基于深度学习的双目相机三维重构

刘谨鸿

( 电子科技大学 信息与通信工程学院)

3D reconstruction of binocular camera based on deep learning

Jinhong Liu

(School of Information and Communication Engineering, UESTC)

**摘 要**

基于计算机视觉的测量技术作为一种非接触式的先进测量技术，具有精度高、效率高、成本低等诸多优点，有着广阔的应用前景，双目立体视觉是计算机视觉的一个重要分支，其原理是由不同位置的两台摄像机(双目或者多目)或者一台摄像机经过移动或旋转拍摄同一幅场景，通过计算机空间点在两幅图像中的视差，获得该点的三维坐标值。由于双目测量具有一定的三维测量精度和测量的实时性，且花费的代价较其他方法要小的多，所以，双目视觉在测量方面应用广泛。三维重构是环境感知的关键技术之一，被广泛应用于无人驾驶、物体导航、无人机避障、医学诊断、遥感测绘、精密仪器测量等方面。本文中用到130度的广角双目相机对所拍摄事物做立体三维重构。

**关键词**: 计算机视觉、非接触式、双目立体视觉、摄像机标定、立体匹配、三维重构

**Abstract**

As a non-contact advanced measurement technology, measurement technology based on computer vision has many advantages such as high precision, high efficiency and low cost, and has broad application prospects. Binocular stereo vision is an important branch of computer vision. The principle is to capture the same scene by moving or rotating two cameras (binocular or multi-eye) or one camera at different positions, and obtain the three-dimensional coordinate value of the point through the parallax of the computer space point. two pictures. Because binocular measurement has certain 3D measurement accuracy and real-time measurement, and the cost is much lower than other methods, binocular vision is widely used in measurement. 3D reconstruction is one of the key technologies of environmental perception, which is widely used in unmanned driving, object navigation, UAV obstacle avoidance, medical diagnosis, remote sensing mapping, precision instrument measurement and other fields. In this paper, a 130-degree wide-angle binocular camera is used to reconstruct the subject in three dimensions.

**Keywords:** computer vision, non-contact, binocular stereo vision, camera calibration, stereo matching, 3D reconstruction

1. **概述**

1.1 项目研究意义

手势识别技术是人机交互领域的典型代表。在各类人机交互可传递的信息种类中，手势是最为特殊的一项。手势作为一种非语言类沟通方式，诞生时间早，表意相对直接，可在较大程度上弥补语言沟通的欠缺，在人日常生活中起到举足轻重的作用。手势识别技术，通过使计算机获取并理解人的手势信息，进而做出相应反应，可打破人机接触式交互的诸多限制，增加人机交互方式，拓宽交互场景。手势识别技术的应用领域十分广泛，如车辆驾驶环境中，通过手势对车内各项功能进行控制，可有效提升驾驶安全性；物联网及智能家居环境中，通过手势控制各类家居使用，如电视换台、音响控制等，可使人居生活更具便利性。

现有较为成熟的手势识别技术，大多通过光学传感器或可穿戴设备获取手势信息，进行手势识别，而这两种信息获取手段均存在其局限性。基于光学传感器的手势识别方案，通过摄像头采集手势信息，以二维图像

的数据形式进行呈现。该方案对手势所处环境的光照强度较为敏感，在光照强度变化剧烈时，识别准确率会显著降低。此外，光学图像的获取无法保证有效防止用户隐私泄漏问题，故采用光学传感器的方案存在较大隐忧。对于可穿戴设备的方案而言，其通过速度传感器、陀螺仪等对手势数据进行测量，存在便携性较差，成本较高等问题。基于上述局限性，研究基于其他传感器的手势识别方案具有很大意义。基于毫米波雷达的手势识别作为新兴技术，克服了传统基于光学图像的诸多问题，应用前景广阔，然而由于动态手势具有复杂多样、时空特征多变的特点，给人机交互和手势识别的研究带来了许多新的挑战，因此该技术在具有极大研究价值的同时，还具有一定挑战性。

