



# 图像处理与机器视觉 大作业报告

基于双目相机系统的  
实时深度图与三维点云生成

任课老师：蒋磊，范睿

学生：2152482 鄂俊伍，2151094 宋正非，2152222 邹阳

班级：10059901

2025 年 7 月

## 摘要

本研究构建了一个基于双目相机系统的三维重建平台，实现了从图像采集到三维点云生成的完整流程。系统采用 HBVCAM-W202011HD V33 双目相机作为硬件核心，通过严格的相机标定流程获取内参矩阵、畸变系数和外参矩阵。研究对比了 Block Matching (BM) 和 Semi-Global Block Matching (SGBM) 两种立体匹配算法，其中 SGBM 算法通过引入路径代价聚合机制，显著提升了视差图的质量。系统实现了深度图的实时生成，并通过反投影计算将二维图像坐标转换为三维点云。实验结果表明，该系统能够有效重建三维场景，物体长度测量误差控制在 10% 以内。研究还探讨了基于 SIFT 特征的双目图像矫正与匹配方法，进一步提高了匹配精度。最终生成的稠密点云可用于三维场景理解和目标测量等应用。

**关键词：**双目视觉、立体匹配、深度图、三维重建、点云生成

# 目 录

摘要 .....	I
第一章 任务概述 .....	1
1.1 问题描述 .....	1
1.2 问题背景 .....	1
1.3 方法描述 .....	1
第二章 实验 .....	3
2.1 建立双目相机系统 .....	3
2.2 相机标定 .....	3
2.2.1 定义世界坐标 .....	3
2.2.2 多角度获取图像 .....	4
2.2.3 找到棋盘的二维坐标 .....	4
2.2.4 标定相机 .....	6
2.2.5 计算重投影误差 .....	7
2.2.6 标定结果 .....	9
2.3 创建立体图像并生成特征对 .....	10
2.4 生成视差图 .....	11
2.5 计算深度图 .....	12
2.6 实物测试 .....	12
2.7 结果讨论 .....	14
第三章 拓展任务 .....	16
3.1 提升深度图质量 .....	16
3.1.1 基于 SIFT 的双目图像矫正与匹配流程 .....	16
3.1.2 关键代码与功能对照 .....	16
3.1.3 立体校正验证原理 .....	16
3.1.4 SIFT 特征提取与匹配原理 .....	17
3.1.5 匹配质量评估指标 .....	17
3.2 算法优化 .....	17
3.2.1 Semi-Global Block Matching (SGBM) 算法原理 .....	17
3.2.2 SGBM 与 BM 的对比分析 .....	19
3.2.3 算法优化后的实物测试 .....	19
3.3 建立 3D 点云 .....	19
3.3.1 图像校正与立体匹配 .....	19
3.3.2 深度计算与点云生成 .....	22

3.3.3 点云可视化 .....	22
3.3.4 系统框架总结 .....	22
第四章 总结 .....	24
第五章 成员分工 .....	25
参考文献 .....	26

# 第一章 任务概述

## 1.1 问题描述

本课设旨在构建一个基于双目相机系统的三维重建平台，能够实时生成深度图（Depth Map）和三维点云（3D Point Cloud）。任务要求仅使用 C++ 语言和 OpenCV 库，从零开始完成整个系统的搭建与算法实现，最终实现对真实物体的三维结构测量，确保测量误差控制在 10% 以内。

## 1.2 问题背景

双目视觉是模仿人类眼睛感知深度的计算机视觉技术，其核心思想是通过两台相机从不同视角获取同一场景的图像，利用视差信息恢复物体的三维结构<sup>[1]</sup>。该技术被广泛应用于自动驾驶、机器人导航、三维建模等领域<sup>[2-4]</sup>。构建一个完整的双目相机系统，需要完成硬件搭建、相机标定、图像校正、立体匹配、深度计算等多个步骤，具有较强的工程综合性与算法实现挑战。

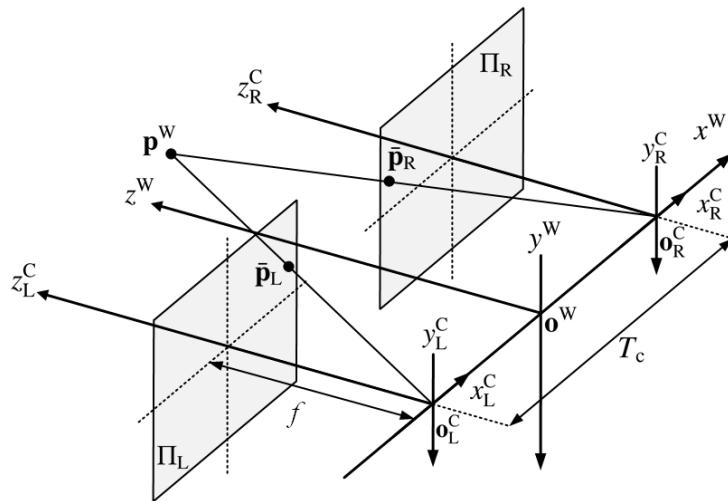


图 1.1: 立体视觉系统

## 1.3 方法描述

传统的立体匹配算法通常包括四个主要步骤：代价计算，代价聚合，视差优化，视差精化。由于其支持区域通常为矩形块状，这类方法也被称为立体块匹配 Block Matching

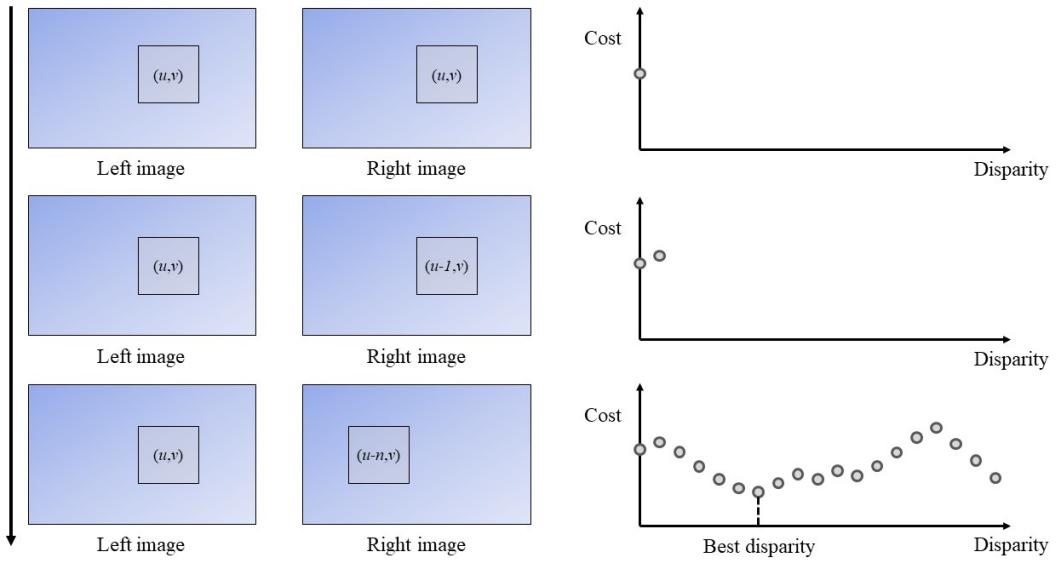


图 1.2: 立体匹配原理

(BM) 算法。当卷积过程采用线性盒式滤波 (linear box filtering) 时,  $c_{AD}$  和  $c_{SD}$  的聚合方式分别对应于绝对差值和 (SAD, Sum of Absolute Differences) 与平方差值和 (SSD, Sum of Squared Differences)<sup>[5]</sup>。其数学表达如下所示:

$$C_{SAD}(p, d) = \sum_{q \in N_p} |i_L(q) - i_R(q - d)| \quad (1.1)$$

$$C_{SSD}(p, d) = \sum_{q \in N_p} (i_L(q) - i_R(q - d))^2 \quad (1.2)$$

其中,  $N_p$  是像素  $p$  的邻域。

## 第二章 实验

### 2.1 建立双目相机系统

我们使用双目相机 HBVCAM-W202011HD V33 作为系统的核心感知设备。该型号相机具有双镜头同步输出、高帧率、USB 接口等特性，适合用于立体匹配与三维重建实验。通过合理设计相机基座并进行精准的标定，可充分发挥该相机的潜力，构建稳定可靠的双目三维感知系统。



