

自动驾驶车辆的基础技术是环境感知。环境感知就是利用传感器获得车辆周围环境信息,再利用计算机技术对获得的信息进行处理,从而为车辆规划路线,控制车辆行驶,提高车辆的智能化水平。对交通标识的测距是自动驾驶环境感知研究的重要部分。交通标识是路面上最常见的道路设施,通常利用文字或符号向人们传递各种交通信息,如警告、限制、引导或指示等。

自动驾驶车辆环境感知常用的测距技术包括超声波测距、微波雷达测距、激光测距、红外测距和视觉测距等。超声波测距可测距离相对较近,不能满足对于交通标识的远距离测距;微波雷达测距和激光测距造价相对较高;红外线测距的应用场景受到限制,主要适用于黑夜或可见度较低的环境。视觉测距采用的是被动式测距,这与其他被动式测距算法完全不同,其不需要向待测目标发射或接收任何信号,只要采集到包含待测物体的录像或图片,就可以根据实际需要目标物体进行距离计算。

视觉测距又分为双目视觉测距和单目视觉测距。双目视觉测距是利用同一目标点在两个相机中成像点不同而产生的视差,得到物体 3D 信息,计算出距离。单目视觉测距则是利用相机拍摄图像中几何关系的投影模型进行测距。测距时前者不依赖外界环境,鲁棒性更好,精度也更高;而后者的投影模型一般对环境或场景有特定要求,这也在一定程度上制约了单目视觉测距技术的发展,综合考虑,本章重点讨论双目视觉测距的原理和基本方法。

5.1 双目视觉

5.1.1 双目视觉定义

双目视觉原理是通过两幅图像中相同目标点在左、右视图中成像点位置不同产生的视差来获得物体 3D 信息的方法。人们在研究生物视觉体系的过程中,发现

所有具有视觉的生物都是两只眼睛,由于两眼存在距离,当视觉生物观察一个目标时,目标在两眼的成像也稍有差别,位置会有些许偏差,这就形成了视差。人类利用视差融合两幅图像,从而产生远近或深度的感觉,视差对于人类的生活和生存有着非常重要的影响。

模仿生物视觉原理,建立由两个相机组成的双目测距系统,将相机间设置一定的距离,利用两台相机同时采集某一目标的两幅图像,通过图像处理技术获得同一目标点在两幅图像中的位置从而计算出视差,根据视差和双目视觉测距原理进行3D重建得到目标的距离。双目视觉有相对稳定的4个步骤,即相机标定、立体校正、立体匹配和3D重建。

1. 相机标定

Marr视觉理论认为从2D图像还原出3D场景,决定性的因素是产生2D图像相机的相关参数,这些参数描述了相机光心位置到2D成像平面和真实3D空间的转换关系。求解参数矩阵的过程就称为相机标定,这一过程的准确性也决定了3D重建的精度。目前常见的相机标定的方法包括摄影测量学的传统标定方法、直接线性变换法、Tsai两步法、基于Kruppa方程的自标定方法及张正友标定法。

以上标定方法中,摄影测量学方法计算十分复杂,直接线性变换法的结果精度不高,Tsai两步法需要特殊的标定设备,基于Kruppa方程的自标定方法精度不高。只有张正友标定法可以使用较为普通的棋盘格标定板,先通过线性成像模型得到相机参数的初值,再使用非线性最优化方法来得到相机参数的最优解,结果鲁棒性较好、精度较高且操作简单,因此被广泛使用。

2. 立体校正

双目立体视觉系统在实际安装过程中,很难保证左右相机光轴严格平行,且相机处于同一水平面上,但涉及后期3D坐标的重建和深度估计,所以需要根据相机标定出的参数对采集到的左、右视图进行立体校正。

图像经过立体校正以后实现左、右相机成像平面共面,相同目标点在左、右视图中处于同一行上。这样在后续的立体匹配中,便可采用极线约束算法,在相同行上搜寻相同目标点在左、右视图中的位置,从而减少立体匹配的工作量和复杂程度。图像立体校正可以分为相机未标定立体校正法和相机已标定立体校正法两类方法。

1) 相机未标定立体校正法

相机未标定立体校正法的代表为Hartley方法。该方法通过计算立体图像对之间的匹配点来求解基本矩阵。其优点是可以直接获得将极点映射到无穷远的单应性矩阵,无须知道相机内部参数。其缺点是匹配点的选取对校正精度影响较大,

且计算量大。

2) 相机已标定立体校正法

相机已标定立体校正法的代表为 Bouguet 方法。该方法需要知道左侧相机和右侧相机的相机内参和两相机之间的外参,根据标定参数计算重投影矩阵。

3. 立体匹配

立体匹配是将双目相机拍摄的两张不同的 2D 图像进行匹配的过程,是立体视觉中最为关键的技术。左、右相机的空间位置不同,其拍摄的图像也存在差异,具体体现在两幅图像的水平 and 深度视差。将左、右图像中的相同点进行匹配,就可以找到差异,然后利用这些差异,根据相关算法计算出 3D 立体信息。当前的立体匹配方法主要包括区域匹配、特征匹配和相位匹配等。

(1) 区域匹配的实质是利用局部窗口之间灰度信息的相关程度来实现图像匹配。该方法在变化平缓且细节较为丰富的图像区域可以达到较高的精度,但在遮挡、弱纹理以及深度不连续区域易出现错误匹配,而且抗噪性能差。

(2) 特征匹配是对图像的灰度信息进行抽象,匹配特征包括图像中的角点、线段、轮廓和边缘等,基于特征的匹配算法稳定性强、计算量小、速度快。常用的特征匹配算法有 SURF 算法、SIFT 算法和 ORB 算法等。

(3) 相位匹配将相位信息作为匹配基元,不易受噪声影响,误匹配概率低,采用该算法的前提条件是图像对中对对应点的局部相位是相等的。现有的相位匹配算法有基于极线校正的亚像素相位匹配方法、Fourier-Mellin 匹配算法等,相位匹配算法中存在的相位缠绕、相位偏差、相位奇点等问题都会影响匹配精度。

4. 3D 重建

3D 重建的原理类似于人们用双眼观察 3D 空间中的物体。根据前期的图像获取、相机标定、立体校正和立体匹配工作,从获得的相机内、外参数以及图像特征点的对应关系,就可以利用视差原理和三角测量原理计算出目标的 3D 坐标,恢复被测物体的深度信息,从而获得目标物体的 3D 信息。

5.1.2 双目视觉面临的问题

双目视觉主要面临的问题是立体匹配和视差计算方面。双目立体匹配的主要目的在于提高像素间的匹配准确度及效率,优化结果得到最终视差图,所以视差计算的问题其实也可以归因于立体匹配。由于实际环境复杂度较高,立体匹配过程中会面临各种不同的挑战。立体匹配是计算机视觉典型的难题,基线宽得到远目标测距准,而基线短得到近目标测距结果好,这就需要根据应用的需要进行折中。

立体匹配方法还可以分为全局法和局部法,实用的是局部法,因为全局法太

慢。深度学习虽然可以取代全局法,但目前还不是很成熟。此外,立体匹配的实际应用中可能存在重复的区域,进而导致匹配过程产生多异性。在拍摄同一物体时就会存在部分区域被遮挡的现象,即某个待匹配像素在匹配图像中找不到对应的像素点。除了以上几点不足外,还有光照、视差不连续等因素影响最终匹配结果。这些都是需要在立体匹配时考虑的因素。

