

東 北 大 學

硕士学位论文开题报告及论文工作计划书

学 号： 2170871

姓 名： 张佳预

导 师： 贾同

学科类别： ☐工学 ☒全日制专业学位

学科/工程领域： 控制工程

所属学院： 信息科学与工程学院

研究方向： 双目结构光三维重建

拟选题目： 基于无监督学习的跨光谱主动
立体匹配方法与系统研究

选题时间： 2022 年 10 月 1 日

东北大学研究生院

2022 年 10 月 27 日

填表说明

- 1、本表应在导师指导下如实填写。
- 2、学生在开题答辩前一周，将该材料交到所在学院、研究所。
- 3、按有关规定，没有完成开题报告的学生不能申请论文答辩。
- 4、全文正文均用小四号宋体，单倍行距，段前段后间距为 0，如果页数不够，可以整页扩页，其他格式要求参见《东北大学硕、博士学位论文格式》。

一、前期工作基础（本节可以整页扩页）

课程学习及选题开题阶段，在导师指导下从事研究工作总结（不少于 2000 字）

开题阶段，在导师指导下研究了单目结构光三维测量技术和双目结构光三维测量技术，并取得了一定成果。单目结构光三维测量技术的研究内容，包括投影仪-相机系统标定和基于相移法的结构光三维测量技术；双目结构光三维测量技术的研究内容，包括双目系统标定与校正和基于半全局方法的双目散斑结构光三维测量技术。下面对前期工作和成果进行简单的阐述。

1.单目结构光三维重建技术

1.1 投影仪-相机系统标定技术

1.1.1 单目相机标定

1.1.1.1 引言

相机标定已经研究多年，标定的算法可以分为基于摄影测量的标定和自标定。其中，应用最为广泛的还是张正友标定法，简称张氏标定法。这是一种简单灵活、高鲁棒性、低成本的相机标定算法。仅需要一台相机和一块平面标定板构建相机标定系统，在标定过程中，相机拍摄多个角度下（至少两个角度，推荐 10 到 20 个角度）的标定板图像，即可对相机的内外参数进行标定。下面对张氏标定法的标定流程进行概括。

1.1.1.2 标定流程

单目相机张氏标定法的标定流程如下：

- （1）准备一个棋盘格标定板，打印或定制皆可，建议定制，因为精度更高；
- （2）移动相机或标定板，相机拍摄多个位置标定板的图像，建议拍摄 10 到 20 张图像；
- （3）检测图像的角点；
- （4）线性方程估计相机的内参封闭解；
- （5）求解线性最小二乘问题，估计畸变系数；
- （6）使用 LM 算法最小化重投影误差，优化相机内参数。

1.1.2 投影仪标定

1.1.2.1 引言

单目结构光系统包括一台相机和一台投影仪，结构光三维重建的原理是投射编码结构光图案到场景中，然后相机拍摄经过物体高度调制的变形结构光图像，通过计算机解码得到投影图案和结构光图像的特征点对应，最后基于三角测量原理得到三维点云。三角测量原理就需要用到相机和投影仪的参数。所以必须对投影仪和相机系统进行标定，要标定的参数包括相机内参、投影仪内参、投影仪-相机的相对外参。

因为投影仪无法对 3D 物体成像，无法像相机直接获得投影图案和三维空间中特征点的对应关系，所以投影仪标定的核心问题就是如何找到投影图案 2D 点和 3D 投影点的对应关系。

虽然投影仪无法对场景成像，但它的成像模型跟相机成像模型是一致的，可以将投影仪视作逆相机模型。这样，只要找到对应关系，就可以利用张氏标定法对投影仪进行标定了。下面介绍投射格雷码图案的标定方法。

1.1.2.2 投影格雷码标定

投影格雷码的标定流程如下：

- （1）获取相机拍摄的标定板图片，保持标定板不动，依序投影结构光编码到标定板上；
- （2）找到标定板图片中的特征点坐标；
- （3）对相机拍摄到的投影结构光进行解码；
- （4）对每个找到的标定板图片中的特征点坐标求局部的单应性矩阵，并使用局部单应性矩阵计算出特征点在投影仪图片上的坐标；
- （5）改变标定板位姿，重复步骤一到步骤四；
- （6）重复步骤五至少三次，并利用找到的特征点在相机和投影仪中的坐标分别计算出相

机内参矩阵和畸变系数，投影仪内参矩阵和畸变系数；

(7) 利用找到的特征点在相机和投影仪中的坐标值进行双目标定计算，得到相机和投影仪之间的外参矩阵：旋转矩阵 R 和平移矩阵 T 。

1.2.基于相移法的结构光三维测量技术

1.2.1 引言

相移法是结构光三维测量技术的一种，它的原理是通过投射多幅具有一定相位差的条纹光栅至物体表面，通过变形的光栅条纹图来计算相位，最后通过相位与高度的映射关系来得到物体的轮廓信息，进而实现物体的三维重建。常用的几种算法有三步相移法、四步相移法，一般来说步数越小效率越高，但大步数的相移法能减小因为光照不均匀而产生的相位误差，因而其测量精度受投影条纹光栅影响。同时，由于需要利用反正切函数关系来求解出相位分布，包裹相位的展开也会影响测量精度。随着近年来 DLP 投影技术的发展，相移可以通过 PC 机编程实现准确相移，测量精度相比于传统的光栅机械移动大大提升。

1.2.2 相移法三维重建过程

首先将计算机生成的一系列标准正弦条纹图由投影仪投射到被测物体表面，初始条纹图收到被测物体表面的调制发生变形，之后对相机采集到的变形条纹图进行折叠相位和展开相位的计算，最后由系统标定得到的“相位-深度”的关系反求物体的深度信息，从而恢复的三维坐标。

1.2.3 存在的问题及其解决办法

(1) 投影仪非线性问题：投影仪存在 Gamma 非线性，正弦光栅编码在经过投影仪和相机拍摄后，真正获得的条纹并不是正弦条纹。投影仪的非线性会导致深度图出现水波纹现象。常见的解决办法是二值条纹离焦、双目结构光的方法。

(2) 包裹相位展开问题：由于包裹相位（折叠相位）的获取是通过四象限反正切函数获得的，相位局限在 $[-\pi, \pi]$ 范围内，最终的绝对相位需要通过相位展开（相位解包裹）的方式获得。由于环境光、噪声、阴影等因素的影响，很容易造成相位展开的错误。因此，如何准确的进行包裹相位的展开成为提高三维测量精度的一个难点。目前求解绝对相位有很多成熟的方法，包括格雷码法、多频外差法、倍频法等。

2.双目结构光三维测量技术

2.1 双目系统标定与图像校正

2.1.1 引言

双目系统一般包括两个参数基本一致的相机组成，双目三维重建的原理是通过不同位置的两个相机或者一台相机经过两个位置对同一个场景进行图像采集，利用空间点在两幅图像中的不同像素坐标位置，计算视差，然后通过三角测量的原理计算该目标的三维坐标信息从而获得整个场景的三维信息。三角测量原理需要用到双目系统的参数。所以必须对双目系统进行标定，要标定的参数包括左右相机内参以及相对外参，双目系统同样使用张氏标定法进行标定。

由于计算视差需要左右两张图像对齐，然而由于制作工艺的限制，一般的双目系统拍摄的图片都没有完美对齐，为了解决这个问题，双目系统进行标定后一般会进行图像校正的步骤。

2.1.2 双目系统标定

基于张氏标定法的双目系统标定流程如下：

(1) 准备一个棋盘格标定板，打印或定制皆可，建议定制，因为精度更高；

(2) 移动相机或标定板，双目左右相机拍摄多个位置标定板的图像，建议拍摄 20 到 30 张图像；

(3) 检测图像的角点；

- (4) 左右相机单独求解线性方程估计各自内参的封闭解；
- (5) 左右相机单独求解线性最小二乘问题，估计各自径向畸变的系数；
- (6) 使用 LM 算法最小化重投影误差，优化左右相机内参数。

(7) 计算左右相机每张图像的相机坐标系和世界坐标系之间的旋转矩阵和平移向量，求解线性最小二乘法问题，估计左右相机之间的旋转矩阵和平移向量，利用奇异值分解得到标准的旋转矩阵，即双目系统外参数。

