

学术型硕士(打印时删除)



同濟大學
TONGJI UNIVERSITY

硕士学位论文

基于 RGB-D 图像的三维物体识别算法 的研究与实现

姓 名：李 勇 奇

学 号：1531620

所在院系：电子与信息工程学院

学科门类：工学

学科专业：控制科学与工程

指导教师：陈启军 教授

二〇一七年十二月



同濟大學
TONGJI UNIVERSITY

A dissertation submitted to
Tongji University in conformity with the requirements for
the degree of Master of Engineering

3D Object Recognition and Pose Estimation Based on RGB-D Images

Candidate : Li Yongqi
Student Number : 1531620
School/Department : College of Electronics and
Information Engineering
Discipline : Engineering
Major : Control Science and Engi-
neering
Supervisor : Prof. Chen Qijun

December, 2017

学位论文版权使用授权书

本人完全了解同济大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版本；学校有权保存学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或者部分的阅览服务；学校有权按有关规定向国家有关部门或者机构送交论文的复印件和电子版；在不以赢利为目的的前提下，学校可以适当复制论文的部分或全部内容用于学术活动。

学位论文作者签名：

年 月 日

同济大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

年 月 日

摘要

在实际工程结构的服役过程中, 由于非线性与随机性的耦合作用, 工程结构特别是混凝土结构的非线性反应具有不可精确预测的性质。因此, 从概率密度演化的角度考察工程结构的非线性性状是准确把握结构非线性性能的必由之路。本文基于随机结构反应概率密度演化的思想对于随机结构分析理论进行了深入的探讨, 初步建立了随机结构反应概率密度演化的基本图景。

结构静力非线性分析是评价结构抗震性能的重要手段。对于具有双线型广义随机本构关系材料的结构, 其塑性截面分布状态的演化过程即非线性损伤构形状态转移过程反映了结构内力演化的性质。无记忆特性结构的非线性损伤构形状态转移过程具有马尔可夫性, 通过结构的力学分析可建立风险率函数与状态转移速率之间的关系, 进一步考虑状态之间的逻辑关系, 即可得到概率转移速率矩阵。对于有记忆特性结构及力-状态联合演化过程, 可通过引入相应的记忆变量构造向量马尔可夫过程, 并采用次序分析方法建立其确定性的概率密度演化方程。关于简单结构的情况进行了解析求解, 并据以探讨了结构非线性构形状态演化的若干特征, 发现了在实际应用中可能具有重要意义的稳定构形现象。讨论了力-状态的解耦问题。基于非线性构形状态本身的性质以及演化过程的规律, 初步研究了可能的简化与近似方法。

.....

最后, 关于进一步工作的方向进行简要的讨论。

关键词: 随机结构, 马尔可夫过程, 非线性构形状态, 差分方法

ABSTRACT

In practical engineering, the structures usually exhibits strong nonlinearity coupled with randomness of the involved parameters. This makes it almost impossible to exactly predict nonlinear response of the structures, particularly for the concrete structures. To tackle the difficulty, it is necessary to capture the nonlinear performance of the structures in the sense of probability, instead of purely deterministic standpoint. The present thesis is the result of the efforts devoted to developing the probability density evolution method for analysis of nonlinear stochastic structures.

.....

In the finality, the problems requiring further studies are discussed.

Key Words: stochastic structure, Markov process, nonlinear configuration state, difference method

目录

第 1 章 引言	1
第 2 章 RGB-D 图像的获取与融合	2
2.1 3D 相机现状与分析	2
2.2 RGB-D 相机	2
2.2.1 RGB-D 相机原理与结构	2
2.2.2 RGB-D 相机的数学模型	3
2.2.2.1 RGB-D 相机的标定流程	6
2.3 对偶 RGB-D 相机	8
2.3.1 对偶 RGB-D 相机原理与结构	8
2.3.2 对偶 RGB-D 相机的数学模型	11
2.3.3 深度图像融合算法	11
2.3.4 对偶 RGB-D 相机的标定	11
第 3 章 基于 RGB-D 图像的目标检测算法	12
3.1 深度信息预处理	12
3.2 3D Faster RCNN	12
3.3 3D MRCNN	12
第 4 章 基于点云的位姿估计算法	13
4.1 ICP	13
4.2 Super4PCS	13
第 5 章 实验验证	14
第 6 章 结论与展望	15
6.1 结论	15
6.2 进一步工作的方向	15
致谢	16
参考文献	17
附录 A 补充资料	18
个人简历、在学期间发表的学术论文与研究成果	19

符号说明

GNU	GNU's Not Unix /'gnu:/
GFDL	GNU Free Documentation License
GPL	GNU General Public License
FSF	Free Software Foundation