5.2 双目视觉的应用

双目视觉广泛应用在机器人导航、工业尺寸测量、目标跟踪和自动驾驶等领域。

1. 机器人导航

对于任何移动设备而言,在其环境中导航的能力都很重要。首先避免如碰撞和不安全条件(温度、辐射、暴露于天气等)的危险情况,但是如果机器人的目的与机器人环境中的特定位置有关,则必须找到这些位置。机器人导航是指机器人确定自己在参考系中的位置,然后规划通往某个目标位置的路径的能力。

导航可以定义为3个基本能力的组合,即自我定位、路径规划、地图构建和地图解释,一些机器人导航系统使用同步定位和地图生成周围环境的3D重建。机器人定位表示机器人在参考框架内建立自己的位置和方向的能力。路径规划实际上是定位的扩展,因为它需要确定机器人的当前位置和目标位置,两者都在同一参考系或坐标系内。地图构建可以采用公制地图的形式,也可以采用任何表示机器人参照系中位置的符号。基于视觉的导航是双目视觉的重要应用领域,其技术实现包括基于激光的测距仪和双目相机,以提取周围环境中定位所需的视觉特征。随着双目视觉技术的发展,机器视觉技术已越来越多地用于机器人导航技术领域。

2. 工业尺寸测量

工业产品的尺寸测量是工业生产中重要的工序环节,传统方法采用的都是人工视觉测量,但是常常出现测量效率低下、精度不高及损伤工件表面等问题。非接触式的双目立体视觉系统是一种测量精度高、生产效率高的测量方式,尤其重要的是,双目立体视觉配合计算机系统的控制,能够实现工件3D尺寸测量的智能化,因此已广泛应用于工业工件质量检测及3D尺寸测量方面,可以提高生产效率和产品质量、方便采集产品数据信息。

机器视觉系统的特点是提高生产的柔性和自动化程度。在一些不适合人作业的危險工作环境或人工视觉难以满足要求的场合,常用机器视觉来替代人工视觉。此外,机器视觉易于实现检测与测量信息的集成。因此,发展更加智能化的机

器视觉测量系统,特别是双目视觉测量系统能够进一步加速生产企业的智能化水平,提高企业的效率。

3. 目标跟踪

在计算机视觉领域,目标跟踪是一个重要的研究方向,能够进一步实现视觉系统对物理世界的理解和感知。基于立体视觉的目标跟踪算法的优点是能够实时获取跟踪过程中被跟踪目标的深度信息,对运动目标的距离信息能够给出直观描述。其单目标跟踪算法处理流程大致包括以下步骤。

(1) 提取目标信息。根据双目图像的信息采集数据,在第 N 帧中提取被跟踪的目标并在图像中做出标记,作为目标模板。

(2) 跟踪目标。应用相关算法,在第 $N+1$ 帧中找到被跟踪的目标并做出标记,求出新目标的图像坐标并存储更新为新的目标模板。

(3) 对双目图像序列中的第 $N+1$ 帧进行立体匹配,用更新过的模板重复步骤(2)和步骤(3),直到跟踪结束。

(4) 输出跟踪目标在 2D 图像像素坐标中的运动轨迹图,也要输出运动目标在深度图像像素坐标系中的运动轨迹图。

4. 自动驾驶

在自动驾驶过程中可以利用双目视觉进行道路识别,主要是采用图像特征法和模型匹配法来进行识别。行驶过程中需要进行障碍物检测和路标路牌识别等,此时车辆上的信息采集也可以运用单目视觉或者多目视觉。相比之下,运用多目视觉更具优势,获取的图像信息可构建成 3D 空间,物体运动以及遮挡等问题对其影响较小。当然,双目视觉和多目视觉的缺点也比较明显,即采集的数据量较大,计算处理的实时性难以得到满足。

近年来,随着计算机技术及硬件的飞速发展,双目视觉成为自动驾驶中环境感知的重要方法。利用双目视觉可以实现道路中交通标识的测距、前方车辆的测距及车道线检测等。

5.3 双目测距原理与优化

5.3.1 双目视觉测距原理

图 5-1 所示为理想双目视觉测距模型的平面测距原理。其中,相机为理想模型,左侧相机和右侧相机光轴相互平行。图中,左、右两个相机焦距均为 f ,左、右相机光学中心为 O_L 、 O_R ,相机之间的距离为 d 。 O_1 、 O_2 分别为左、右相机成像平面中心。 P 是待测目标点, Z 为 P 到相机的垂直距离, P 在左、右相机上形成的像点分别是 P_1 和 P_2 , P_1 、 P_2 到 O_1 、 O_2 的横坐标为 x_1 、 x_2 。

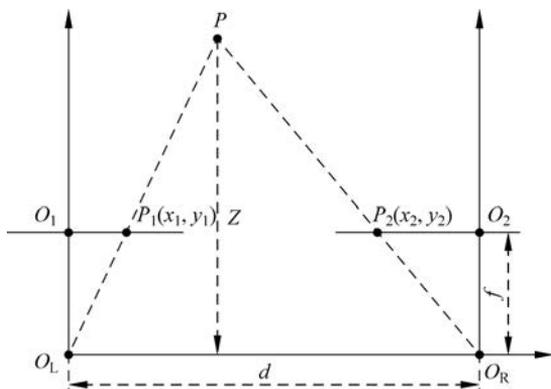


图 5-1 双目视觉测距原理

根据图 5-1,利用相似三角形性质,可得

$$\frac{d - (x_1 - x_2)}{Z - f} = \frac{d}{Z} \quad (5-1)$$

化简后,可得

$$Z = \frac{df}{x_1 - x_2} \quad (5-2)$$

式(5-2)为双目视觉测距模型的基本原理。其中 $x_1 - x_2$ 为视差,是双目立体视觉测距系统进行测距时主要求解的变量。

5.3.2 双目视觉测距优化

为了达到实时测距要求,需对算法进行两方面的改进,一方面是构建校正映射表,对坐标进行实时快速校正,另一方面是采用金字塔模板匹配算法完成相同目标的匹配。

1. 构建校正映射表

传统算法测距时一般会对全图进行立体校正,然后在校正后的图像上进行立体匹配,得到图像的深度图,这样会耗费较多的时间。传统的双目视觉测距流程如下。

- (1) 对相机进行标定。
- (2) 利用标定参数进行立体校正。
- (3) 对校正以后的图像进行立体匹配,得到视差。
- (4) 利用视差,根据双目视觉原理进行 3D 重建。

为了达到对交通标识实时测距的目的,需对传统的双目测距模型进行改进。由于针对图像中的交通标识进行实时测距,并不需要将图像所包含的所有物体都进行测距,故只需通过对原图进行匹配,得到交通标识在左、右视图中相对应的匹

配点,并将匹配点单独进行校正,会大大缩短测距时间。

Bouguet 算法原理是通过建立两个投影矩阵:将由于实际安装过程中形成的双目相机光轴不平行和镜头不在一个水平面上等问题带来的误差,通过两个投影矩阵投影到双目相机光轴平行、镜头在同一水平面上的理想模型中。

利用 Bouguet 原理得到投影矩阵,并求得其反投影矩阵,在假设已知非畸变图像的情况下得到相机真实的安装模型,再利用相机的畸变原理,根据非畸变坐标求出畸变坐标,得出理想双目测距模型到实际双目测距模型的映射矩阵,通过对其进行重建,得到由实际双目测距模型到理想双目测距模型的映射表。