2.1.3 双目图像校正

双目图像校正包括畸变校正和极线校正，其校正步骤如下：

- (1) 原始立体图像对结合各自标定的畸变参数去除畸变。
- (2) 去畸变的图像对结合系统内外参数进行极线校正，一般使用 bouguet 校正法，首先将左右图像各旋转外参中旋转矩阵的一半，使得左右图像共面，然后再将左右图像沿平移向量方向极线对齐。
- (3) 裁剪图像保留左右图像最大公共视区。

2.2 双目散斑结构光三维测量技术

2.2.1 引言

双目立体视觉是计算机视觉中一个重要的组成部分，由于双目视觉系统具有易实施性、原理简单等优点，因此在对室内外场景进行重建中具有很大的优势。但双目视觉在弱纹理区域以及重复纹理区域进行立体匹配时容易产生误匹配点，导致重建精度较低；结构光技术不依赖于物体表面的纹理信息，可根据实际应用的需求选择不同的结构光，测量精度高，但其易受环境光影响，而且标定技术较复杂。近年来将双目视觉和结构光方法相结合的方法不断发展，基于散斑结构光的双目立体匹配算法就是其中的代表，由主动光投影仪投射随机散斑增强了场景纹理，再结合双目立体匹配方法，对于纹理较弱的场景的测量精度大大提升。

2.2.2 双目散斑结构光三维重建过程

首先通过张氏标定法对双目相机进行标定，得到双目系统的内外参数，然后进行图像校正，包括镜头畸变校正和极线校正，校正完成后，使用校正后的立体图像对进行对应点匹配计算视差图，将视差图转化为深度图同时利用视差传播方法对原始深度图进行去噪处理，然后通过亚像素插值、左右一致性检测与联通域检测等一系列方法对所得深度图像进行平滑处理，得到高质量的深度图，最后结合双目系统标定参数利用三角测量原理进而恢复三维坐标信息。

2.2.3 存在的问题及其解决方法

(1) 物体表面干扰问题：待测物体有弱纹理、重复纹理、高反光等情况时，导致无法对特征点进行精确的定位从而影响深度估计效果。目前的解决方法有投射结构光图案来增加纹理信息，还有采用自适应结构光的方法来解决高反光等问题。

(2) 实时性问题：相机技术不断发展，图像的分辨率越来越大，立体匹配的计算复杂度也越来越大，导致了实时性较差。现有的解决方法有硬件加速，如使用 FPGA 和 GPU 等；还有图像处理的方法，如降低匹配分辨率，然后使用滤波的方法恢复高分辨率深度图。

3.前期工作成果

软件著作权：基于 RGB-D 深度相机的《多模态监控相机感知监控系统 V1.0》，登记号，2022SR0965725。

二、选题依据（本节可以整页扩页）

课题背景、选题依据、课题研究目的、理论意义和应用价值（工学硕士）/工程背景和实用价值（专业学位硕士）（不少于 1000 字）

1.课题背景

光学三维测量技术在工业自动检测、生物医学、虚拟现实、文物复制、人体测量和自动驾驶等领域具有广泛应用^[1]。这种巨大的应用需求，促使了多种光学测量技术的快速发展。随着计算机技术、数字图像传感器和光学器件的发展，很多光学三维测量技术已经进入商业应用的成熟阶段，同时新的三维测量方法还在不断涌现。

按照成像照明方式的不同，光学三维测量技术可分为被动三维测量和主动三维测量两大类^[2]。被动三维测量技术无需结构光照明，直接从一个或多个摄像系统获取的二维图像中提取物体的三维信息^[3]，比较常用的技术是立体视觉^[4]。主动三维测量技术采用不同种类的结构光，并拍摄经被测物体表面调制而发生变形的结构光图像，然后从携带有被测物体表面三维形貌信息的图像中计算出被测物体的三维形貌数据^[5]，典型的技术是单目结构光^[6]和主动立体匹配^{[7][8][9]}。

主动立体匹配是立体视觉和结构光相结合的技术，结构光为被测物体表面添加了携带深度信息的变形纹理，因此该方法的深度估计比单纯的立体视觉精度更高，鲁棒性更强；而立体视觉无需结构光投影仪的标定，相比于结构光系统，立体视觉系统标定的精度更高，从而使三角测量恢复深度的精度更高。因此主动立体三维重建技术以其大视场、非接触、高精度、强鲁棒、高效率的优点，已经成为光学三维测量技术中可靠的方法，在学术领域和工业三维测量领域得到广泛而深入的研究。

近年来，跨光谱图像（如 RGB-NIR 图像对）常用于自动驾驶的立体视觉系统。因此，人们积极提出了利用多光谱立体图像进行深度估计的方法。例如，一对跨光谱立体图像可以用作单目深度估计的训练数据集^[10]。人们对跨光谱图像中的匹配问题越来越感兴趣，因此针对跨光谱立体图像的研究不仅有单目深度估计，还有跨光谱立体匹配。

2.选题依据

实践依据：自 2010 年微软 Kinect^[11] 发布以来，消费级深度传感器得到了广泛应用。Kinect 基于单目结构光方法，2017 年发布的 iPhoneX 也是相同的原理。然而，单目结构光方法可能无法对远距离物体或强光下的室外场景进行深度测量。双目立体视觉系统的测量范围比单目结构光系统更大，也可以在阳光强烈的户外环境中工作，但它容易受到物体表面纹理的影响。近年来，基于深度学习的立体匹配方法取得了显著的进展。但是，在纹理较弱的场景中，这些方法仍可能失效，而且这种弱纹理物体在室内环境中非常常见。双目主动结构光系统（例如 Intel D435）依靠两个近红外（NIR）相机和一个红外投影仪进行深度估计，在室内和室外环境中都具有良好的适应性。获取纹理，需要第三个 RGB 相机。由于 RGB 相机和 NIR 相机之间存在基线，因此需要进行坐标系转换步骤以使深度图像与 RGB 图像对齐。由于深度图存在噪声和校正误差，因此 RGB 图像和深度图存在较大对齐误差。在硬件方面，需要三台相机和一台投影仪，没有充分利用所有设备，结构并不高效。因此双目主动结构光的硬件系统还有进一步开发的空间。本课题拟提出一种新型的系统结构，由一个 RGB 相机、一个 NIR 相机和一个红外投影仪组成了，整个系统构成 RGB-NIR 跨光谱主动立体匹配系统，NIR 相机和红外投影仪组成单目结构光子系统。系统具有多种模组，可以根据不同的场景自适应的选择合适的模组，提高系统的环境适应性和深度估计精度。

理论依据：近些年随着深度学习的广泛应用，特别是在卷积神经网络应用在图像领域取得巨大成功后，基于深度学习的立体匹配方法已明显超过传统算法。根据是否需要地面真值视差图作为标签，这些方法可分为有监督类和无监督类。在标签不易获得的许多真实场景中，无监督方法可以在没有真值信息的情况下进行学习，其中大多数方法利用光度一致性假设来表示光度损失^[12]。然而，当立体图像变得不对称时，如本课题的 RGB-NIR 跨光谱图像对，这

一假设将不在适用。为了解决这个问题现有的方法是在将 RGB 图像转换为 NIR 图像后执行匹配。然而如果图像转换效果不佳，那么立体匹配的效果将大大降低。针对上述问题，本课题拟采用无监督学习的方法，利用 CNN 强大的特征提取能力，从 RGB-NIR 跨光谱输入图像中提取对称特征，在特征空间进行匹配，来达到更好的效果。

3.课题研究目的、理论意义和应用价值

课题研究目的：针对现有主动立体匹配方法中存在的问题，搭建一套能够根据不同场景自适应选择相机模组的新型跨光谱主动立体匹配硬件系统，并开发基于无监督学习的跨光谱主动立体匹配算法，实现更强的环境适应性和较高的测量精度。

课题研究理论意义：经过多年研究和发展，传统的对称主动立体匹配技术理论趋于成熟，然而随着跨光谱相机模组的出现和广泛应用，基于跨光谱非对称的主动立体匹配仍然是一个开放性的问题。