第 1 章 引言

本文 (Knuth 1989)

第 2 章 RGB-D 图像的获取与融合

2.1 3D 相机现状与分析

2.2 RGB-D 相机

2.2.1 RGB-D 相机原理与结构

RGB-D 相机获取深度的原理大致可以分为三种：

- Structure Light
- Time of Flight(ToF)
- Stereo

Structure Light 获取深度信息的原理是通过激光发射器投射带有特定编码的结构光到物体表面后,由 IR Camera 采集,根据采集到的光信号量的变化来计算物体的深度。举一个形象的例子,将手电筒照向墙面,手电筒离墙面越远,墙面上所形成的光斑的直径就越大,所以可以通过光斑的直径来计算手电筒距离墙面的距离。ToF 获取深度信息的原理是通过专有的传感器捕捉红外光发射到接收的飞行时间来计算物体的深度。Stereo 是通过双摄像头拍摄物体,再通过特征点匹配,根据三角测量原理来计算物体的深度。

三种原理的深度相机各有其特点,采用 Structure Light 原理的深度相机一般精度比较高,但景深比较短并且受光线影响比较大,适合室内场景;ToF 原理的深度相机获取深度图的精度和分辨率一般都比较低,但帧率高,并且具有一定的抗光照性能;Stereo 获取深度精度适中,帧率相对来说较低,并且需要较强的计算性能,但抗光照能力强,适合室外场景。

本文所使用的 RGB-D 相机是 Intel 的 Realsense SR300 相机,SR300 采用的结构光的原理获取深度^①,其内部结构如图2.1所示。从图2.1可以看出,SR300 内部的传感器主要有彩色摄像头(Color Camera)、红外激光发射器(Infrared Laser Projector)和红外摄像头(Infrared Camera)。Color Camera 是 1920×1080 像素的普通针孔摄像头,用来获取彩色图像;Infrared Laser Projector 和 Infrared Camera 用来获取深度图像或者红外成像图,两种成像流程如图2.2所示。其中当 Infrared Laser Projector 投射带有编码的结构光时,Infrared Camera 可以获取深度图;当投

