

中图分类号： TP391.41

学校代码： 10081

U D C:

密 级： 公开



# 硕 士 学 位 论 文

## 基于双目视觉的物体尺寸测量研究

论文作者： 王然

学生类别： 全日制专业型

学科专业： 计算机技术

学位类别： 电子信息硕士

指导教师： 于复兴 副教授

企业教师： 雷益龙 高级工程师

唐山 华北理工大学

2023 年 10 月

# **Research on Object Size Measurement Based on Binocular Vision**

Dissertation Submitted to  
**North China University of Science and Technology**  
in partial fulfillment of the requirement  
for the degree of  
**Master of Electronic Information**

By  
**Wang Ran**  
(Computer Technology)

Supervisor: **Professor Yu Fuxing**  
**Senior Engineer Lei Yilong**

**October, 2023**

## 摘 要

随着计算机视觉方面算法的不断提高以及相关硬件设备的发展，基于双目视觉的尺寸测量技术广泛应用于工业生产中。本课题基于双目视觉的相关原理来实现对物体尺寸测量的研究，主要的研究内容包含以下几个方面：

1、针对立体匹配算法中传统 Census 变换的局限性，提出了一种边缘自适应 Census 变换与颜色、梯度信息结合的融合匹配代价。通过使用 Sobel 获取图像的边缘信息，并根据初始窗口内边缘像素总和与灰度均值自适应 Census 窗口大小，最后融合三通道的 AD 算法和基于梯度的代价计算，获取可靠性更高的匹配代价，提高立体匹配精度。

2、在立体匹配过程的代价聚合阶段采用具有边缘的最小生成树聚合算法，并针对最小生成树聚合在纹理模糊、重复结构或遮挡的图像区域的聚合效果不好的问题，提出了扫描线优化算法，从四个方向分别优化获取较为准确的聚合代价，最后根据多尺度思想，融合多尺度聚合代价，提高立体匹配的准确度。

3、在特征点获取过程中使用图像分割算法对目标物体进行初步分割，为了获取更精准的图像边缘信息，提出边缘搜索优化方法，获取较为准确的目标物体轮廓信息，再使用 Harris 角点检测算法提取目标物体特征点。最后根据相机参数信息、视差图和特征点坐标，计算出特征点的三维坐标，根据欧氏距离方程得到物体尺寸信息。

图 28 幅；表 17 个；参 44 篇。

**关键词：**双目视觉测量；相机成像模型；相机标定；立体校正；立体匹配

**分类号：**TP391.41

## Abstract

With the continuous improvement of computer vision algorithms and the development of related hardware devices, size measurement technology based on binocular vision is widely used in industrial production. This project realizes the research on object size measurement based on the relevant principles of binocular vision, and the main research content includes the following aspects:

1. Aiming at the limitations of the traditional Census transformation in the stereo matching algorithm, a fusion matching cost combining the edge adaptive Census transformation with color and gradient information is proposed. By using Sobel to obtain the edge information of the image, adaptive Census window size according to the sum of pixels at the edge of the initial window and the mean grayscale, and finally fusing the three-channel AD algorithm and gradient-based cost calculation to obtain a more reliable matching cost and improve the stereo matching accuracy.

2. In the cost aggregation stage of the stereo matching process, the minimum spanning tree aggregation algorithm with edges is adopted, and aiming at the problem that the minimum spanning tree aggregation has a poor aggregation effect in the image area with blurred texture, repeated structure or occlusion, a scan line optimization algorithm is proposed to optimize the more accurate aggregation cost from four directions, and finally according to the multi-scale idea, the multi-scale aggregation cost is integrated to improve the accuracy of stereo matching.

3. In the process of feature point acquisition, the image segmentation algorithm is used to initially segment the target object, in order to obtain more accurate image edge information, an edge search optimization method is proposed to obtain more accurate target object contour information, and then Harris corner point detection algorithm is used to extract the target object feature point. Finally, according to the camera parameter information, disparity map and feature point coordinates, the three-dimensional coordinates of the feature points are calculated, and the object size information is obtained according to the Euclidean distance equation.

Figure 28; Table 17; Reference 44

**Keywords:** Binocular visual measurement, Camera imaging model, Camera calibration, Stereo correction, stereo matching

**Chinese books catalog:** TP391.41

# 目 录

第 1 章 绪论.....	1
1.1 课题的研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 机器视觉的研究现状.....	2
1.2.2 双目视觉测量的研究现状.....	2
1.3 本文主要研究内容.....	4
第 2 章 相机标定及立体校正 .....	6
2.1 双目立体视觉模型.....	6
2.1.1 相机成像模型.....	6
2.1.2 相机标定基础.....	9
2.2 相机标定及立体校正原理 .....	10
2.2.1 张正友标定法.....	10
2.2.2 最大似然估计.....	12
2.2.3 双目立体标定.....	14
2.2.4 立体校正原理.....	14
2.3 相机标定和立体校正的实现 .....	15
2.3.1 标靶图像采集.....	15
2.3.2 基于 OpenCV 的相机标定及立体校正 .....	16
2.4 本章小结.....	18
第 3 章 半全局立体匹配算法研究及改进 .....	19
3.1 立体匹配算法的基本流程 .....	19
3.2 基于 Census 的半全局立体匹配算法.....	19
3.2.1 匹配代价计算.....	19
3.2.2 匹配代价聚合.....	21
3.2.3 视差计算与优化.....	23
3.3 半全局立体匹配算法的改进 .....	24
3.3.1 算法流程.....	24
3.3.2 匹配代价计算.....	25
3.3.3 匹配代价聚合.....	28
3.4 实验结果与分析.....	31
3.4.1 立体匹配评价标准.....	31
3.4.2 改进代价计算验证.....	32
3.4.3 改进代价聚合验证.....	34
3.4.4 改进算法对比分析.....	36

3.5 本章小结.....	37
第4章 双目视觉物体尺寸测量实现 .....	38
4.1 测量方案设计.....	38
4.2 基于双目视觉的物体尺寸测量研究 .....	39
4.2.1 图像采集与校正.....	39
4.2.2 目标物体轮廓提取.....	39
4.2.3 目标物体特征点提取.....	42
4.2.4 目标物体尺寸测量.....	44
4.3 物体尺寸测量的实验与分析 .....	48
4.3.1 物体距离测量实验.....	48
4.3.2 算法可靠性验证.....	49
4.3.3 算法有效性验证.....	51
4.3.4 物体尺寸测量误差分析.....	52
4.4 本章小结.....	53
结论.....	54
参考文献.....	55

## 第1章 绪论

### 1.1 课题的研究背景及意义

在生活、生产和科研等领域人们对物体尺寸测量都有着巨大的需求，但传统的测量基本都是借助一些测量工具，通过人工的接触式测量方法。其中需要人工操作的游标卡尺、千分尺、量规等工具的测量精度低、效率低，不适用于大规模工业生产中的批量测量，且容易受到疲劳、走神、不仔细等人为因素影响。三坐标测量机具有高精度和高效率的特点，但由于需要接触被测对象，可能会对其造成损伤。超声波测量法的应用受到环境条件的限制，尤其是空间湿度对其灵敏度的影响较大<sup>[1]</sup>。因此，采用立体测量技术通过使用摄像机采集目标物体的图像，对其特征进行提取和分析计算，从而实现对物体的检测和测量受到了广泛的关注<sup>[2]</sup>。

双目视觉是人类视觉系统中的一项重要功能。通过两只眼睛同时观察物体，可以使我们感知并理解物体的三维形状、尺寸和深度信息。在计算机视觉领域，使用双目视觉进行物体尺寸测量已经成为一种常见而有效的方法。双目视觉是利用两个摄像头采集同一场景下的物体信息，再从采集到图像中根据三角测量原理获取三维场景中相对于相机成像平面的纵深信息。

双目物体尺寸测量在许多领域有着广泛的应用和重要的意义。在工业制造过程中，需要准确测量产品的尺寸以保证对质量的良好控制，双目视觉提供高效的非接触式的尺寸测量方法<sup>[3]</sup>，大大提高了生产效率和成产质量。在建造设计和施工过程中，需要对空间进行准确的尺寸测量和规划，双目视觉可以帮助工程师快速测量建筑物的高度、长度等参数，并进行精确的空间分析。在医学诊断中，精确的尺寸测量对于病变评估和治疗规划至关重要，通过双目视觉技术可以辅助医生对肿瘤、器官等病变进行精确的尺寸测量。在自动驾驶汽车中，准确测量道路上的障碍物尺寸是实现安全行驶的基础，双目视觉可以帮助车辆感知并测量周围环境中的物体尺寸，从而提供可靠的决策依据。

综上所述，传统的测量方法具有一定的局限性，如测量精度低、效率低、容易受到人为因素影响等特点。而基于双目视觉的物体尺寸测量具有高效、精确、非接触等特点，在许多领域具有广泛的应用和重要的意义，因此对双目物体尺寸测量的研究和应用具有重要的价值和意义。

## 1.2 国内外研究现状

### 1.2.1 机器视觉的研究现状

在 20 世纪 50 年代到 70 年代, 机器视觉技术开始出现, 当时的研究主要集中在低级的图像处理任务, 如边缘检测、光学字符识别、工件表面分析等, 这个阶段的主要目标是开发基本的图像处理算法, 来提取图像中的特征。在 20 世纪 70 年代到 90 年代中期, 机器视觉开始从图像中提取更加高级的特征, 并进行模式识别和分类。这个阶段主要进行特征提取算法的研究, 同时机器学习技术也开始在机器视觉领域应用, 使得机器视觉能够学习和识别更复杂的模式。在这个阶段麻省理工学院提出了著名的 Marr 视觉理论<sup>[5]</sup>, 该理论成为机器视觉领域的重要理论框架。国内外学者根据此框架为基础对双目视觉立体测量方法进行了深入的研究, 并取得不少有价值的成果。

在智能机器人领域, 学者 Jung 等人<sup>[6]</sup>通过将双目视觉测量技术与机器学习相结合, 从图像中检测出运动物体并计算出周围障碍物的距离。这使得机器人能够在复杂的环境中, 对运动物体进行有效的跟踪和交互。学者 Okada 等人<sup>[7]</sup>通过整合视觉系统的楼道区域信息和全身姿态信息, 围绕机器人构建了局部楼层地图, 令人形机器人能够在行进时躲避障碍物。张亚斌<sup>[8]</sup>通过机器视觉的目标检测算法和双目测距, 根据双目相机拍摄的图像, 对路上行人进行检测和测距, 实现了移动机器人对路上行人距离的精准测量。广西师范大学的郭海州<sup>[9]</sup>对室内服务机器人进行研究, 通过双目视觉系统, 完成了对室内物体的检测和测距, 使得室内服务机器人可以根据这些信息进行规划和反应。

在自动驾驶领域, 学者 Chen<sup>[10]</sup>通过 3D 测量方法, 实时检测驾驶员的头部姿势, 用以提醒安全驾驶。Xie 等人<sup>[11]</sup>提出了一种基于边缘的快速估计算法, 通过实时计算障碍物距离并规划路径, 可以用于高级辅助驾驶。魏新华等人<sup>[12]</sup>通过双目视觉技术对田间作物的高度和收割边界进行提取, 实现了收获无人机的自适应调控。

浙江大学的蒋焕煜等人<sup>[13]</sup>通过双目立体视觉技术利用成熟番茄与背景之间的颜色特征差异进行图像分割, 来识别成熟番茄并定位番茄中心位置。贵州民族大学的陈红<sup>[14]</sup>使用双目视觉立体测量技术, 实现了机械零件的高精度测量。

### 1.2.2 双目视觉测量的研究现状

视觉测量技术是利用计算机视觉和图像处理技术, 通过对图像进行分析和处理, 完成对物体的尺寸、形状、位置等进行测量和计算的技术。视觉测量技术可以根据



在图像采集时所用到的相机的数量分为：单目视觉测量技术、双目视觉测量技术和多目视觉测量技术<sup>[15]</sup>。单目视觉是指利用单台相机获取场景图像的视觉测量技术。由于单目视觉只有一个视角，图像中丢失了场景的深度信息，因此无法直接恢复空间点的三维坐标。双目视觉利用两个相机从不同角度拍摄图像，通过视差计算可以获取目标的深度信息，但需要对两个相机进行精确的校准，系统配置复杂一些。多目视觉则利用多台相机同时拍摄图像，可以获得更多的视角和深度信息，但系统配置更复杂，成本和计算量也会相应增加。单目视觉测量无法直接获取图像的深度信息，多目视觉测量与双目相机相比，系统的配置更加复杂，成本和计算量也会相应增加，因此，双目视觉技术受到了广泛的欢迎。

国内外学者针对双目视觉技术与应用也做了不少相关研究。学者 Chen<sup>[16]</sup>提出了基于两级双目视觉对接大型部件的姿态测量方法，通过降低了坐标变换的误差，克服了范围与精度的矛盾。学者 Zhang<sup>[17]</sup>构建了一种基于双目视觉的形状重建和测量系统，为检测技术在自动修整研究中提供了重要的技术支持。宋超群等人<sup>[18]</sup>通过对传统角点检测算法 FAST 进行改进，并使用改进 BRIEF 描述特征点，完成图像特征点的匹配，最后根据视差和三角测量原理完成了物体尺寸测量。

相机标定作为双目视觉中最基础的部分，其主要目的是完成双目相机模型参数的估计，从而获取空间坐标与像素坐标的换算关系<sup>[19]</sup>。所以，相机标定的精度及其标定的稳定性对后期的研究工作具有一定的影响。为此，国内外众多学者做了大量研究。由最早的 Horaud 标定方法到 Zhuang 等增加畸变进行改进<sup>[20]</sup>，然而应用最广泛的还是张氏标定法。张氏标定法是由张正友博士<sup>[21]</sup>于 2000 年提出的一种相机标定方法。该方法通过使用标定板上的已知特征点，通过摄像机拍摄的图像与实际世界坐标系之间的对应关系，来确定相机的内部参数和外部参数。这种方法具有高效、精准的特点，并且在机器视觉领域的实际应用中起到了重要的推动作用。Dattar 等人<sup>[22]</sup>在张氏标定法的基础上进行了改进，他们利用传统标定算法获得的参数作为初始值，然后使用规范平面上的特征点进行标定板的定位，并通过迭代优化的方式重新计算相机参数，直到达到收敛的效果。这种改进方法可以提高相机标定的精度，使得标定结果更加准确可靠。

立体匹配作为双目视觉中最为困难且最有挑战性的部分，其目的是对左右平面上同名点进行匹配，计算视差图，从而获取深度信息。传统立体匹配方法通常包含以下四个步骤：匹配代价计算、代价聚合、视差计算和视差优化<sup>[23]</sup>。一直以来，如何提高匹配算法的实时性、准确性、鲁棒性、适应性等问题是众多科研人员的目标。学者 Lu 等人<sup>[24]</sup>针对在弱纹理或重复性纹理匹配效果较差的问题，通过引入

Census 变换后左右图像中相应像素的汉明距离作为能量函数数据项的相似的度量，减少了对像素值的依赖。学者李迎松<sup>[25]</sup>为了提高立体匹配算法的效率，基于 Census 变换的 SGM 立体匹配算法做了大量并行算法优化。

### 1.3 本文主要研究内容

论文主要以现实生活场景中的物体为测量目标，通过双目视觉系统完成物体尺寸测量。本文利用开源机器视觉库（Open Source Computer Vision Library, OpenCV）完成对双目相机的标定，重点研究双目视觉的立体匹配算法，结合传统匹配代价算法的优缺点，提出自适应 Census 变换与图像颜色、梯度特征结合的匹配代价计算方法，针对物体测量过程对目标物体边缘区域的精确度要求，提出采用多尺度最小生成树（Minimum Spanning Tree, MST）与扫描线优化的代价聚合算法。在目标图像特征点提取过程中，研究现有分割算法的优缺点，采用基于深度学习的图像分割并提出边缘搜索优化，获取准确的目标图像轮廓，后采用 Harris 角点检测获取角点，并利用特征点特征筛选目标物体特征点。最后，整合各研究章节内容，完成物体尺寸测量。论文内容安排如下：

第一章：绪论。介绍了课题的研究背景及意义，对双目视觉的来源和应用进行了简要的介绍。然后介绍了机器视觉在各个领域的研究现状，并介绍了国内外学者针对双目视觉测量过程中的技术进行的研究。最后介绍了本文的主要研究内容以及章节安排。

第二章：双目相机标定及立体校正。通过描述四个坐标系之间的换算关系建立相机成像模型，然后介绍了相机标定及校正的基本原理，最后使用张正友相机标定法获取了左右摄像机的内参数、外参数以及畸变系数，并根据摄像机参数信息对双目相机进行了立体标定以及立体校正。

第三章：半全局立体匹配（Semi-Global Matching, SGM）算法研究及改进。首先对传统的半全局立体匹配算法进行了简要介绍，然后在代价计算阶段，并针对经典算法的不足，提出了具有自适应窗口的 Census 代价和图像梯度信息、颜色信息相结合的融合匹配代价。在代价聚合阶段，针对双目视觉测量对图像边缘的匹配要求，采用多尺度下的最小生成树聚合策略，并引入扫描线算法，减少误匹配点并提高了整体匹配精度，最后采用视差计算和视差优化方法获取视差图。最后采用 Middlebury 测试平台的标准数据将本文所提算法与同类改进算法和经典算法进行对比，可以看出算法改进能够获取较高匹配精度的视差图。

第四章：双目视觉物体尺寸测量实现。首先介绍了双目视觉系统的测量方案设计，基于前两章双目视觉测量相关技术的研究结果，搭建双目视觉测量系统。在完成目标物体的图像采集后，根据第二章的理论，完成图像的立体校正。然后简要介绍图像分割算法，并采用基于神经网络的图像分割算法对校正后图像进行初分割，并提出边缘搜索优化方法得到目标物体轮廓，然后使用 Harris 角点检测并过滤无效特征点，得到目标物体特征点。最后根据相机内外参数、视差信息和特征点坐标获取特征点三维坐标，完成对目标物体尺寸测量实验，通过实验验证测量系统的可靠性与改进算法的有效性，并对测量误差进行分析。

## 第 2 章 相机标定及立体校正

### 2.1 双目立体视觉模型

#### 2.1.1 相机成像模型

相机成像模型<sup>[26]</sup>是用数学公式刻画成像点在世界坐标系与图像坐标系之间映射关系的抽象数学模型，即被拍摄的物体在三维空间中的坐标点与图像中的像素点之间的几何变换关系。建立三维空间坐标点与图像像素点之间的映射关系，涉及到坐标系的转换。在进行相机标定之前需要明确坐标系的关系，共包含四个坐标系：世界坐标系、相机坐标系、图像坐标系和像素坐标系，相机的成像可以分为刚体变换、透视投影、和仿射变换<sup>[27]</sup>三个步骤，如图 1 所示。

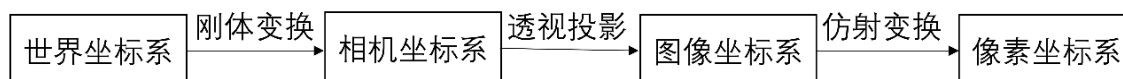


图 1 相机成像流程图

假设三维空间中任意点的坐标为：世界坐标系 $(o_w, x_w, y_w, z_w)$ 、相机坐标系 $(o_c, x_c, y_c, z_c)$ 、图像坐标系 $(o_i, x, y)$ 和像素坐标系 $(o_0, x, y)$ 。