在理想相机模型中(即非畸变的情况下),有某点 q ,其在图像中的像素坐标为 (u, v) ,在图像中的物理坐标为 (x, y) ,它们之间的相互转化关系为

$$u = u_0 + \frac{f}{dx} \left(\frac{X_c}{Z_c} \right) \quad (5-3)$$

$$v = v_0 + \frac{f}{dy} \left(\frac{Y_c}{Z_c} \right) \quad (5-4)$$

化简后,可得

$$u = u_0 + f_x x' \quad (5-5)$$

$$v = v_0 + f_y y' \quad (5-6)$$

式中, f_x 、 f_y 为相机标定得出的在 x 和 y 方向的焦距; x' 和 y' 为归一化的世界坐标。

对式(5-5)和式(5-6)进行变化,可得

$$x' = \frac{u - u_0}{f_x} \quad (5-7)$$

$$y' = \frac{v - v_0}{f_y} \quad (5-8)$$

相机畸变校正公式为

$$x_d = x(1 + k_1 r^2 + k_2 r^4) + 2p_1 xy + p_2(r^2 + 2x^2) \quad (5-9)$$

$$y_d = y(1 + k_1 r^2 + k_2 r^4) + 2p_2 xy + p_1(r^2 + 2y^2) \quad (5-10)$$

式中, x_d 、 y_d 为实际相机成像点,即畸变坐标; x 、 y 为理想相机成像点,即非畸变坐标; k_1 、 k_2 为相机径向畸变校正系数; p_1 、 p_2 为相机切向畸变校正系数; r 定义为

$$r^2 = x^2 + y^2 \quad (5-11)$$

把 x' 和 y' 代入式(5-9)和式(5-10),可得到 p 点对应的相机畸变图像坐标为

$$x'' = x'(1 + k_1 r'^2 + k_2 r'^4) + 2p_1 x' y' + p_2(r'^2 + 2x'^2) \quad (5-12)$$

$$y'' = y'(1 + k_1 r'^2 + k_2 r'^4) + 2p_2 x' y' + p_1 (r'^2 + 2y'^2) \quad (5-13)$$

其中, r' 的定义为

$$r'^2 = x'^2 + y'^2 \quad (5-14)$$

将上述公式代入式(5-5)和式(5-6),得到 p 点非畸变坐标 (u, v) 对应的畸变坐标为

$$u_d = u_0 + f_x x'' \quad (5-15)$$

$$v_d = v_0 + f_y y'' \quad (5-16)$$

式中, (u_d, v_d) 为实际相机成像的畸变坐标。

由式(5-15)和式(5-16)可以计算出由理想相机成像时的非畸变坐标计算出实际相机成像时对应的畸变坐标,进而得到整幅图像 u 和 v 对应的映射矩阵 $\text{map } u$ 和 $\text{map } v$, 计算原理如图 5-2 所示。

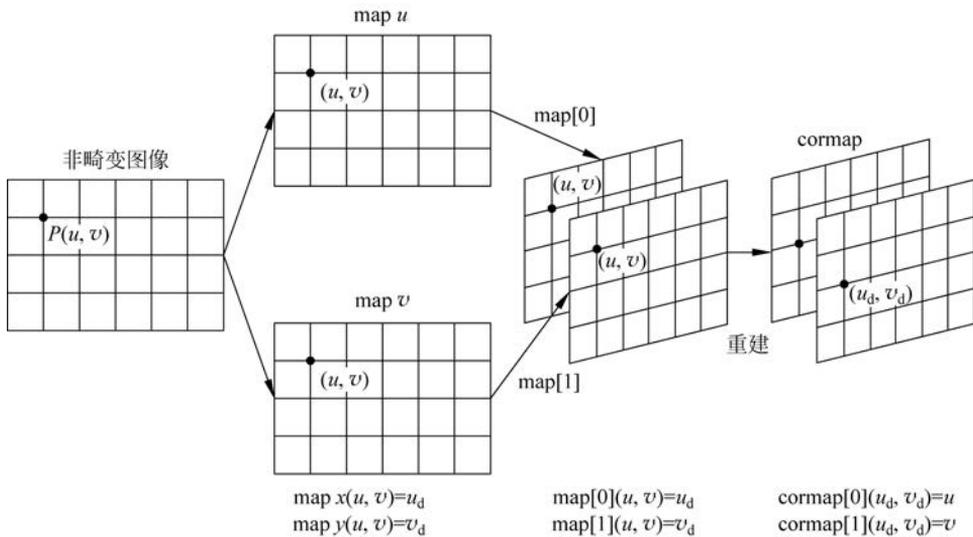


图 5-2 校正映射表原理

如图 5-2 所示,非畸变图像经过 Bouquet 计算和上述计算后,可得矩阵 $\text{map } u$ 和 $\text{map } v$, 两矩阵的大小与原图一致,矩阵的索引为校正后坐标 (u, v) 。 $\text{map } u$ 中存储的内容为 u_d , 为索引坐标中 u 对应的畸变坐标。 $\text{map } v$ 中存储的内容为 v_d , 为索引坐标中 v 对应的畸变坐标。再通过对 $\text{map } u$ 和 $\text{map } v$ 进行重建,得到校正映射表 cormap , 其索引为畸变坐标,存储的内容为对应的校正坐标。

2. 金字塔模板匹配

为了降低时间复杂度,采用图像金字塔的方法来进行模板匹配是比较合适的选择。图像金字塔也叫多分辨率搜索法,其原理是构建金字塔层数相同的搜索图像 S 和模板图像 T ,然后在分辨率最低的金字塔高层开始匹配,得到多个可能匹配

结果后,在高分辨率图像上对可能结果进行验证,直至找到最佳匹配结果。

构建图像金字塔的方法有很多,不同方法得到的采样效果也各不相同。相对简单的是隔点采样法,实现也比较简单,但其容易在匹配过程中出现误匹配,因为该采样方法会产生锯齿效应,导致梯度信息的丢失;均值滤波也有缺陷,它在做下采样时图像细节会被破坏,从而导致图像变得模糊;高斯滤波能够有效保留采样信息,所以本书采用高斯滤波方法构建金字塔图像。图 5-3 所示为高斯金字塔示意图,每层的尺寸从下至上依次减小。

高斯降采样的原理:对于一幅灰度图像 $F(x, y)$,其大小为 $M \times N$,令高斯金字塔 L_0 层图像为

$$L_0(x, y) = F(x, y) \quad (5-17)$$

则 L_l 层图像为

$$L_l(x, y) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) L_{l-1}(2x + m, 2n + y) \quad (5-18)$$

式中, $w(m, n)$ 为高斯加权函数。

经过高斯金字塔降采样得到的图像尺寸会缩减,此时得到的特征点坐标需要与原图进行关联。假设原图中 p 点的坐标为 (x, y) ,创建第 L 层高斯金字塔,此时特征点坐标 $p^L(x, y)$ 与原图的关系为

$$p^L(x, y) = p\left(\frac{x}{2^L}, \frac{y}{2^L}\right) \quad (5-19)$$

采用高斯滤波创建的图像金字塔的效果如图 5-4 和图 5-5 所示。

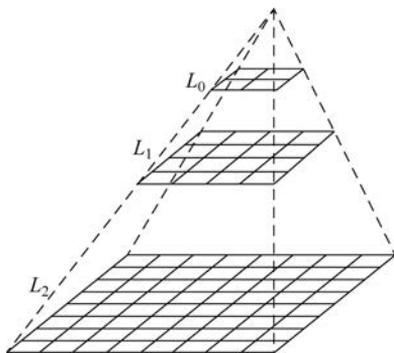


图 5-3 高斯金字塔示意图



(a) S(1)