① 此后所提到的 RGB-D 相机均指与 SR300 相机类似的采用结构光原理获取深度的 RGB-D 相机

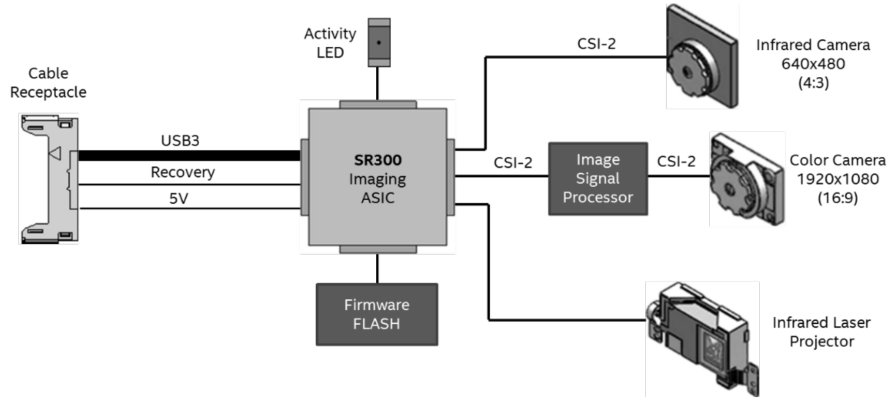


图 2.1 Realsense SR300 内部结构图

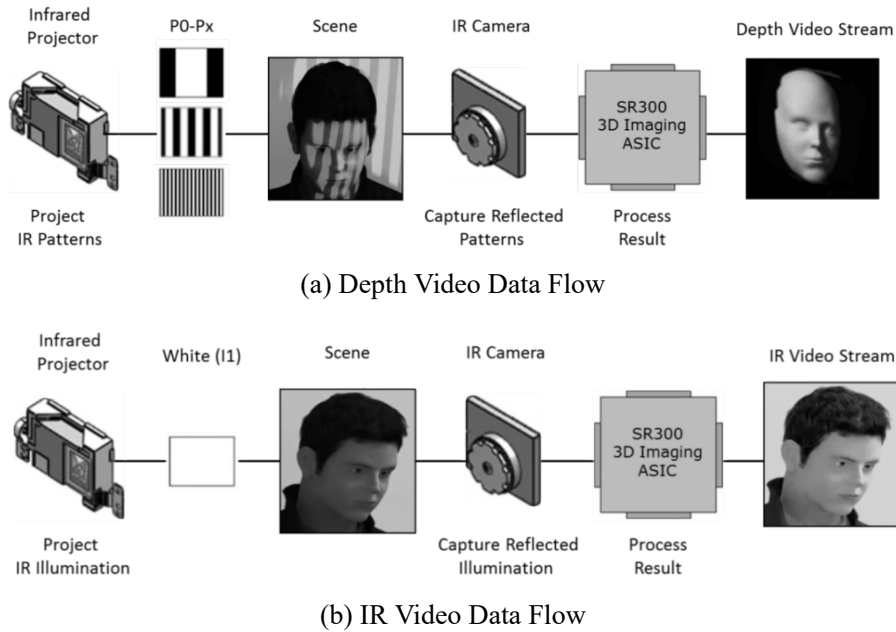


图 2.2 Realsense SR300 深度成像流程

射不带编码的红外光时, Infrared Camera 可以获取红外成像图。正常使用时, 往往设置 Infrared Laser Projector 投射带有编码的结构光来获取深度信息。因此, 从 RGB-D 相机的使用来看, 可以忽略其内部具体结构, 将其看成由一个彩色摄像头和一个深度摄像头构成, 其中彩色摄像获取彩色 (RGB) 信息, 深度摄像头获取深度 (depth) 信息。

2.2.2 RGB-D 相机的数学模型

图2.3展示了本文所使用的 RGB-D 相机的基本物理模型, 其中彩色摄像头和深度摄像头都使用了针孔 (pin-hole) 相机模型 (Heikkilä 2000)。先考虑普通针孔相机的模型, 相机图像坐标系下一点 $\mathbf{u} := [u, v]^T$, 对应的三维世界中的一点在相

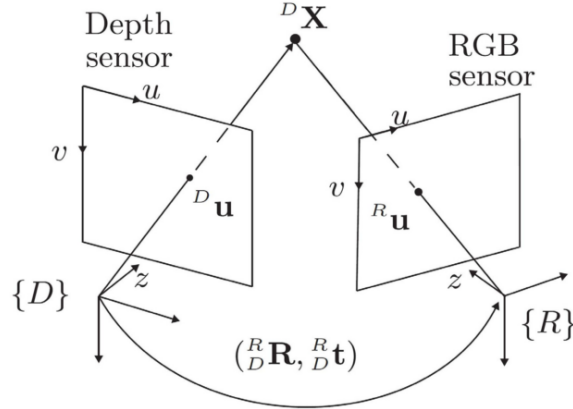


图 2.3 RGB-D 相机模型

机坐标系下表示为 $\mathbf{X} := [x, y, z]^T$ 。根据针孔相机模型有：

$$z\tilde{\mathbf{u}} = \mathbf{K}\mathbf{X} \quad (2.1)$$

其中 $\tilde{\mathbf{u}}$ 表示 \mathbf{u} 的齐次变换形式，彩色相机的内参矩阵 \mathbf{K} 的定义如下：

$$\mathbf{K} := \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$

其中 f_u 和 f_v 分别表示彩色相机在图像坐标轴上的焦距（以像素为单位）， u_0 和 v_0 表示彩色相机光心在图像平面的投影中心。

公式2.1还未考虑镜头的畸变，为了提高相机的精度，现引入径向畸变（radial distortion）和切向畸变（tangential distortion）：

- 径向畸变是由相机透镜的不完善和表面曲率存在误差造成的，径向畸变的数学模型可以表示为：

$$\begin{cases} \hat{x} = \bar{x}(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ \hat{y} = \bar{y}(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \quad (2.3)$$

其中

$$\bar{x} = x/z \quad (2.4)$$

$$\bar{y} = y/z \quad (2.5)$$

$$r = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (2.6)$$

\bar{x}, \bar{y} 表示点 \mathbf{X} 在归一化平面上的坐标， \hat{x}, \hat{y} 表示修正径向畸变后的坐标， k_1, k_2, k_3 表示径向畸变的参数。

- 切向畸变是由于相机透镜与图像平面不平行造成的,其数字模型可以表示为:

$$\begin{cases} \hat{x} = \bar{x} + (2p_1\bar{x}\bar{y} + p_2(r^2 + 2\bar{x}^2)) \\ \hat{y} = \bar{y} + (p_1(r^2 + 2\bar{y}^2) + 2p_2\bar{x}\bar{y}) \end{cases} \quad (2.7)$$