#### (1) 世界坐标系转换为相机坐标系

在双目视觉系统中，需要将世界坐标系中的点转换为相机坐标系的点。这个转换过程涉及到平移和旋转的变换。首先，平移变换是将世界坐标系中的点从原点平移到相机所在的位置。这个平移变换是根据相机的位置和方向来确定的。平移变换只改变了点的位置，而不改变其形状。其次，旋转变换是根据相机的方向来确定的。相机有一个观察方向和一个向上方向。根据这两个方向，可以确定一个旋转矩阵，将世界坐标系中的点旋转到相机坐标系中。平移变换和旋转变换可用两个变量来描述，旋转矩阵 $R$ 和平移向量 $t$ 。

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + t \quad (1)$$

式中：

$R$ —正交旋转矩阵，描述世界坐标系与相机坐标系之间的旋转方向；

$t$ —平移矢量，描述世界坐标与相机坐标之间的平移关系。

齐次坐标下，公式(1)可以表示为：

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2)$$

## (2) 相机坐标系转换为图像坐标系

从相机坐标系转换为图像坐标系的过程，也就是使用透视投影将相机坐标系投影到成像平面上，得到相机坐标系中的点。这个过程满足小孔成像模型，是从3D转换到2D的过程。通过简单的相似三角形原理，可得相机坐标系转图像坐标系原理图，如图2所示。

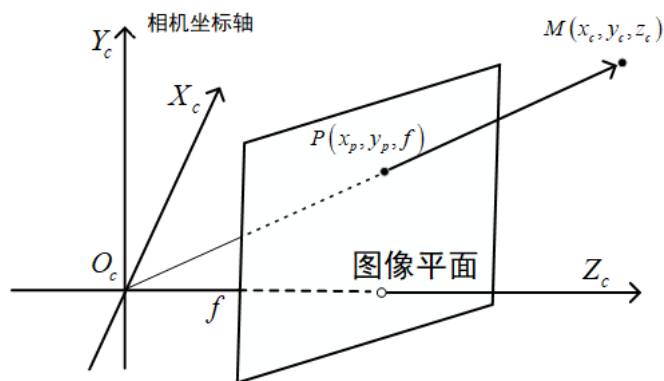


图2 小孔成像模型

从图中的相似三角形可得， $\frac{x_c}{x} = \frac{y_c}{y} = \frac{z_c}{f}$ ，其中  $f$  为焦距。通过变换可得

$x = \frac{x_c}{z_c} f$ ， $y = \frac{y_c}{z_c} f$ ，在齐次坐标系下表示为：

$$z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (3)$$

## (3) 图像坐标系转换为像素坐标系

在相机中，光线经过镜头后最终成像在感光阵列上。感光阵列是由许多光敏元件组成的，每个光敏元件对应图像中的一个像素。在感光阵列上，每个像素的物理尺寸在  $x$  和  $y$  方向上分别表示为  $dx$  和  $dy$ 。通常情况下，这两个尺寸是接近的，但也存在一定的误差。此外，感光阵列的法向和相机光轴之间也存在一定的偏差。这意味着在成像过程中，光线可能不是垂直地投射到感光阵列上。这种偏差也会对成像结果产生一定的影响，原理图如3所示。

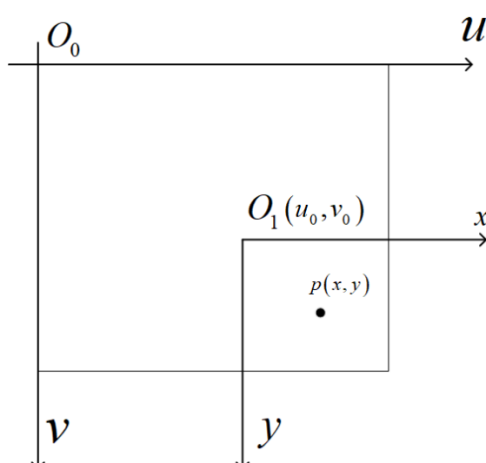


图3 图像坐标系到像素坐标系

从图中可知，两坐标系之间的变换满足  $u = u_0 + \frac{x}{dx}$ ,  $v = v_0 + \frac{y}{dy}$ 。其中， $dx$ 、 $dy$  表示像素坐标系中每个像素点的宽和高，而图像坐标系原点在像素坐标系中的横纵坐标为  $u_0$  和  $v_0$ 。在齐次坐标下：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4)$$

当知道四个坐标系之间的变换关系，就可以得到世界坐标系到像素坐标系的关系：

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (5)$$

式中：

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{—相机内参；}$$

$$\begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \text{—相机外参。}$$

相机成像模型是一种理想化的模型，然而在实际应用中，由于各种原因，摄像

机在采集图像时会出现一定程度的失真，这被称为非线性畸变。镜头畸变在实际相机中主要分为径向畸变和切向畸变两种类型。

径向畸变是由于镜头制造工艺的不完美导致镜头形状存在缺陷而产生的。它会使得离图像中心较远的像素在图像平面上的实际位置与理想位置之间存在差异。具体而言，枕形畸变会使得距离图像中心较远的像素比理想位置更远，而桶形畸变会使得距离图像中心较近的像素比理想位置更近。径向畸变通常可以使用二阶或三阶多项式函数进行建模和校正，径向畸变的数学表达式如下所示：

$$\delta_{xr} = x(k_1 r^2 + k_2 r^4 + k_3 r^6 + K) \quad (6)$$

$$\delta_{yr} = y(k_1 r^2 + k_2 r^4 + k_3 r^6 + K) \quad (7)$$

切向畸变则是由于镜头和图像平面之间的非平行性造成的。它会导致图像中的像素沿着水平和垂直方向出现拉伸或压缩的情况。切向畸变中常见的是薄透畸变和离心畸变两种情况。薄透畸变是投影的倾斜导致的，当光线透过倾斜的透镜会导致传感器上的像素位置发生畸变，造成图像出现拉伸或压缩的现象。而离心畸变是由多个透镜组成的镜头中，没有实现所有的透镜都在同一条线上导致的，这种畸变会导致不同位置的像素被放大或缩小，从而产生图像的不均匀形变。切向畸变通常可以使用两个参数来表达水平和垂直方向的畸变程度，参数的数学表达式如下所示：

$$\delta_{xd} = 2p_1 xy + p_2(r^2 + 2x^2) + K \quad (8)$$

$$\delta_{yd} = 2p_1(r^2 + 2y^2) + 2p_2 xy + K \quad (9)$$

在引入镜头畸变后，真实的图像坐标系与理想的图像坐标系关系可以表示为：

$$x' = x + \delta_{xr} + \delta_{xd} \quad (10)$$

$$y' = y + \delta_{yr} + \delta_{yd} \quad (11)$$

### 2.1.2 相机标定基础

相机标定是机器视觉中非常重要的一步，它的目的是确定相机的内外参数，以便进行准确的图像测量和三维重建等任务。相机标定方法主要包括传统相机标定法、主动视觉相机标定方法和相机自标定法<sup>[28]</sup>。

传统相机标定法是最常用的方法之一。通常采用一个特定的标定板（例如棋盘格）作为标定目标。通过将标定板放置在不同的位置和姿态下，采集一系列图像，并对这些图像进行处理，可以计算出相机的内外参数。常用的传统相机标定方法包括 Tsai 标定法、Zhang 标定法等。传统相机标定法在标定过程中需要使用标定物体，且标定物体的制作精度会影响标定结果。同时，一些场景不适合放置标定物体，限制了传统相机标定法的应用。

主动视觉相机标定方法<sup>[29]</sup>是一种相对于传统方法更灵活和自动化的方法。主动视觉相机标定通常利用机器视觉系统内安装的辅助设备（如机器人手臂、面部识别等）来辅助标定。通过与辅助设备的联动以及特定的标定模式和算法，可以实时地获取相机的内外参数。基于主动视觉的相机标定方法具有算法简单、能够获得线性解、鲁棒性较高的优点，但系统成本较高、实验设备昂贵、实验条件要求较高，且不适用于运动参数未知或无法控制的场合。

相机自标定法<sup>[30]</sup>是指相机可以自行进行标定而无需外界标定目标的方法。这种方法通常需要相机本身具备一定的自标定能力，例如在相机内部装置了校准结构或使用特殊材质制成的标定棋盘等。相机自标定法通常通过在图像中检测和分析特定的模式或特征，从而实现标定参数的推导。相机自标定方法具有灵活性强、可进行在线定标的优点，但由于其基于绝对二次曲线或曲面的方法，算法的鲁棒性较差。

这些相机标定方法具有各自的优势和适用场景，由于双目物体尺寸测量对相机标定的精度要求较高，因此本文采用了传统相机标定法作为双目视觉测量的相机标定方法。传统相机标定法中，Tsai 两步法<sup>[31]</sup>是一种常用的方法。该方法先线性求得相机参数，然后考虑畸变因素，得到初始的参数值，再通过非线性优化得到最终的相机参数。Tsai 两步法的速度较快，但只考虑径向畸变，当相机畸变严重时，该方法可能不适用。另一种常用的标定方法是张氏标定法。该方法使用由二维方格组成的标定板进行标定，通过采集标定板在不同位姿下的图片，并提取图片中的角点像素坐标，可以计算出相机的内外参数的初始值。然后利用非线性最小二乘法估计畸变系数，并最后使用极大似然估计法来优化参数。该方法简单易操作，并且精度较高，适用于大多数实际场景。综合考虑双目视觉测量的需求和实验条件，本文采用张氏相机标定法进行双目相机的标定。

## 2.2 相机标定及立体校正原理

### 2.2.1 张正友标定法

张正友标定法的基本原理是基于特制的模板平面来构建目标点与图像点之间的单应性矩阵，即世界坐标与像素坐标之间的映射关系。该方法使用了特制的棋盘格标定板，因此可以知道棋盘格角点在三维场景中的坐标。通过提取棋盘格角点在像素坐标系下的坐标，就能确定棋盘平面（三维坐标）和图像平面（像素坐标），从而构建单应性矩阵。接下来，利用正交约束条件将相机内参和外参的初始值分离出来。最后，利用极大似然估计算法对结果进行优化，以获取左右相机的内外参矩阵。



假设空间内存在任意一点  $P$ ，在世界坐标系下的坐标为  $(X_w, Y_w, Z_w)$ 。点  $P$  在摄像机上的投影点为  $p$ ，投影点  $p$  在图像像素坐标系下的坐标为  $(u, v)$ 。根据公式 (5) 可得：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A[R \quad t] \begin{bmatrix} X_w & Y_w & Z_w & 1 \end{bmatrix} \quad (12)$$

式中：

$s$  — 为比例因子，表示实际世界尺寸与图像中像素尺寸间的比例关系；

$A$  — 为内参矩阵；

$R$  — 为旋转矩阵；

$t$  — 为平移矩阵。

根据张氏标定法理论，假设标定板平面位于世界坐标系  $Z_w = 0$  的平面上。若将旋转矩阵  $R$  的第  $i$  行向量用  $r_i$  表示，则公式 (12) 可以变为：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A[r_1 \quad r_2 \quad r_3 \quad t] \begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = A[r_1 \quad r_2 \quad t] \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad (13)$$

设单应矩阵  $H = A[r_1 \quad r_2 \quad t]$ ，它用来描述物体在世界坐标系和像素坐标系的位置映射关系，则标定板上的点到图像上的映射关系为：

$$sp = HP \quad (14)$$

令  $H = [h_1 \quad h_2 \quad h_3]$ ，可得：

$$[h_1 \quad h_2 \quad h_3] = \lambda A[r_1 \quad r_2 \quad r_3] \quad (15)$$

式中：

$\lambda$  — 比例因子， $\lambda = 1/s$ 。

因为旋转矩阵为正交矩阵，因此有  $r_1^T r_2 = 0$ ， $|r_1| = |r_2| = 1$ ，将公式 (15) 带入可得：

$$h_1^T A^{-T} A^{-1} h_2 = 0 \quad (16)$$

$$h_1^T A^{-T} A^{-1} h_2 = h_2^T A^{-T} A^{-1} h_1 \quad (17)$$

融合单应矩阵的两个约束条件，并令  $B = A^{-T} A^{-1}$ ，其公式表达如下所示：

$$B = A^{-T} A^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \quad (18)$$

式中：

$B$  —  $3 \times 3$  的对阵矩阵。

则可以表示为:

$$b = [B_{11} \ B_{12} \ B_{13} \ B_{22} \ B_{23} \ B_{33}]^T \quad (19)$$

$$h_i = [h_{i1} \ h_{i2} \ h_{i3}]^T \quad (20)$$

式中:

$h_i$ —表示  $H$  的第  $i$  向量。

则可以表示为:

$$h_i^T B h_j = v_{ij}^T b \quad (21)$$

$$v_{ij} = [h_{i1}h_{j1} \ h_{i2}h_{j1} + h_{i1}h_{j2} \ h_{i2}h_{j2} \ h_{i3}h_{j1} + h_{i1}h_{j3} \ h_{i3}h_{j2} + h_{i2}h_{j3} \ h_{i3}h_{j3}]^T \quad (22)$$

将公式 (16) 和公式 (17) 的两个约束条件整理可得:

$$\begin{bmatrix} v_{12}^T \\ (v_{12} - v_{22})^T \end{bmatrix} b = 0 \quad (23)$$

如果相机进行标定时采集了  $n$  对标定图像 ( $n > 3$ )，则将会有  $n$  组公式 (23) 方程。通过  $n$  组方程式联立可以求出参数:

$$\begin{cases} v_0 = (B_{12}B_{13} - B_{11}B_{23}) / B_{11}B_{22} - B_{12}^2 \\ \lambda = B_{33} - [B_{33}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})] / B_{11} \\ f_x = \sqrt{\lambda / B_{11}} \\ f_y = \sqrt{\lambda B_{11} / (B_{11}B_{22} - B_{12}^2)} \\ u_0 = -B_{13}f_x^2 / \lambda \end{cases} \quad (24)$$

根据 (24) 可以求出外部参数为:

$$\begin{cases} r_1 = \lambda A^{-1} h_1 \\ r_2 = \lambda A^{-1} h_2 \\ r_3 = r_1 \times r_2 \\ t = \lambda A^{-1} h_3 \end{cases} \quad (25)$$

## 2.2.2 最大似然估计

由于相机的内参数和外参数是在没有外界干扰的情况下推导得出的，实际情况中会存在一定的偏差。因此，为了增强相机标定结果的可靠性，可以采用最大似然估计方法。假设相机标定图像为  $n$  幅，每幅图像上有  $m$  个角点，则第  $i$  幅图像上的第  $j$  个角点投影点  $\tilde{m}(A, R_i, t_i, M_{ij})$  与其在标定板上所对应的点  $M_{ij}$  的关系可以表示为:

$$\tilde{m}(A, R_i, t_i, M_{ij}) = A[R \ t]M_{ij} \quad (26)$$

式中:

$A$ —相机的内参;

$R_i$ 、 $t_i$ —表示第 $i$ 幅图像的旋转矩阵和平移矩阵。

则概率密度函数可以表示为:

$$f(m_{ij}) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\sum_{i=1}^n \sum_{j=1}^m (\tilde{m}(A, R_i, t_i, M_{ij}) - m_{ij})^2}{\sigma^2}} \quad (27)$$

构造似然函数为:

$$L(A, R_i, t_i, M_{ij}) = \prod_{i=1}^n \prod_{j=1}^m f(m_{ij}) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\sum_{i=1}^n \sum_{j=1}^m (\tilde{m}(A, R_i, t_i, M_{ij}) - m_{ij})^2}{\sigma^2}} \quad (28)$$

从公式(28)中可知, 当 $\sum_{i=1}^n \sum_{j=1}^m (\tilde{m}(A, R_i, t_i, M_{ij}) - m_{ij})^2$ 取得最小值时,  $L$ 可以取得最大值。

在只考虑径向畸变的情况下, 可以得出畸变多项式:

$$\begin{aligned} x &= x' + x' [k_1(x'^2 + y'^2) + k_2(x' + y')^2] \\ y &= y' + y' [k_1(x'^2 + y'^2) + k_2(x' + y')^2] \end{aligned} \quad (29)$$

式中:

$(x', y')$ —为理想情况下的归一化图像坐标点;

$(x, y)$ —为有畸变的归一化图像坐标点。

根据相机内参模型可得:

$$\begin{cases} u = u_0 + \alpha x + \gamma y \\ v = v_0 + \beta y \end{cases} \quad (30)$$

为了能够最好地描述相机坐标系与真实世界坐标系之间的关系, 采用零假设求解最优的参数值。假设 $\gamma = 0$ , 将公式(29)和公式(30)联立并化为矩阵形式可得以下公式:

$$\begin{bmatrix} (u' - u_0)(x'^2 + y'^2) & (u' - u_0)(x'^2 + y'^2)^2 \\ (v' - v_0)(x'^2 + y'^2) & (v' - v_0)(x'^2 + y'^2)^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u - u' \\ v - v' \end{bmatrix} \quad (31)$$

式中:

$(u_0, v_0)$ —为相机主点;

$(u', v')$ —为理想模型下的像素点坐标;

$(u, v)$ —为实际情况下的像素点坐标。

则径向畸变为:

$$k = [k_1 \quad k_2]^T = (D^T D)^{-1} D^T d \quad (32)$$

采用最大似然估计优化结果, 使用 Levenberg-Marquardt 算法求取的最优目标函

数如下式所示：

$$\sum_{i=1}^n \sum_{j=1}^m (\tilde{m}(A, k_1, k_2, R_i, t_i, M_{ij}) - m_{ij})^2 \quad (33)$$

最终得到相机的畸变系数  $k_1$ 、 $k_2$ 。

### 2.2.3 双目立体标定

通过使用张正友标定法对左右摄像机进行标定后，可以联合使用两个摄像机的外参数进行立体标定，以获取两个摄像机之间的相对位置。相对位置可以用旋转矩阵  $R$  和平移向量  $T$  来表示。

假设空间内存在一点  $P$ ，在世界坐标系下的坐标为  $P_w$ ， $P_l$  表示  $P$  点在左相机坐标系下的投影坐标， $P_r$  表示  $P$  点在右相机坐标系下的投影坐标，以及平移矩阵  $t_l$  和  $t_r$ 。则  $P$  点在左右相机坐标系中的投影坐标用数学表达式表示为：

$$\begin{cases} P_l = R_l P_w + t_l \\ P_r = R_r P_w + t_r \end{cases} \quad (34)$$

考虑左相机为主坐标系，则  $P_l$  和  $P_r$  存在以下关系：

$$P_l = R P_r + T \quad (35)$$

由公式 (34) 和公式 (35) 可以推出双目相机的外参为：

$$\begin{cases} R = R_r R_l^{-1} \\ T = t_r - R t_l \end{cases} \quad (36)$$

### 2.2.4 立体校正原理

双目视觉采用的为双目平行结构模型，然而在现实情况中，双目相机很难保证是完全平行。为了满足后续的图像处理及计算的条件，需要对图像进行立体校正。立体校正是利用双目摄像机的内外参数来实现的，通过相机的内参消除图像的畸变，通过相机的外参消除系统中由于相机不平行造成的误差，OpenCV 中不仅有相机标定的方法，也包含立体校正的函数，本文通过 OpenCV 完成对双目相机的标定和图像的校正。

立体校正首先将旋转矩阵  $R$  分解成两个矩阵  $R_l$  和  $R_r$ ，且满足  $R_l$  和  $R_r$  互为逆矩阵。 $R_l$  和  $R_r$  的数学表达式为：

$$\begin{cases} R_l = R^{\frac{1}{2}} \\ R_r = R^{\frac{1}{2}} \end{cases} \quad (37)$$