图 2.1: 双目相机

### 2.2 相机标定

相机标定是使用包含标定图案的图像来估算相机参数的过程。这些参数包括相机固有矩阵、畸变系数和相机外参矩阵。使用这些相机参数可以消除图像中的镜头畸变效应、测量平面物体、从多个相机重建三维场景，以及执行其他计算机视觉应用。

#### 2.2.1 定义世界坐标

在双目相机系统的标定过程中，必须建立图像像素坐标与真实世界坐标之间的映射关系。为此，我们采用棋盘格标定板作为标准的二维标定参考物，定义真实世界坐标。通过检测棋盘格的角点，可以在图像平面上精确提取特征点的位置，并与已知的三维坐标对应，从而完成坐标系定义和相机几何模型的估计。

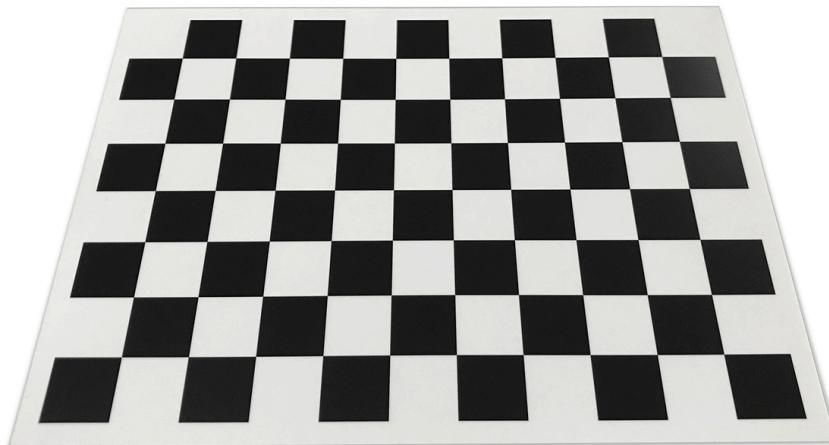


图 2.2: 使用棋盘格图案定义真实世界坐标系

### 2.2.2 多角度获取图像

为了实现高精度的双目相机标定，必须从多个不同视角采集棋盘格图案的图像。这一步骤是整个标定流程中至关重要的一环，直接影响到内参、畸变参数和外参的求解精度。

通过多视角图像采集，我们可以构建更全面的图像坐标点与真实世界坐标点之间的对应关系，进而增强相机几何模型的鲁棒性和拟合效果。不同的拍摄角度和位置能覆盖更大的图像区域，有助于准确估计镜头畸变、焦距偏移以及光轴偏差等误差。

我们同步采集左右图像，共采集了三十几组图像对。每一组图像来自不同的观察角度和空间位置，从正前方、斜上方、斜下方、左侧、右侧等多种角度观察，并改变棋盘格与相机之间的距离。在拍摄过程中保证图像质量，保持清晰，光照均匀，图像未过曝或欠曝；棋盘格完整可见，且角点清晰分明。在图像采集的同时，我们通过代码直接把图像进行裁切并按左右分开存储到两个文件夹并命名，例如：left\_0.png, right\_0.png。每一张图片的尺寸为  $1280 \times 720$ ，后续代码处理的尺寸需要与之统一，否则由于尺寸异常而无法正常获取信息。

### 2.2.3 找到棋盘的二维坐标

在完成多组视角下棋盘格图像的采集后，接下来的关键任务是从图像中提取棋盘格的角点位置，即获得每幅图像中棋盘格角点的二维图像坐标。

我们使用了 MATLAB 的 Stereo Camera Calibrator Toolbox 来完成角点提取与数据处

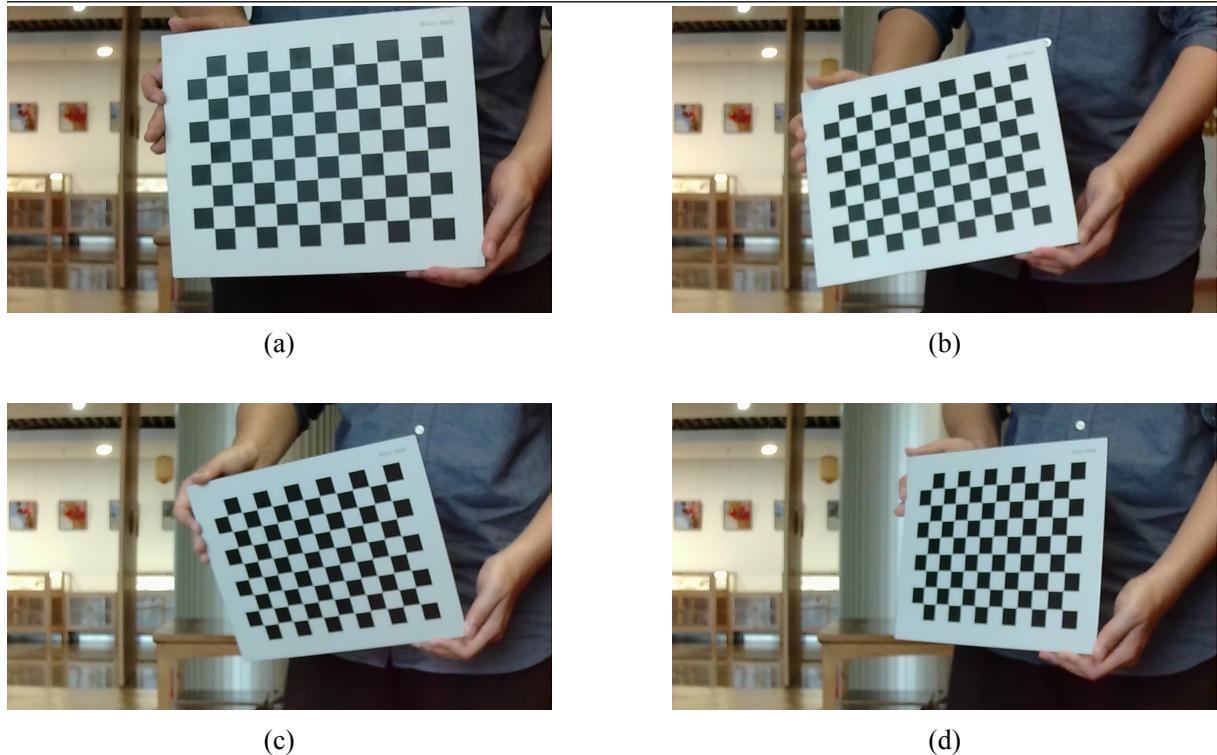


图 2.3: 采集图像示例

理。该工具箱提供了图形化的用户界面，能够自动检测棋盘格角点，并进行角点对齐、重投影误差计算和参数优化。

具体步骤包括：

- 导入左右相机采集的多组棋盘格图像对。
- 设置角点数量、格子尺寸等棋盘格的实际几何参数。
- 自动在每张图像中识别棋盘格图案，并提取角点。
- 对检测结果进行可视化校验，剔除检测失败的图像。

角点检测的基本原理是寻找棋盘格中黑白方块交界处的内角点，这些角点通常在图像中呈现为灰度变化明显的区域。MATLAB Stereo Camera Calibrator Toolbox 工具箱内部基于图像梯度与模板匹配方法，结合 Harris 或 FAST 算法，对图像进行处理并提取所有角点的亚像素精度位置。

由于左右相机同时观测到同一个棋盘格图案，系统会自动将左图与右图中的对应角点进行配对。这些匹配对在立体标定过程中用于估计双目系统的外参，即旋转矩阵和平移向量，进而实现三维重建。

