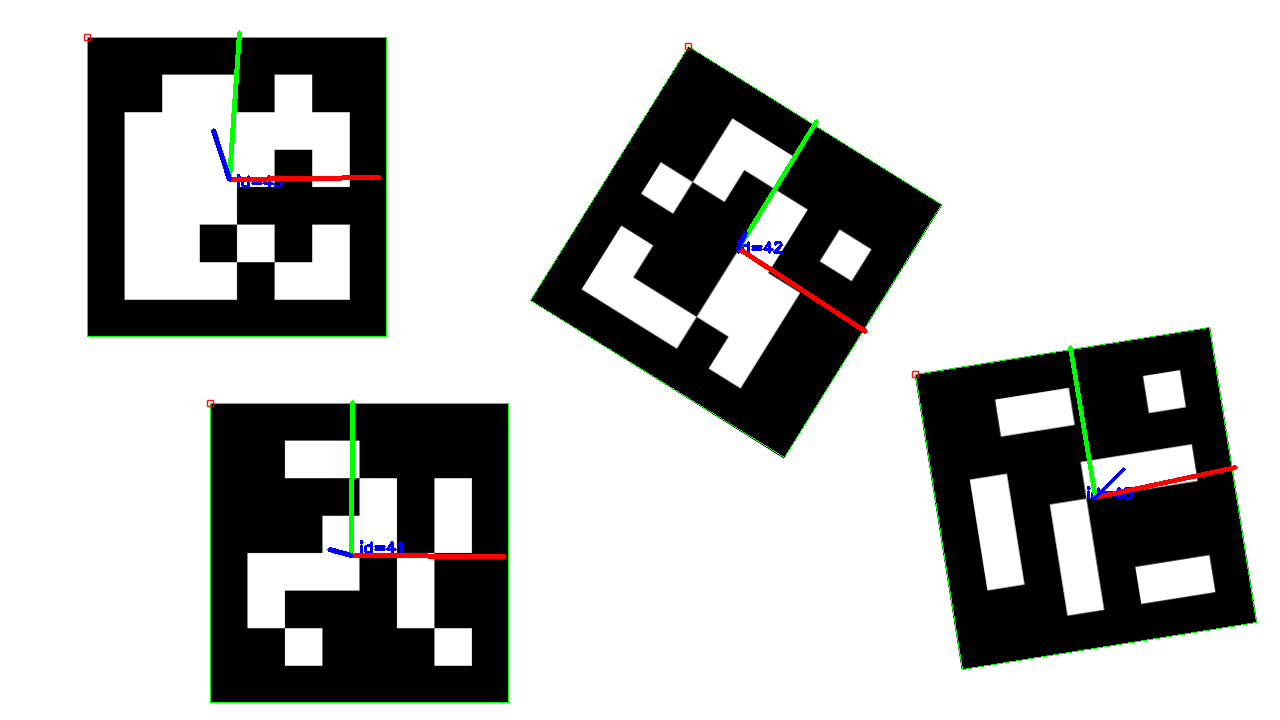


图2.3 ArUco二维码识别结果示意图

2.2.3 估计相机外参

对于传入的单帧图像，在提取完二维码的四边形轮廓，4个角点以及检测对应的唯一ID值后，接下来会估计检出的二维码位姿，包括平移向量和旋转向量，相机的外参主要是通过PnP方法，利用4对点来求解。

为了简化定位的过程，以二维码的坐标系作为世界坐标系，二维码边长为S，则四个角点的坐标分别为A（-S/2，S/2，0），B（S/2，S/2，0），C（S/2，-S/2，0），D（-S/2，-S/2，0），在图像中的对应像素点为a（）， b（）， c（）， d（）。因为相机的内参K提前标定，则三维空间中的点和像素坐标中的点之间的转换关系可以表示为：

通过以上公式，可以通过二维码到的四个角点求解出相机的旋转矩阵和平移矩阵。

2.3 基于二维码的SLAM系统描述

2.3.1 包含二维码的地图描述

在SLAM运行的过程中，可以生成一套地图系统，对于同一场景，该系统能够复用于后续的建图和定位过程，通过这样得到的方式可以添加对系统的约束，提高待估计量的精度。传统SLAM系统组成地图的元素主要包括关键点和关键帧，通过对关键点的描述符进行匹配，可以较好的的估计出相机的位姿。但关键点描述符的计算和匹配过程一般都比较耗时，而且对于重复场景，非常用于出现匹配错误的情况，考虑到这一情况，本文在此基础上，又添加二维码信息作为构成地图的元素，来进一步优化地图以获取更加准确得到相机位姿。本文中地图的构成包括以下3个集合：关键点集合p，关键帧集合f和二维码m集合，每个集合之间的数据元素相互耦合关联。