为了立体校正后的重投影畸变最小，将左右相机各旋转一般，此时左右相机的

光轴和左右成像平面保持平行。根据平移矩阵  $T$  构建变换矩阵  $R_{rect}$ ，令左视图的极点变换至无穷远处，此时左目和右目相机的连线与成像平面平行。

构造  $e_1$  的数学表达式为：

$$e_1 = \frac{T}{\|T\|}, T = [T_x, T_y, T_z]^T \quad (38)$$

构造  $e_2$ ，方向与光轴正交，且与  $e_1$  垂直，其数学表达式为：

$$e_2 = \frac{\begin{bmatrix} -T_y & T_x & 0 \end{bmatrix}}{\sqrt{T_x^2 + T_y^2}} \quad (39)$$

得到  $e_1$  和  $e_2$  后，定义  $e_3$  分别于  $e_1$ 、 $e_2$  正交，因此：

$$e_3 = e_1 \times e_2 \quad (40)$$

根据公式 (38)、公式 (39) 和公式 (40)，得到变换矩阵  $R_{rect}$  的表达式为：

$$R_{rect} = [e_1^T \quad e_2^T \quad e_3^T]^T \quad (41)$$

左右摄像机整体旋转矩阵为合成旋转矩阵与变换矩阵的乘积，其表达式为：

$$\begin{cases} R_l' = R_{rect} R_l \\ R_r' = R_{rect} R_r \end{cases} \quad (42)$$

## 2.3 相机标定和立体校正的实现

### 2.3.1 标靶图像采集

常用的标靶图像包含棋盘格和同心圆等，在本文中，选择了规格为  $11 \times 8$ 、单元格边长为 30mm 的棋盘格作为标靶图像，用于双目相机标定，如图 4 所示。

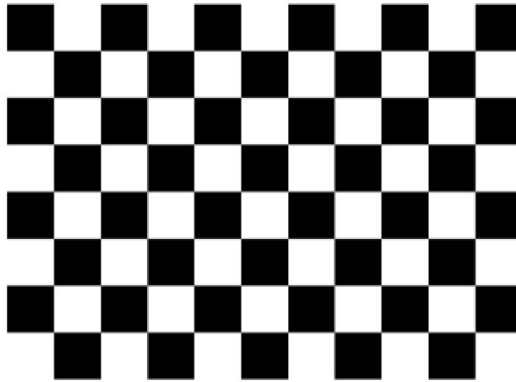


图4 标靶图像

在进行相机标定的过程中，标靶图像的采集非常重要。为了获得准确的相机内外参数，需要在不同方位和不同视角下拍摄标靶图像，以尽可能覆盖相机的成像视野。对于张氏标定法来说，至少需要采集 3 张不同的标靶图像才能进行参数求解。

随着标靶图像数量的增多，标定参数的稳定性会逐渐增强。然而，当图像数量过多时，可能会出现图像相似度过高的问题，导致标定结果出现较大误差。因此，在选择标靶图像数量时需要进行权衡，尽量选择适量的图像来平衡稳定性和准确性。

在进行相机标定实验时，外界因素对标定参数精确性的影响是一个重要的考虑因素。为了减小外界环境引起的非必要标定误差，我们应尽量避免外界因素的干扰。例如，选择在室内环境稳定的条件下进行图像拍摄，避免光照变化、遮挡物、反射等因素对图像质量的影响。另外，最好选择具有丰富纹理和对比度的标定板，以便更准确地检测并提取角点。有学者的研究表明，在标定图像数量在 15 至 25 张之间时，可以得到最精确的相机内外参数。因此，本文采集了 18 组标定图像进行相机标定实验，以获得较为准确的标定参数。这些标定图像如图 5 所示。



图 5 双目相机采集标靶图像

### 2.3.2 基于 OpenCV 的相机标定及立体校正

OpenCV 提供了丰富的函数和工具，用于处理图像和视频、进行计算机视觉任

务和机器学习。它提供了丰富的图像处理和计算机视觉算法，方便开发人员进行各种任务，如图像滤波、边缘检测、物体检测、特征提取等。在使用 OpenCV 进行双目相机的标定过程中，首先循环读取拍摄的每一个标靶图像，将其转换成灰度图像，并使用 `findChessboardCorners()` 函数检测棋盘格的角点，将检测到的角点坐标保存起来，作为标定的输入。然后使用 `cornerSubPix()` 函数对角点进行亚像素级别的优化，提高角点的精度。最后，使用 `calibrateCamera()` 函数对相机进行标定。将检测到的角点坐标以及相机的内参数初值作为输入，计算相机的内参数和畸变参数。基于 OpenCV 的双目相机标定的角点检测效果如图 6 所示。



图6 角点检测效果图

最后使用 `calibrateCamer()` 函数对所有标定板的图像进行相机标定，得到相机的内部和外部参数。标定的相机内外参数如表 1 所示：

表1 双目相机标定参数

	左摄像头	右摄像头
内参矩阵	$\begin{bmatrix} 644.8193 & -0.6612 & 315.0428 \\ 0 & 644.0194 & 239.3396 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 642.2449 & -1.0136 & 324.1832 \\ 0 & 641.4311 & 255.0053 \\ 0 & 0 & 1 \end{bmatrix}$
畸变矩阵	$[-0.433 \ 0.209 \ -0.001 \ 0.003 \ -0.034]$	$[-0.425 \ 0.192 \ -0.003 \ 0.001 \ -0.01]$
旋转矩阵	$\begin{bmatrix} 1 & 0.0026697 & -0.00265969 \\ -0.00265969 & 1 & -0.00866214 \\ 0.002263646 & 0.00866924 & 1 \end{bmatrix}$	
平移矩阵	$[-63.20819762 \ -0.07061582 \ -0.34382417]$	

在实验中使用的双目相机存在明显的畸变效应。为了准确地处理畸变，根据相机标定得到的内外参数，在双目相机立体校正的原理下，使用 `undistort` 函数对图像



进行去畸变处理，得到更加准确的图像结果。校正前后的图像对比可参见图 7，可以清晰地观察到未校正的图像具有明显的相机畸变效果，而校正后的图像则能满足后续处理的需要。

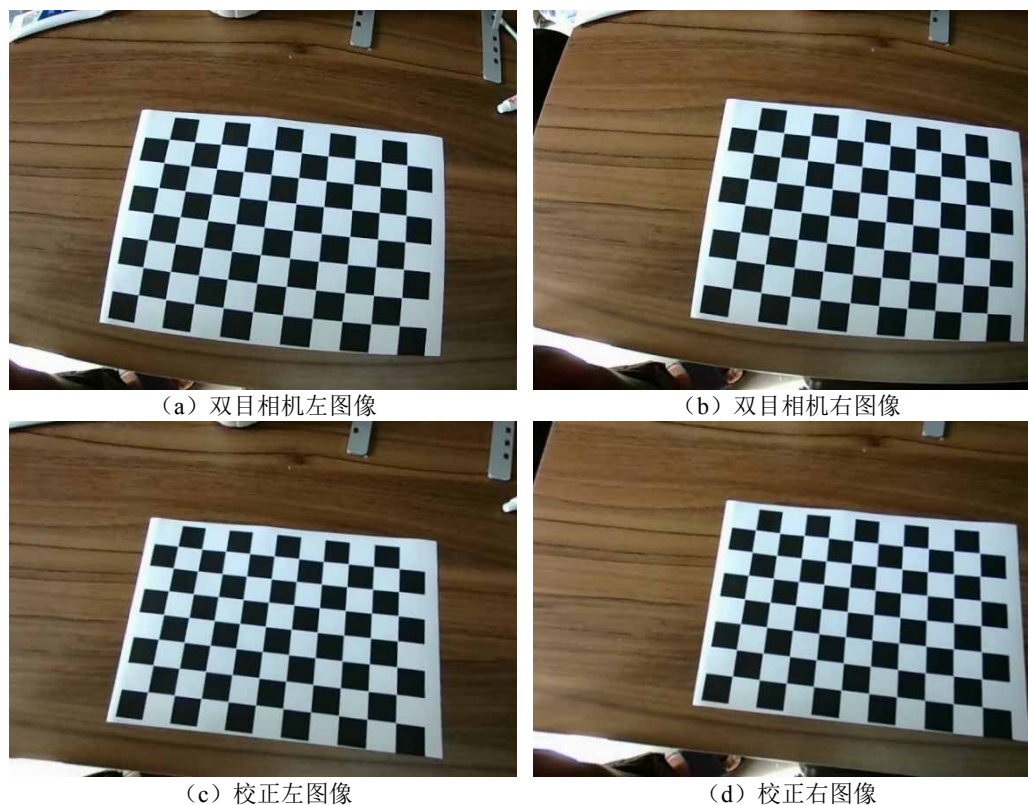


图 7 立体校正效果图

## 2.4 本章小结

本章介绍了双目相机标定及校正的基础理论和实现方法。首先，介绍了相机的成像原理，并建立了世界坐标系到像素坐标系的数学映射模型。这样可以将三维空间中的点映射到二维图像上的像素点，实现了空间坐标与图像像素点之间的对应关系。接着，介绍了张正友等人提出的相机标定方法的基本原理和步骤。在标定过程中，利用最大似然估计的方法来估计相机的内外参数和畸变系数。通过使用特定的标定板，采集一系列具有已知空间坐标的图像，求解出相机的标定参数。最后，利用 OpenCV 库来实现双目相机的标定及校正。标定和校正的目的是消除双目相机位置和视角不同引起的误差，使得双目图像可以更准确地进行深度信息提取和三维重构。校正算法采用了 Bouguet 极线校正算法，通过对图像进行几何变换来校正双目图像。



## 第3章 半全局立体匹配算法研究及改进

### 3.1 立体匹配算法的基本流程

立体匹配是计算机视觉领域中一项重要的任务，旨在从两幅或多幅图像中找到对应的像素点，用于恢复场景的三维信息。立体匹配的应用包括立体视觉、三维重建和深度估计等。立体匹配算法的基础流程如图8所示：

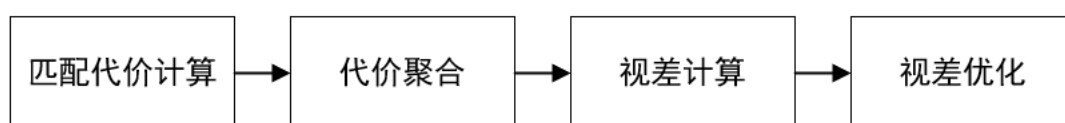


图8 立体匹配流程图

匹配代价计算计算两个像素点之间的相似性，通常使用两个像素邻域内一定大小窗口的像素来进行计算。但这样计算只考虑了局部信息，很容易受到噪声等因素的影响，这样就导致同名点不是最小，因此就需要代价聚合让代价值能够准确反映像素间的相关性，它依据相邻的像素应该具有连续的视差值，建立像素间的联系，来对代价矩阵进行优化。代价计算时无论是否是同名点，都可以进行代价匹配，但是非同名的两个像素点的代价更高，而视差计算就是通过搜索代价矩阵找到最小代价的像素点。通过视差计算就得到了一个视差图，但得到的视差图同样可能受到噪声、遮挡等因素的印象，因此需要对像素点进行检查和优化。

### 3.2 基于 Census 的半全局立体匹配算法

#### 3.2.1 匹配代价计算

在半全局立体匹配算法<sup>[32]</sup>中,使用互信息来进行匹配代价计算。互信息是一种对图像明暗变化不敏感的相关性测度，它通过计算两幅影像各自的熵和它们的联合熵来定义。熵代表了影像的信息量，是基于灰度的概率分布所得到的统计量，影像的熵越大则代表它包含的像素灰度越丰富，灰度分布越均匀。为了计算互信息，首先需要预先知道视差图来对匹配影像进行纠正。SGM 采用的是一种分层迭代的方案，通过对影像进行逐级降采样得到多层影像对。对于最高层的影像对，随机生成一张视差图作为初始视差。然后，利用初始视差计算得到的代价数组作为初始代价数组，通过代价聚合得到新的视差图。接着，将新的视差图作为下一层影像对的视差图，

并重复上述步骤。这个迭代过程持续迭代到最底层影像，一般迭代三次即可获得较好的匹配结果。

由于基于互信息的匹配代价计算由于需要初始视差值，所以需要通过分层迭代的方式得到较为准确的匹配代价值，而且概率分布计算稍显复杂，这导致代价计算的效率并不高。学者 Zabih 和 Woodfill<sup>[33]</sup>提出了基于 Census 变换法被广泛用于匹配代价计算。Census 变换，是一种常用的图像特征描述方法，在立体匹配中被广泛应用。Census 变换通过比较像素灰度之间的关系，将图像转换为二进制的比特串，从而捕捉到了像素之间的结构信息。在匹配代价计算中，Census 变换可以用来度量两个像素之间的相似度。

Census 变换过程可通过如下公式表达：

$$T(p) = \bigotimes_{q \in N_p} \zeta[I(p), I(q)] \quad (43)$$

式中：

$I(p)$ —为参考像素灰度值；

$I(q)$ —为变换窗口内非中心像素灰度值；

$N_p$ —表示中心像素  $p$  的邻域；

$\bigotimes$ —为比特位的逐位连接运算。

像素标记运算则由下面公式定义：

$$\zeta[I(p), I(q)] = \begin{cases} 1, I(p) \geq I(q) \\ 0, I(p) < I(q) \end{cases} \quad (44)$$

如图 9 所示，设邻域窗口为  $5 \times 5$  的大小，在视图选取任一点，以该点为中心画出一个  $5 \times 5$  的窗口，将窗口邻域像素的灰度值与中心像素的灰度值进行比较，Census 变换后的比特串为 1111010110111111111100。

35	31	22	20	19
40	17	25	30	18
30	25	20	23	26
25	35	24	24	27
30	42	40	17	19

→

1	1	1	1	0
1	0	1	1	0
1	1	X	1	1
1	1	1	1	1
1	1	1	0	0

图 9 Census 变换示意图

通过公式 (44) 可以得到了 Census 变换的一串二进制，基于 Census 变换的匹配代价计算方法是计算左右影像对应的两个像素的 Census 变换值的汉明 (Hamming) 距离，即匹配代价为：

$$C_{Census}(p, q) = Ham \min g(T(p), T(q)) \quad (45)$$

式中:

$T(p)$ —左图产生的二进制串;

$T(q)$ —右图产生的二进制串。

Hamming 距离是一种度量两个比特串之间差异的方法,其计算方式为将两个比特串进行异或运算,然后统计异或运算结果中为 1 的比特位的数量。Hamming 距离越小,意味着两个比特串越相似。

### 3.2.2 匹配代价聚合

由于代价计算只考虑了图像的局部相关性,直接使用计算得到的代价数组进行视差图计算容易受到噪声、光照变化等因素的干扰,导致匹配错误。所以需要代价聚合对像素周围邻域的代价进行统计和整合,通过考虑多个像素间的一致性和相互约束来消除噪声和主观差异。

为了获得更好的匹配效果,SGM 算法采用全局立体匹配算法的思想,通过对代价数组进行聚合操作来得到最终的视差图。全局能量函数的定义如下所示。

$$E(d) = E_{data}(d) + E_{smooth}(d) \quad (46)$$

式中:

$E_{data}(d)$ —视差图匹配代价的测度;

$E_{smooth}(d)$ —对视差图约束的平滑项。

能量函数最小化是一个二维最优问题,这是一个多项式复杂程度的非确定性问题,有很多近似的较为高效的能量最优策略如图割、置信度传播、合作优化等算法被用来解决这个问题,但是效率上依旧需要进一步改进。为了更高效的解决这个二维最优问题,SGM 算法采用基于类似于扫描线或者叫单方向动态规划的方法,使用一维路径聚合的方式来近似二维最优,相比其他解决方法效率更高,效果相当。

SGM 算法提出的能量函数如下所示:

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \quad (47)$$

式中:

$C$ —匹配代价,表示视差图为  $D$  时,所有匹配代价的累加;

$\sum_{q \in N_p} P_1 T[|D_p - D_q| = 1]$ —平滑项,对像素  $p$  邻域内视差变化较小的情况进行

惩罚,  $P_1$  是一个惩罚常数;

$\sum_{q \in N_p} P_2 T[|D_p - D_q| > 1]$ —平滑项，对像素  $p$  邻域内视差变化较大的情况进行

行惩罚； $P_2$  是非固定惩罚力度。

较小的惩罚项可以让算法能够适应视差变化小的情形，如倾斜的平面或者连续的曲面，较大的惩罚项可以让算法正确处理视差非连续情况，由于影像的亮度边缘位置是前景背景交界的可能性较大，这些位置的像素邻域内往往不是视差连续的，相差较大，为了保护真实场景中的视差非连续情况， $P_2$  往往是根据相邻像素的灰度差进行动态调整，其数学表达式如下所示：

$$P_2 = \frac{P_2'}{|I_{bp} - I_{bq}|}, P_2 > P_1 \quad (48)$$

式中：

$P_2'$ — $P_2$  初始值；

$\frac{P_2'}{|I_{bp} - I_{bq}|}$ —邻域像素灰度值之差。

根据公式 (48)， $P_2$  的惩罚力度和灰度值差成反比，可以看出当像素与邻域像素灰度值差很大的情况下，则这个像素可能在视差非连续区域，就会对超过 1 个像素的惩罚力度适当减小。

针对能量函数的优化问题，SGM 算法提出了一种路径代价聚合思路，将像素所有视差下的匹配代价进行像素周围所有路径上的一维聚合得到路径下的路径代价值，然后将所有路径代价值相加得到该像素聚合后的匹配代价值。像素点  $p$  沿着  $r$  的路径代价计算公式如下所示：

$$L(p, d) = C(p, d) + \min \left\{ \begin{array}{l} L_r(p-r, d) \\ L_r(p-r, d-1) + P_1 \\ L_r(p-r, d+1) + P_1 \\ \min_i L_r(p-r, i) + P_2 \end{array} \right\} - \min_i L_r(p-r, i) \quad (49)$$

式中：

$C(p, d)$ —匹配代价值；

$\min_i L_r(p-r, i)$ —前一个像素点的最小路径代价。

公式 (49) 通过减去前一个像素点的最小路径代价的方式，保证新的路径代价值不超过一定上限，即：

$$L \leq C_{\max} + P_2 \quad (50)$$

将所有路径上的路径代价累加求和，就是最终的聚合代价。相比起全局匹配复杂的匹配代价计算过程，仅包含了简单的加减和求一维最优解，就达到了相当的匹配精度。

### 3.2.3 视差计算与优化

视差计算用于确定不同视点下同一场景点之间的像素位移，以获得场景的深度信息。WTA<sup>[34]</sup>（Winner-Takes-All）是立体匹配中一种常用的视差计算方法，它基于代价聚合所得的代价立方体来确定最优匹配的像素对应的视差值。对于代价立方体中的每个像素位置，选取其中最小代价对应的视差值。即在每个像素位置上，选择代价最小的视差值作为最终的视差结果，原理如图10所示。