(b) S(2)



(c) S(3)

图 5-4 高斯滤波搜索图下采样

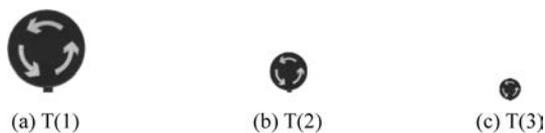


图 5-5 高斯滤波模板图下采样

由图 5-4 和图 5-5 可以看出,经过高斯滤波采样后的 $S(3)$ 和 $T(3)$,虽然分辨率降低,但是图像细节和图像轮廓仍然比较清晰。

金字塔模板匹配时,需对每层图像都进行模板匹配,本书采用的相似度量函数是 NCC 算法。金字塔模板匹配具体流程如下。

(1) 采用高斯降滤波方法对模板图像 T 和搜索图像 S 进行下采样,构建图像金字塔,且金字塔层数一致。

(2) 从金字塔的最顶层 $T(3)$ 和 $S(3)$ 间进行模板匹配,匹配算法选择基于灰度匹配的归一化互相关算法,由于像素较低,容易产生误匹配,因此保留置信度在 70% 以上的多个位置。

(3) 将步骤(2)中保留的位置还原到 $T(2)$ 层中,并在 $T(2)$ 层与 $S(2)$ 层间进行匹配,保留置信度在 80% 以上的位置。

(4) 重复步骤(3),在 $T(1)$ 层与 $S(1)$ 层间进行匹配,取置信度最大值处作为模板匹配的结果。

金字塔模板匹配算法通过在低像素图像上匹配,快速筛选出可能的匹配结果,还原到原图后,只需在可能的结果上再次进行匹配,便可找到模板在搜索图中的位置,有效缩短了匹配时间。

5.4 交通标识测距

针对交通标识实时计算测距法设计分为以下四部分。

(1) 相机标定及校正映射表计算:主要为相机标定试验并根据标定结果进行校正映射表的计算。

(2) 图像预处理:对采集到的测距图像进行预处理。

(3) 模板匹配:利用金字塔模板匹配算法对交通标识进行匹配。

(4) 坐标校正及测距:把模板匹配得到的交通标识坐标通过校正映射表进行校正,根据双目视觉测距原理进行距离计算。

5.4.1 相机标定及校正映射表计算

1. 相机标定

采用传统的张正友标定法。在标定前首先要拍摄用于标定的图像,拍摄标

定图像时要尽量使标定板成像位置在移动的过程中占据相机视野的每一个位置,且在不同位置有一定的角度,这样做是为了给张正友算法提供更多不同的角点,有利于求解。图 5-6 和图 5-7 分别展示了用于标定左侧相机和右侧相机的部分图像。

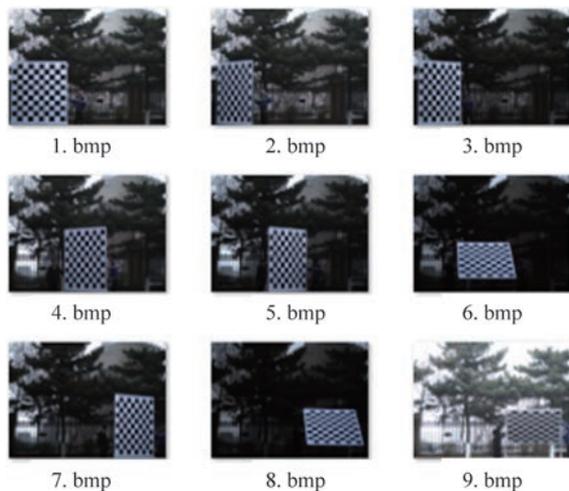


图 5-6 左侧标定图像

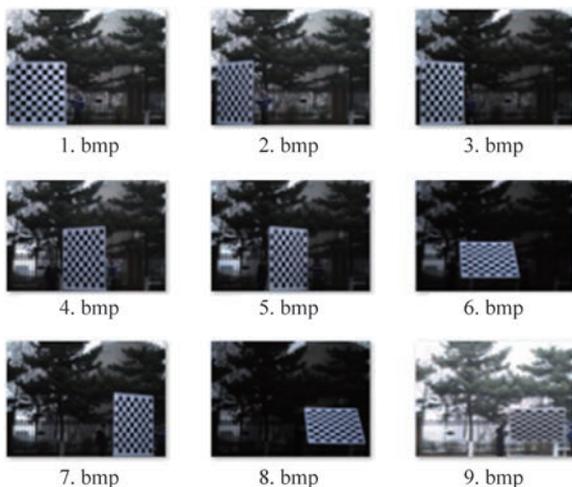


图 5-7 右侧标定图像

标定过程采用的是 MATLAB 2016 软件,它具有封装好的 Stereo Camera Calibrator 工具箱,该工具箱标定结果稳定,且可视化功能较强。图 5-8 所示为 Stereo Camera Calibrator 工具箱标定流程。

表 5-1 所示为左侧相机标定后得到的相机内部参数和畸变参数。



图 5-8 工具箱标定流程框图

表 5-1 左侧相机内部参数

参数名称	参数数值
焦距 (f_x, f_y)	(5662.03, 5648.27)
图像中心坐标 (u_0, v_0)	(599.56, 569.88)
畸变系数 k_c	(0.10021608, -1.03592668, -0.00470609, -0.0112025)

表 5-2 所示为右侧相机标定后得到的相机内部参数和畸变参数。

表 5-2 右侧相机内部参数

参数名称	参数数值
焦距 (f_x, f_y)	(5665.82, 5655.74)
图像中心坐标 (u_0, v_0)	(582.30, 540.54)
畸变系数 k_c	(0.11282374, -1.18606006, -0.00477237, -0.01115696)

得到相机内部参数后,根据单应性矩阵,对相机的外部参数进行求解,结果为

$$\mathbf{T} = [-359.08218 \quad 2.84432 \quad 37.08946] \quad (5-20)$$

$$\mathbf{R} = \begin{bmatrix} 0.99974 & -0.00457 & 0.02191 \\ 0.00455 & 0.99998 & 0.00103 \\ -0.02192 & -9.38133 \times 10^{-4} & 0.99975 \end{bmatrix} \quad (5-21)$$

根据标定得到的参数建立重投影矩阵,可得到用于标定的每张图像的重投影误差和所有标定图像的平均重投影误差,如图 5-9 所示。

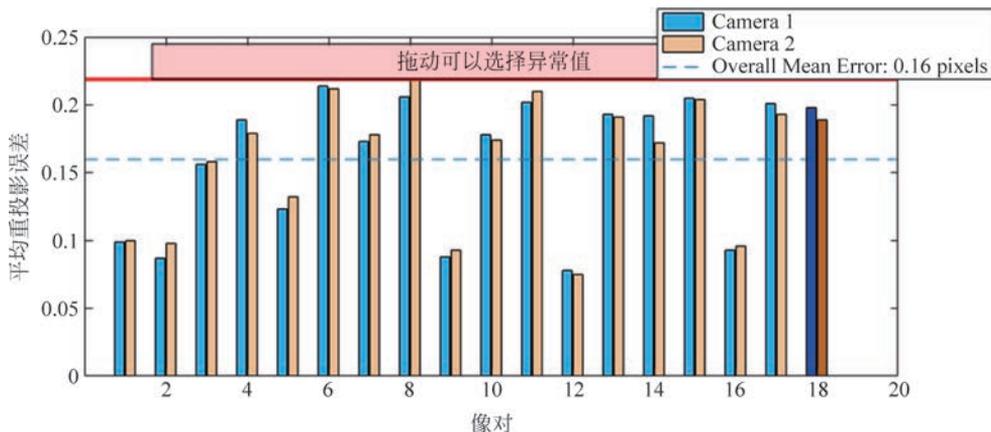


图 5-9 重投影误差表

重投影误差是评判相机标定结果好坏的一个重要标准,若最后平均重投影误差在 0.5 以内,表示标定结果相对准确,能够用于后续的测距程序。若平均重投影误差大于 0.5,则建议重新拍摄标定图像;否则会影响后续测距程序。