## 第二章 实验

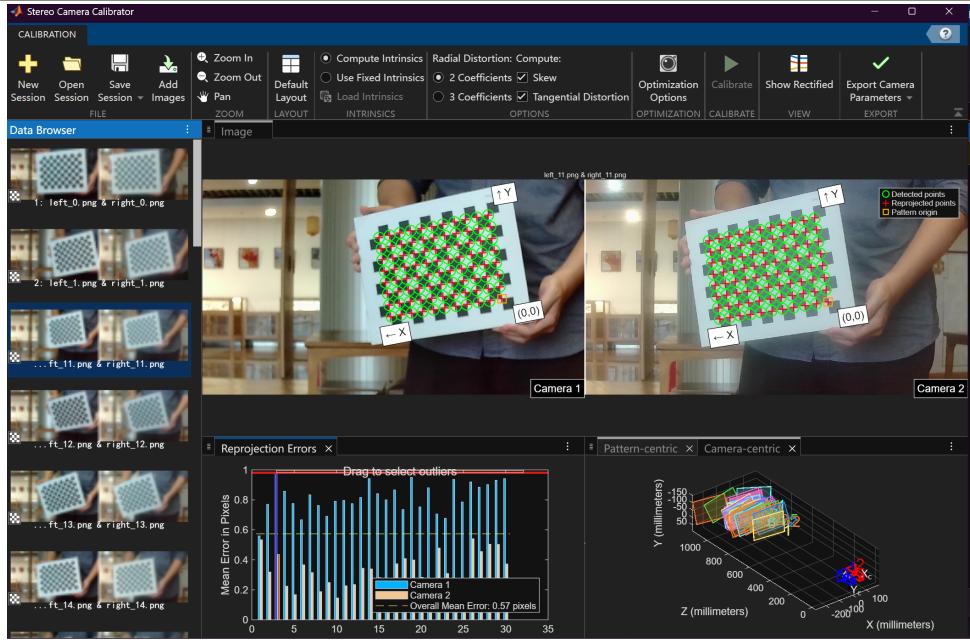


图 2.4: 相机标定

### 2.2.4 标定相机

完成角点提取后，我们进入了整个立体相机系统构建中最关键的一步——相机标定。标定的目的是求解出相机的内参、外参以及畸变系数等，从而建立真实世界与图像像素之间的几何映射关系。

内参矩阵用于连接成像平面坐标系与图像坐标系。其可表示为，

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & u_o \\ 0 & f_y & v_o \\ 0 & 0 & 1 \end{bmatrix}$$

其中， $[u_o, v_o]$  为主点坐标， $f_x = f \cdot s_x$ ,  $f_y = f \cdot s_y$ ,  $f$  为焦距， $s_x$  和  $s_y$  分别是水平和垂直方向上测量的有效尺寸（单位为 pixel/mm）。

畸变系数用于修正镜头产生的径向畸变和切向畸变。畸变系数可表示为

$$\text{distCoeffs} = [k_1, k_2, p_1, p_2, k_3]$$

$k_1, k_2, k_3$  为径向畸变系数；径向畸变主要由于镜头的球面结构引起，表现为图像中心向外放射状弯曲或收缩； $p_1, p_2$  为切向畸变系数；切向畸变由镜头组件在安装时发生偏心或倾斜引起，表现为图像点沿非径向方向偏移，导致图像整体出现“歪斜”或“拉伸”。

外参矩阵描述了相机相对于棋盘格世界坐标系的姿态，包括旋转和平移。

在以相机为中心 (camera-centric) 的表示中，棋盘格的世界坐标被转换到每个相机的坐标系下，目的是了解标定板在相机视角下的空间位置。

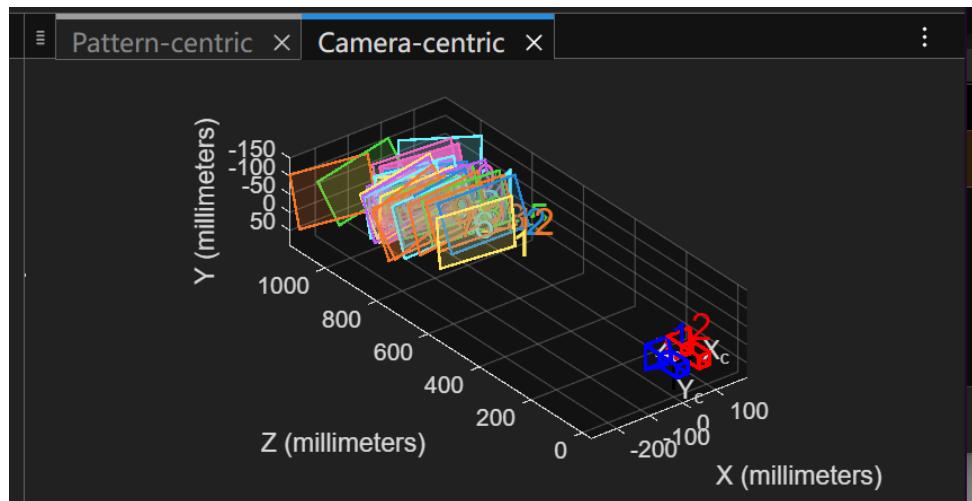


图 2.5: camera-centric result

在以棋盘格为中心 (pattern-centric) 的表示中，我们将棋盘格坐标系作为世界坐标系的参考原点，计算两个相机相对于棋盘格的位姿。

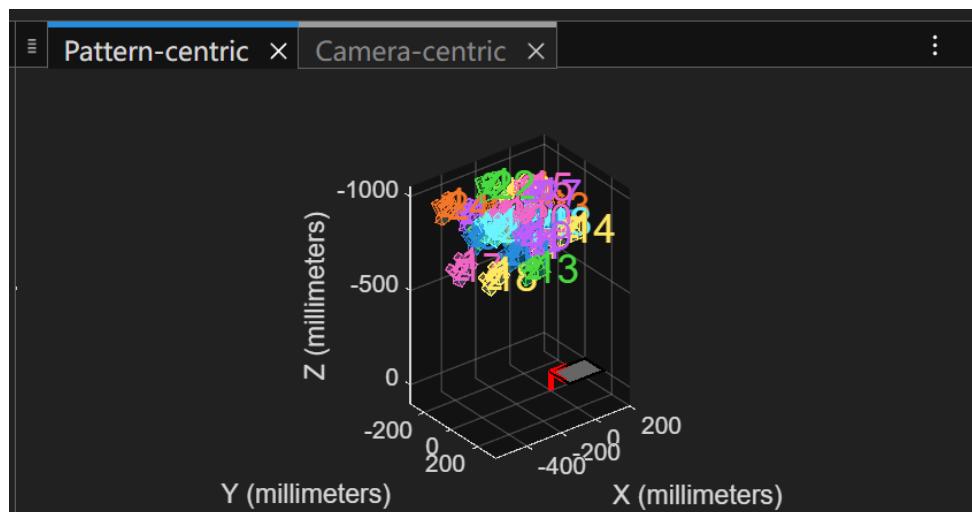


图 2.6: pattern-centric result

## 2.2.5 计算重投影误差

工具箱会输出每一对图像的重投影误差 (Reprojection Error)，表示预测的角点位置与实际检测位置之间的偏差。较小的重投影误差表示标定结果更准确。我们通过剔除误差较大的图像对，提高了整体标定的精度。