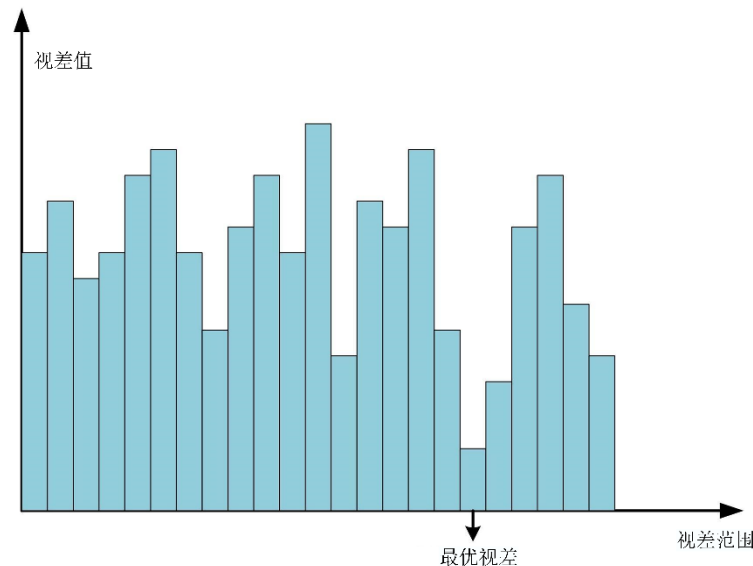


图10 赢家通吃示意图

视差优化是在初步得到视差图之后，对视差图的结果进行判断，对发现的匹配错误情况进行改正，其方法旨在通过考虑场景中的空间一致性和平滑项来修正计算出的视差值。

#### （1）唯一性检测

唯一性检测是立体匹配中的一个重要步骤，用于判断视差计算结果的唯一性和可靠性。唯一性检测的基本思想是通过分析局部窗口内的图像信息，判断观察点周围的像素在不同视差值下的响应程度，即选取与其他视差值差异最大的视差值作为最终视差值。

#### （2）左右一致性检查

左右一致性检查用于验证计算得到的左视图与右视图的匹配结果的一致性。其

基本思想是通过将右视图中的像素在左视图中进行反向投影，然后检查反向投影的像素位置上的视差值是否与右视图中的原始视差值一致，以判断匹配结果的准确性。

### （3）视差填充

视差填充用于在视差图中填补视差值缺失的区域，以便形成完整的深度图。在立体匹配过程中，由于一些像素可能由于纹理缺失、遮挡或其他原因无法匹配到对应像素，导致视差图中存在一些视差值缺失的区域。而视差填充的目的就是通过利用已有的视差信息，推断和填补这些缺失值，使得深度图更加完整。

## 3.3 半全局立体匹配算法的改进

### 3.3.1 算法流程

基于 Census 变化的半全局立体匹配算法中，使用固定窗口的 Census 变化，无法完美匹配图像中的所有场景，且作为匹配代价较为单一，使得匹配代价可靠性低，扫描线聚合方法没有全局的视差一致性约束，容易在存在遮挡、纹理模糊的区域出现误匹配现象，为了提高立体匹配的精度，减少在纹理不连续区域、遮挡区域和边界区域误匹配率，提高立体匹配的准确性，本文提出了一种高精度的半全局立体匹配算法。

对传统 Census 变化进行改进，通过图像边缘信息和像素灰度，来自适应 Census 变化的匹配窗口大小，可以使 Census 变化结果更加可靠。单一 Census 变化计算匹配代价在重复纹理区域和弱纹理区域容易出现误匹配，而绝对值差（Absolute Differences, AD）算法使基于单个像素点计算的匹配代价，受光照不均、图像噪声影响较大，但对纹理丰富区域有较好的匹配效果，AD 算法考虑的单像素，可以在一定程度上缓解 Census 算法对重复纹理的效果不好的问题，加入梯度信息可以降低在视差不连续区域的误匹配率。代价聚合阶段，利用具有边缘保持的最小生成树聚合方法，得到一个初始的视差估计结果。然后再使用四方向的扫描线优化方法对初始结果进行进一步的优化和修正，以提高视差估计的准确性。当观察分辨率低的图像时更加关注图像整体轮廓信息，观察分辨率高的图像时有更多细节信息被关注。仅使用原始分辨率图像进行代价聚合算法容易在弱纹理区域或重复纹理区域产生误匹配，因此采用多尺度聚合思想，使用高斯金字塔获取不同尺度的图像后，对不同图像进行代价计算和代价聚合，然后融合各个尺度的聚合代价值获取最终聚合代价。最后视差计算和优化得到最终视差图，流程图如图 11 所示。

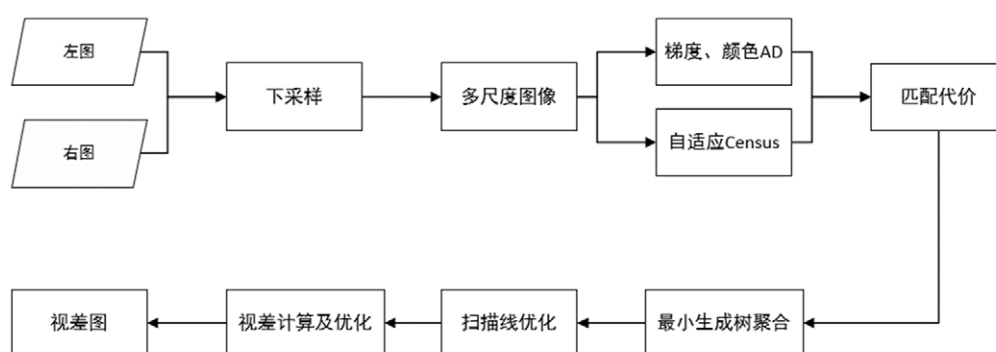


图 11 本文算法流程图

### 3.3.2 匹配代价计算

在 Census 变换的过程中, 图像的边缘区域包含大量的图像信息, 但干扰信息也会增加, 较大的窗口也会引入更多的干扰信息, 可能导致误匹配率增加。另一方面, 在弱纹理区域和重复纹理区域, 图像的信息较少, 如果使用较小的窗口进行匹配, 可能无法获取足够的图像信息进行准确的匹配, 因此对具有不同特点的区域应采用不同大小的窗口。

针对双目视觉的物体尺寸测量对边缘区域的需要, 本课题引入了边缘纹理信息。首先对图像进行 Sobel 算子边缘检测, 得到图像在  $x, y$  方向的梯度值, 梯度幅值计算公式如下所示:

$$G(p) = \sqrt{g_x^2 + g_y^2} \quad (51)$$

式中:

$G(p)$ —像素点  $p$  处的梯度幅值;

$g_x$ 、 $g_y$ — $x, y$  方向的梯度值。

然后, 选取一个固定窗口, 计算窗口内灰度均方差, 计算公式如下所示:

$$\sigma(p) = \sqrt{\frac{1}{N} \sum_{i=1}^N [I(p_i) - M(p)]^2} \quad (52)$$

式中:

$N$ —初始窗口像素个数;

$I(p_i)$ —初始窗口中各像素的灰度值;

$M(p)$ —初始窗口的灰度均值。

通过将梯度幅值和窗口内灰度均方差融合, 自适应地选择合理的变换窗口大小以获取更精准的匹配代价。设定权重阈值  $T1$  和  $T2$ , 当权重小于  $T1$  时, 说明匹配窗口内图像信息较少, 需要扩大匹配窗口, 当权重大于  $T1$  小于  $T2$ , 说明设

置的匹配窗口可以很好的进行匹配，当权重大于  $T_2$  时，说明匹配窗口内图像信息较多，需要缩小窗口以避免干扰信息的影响。从 3.4 的实验结果与分析中，可以看出当固定窗口  $5 \times 5$  时，可以获得较其他窗口精确的匹配代价，因此本文采用  $5 \times 5$  为初始匹配窗口，自适应匹配窗口变换窗口大小公式如下所示：

$$W_{\text{adapt}}(p) = \begin{cases} 7 \times 7 & G(p) + \sigma(p) < T_1 \\ 5 \times 5 & T_1 \leq G(p) + \sigma(p) < T_2 \\ 3 \times 3 & T_2 \leq G(p) + \sigma(p) \end{cases} \quad (53)$$

Census 变换是基于比较像素灰度之间的关系来计算的，不是依赖于绝对灰度值，因此对图像的明暗变化不敏感，但在重复纹理区域的效果不好。而且 Census 变换使用的是图像的灰度图，但灰度图包含的图像信息有限，无法获得较为可靠的匹配代价。针对 Census 变换的缺点，为了增加匹配代价的鲁棒性，提高弱纹理区域和重复纹理区域的匹配精度，本文在匹配代价计算时，在自适应 Census 变换的基础上，引入 AD 变换。

AD 算法是图像处理中常用的匹配代价计算方法之一。它通过比较像素灰度之间的绝对差异来度量两个像素之间的相似度。与其他算法相比，AD 算法是基于单个像素点进行匹配代价计算的，因此在光照不均、图像噪声较多的情况下，其效果会受到较大的影响。但是，在纹理丰富的区域，AD 算法的匹配效果较好。为了缓解 Census 算法在处理重复纹理时的问题，提高 AD 算法在抗噪性方面的表现，将 AD 算法与 Census 算法结合使用。这样可以在一定程度上互补两种算法的优缺点，提高匹配结果的准确性。

颜色对于图像的表达和识别具有重要意义，通过利用颜色信息，可以更准确地区分图像的前景和背景，提高匹配的精度。对纹理丰富区域有较好的匹配效果，纹理丰富的区域往往包含了更多的颜色变化。传统 AD 算法采用图像的灰度信息，无法充分利用图像的颜色信息。而基于 RGB 颜色的 AD 变换可以更敏感地捕捉颜色的变化，因此在纹理丰富的区域具有较好的匹配效果。噪声通常会引入图像中的颜色变化，基于 RGB 颜色的 AD 变换可以通过比较像素之间的颜色差异来判断相似度，相对于灰度图像的方法，它可以更好地抵抗噪声的影响，提高匹配的鲁棒性。本文采用图像像素点的 RGB 三通道颜色分量的绝对值均值作为 AD 变换值，基于 RGB 的颜色变换公式如下所示：

$$C_{AD}(p, d) = \frac{1}{3} \sum_{i \in \{R, G, B\}} |I_i(p) - I'_i(p_d)| \quad (54)$$

式中：



$C_{AD}(p, d)$ —像素  $p$  在视差  $d$  时的匹配代价;

$i$ —图像的 RGB 三通道;

$I_i(p)$ —像素  $p$  在通道  $i$  的像素值;

$I'_i(p_d)$ —像素  $p$  偏移  $d$  时在通道  $i$  的像素值。

自适应 Census 变换和 AD 变换都是基于一个邻域窗口, 对于图像边缘区域没有有效的约束。基于梯度的代价计算方法可以进一步提供更丰富的几何和细节信息。结合自适应 Census 和 AD 的代价计算方法, 可以辅助梯度信息, 从而提供更准确的匹配结果。本文采用自适应 Census 变换、AD 和基于梯度的代价计算方法结合的融合匹配代价, 得到更可靠的匹配代价。

Census 变换使用的是图像的灰度图, 但灰度图包含的图像信息有限, 无法获得较为可靠的匹配代价。为了提高匹配精度, 使用具有截断阈值的颜色特征和梯度特征构建联合代价函数, 联合代价函数如下所示:

$$C_{AG}(p, d) = \alpha \left[ \min \left[ \frac{1}{3} \sum_{i \in \{R, G, B\}} |I_i(p) - I'_i(p_d)|, T_{AD} \right] + (1 - \alpha) \left[ |\nabla I(p) - \nabla I'(p_d)|, T_{GRD} \right] \right] \quad (55)$$

式中:

$C_{AG}(p, d)$ —表示通过颜色信息和梯度信息得到的匹配代价融合后的联合匹配代价;

$I_i$ 、 $I'$ —分别表示左右图像;

$\alpha$ —分别为颜色、梯度的权重划分;

$p_d$ —表示右图中与  $p$  视差为  $d$  的像素;

$\nabla$ —表示图像的梯度;

$T_{AD}$ 、 $T_{GRD}$ ——分别为颜色、梯度信息的截断阈值。

由于联合代价和 Census 计算式不同的度量, 所以它们得到的代价图像具有不同的范围。在进行融合时, 通常需要对代价图像进行归一化处理, 确保它们在相同的范围内, 以便更好地进行后续的匹配和处理操作。采用自然指数函数归一化后得到最终的匹配代价函数, 其公式定义如下:

$$C(p, d) = 2 - \exp[-C_{AG}(p, d) / \beta_1] - \exp[-C_{Census}(p, d) / \beta_2] \quad (56)$$

式中:

$\exp$ —自然指数函数, 使匹配代价归一化;

$\beta_1$ 、 $\beta_2$ —联合匹配代价和自适应 Census 变换的权重。

通过使用指数进行代价计算的归一化, 可以调整代价图像的范围, 并增强图像中不同区域的对比度, 从而有助于后续的立体匹配过程。这种归一化方法通常能够

提高立体视觉算法的稳定性和性能。改进匹配代价计算流程如表 2 所示。

表 2 改进匹配代价计算流程

算法 1: 改进匹配代价计算
输入: 校正后左图像 leftImg, 校正后右图像 rightImg
输出: 匹配代价矩阵 costVol
1: 获取左图像边缘信息 leftEdges
2: 初始化匹配代价 costCensus
3: for 左图像中像素:
4:     统计匹配窗口内边缘像素个数
5:     计算匹配窗口内均方误差
6:     根据边缘像素个数和均方误差判断自适应匹配窗口大小
7:     计算匹配窗口内 Census 变换
8:     计算 Hamming 距离
9:     计算自适应 AD 匹配代价 censusCost
10: 计算基于梯度的匹配代价 gradCost
11: 结合 AD 和基于梯度的匹配代价 costAg=weightAD*adCost+weightGrad*gradCost
12: 结合归一化代价 costVol=weightCensus*exp(censusCost)+weightAG*exp(costAg)

### 3.3.3 匹配代价聚合

在经典的代价聚合算法中, 最小生成树聚合模型<sup>[35]</sup>具有边缘保持的特点, 且聚合精度高、效率高等优点, 因此本课题采用最小生成树进行代价聚合。MST 的聚合模型采用将图像看作一个加权无向图, 将两点间的灰度差值作为两点所成边的权值, 然后, 通过从相邻像素的边的集合中选择边的子集来构建树, 最后所有像素的总成本可以通过两次连续遍历树结构  $T$  来获取, 分别为自底向下聚合和自顶向上聚合, 两次聚合公式为:

$$C_D^{A\uparrow}(p) = C_d(p) + \sum_{q \in Ch(p)} S(p, q) \square C_D^{A\uparrow}(q) \quad (57)$$

$$C_d^A(p) = S(Pr(p), p) \square C_d^A(Pr(p)) + (1 - S^2(Pr(p), p)) \square C_d^{A\uparrow}(p) \quad (58)$$

式中:

$C_d^{A\uparrow}(p)$ 、 $C_d^A(p)$ —分别表示叶节点到根节点与根节点到叶节点的聚合结果;

$Ch(p)$ —包含像素  $p$  所有子项;

$Pr(p)$ —为节点  $p$  的父节点。

根据最小生成树的原理, 可以将图像视为一个四联通区域。图像中的每个像素点可以看作是一个节点, 而图像中相邻像素点之间的边的权值则表示它们的灰度值之差, 即表示它们的相似度。针对图像生成的最小生成树, 其聚合方法如图 12 所示。

在图像的最小生成树中，有两种聚合方法。

首先是自底向上的聚合。以节点 V4 为例，计算 V4 节点的聚合代价值可以直接计算其子节点 (V3,V4) 和 (V5, V4) 的聚合值与它们各自权值的乘积的集合。从最小生成树的底部开始，逐层计算直到根节点。这样，就可以得到一个自底向上聚合后的最小生成树。

接下来是自顶向下的聚合。以节点 V3 为例，假设 V3 为根节点，它的父节点 V4 被转换为它的子节点。V3 节点的聚合代价的计算方式与 V4 节点的聚合代价相同，只是需要将 V3 节点的聚合代价计算结果加上 V4 节点的聚合值与它们的权值的乘积。其他节点依次进行自顶向下的聚合计算，最终得到最小生成树的代价。

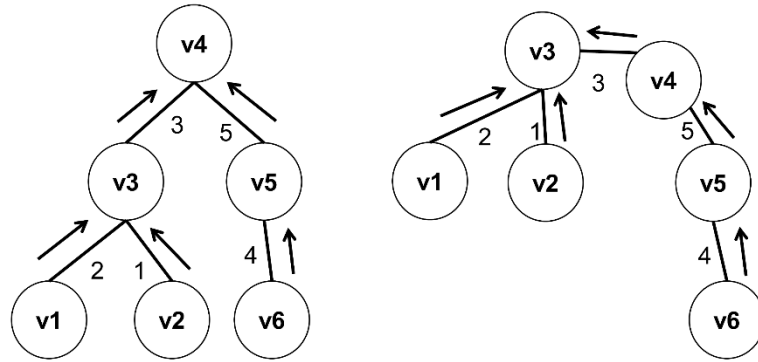


图 12 最小生成树聚合模型图

最小生成树聚合方法可能存在误匹配的问题，即错误地合并了不应该匹配的像素，导致错误的视差结果。这可能发生在存在纹理模糊、重复结构或遮挡的图像区域。为了进一步提高代价精度且减少匹配错误，本文引入扫描线算法对聚合结果进行优化。先使用最小生成树聚合方法进行代价聚合，得到一个初始的视差估计结果。然后再使用扫描线聚合方法对初始结果进行进一步的优化和修正，以提高视差估计的准确性。

扫描线算法将像素所有视差下的匹配代价进行像素周围所有路径上的一维聚合得到路径下的路径代价值，然后将所有路径代价值相加得到该像素聚合后的匹配代价值。四个扫描线优化是相互独立的，像素  $p$  在视差  $d$  处的代价  $G_r(p,d)$  的优化公式如下：

$$C_r(p,d) = C_1(p,d) + \min(C_r(p-r,d)), C_r(p-r,d \pm 1) + P_1, \min C_r(p-r,k) + P_2) - \min C_r(p-r,k) \quad (59)$$

式中：

$p-r$  —表示沿同一方向的前一个像素；

$p_1$ 、 $p_2$  —惩罚项，根据左右图色差进行动态调整。

公式 (59) 中惩罚项的具体设定规则如下:

$$\begin{aligned}
 &1. P_1 = \Pi_1, P_2 = \Pi_2, \text{if } D_1 < \tau_{so}, D_2 < \tau_{so} \\
 &2. P_1 = \Pi_1 / 4, P_2 = \Pi_2 / 4, \text{if } D_1 < \tau_{so}, D_2 > \tau_{so} \\
 &3. P_1 = \Pi_1 / 4, P_2 = \Pi_2 / 4, \text{if } D_1 > \tau_{so}, D_2 < \tau_{so} \\
 &4. P_1 = \Pi_1 / 4, P_2 = \Pi_2 / 10, \text{if } D_1 > \tau_{so}, D_2 > \tau_{so}
 \end{aligned} \tag{60}$$

式中:

$\Pi_1$ 、 $\Pi_2$ —设定的固定阈值;

$\tau_{so}$ —设定的颜色差阈值。

最后, 将 4 个方向的优化代价取平均值赋值给像素  $p$  :

$$C(p, d) = \frac{1}{4} \sum_r C_r(p, d) \tag{61}$$

根据人眼视觉的习惯, 轮廓信息和细节信息在不同尺度观察下不同。当观察分辨率低的图像时更加关注图像整体轮廓信息, 观察分辨率高的图像时有更多细节信息被关注。仅使用原始分辨率图像进行代价聚合算法容易在弱纹理区域或重复纹理区域产生误匹配, Zhang 等提出了基于多尺度代价聚合的立体匹配算法, 使用图像的多尺度信息进行立体匹配, 得到了的视差图更加精准。因此本文采用多尺度聚合思想, 使用高斯金字塔获取不同尺度的图像后, 对不同图像进行代价计算和代价聚合, 然后融合各个尺度的聚合代价值获取最终聚合代价。改进代价聚合流程如表 3 所示。