课题研究应用价值：目前国内外已推出了多种主动立体三维测量系统，并得到了广泛的应用，包括 Intel 开发的用于深度感知、环境理解、环境交互的 RealSense D400 系列相机，德国 GOM 公司研发的全自动三维扫描仪 ATOS ScanBox，Google 公司开发的 Pixel 4 手机上的 uDepth 传感器。同时在许多现代设备中跨光谱系统变得非常常见，如 Realsense、Kinect 和 iPhoneX。此外，跨光谱立体匹配在自动驾驶、人脸识别、检测和场景解析方面非常有用，这充分说明了跨光谱主动立体匹配技术具有强大的生命力和巨大的应用潜力。

三、文献综述（本节可以整页扩页）

国内外研究现状、发展动态描述（不少于 1000 字）；所阅文献的查阅范围及手段，附参考文献（不少于 10 篇，其中近 3 年文献不少于 5 篇，英文文献不少于 3 篇，全部按照标准格式列出，并在文中顺序标注）

1.国内外研究现状及发展动态

1.1 立体匹配现状研究

自 Scharstein 和 Szeliski 在 2002 年提出了系统的双目立体视觉算法的分类和评估方式，对双目立体视觉的研究已经进入高速发展的阶段^[13]。

立体匹配算法根据视差原理，获得视差从而恢复场景的三维信息，算法的核心问题在于找到左视图中所有像素点在右视图中对应的点，根据匹配策略的不同，可以将立体匹配算法分为局部算法、全局算法和半全局算法三大类。局部算法又被称为基于窗口的算法，通常计算速度较快，匹配精度较低；全局算法需要根据全局优化理论最小化能量函数，需要更多的运算时间，结果精度较高；半全局（SGM）算法依旧采用类似于全局算法基于能量函数最小化的框架，但通过聚合邻域内多个路径方向的一维路径代价，来近似二维最优，大大提高了匹配效率，精度在局部算法和全局算法之间。

典型的局部算法包括绝对误差和（SAD）^[14]，差方和（SSD）^[15]，归一化互相关（NCC）^[16]，零归一化互相关（ZNCC）^[17]，Census 变换^[18]等。这些算法都在固定矩形窗口内计算匹配代价，会在视差不连续的区域引入前景膨胀效应^[19]。针对该问题，国内外学者展开了广泛而深入的研究，提出了各种局部算法的变体，包括可移动窗口法^[20]、多窗口法^[21]、可变窗口法^[22]、基于图像分割法^[23]、自适应权重法^[24]、基于图像分割的自适应窗口法^[25]、快速聚合算法^[26]、快速双边匹配算法^[27]、快速分割匹配算法^[28]，这些算法都对窗口的大小、形状、权重中的一个或多个进行调整，从而提高局部算法的匹配性能。

全局算法则会综合考虑图像的局部信息和全局约束，代表方法包括动态规划（DP）^[29]，图割（GC）^[30]，置信传播（BP）^[31]，协同优化（CO）^[32]等。全局算法的核心是求解使得全局能量函数最小化的视差，能量函数通常除了局部算法使用的匹配代价外，还有一个平滑项用来评估相邻像素的视差平滑度。目前在 Middlebury 数据集^[33]上，全局算法中匹配精度最高的是 PatchMatch 算法^[34]及其变体 LocalExp 算法^[35]和 3DMST 算法^[36]，这类算法核心思想是用近似平面来模拟像素区域，为区域内每个像素初始化一个随机平面，假设区域内至少有一个像素初始化平面近似于正确的平面，迭代地求解最优参数，代价是计算开销远远高于传统算法。

半全局算法^[37]将全局能量函数的二维优化问题简化为聚合多个一维路径方向优化问题，具有很高的匹配效率，它中和了局部算法和全局算法的优缺点，在精度和效率上有较好的平衡。

1.2 立体匹配网络研究现状

近年来，基于深度学习的立体匹配方法在匹配精度和匹配效率上均取得了远超传统方法的显著性能^[38]。随着深度学习的发展，Mayer 等人^[39]提出了第一个端到端立体匹配网络，尽管相对于传统立体匹配方法而言，在当时并没有达到最为精准的匹配效果，但它却引发了传统立体匹配方法向基于 CNN 模型的巨大转变。以 Mayer 等人的工作为基础，引发了一系列相似的网络架构^{[40][41][42][43][44]}，在网络中引入一个自定义层，即相关性层（correlation layer），负责计算两个图像中提取的特征之间的相似度分数。其中^{[40][41]}利用逐步细化或残差学习的多级级联结构，显著提高了基于深度学习的深度估计性能，并使其良好性能可以出现在各排行榜上。此外，得益于现代 GPU 上 2D 卷积运算的效率，某些模型^[42]每秒可实现几帧到几十帧的推理速度，而代价仅仅是损失了可忽略不计的精度。更有甚者通过深度学习将深度估计的任务与其余相关任务（例如结合边缘检测^[43]、语义分割^[44]等任务）联合学习，多个任务之间相辅相成，达到共同提升性能的效果。

与此同时，GC-Net^[45]通过拼接等方式构建 4D 代价体，再利用 3D 卷积显式地对特征处理

过程中的上下文信息进行编码,使得基于深度学习方法的性能首次超越了基于“手工设定”的传统方法,也带动了这一系列工作^{[46][47][48][49]}的产生。然而这种策略是以更高的内存需求和运行时间为代价的,因此随后的这些工作基本以降低内存需求以及提高准确性为目标来提出创新性工作。

以上这些工作基本都有相同的训练过程。具体来说,网络首先在大型且注释完美的合成 FlyingThings3D 数据集上进行预训练,然后在较小的 KITTI 训练集上进行微调。最近一些工作^{[50][51][52]}提出无监督的立体匹配方法,在标签不易获得的许多现实世界系统中,缺乏足够的带有标签的真实图像对来微调网络,无监督方法可以在没有基本事实信息的情况下进行学习,其中大多数方法利用光度一致性假设来表示光度损失。

1.3 跨光谱立体匹配现状研究

最近许多立体视觉研究针对的是一种跨光谱相机系统,该系统捕捉具有不同光谱带的跨光谱图像,例如一对近红外(NIR)和RGB图像。通过NIR相机拍摄的图像,与RGB图像相比,可以识别夜间的物体。为了利用这一优势,已经提出了在NIR和RGB图像之间执行对象匹配的算法^[53]。然而,由于RGB和NIR图像的光谱带不同,传统匹配方法在RGB和近红外图像中的应用受到限制。许多经典的方法都致力于设计好的描述子来克服交叉光谱图像之间的差异。传统立体匹配的关键是在立体图像对之间设计鲁棒的描述子或特征。HEO等人^[54]直接使用了颜色形成模型,并提出了自适应归一化互相关(ANCC)解决光照变化和相机参数差异。使用基于窗口的局部自相似性描述子(LSS)来估计热图像和可见光图像之间的密集对应。Pinggera等人^[55]表明,基于定向梯度直方图的稠密梯度特征(HOG)比LSS描述子具有更好的性能。Aguilera等人^[56]提出了一种基于非线性强度变化的特征点匹配的特征描述子。Kim等人^[57]通过随机感受野池化对LSS描述子进行改进,提出了稠密自适应自相关描述子(DASC)。然而,如果不考虑光谱特性,它们就无法解决跨光谱差异。Zhi等人^[58]提出了基于无监督学习的深度材料感知交叉光谱立体匹配来解决这一问题。它们通过对光谱差异的处理,在跨光谱立体匹配中取得了良好的性能。Liang等人^[59]提出了一种交替优化光谱平移网络和立体匹配网络的跨光谱立体匹配新框架。而且,Zhu等人^[60]提出了一种新的基于循环GAN的风格自适应网络,以提高光谱变换的鲁棒性。

近期的方法都是基于光谱图像变换的,匹配过程在图像转换之后执行,这显著提高了匹配性能。但是,如果图像没有适当转换,则交叉光谱图像的匹配精度将降低。迄今为止提出的方法都不涉及将RGB图像转换为与实际NIR图像完美匹配。

2.文献的查阅范围及手段

文献查阅范围:中国学术期刊全文数据库、EI Compendex (EI)、Web of Science (SCI)、万维数字资源系统、维普中国科技期刊, Institute of Electrical and Electronics Engineers (IEEE)、Google 学术、百度学术等国内外电子版书籍、会议、期刊等文献,以及东北大学图书馆馆藏文件等。