MATLAB 工具箱可以根据得到的相机参数进行逆投影变换,重建出相机与标定物之间的位置关系,如图 5-10 所示。

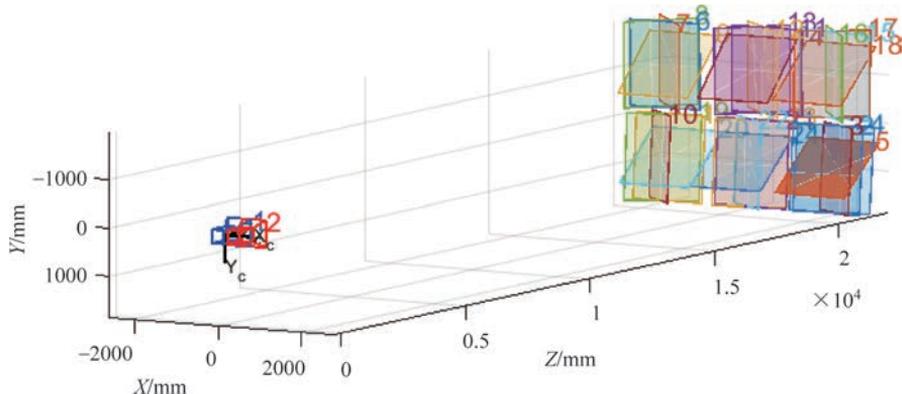


图 5-10 还原位置关系

从图中 5-10 中可以看出,相机与标定板之间的位置关系和两个相机之间的位置关系符合实际相机的安装情况。

MATLAB 中用于双目标定的 Stereo Camera Calibrator 工具箱实用性比较高,在标定的同时,可以随时观察重投影误差,并单击误差较大的图像进行删除,重

新进行标定,直至得到理想的结果。

2. 校正映射表计算

当得到相机参数后,根据本小节介绍的原理计算校正映射表,得到 cormap_L 和 cormap_R ,其分别为左侧视图和右侧视图对应的校正映射表,畸变坐标对应的校正后坐标为

$$u_L = \text{cormap}_L(u, v)[0] \quad (5-22)$$

$$u_R = \text{cormap}_R(u, v)[0] \quad (5-23)$$

式中, (u, v) 为畸变坐标; u_L 和 u_R 分别为左侧视图和右侧视图中 u 对应的校正后坐标,因为视差只与 u 有关,所以不对 v 进行校正。

5.4.2 图像预处理

图像预处理主要针对在车载环境中出现的曝光过度或曝光不足的问题,提出相应的解决方法,从而增强算法的鲁棒性。首先对图像进行灰度化,以减少计算量;然后建立判断条件,判断图像是否存在曝光过度或曝光不足,对需要预处理的图像进行直方图均衡化和拉普拉斯图像锐化。

1. 曝光问题与灰度化

车载环境采集图像时,有时会出现动态范围较窄的情况。CCD 或 CMOS 相机在采集图像时,由于背景与拍摄目标反差过大而导致高亮区域图像细节变得模糊,低亮区域曝光不足而过于黑暗的情况,如图 5-11(a)和图 5-11(b)所示。

图 5-11(c)所示为曝光不足对应的灰度直方图,图 5-11(d)所示为曝光过度对应的灰度直方图,由直方图可以看出图像有溢出现象,图像溢出会给后期模板匹配造成很大的干扰。所以,对于这种图像,在进行测距之前,需要进行预处理。为了加快预处理时的处理速度,对采集到的图像要先进行图像灰度化。

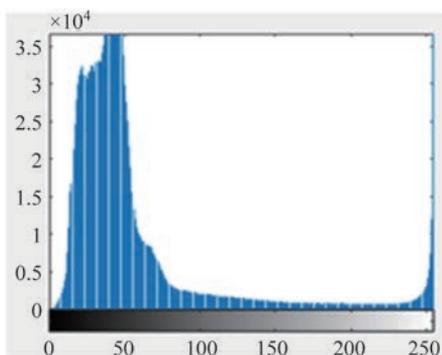
在现实生活中颜色的种类数不胜数,但经过研究发现,所有的颜色都可以由红、绿、蓝 3 种颜色按一定的比例组成。根据这一原理,相机采集到的图像也由 3 个通道组成,每个通道都有 $0 \sim 255$ 共 256 个数量级,因此在彩色图像中的每个像素点数占据的 3 个通道可由 $(0, 0, 0) \sim (255, 255, 255)$ 来表示,共约 1600 多万种颜色。在直接对彩色图像进行处理的情况下,尽管有信息丰富的优点,但会产生巨大的计算量,这难以满足实时性的要求。在处理图像的过程中,首先进行将彩色图像转换到灰度图像的操作,灰度图像里面的每个像素点仅包含一个分量,像素值在 $0 \sim 255$ 内进行变化,即灰度图像是特殊的彩色图像。转换为灰度图像后会明显减小相应的计算量,从而使处理图像的过程变得相对简单,同时灰度图像也可以显示彩色图像中色度和亮度的相关信息。将彩色图像转换为灰度图像主要有以下 3 种



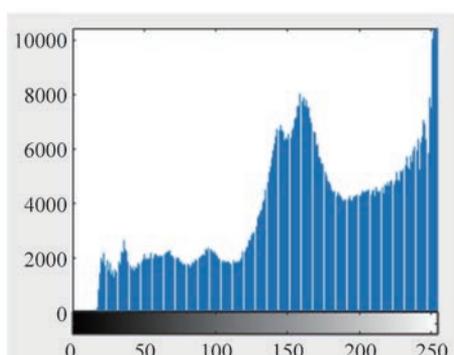
(a) 曝光不足图像



(b) 曝光过度图像



(c) 曝光不足图像直方图



(d) 曝光过度图像直方图

图 5-11 曝光不足和曝光过度图像

方法。

(1) 对比像素点上的 3 种颜色,将亮度最大的颜色分量定义为该像素的新灰度值,该方法为最大值法。

$$\text{gray}(i, j) = \max(R(i, j), G(i, j), B(i, j)) \quad (5-24)$$

(2) 将像素点上的 3 种颜色分量相加求平均值,将平均值定义为新的灰度值,该方法为平均值法。

$$\text{gray}(i, j) = \frac{R(i, j) + G(i, j) + B(i, j)}{3} \quad (5-25)$$

(3) 由于人眼观察物体时,对 3 种颜色的敏感程度不同,模仿人眼对 3 种颜色的敏感程度,对 3 种颜色赋予权值,将加权后的图像亮度值定义为新的灰度值,该方法为加权平均法,3 种颜色的权值为分别为 0.299、0.587、0.114。

$$\text{gray}(i, j) = 0.299R(i, j) + 0.587G(i, j) + 0.114B(i, j) \quad (5-26)$$

本书采用加权平均法对图像进行灰度化,效果如图 5-12 所示。

经过灰度化处理的图像,利用一个分量的灰度信息代替 3 个分量的彩色数据



图 5-12 灰度化效果

信息,能够有效减少图像处理时的数据计算量。

2. 判断是否存在曝光问题

图像灰度化以后,需要建立判断机制,判断图像是否存在曝光过度或曝光不足的现象。并非所有图像都存在曝光过度或曝光不足,如果存在曝光过度或者曝光不足,需进行后续预处理;如果不存在,则不处理。本书拟通过计算直方图中最大灰度级数、最小灰度级数以及灰度均值之间的关系来确定图像是否需要预处理。计算图像直方图右端像素数量低于总像素数 5% 的最小灰度级数 G_{\max} ,左端像素数量低于总像素数 5% 的最大灰度级数 G_{\min} 、图像灰度均值 G_{mean} ,得到直方图两端像素个数低于 5% 区域所占的比例关系 E_{rat} ,即

$$E_{\text{rat}} = \frac{G_{\min} + 1}{255 - (G_{\max} - G_{\min} - 1)} \quad (5-27)$$