表 3 改进代价聚合流程

算法 2: 改进代价聚合
输入: 校正后左图像 leftImg, 校正后右图像 rightImg
输出: 代价聚合矩阵 aggVol
1: 初始化参数 i, 用于统计多尺度图像信息
2: if i<5 do
3:     改进匹配代价计算 costVol
4:     将图像构成四联通图, 权重为像素灰度
5:     自底向上聚合后自顶向下聚合
6:     分别从四个方向聚合优化
7:     存储多尺度聚合代价 aggVol[i]
8:     左右图像下采样
9: end
10: 融合多尺度聚合代价 aggVol

### 3.4 实验结果与分析

#### 3.4.1 立体匹配评价标准

在对双目视觉的立体匹配结果进行评估时，常采用两种方法。一种方法是通过对比标准视差图，判断是否产生了误匹配，但这种方法无法进行定量判断和分析。另一种方法是使用立体匹配算法进行测距估计，先构建一个真实的三维场景，在现实世界中进行标准测距，然后将从视差图中计算得到的深度值与标准测量的数值进行比对，以此来评估视差图的准确性。这种方法可以提供定量的评价结果，并对立体匹配算法的性能进行分析。

Middlebury 立体匹配评价平台是一个广泛应用于计算机视觉领域的立体匹配算法评估平台，是现在最常用的立体匹配算法平台。Middlebury 提供了一系列的立体匹配数据集，包括了具有不同场景和难度的图像对。这些数据集中的图像对包含了真实的深度信息作为真实视差图，方便算法进行和验证。它还提供了用于评价立体匹配结果的标准，是用来度量立体匹配算法性能的重要指标，可帮助研究人员进行算法比较和分析。Middlebury 中提供的评价标准是基于均方根误差（root-mean-square, RMS）和错误匹配率（Percentage of bad matching pixels, PBM）：

均方根误差 RMS：

$$R = \left( \frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2 \right)^{\frac{1}{2}} \quad (62)$$

误匹配率 PBM：

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_T(x,y)| > \delta_d) \quad (63)$$

式中：

$N$ —像素点总量；

$d_C(x,y)$ —算法输出的视差估计值；

$d_T(x,y)$ —标准视差值；

$\delta_d$ —判断阈值。

误匹配率是 Middlebury 立体匹配评价平台中最常用的评价指标之一，用于评估立体匹配算法在生成视差图时产生的误匹配情况。误匹配率的计算基于比较算法得到的视差估计值与标准视差值之间的差异，如果差异大于设定的阈值，则将其判定为误匹配点。通过计算误匹配点的数量，可以获得误匹配点在整个视差图中的占比。在 Middlebury 评价平台的第三版中，针对不同的图像，设定了不同的阈值来评估误

匹配率。此外，对于误匹配率的评价，还进一步考虑了不同区域的情况，包括非遮挡区域、全部区域和视差不连续区域。这样的评价方式更加细致地考虑了不同区域的特点，能够更准确地评估算法在不同场景下的误匹配情况。

### 3.4.2 改进代价计算验证

为了客观公正的评价本文提出的算法性能，实验采用 Middlebury 测试平台第三版标准数据集 Tsukuba、Venus、Teddy、Cones 在全部区域(All)、非遮挡区域(Nocc)和视差不连续区域(Disc)的误匹配率作为评价指标，数据的最大视差为 15, 19, 59, 59。根据参考文献[36]实验参数为基础，并进行微调后的实验参数设置为： $T_{AD} = 7$ ， $T_{GRD} = 2$ ， $\alpha = 0.11$ ， $\beta_1 = 35$ ， $\beta_2 = 15$ 。

为验证算法改进的有效性，分别采用 11×11、9×9、5×5 和 3×3 的固定窗口，与自适应窗口、自适应融合代价的匹配代价计算方法进行对比，在代价聚合阶段均采用最小生成树聚合算法，不进行视差优化。使用 Middlebury 测试平台第三版标准数据集，并分别计算了全部区域误匹配率、非遮挡区域误匹配率和视差不连续区域误匹配率。实验结果如表 4、5 和 6 所示。

表 4 代价计算全部区域误匹配率

	Tsukuba	Venus	Teddy	Cones	Average
11x11	4.49	2.88	16.88	14.70	9.73
9x9	4.23	2.78	16.62	14.23	9.46
7x7	3.88	2.65	16.47	13.78	9.19
5x5	4.01	2.50	16.19	13.41	9.02
3x3	6.61	2.62	16.50	13.02	9.68
自适应窗口	3.76	2.19	15.73	12.91	8.64
自适应融合代价	<b>3.34</b>	<b>1.23</b>	<b>13.82</b>	<b>10.64</b>	<b>7.25</b>

表 5 代价计算非遮挡区域的误匹配率

	Tsukuba	Venus	Teddy	Cones	Average
11x11	3.57	1.70	7.88	4.92	4.51
9x9	3.32	1.62	7.64	4.48	4.26
7x7	3.04	1.53	7.50	4.05	4.03
5x5	3.02	1.44	7.19	3.67	3.83
3x3	5.67	1.60	7.59	3.54	4.60
自适应窗口	2.83	1.16	6.82	3.30	3.52
自适应融合代价	<b>2.58</b>	<b>0.49</b>	<b>5.77</b>	<b>2.46</b>	<b>2.82</b>

表6 代价计算视差不连续区域的误匹配率

	Tsukuba	Venus	Teddy	Cones	Average
11x11	11.56	13.46	21.10	13.84	14.99
9x9	10.65	12.69	20.40	12.65	14.09
7x7	8.63	11.57	19.78	11.30	12.82
5x5	7.92	10.02	18.79	10.13	11.71
3x3	7.39	9.34	19.46	9.78	11.49
自适应窗口	7.22	8.18	18.30	9.28	10.74
自适应融合代价	<b>6.49</b>	<b>4.23</b>	<b>14.97</b>	<b>7.07</b>	<b>8.19</b>

通过观察表4、5和6中的数据，可以发现在非遮挡区域、视差不连续区域甚至是全部区域，随着匹配窗口逐渐缩小，误匹配率呈现逐渐下降的趋势。当固定窗口的大小为 $5\times 5$ 时，整体的误匹配率是最低的。然而，即使使用固定 $5\times 5$ 的窗口，也存在误匹配率并非最优的情况。因此，为了获取最优的匹配代价，需要采用自适应的方式来改变Census变换窗口的大小。由于 $5\times 5$ 窗口的整体误匹配率最低，所以在自适应窗口的Census变换中采用了初始 $5\times 5$ 的窗口。

基于双目视觉的物体尺寸测量主要依赖于图像视差不连续区域。通过观察表6，可以发现使用基于自适应窗口的Census代价相比固定窗口的Census代价得到的视差图在不连续区域的误匹配率更低。此外，当融合了梯度信息和颜色信息后，误匹配率进一步降低，这证明了采用具有边缘信息的自适应Census变换代价，并与梯度信息和颜色信息融合所得到的计算代价是可行的。

为了进一步验证本文所提改进匹配代价计算方法的有效性。将所提算法与三种同类改进匹配代价算法进行对比。第一种改进匹配代价方法将AD算法和Census变换融合形成的融合代价函数（ADCensus）；第二种改进匹配代价方法将AD、x方向梯度融合形成的融合代价函数（GRD）；第三种改进匹配代价方法将AD算法、Census变换和x方向梯度融合形成的融合代价函数（CG）；对照实验均使用相同的参数权重及相同代价聚合，但不进行视差优化，直接利用代价计算匹配代价，通过WTA策略得到视差图。对照实验以Cones图像为例，视差图如图13所示。

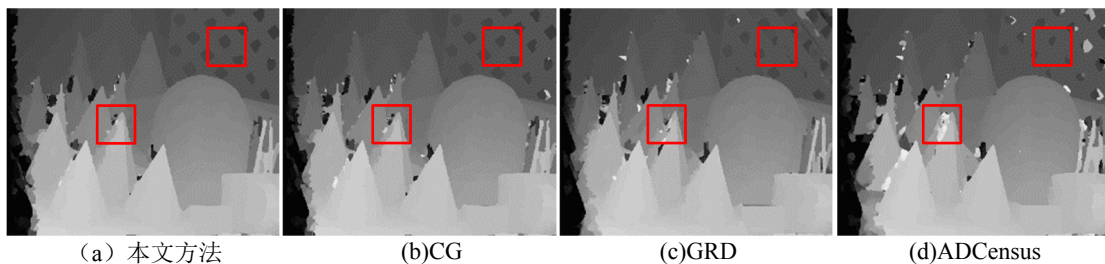


图13 代价计算视差图对比

从图中可以看出，所提算法得到的视差图中错误匹配更少，在锥体和背景格周围区域有一定的提升，在边缘交界处更为平滑。为了算法评估的准确性，进一步使用 Middlebury 数据集的 Tsukuba、Venus、Teddy、Cones 标准测试图像的视差图，获取图像的所有区域平均误差（Avg\_all）、非遮挡区域平均误差（称为 Avg\_nocc）以及非连续区域平均误差(Avg\_disc)。图 3 展示了所得结果，从图中可以看出本文所提融合代价计算方法在非遮挡区域有较好的表现，其整体匹配精度也比其他算法有一定的提升。

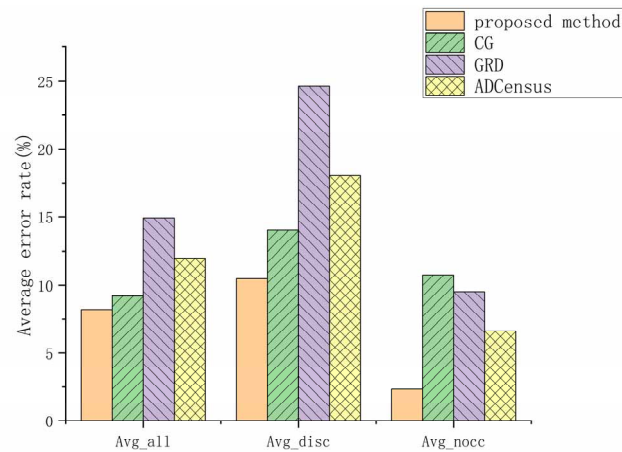


图 14 代价计算平均误匹配率

3.4.3 改进代价聚合验证

为验证算法改进的有效性，分别采用最小生成树（ST）、最小生成树加扫描线优化（STS）和多尺度最小生成树加扫描线优化（MSTS）进行对比，在匹配代价计算阶段使用 5×5 固定窗口的 Census 变换进行代价计算，不进行视差优化。使用 Middlebury 测试平台第三版标准数据集，并分别计算了全部区域误匹配率（all）、非遮挡区域误匹配率（Disc）和视差不连续区域误匹配率（Nocc）。实验结果如表 7 所示。

表 7 代价聚合误匹配率

		Tsukuba	Venus	Teddy	Cones	Average
ST	All	6.41	5.22	17.72	15.37	11.18
	Disc	9.10	13.90	20.98	12.94	14.23
	Nocc	5.48	4.21	8.86	5.64	6.04
STS	All	6.24	<b>5.17</b>	17.62	<b>15.24</b>	11.06
	Disc	8.60	13.62	20.79	<b>12.61</b>	13.90
	Nocc	5.31	<b>4.15</b>	8.74	<b>5.50</b>	5.92



表7 (续)

		Tsukuba	Venus	Teddy	Cones	Average
	All	<b>6.02</b>	5.38	<b>17.15</b>	15.26	<b>10.95</b>
MSTS	Disc	<b>8.51</b>	<b>13.49</b>	<b>19.83</b>	12.65	<b>13.62</b>
	Nocc	<b>5.2</b>	4.4	<b>8.42</b>	5.62	<b>5.91</b>

从表中可以观察到, 仅使用最小生成树进行代价聚合时, 所获得的视差图在全部区域、非遮挡区域和视差不连续区域的误匹配率都是最高的。当引入扫描线优化和多尺度图像信息后, 误匹配率逐渐降低。这一结果证明了引入扫描线优化和多尺度图像信息对于提高视差算法的准确性和鲁棒性的重要性。扫描线优化可以更好地利用视差与各像素之间的空间连续性, 从而降低误匹配率。多尺度图像信息的引入则利用了不同尺度下的特征差异, 帮助解决视差不连续区域的问题, 进一步提高视差估计的准确性。

为验证改进代价聚合方法的有效性, 将改进代价聚合方法与双边滤波器 (BF)<sup>[37]</sup>、引导滤波器 (GF)<sup>[38]</sup>、方框滤波器 (BOX)、非局部代价聚合方法 (NL)<sup>[39]</sup> 及基于分割树 (ST)<sup>[40]</sup> 的代价聚合方法进行对比。它们使用相同的代价计算方法及相同的参数设置, 且均不进行视差优化后处理, 对照实验以 Cones 图像为例, 视差图如图 15 所示。

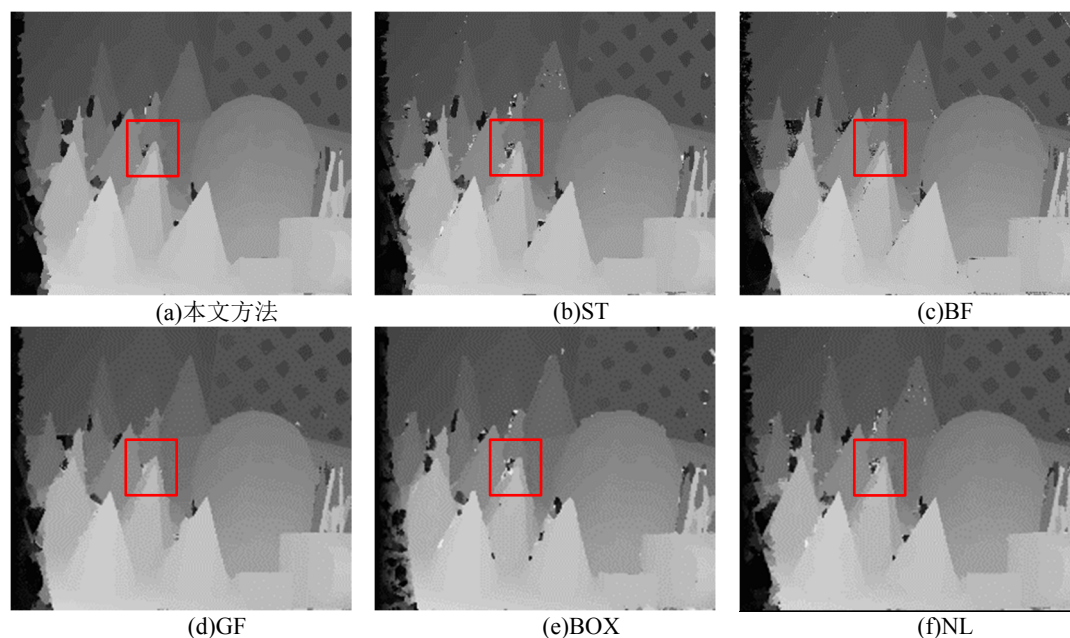


图 15 不同代价聚合下的视差图

从图中可以看出, 使用本文所提代价聚合算法获得的视差图更加的平滑, 且在视差不连续区域和非遮挡区域精度更高。为了更直观的体现所提代价聚合算法的性能, 同样对比各个代价聚合算法在 Middlebury 数据集 Tsukuba、Venus、Teddy、

Cones 的 Avg\_all、Avg\_nocc 和 Avg\_disc 指标，结果如下图所示，所提代价聚合算法的整体效果优于其他算法。

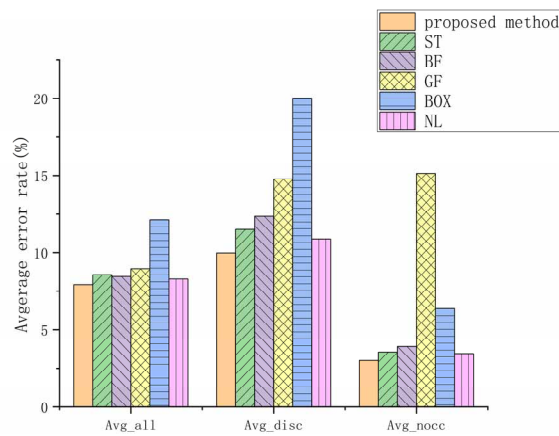


图 16 代价聚合算法视差图的平均误匹配率

3.4.4 改进算法对比分析

为了验证本文算法的整体效果，将本文算法与经典 SGM、ST 和 AD-Census 算法进行对比，并与三种改进算法进行对比，对比数据如表 8 所示。改进算法 1<sup>[41]</sup>采用了 HSV 色彩信息替代常见的 RGB 进行代价计算，并在匹配窗口中引入了距离颜色加权的 Census 变换值。最后，该算法融合了 AD 算法和改进的 Census 变换作为最终的匹配代价，并采用基于峰度系数的自适应引导滤波在代价聚合过程中进行自适应处理。改进算法 2<sup>[42]</sup>在代价计算过程中首先使用 Sobel 算法获取图像边缘信息，然后将其与 Census 变换和 AD 算法的代价值进行非线性融合。在视差优化过程中采用 MeanShift 算法对图像进行分割，并对视差图进行优化。改进算法 3<sup>[43]</sup>在匹配代价计算过程中，通过去除 Census 变换匹配窗口中的异常值，并计算剩余像素的平均值代替匹配窗口中心像素值，以获取更稳定的 Census 匹配代价。最后，该算法融合了颜色和梯度信息构建融合匹配代价来进行立体匹配。

表 8 整体算法的误匹配率

		Tsukuba	Venus	Teddy	Cones	Average
SGM	All	5.66	21.61	17.25	13.63	14.54
	Disc	16.38	24.75	21.01	14.14	19.07
	Nocc	2.51	21.04	9.65	7.23	10.11
ST	All	8.04	8.04	24.29	20.40	15.19
	Disc	13.58	24.72	27.81	21.32	21.86
	Nocc	6.21	6.50	15.75	10.53	9.75

表8 (续)

	All	2.58	2.94	15.88	14.23	8.91
AD-Census	Disc	10.15	5.46	20.6	14.4	12.65
	Nocc	<b>1.94</b>	2.64	10.49	8.61	5.92
	All	3.91	1.44	12.95	9.6	6.98
改进算法 1	Disc	12.06	9.25	19.32	10.42	12.76
	Nocc	3	0.81	7.31	3.58	3.68
	All	4.35	0.89	11.8	8.81	6.46
改进算法 2	Disc	8.94	2.81	15.6	7.57	8.73
	Nocc	3.23	0.28	6.12	2.63	3.07
	All	<b>2.26</b>	0.57	11.24	9.28	5.84
改进算法 3	Disc	8.43	2.79	14.94	9.5	8.92
	Nocc	<b>1.98</b>	0.28	5.72	3.36	2.84
	All	2.81	<b>0.44</b>	<b>10.91</b>	<b>7.72</b>	<b>5.47</b>
本文算法	Disc	<b>7.88</b>	<b>1.78</b>	<b>14.37</b>	<b>7.02</b>	<b>7.76</b>
	Nocc	2.51	<b>0.21</b>	<b>5.49</b>	<b>2.42</b>	<b>2.66</b>

从表中可以看出, 本文算法在全部区域、视差不连续区域、非遮挡区域的误匹配率平均值分别为 5.47%、7.76%和 2.66%。较经典算法中全部区域最优的 8.91%降低了 3.44%, 近期的改进算法中最优的改进算法 3 的 5.84%降低了 0.37%。视差不连续区域较经典算法中最优的 12.65%降低了 4.89%, 较近期的改进算法中最优的改进算法 2 的 8.73%降低了 0.97%。非遮挡区域较经典算法中最优的 5.92%降低了 3.26%, 较近期的改进算法中最优的改进算法 3 的 2.84%降低了 0.18%。证明了本文算法改进的有效性。