文献查阅手段:首先阅读综述性文章,然后选择文章列出的参考文献中有代表性的进行筛选和精读,针对阅读时遇到的新颖的算法和技术问题进行查阅,涉及论文的相关代码通过GitHub 查阅。

3.参考文献

- [1] Luhmann T. Close range photogrammetry for industrial applications[J]. ISPRS journal of photogrammetry and remote sensing, 2010, 65(6): 558-569.
- [2] 苏显渝, 李继陶. 信息光学[M]. 科学出版社, 1999.
- [3] 段峰, 王耀南, 雷晓峰, 等. 机器视觉技术及其应用综述[J]. 自动化博览, 2002, 19(3): 59-61.
- [4] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International journal of computer vision, 2002,

- 47(1): 7-42.
- [5] Li B , An Y , D Cappelleri, et al. High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics[J]. International Journal of Intelligent Robotics & Applications, 2017, 1(1):86-103.
 - [6] Salvi J , J Pagès, Batlle J . Pattern codification strategies in structured light systems[J]. Pattern Recognition, 2004, 37(4):827-849.
 - [7] Zhang C, Wang L, Meng Z. Binocular Depth Estimation Based on Diffractive Optical Elements and the Semiglobal Matching Algorithm[C]//2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2018: 504-508.
 - [8] Gu F , Song Z , Zhao Z . Single-Shot Structured Light Sensor for 3D Dense and Dynamic Reconstruction[J]. Sensors, 2020, 20(4):1094.
 - [9] Yin W, Zuo C, Feng S, et al. An openCL-based speckle matching on the monocular 3D sensor using speckle projection[C]//Fourth International Conference on Photonics and Optical Engineering. SPIE, 2021, 11761: 517-524.
 - [10] Lu Y, Lu G. An alternative of lidar in nighttime: Unsupervised depth estimation based on single thermal image[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3833-3843.
 - [11] Zhang Z. Microsoft kinect sensor and its effect[J]. IEEE multimedia, 2012, 19(2): 4-10.
 - [12] Chen X, Xiong Z, Cheng Z, et al. Degradation-agnostic Correspondence from Resolution-asymmetric Stereo[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12962-12971.
 - [13] D Scharstein, Szeliski R . A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms[J]. International Journal of Computer Vision, 2002, 47(1):7-42.
 - [14] Mühlmann K, Maier D, Hesser J, et al. Calculating dense disparity maps from color stereo images, an efficient implementation[J]. International Journal of Computer Vision, 2002, 47(1): 79-88.
 - [15] Okutomi M , Kanade T . A locally adaptive window for signal matching[J]. International Journal of Computer Vision, 1992, 7(2):143-162.
 - [16] Mordohai P , Medioni G . Stereo Using Monocular Cues within the Tensor Voting Framework[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(6):968-982.
 - [17] Di Stefano L, Mattoccia S, Tombari F. ZNCC-based template matching using bounded partial correlation[J]. Pattern recognition letters, 2005, 26(14): 2129-2134.
 - [18] Zabih R, Woodfill J. Non-parametric local transforms for computing visual correspondence[C]//European conference on computer vision. Springer, Berlin, Heidelberg, 1994: 151-158.
 - [19] Zhang Y, Kambhamettu C. Stereo matching with segmentation-based cooperation[C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2002: 556-571.
 - [20] Bobick A F, Intille S S. Large occlusion stereo[J]. International Journal of

- Computer Vision, 1999, 33(3): 181–200.
- [21] Hirschmüller H, Innocent P R, Garibaldi J. Real-time correlation-based stereo vision with reduced border errors[J]. International Journal of Computer Vision, 2002, 47(1): 229–246.
 - [22] Veksler O. Fast variable window for stereo correspondence using integral images[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. IEEE, 2003, 1: I–I.
 - [23] Gerriets M, Bekaert P. Local stereo matching with segmentation-based outlier rejection[C]//The 3rd Canadian Conference on Computer and Robot Vision (CRV'06). IEEE, 2006: 66–66.
 - [24] Yoon K J, Kweon I S. Adaptive support-weight approach for correspondence search[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(4): 650–656.
 - [25] Tombari F, Mattoccia S, Stefano L D. Segmentation-based adaptive support for accurate stereo correspondence[C]//Pacific-Rim Symposium on Image and Video Technology. Springer, Berlin, Heidelberg, 2007: 427–438.
 - [26] Tombari F, Mattoccia S, Di Stefano L, et al. Near real-time stereo based on effective cost aggregation[C]//2008 19th International Conference on Pattern Recognition. IEEE, 2008: 1–4.
 - [27] Mattoccia S, Giardino S, Gambini A. Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering[C]//Asian Conference on Computer Vision. Springer, Berlin, Heidelberg, 2009: 371–380.
 - [28] Mattoccia S, De-Maeztu L. A fast segmentation-driven algorithm for accurate stereo correspondence[C]//2011 International Conference on 3D Imaging (IC3D). IEEE, 2011: 1–6.
 - [29] Gong M , Yang Y H . Fast unambiguous stereo matching using reliability-based dynamic programming.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(6):998–1003.
 - [30] Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts[C]//Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. IEEE, 2001, 2: 508–515.
 - [31] Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure[C]//18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 3: 15–18.
 - [32] Wang Z F, Zheng Z G. A region based stereo matching algorithm using cooperative optimization[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1–8.
 - [33] D. Scharstein and R. Szeliski. Middlebury online stereo evaluation. www.middlebury.edu/stereo.
 - [34] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Trans. Graph., 2009, 28(3): 24.
 - [35] Taniai T, Matsushita Y, Sato Y, et al. Continuous stereo matching using local expansion moves[J]. arXiv preprint arXiv:1603.08328.

- [36] Li L , Yu X , Zhang S , et al. 3D cost aggregation with multiple minimum spanning trees for stereo matching[J]. *Applied Optics*, 2017, 56(12):3411.
- [37] H Hirschmüller. Stereo Processing by Semiglobal Matching and Mutual Information[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 30.
- [38] 尹晨阳, 职恒辉, 李慧斌. 基于深度学习的双目立体匹配方法综述[J]. *计算机工程*, 2022, 48(10):1-12.
- [39] Mayer N, Ilg E, Hausser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 4040-4048.
- [40] Pang J, Sun W, Ren J S J, et al. Cascade residual learning: A two-stage convolutional neural network for stereo matching[C]//*Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017: 887-895.
- [41] Liang Z, Feng Y, Guo Y, et al. Learning for disparity estimation through feature constancy[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 2811-2820.
- [42] Yin Z, Darrell T, Yu F. Hierarchical discrete distribution decomposition for match density estimation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019: 6044-6053.
- [43] Song X, Zhao X, Hu H, et al. Edgestereo: A context integrated residual pyramid network for stereo matching[C]//*Asian Conference on Computer Vision*. Springer, Cham, 2018: 20-35.
- [44] Yang G, Zhao H, Shi J, et al. Segstereo: Exploiting semantic information for disparity estimation[C]//*Proceedings of the European conference on computer vision (ECCV)*. 2018: 636-651.
- [45] Kendall A, Martirosyan H, Dasgupta S, et al. End-to-end learning of geometry and context for deep stereo regression[C]//*Proceedings of the IEEE international conference on computer vision*. 2017: 66-75.
- [46] Yu L, Wang Y, Wu Y, et al. Deep stereo matching with explicit cost aggregation sub-architecture[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2018, 32(1).
- [47] Chang J R, Chen Y S. Pyramid stereo matching network[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 5410-5418.
- [48] Tulyakov S, Ivanov A, Fleuret F. Practical deep stereo (pds): Toward applications-friendly deep stereo matching[J]. *Advances in neural information processing systems*, 2018, 31.
- [49] Khamis S, Fanello S, Rhemann C, et al. Stereonet: Guided hierarchical refinement for real-time edge-aware depth prediction[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 573-590.
- [50] Godard C, Mac Aodha O, Brostow G J. Unsupervised monocular depth estimation with left-right consistency[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 270-279.
- [51] Zhang Y, Khamis S, Rhemann C, et al. Activestereonet: End-to-end self-supervised learning for active stereo systems[C]//*Proceedings of the European*