1.2 国内外研究现状

国外：Jung等基于双目立体视觉技术研制了移动机器人，该机器人能够感知外界物体的运动情况，然后根据接受到的环境信息进行动态场景的处理。美国波古顿动力公司设计开发了基于立体视觉导航技术的仿生四足机器人Big Dog，该机器人能辨识周围环境，通过获取各种障碍物的距离信息进行移动和避障。 华盛顿大学与微软公司合作研制了宽基线立体视觉系统，使“探测者”号能够在火星上进行精确的定位与导航。系统使用同一个相机在“探测者”号的不同位置上拍摄图像对，然后采用非线性优化算法、最大似然概率法及高效的立体搜索得到亚像素精度级别的视差，再据此计算图像对中各点的三维坐标，实现实时的火星表面地形的三维重建。日本大阪大学自适应机械系统研究院设计并研制了自适应双目视觉伺服系统，以每幅图像中相对静止的三个标志点为参考，实时计算目标图像的雅可比矩阵，对目标下一步运动方向进行预测，实现了运动目标的自适应跟踪。与传统的视觉跟踪伺服系相比，该系统不需要事先知道相机的相关参数和目标的运动方式，仅要求拍摄的两幅图像中都有静止的参考标志物，操作简单，大大提高了系统工作效率。

国外：哈尔滨工业大学的高庆吉等开发出异构双目视觉系统，使机器人能够捕捉多个有效目标，实现了足球机器人的全自主导航。中科院自动化研究所开发了一款自主移动机器人，通过对周围环境进行图像采集并立体匹配，能够得到相关位置和距离信息，进而对自身的移动进行导航。南京航空航天大学的张凤静等利用双目立体视觉测量汽车安全车距，根据车辆分别在左右相机拍摄的图像中的坐标计算得到车辆到相机的距离。浙江大学左爱秋等利用双目立体视觉，只需从采集到的两幅图像中得到必要的特征点的三维坐标，就能实现对多自由度机械装置精确地进行动态位姿检测，该方法处理速度快，特别适合对动态情况的检测。天津大学精密测试技术与仪器国家重点实验室研制了“夏利轿车车身尺寸视觉检测站”，可以用于汽车工业中在线测试车身总长。该系统用多个视觉传感器分别测量车身上一个关键点在传感器坐标系 中的三维坐标，再将所有测量结果统一到工件坐标系中，得到最终的测量结果。清华大学智能技术与系统国家重点实验室采用典型的平行双目立体视觉系统开发了拟人机器人TBIPR-1，身高1.60m，体重100～210 kg，能根据视觉、触觉和听觉自主实现步行、慢跑、跨越障碍、拾 取物体以及其他与人类相似的较高级运动。[1]

1.3 项目难点和创新点

双目相机虽然成本低廉，在三维重构方面取得了一定的成果，但依旧有着计算量大，重建效果不好，价格高，自动化程度不高等缺点。后来随着深度神经网络的发展，许多基于CNN的方法被提出，然而，大多数技术只能生成一个深度值数组，没有平面检测或者分割，达不到的物体或场景重建。因为分段平面深度图的重建需要结构化的几何表示，并且我们不知道要预测的平面数，也不知道在输出特征向量中要回归的平面数顺序，这对深度神经网络也是十分困难的。

**创新点：**

(1) 在众多三维重建论文中，暂没有一个完整的项目代码只用双目相机进行双目立体视觉的一个三维重构，许多都用结构光相机或者深度相机去进行模型的重构。

(2) 使用SGC模型做类似滤波的处理，滤除比较尖锐的噪声点。

**2 系统设计原理与方案**

2.1 原理

相机通过目标物体表面的反射光将空间三维物体转化为平面二维图像，而双目立体视觉是相机成像的逆过程，即以两台相机模拟人眼，通过计算两幅图像对应点间的位置差异并基于视差原理得到物体的三维几何信息，实现三维重建及还原。三维重构的步骤主要分为：图像获取、摄像机标定、畸变校正、立体匹配、三维重构等方面。

# 2.1.1 标定

现在常用的标定方法是用棋盘格进行标定，是张正友教授于1998年提出的单平面棋盘格的摄像机标定方法。世界坐标系固定于棋盘格上，则棋盘格上任一点的物理坐标 W=0，由于标定板的世界坐标系是人为事先定义好的，标定板上每一个格子的大小是已知的，我们可以计算得到每一个角点在世界坐标系下的物理坐标 (U,V,W=0)。我们将利用这些信息：每一个角点的像素坐标 (u,v)、每一个角点在世界坐标系下的物理坐标 (U,V,W=0)，来进行相机的标定，获得相机的内外参矩阵、旋转矩阵、平移矩阵、畸变参数(一般会有畸变)等。