其中 p_1, p_2 是切向畸变的参数。

- 结合公式2.3和2.7可以得到修正径向畸变和切向畸变的 Brown-Conrady 模型 (BROWN 1966):

$$\begin{cases} \hat{x} = \bar{x}(1 + k_1r^2 + k_2r^4 + k_3r^6) + (2p_1\bar{x}\bar{y} + p_2(r^2 + 2\bar{x}^2)) \\ \hat{y} = \bar{y}(1 + k_1r^2 + k_2r^4 + k_3r^6) + (p_1(r^2 + 2\bar{y}^2) + 2p_2\bar{x}\bar{y}) \end{cases} \quad (2.8)$$

通过以上分析,根据公式2.1和2.8可以推导出带有畸变的针孔相机模型:

$$\begin{cases} u = f_u(\bar{x}(1 + k_1r^2 + k_2r^4 + k_3r^6) + (2p_1\bar{x}\bar{y} + p_2(r^2 + 2\bar{x}^2))) + u_0 \\ v = f_v(\bar{y}(1 + k_1r^2 + k_2r^4 + k_3r^6) + (p_1(r^2 + 2\bar{y}^2) + 2p_2\bar{x}\bar{y})) + v_0 \end{cases} \quad (2.9)$$

为方便起见,记 $\mathbf{d} := [k_1, k_2, p_1, p_2, k_3]^T$, 定义函数

$$f_{undist}(\mathbf{d}, \mathbf{X}) := \begin{bmatrix} \bar{x}(1 + k_1r^2 + k_2r^4 + k_3r^6) + (2p_1\bar{x}\bar{y} + p_2(r^2 + 2\bar{x}^2)) \\ \bar{y}(1 + k_1r^2 + k_2r^4 + k_3r^6) + (p_1(r^2 + 2\bar{y}^2) + 2p_2\bar{x}\bar{y}) \end{bmatrix} \quad (2.10)$$

$$\tilde{f}_{undist}(\mathbf{d}, \mathbf{X}) := \begin{bmatrix} f_{undist}(\mathbf{d}, \mathbf{X}) \\ 1 \end{bmatrix} \quad (2.11)$$

则公式2.9可简化为:

$$\tilde{\mathbf{u}} = \mathbf{K} \cdot \tilde{f}_{undist}(\mathbf{d}, \mathbf{X}) \quad (2.12)$$

其中需要标定的参数有相机内参矩阵 \mathbf{K} (包含未知参数 f_u, f_v, u_0, v_0) 以及畸变参数 \mathbf{d} (包含未知参数 k_1, k_2, p_1, p_2, k_3), 共 9 个参数。

明确了针孔相机的数学模型后,很容易推出 SR300 的相机模型:

$$\begin{cases} {}^R\tilde{\mathbf{u}} = {}^R\mathbf{K} \cdot \tilde{f}_{undist}({}^R\mathbf{d}, {}^R\mathbf{X}) \\ {}^D\tilde{\mathbf{u}} = {}^D\mathbf{K} \cdot \tilde{f}_{undist}({}^D\mathbf{d}, {}^D\mathbf{X}) \\ {}^R\mathbf{X} = {}^R_D\mathbf{R} {}^D\mathbf{X} + {}^R_D\mathbf{t} \end{cases} \quad (2.13)$$

其中左上标 $\{R\}$ 表示 SR300 相机中的彩色相机 (RGB), $\{D\}$ 表示 SR300 相机中的深度相机 (Depth), ${}^R_D\mathbf{R}$ 和 ${}^R_D\mathbf{t}$ 表示了彩色相机坐标系和深度相机坐标系之间的齐次变换关系。

2.2.2.1 RGB-D 相机的标定流程

根据上文所述的 RGB-D 相机的结构及数学模型,RGB-D 相机的标定主要涉及到彩色摄像头内参和畸变的标定,深度摄像头内参和畸变的标定,以及彩色摄像头和深度摄像头之间位姿变换的标定。由于 RGB-D 相机是一种较为新颖的相机,所以市面上基本上没有较为成熟通用的标定 RGB-D 相机的方法以及对应的工具。因此本文针对所使用的 Realsense SR300 相机,设计了一套标定方法。

根据公式2.13可知相机需要标定的参数有彩色相机内参和畸变参数 9 个,深度相机内参和畸变参数 9 个,彩色相机和深度相机之间的位姿关系 6 个,一共 24 个参数。一起标定这 24 个参数理论上是相当困难的,考虑到普通针孔相机的标定技术已经相当成熟(如张正友的棋盘格标定 (Zhang 2002),以及 RGB-D 相机中彩色相机和深度相机的解耦性,因此所设计的标定方法分为三步:

- Step 1 标定彩色相机内参以及畸变参数
- Step 2 标定深度相机内参以及畸变参数
- Step 3 标定彩色相机和深度相机之间的齐次变换关系

步骤 1 标定彩色相机内参以及畸变参数相对来说比较简单,主要参考文献 (Zhang 2002),但所使用的标定板是不对称圆盘标定板 (Asymmetrical Circle Board),如图2.4是 4×11 的不对称圆盘标定板。使用圆盘标定板而非棋盘格标定

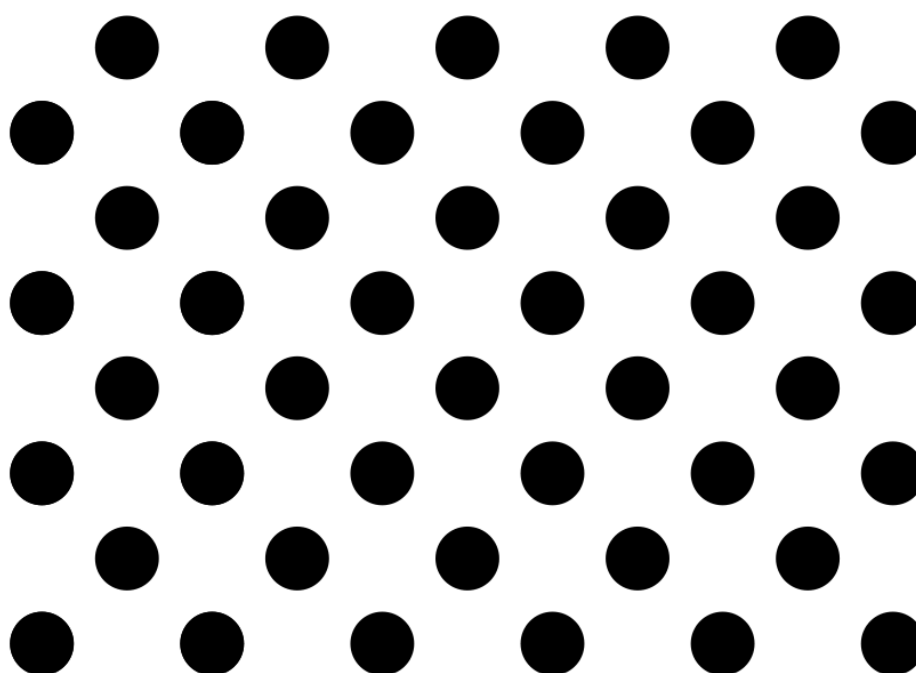


图 2.4 Asymmetrical Circle Board

板的原因是圆盘相对于棋盘格有更高的检测精度,在某些情况下可以达到 0.1 到 0.01 像素的亚像素精度,当然代价是相比计算棋盘格的角点,计算椭圆(圆形经过投影变换后退化为椭圆)的中心会涉及到较为复杂的数学运算,这也是为什么工业上大多使用圆盘作为标定板的原因。

步骤 2 标定深度相机内参以及畸变参数的方法和步骤 1 类似,区别在于深度相机并不能直接获得颜色信息,因此也不能直接检测图 2.4 所示的标定板。但是,幸运的是,根据前文所述的 SR300 深度相机的原理,其本质上也是个普通的针孔相机,只不过在其镜头上加上了滤波片,可以认为其只对红外光成像。因此,只要使用图 2.2 中的红外成像模式获取红外成像图,在红外成像图上检测标定板。如图 2.5 所示,在红外成像图中检测出了标定板。