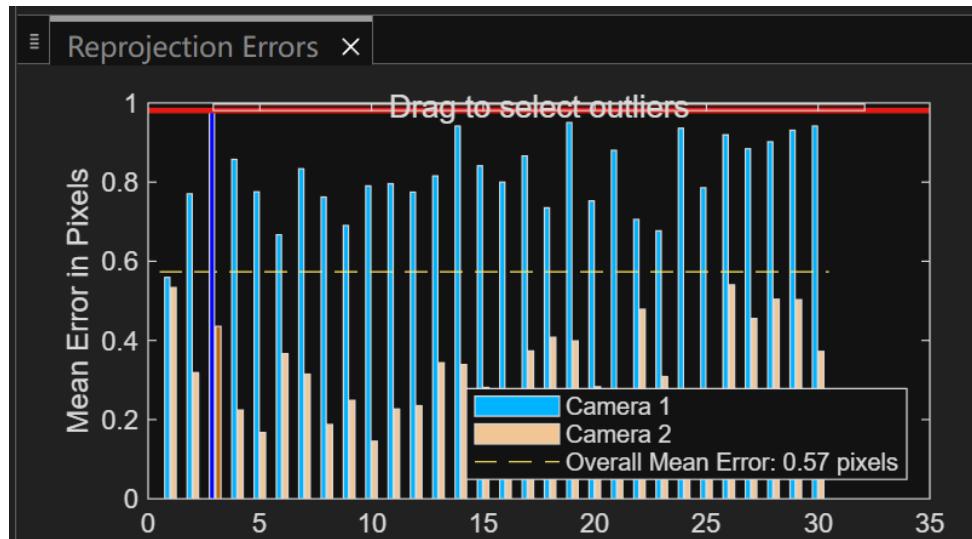


图 2.7: 重投影误差

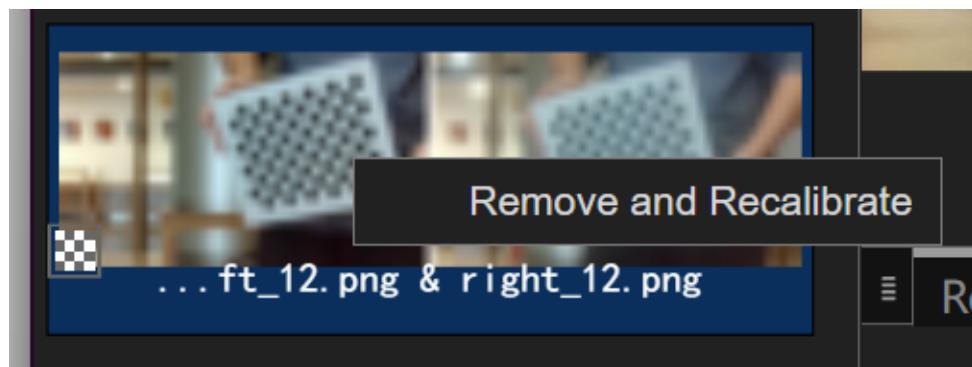


图 2.8: 移除异常值

## 第二章 实验

### 2.2.6 标定结果

通过上述步骤的双目相机标定，我们得到了内参矩阵、畸变系数、外参矩阵。

参数	值	维度	类型
(a) stereoParams.CameraParameters1.K	$\begin{bmatrix} 1997.2 & 3.0610 & 748.5226 \\ 0 & 1991.2 & 385.5766 \\ 0 & 0 & 1 \end{bmatrix}$	$3 \times 3$	double
(b) stereoParams.CameraParameters1.Intrinsics.K	$\begin{bmatrix} 1991.7 & 4.9914 & 672.5531 \\ 0 & 1982.5 & 390.8599 \\ 0 & 0 & 1 \end{bmatrix}$	$3 \times 3$	double
(c) stereoParams.CameraParameters1.Distortion	$\begin{bmatrix} [720, 1280] \\ [-0.377, -0.104] \\ [0.001, -0.007] \end{bmatrix}$	$1 \times 2$	double
(d) stereoParams.CameraParameters1.TangentialDistortion	$\begin{bmatrix} [720, 1280] \\ [0.410, 0.2374] \\ [0.001, -0.0070] \end{bmatrix}$	$1 \times 2$	double
(e) stereoParams.PoseCamera2.T	$\begin{bmatrix} 1 & 2 & 3 \\ 0.9999 & -0.0011 & 0.0104 \\ 0.0011 & 1.0000 & 0.0001 \\ -0.0104 & -0.0001 & 0.9999 \end{bmatrix}$	$3 \times 3$	double
(f) stereoParams.PoseCamera2.R	$\begin{bmatrix} 1 & 2 & 3 \\ -68.0171 & 2.1145 & 1.8585 \end{bmatrix}$	$1 \times 3$	double

图 2.9: 标定结果

左相机内参矩阵  $\text{cameraMatrixL}$ ,

$$\mathbf{K}_L = \begin{bmatrix} 1997.2000 & 3.0610 & 748.5226 \\ 0.0000 & 1991.2000 & 385.5766 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

其中,  $f_x$  为 1997.2,  $f_y$  为 1991.2,  $u_o$  为 748.5226,  $v_o$  为 385.5766。

右相机内参矩阵  $\text{cameraMatrixR}$ ,

$$\mathbf{K}_R = \begin{bmatrix} 1991.7000 & 4.9914 & 672.5531 \\ 0.0000 & 1982.5000 & 390.8599 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

其中,  $f_x$  为 1991.7,  $f_y$  为 1982.5,  $u_o$  为 672.5531,  $v_o$  为 390.8599。

在理想情况下, 摄像机的像素坐标轴应相互垂直, 即主点坐标系的  $x$  轴与  $y$  轴应呈  $90^\circ$  角, 对应的偏斜因子 (skew) 应为 0。然而在实际相机系统中, 由于制造工艺误差、成像器件的微小倾斜或镜头安装偏差, 可能导致图像平面上的  $x$  与  $y$  轴存在轻微的不正交性, 从而出现非零的 skew 值。在本课设中, 左相机内参矩阵中的 skew 值为 3.061, 右相机内参矩阵中的 skew 值为 4.9914。考虑到该相机的主点坐标值与焦距参数的数量级都很大, skew 值在该数值量级下属于较小的偏差。其影响远小于焦距和主点位置的

## 第二章 实验

绝对值，因此可以粗略认为像素坐标轴近似正交，该偏斜值在一定精度要求下可以忽略不计。

左相机畸变系数，

$$distCoeffsL = [-0.3771, -0.1041, 0.0001, -0.0070, 0.0000]$$

右相机畸变系数，

$$distCoeffsR = [-0.4105, 0.2374, 0.0011, -0.0070, 0.0000]$$

相机外参描述右相机相对左相机的空间位置与姿态。旋转矩阵可表示为，

$$\mathbf{R} = \begin{bmatrix} 0.9999 & -0.0011 & 0.0104 \\ 0.0011 & 1.0000 & 0.0001 \\ -0.0104 & -0.0001 & 0.9999 \end{bmatrix}$$

该矩阵表示右相机在三维空间中绕三个坐标轴的微小旋转，表明两个相机之间的视角几乎一致，仅存在非常小的偏差。

平移向量可表示为，

$$\mathbf{T} = \begin{bmatrix} -66.0171 \\ 2.1145 \\ 1.8565 \end{bmatrix}$$

表示右相机相对于左相机在 X、Y、Z 方向上的相对位移。可以看出两个相机的基线距离约为 66mm。

### 2.3 创建立体图像并生成特征对

为了实现准确的立体匹配，需要首先进行图像的极线校正（epipolar rectification）。这是因为在实际拍摄过程中，由于摄像头间的角度和位置误差，同一物体在左右图像中的位置往往并不完全在同一水平线上，给匹配带来困难。

校正的基本原理包括：

- 计算旋转矩阵和平移向量。通过预先的相机标定步骤，获得左右摄像头之间的相对姿态信息，即旋转矩阵  $\mathbf{R}$  与平移向量  $\mathbf{T}$ 。

- 构建新的投影矩阵。通过几何计算，构造两个新的投影矩阵，使得经校正后的左右图像具有统一的投影模型。校正的目标是将对应点变换至同一行上，简化匹配为一维搜索。
- 图像重映射。通过插值方式对图像每个像素进行重定位，使其对应点在左右图像中处于同一扫描线上，形成所谓的“极线对齐”图像。

经过校正后，左右图像的几何畸变被消除，对应点严格位于同一行中，便于后续的立体匹配。

之后，为提取图像中的显著点对，系统采用尺度不变特征变换（SIFT）方法进行特征点提取与描述。该方法在图像的多尺度空间中检测稳定的关键点，并对其周围区域进行梯度方向分析，从而生成高维描述子，用于表达关键点的局部纹理结构。

接着，对左右图像中的特征进行两两匹配。为了提高匹配的准确性，系统采用双向最近邻比率测试（ratio test）的方式，仅保留两个最近匹配中距离最小者远小于次近距离的匹配对，从而排除大量错误匹配。