- Conference on Computer Vision (ECCV). 2018: 784–801.
- [52] Zhou C, Zhang H, Shen X, et al. Unsupervised learning of stereo matching[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1567–1575.
 - [53] Chang Y J, Lee B G, Jeon M. Warping-based spectral translation network for unsupervised cross-spectral stereo matching[J]. Information Sciences, 2022, 588: 214–230.
 - [54] Heo Y S, Lee K M, Lee S U. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(5): 1094–1106.
 - [55] Pinggera P, Breckon T, Bischof H. On cross-spectral stereo matching using dense gradient features[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. 2012, 2: 3.
 - [56] Aguilera C A, Aguilera F J, Sappa A D, et al. Learning cross-spectral similarity measures with deep convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016: 1–9.
 - [57] Kim S, Min D, Ham B, et al. DASC: Robust dense descriptor for multi-modal and multi-spectral correspondence estimation[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(9): 1712–1729.
 - [58] Zhi T, Pires B R, Hebert M, et al. Deep material-aware cross-spectral stereo matching[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1916–1925.
 - [59] Liang M, Guo X, Li H, et al. Unsupervised cross-spectral stereo matching by learning to synthesize[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 8706–8713.
 - [60] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125–1134.
 - [61] Xu Y, Yang X, Yu Y, et al. Depth Estimation by Combining Binocular Stereo and Monocular Structured-Light[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 1746–1755.
 - [62] Chen X, Xiong Z, Cheng Z, et al. Degradation-agnostic Correspondence from Resolution-asymmetric Stereo[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12962–12971.
 - [63] Liang X, Jung C. Deep Cross Spectral Stereo Matching Using Multi-Spectral Image Fusion[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 5373–5380.

四、研究内容（本节可以整页扩页）

1. 研究构想与思路、主要研究内容及拟解决的关键问题（不少于 1000 字）

1.1 研究构想与思路

深度相机已成为三维场景重建和场景理解的必要条件，而主动立体技术又是深度相机中极具潜力的技术，目前基于结构光的主动立体三维重建技术已经在各种高新技术领域和新型场景中得到应用，包括机器人技术、自主无人机、导航、驾驶员监控、人机交互、虚拟和混合现实以及远程会议。

现有的消费级深度相机产品主要有单目结构光、双目立体视觉、主动双目立体视觉和 TOF。然而，单目结构光方法可能无法对远距离物体或强光下的室外场景进行深度测量。双目立体视觉容易受到物体表面纹理的影响。双目主动结构光系统（例如 Intel D435）依靠两个近红外（NIR）相机和一个红外投影仪进行深度估计要获取纹理，需要第三个 RGB 相机。由于 RGB 相机和 NIR 相机之间存在基线，因此需要进行坐标系转换步骤以使深度图像与 RGB 图像对齐。由于深度图存在噪声和校正误差，因此 RGB 图像和深度图存在较大对齐误差。在硬件方面，需要三台相机和一台投影仪，没有充分利用所有设备，结构并不高效。针对上述存在的问题，本课题旨在进一步开发主动双目立体匹配系统，搭建新型硬件系统结构，充分利用当前相机模组中的 RGB 相机、NIR 相机和红外投影仪，在此基础上解决 RGB-NIR 跨光谱立体对的匹配问题，使得主动双目立体匹配具有更强的环境适应能力和较高的精度，同时结构高效，有更高的经济效益。

本课题拟采用图 1 所示系统结构，整个系统由一个 RGB 相机、一个 NIR 相机和一个红外投影仪组成，整个系统模组构成主动双目立体匹配系统，NIR 相机和红外投影仪模组组成单目结构光子系统。该系统相较于传统的硬件结构具有更加高效的结构，和更加灵活的深度感知模组。基于该系统模组，RGB 图像和深度图天然对齐，因此无对齐误差；具有更强的环境适应能力，能够根据场景自适应切换合适的模组来达到最优的效果；还可以结合两种模组来达到更加精确的结果。

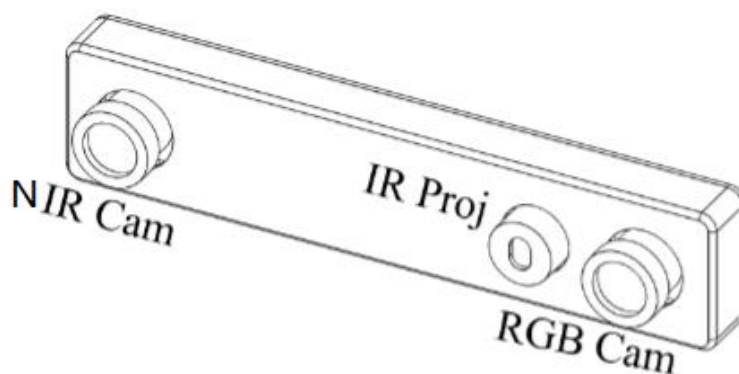


图 1 跨光谱主动立体匹配系统简图

近年来针对立体匹配问题的研究层出不穷，基于学习的方法在性能上已经超过传统方法，针对真实世界视差真值难以获取的特性，无监督的学习方法不需要视差真值信息，具有更强的适应能力，因此本课题拟采用无监督学习的方法。另外现有的方法的输入大多为对称立体对，而本课题的系统结构中同时存在 RGB 相机和 NIR 相机，RGB 相机接收可见光形成 RGB 图像，NIR 相机接收近红外光形成 NIR 图像，并且这两个相机组成立体匹配系统，因此本课题立体匹配的输入是跨光谱的 RGB-NIR 图像对。如图 2 所示，RGB-NIR 跨光谱立体对因其成像光谱不同而具有明显的不对称性。因此，现有的主流方法将不再适用。卷积神经网络具有强大的学习能力和特征提取能力，于是，本课题拟采用特征提取器提取对称的特征，然后基于特征空间构建特征度量来对网络进行训练。

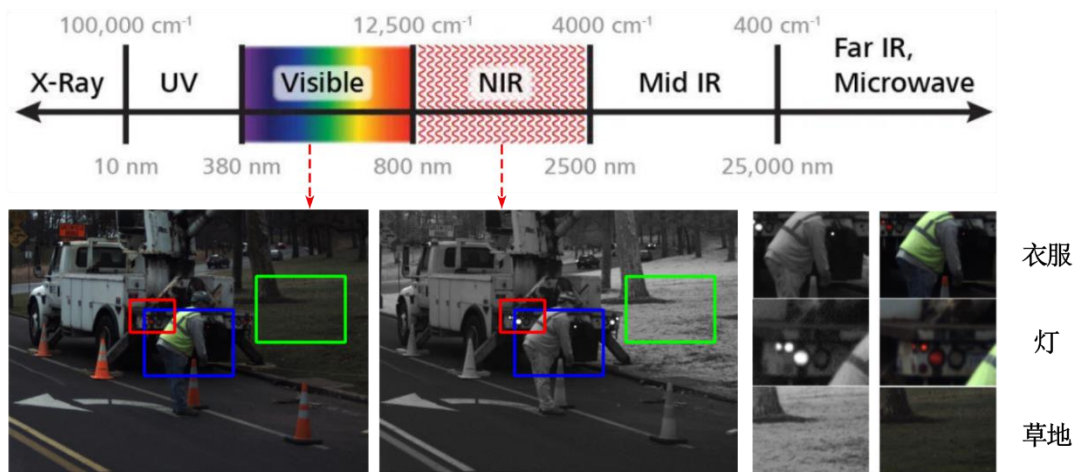


图 2 RGB-NIR 跨光谱立体对

基于学习的方法对数据有很强的依赖性，因此本课题需要构建训练数据集，对网络进行充分的训练。本课题拟采用采集真实场景数据集和基于现有数据集合成训练数据集两种方式构建数据集。

本课题的系统中存在单目结构光子系统，单目结构光方法获得的深度图虽然具有许多无效点，但是其有效的部分深度值是比较可靠的。为了进一步提高深度估计的精度，高效利用结构优势，本课题拟采用稀疏深度引导的方式，利用单目结构光子系统的稀疏深度作为引导信息，进一步优化网络的预测结果。