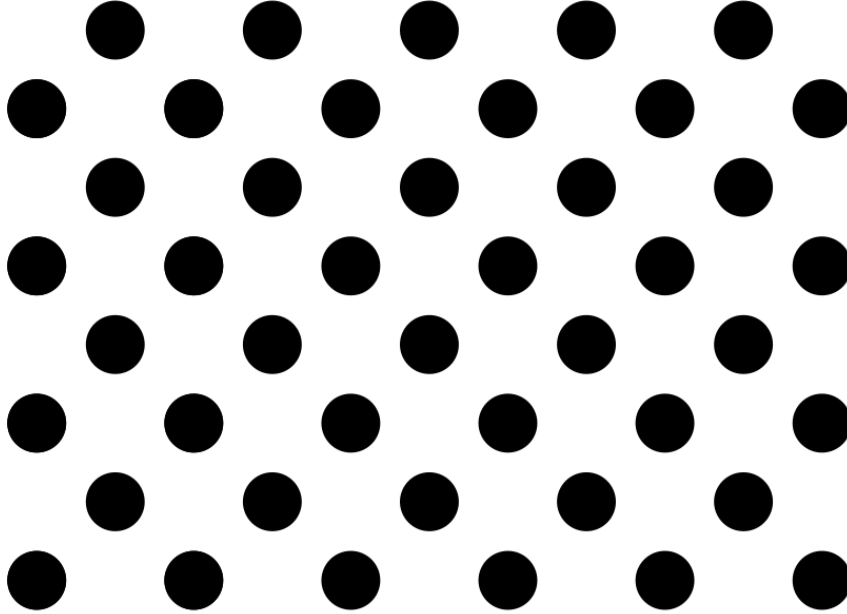


图 2.5 红外成像图中检测标定板

步骤 3 标定彩色相机和深度相机之间的齐次变换关系需要依赖于步骤 1 和步骤 2 中标定出的彩色相机和深度相机的内参和畸变参数,具体做法是将标定板放在彩色相机和深度相机下,使彩色相机和深度相机能够同时检测到标定板,然后分别根据各自的内参和畸变参数计算出标定板的位姿 ${}^R_B\mathbf{H}$ 和 ${}^D_B\mathbf{H}$, 其中 ${}^R_B\mathbf{H}$ 是 4×4 的齐次变换矩阵,表示标定板在彩色相机坐标系下的位姿,也是彩色相机坐标系变换到标定板坐标系的齐次变换矩阵; ${}^D_B\mathbf{H}$ 也是 4×4 的齐次变换矩阵,表示标定板在深度相机坐标系下的位姿,也是深度相机坐标系变换到标定板坐标系的齐次变换矩阵。从而所要求的彩色相机坐标系变换到深度相机坐标系的齐次变

换矩阵为:

$${}^R_D\mathbf{H} = {}^R\mathbf{H}_B {}^D_B\mathbf{H}^{-1} \quad (2.14)$$

其中

$${}^R_D\mathbf{H} := \begin{bmatrix} {}^R_D\mathbf{R} & {}^R_D\mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (2.15)$$

当然,实际标定时,往往会采取多组 ${}^R_B\mathbf{H}$ 和 ${}^D_B\mathbf{H}$ 来提高标定的精度。

2.3 对偶 RGB-D 相机

使用 SR300 相机时,发现相机在某些情况下,对一些反光的物体的深度图有严重的缺失,具体如图2.6所示。经过实验,发现这种缺失情况的出现和拍摄的

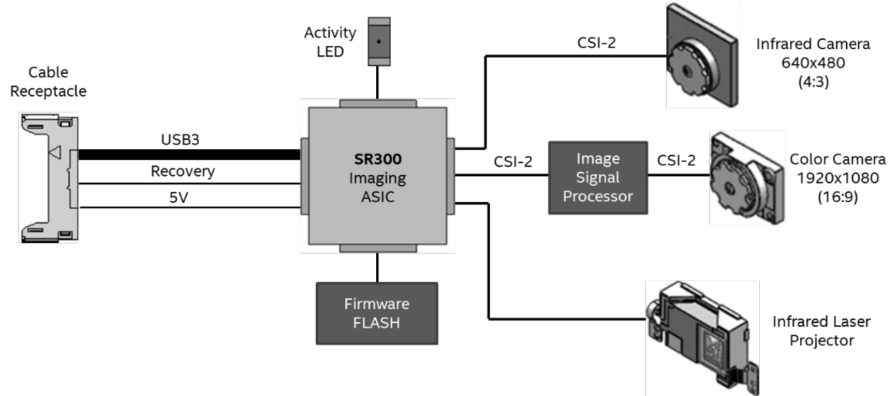


图 2.6 SR300 采集的物体深度信息部分缺失情况下的深度图

角度以及光线有关,因此本文提出一种组合相机对偶 RGB-D 相机 (Dual RGB-D Camera)。

2.3.1 对偶 RGB-D 相机原理与结构

对偶 RGB-D 相机在原 RGB-D 相机的基础上,通过增加一个与原相机呈 180 度夹角的 RGB-D 相机构成,实际物理结构如图2.7所示。

对于对偶 RGB-D 相机,当其中一个相机深度图出现严重缺失时,另外一个相机的深度图往往不会在相同的地方深度信息出现严重的缺失,如图2.8所示,有效的避免了单个 RGB-D 相机某些情况下深度信息严重缺失的情况。