最后，为进一步去除错误匹配，还引入了极线几何约束，即要求两个匹配点在竖直方向上的像素坐标差值小于阈值。这样可以过滤掉不满足几何一致性的异常点对，得到更为可靠的匹配结果。这些高质量的特征匹配对为后续深度估计提供了精确的对应关系基础。

## 2.4 生成视差图

视差图的生成基于双目图像中匹配点之间的水平位置差。由于经过极线校正，所有匹配点应位于同一行上，因此可以简化为一维搜索问题，即仅在水平方向上查找对应点。

为了简化匹配运算，首先将左右图像转换为单通道的灰度图像。灰度图在亮度变化上更稳定，且计算效率更高。

随后，使用块匹配 Block Matching (BM) 实现视差估计。通过滑动窗口在一张图像中选取参考块，在对应行的另一图像中移动搜索窗口，通过计算两个窗口之间的相似性度量（如绝对差、互相关等），确定最佳匹配位置，并据此计算视差值。

## 2.5 计算深度图

生成的 disparity 视差图像中，每个像素的值表示其在左右图像之间的水平位移。视差值越大，表示该像素对应的物体越靠近相机；视差越小，则表示距离较远。

对某一空间点  $P$ ，它在左、右相机图像中的投影点为  $p_L$  和  $p_R$ ，其横坐标分别为  $x_L$  和  $x_R$ ，则该点的视差  $d$  定义为： $d = x_L - x_R$ 。

由针孔相机模型可得深度  $Z$ （即从相机出发到该点的距离）计算公式如下：

$$Z = \frac{f \cdot B}{d}$$

其中， $f$  为相机的焦距， $B$  为左右相机的基线距离， $d$  为该像素的视差值， $Z$  为输出深度值。

## 2.6 实物测试

我们在多场景下进行了多个实物测试，包括举着水杯的人、静坐的人与桌上的书包，如图2.10、2.11与2.12所示。

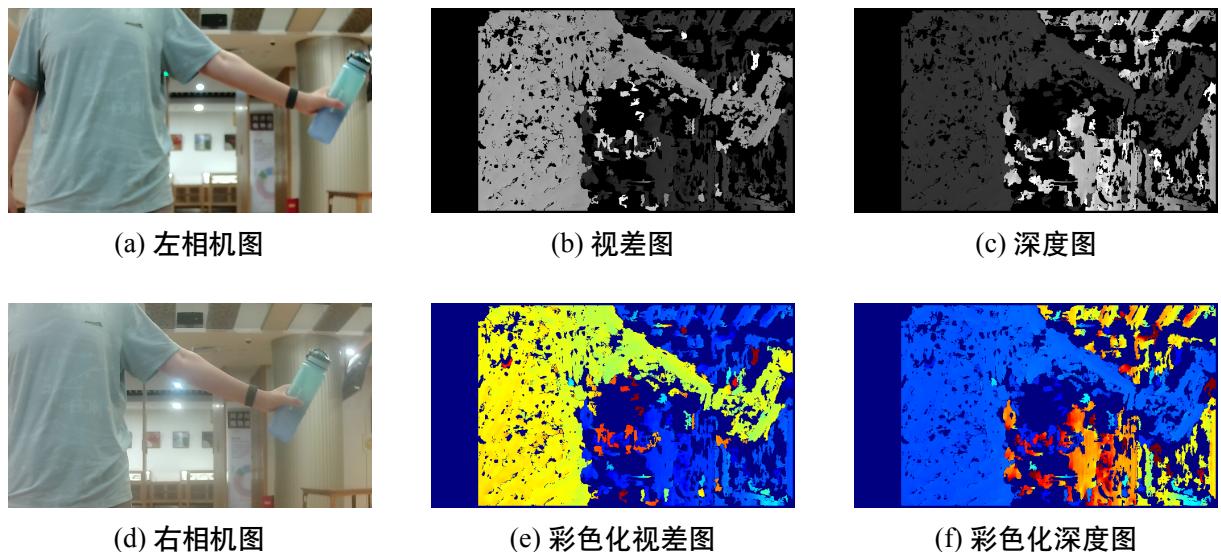


图 2.10: 实物测试 1

在已知相机内参与像素级深度信息的前提下，可实现从图像中精确估计物体的实际物理长度。该过程依赖于摄像机成像模型与深度反投影原理，将图像中的像素点坐标转换为相机坐标系下的三维坐标，并计算空间距离。

随后进行成像模型与深度反投影，相机的针孔模型将三维世界坐标点  $(X, Y, Z)$  投影到图像平面像素点  $(u, v)$ ，满足以下关系：

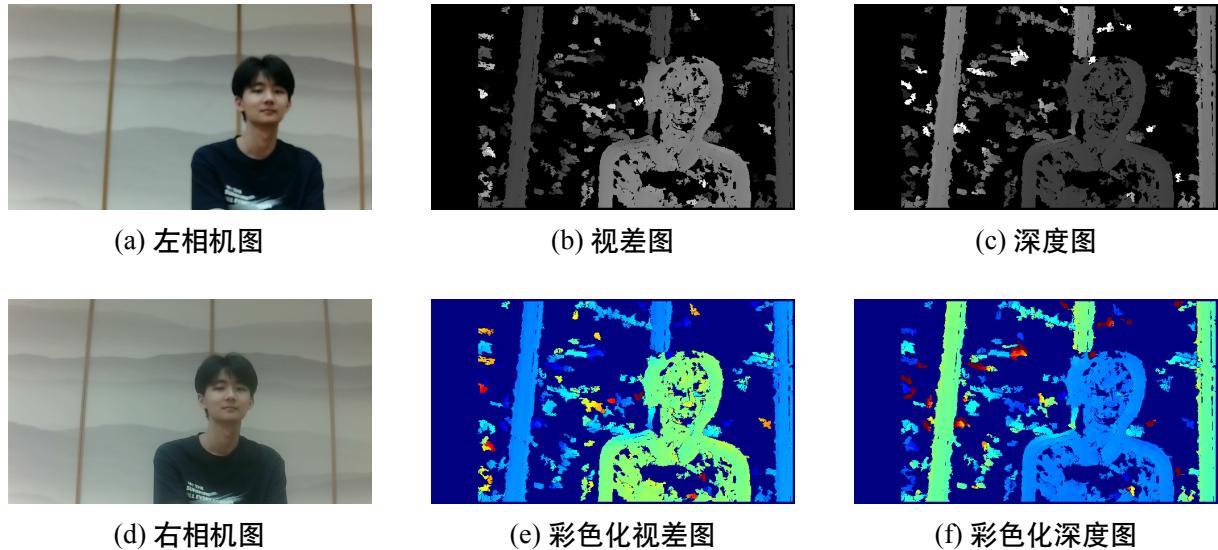


图 2.11: 实物测试 2

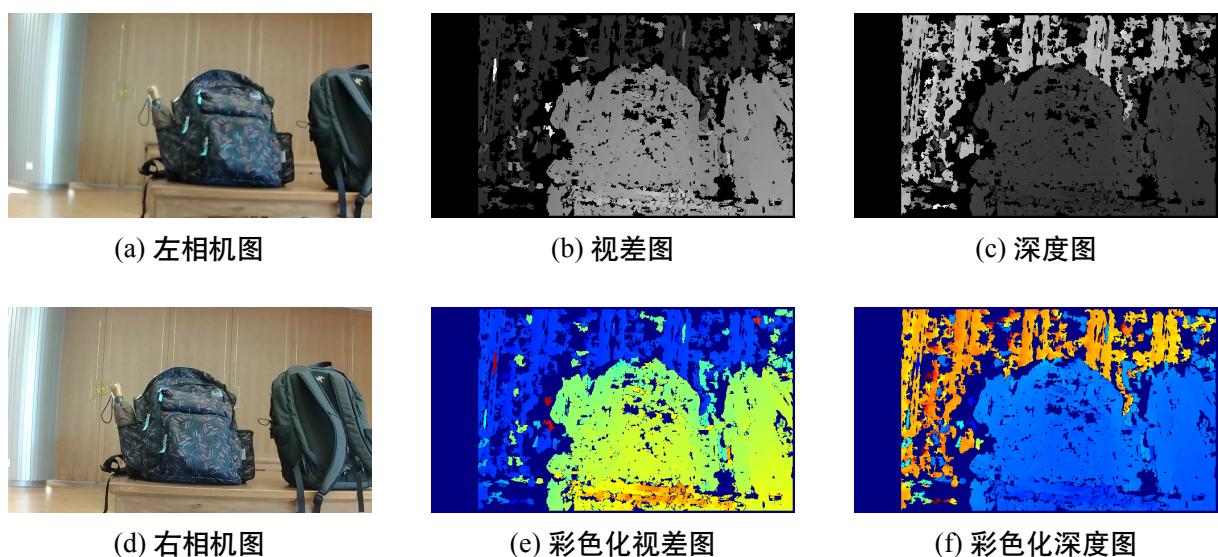


图 2.12: 实物测试 3

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \Rightarrow \begin{cases} X = \frac{(u - c_x) \cdot Z}{f_x} \\ Y = \frac{(v - c_y) \cdot Z}{f_y} \\ Z = \text{Depth}(u, v) \end{cases}$$