1.2 主要研究内容

经过上面的思考，本课题有四个方面的内容。第一，针对现有硬件系统结构存在的问题进一步开发，构建新型硬件系统；第二，采用无监督学习的方式解决视差真值难以获取问题，采用学习的方法需要大量数据的支撑，需要构建适用于本课题的训练数据集；第三，针对新系统带来的跨光谱不对称输入，基于 CNN 强大的特征提取和学习能力，在特征空间构建特征度量进行网络训练，来解决这个问题；第四，利用单目结构光子系统可靠的稀疏深度引导网络进行视差优化，进一步提高深度估计的精度。四个研究内容具体展开如下：

（1）跨光谱主动立体匹配硬件系统搭建

针对现有主动立体匹配系统存在的问题，本课题在现有系统基础上进行改进，搭建新型跨光谱主动立体匹配系统，系统实物硬件图如图 3 所示，整个系统包括 RGB 相机、NIR 相机和红外投影仪组成，其中 NIR 相机和红外投影仪还可以构成单目结构光子系统。该系统具有，结构高效，同时存在双目系统和单目结构光子系统；环境适应性强，能根据不同场景自适应调节相机模组；没有 RGB-深度对齐误差。



图 3 跨光谱主动立体匹配系统实物图

（2）训练数据集构建

本课题拟采用无监督学习方法进行算法设计，基于学习的方法对数据具有很强的依赖

性，因此构建训练数据集十分必要，本课题需要的训练数据集包括 RGB-NIR 跨光谱被动立体对、RGB-NIR 跨光谱主动立体对（带散斑结构光图案），同时合成数据集要尽可能接近真实场景。因此本课题拟采用采集真实场景数据集与合成数据集两张方式构建数据集。

合成数据集需要尽可能接近真实场景，于是本课题基于波动光学和几何光学模拟实际的成像过程，整体的合成方法流程如图 4 所示：

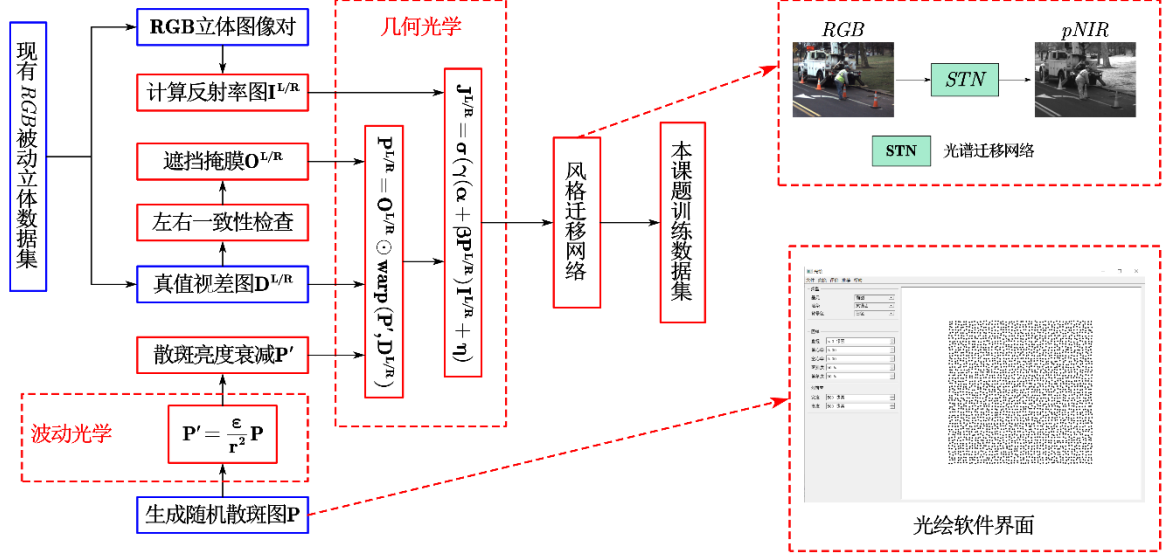


图 4 合成数据集流程图

图中的符号及公式： $D^{L/R}$ 左右视图视差图； $I^{L/R}$ ，左右视图反射率； $O^{L/R}$ ，左右视图关于照明模块的遮挡掩膜；结构光图案 P ；衰减后的结构光图案 P' ； $J^{L/R}$ 是经过波动光学和几何光学合成的图像； γ 是描述相机传感器对光的响应的一个标量； α 是环境光强度； β 是激光投射器的功率系数； η 是高斯噪声； σ 是强度截断函数； \odot 是逐像素乘积的操作。

波动光学模拟光线传播随距离衰减现象：

$$P' = \frac{\epsilon}{r^2} P \quad (1)$$

几何光学模拟实际成像过程：

$$P^{L/R} = O^{L/R} \odot \text{warp}(P', D^{L/R}) \quad (2)$$

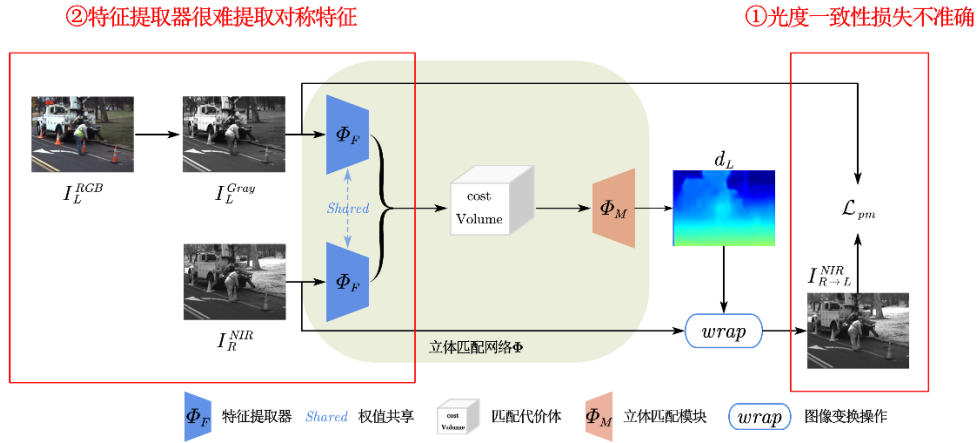
$$J^{L/R} = \gamma(\alpha + \beta P^{L/R}) I^{L/R} + \eta \quad (3)$$

（3）无监督跨光谱主动立体匹配模型搭建

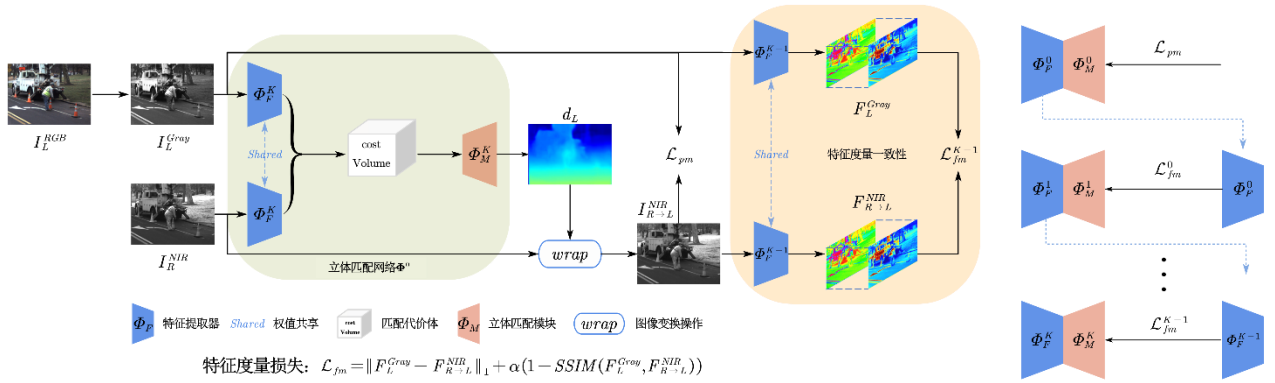
如今，主流的立体匹配方法都是基于对称立体对进行的，而在本课题中输入为 RGB-NIR 跨光谱立体对，由于输入的不对称性，这使得现有的无监督立体匹配网络不再适用，具体问题图 5 所示，跨光谱的立体对会导致光度一致性假设不在准确；同时特征提取器能否从不对称的输入中提取对称的特征。