### 3.5 本章小结

本章节主要对双目视觉的立体匹配算法进行研究, 基于半局部立体匹配算法流程, 包含代价计算、代价聚合、视差计算和视差优化进行研究。在代价计算阶段, 并针对经典算法的不足, 提出了具有自适应窗口的 Census 代价和图像梯度信息、颜色信息相结合的融合匹配代价。在代价聚合阶段, 针对双目视觉测量对图像边缘的匹配要求, 采用多尺度下的最小生成树聚合策略, 并引入扫描线算法, 减少误匹配点并提高了整体匹配精度, 最后采用视差计算和视差优化方法获取视差图。最后采用 Middlebury 测试平台的标准数据将本文所提算法与同类改进算法和经典算法进行对比, 可以看出算法改进能够获取较高匹配精度的视差图。

## 第4章 双目视觉物体尺寸测量实现

### 4.1 测量方案设计

在硬件方面，本文使用双目相机和笔记本电脑进行图像的采集和处理。双目相机的采集图像分辨率为 640\*480。在软件方面，采用了 Visual Studio 2019 和 PyCharm 作为开发环境。Visual Studio 2019 用于双目相机的标定和图像的立体匹配算法的研究，PyCharm 用于对双目相机采集的图像进行批量切分和目标物体的特征点提取。

基于双目视觉的物体尺寸测量流程图如图 17 所示：

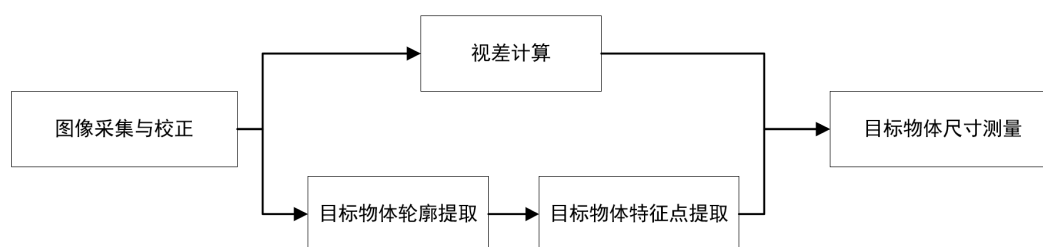


图 17 基于双目视觉的物体尺寸测量流程图

1、图像采集与校正：使用双目相机进行标靶图像和测量物体图像的采集，得到双目图像。在 PyCharm 中，将双目图像切分为左、右图像。然后使用 OpenCV 和标靶图像对，完成相机的标定，获取相机的内外参数。对采集到的图像进行立体校正，使得左右图像的视点在同一水平线上。

2、视差计算：在双目视觉中，通过比较左右相机对同一物体的采集图像，可以计算出每个像素点的视差。视差是左右图像中对应像素点的水平偏移量，它与物体距离成反比关系。根据第四章中对立体匹配算法的改进，对校正后的图像进行立体匹配，得到视差图。

3、目标物体轮廓及特征点提取：采集到的图像包含无用的背景和目标物体纹理信息，因此需要提取目标物体轮廓信息，采用 Segment Anything 图像分割算法<sup>[44]</sup>，从校正好的图像中提取目标物体的轮廓信息并进行优化，得到目标物体的轮廓信息。目标物体的特征点为轮廓边缘的交点，因此对目标物体轮廓采用 Harris 角点检测算法，从轮廓信息中提取特征点，过滤无效特征点，得到目标物体边缘的交点。

4、物体尺寸测量：使用 OpenCV 的 getRectifyTransform() 函数获取用于畸变校正和立体校正的映射矩阵以及用于计算像素空间坐标的重投影矩阵，使用 reprojectImageTo3D() 函数根据视差图生成三维深度图。在对物体尺寸进行测量时，

根据待测点的三维信息，计算两个待测点间的欧式距离，即可得到物体的尺寸信息。令两个待测点的三维坐标分别为 $(X_i, Y_i, Z_i)$ 和 $(X_j, Y_j, Z_j)$ ，则两点间的计算公式为：

$$d = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2} \quad (64)$$

## 4.2 基于双目视觉的物体尺寸测量研究

### 4.2.1 图像采集与校正

本文第二章已经完成了对双目相机的标定，相机内外参数为表 1 中数据。本节采用现实场景中的图像为例进行物体尺寸测量，使用双目相机拍摄目标物体，然后对双目图像进行切割为左图像和右图像，并进行立体校正，左右相机拍摄的原图像与校正图像如图 18 所示。

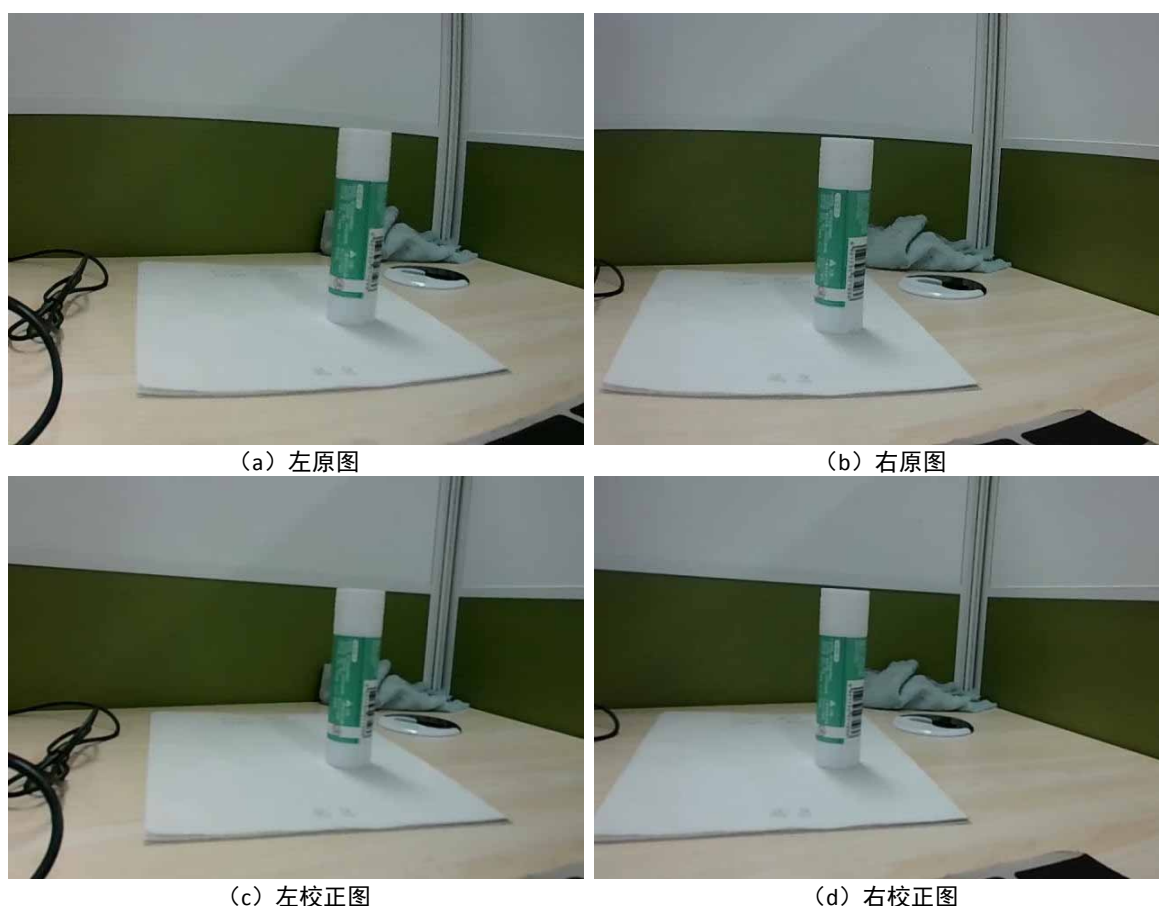


图 18 测量物体校正对比图

### 4.2.2 目标物体轮廓提取

双目视觉系统采集到的图像中包含目标物体以外的背景信息，这些信息可能会对后续的图像处理和立体匹配造成干扰。因此，在进行双目视觉测量时，需要对采

集到的左右图像进行预处理，从中检测出目标物体的图像区域，并去除图像的冗余信息。

本文研究的目标物体是现实场景中的复杂物体，获取的图片背景也具有复杂的场景，因此使用传统的图像分割算法无法较为准确的获取目标物体区域。随着算力的增加以及深度学习的不断发展，基于深度学习的图像分割算法的分割效果已经远超传统的图像分割算法。其中 Segment Anything 就是其中的佼佼者，它的核心思想是通过神经网络模型实现对图像中任意目标物体的精确分割，从而实现对任意物体分割的效果。Segment Anything 算法的核心思想是采用深度学习模型实现物体检测和分割的一体化，从而能够实现对图像中任意物体的精确分割，而本文需要对现实场景中的物体进行分割提取，因此本文采用 Segment Anything 图像模型对目标物体进行初步分割，分割结果如图 19 所示。

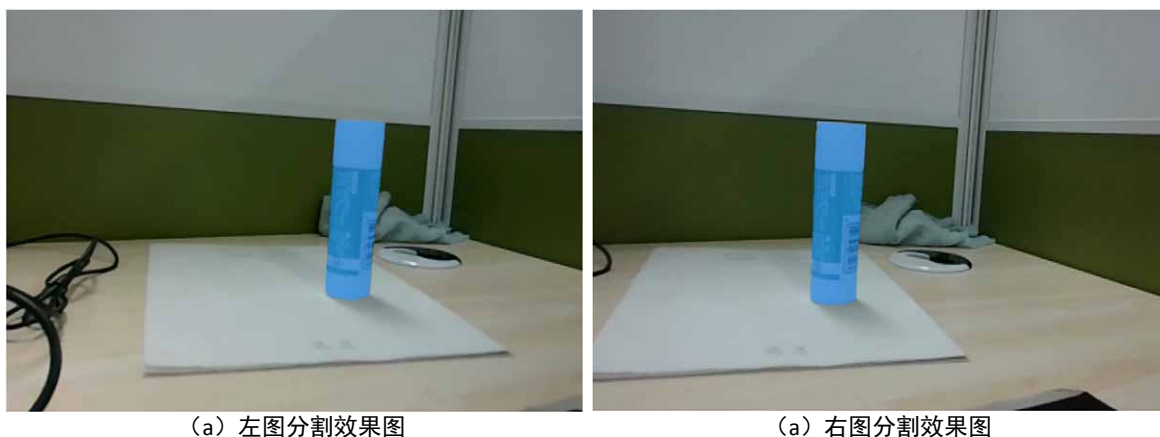


图 19 图像分割效果图

从图中可以看出，使用 Segment Anything 进行图像分割后，得到了较为准确的分割蒙版，但在目标物体边缘区域仍然有一定的误差，根据相邻像素间具有相近的颜色的特点，本文提出了边缘搜索优化。

边缘搜索优化方法首先将图像分割区域标定为 255，以避免搜索算法向目标物体内部搜索，并确定目标物体外部为搜索方向。然后以边缘像素开始，向目标物体外部延伸，当遇到颜色与该像素差别较大时停止延伸。此外，还需要设置一个最大延伸长度，避免无限制的延伸。边缘搜索优化规则的数学表达为：

$$D_c(p, p') < \tau; D_c(p, p') = \max_{i=R, G, B} |I_i(p) - I_i(p')| \quad (65)$$

$$D_s(p, p') < L; D_s(p, p') = |p - p'| \quad (66)$$

式中：

$D_c(p, p')$ —为  $p$  和  $p'$  的颜色差异，取 RGB 三分量差值的最大值；

$\tau$ —为设置的阈值；

$D_s(p, p')$ —为 $p$ 和 $p'$ 的空间长度。

采用边缘搜索优化后的效果图如图 20 所示。

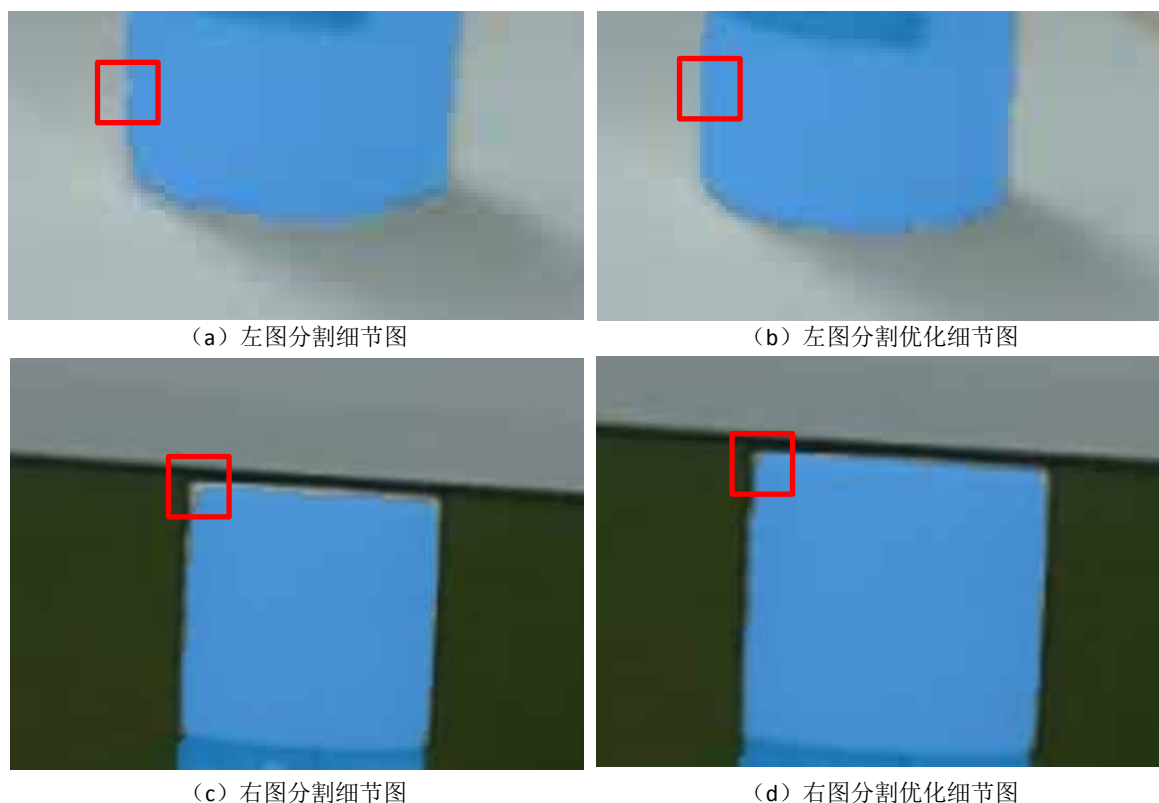


图 20 边缘搜索优化效果图

通过对比优化前后的效果图，可以观察到经过边缘搜索优化后，分割出的目标物体区域在边缘区域更加准确，提高目标物体的边缘定位和边界精度。将分割图像的目标物体区域设置为白色，背景区域设置为黑色，就得到了目标物体轮廓信息，目标物体轮廓图如图 21 所示。

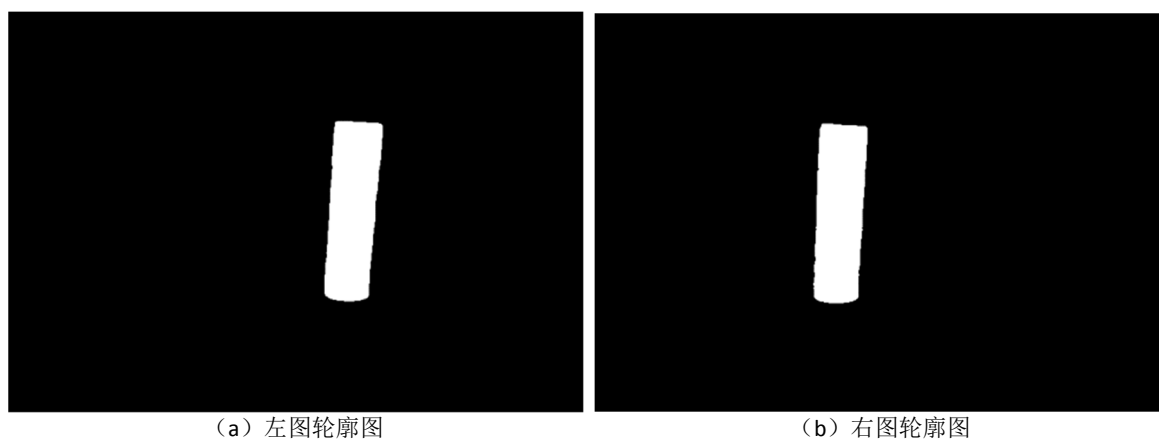


图 21 目标物体轮廓图



### 4.2.3 目标物体特征点提取

在双目视觉物体尺寸测量系统中，为了完成物体的尺寸测量，需要获取目标物体表面的特征点，这些特征点可以代表物体上的一些关键位置，比如物体的边缘、角点等。Harris 角点检测方法具有旋转不变性、尺度不变性、光照不变性，且它的抗干扰能力强，因此受到了广泛的应用，Harris 角点检测的思想是在图像中寻找像素灰度变化最大的区域，如下图所示：

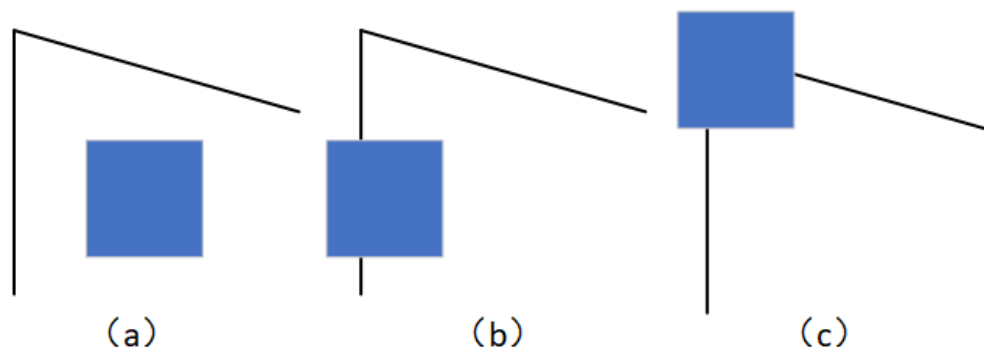


图 22 Harris 窗口检测特征

如图 22 所示，处于情况 a 时，窗口位于图像特征相对均匀的灰度平坦区域，此时窗口沿任意方向变化，窗口内的像素值变化较小。情况 b 时，窗口位于图像边缘区域，边缘是图像中灰度值变化显著的地方，而在窗口垂直于边缘的方向上，窗口内的像素值变化较大，而在平行于边缘的方向上，窗口内的像素值变化较小。情况 c 时，窗口位于角点区域，窗口内的像素值在不同方向上都会有明显的变化，因此，通过分析窗口灰度值的变化，可以判断出角点。

将上述思想转换为数学形式，即将局部窗口向各个方向移动 $(u, v)$ 并计算灰度值差异的总和，表达式为：

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2 \quad (67)$$

式中：

$I(x, y)$ —为局部窗口的图像灰度；

$I(x+u, y+v)$ —为平移后的图像灰度；

$w(x, y)$ —为滑动窗口的加权函数。

当检测出是角点时， $E(u, v)$ 的值是最大。利用一阶泰勒公式展开可得：

$$I(x+u, y+v) = I(x, y) + I_x u + I_y v \quad (68)$$

式中：

$I_x$  和  $I_y$ —沿  $x$  和  $y$  方向的导数，可以用 Sobel 算子计算。



由公式(67)和公式(68)可得:

$$\begin{aligned}
 E(u, v) &= \sum_{x, y} w(x, y) [I_x^2 u^2 + 2I_x I_y uv + I_y^2 v^2] \\
 &= \sum_{x, y} w(x, y) [u, v] \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\
 &= [u, v] \sum_{x, y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\
 &= [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}
 \end{aligned} \tag{69}$$