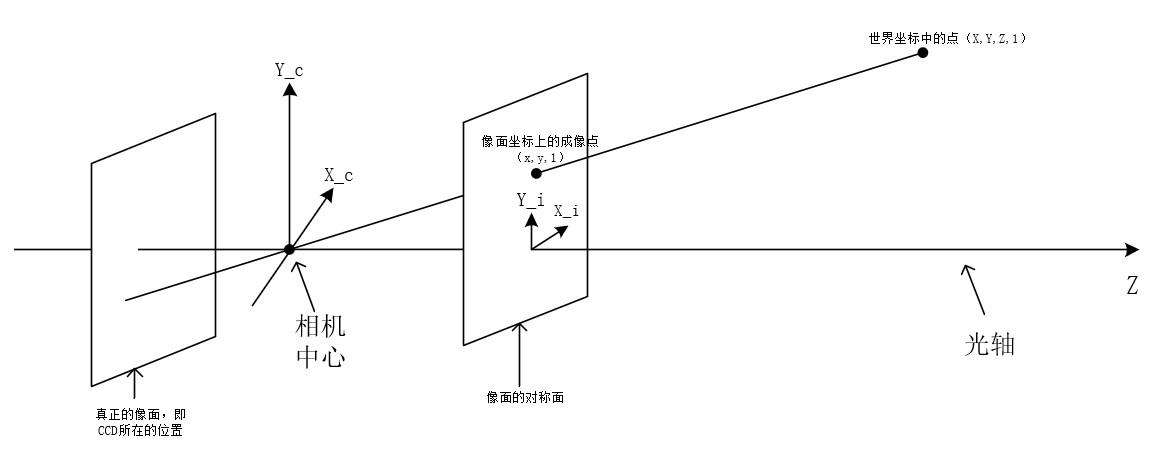


图 2-1 标定原理图

# 2.1.2 畸变校正和立体校正

镜头畸变实际上是光学透镜固有的透视失真的总称，也就是因为透视原因造成的失真，这种失真对于照片的成像质量是非常不利的，毕竟摄影的目的是为了再现，而非夸张，但因为这是透镜的固有特性（凸透镜汇聚光线、凹透镜发散光线），所以无法消除，只能改善。畸变校正就用标定得到的畸变系数即可校正。

立体校正的目的是将拍摄于同一场景的左右两个视图进行数学上的投影变换，使得两个成像平面平行于基线，且同一个点在左右两幅图中位于同一行, 简称共面行对准。只有达到共面行对准以后才可以应用三角原理计算距离。这需要对其中某一个相机的坐标系进行旋转或者平移使他们达到这样的效果。校正后可以用对极线查看是否校正。