本课题拟在特征空间构建度量假设，使用特征度量一致性来构造损失函数，如式（4）所示对网络进行训练。针对特征提取器的性能问题，本课题拟采用一种自增强策略，基于渐进损失更新来逐步优化特征提取器和增强特征度量一致性。

$$\mathcal{L}_{fm} = \|F_L^{Gray} - F_{R \rightarrow L}^{NIR}\|_1 + \alpha(1 - SSIM(F_L^{Gray}, F_{R \rightarrow L}^{NIR})) \quad (4)$$

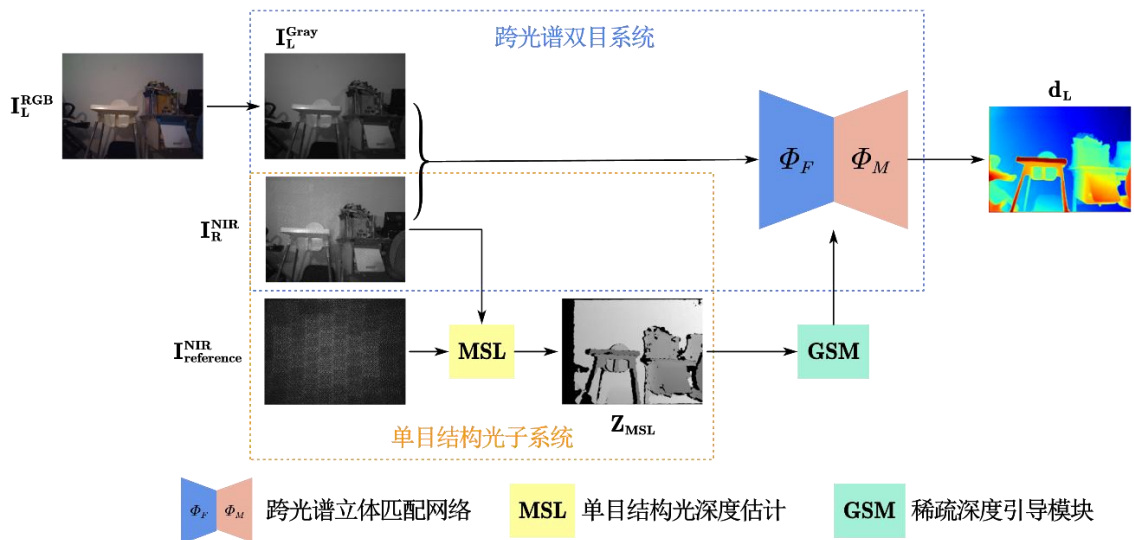


本课题所搭建的网络结构图如图 6 所示：



(4) 基于稀疏深度引导的视差优化

结构光方法估计出来的深度图会有许多无效点，但其有效的部分是可靠的。由于本课题的硬件系统存在单目结构光子系统，可以将这一系统优势合理利用起来，于是本课题拟采用单目结构光子系统估计出来的稀疏可靠深度值作为引导信息，加入网络结构中，通过稀疏深度引导模块进行视差优化，整个网络结构如图 7 所示。



1.3 拟解决的关键问题

(1) 跨光谱主动立体匹配硬件系统搭建：针对现有三维测量方法的不足，本课题采用双目主动立体匹配方法，提出新的系统结构，能解决现有系统结构冗余，存在环境适应能力不强，而且有对齐误差的问题；

(2) 训练数据的构建：本课题利用新型系统采集真实场景数据集并基于现有数据集结合波动光学和几何光学合成训练数据集，能解决跨光谱数据集不足与合成数据集需要接近真实场景的问题；

(3) 无监督跨光谱主动立体匹配模型搭建：本课题采用基于特征度量一致性的无监督学习方式，并提出一种自增强策略加强特征度量一致性和优化特征提取器性能，能解决 RGB-NIR 跨光谱不对称输入的立体匹配问题；

(4) 稀疏深度引导视差优化：本课题充分利用系统优势，结合单目子系统的稀疏深度信息作为引导，指导立体匹配网络进行视差优化，能进一步优化网络的视差结果，提高后续三维重建精确。

2. 拟采取的研究方法、技术路线、实施方案及可行性分析

2.1 拟采取的研究方法

本课题采取的研究流程如图 8 所示。

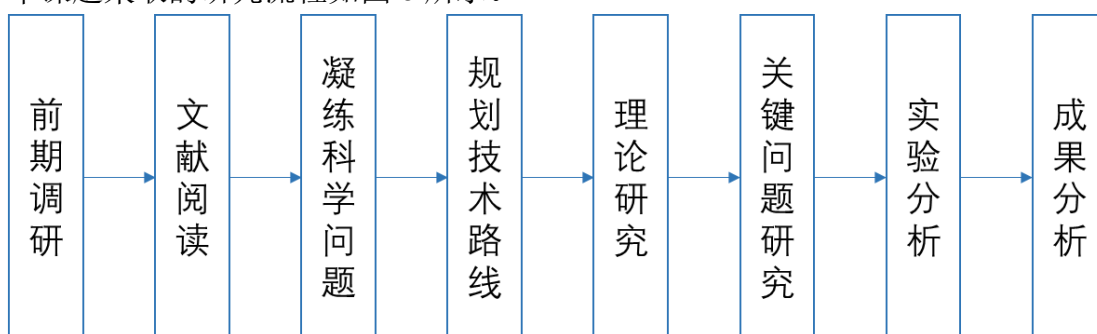


图 8 研究流程图

具体：前期调研（关于主动立体视觉系统的搭建和标定、RGB-NIR 跨光谱立体匹配的研究现状）→阅读相关文献（关于波动光学、几何光学、主动立体网络、无监督网络、跨光谱立体匹配方法）→凝练科学问题（如何基于跨光谱立体对进行主动立体匹配，如何充分利用系统优势进一步优化结果）→规划技术路线（每一步需要哪些技术，实现什么功能）→理论研究（针对技术路线中的各种技术精读相关文献）→关键问题研究（针对关键问题、难题与老师和同门展开讨论）→仿真实验和实物实验验证（采用具体的深度学习框架对构建的网络进行训练和测试）→论文撰写（整理实验数据，完成论文写作）。

2.2 技术路线

本课题的技术路线规划如图 9（下页）所示。

2.3 实施方案

（1）搭建硬件系统，并进行系统标定；

（2）构建训练数据集，利用新型硬件系统采集真实场景数据集，基于波动光学和几何光学知识，理解实际成像过程，基于现有数据集合成的本课题所需的训练数据集。

（3）构建无监督跨光谱主动立体网络：搜集三大顶会中基于深度学习的立体匹配技术的文献，精读并进行归纳总结，明确特征度量一致性和自我提升策略。基于 Pytorch 等开源框架进行网络构建，利用第（2）步合成的训练数据集进行仿真训练，联合优化结构光图案和重建算法，并与目前的其他算法进行对比实验，分析实验结果。

（4）稀疏深度引导视差优化：搜集三大顶会中基于视差优化方法，并进行归纳总结，编写视差优化模块代码。进行仿真和实物实验，分析算法的深度重建精度和效率，并与其它算法进行对比实验，整理实验数据。

（5）三维重建：利用第（4）步获得视差图以及第（1）步获得标定参数，基于三角测量原理计算点云，在 PCL 上进行显示。

2.4 可行性分析

（1）理论基础：在计算机视觉领域 CVPR、ICCV、ECCV 三大顶会和期刊上已有较多的关于立体视觉网络、开源数据集、无监督学习的文献，近年来，跨光谱立体匹配的文献也不断涌现，说明了本课题跨光谱主动立体匹配算法的可行性。

（2）应用基础：目前深度相机已经商业化，RealSense D400 系列相机、iPhone X 的 faceID 等都具有与本课题所搭建系统有相似之处，而且已经获得广泛应用，同时在自动驾驶领域，跨光谱图像更是感知部分的重要组成部分，证明了跨光谱主动立体相机三维重建应用的可行性。