当 $E_{\text{rat}} \geq 0.5$ 且 $G_{\text{mean}} \geq 128$ 时,认为图像曝光过度,需要进行预处理;当 $E_{\text{rat}} \leq 0.5$ 且 $G_{\text{mean}} \leq 128$ 时,认为图像曝光不足,需要进行预处理。其他情况下,图像效果比较满意,则不需要进行预处理。

3. 直方图均衡化

直方图均衡化就是对整幅图像的灰度值进行重新分配。由于图像的采集过程光线不均匀,导致图像的某些灰度级数的像素数量过多或某些灰度级数的像素数量过少,将这些灰度级数进行非线性拉伸,降低像素数量多的灰度级数、增大像素数量少的灰度级数,从而使整幅图像的灰度值在一定范围内分布均匀。以图 5-11(c) 为例,横坐标代表灰度级数,用 r 表示,纵坐标代表该灰度级数的像素个数,用 r_i 表示,出现灰度级数 r 的概率用 $p_r(r_i)$ 表示,即

$$p_r(r_i) = \frac{r_i}{\text{图像像素总个数}} \quad (5-28)$$

式中, $\sum_{i=0}^{k-1} p_r(r_i) = 1$, k 为一幅图像对应的灰度级数。

对 r 进行归一化, 则图像的灰度级数应分布在 $[0, 1]$ 区间内, 所以任意一灰度级数 r 都可以产生一个 s 值, 按 $s = T(r)$ 变换。

假定 $s = T(r)$ 满足两个条件: 一是在 $0 \leq r \leq 1$ 区间内有 $0 \leq T(r) \leq 1$; 二是在 $0 \leq r \leq 1$ 区间内是单调增加函数, 那么 $r = T^{-1}(S)$ 也应满足上述条件。同时假定 $F_s(s)$ 为随机变量 s 的分布函数, 其表达式为

$$F_s(s) = \int_{-\infty}^s p_s(s) ds = \int_{-\infty}^r p_r(r) dr \quad (5-29)$$

又由式(5-29)推导出 $ds = p_r(r) dr$, 假定 $p_s(s) = 1$, 两边积分得

$$s = T(r) = \int_{-\infty}^r p_r(r) dr \quad (5-30)$$

用灰度级数频率 L 来代替概率, 其离散形式的变换函数 $T(r_k)$ 可表示为

$$s_k = T(r_k) = (L - 1) \sum_{i=0}^k p_r(r_i) = (L - 1) \sum_{i=0}^k \frac{n_i}{n} \quad (5-31)$$

式中, $0 \leq r_k \leq 1, k = 0; 1, \dots, L - 1$ 。

直方图均衡化后的结果如图 5-13 所示。



(a) 曝光不足图像均衡化结果



(b) 曝光过度图像均衡化结果

图 5-13 图像直方图均衡化后的结果

由于像素被均匀拉伸, 经过直方图均衡化后的图像灰度分布更加均匀, 但同时均衡化后的图像会有一定程度的模糊, 需对其进行图像锐化。

4. 图像锐化

图像锐化(Image Sharpening)的原理是根据图像像素的变化程度, 也就是说, 依据像素点与周围像素的突变程度, 强调突变像素点达到使图像清晰的目的。一般像素值发生明显变化的位置在灰度跳变处和图像边缘处, 其余位置像素值的变化相对比较平缓, 为了使图像变得清晰, 可以增强图像的灰度跳变处和边缘处, 即

强调灰度跳变处和边缘处的像素值。

图像锐化的依据是函数的一阶微分和二阶微分,图像像素值向哪个方向变化是由函数的一阶微分描述的,即一阶微分描述了像素值的增长或者降低趋势;图像像素增长速度的快慢、下降速度的快慢则是由二阶微分描述的。

1) Robert 算子

Robert 是基于一阶微分的算子,其定义为

$$g_x = f(x+1, y) - f(x, y+1) \quad (5-32)$$

$$g_y = f(x, y) - f(x+1, y+1) \quad (5-33)$$

常用的检测模板为

$$\mathbf{g}_x = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (5-34)$$

$$\mathbf{g}_y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (5-35)$$

对交通标识进行边缘锐化试验,结果如图 5-14 所示。

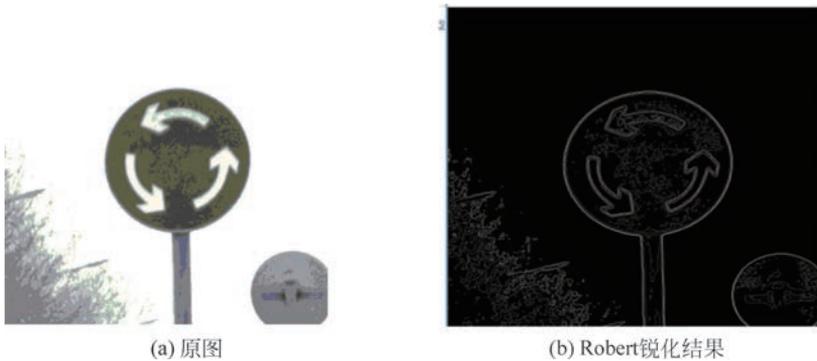


图 5-14 Robert 锐化结果

2) Sobel 算子

Robert 交叉算子的尺寸是偶数,偶数尺寸滤波器没有对称中心,计算效率较低,所以通常滤波器的模板尺寸设计的都是奇数。Sobel 算子的卷积模板都是奇数的,有对称中心,理论上计算效率应比 Robert 算子快。

Sobel 算子定义为

$$g_x = (f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)) - (f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)) \quad (5-36)$$

$$g_y = (f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)) - (f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1)) \quad (5-37)$$

利用上述公式可以得到以下两个卷积模板,分别计算图像在 x 方向和 y 方向的梯度,卷积模板为

$$\mathbf{g}_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (5-38)$$

$$\mathbf{g}_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (5-39)$$

对交通标识进行边缘锐化试验,结果如图 5-15 所示。



图 5-15 Sobel 锐化结果

3) Laplacian 算子

二阶微分算子的代表就是拉普拉斯算子,定义为

$$\nabla^2 f(x, y) = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y) \quad (5-40)$$

其得到的模板为

$$\mathbf{g} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (5-41)$$

对交通标识进行边缘锐化试验,结果如图 5-16 所示。

4) Log 算子

Log 算子是拉普拉斯算子与高斯低通滤波相结合的算子,Log 定义为

$$\nabla^2 G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} = \frac{-2\sigma^2 + x^2 + y^2}{2\pi\sigma^6} e^{-(x^2+y^2)/2\sigma^2} \quad (5-42)$$