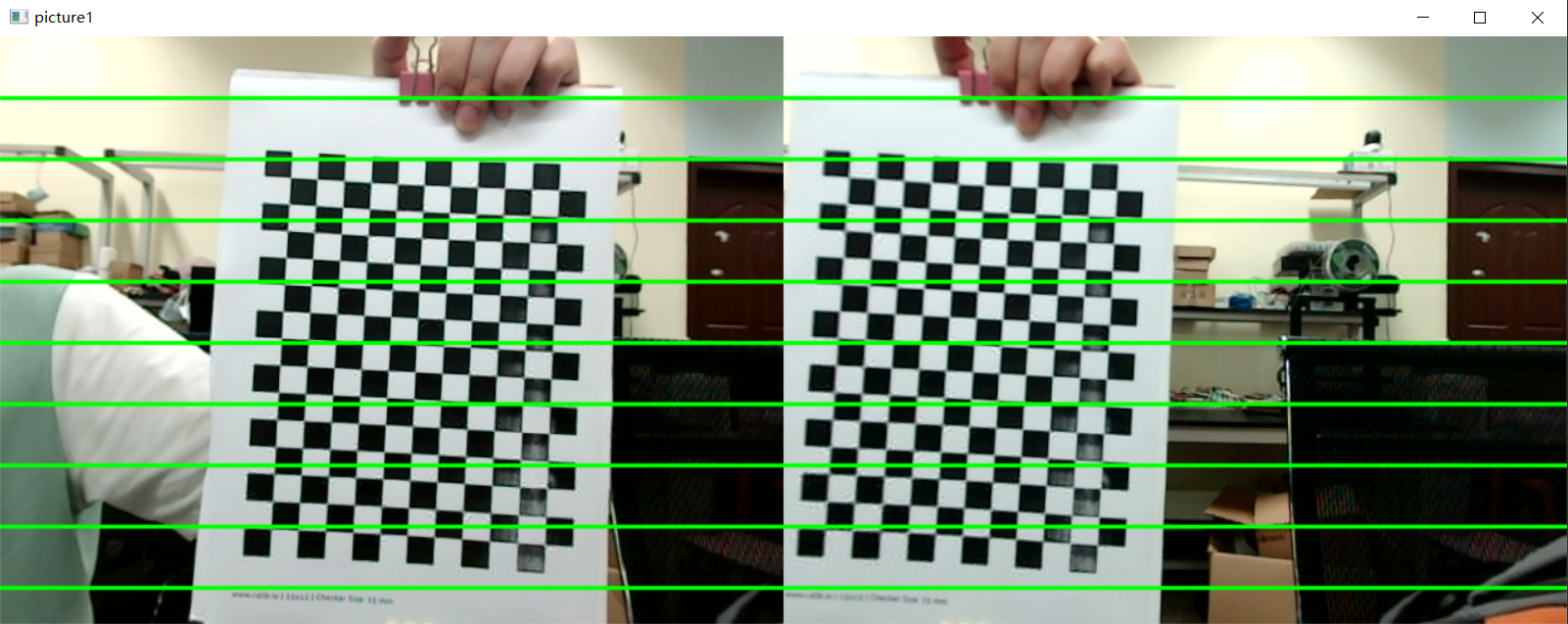


图 2-2 立体校正图

# 2.1.3 立体匹配

立体匹配的原理类似于人类的视觉原理：分别单独使用左眼和单独使用右眼去观察一个物体，该物体在两个视角中的相对位置会存在一定的偏移，人类视觉系统可以根据该偏移去判断该物体的距离远近。同样的，在双目立体匹配中，只要能在两张图像中正确地找到匹配点，结合相机的内部参数和外部参数，就能精准地计算出空间点距离拍摄相机的距离。

立体匹配的目的是为左图中的每一个像素点在右图中找到其对应点（世界中相同的物理点），这样就可以计算出视差，大致可以分为匹配代价计算、代价聚合、视差优化、视差细化几个阶段，其难点在