（3）实验基础：实验室有较好光学实验平台，和许多立体相机原型；同时实验室具有较好的服务器集群，用于网络训练。这些实验条件为实物实验和仿真实验提供保障。

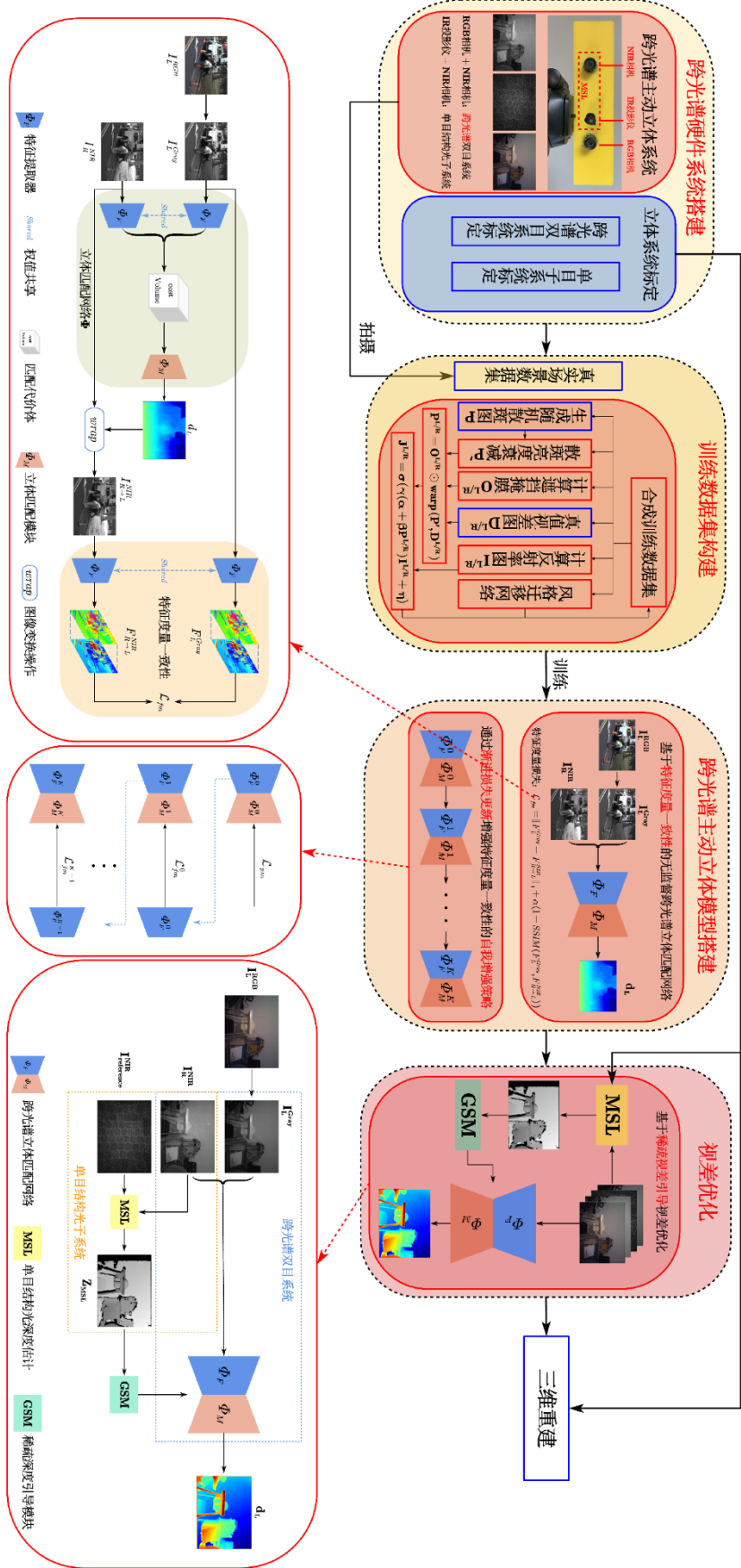


图 9 技术路线图

五、预期研究成果（本节可以整页扩页）

对所研究的成果进行阐述，同时要对与前文研究内容的相关性及与前人（他人）研究成果的差异性进行描述

1.预期研究成果

- （1）搭建跨光谱主动立体匹配硬件系统。
- （2）使用本课题系统采集真实场景数据集，并根据波动光学和几何光学在开源被动数据集上进行拓展，形成可用于深度学习训练的跨光谱主动立体数据集。
- （3）构建基于特征度量一致性的无监督跨光谱主动立体匹配模型。
- （4）采用渐进损失更新的自我增强策略训练网络，加强特征度量一致性。
- （5）利用单目子系统的稀疏深度引导网络进行视差优化。
- （6）基于所研究的成果，撰写论文，并发表一篇高水平论文。

2.与前人研究的相关性

- （1）Xu 等人^[61]搭建了一个与本课题类似的硬件系统，使用有监督学习进行网络训练，最后进行视差优化。
- （2）Zhang 等人^[51]提出了一种无监督的主动立体网络，无需训练数据的真值便可对网络进行优化。
- （3）liang 等人^[63]同样使用无监督学习研究跨光谱立体匹配问题。
- （4）Chen 等人^[62]构建特征空间的损失函数来训练分辨率不一致输入的立体匹配网络。

3.与前人研究的差异性

- （1）本课题搭建的硬件系统具备更强的环境适应性，系统内存在多个模组，可以根据使用场景自适应切换相机模组。
- （2）本课题所构建的数据集充分考虑了实际的光学成像过程，基于波动光学和几何光学合成训练数据集。
- （3）本课题采用的是在特征空间基于特征度量一致性的无监督学习训练跨光谱主动立体匹配网络，而前人大多基于图像变换的方法进行。

六、研究条件（本节不允许扩页）

1. 所需实验手段、研究条件和实验条件

（如果没有，可以空白）

1.1 实验手段

- （1）理论分析与研究；
- （2）搭建跨光谱主动立体匹配硬件系统；
- （3）构建跨光谱训练集；
- （4）跨光谱立体网络模型的构建和仿真训练；
- （5）实物场景的算法测试。

1.2 研究条件和实验条件

- （1）PC 机
- （2）2080Ti 服务器
- （3）跨光谱立体系统
- （4）Nvidia Jetson TX2 边缘计算平台
- （5）实物实验模具
- （6）光学实验平台

2. 所需经费，包含经费来源、开支预算（工程设备、材料须填写名称、规格、数量）

（如果没有，可以空白）

1.1 所需经费及预算

- （1）跨光谱主动立体相机：2000 元；数量：1 个
- （2）Nvidia Jetson TX2 边缘计算平台：5000 元；数量：1 个
- （3）系统配套显示器：200 元；数量：1 个
- （4）系统配套键盘鼠标：200 元；数量：1 套

总计：7400 元

1.2 经费来源

国家自然科学基金项目：自适应环境结构光生成方法研究

七、工作计划（本节不允许扩页）

序号	阶段及内容	工作量估计 (时数)		起止日期	阶段研究成果
1	搜集资料，阅读文献	400		2022 年 9 月-2022 年 10 月	确立研究思路和技术路线
2	搭建跨光谱主动立体三维重建系统，进行系统标定	150		2022 年 10 月-2022 年 11 月	搭建系统并完成系统标定
3	光学知识研究	400		2021 年 11 月-2021 年 12 月	模拟实际光学成像过程
4	采集与合成训练数据集	400		2022 年 12 月-2023 年 2 月	构建本课题所需要的训练数据集
5	无监督跨光谱主动立体网络研究	400		2023 年 2 月-2023 年 4 月	构建一个无监督跨光谱主动立体匹配网络
6	基于单目结构光稀疏深度引导的视差优化	400		2023 年 4 月-2023 年 5 月	进一步优化网络预测的视差结果
7	在跨光谱主动立体系统上进行实验，对设定的场景进行三维测量	400		2023 年 5 月-2023 年 6 月	完成本课题研究内容
8	对比实验	400		2023 年 6 月-2023 年 8 月	完成对比实验并论证本课题的创新性
9	后续改进并撰写论文	600		2023 年 8 月-10 月	完成论文撰写
		合计	3550		

八、导师意见

<p>基于上述内容及学生前期研究情况，导师是否同意其开题</p> <p style="text-align: right;">导师签字（盖章）： 贾同</p> <p style="text-align: right;">2022 年 10 月 29 日</p>
--