常用模板为

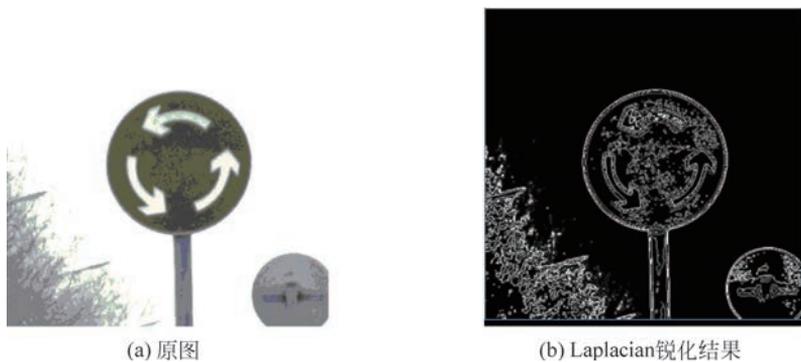


图 5-16 Laplacian 锐化结果

$$g = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & -16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (5-43)$$

对交通标识进行边缘锐化试验,结果如图 5-17 所示。由图 5-14~图 5-17 可以看出,这 4 种算子都能有效地提取交通标识的边缘,其中 Robert 算子和 Sobel 算子的检测效果相对更加清晰,但需要考虑运算时间。对 4 种算子运行时间进行对比,如表 5-3 所示。



图 5-17 Log 锐化结果

表 5-3 算子运行时间对比表

算 子	运行时间/ms
Robert	25.4
Sobel	2.7
Laplacian	0.4
Log	23.3

从表 5-3 可以看出, Laplacian 算子速度最快, Robert 算子运行速度最慢, 综合考虑锐化结果和运行时间, 决定采用 Laplacian 算子对图像进行锐化。

对直方图均衡化后的图像采用 Laplacian 算子进行锐化, 锐化结果如图 5-18 所示。



图 5-18 锐化结果

由图 5-18 可以看出, 经过锐化后的图像, 明显比均衡化后的图像清晰很多。而经过预处理操作以后, 交通标识明显较原图更加清晰。

5. 预处理流程及试验

1) 预处理流程

采集到图像以后, 首先对图像进行灰度化, 然后通过判断机制, 判断图像是否需要预处理, 若需要则进行直方图均衡化和 Laplacian 锐化, 流程如图 5-19 所示。

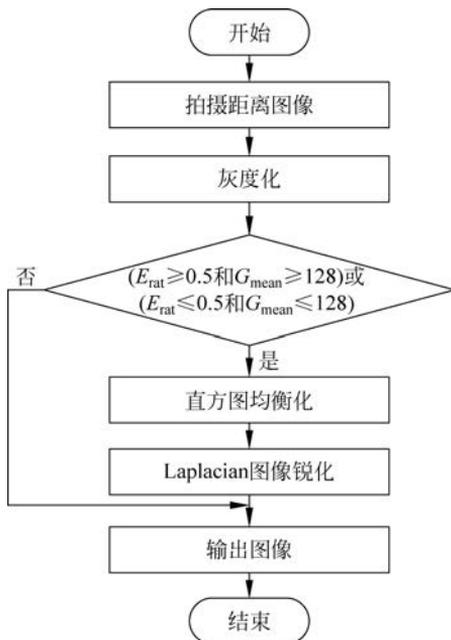


图 5-19 预处理流程框图

2) 预处理试验

为了验证预处理算法是否真的能有效解决曝光过度或曝光不足问题,减少因为曝光过度或曝光不足对模板匹配带来的影响,进行试验验证,匹配结果如图 5-20 所示。



图 5-20 预处理试验结果

图 5-20(a)、图 5-20(c)、图 5-20(e) 所示为在未经过预处理的曝光过度或曝光不足图像上进行的模板匹配结果,图 5-20(b)、图 5-20(d)、图 5-20(f) 所示为经过预处理以后模板匹配的结果。

由图 5-20 可以看出,未经预处理的曝光过度或曝光不足图像匹配不出图中的交通标识,而经过预处理后的图像能匹配到,说明预处理算法有效增加了算法的鲁棒性。

5.4.3 金字塔模板匹配

金字塔模板匹配时,手动获取左侧视图中交通标识的位置坐标,对其进行裁剪作为模板图像,根据金字塔模板匹配原理在右侧视图上进行匹配,匹配结果如图 5-21(c) 所示。



图 5-21 模板匹配结果

5.4.4 坐标校正及测距

得到左侧视图和右侧视图中交通标识坐标后,可以直接根据校正映射表读出其校正后坐标。考虑到距离主要根据视差进行计算,为了排除单一坐标点校正坐标可能出现的误差,所以选择交通标识中心坐标周围的 24 个坐标一起进行校正求平均值,再根据双目视觉测距原理求距离,即

$$\bar{u}_L = \frac{1}{25} \sum_{i=-2}^2 \sum_{j=-2}^2 \text{cormap}_L(u+i, v+j)[0] \quad (5-44)$$

$$\bar{u}_R = \frac{1}{25} \sum_{i=-2}^2 \sum_{j=-2}^2 \text{cormap}_R(u+i, v+j)[0] \quad (5-45)$$

式中, \bar{u}_L 、 \bar{u}_R 分别为左侧视图和右侧视图中 u 对应的校正后坐标的平均值。

根据双目视觉测距原理进行测距,有

$$Z = \frac{df_{xL}}{\bar{u}_L - \bar{u}_R} \quad (5-46)$$

式中, d 为双目相机之间的距离, 本书为 359mm; f_{xL} 为相机标定出的左侧相机在 x 轴方向的焦距(像素)。

5.4.5 测距流程

本书针对交通标识设计的测距算法流程如图 5-22 所示。运行测距程序时首先对系统进行初始化, 根据标定好的相机参数计算出校正映射表; 然后读取右侧视图, 判断是否需要预处理, 读取左侧视图, 截取交通标识作为模板图像; 接着利用金字塔模板匹配算法在右侧图像中匹配到交通标识的位置坐标, 再结合校正映射表得到左侧视图和右侧视图中交通标识校正后的位置坐标; 最后根据双目测距原理计算出距离。