1.图像中可能存在重复纹理和弱纹理，这些区域很难匹配正确；

2.由于左右相机的拍摄位置不同，图像中几乎必然存在遮挡区域，在遮挡区域，左图中有一些像素点在右图中并没有对应的点，反之亦然；

3.左右相机所接收的光照情况不同；

4.过度曝光区域难以匹配；

5.倾斜表面、弯曲表面、非朗伯体表面；

6.较高的图像噪声等。

本文中使用了较为优秀的代价匹配算法，这里不做深究。

## 2.1.4 得到视差图和深度图

立体匹配之后即可得出视差(两个对应点在图像中的列坐标的差)，深度图可以由下面的式子得到

式2-1

## 2.1.5 得到视差图和深度图

得到深度图(包含距离信息)以后，就可以根据距离画出点云图，且可以根据距离给深度图附上颜色。

2.2 方案设计

本文只用到了一个130°的广角双目相机 。

2.2.1 双目相机标定

在原先方案上想直接用Python进行标定，也方便直接写入配置文件，但Python标定程序标定后误差过大，没有采取使用，最后使用matlab标定得到的参数。

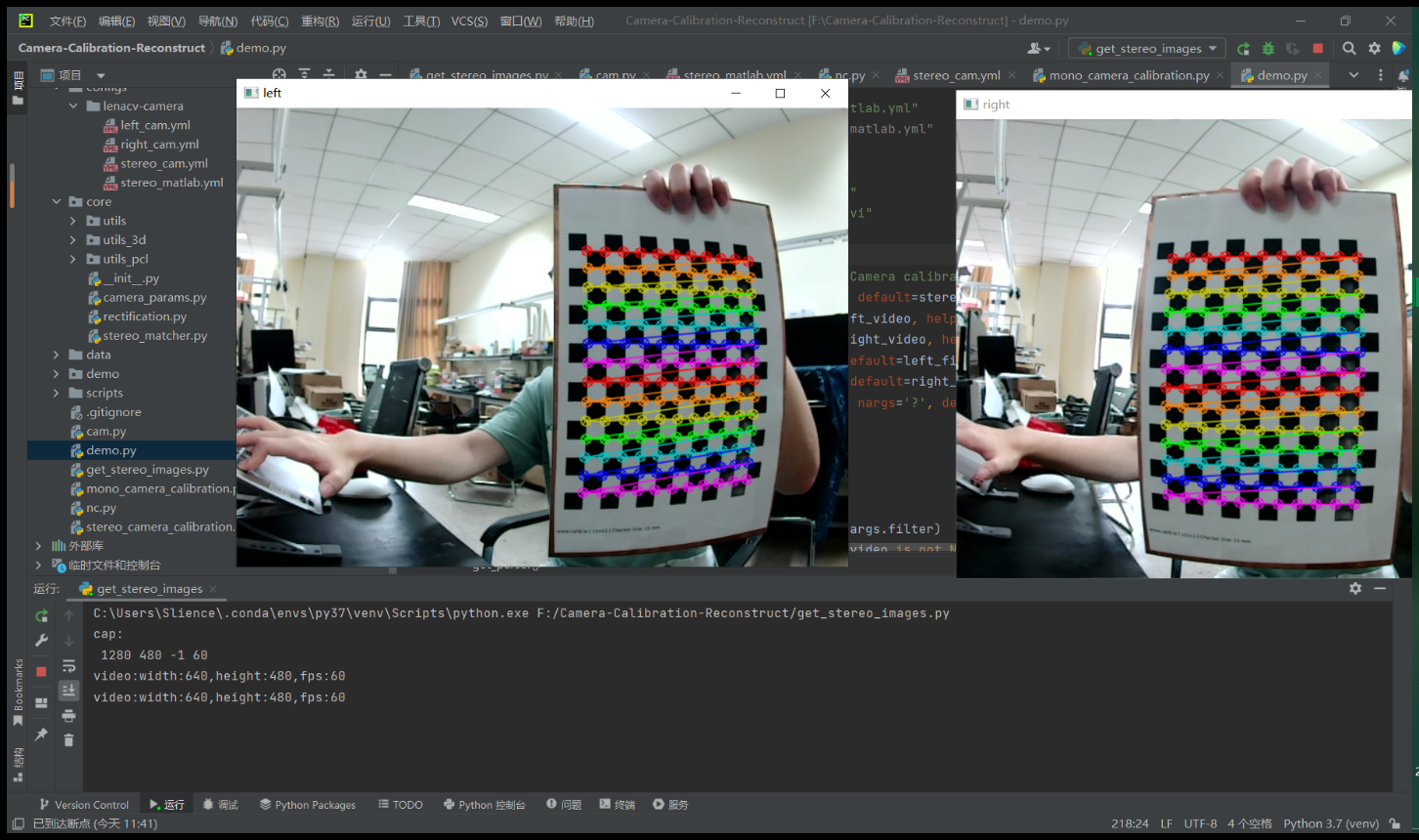


图 2-3 标定实例图

2.2.2 畸变校正、立体校正及立体匹配

此部分用Python库函数直接计算得到。

# 2.2.3 视差图空洞填补

因为得到的视差图不知由于什么原因，有较多处空洞，这里用多层均值滤波填补空洞(将大区域的空洞赋值。然后下次滤波时，将窗口尺寸缩小为原来的一半，利用**原来的积分图**再次滤波，给较小的空洞赋值覆盖原来的值)。