从公式(69)中可知  $M$  矩阵决定了  $E(u, v)$  的取值,  $M$  矩阵是  $I_x$  和  $I_y$  的二次函数, 可以表示成椭圆形, 设椭圆的长短半轴由特征值  $\lambda_1$  和  $\lambda_2$  决定,  $\lambda_1$  和  $\lambda_2$  的大小关系和图像的结构有如下关系:

当  $\lambda_1$  和  $\lambda_2$  都很小时, 图像窗口在所有方向上移动都无明显的灰度变化, 代表了图像中的平面区域。

当  $\lambda_1$  和  $\lambda_2$  数值相差较大, 图像窗口在某一方向上的灰度变化较小, 正交方向灰度变化较大, 代表了图像中的边缘区域。

当  $\lambda_1$  和  $\lambda_2$  都很大时, 图像窗口在所有方向上移动都产生明显的灰度变化, 代表了图像中的角点。

所以只需要确定  $M$  矩阵的两个特征值的大小, 就能够确定是否为角点。为了提高算法效率, Harris 和 Stephens 利用矩阵的性质, 通过计算一个角点响应值  $R$  来判断角点,  $R$  的计算公式为:

$$R = \det M - \alpha (\text{trace} M)^2 \tag{70}$$

式中:

$\det M$  — 矩阵  $M$  的行列式;

$\text{trace} M$  — 矩阵  $M$  的迹;

$\alpha$  — 常数, 取值范围为 0.04~0.06。

$\det M$  和  $\text{trace} M$  可以表达为:

$$\det M = \lambda_1 \lambda_2 \tag{71}$$

$$\text{trace} M = \lambda_1 + \lambda_2 \tag{72}$$

因此, 当  $R$  为大数值的正数时是角点, 当  $R$  为大数值的负数时是边界, 当  $R$  为小数值时是平坦区域。使用 Harris 角点检测算法对图像轮廓进行提取的效果图如图 23 所示。

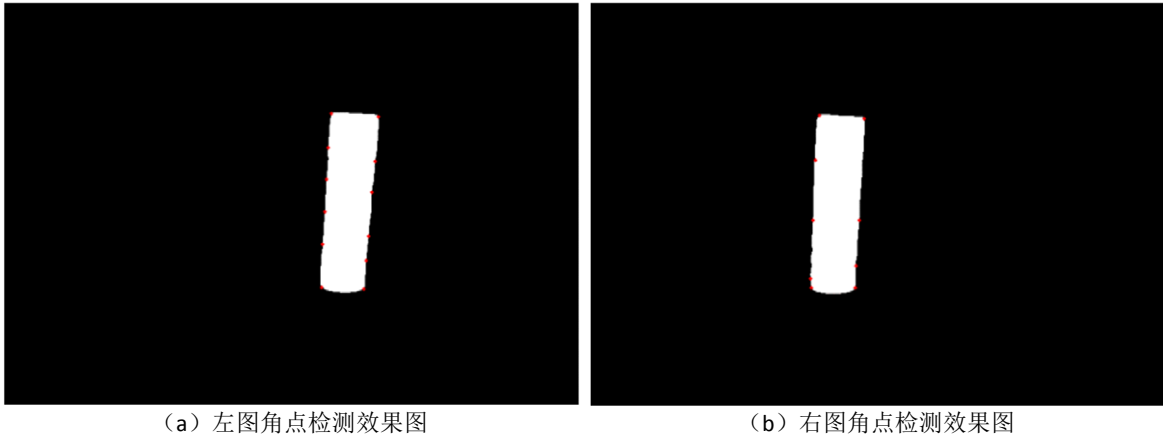


图 23 Harris 角点检测效果图

在进行角点检测后，不止得到了需要的特征点，也包含了不需要的特征点，因此需要对检测的角点进行过滤。图像中包含的特征点可以划分为两类，一种是轮廓边缘区域的特征点，一种为轮廓交点的特征点，而本文所需的特征点为轮廓交点的特征点，考虑轮廓边缘角点的特征，以角点为圆心，半径  $R$  作圆，统计圆形边缘中黑白像素的个数，即可排除边缘区域的角点。进行角点过滤后的图像如下所示：

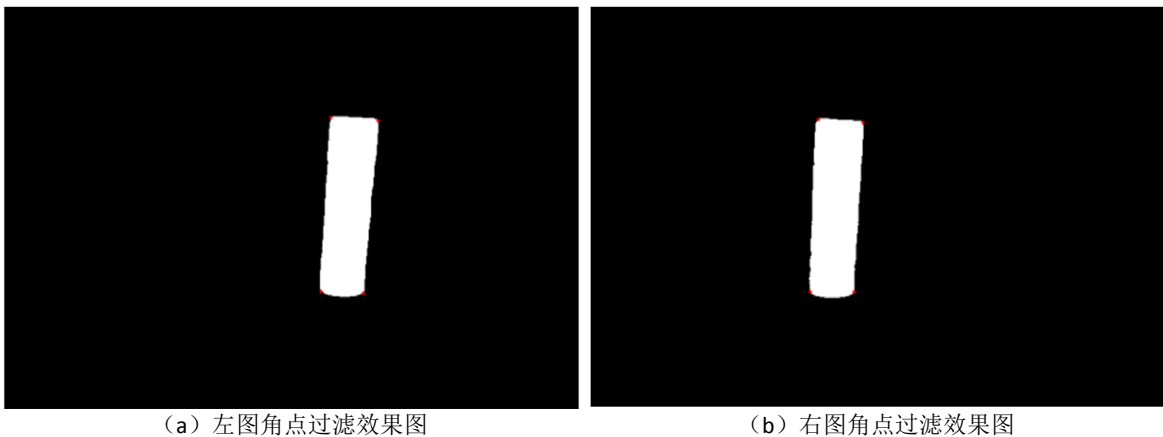


图 24 角点过滤效果图

#### 4.2.4 目标物体尺寸测量

双目测距原理是基于人类双眼视觉的工作原理而设计的一种测距方法。基于双目视觉的物体尺寸测量，通过利用两个摄像机模拟人类的双眼观察场景。首先，利用具有固定边长的标靶图像进行双目相机的标定，获取相机的内外参数，然后根据相机的内外参数对测量物体图像进行校正，对校正后的图像立体匹配，把同一场景在左右视图上的像素点匹配起来，得到了视差图。将视差图通过相似三角形原理可获取目标的三维信息。理想双目相机成像模型如下图所示。

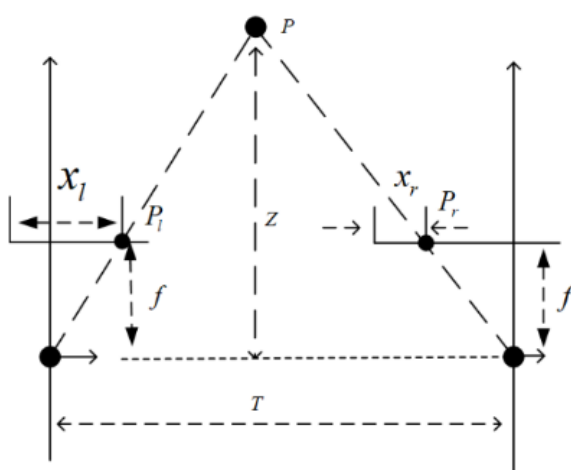


图 25 理想相机成像模型

图 25 中， $T$  为基线，是两个相机的投影中心的距离，在双目测距中基线是一个重要的参数，它的大小直接影响到测量的距离精度，一般情况下，基线越长，测量的距离范围也会相应增大。 $p_l$  和  $p_r$  分别为三维空间任意一点  $P$  在左相机和右相机的成像点，点  $P$  在左右相机的视差可以定义如下：

$$d = |x_l - x_r| \quad (73)$$

式中：

$d$ ——视差，为两个相机的成像点的距离；

$x_l$ ， $x_r$ ——分别是左右相机成像点到左成像面的距离。

从图 25 中可以看出，目标像素点与两个相机的成像平面点构成的三角与目标像素点与相机构成的三角是相似的，则可以计算出像素点距离相机的深度  $z$ ，数学表达式为：

$$Z = \frac{fT}{x_l - x_r} \quad (74)$$

式中：

$f$ ——相机的焦距；

$Z$ ——空间点  $p$  到两个相机基线的距离；

$x_l$ ， $y_l$ ——左相机的二维坐标。

根据三角形相似性原理可得：

$$\frac{Z}{f} = \frac{T}{d} = \frac{X}{x_l} = \frac{Y}{y_l} \quad (75)$$

一般计算像素的三维坐标，都是先求出深度  $z$ ，然后根据公式（75），可以计算出像素的  $X$  和  $Y$  轴坐标，数学表达式定义如下：

$$\begin{cases} Z = \frac{fT}{x_l - x_r} \\ X = \frac{Zx_l}{f} \\ Y = \frac{Zy_l}{f} \end{cases} \quad (76)$$

对于空间中任意一点  $P$ ，都可以计算出相应的视  $d$ ，视差  $d$  根据  $P$  点位置的不同而大小不同，即可形成视差图，再根据相机的参数，就可以得到该点的三维坐标。在进行物体尺寸测量时，只需要得出待测物体轮廓的两个特征点的三维坐标，即可完成物体尺寸测量。使用第三章改进的立体匹配算法对校正后的双目图像进行立体匹配，得到视差图，如图 26a 所示

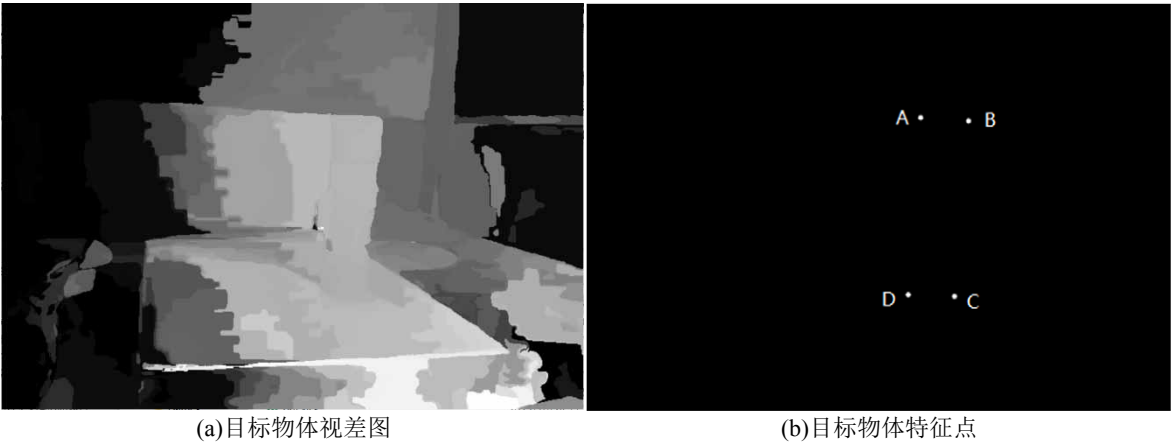


图 26 目标物体特征图

利用相机参数和视差图，根据三角测量原理，能够有效计算特征点的三维坐标。根据图 24 所示特征点信息，以左上角特征点为起始点，顺时针依次命名为点 A、B、C 和 D，目标特征点位置如图 26b 所示。根据双目视觉系统的测量原理，求出特征点二维坐标与对应空间三维坐标数据，特征点坐标信息如表 9 所示。为验证目标轮廓提取的边缘搜索优化的有效性，同时对没有进行边缘优化搜索的轮廓信息进行角点检测和过滤，表 10 中数据即为非进行边缘搜索优化的分割轮廓提取的特征点坐标。

表 9 特征点坐标值及三维坐标

匹配点	二维坐标点		三维坐标点		
	x 轴	y 轴	x 轴	y 轴	z 轴
A	405	141	54.42	-51.95	308.96
B	463	145	84.19	-50.35	311.43
C	446	357	77.45	58.25	319.09
D	390	355	48.04	56.75	316.50

表 10 非边缘搜索优化特征点坐标值及三维坐标

匹配点	二维坐标点		三维坐标点		
	x 轴	y 轴	x 轴	y 轴	z 轴
A	405	140	54.42	-51.46	308.96
B	462	144	83.02	-50.49	308.96
C	444	353	76.20	57.31	313.95
D	391	354	49.32	56.20	316.83

对比图 26b 和表 7，在 X 轴方向上依次是 A、D、B 和 C,在 Y 轴方向上依次是 B、A、C 和 D，图和表中数据符合特征点特征。表中 Z 轴数据无异常值，表明系统在特征点区域获取了较为准确的视差值。使用表中三维坐标数据，根据欧式距离公式，可以计算出坐标点间的真实距离。实验中，被测物体尺寸顶端尺寸为 30.1mm，底端尺寸为 29.7mm，高度为 110.3mm，目标物体的真实尺寸采用游标卡尺获得，真实尺寸测量图如图 27 所示。改进算法所得物体尺寸和非边缘优化的物体尺寸如表 11 所示。



图 27 目标物体真实尺寸测量图

表 11 测量结果与测量物体的真实尺寸对比

实际尺寸(mm)		AB	AD	BC	CD
		30.10	110.30	110.30	29.70
边缘优化	测量尺寸(mm)	29.91	109.15	109.08	29.56
	绝对误差(mm)	0.19	1.15	0.92	0.14
	相对误差(%)	0.63	1.04	1.11	0.40
	测量尺寸(mm)	28.62	108.07	108.13	27.10
未边缘优化	绝对误差(mm)	1.48	2.23	2.17	2.60
	相对误差(%)	4.92	2.02	1.97	8.75

从表 11 中可以得出，对物体尺寸测量的最低误差为 0.4%，最高误差为 1.11%，平均误差为 0.80%。通过对比轮廓信息边缘优化的物体尺寸测量结果，可以看出，对目标物体轮廓边缘优化后，得到的特征点坐标更加精准。

4.3 物体尺寸测量的实验与分析

为验证本文算法及改进的可靠性与有效性，通过物体进行多次测量，验证本文算法的可靠性，通过对比经典立体匹配算法，验证本文改进算法的有效性。测量物体为螺丝的螺纹区域，测量数据为螺纹顶端为 12.5mm,螺纹底端为 12.8mm，高度为 75.5mm。

4.3.1 物体距离测量实验

在通过双目视觉进行物体尺寸测量时，当物体与摄像机距离增加时，物体在图像中的大小会减小，反之亦然。这是由于物体与摄像机距离增加会导致视角变小，在图像上表现为物体变小。而物体与摄像距离越近，测量结果中物体的尺寸越大。

为了验证本文算法在不同距离的测量准确性，设置多倍目标物体高度的拍摄距离进行实验。在实验过程中，当目标物体与相机的距离小于 3 倍目标物体高度时，视差图中的目标物体存在大量误差，无法完成目标物体尺寸测量，这是由于目标物体与相机距离过近时，镜头的畸变效应会更显著地影响图像的质量，物体在图像上所占的像素数量相对较多，会导致目标物体的纹理信息不明显或被混淆，使得双目视觉算法难以稳定地提取到准确的视差信息。实验以 80mm 为基准，分别实验 3~6 倍距离的测量，不同距离目标物体特征点坐标如表 12 所示，测量数据及误差表如表 13 所示。

表 12 不同距离目标物体特征点坐标

特征点		二维坐标点		三维坐标点		
		x 轴	y 轴	x 轴	y 轴	z 轴
3 倍	A	472	212	68.90	-12.78	241.80
	B	504	212	81.46	-12.78	241.80
	C	507	399	84.73	62.18	247.76
	D	490	400	72.12	62.80	247.76
4 倍	A	432	178	70.78	-34.77	321.73
	B	456	178	83.31	-34.77	321.73
	C	458	319	87.24	40.21	332.73
	D	430	319	74.41	40.21	332.73

表 12（续）

特征点		二维坐标点		三维坐标点		
		x 轴	y 轴	x 轴	y 轴	z 轴
5 倍	A	371	266	44.42	12.78	402.28
	B	392	268	56.94	13.98	402.28
	C	391	396	55.82	89.46	393.83
	D	369	391	42.82	88.51	393.83
6 倍	A	417	223	95.2	-17.04	486.62
	B	432	223	107.05	-17.04	486.62
	C	425	315	99.04	54.3	474.75
	D	415	315	92.46	54.97	480.61

表 13 不同距离目标物体测量对比

实际尺寸(mm)		AB	AD	BC	CD
		12.50	75.50	75.50	12.80
3 倍	测量尺寸(mm)	12.56	75.29	75.89	12.64
	绝对误差(mm)	0.06	0.21	0.39	0.16
	相对误差(%)	0.48	0.28	0.52	1.25
4 倍	测量尺寸(mm)	12.53	75.87	75.88	12.83
	绝对误差(mm)	0.03	0.37	0.38	0.03
	相对误差(%)	0.24	0.49	0.50	0.23
5 倍	测量尺寸(mm)	12.58	75.53	75.79	13.03
	绝对误差(mm)	0.08	0.03	0.29	0.23
	相对误差(%)	0.64	0.04	0.38	1.80
6 倍	测量尺寸(mm)	11.85	72.76	72.31	8.84
	绝对误差(mm)	0.65	2.74	3.19	3.96
	相对误差(%)	5.20	3.63	4.23	30.94

从表 12 中可以看出，目标物体距离相机 3~6 倍距离时，得到的 z 轴信息与真实距离相近，证明了本文改进立体匹配算法的稳定性。表 13 中 3~5 倍的测量相对误差为 0.03%~1.80%，表明测量距离为目标物体长度 3~5 倍时可以得到较为准确的测量值。当目标物体与相机的距离达到目标物体 6 倍时，由于相机分辨率较低，且目标物体占图像内容较少，导致测量准确性下降。

4.3.2 算法可靠性验证

为了验证本文算法的可靠性，进行了对同一物体的多次测量实验。在实验中，

对同一物体拍摄三组图像，分别称为实验 a、b 和 c，通过对这些图像进行校正、立体匹配、特征点提取，得到校正图像、视差图和特征点位置信息，如图 28 所示。实验中特征点三维坐标和所得的物体尺寸信息，在表 14 和表 15 中描述。