其中,  $f_x$ 、 $f_y$  是相机的水平与垂直焦距 (单位: 像素),  $c_x$ 、 $c_y$  为主点坐标,  $Z$  为当前像素的深度值 (单位: 米)。

通过上述反投影公式, 可将图像上任意像素点  $(u, v)$  转换为相机坐标系下的三维点  $(X, Y, Z)$ 。

评价过程以长度预测的准确性评价, 给定两个物体关键点的像素坐标  $(u_1, v_1)$  与  $(u_2, v_2)$ , 以及其对应的深度值  $Z_1$  与  $Z_2$ , 通过以下步骤估计两点间实际长度:

- 根据相机内参  $(f_x, f_y, c_x, c_y)$  和深度图, 将两个像素点反投影为三维坐标:

$$\mathbf{P}_1 = \left( \frac{(u_1 - c_x)Z_1}{f_x}, \frac{(v_1 - c_y)Z_1}{f_y}, Z_1 \right)$$

$$\mathbf{P}_2 = \left( \frac{(u_2 - c_x)Z_2}{f_x}, \frac{(v_2 - c_y)Z_2}{f_y}, Z_2 \right)$$

- 计算三维空间中两点之间的欧氏距离:

$$d = \|\mathbf{P}_1 - \mathbf{P}_2\| = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}$$

- 根据书包的视差和内参信息, 我们按照上述计算方法得到估计长度 0.322, 单位为米 (实际长度为 0.3m), 误差小于 10%。

## 2.7 结果讨论

从实物测试图中可以看出, 以上基于块匹配 BM 算法的立体匹配方法, 已经能够较为有效地提取视差图并计算出对应的深度图, 为后续的三维重建和目标定位提供了基础。然而, 该方法仍存在一定的局限性。例如, 生成的视差图中噪点较多, 部分区域存在匹配误差, 且在实际物体的边缘处, 视差变化不够平滑, 出现了锯齿状或模糊边缘的情况。这些问题在一定程度上影响了深度图的精度和稳定性。

造成上述问题的主要原因在于 BM 方法在代价聚合阶段仅使用固定窗口进行像素强度匹配, 难以适应复杂背景或纹理较弱区域。此外, BM 方法未能充分考虑像素间的

## 第二章 实验

全局一致性，也缺乏边缘保持能力。

## 第三章 拓展任务

### 3.1 提升深度图质量

#### 3.1.1 基于 SIFT 的双目图像矫正与匹配流程

本系统在完成立体相机标定与图像校正的基础上，引入 SIFT 特征提取与匹配技术对图像对进行几何一致性验证，并辅助视差图计算与深度生成。其主要步骤如下：

- 对立体校正后的左右图像分别进行 SIFT 特征点检测和描述符提取。
- 使用基于欧氏距离的暴力匹配器（BFMatcher）进行特征匹配，并采用  $k = 2$  的近邻匹配方式。
- 引入比率测试（Ratio Test）以剔除歧义匹配，具体条件为：第一匹配距离小于第二匹配距离的 0.75 倍。
- 基于极线约束进一步过滤匹配点对，即剔除  $y$  坐标差值大于阈值（如 10 像素）的匹配点。
- 将筛选后的匹配结果用于计算视差图和重建深度图。

#### 3.1.2 关键代码与功能对照

表 3.1: SIFT 匹配关键函数与作用

函数或参数	功能说明
SIFT::detectAndCompute()	提取图像的关键点和 128 维描述符
BFMatcher::knnMatch(..., 2)	执行 k 近邻特征点匹配
ratio_thresh = 0.75	Lowe 比率测试，过滤不良匹配
y_threshold = 10	极线约束过滤阈值
reprojectImageTo3D()	根据视差图生成三维深度图

#### 3.1.3 立体校正验证原理

立体校正的目标是将原始图像对几何变换为极线对齐的形式，使得匹配点在左右图像中的垂直坐标一致，即：

$$y_l = y_r$$

匹配点对之间的  $y$  坐标差值可用于评估校正精度：

$$\text{平均极线误差} = \frac{1}{N} \sum_{i=1}^N |y_{l,i} - y_{r,i}|$$

误差越小，表示图像对校正效果越好。

### 3.1.4 SIFT 特征提取与匹配原理

SIFT 特征点具有旋转不变性、尺度不变性等优点，包含以下关键步骤：

- 尺度空间构建：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

- DoG 极值检测：

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

- 关键点定位（亚像素）：

$$\begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = - \begin{bmatrix} D_x \\ D_y \end{bmatrix}$$

- 梯度方向计算：

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

- 特征描述子生成：将邻域划分为  $4 \times 4$  个子块，对每块计算 8 个方向直方图，共形成  $4 \times 4 \times 8 = 128$  维特征向量。

### 3.1.5 匹配质量评估指标

## 3.2 算法优化

### 3.2.1 Semi-Global Block Matching (SGBM) 算法原理

在立体视觉中，视差计算是三维重建与深度估计的核心步骤。Block Matching (BM) 与 Semi-Global Block Matching (SGBM) 是两种广泛使用的视差估计算法，分别代表了传统局部方法与优化型方法的典型思路。

表 3.2: SIFT 匹配常用质量指标

指标名称	计算方式
匹配正确率	$\frac{\text{正确匹配数}}{\text{总匹配数}}$
平均极线误差	$\frac{1}{N} \sum  y_l - y_r $
视差一致性	$\frac{\text{std}(disparity)}{\text{mean}(disparity)}$

**Block Matching (BM)** 算法是一种典型的局部匹配方法，其基本思想是在一幅图像中以固定窗口为单位，在另一幅图像的对应扫描线上滑动窗口，通过计算窗口内像素灰度差的某种代价函数（如绝对差值和 SAD 或相关系数）来寻找最佳匹配位置，从而获得视差值。该方法计算简单、易于实现，适合资源受限或实时性要求高的场景。然而，BM 算法存在明显的局限性：对光照变化、弱纹理区域、边缘与遮挡处理能力较差，容易出现块状伪影和匹配错误，整体匹配精度有限。

为弥补 BM 的不足，**Semi-Global Block Matching (SGBM)**<sup>[6]</sup> 被提出作为一种权衡精度与效率的优化方法。SGBM 的核心思想是在局部匹配的基础上，引入全图范围的路径代价聚合机制。它不再仅依赖单一窗口内的局部信息，而是结合图像中多个方向（通常为 8 或 16 个）上的像素路径，沿每个方向对匹配代价进行累积，通过惩罚视差不连续的跳变，从而在不显著增加计算负担的前提下逼近全局最优解。

SGBM 使用的能量函数可表示为：

$$E(D) = \sum_{\mathbf{p}} \left( C(\mathbf{p}, D_{\mathbf{p}}) + \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} P_1 \cdot [|D_{\mathbf{p}} - D_{\mathbf{q}}| = 1] + P_2 \cdot [|D_{\mathbf{p}} - D_{\mathbf{q}}| > 1] \right)$$