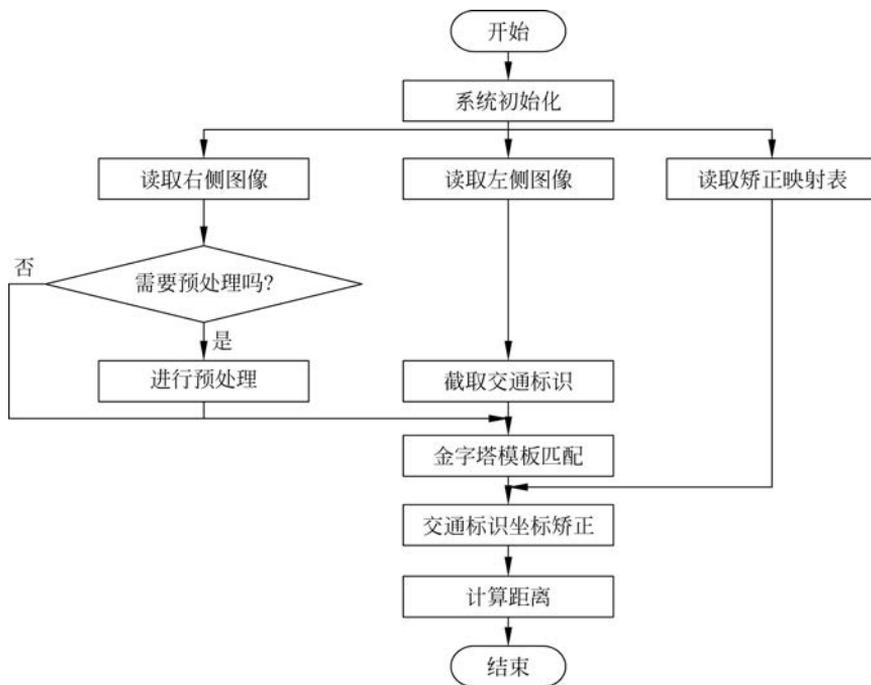


图 5-22 测距流程框图

5.5 试验与结果分析

双目视觉测距试验的基础是传感器设备, 视觉传感器的选择影响采集图像质量的高低, 所以要选择与算法和应用场景相适合的视觉传感器。车载环境下一般采用嵌入式系统, 试验还要验证双目测距系统和算法在嵌入式系统下的测距时间及测距精度是否能够满足自动驾驶车辆的要求。

5.5.1 双目视觉传感器

双目视觉测距系统主要由软件和硬件两部分构成。软件部分由视觉处理算法和软件开发环境组成。硬件部分由计算机和视觉传感器组成,视觉传感器如图 5-23 所示,车载环境下的配置如图 4-19 所示。视觉传感器由相机和镜头两部分构成。本书的应用场景为对远距离小目标交通标识进行拍摄,视觉传感器选择如图 4-20 所示。

远距离交通标识进行测距时,常见的测距标识有直行、环岛、左转弯、右转弯等,一般标准尺寸为 $120\text{cm} \times 120\text{cm}$,在远距离成像中所占的像素一般很少,所以尽量选择畸变小、远距离成像清晰、视场角较小的镜头和相机。镜头的参数如表 4-3 和表 4-4 所示。利用上述设备,采集 60m 处的交通标识时,交通标识在图像中大约占 55×55 像素,可以有效地进行模板匹配,从而进行测距。



图 5-23 视觉传感器

5.5.2 视觉传感器参数计算

使用的相机采用的是手动调节光圈和对焦,所以需确定光圈从而确定对焦平面,以保证相机在成像清晰的情况下有最大的景深。

光圈半径越大,进光量就越大,为了能够保证相机在阴天、多云和晴天的情况下都能对 60m 处的交通标识有清晰的成像,同时又不过多增加曝光(相机增加曝光的同时会降低采集帧率),光圈的半径不能太小,但也不能太大,光圈半径过大将影响景深的有效范围(光圈半径越大景深越小),最终经过多次试验测试后,决定采用光圈相对较大的 $F=2.8$ 。

确定光圈后计算相机的对焦平面,光圈与景深的关系为

$$\Delta_1 = \frac{g^2 \zeta}{2a - g\zeta} \quad (5-47)$$

$$\Delta_2 = \frac{g^2 \zeta}{2a + g\zeta} \quad (5-48)$$

式中, Δ_1 、 Δ_2 分别为后景深和前景深; g 为对焦面到相机的距离; ζ 为人眼明视距离,值为 0.00029; $2a$ 为光圈直径。

光圈 F 与光圈半径的关系为

$$F = \frac{f}{2a} \quad (5-49)$$

令 $2a - g\zeta = 0$, 便可得到唯一的一个对焦平面, 其后景深的有效范围为无穷远。如果当相机的光圈和焦距确定以后, 对焦时不对在焦面 g 处, 而是对焦在别的距离上, 景深将会大幅度缩小, 所以采用在 g 处对焦。

光圈为 $F = 2.8$ 时, 由式(5-48)、式(5-49)可得

$$2a = \frac{f}{F} = \frac{25}{2.8} \text{mm} \quad (5-50)$$

$$g = \frac{2a}{\zeta} = \frac{25}{2.8 \times 0.00029} \text{m} = 30.788 \text{m} \quad (5-51)$$

由上述公式可知, 当采用 25mm 镜头、光圈设置为 $F = 2.8$ 时, 在 30.788m 处对焦将有最大的景深。采用上述设备对距离相机 60m 的交通标识进行拍摄, 如图 5-24 所示。



(a) 左侧视图



(b) 右侧视图

图 5-24 采集图像

5.5.3 车载环境试验与分析

车载环境下的嵌入式系统硬件为 Nvidia Jetson TX2, 软件为 Ubuntu 18.04 定制版。车载环境下整体设备安装如图 4-19 所示。在该车载环境下, 在 20~60m 内对交通标识进行随机测距, 测距结果如表 5-4 所示。

表 5-4 车载环境测距结果表

实际值/m	测量值/m	绝对误差/m	相对误差/%
20.50	20.81	0.31	1.512
27.15	27.47	0.32	1.178
37.60	38.15	0.55	1.462
39.70	39.62	0.08	0.201
41.80	41.40	0.4	0.956
42.00	42.04	0.04	0.095

续表

实际值/m	测量值/m	绝对误差/m	相对误差/%
44.00	43.83	0.17	0.386
45.10	44.79	0.31	0.687
48.35	47.91	0.44	0.910
51.40	51.51	0.11	0.214
54.70	54.22	0.48	0.877
59.50	58.86	0.64	1.075

由表 5-4 可知,在嵌入式环境下,本书算法测距最大误差为 1.512%,整体误差略比实验室环境下大一些,是由于在实际车距测量过程中,移动的车辆随机停止,用卷尺测量前方交通标识至车辆上采集设备的距离,在操作过程中车载上双目设备到交通标识的距离无法被准确测量。因此,实际交通标识测距时存在标准差不准从而引入的人为误差较大,但整体满足对交通标识的测距精度需求。

算法在车载嵌入式环境下完成一次测量的时间在 20ms 以内,能够实现实时测距,满足在自动驾驶车辆中对交通标识进行实时测距的要求。

参考文献

- [1] 黄鹏程,江剑宇,杨波. 双目立体视觉的研究现状及进展[J]. 光学仪器,2018,40(4): 81-86.
- [2] 裴文慧. 面向自动驾驶的交通标识测距[D]. 沈阳: 沈阳工业大学,2020.
- [3] 赵勋杰,李成金. 双目立体实时测距系统的关键技术研究[J]. 激光与红外,2006(9): 874-877.
- [4] 祝琨,杨唐文,阮秋琦,等. 基于双目视觉的运动物体实时跟踪与测距[J]. 机器人,2009,31(4): 327-334.
- [5] 李海军. 基于双目视觉的自动跟踪系统应用[J]. 火力与指挥控制,2010,35(1): 150-152.
- [6] 朱先锋. 基于双目视觉的工件尺寸三维测量[D]. 舟山: 浙江海洋学院,2014.
- [7] ZHANG Z. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,22(11): 1330-1334.
- [8] DELCLÒS-ALIÓ X, MARQUET O, MIRALLES-GUASCH C. Keeping track of time: a smartphone-based analysis of travel time perception in a suburban environment[J]. Travel Behaviour and Society,2017,9: 1-9.
- [9] TOULMINET G, BERTOZZI M, MOUSSET S, et al. Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis [J]. IEEE Transactions on Image Processing,2006,15(8): 2364-2375.
- [10] 韩峻峰,王帅. 基于双目立体视觉技术的汽车测距系统实现[J]. 计算机应用与软件,2016,33(9): 227-230.