关键点集合p = {x,v,d},其中每个元素代表三维空间中的一个点，该点的描述包括在地图坐标系中的三维坐标x，观测方向v以及该点的描述符信息d，考虑到SLAM系统的实时性，选择BRIEF描述符来加快匹配过程。关键帧集合f = {T，&}，其中每一个关键帧包含一个外参矩阵矩阵T（旋转矩阵，平移矩阵），该外参矩阵是全局参考坐标系到相机参考坐标系的转化，&是相机的内参矩阵，该参数包括相机焦距，光学中心以及畸变参数，这些参数需要在运行SLAM系统之前就准确测量得到。二维码集合m={s,M,x1,x2,x3,x4}，其中的每一个二维码包括边长s（需要保证场景中的所有二维码的为正方形且所有二维码的尺度完全一致），二维码由其自身坐标系到全局坐标系的转换矩阵M，x1,x2,x3,x4分别代表二维码四个角点在其自身坐标系下的坐标位置。通过对包含二维码的自然场景进行离线建图，可以得到如图2.4所示的地图，其中正方形代表检出的二维码，蓝色矩形代表检出关键帧，红色点代表检出关键点。

其中每一类地图元素之间还具备一定的耦合关系，1）对于任意一个关键点，除了自身的属性外，还包括被观测帧序列，即所有可以观测到该关键点的关键帧的集合，以及在这些帧中出现的像素坐标位置；2）对于任意一个关键帧，还包与其关联的其他关键帧序列，以及在该关键帧中所有观测到的关键点和坐标，通过这些约束可以使得估计结果有更好的鲁棒性。

2.3.2 引入二维码的影响

SLAM是一个在导航的过程中，进行建图的工作，在使用单目相机时，即使目前表现较好的ORB-SLAM或者LDSO等算法也很难达到预期想要的效果，或者说存在很多的限制。第一个问题即是，所构成的地图的尺度是未知的，这样就造成无法在实际的导航任务中发挥作用，其次，在纯旋转的移动过程中，会导致算法失效，第三个就是单目视觉SLAM往往需要场景中存在比较丰富的材质便于跟踪，最后还存在一个问题，当场景中的视点发生变化，或者场景重复度较高时，在使用BOW词袋方法进行重定位时表现效果往往会很受到很大的限制。

除了用自然点来解决SLAM工作外，还有部分方法提出了仅仅使用人为设定的二维码来估计相机的位姿的方法，例如SPM-SLAM方法等，这种方式就可以较好的解决上述问题。因为二维码的放置不需要按照特定的规则，可以按照实验人员根据具体场景来布置，这样做就可以的得到一个由二维码构成的地图，并且，地图中的二维码都是具有尺度概念的。但是这样做存在的问题就是，我们需要在场景中布置足够多的二维码标志物，那么在大尺度的场景中，这样做就会提高实验的难度，因为在相机移动的过程中，要时刻保证至少有两张二维码出现在本帧中。

基于上述两种情况，本文提出了一种融合自然特征点和人工放置的二维码的SLAM方法来解决各自的问题和限制。首先，对于实际实验环境中，能够获取到真实的尺度概念，其次结合了自然点和二维码的SLAM方法也还可以可以只仅仅检测自然点，或者只检测二维码来进行位姿估计，因此对于大多数场景，都可以有更好的鲁棒性，第三个优点在于每一个二维码都有一个单独准确得到的ID值，那么在高度重复的环境中，也能保证匹配的正确性，最后对于大尺度的场景，可以结合自然点进行跟踪，二维码进行重定位来保证SLAM系统的长期稳定。

如图所示，对自然场景的局部区域进行建图，上图为自然区域的原图，在该区域中，有一部分的纹理较为丰富，其余部分则为贫纹理的白墙面，在白墙面上人为布置二维码标志，下图为根据该区域重建出的3D地图，由图可见，在丰富纹理的区域，自然点和人为布置的二维码都可以检出，但在贫纹理区域则，则基本只有二维码被检出，因此在贫纹理的区域就可以依靠二维码发挥作用。

2.3.3 基于二维码的SLAM过程描述

本文所提出的基于二维码的SLAM系统运行流程如图2.5所示，与一般常见SLAM系统相比较，本文主要提出了结合关键点和二维码标记进行结合使用。本方法中会一直对地图进行维护，有新的信息加入时，则会对地图进行不断的更新。在对某一个自然场景运行SLAM时，该地图为空，所以需要对其进行初始化的工作。

当地图初始化完成后，系统进行跟踪和重定位的环节，如果估算出来的相机位姿在某一帧被确定，那么系统会将该帧作为起始帧来估计的位置。对于跟踪环节，是依靠联合优化关键点和二维码角点的重投影误差进行的。地图的关键帧在系统跟踪前就被选定，并在整个系统中都起着很重要的作用。