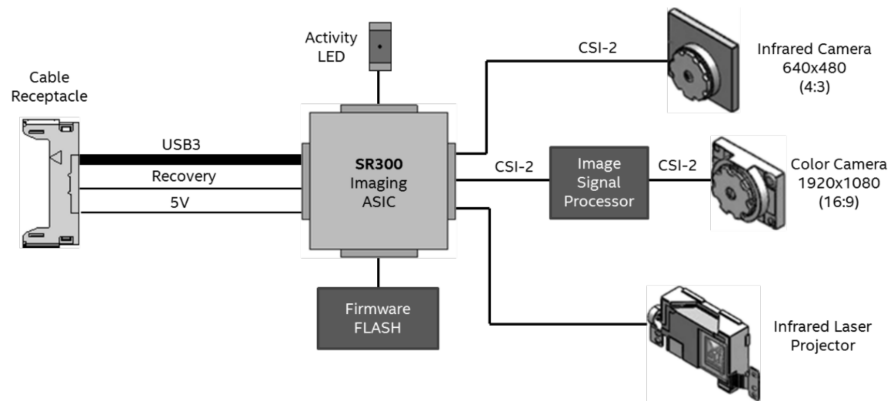


图 2.7 对偶 RGB-D 相机结构



图 2.8 对偶 RGB-D 相机采集的左右两张深度图

除此之外,对偶 RGB-D 相机还可以利用两个相机的彩色图构成双目,生成第三张深度图,从而通过设计的深度的融合算法将三张深度图融合成为一张质量更高的深度图,其内部原理如图2.9所示。

从外部使用来看,对偶 RGB-D 相机也输出一张彩色图、一张深度图。输出的彩色图就是从左相机采集到的彩色图;输出的深度图是由三张深度图融合而成,并且与输出的彩色图相对齐,对齐的意思是彩色图和深度图相同图像坐标下的颜色信息和深度信息对应的实际物理世界中相同的一点,对齐的意义在于方便后续的一些图像处理的算法。

从内部实现来看,主要涉及到三个部分:

- 将深度图与输出的彩色图对齐 (Align to RGB)
- 利用左相机采集的彩色图和右相机采集的彩色图,通过双目匹配算法形成一张新的深度图
- 融合左相机对齐后的深度图、右相机对齐后的深度图和双目匹配得到的深度图

将深度图与彩色图对齐,相对来讲实现还是比较简单的,以匹配左相机深度图到左相机彩色图为例,需要分别知道左相机深度相机和彩色相机的内参和畸变参数以及深度相机与彩色相机之间的齐次变换关系(通过相机标定这些参数都可以得到)。具体流程如算法1所示。