其中， $\mathbf{p}$  表示像素位置， $D_{\mathbf{p}}$  表示像素的视差值， $C(\mathbf{p}, d)$  为初始匹配代价， $P_1$  与  $P_2$  为惩罚项，控制相邻像素视差变化的平滑性强度。若两个相邻像素的视差差值为 1，则施加较小惩罚  $P_1$ ，若差值大于 1，则施加较大惩罚  $P_2$ 。这种方式允许对物体边缘保持一定的视差变化，而对平坦区域则鼓励视差一致性，从而在保留边缘的同时抑制噪声。

为了避免计算复杂度飙升，SGBM 采用沿若干固定路径方向的动态规划进行代价聚合，并最终取所有方向的最小代价和作为最终匹配代价。这种“半全局”的策略相比于传统的全局优化方法（如 Graph Cuts 或 Belief Propagation）大幅降低了计算资源需求，同时在准确性上远超 BM。

### 3.2.2 SGBM 与 BM 的对比分析

与 BM 相比，SGBM 在多个方面具有显著优势：

- **匹配准确性更高：**通过路径代价聚合机制，SGBM 能显著降低弱纹理区域和遮挡区域的误匹配概率，增强边缘保持能力。
- **平滑性与结构性更强：**借助惩罚项  $P_1$  和  $P_2$  的调节，SGBM 在光滑区域生成更加一致的视差值，同时在物体边界处保持清晰结构。
- **更强的鲁棒性：**对光照变化、纹理模糊等干扰具有更高的适应能力。
- **适中的计算开销：**相比于全局优化方法，SGBM 不需大规模迭代优化，在可接受的计算资源下实现了更优的匹配结果。

然而，SGBM 也并非毫无缺点。其计算复杂度和内存消耗高于 BM，在嵌入式平台或硬实时系统中仍需权衡使用。同时，路径方向的选择可能引入一定的方向偏差，对某些结构倾斜的区域匹配存在一定误差。我们可以认为，SGBM 是一种在精度与效率之间取得良好平衡的立体匹配算法，特别适用于对匹配质量要求较高但又无法承受全局优化成本的中高端应用场景。

### 3.2.3 算法优化后的实物测试

在算法优化后，再次进行实物测试，如图3.1、3.2、3.3所示。与优化前的结果相比，可以明显看出视差图和深度图的整体质量有了显著提升。首先，图像中的噪点数量明显减少，尤其是在低纹理区域和物体边缘处，原本较为严重的误匹配问题得到了有效抑制，使得深度估计更加稳定可靠。其次，物体轮廓的边缘过渡更加平滑且清晰，较好地保留了目标的几何结构与边界信息，说明优化后的算法在边缘保持方面具有更强的能力。

## 3.3 建立 3D 点云

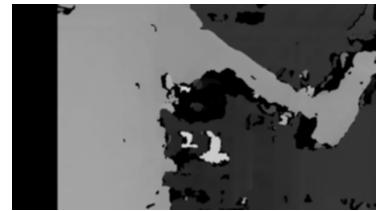
本系统通过双目立体视觉技术，从左右相机获取同步图像对，并基于视差计算原理生成稠密点云，实现对场景的三维重建。其主要流程与原理如下：

### 3.3.1 图像校正与立体匹配

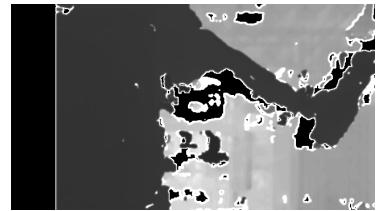
首先，利用事先标定获得的相机内参和外参，对图像对进行立体校正（Stereo Rectification），将左右图像的扫描线对准，从而使得对应点位于同一行。