跟踪成功后，SLAM系统会通过二维码来查询闭环，和单纯基于关键点的闭环检测相比较，二维码的检测必须提前执行，以防止在没有漂移校正的情况下使用二维码进行跟踪。如果成功检测到闭环，地图的两边需要进行适当的融合，随后将当前帧作为关键帧插入到地图中，并对整个地图进行全局的优化矫正，提出地图中冗余的信息，保证地图的内存大小维持在一个可以维护的程度。

随后，SLAM系统通过检测关键点来判断是否有闭环出现，若没有检测到，将利用局部优化来整合新的信息，若检测到，则进行全局优化。

如果SLAM系统在跟踪的过程中失效，则进入重定位模式，重定位模式即首先查找地图中已经存在的二维码标记，若找到二维码，则则可以通过二维码的位置估计出相机的位姿，并通过地图信息进行进一步细化，重新估计相机位姿。若没有找到可见的二维码标记，则在通过传统的BoW方法来进行重定位。在闭环检测中，将当前帧的关键帧和所有地图中的关键帧进行匹配，并采用RANSAC方法过滤误匹配。如果匹配的成功数量足够高，则认定重定位成功，并重新进入跟踪模式。

2.4 坐标系统转换方法

2.4.1 坐标变换的数学描述

在SLAM系统中，常用的坐标系存在2个概念，分别是相机坐标系和世界坐标系。其中世界坐标系是我们最需要考虑的一个坐标系，世界坐标系代表的是真实世界下，物体所参考的坐标系，一般都是固定不动的，可以给运动的物体提供控制方法，对于世界中的任意一点，可以描述为P（（Xw,Yw,Zw）。相机坐标系代表的是移动的物体自身的坐标系，由于相机是刚性物体，一般以相机的坐标为准，对于相机中的任何一点，可以描述为P（Xc,Yc,Zc)。

因为SLAM系统中会涉及到多个坐标系，不同的坐标系之间需要进行转化，空间中点的坐标系变换是一个欧式变换，一般由旋转加平移构成。首先只考虑坐标的纯旋转问题，即只有旋转，没有平移产生，两个坐标系的基底分别为（e1，e2，e3），（e1，e2，e3），对应的坐标分别(a1,a2,a3),(a1,a2,a3)。则有：

其中旋转矩阵R由两组基之间的内积组成，表示坐标系旋转前后同一个向量的坐标变换关系。旋转矩阵是一个行列式为1的正交矩阵：满足 特点，则有

a = Ra

在欧式变换中，除了旋转还有平移：

a=Ra+t

可以用旋转矩阵R和平移向量t 就能描述空间中的坐标变换关系,在三维向量的末尾添加1，称为齐次坐标。矩阵T同时包含了旋转和平移，称为变换矩阵。

坐标系代表的是移动的物体自身的坐标系，由于相机是刚性物体，一般以相机的坐标为准，对于相机中的任何一点，可以描述为P（Xc,Yc,Zc)。

2.4.2 获取真实世界坐标系方法

在SLAM系统中所描述的定位信息都是基于地图坐标系或者相机坐标系，但是在实际的应用中，就需要将所有的定位信息全部转换至真实世界坐标系，才能为物体后续的移动提供控制方法。我们通过传统的SLAM方法，只能够得相机在地图坐标系中的位姿，现得益于二维码自带坐标系的特性，可以经过一定的转换可以获取相机在真实世界坐标系下的位姿，可以为后续控制提供真实的位姿信息。

首先，声明所有可以得到的坐标信息，可以通过SLAM系统得到，二维码在SLAM地图坐标系下的位姿Tmarker（camera），相机在SLAM地图坐标系下得到位姿Tcamera（camera），二维码在真实世界坐标系的位姿Tmarker（world）。经过坐标变化，可以的得到相机在二维码下的位姿，Tcamera（marker） = Tmarker（camera）（-1）Tcamera（camera），真实世界坐标系在二维码坐标系下的位姿Tworld（marker）= Tmarker（world）（-1），那么相机在真实世界坐标系下的位姿Tcamera（world） = Tworld（marker）（-1）Tcamera（marker）

= Tmarker（world）Tmarker（camera）（-1）Tcamera（camera）

其中，二维码在真实世界坐标系的位姿可以按照实际场景进行设置，为了计算方便，直接设定第一个二维码的中心点为世界坐标系三维坐标的中心点，且x轴朝右，y轴朝上，z轴垂直于二维码平面向外，如图所示，那么第一个二维码的在世界坐标系下的坐标为（0,0,0），其余二维码则可以按照真实尺度获取。

在SLAM的实际运行前，应该保证二维码的实际布置坐标和设定的二维码在实际世界坐标系下的坐标一致，在SLAM的运行过程中，每一帧都可以看到至少一个二维码，那么每一帧对相机位姿的估计都会收到多个二维码的影响，考虑到

2.5 本章小结