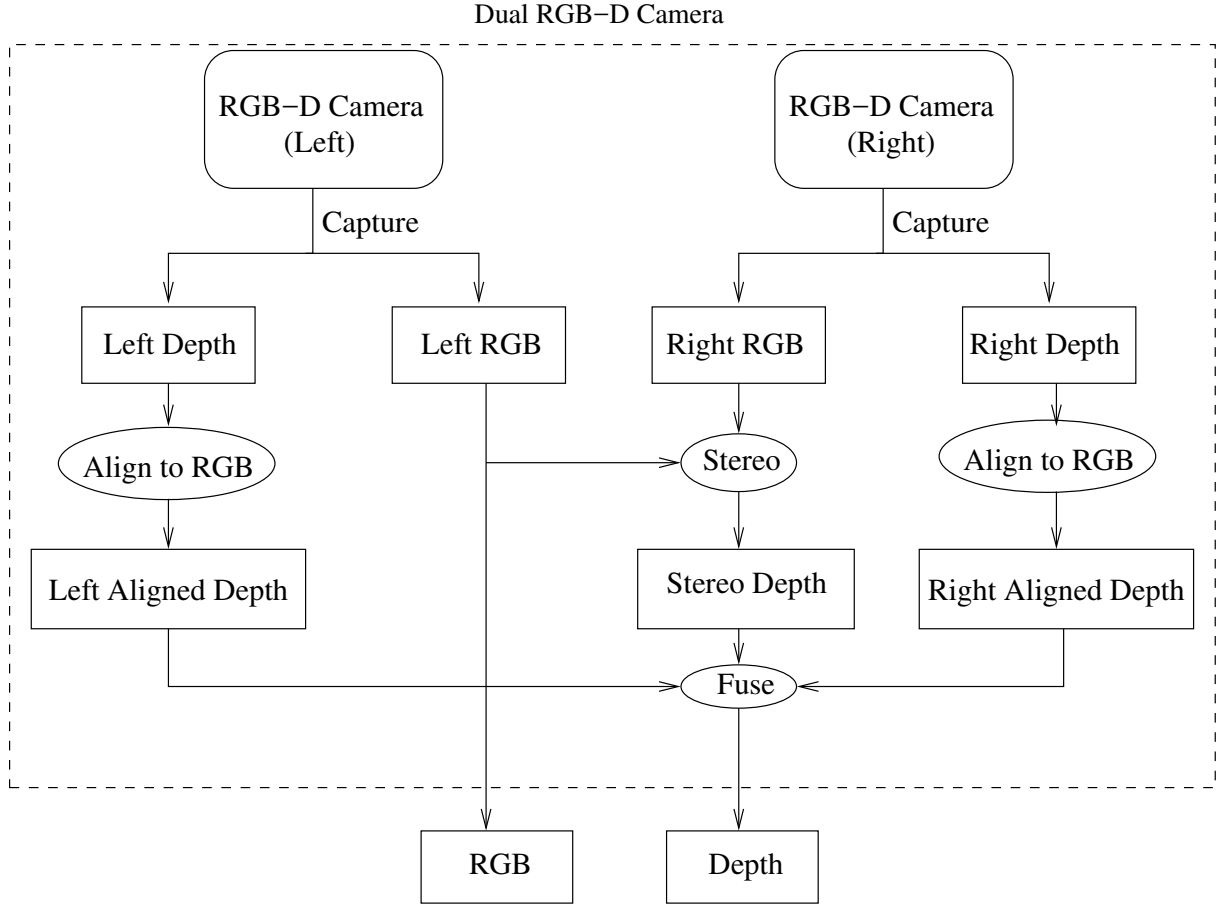


图 2.9 对偶 RGB-D 相机内部原理图

算法 1: Align Depth Frame

Input: Raw Depth Frame $Raw_D_{dh \times dw}$

Output: Aligned Depth Frame $Aligned_D_{ch \times cw}$

for p **in** $Aligned_D$ **do**

$p = 0$

for $dy = 1; dy \leq dh; ++dy$ **do**

for $dx = 1; dx \leq dw; ++dx$ **do**

 通过深度相机内参将点 (dx, dy) 反投影到三维空间一点 ${}^D\mathbf{X}$;

 坐标变换 ${}^R\mathbf{X} = {}^R_D\mathbf{R} {}^D\mathbf{X} + {}^R_D\mathbf{t}$;

 通过彩色相机内参将点 ${}^R\mathbf{X}$ 投影变换到彩色图像坐标系下一点

(cx, cy) ;

if cx **in** $(0, cw]$ **and** cy **in** $(0, ch]$ **then**

$Aligned_D(cx, cy) = Raw_D(dx, dy)$;

2.3.2 对偶 RGB-D 相机的数学模型

2.3.3 深度图像融合算法

2.3.4 对偶 RGB-D 相机的标定

第 3 章 基于 RGB-D 图像的目标检测算法

3.1 深度信息预处理

3.2 3D Faster RCNN

3.3 3D MRCNN

第 4 章 基于点云的位姿估计算法

4.1 ICP

4.2 Super4PCS

第 5 章 实验验证

第 6 章 结论与展望

6.1 结论

6.2 进一步工作的方向

致谢

逾尺的札记和研究纪录凝聚成这么薄薄的一本,高兴和欣慰之余,不禁感慨系之。记得鲁迅在一篇文章里写道:“人类的奋战前行的历史,正如煤的形成,当时用大量的木材,结果却只是一小块”。倘若这一小块有点意义的话,则是我读书生活的最好纪念,也令我对于即将迈入的新生活更加充满信心。回想读书生活,已经整整二十个年头,到同济求学将近五年,攻读博士学位也已三年了。进入同济大学以来,深深醉心于一流学府的大家风范。名师巨擘,各具特点;中西融合,文质相顾。处如此佳境以陶铸自我,实乃人生幸事。

2017 年 12 月

参考文献

- [1] BROWN D C. Decentering distortion of lenses[J/OL]. Photogrammetric Engineering and Remote Sensing, 1966. <https://ci.nii.ac.jp/naid/10022411406/en/>.
- [2] HEIKKILÄ J. Geometric camera calibration using circular control points[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(10): 1066–1077. DOI: 10.1109/34.879788.
- [3] KNUTH D E. The T_EX book[M]. 15th ed. Reading, MA: Addison-Wesley Publishing Company, 1989.
- [4] ZHANG Z. A Flexible New Technique for Camera Calibration (Technical Report)[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 22(11): 1330–1334. DOI: 10.1109/34.888718.

附录 A 补充资料

可能需要补充的内容……

个人简历、在学期间发表的学术论文与研究成果

个人简历

同济人,男/女,xxxx 年 xx 月生。

xxxx 年 xx 月毕业于 xxxx 大学 xxxx 专业获 xx 学位。

xxxx 年 xx 月入同济大学攻读 xx 学位。

已发表论文

[1] ...

[2] ...

[3] ...

已获得专利

[1] ...

[2] ...