(a) 左相机图



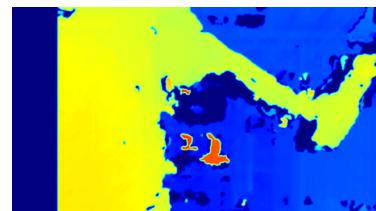
(b) 视差图



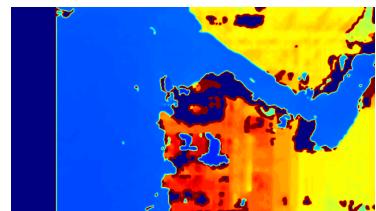
(c) 深度图



(d) 右相机图



(e) 彩色化视差图

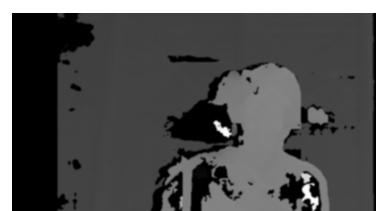


(f) 彩色化深度图

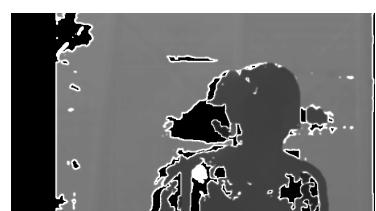
图 3.1: 实物测试 1



(a) 左相机图



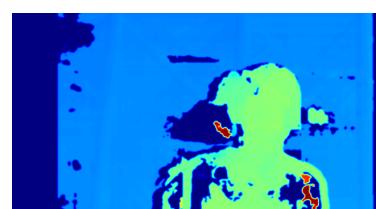
(b) 视差图



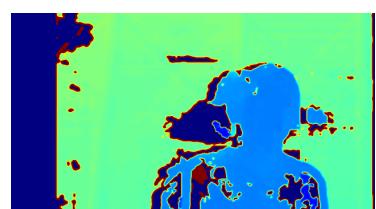
(c) 深度图



(d) 右相机图



(e) 彩色化视差图



(f) 彩色化深度图

图 3.2: 实物测试 2

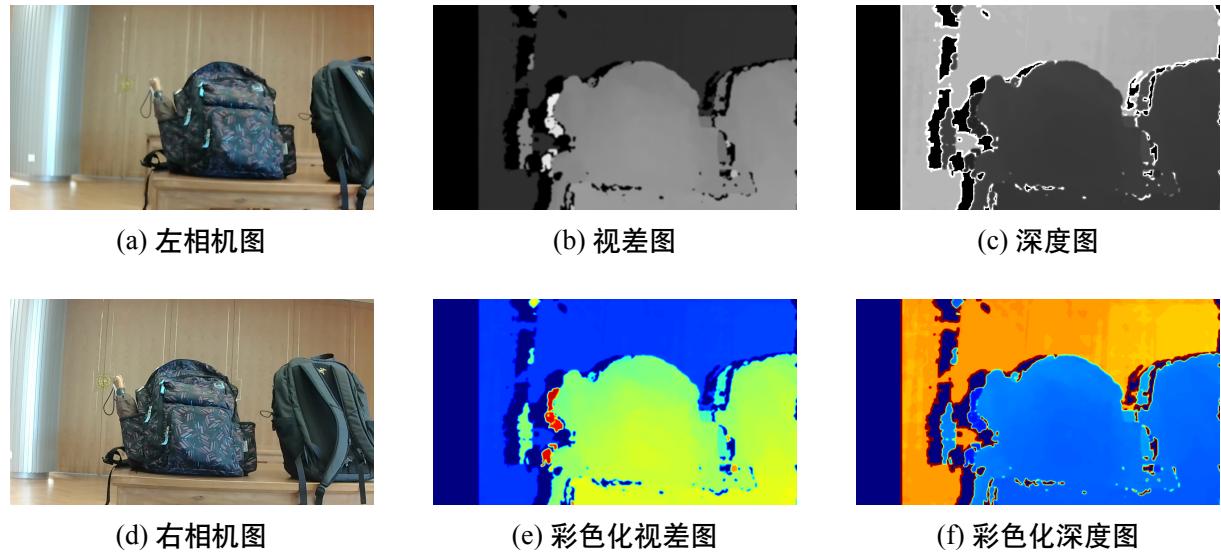


图 3.3: 实物测试 3

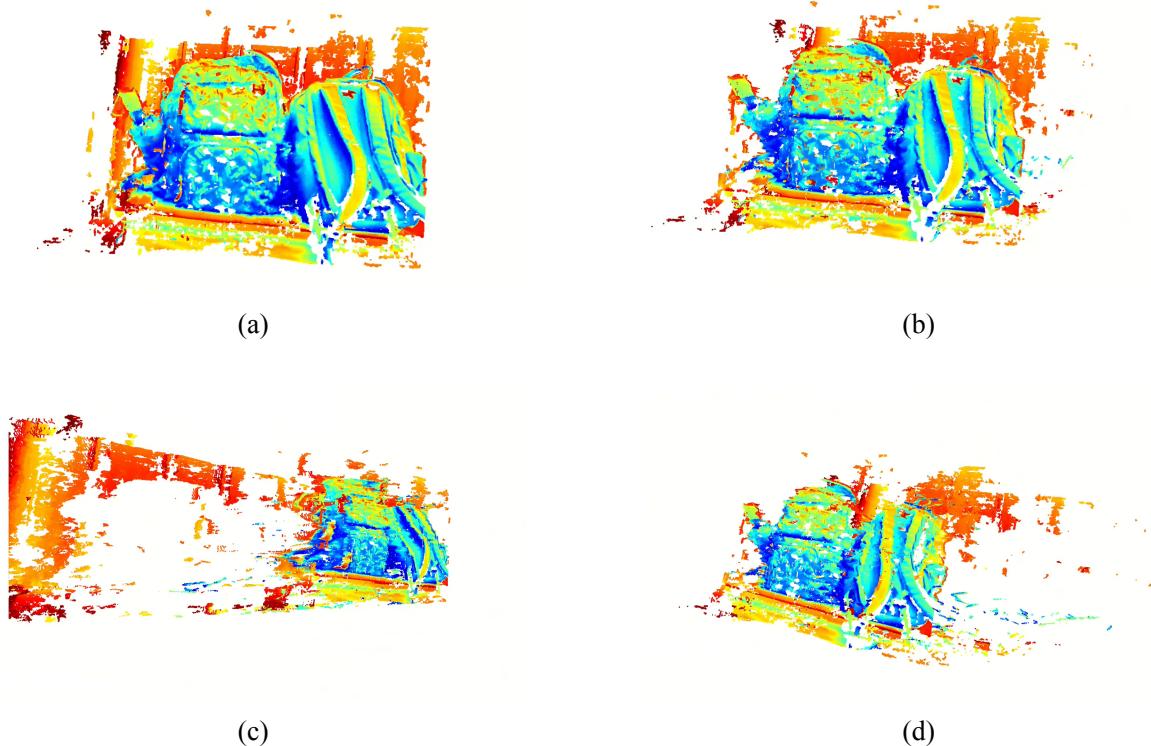


图 3.4: 点云结果示例

随后，将校正后的彩色图像转为灰度图，并使用 OpenCV 的 StereoSGBM 算法计算视差图。SGBM (Semi-Global Block Matching) 算法在局部块匹配的基础上，通过代价聚合方式优化匹配路径，取得更平滑和一致的视差结果。视差图中每个像素的值表示该点在左右图像中的水平偏移量。

### 3.3.2 深度计算与点云生成

已知相机的基线  $b$ 、焦距  $f$  以及主点坐标  $(c_x, c_y)$ ，可以利用三角测量原理将像素点的视差  $d$  转换为深度  $z$ 。

图像坐标系下的像素位置  $(u, v)$  可进一步转换为三维空间坐标  $(x, y, z)$ ：

$$x = \frac{(u - c_x) \cdot z}{f}, \quad y = \frac{(v - c_y) \cdot z}{f}$$

最终，我们将每个有效像素位置  $(u, v)$  对应的三维点  $(x, y, z)$  组成四维向量：

$$\mathbf{p} = \begin{bmatrix} x & y & z & i \end{bmatrix}^T$$

其中  $i$  为灰度图像的归一化强度值（即  $i = \text{gray}(u, v)/255.0$ ），可用于后续点云渲染中的颜色编码。

### 3.3.3 点云可视化

构造完成的点云数据以 Eigen::Vector4d 格式存储，调用封装好的 `showPointCloud` 函数使用 Pangolin 库进行三维可视化，用户可以通过鼠标旋转与缩放查看生成的三维结构。

### 3.3.4 系统框架总结

整体点云生成流程如下：

- 加载双目相机标定参数；
- 实时采集双目图像并进行图像校正；
- 使用 SGBM 算法计算视差图；
- 根据视差图计算像素对应的三维坐标；
- 构建稠密点云并进行可视化。

### 第三章 拓展任务

该方法实现了从原始双目图像到稠密点云的自动化处理，为三维场景理解和重建提供了基础。

## 第四章 总结

本实验围绕双目视觉系统的构建与深度信息获取展开，依次完成了双目相机标定、图像校正、立体匹配、视差图生成、深度图计算、实物测试以及点云重建等关键任务，全面验证了基于几何原理的三维重建流程。

首先，借助 MATLAB 工具箱完成相机标定，获取了双目相机的内外参数与畸变参数，并基于极几何原理进行了图像的立体校正，使得左右图像在极线方向上严格对齐，为后续视差计算打下了基础。随后，采用基于块匹配的视差估计方法 BM 得到了稠密的视差图，利用相机参数与三角测量模型，成功将像素级的视差信息转化为物理世界中的实际深度值。在实物测试部分，通过对深度图的分析和内参矩阵的反投影，实现了对目标物体长度的估计，进一步验证了深度图的几何精度与实际应用价值。

扩展任务中，我们引入 SIFT 特征用于图像匹配，探讨了特征匹配精度对视差质量的影响，并分析了 SGBM 等更高级算法的改进效果，展示了算法在细节恢复与边缘保留方面的优势。

最后，我们基于深度图重建了稠密点云，实现了对空间场景的三维建模，为后续的三维识别、目标检测、路径规划等高层任务提供了数据基础。

总体来看，本次课设我们系统地完成了从二维图像采集到三维信息恢复的完整流程，较为深入地掌握了双目视觉的核心原理与工程实现方法，并具备将理论转化为实际应用的能力。

## 第五章 成员分工

表 5.1: 项目成员分工表

学号	姓名	主要工作内容
2152482	鄂俊伍	负责立体匹配算法的实现与优化，包括 Block Matching (BM) 与 Semi-Global Block Matching (SGBM) 的对比分析、SGBM 代价聚合机制研究、视差图生成与深度图实时计算，以及反投影生成三维点云。负责报告撰写。
2152222	邹阳	负责基于 SIFT 特征的图像矫正与匹配方法的研究，提高双目图像对齐精度。同时完成从图像采集、校正、匹配、重建到点云生成的整体系统集成与调试，以及实验验证和测量误差分析工作。
2151094	宋正非	负责双目相机标定与立体匹配算法的实现，包括图像采集模块构建、HBVCAM-W202011HD V33 双目相机参数配置、内参矩阵、畸变系数和外参矩阵的获取，实验验证和测量误差分析工作，以及 Block Matching (BM) 等算法的实现。负责报告撰写。

## 参考文献

- [1] K Konolige. Small vision systems: Hardware and implementation[C]//Robotics Research: The Eighth International Symposium. Springer. [S.l.]: [s.n.], 1998: 203–212.
- [2] R Fan, L Wang, M J Bocus, et al. Computer stereo vision for autonomous driving[J]. ArXiv preprint arXiv:2012.03194, 2020.
- [3] R Fan, Y Liu, X Yang, et al. Real-time stereo vision for road surface 3-d reconstruction[C]//2018 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE. [S.l.]: [s.n.], 2018: 1–6.
- [4] C.-W Liu, H Wang, S Guo, et al. Stereo matching: fundamentals, state-of-the-art, and existing challenges[G]//Autonomous Driving Perception: Fundamentals and Applications. [S.l.]: Springer, 2023: 63–100.
- [5] R A Hamzah, R Abd Rahim, Z M Noh. Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application[C]//2010 3rd International Conference on Computer Science and Information Technology. Vol. 1. IEEE. [S.l.]: [s.n.], 2010: 652–657.
- [6] R Spangenberg, T Langner, R Rojas. Weighted semi-global matching and center-symmetric census transform for robust driver assistance[C]//International Conference on Computer Analysis of Images and Patterns. Springer. [S.l.]: [s.n.], 2013: 34–41.