同时用模型(简化图像卷积)对深度图进行更深一步处理(相当于滤波器)，使点更加平滑。

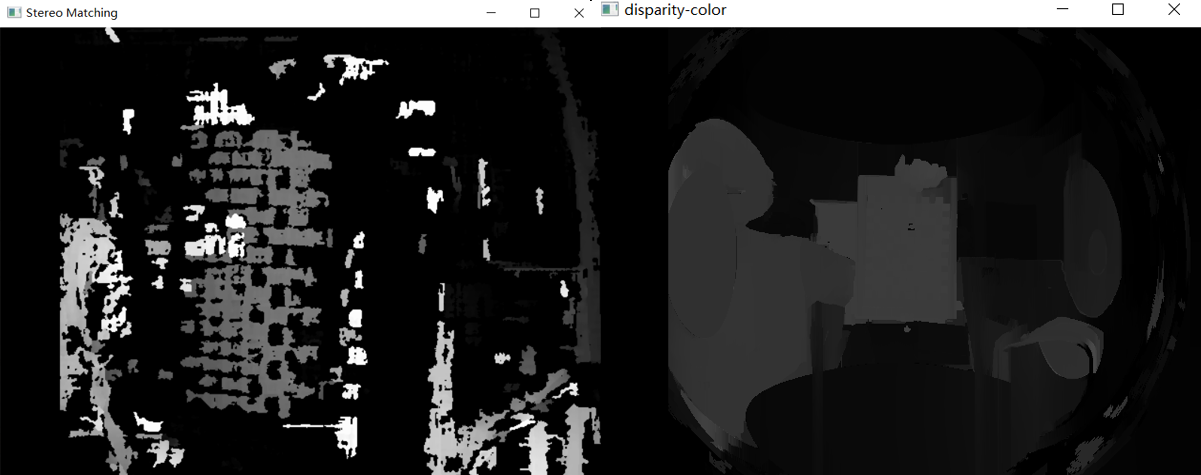


图 2-4 视差图

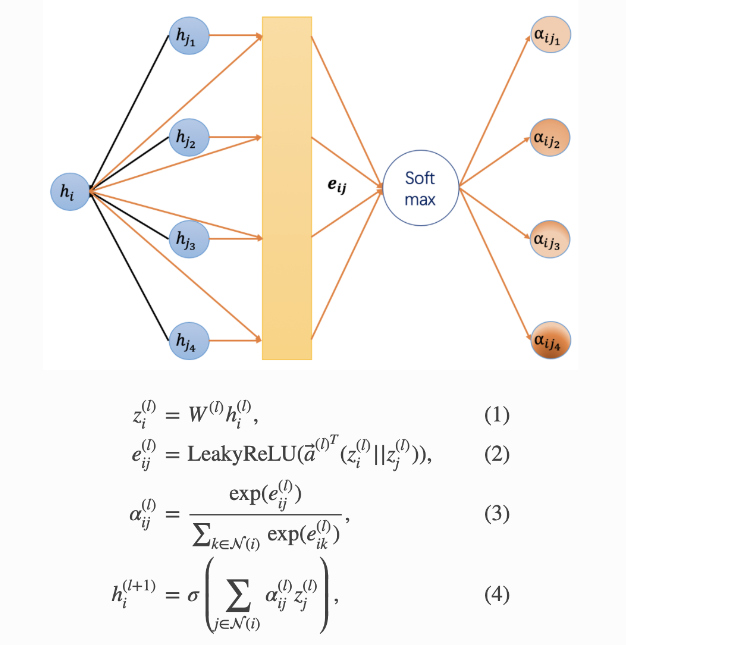


图 2-5 SGC模型

SGC模型：

式2-2

式2-3

第一个公式表示一个特征提取器,s表示对图中所有节点求次1跳距离的平均值、代表节点的特征值。可以看到在求过程中,不需要有参数参与,该过程可以放在样本的预处理环节

第二个式子表示一个分类器，该过程与深度学习中的分类器完全一致。对特征提取后的数据进行全连接神经网络处理，然后再通过SoftMax进行分类。SGC模型突破了GCN的层数限制，将GCN每层的激活函数去掉(不需要非线性变换)。利用图中的节点关系，直接计算图节点间局部邻居的平均值。通过多次计算节点间1跳距离的平均值来实现卷积叠加的效果。