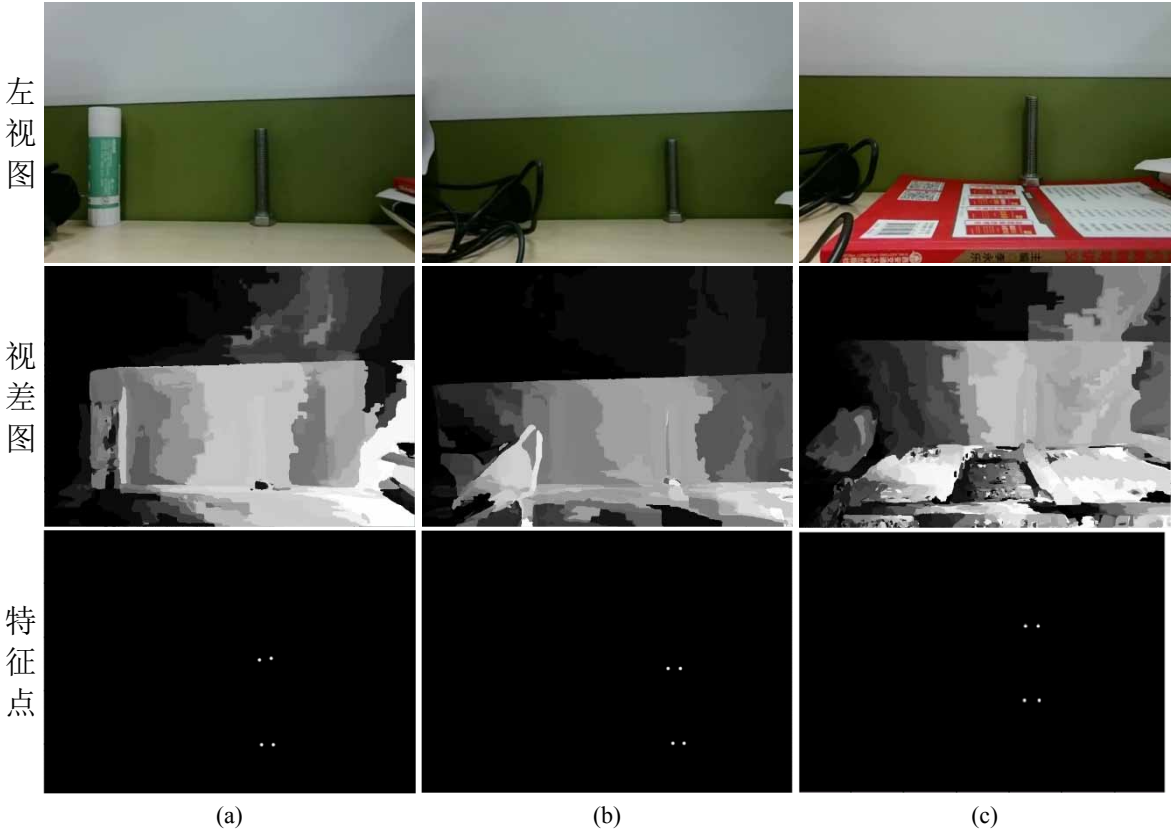


图 28 实验测量图

表 14 实验测量坐标数据表

特征点		二维坐标点		三维坐标点		
		x 轴	y 轴	x 轴	y 轴	z 轴
a	A	399	242	46.94	-1.17	282.1
	B	425	241	59.28	-1.64	284.16
	C	432	406	62.06	73.94	282.1
	D	407	406	49.18	72.86	283.85
b	A	463	260	90.72	8.41	335.6
	B	486	260	103.25	8.41	335.6
	C	493	400	105.25	83.26	329.91
	D	472	400	93.21	82.56	327.14
c	A	432	178	70.78	-34.77	321.73
	B	456	178	83.31	-34.77	321.73
	C	458	319	87.24	40.21	332.73
	D	430	319	74.41	40.21	332.73



表 15 实验测量值与真实值对比

实际尺寸(mm)		AB	AD	BC	CD
		12.5	75.5	75.5	12.8
a	测量尺寸(mm)	12.52	74.08	74.66	13
	绝对误差(mm)	0.02	1.42	0.84	0.2
	相对误差(%)	1.6	1.88	1.11	1.56
b	测量尺寸(mm)	12.53	74.67	75.09	12.37
	绝对误差(mm)	0.03	0.83	0.41	0.43
	相对误差(%)	0.24	1.1	0.54	3.36
c	测量尺寸(mm)	12.53	75.87	75.88	12.83
	绝对误差(mm)	0.03	0.37	0.38	0.03
	相对误差(%)	0.24	0.49	0.5	0.23

从表 14 和表 15 中可以看出，测量结果的相对误差在 0.02%到 3.36%之间，对目标物体多次测量的平均误差为 1.07%，证明了本文方法可以满足日常场景中的物体尺寸测量，测量结果在合理范围内并与实际尺寸相当接近。

4.3.3 算法有效性验证

为验证本文改进算法的有效性，使用不同的立体匹配算法对同一场景下的同一物体进行测量，并与本文算法进行对比。实验采用 4.3.1 的实验 a 场景进行对比，通过对比本文改进算法和本文基准模型 SGM 和 ST，得到的视差图如图 28 所示。

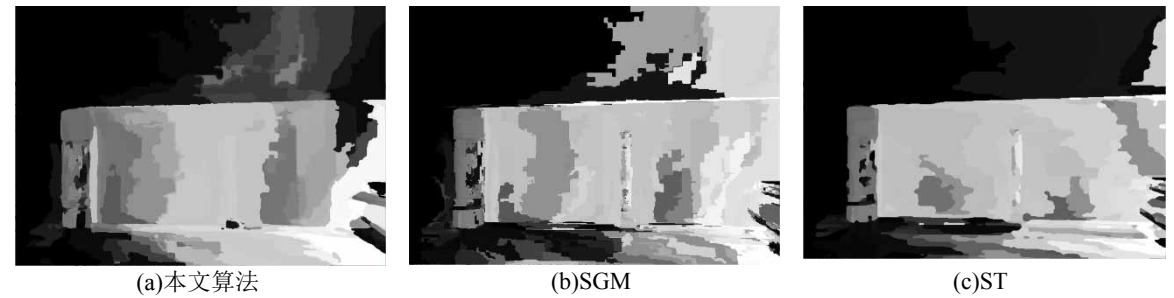


图 28 实验视差图对比

表 16 特征点三维坐标表

算法坐标		三维坐标点		
		x 轴	y 轴	z 轴
本文算法	A	46.94	-1.17	282.10
	B	59.28	-1.64	284.16
	C	62.06	73.94	282.10
	D	49.18	72.86	283.85

表 16 (续)

算法坐标		三维坐标点		
		x 轴	y 轴	z 轴
SGM	A	43.39	-0.7	276.09
	B	58.39	-1.17	282.1
	C	60.7	72.5	280.07
	D	48.39	70.92	335.2
ST	A	46.27	-1.16	278.07
	B	59.72	-3.52	286.25
	C	61.61	73.41	280.07
	D	48.5	70.86	270.34

表 16 中, SGM 算法计算得出的三维坐标中 D 点的 z 轴坐标为 335.20, 与其他特征点相差较大, 因此使用临近像素的 z 轴作为特征点坐标值, 其值为 272.23, 用于测量物体尺寸。所得测量结果如表 17 所示。

表 17 测量结果对比

实际尺寸(mm)		AB	AD	BC	CD
		12.50	75.50	75.50	12.80
本文算法	测量尺寸(mm)	12.52	74.08	74.66	13.00
	绝对误差(mm)	0.02	1.42	0.84	0.20
	相对误差(%)	1.60	1.88	1.11	1.56
SGM	测量尺寸(mm)	13.43	73.74	71.75	14.67
	绝对误差(mm)	0.93	1.76	3.75	1.87
	相对误差(%)	7.44	2.33	4.97	14.61
ST	测量尺寸(mm)	15.91	72.47	76.88	16.52
	绝对误差(mm)	3.41	3.03	1.38	3.72
	相对误差(%)	27.28	4.01	1.83	29.06

从图 28 中可以看出, 本文所提算法生成的视差图更为平滑, 因此获取的三维坐标更为准确, 有助于更精确的计算物体的尺寸。而基准模型 SGM 和 ST 在测量物体区域的视差不连续, 会导致目标物体特征点的三维坐标不准确。从表 15 中可以看出, 基准模型的测量相对误差最小为 0.93%, 最高相对误差达到了 29.06%。在本次实验中, 虽然本文算法的测量结果不一定时最优的, 但是整体更稳定, 误差幅度小, 证明了本文改进算法的有效性。

#### 4.3.4 物体尺寸测量误差分析

本文针对物体尺寸测量过程中可能出现的误差进行分析, 主要包含以下几点:

(1) 相机标定误差：在相机标定过程中，选择标定板的质量和特征以及拍摄角度都会对标定精度产生影响。标定板的质量和特征能够提供足够准确的控制点，以便于进行准确的匹配和计算。拍摄角度的选择要尽量保证标定板在不同位置和姿态下都能够被充分观测到，以提高标定的全局精度。

(2) 背景与纹理复杂度：在实际的物体尺寸测量中，场景图像可能包含复杂的背景信息和纹理信息，而测量需要的是待测物体的轮廓信息。因此，在轮廓分割的过程中，场景图像的背景模板、待测物体的纹理复杂度、分割效果的优劣以及特征点的准确度都会对物体尺寸测量的精度产生影响。

(3) 立体匹配误差：立体匹配是双目视觉测量系统中的一个关键步骤。在实际的立体匹配过程中，会受到多个因素的影响，包括光照强度、噪声和场景图像的复杂程度等。为了提高测量精度，需要采用合适的方法来降低这些影响，例如采用合适的光照条件、进行噪声抑制和提取更多的场景特征等。

(4) 视差范围变换：当被测物体与相机的距离发生变化时，左右图像中对应点的视差会发生变化。根据三角测量原理，视差与物体与相机之间的距离成反比关系。如果物体与相机的距离增大，视差范围会变小；反之，如果距离减小，视差范围会变大。因此，当物体与相机的距离发生微小变化时，视差值会有较大的变化。为了提高测量精度，需要采取措施来减小视差范围内的变化对测量结果的影响。

#### 4.4 本章小结

本章针对双目视觉测量整体流程进行研究。首先，根据双目相机标定获取的相机内外参数对双目图像进行校正，以消除图像畸变和对齐双目视角。然后，采用第三章介绍的立体匹配算法对校正后的图像进行立体匹配，以获取物体的视差信息。在双目物体尺寸测量流程中，针对图像轮廓获取和特征点提取过程，利用 Segment Anything 的图像分割算法对目标物体进行初步分割，然后提出边缘搜索优化方法，获得较为准确的物体轮廓信息。对得到的目标物体轮廓采用 Harris 角点检测算法提取特征点，并通过过滤无意义特征点来提高特征点的质量。最后，基于视差信息和相机参数，计算特征点的三维坐标，利用三角测量原理完成对物体尺寸的测量。通过多次测量同一物体，验证本文算法的可靠性，通过对比本文改进算法与基准模型的测量结果，验证了本文改进的有效性。

## 结 论

随着计算机视觉方面算法的不断提高以及相关硬件设备的发展，基于双目视觉的尺寸测量技术广泛应用于生产生活中。本课题以日常生活场景中的物体为测量目标，对相机标定、极线校正、立体匹配、特征点提取等相关算法进行了研究，最终实现了对目标物体尺寸信息的测量。本文所做的主要工作总结如下：

（1）通过描述双目视觉成像模型的四个坐标系间的变换过程，反应空间坐标与图像像素点间的对应关系。介绍了张氏标定法原理，并选取棋盘格标靶图像基于 OpenCV 实现相机标定，获取相机的内外参数和畸变系数，最后基于 Bouguet 极线校正算法对双目图像进行校正，消除两个相机位置和视角不同导致的误差。

（2）在立体匹配步骤中，本文采用半全局立体匹配算法思想。在代价计算阶段，考虑传统匹配代价计算方法的优缺点，提出了自适应 Census 变换与颜色、梯度信息结合的融合匹配代价，得到了可靠性更好的匹配代价。在代价聚合阶段，针对物体尺寸测量过程中对视差图边缘区域的要求，采用具有边缘保持的最小生成树聚合算法，并引入特征金字塔和扫描线优化，增强了图像边缘的视差精确性和减少误匹配。

（3）根据相机标定、立体校正、立体匹配等研究，完成物体尺寸测量。在测量过程中需要确定目标物体的特征点，本文考虑图像分割技术的优缺点，采用基于神经网络的 Segment Anything 模型完成对目标物体初步分割，然后提出边缘搜索优化方法，调高了图像分割的精确性，获取目标物体轮廓信息，最后使用 Harris 角点检测算法获取角点，并根据特征点特征进行筛选。最后根据标定的相机参数、立体匹配的视差图和目标物体的特征点，完成对目标物体的尺寸测量。并通过多次测量同一物体，验证本文算法的可靠性，通过对比本文改进算法与基准模型的测量结果，验证了本文改进的有效性。

## 参考文献

- [1] Charan A, Karthik C C, Komal P. The Future of Machine Vision in Industries-A systematic review[J]. IOP Conference Series: Materials Science and Engineering, 2022, 1224(1).
- [2] Genevieve D. The Rise of Machine Vision[J]. Quality, 2022, 61(1): 37-39.
- [3] 王正家,景嘉宝,王思宇. 基于双目视觉的车辆外廓尺寸测量方法[J]. 电子测量技术,2023,46(12):150-156.
- [4] 王昊,刘雍翡. 基于双目视觉计算的车辆跟驰状态实时感知系统[J]. 中国公路学报,2019,32(12):88-97+105.
- [5] Martd C. A Computational Investigation into the Human Representation and Processing of Visual Information[J]. Free-man, San Francisco, CA, 1983, 27(1).
- [6] Jung B, Sukhatme G S. Real-time Motion Tracking from a Mobile Robot[J]. International Journal of Social Robotics, 2010, 2(1): 63-78.
- [7] Okada K , Inaba M, Inoue H. Integration of Real-Time Binocular Stereo Vision and Whole Body Information for Dynamic Walking Navigation of Humanoid Robot[C]//IEEE International Conference on Multisensor Fusion & Integration for Intelligent Systems. IEEE, 2003.
- [8] 张亚斌. 基于 ROS 的轮式移动机器人行人目标视觉感知与跟随系统研究[D]. 中国矿业大学,2019.
- [9] 郭海洲. 基于双目视觉的室内目标检测与测距系统研究[D]. 广西师范大学,2022.
- [10] Chen X, Lu C. An Experimental System Development for Head Posture Estimation Based on 3-D Images Measurement [J]. Journal of Computer & Communications, 2016, 02(2): 57-63.
- [11] Xie Q W, Liu R R, Sun Z, Pei S S, et al. A Flexible Free-space Detection System Based on Stereo Vision[J]. Neurocomputing, 2022, 485: 252-262.
- [12] 魏新华,张敏,刘青山等. 基于双目视觉的田间作物高度和收割边界信息提取[J]. 农业机械学报,2022,53(03):225-233.
- [13] 蒋焕煜,彭永石,申川等. 基于双目立体视觉技术的成熟番茄识别与定位[J]. 农业工程学报,2008(08):279-283.
- [14] 陈红. 基于双目立体视觉几何量的高精度测量技术研究[D]. 贵州民族大学,2022. DOI:10.27807/d.cnki.cgzmz.2022.000360.
- [15] 张宗华,刘巍,刘国栋等. 三维视觉测量技术及应用进展[J]. 中国图象图形学报,2021,26(06):1483-1502.

- [16] Chen Y Z, Zhou F Q. Pose Measurement Approach Based on Two-stage Binocular Vision for Docking Large Components [J]. Measurement Science and Technology, 2020, 31(12).
- [17] Zhang Y W, Gu J. A Shape Reconstruction and Measurement Method for Spherical Hedges Using Binocular Vision[J]. Frontiers In Plant Science, 2022.
- [18] 宋超群,许四祥,杨宇等. 基于改进 FAST 和 BRIEF 的双目视觉测量方法[J]. 激光与光电子学进展,2022,59(08):173-180.
- [19] Chen H, Wang D Q, Chen Y Q. Research on the Influence of Calibration Image on Reprojection Error[C]; 2021 International Conference on Big Data Engineering and Education(BDEE), 2021: 60-66.
- [20] 郑冬,冯鹏,龙邹荣等. 面向多因素工况下的相机标定精度综述[J]. 国外电子测量技术,2020,39(08):109-116. DOI:10.19652/j.cnki.femt.2002099.
- [21] Zhang Z. A Flexible New Technique for Camera Calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [22] Datta A , Kim J , Kanade T. Accurate Camera Calibration using Iterative Refinement of Control Points[J]. workshop on visual surveillance, 2009.
- [23] 陈炎,杨丽丽,王振鹏. 双目视觉的匹配算法综述[J]. 图学学报,2020,41(05):702-708.
- [24] Lu B, Sun L, Yu L, et al. An Improved Graph Cut Algorithm in Stereo Matching[J]. Displays, 2021, 69(1):102052.
- [25] Li Y S, Zheng S, Wang X, et al. An Efficient Photogrammetric Stereo Matching Method for High Resolution Images[J]. Computers & Geosciences, 2016, 97(dec.): 58-66.
- [26] 石岩青,常彩霞,刘小红等. 面阵相机内外参数标定方法及进展[J]. 激光与光电子学进展,2021,58(24): 2400001.
- [27] 刘丽伟,赵子君,朱绪康等. 面向温度变化的相机光学系统参数模型研究[J]. 计算机仿真,2021,38(3): 180-185.
- [28] 郑太雄,黄帅,李永福等. 基于视觉的三维重建关键技术研究综述[J]. 自动化学报,2020,46(4):631-652.
- [29] 朱嘉,李醒飞,徐颖欣. 摄像机的一种主动视觉标定方法[J]. 光学学报,2010, 30(5): 1297-1303.
- [30] 张美航,张华,鄢威. 结构光自标定方法综述[J]. 传感器与微系统,2023,42(02):1-4+12. DOI:10.13873/J.1000-9787(2023)02-0001-04.
- [31] 黄及远,李敏,谢兵兵等. 双目视觉关键技术研究综述[J]. 制造业自动化,2023,45(05):166-171.
- [32] 曹毅,包祥威,吴翔. 基于重排序 Census 变换的半全局立体匹配算法[J]. 电子测量技术,2021,44(24):40-46. DOI:10.19651/j.cnki.emt.2107931.

- [33] Zabih R, Woodfill J. Non-parametric Local Transforms for Computing Visual Correspondence[M]. 2006.
- [34] 李涵,黄妙华. 自适应区域划分立体匹配算法[J]. 激光与光电子学进展,2023,60(10):172-182.
- [35] 杨科,刘凯. 多测度融合的树形滤波立体匹配算法[J]. 计算机工程与设计,2021,42(08):2277-2283. DOI:10.16208/j.issn1000-7024.2021.08.025.
- [36] Zhang K, Fang Y, Min D, et al. Cross-Scale Cost Aggregation for Stereo Matching[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(5):965-976. DOI:10.1109/TCSVT. 2015. 2513663.
- [37] 潘卫华,杜旭. 基于改进自适应权重的三边滤波立体匹配算法[J]. 计算机工程与应用,2020,56(18):186-192.
- [38] 刘旺,陈焱涛,刘芳. 基于改进 Census 代价和优化引导滤波立体匹配算法[J]. 电子测量技术,2023,46(07):80-87. DOI:10.19651/j.cnki.emt.2211055.
- [39] Yang Q X, "A Non-local Cost Aggregation Method for Stereo Matching," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 2012, pp. 1402-1409, DOI: 10.1109/CVPR.2012.6247827.
- [40] Mei X, Sun X, Dong W, et al. Segment-Tree based Cost Aggregation for Stereo Matching[J]. PROC CVPR IEEE, 2013.
- [41] 乔景慧,韩玉明,张啸涵. 基于改进 Census 变换的鲁棒立体匹配算法[J]. 计量学报,2023,44(05):694-700.
- [42] 王道累,韩洋. 基于改进匹配代价和均值分割的最小生成树立体匹配算法[J]. 激光与光电子学进展,2023,60(04):333-342.
- [43] 虞文杰,叶嵩,郭毓等. 基于改进 Census 变换与多特征融合的立体匹配算法[J]. 激光与光电子学进展,2022,59(08):157-163.
- [44] Alexander K, Eric M, Nikhila R, et al. Segment Anything. arXiv:2304.02643. 2023. DOI: 10.48550/arXiv.2304.02643