**3 测试分析**

# 3.1 测试

在多次录制视频后进行一个三维重构，用卷尺对一些位置信息进行测量，测量误差20mm内浮动，偶尔会有突然距离值的跳变。

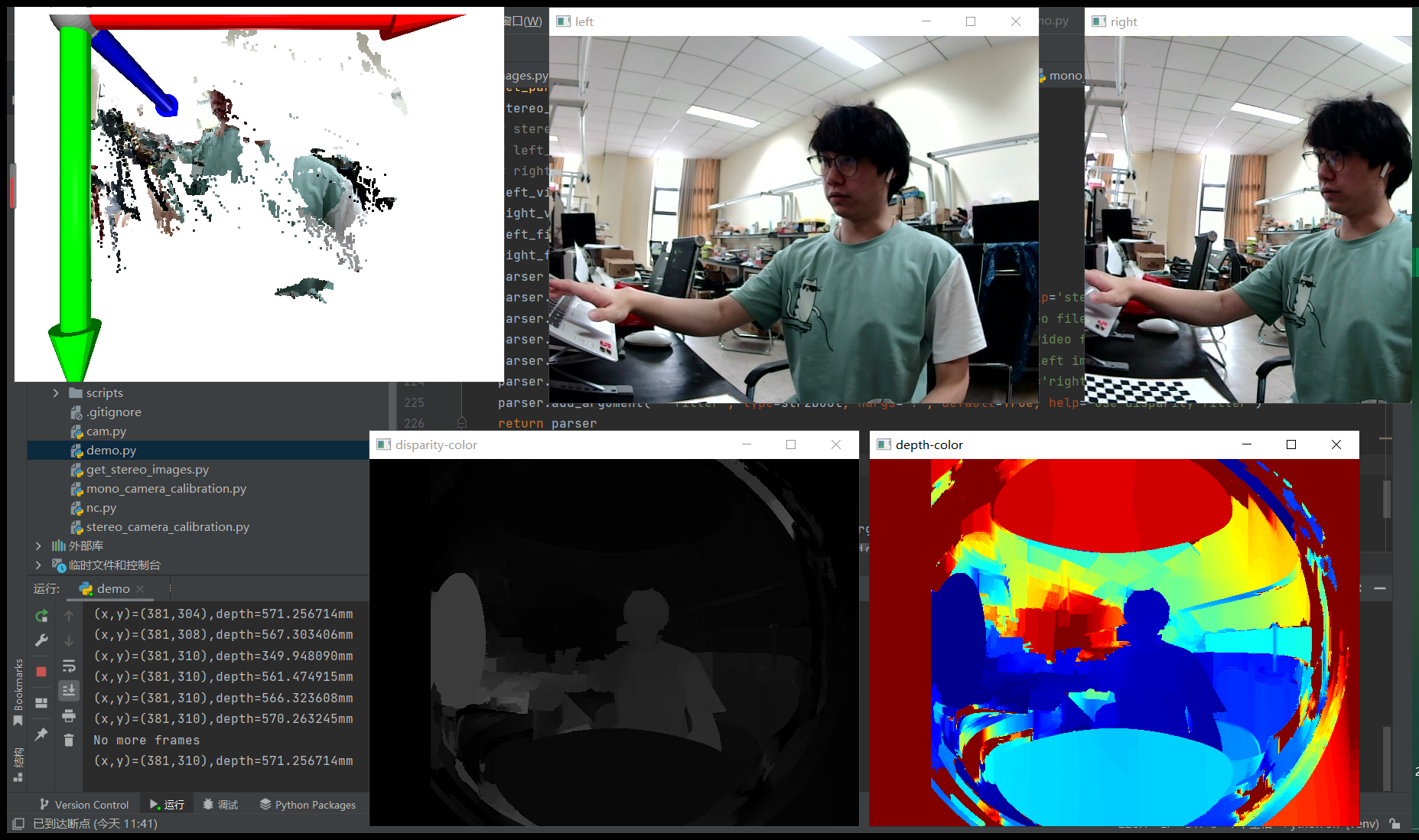


图 3-1 测试

# 3.2 分析

由图3-1可以看出，图像正上，正下和正左有几个相对平滑的点，因为都处于边缘部分，可能是由于边缘畸变无法完全消除，在进行matlab标定时已经可以看出，误差在0.12就降不下来了，反观网络上很多博客写的尽量让校准误差小于0.1，我也看到很多误差只有0.05的相机，后续工作中可能会尝试畸变稍小的相机。不使用多层均值滤波就不会得到较为良好的视差图，这也无法判定是否是后续过程出现问题，本文这里只能处理成这样。

4 总结

本项目基本实现了开题的目标，但open-3d显示点云数量过少，之前用过PCL显示点云，效果不错，显示范围较广，但后续电脑清理了一次之后开始报错，暂时没考虑。摄像头实时显示距离信息打印点云暂时没有实现。

**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. DOI:10.3969/j.issn.1005-5630.2018.04.013.

[2] X. Lin, J. Wang and C. Lin, "Research on 3D Reconstruction in Binocular Stereo Vision Based on Feature Point Matching Method," 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), 2020, pp. 551-556, doi: 10.1109/ICISCAE51034.2020.9236889.