

東 北 大 學

信息科学与工程学院  
硕士研究生学位论文中期报告

学 号：	2170871
姓 名：	张佳预
导 师：	贾同
学科类别：	<input type="checkbox"/> 工学 <input checked="" type="checkbox"/> 全日制专业学位
学科/工程领域：	控制工程
研究方向：	计算机视觉
论文题目：	基于无监督学习的多光谱 立体匹配方法与系统研究
完成学分：	28

东北大学研究生院  
2023 年 11 月 25 日

## 排版格式

一、中期检查报告首页各项按要求填写。

二、除原有标题外，中期检查报告正文部分使用非加粗宋体、小四号字；如需设置小标题，使用加粗宋体、小四号字。

三、段落首行缩进 2 字符。

四、撰写时，将各部分标题下括号中的说明文字删除。

五、如需打印，请 A4 纸正反面打印。

## 一、论文工作总体完成情况

### 1. 开题报告中的论文主要研究内容

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需要额外 RGB 相机获取纹理，这种结构存在较大对齐误差且硬件结构并不高效。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。针对以上问题，开题报告主要有以下三个研究内容：

#### （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，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。

#### （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种方式构建数据集。

#### （3）无监督跨光谱主动立体匹配模型搭建

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，跨光谱的立体对会导致光度一致性假设不再准确；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特征度量一致性来构造损失函数对网络进行训练。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。最后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。

## **2. 论文主要研究内容调整情况**

### **2.1 调整后的题目**

开题报告题目：《基于无监督学习的跨光谱主动立体匹配方法与系统研究》

中期报告题目：《基于无监督学习的多光谱立体匹配方法与系统研究》

**题目调整的理由与依据：**“多光谱”相比于“跨光谱”表述更加科学、准确，更能直接突显出不同本课题的研究特色，后续表述均以“多光谱”为准。在开题后的进一步研究过程中，发现相比于主动结构光多光谱应用，被动多光谱应用场景更加广泛，且更具有挑战性，硬件结构设计也做出了相应调整，不再局限于主动结构光方式，提高了本课题研究方法的通用性，因此将“主动”去掉。

### **2.2 调整后的主要研究内容**

目前多光谱传感器普遍存在于手机、无人机和汽车等，多光谱图像相比单一彩色图像具有更加丰富的信息，广泛应用于夜间自动驾驶、显著性检测，人脸重建和图像恢复等，多光谱图像应用的基础是不同光谱图像之间的对应点匹配，因此多光谱立体匹配技术至关重要。但多光谱图像由于成像光波长不同而造成明显的外观差异，另外由于特殊光谱（如热红外）传感器的制造工艺问题，不同光谱图像常常存在分辨率差异。由于这两个问题的限制，传统立体匹配方法天然失效，然而由于现有公开数据集少，多光谱数据集真值获取困难，导致基于深度学习的多光谱立体匹配方法发展缓慢，针对以上问题，调整后主要有以下两个研究内容：

#### **（1）多光谱硬件系统与数据集构建**

针对现有多光谱数据集少，视差真值标签获取成本高的问题，本课题搭建新型多光谱立体匹配系统，整个系统包括两个 RGB 相机和一个 NIR 相机组成，该系统具有，结构高效，同时存在 RGB-NIR 多光谱双目系统和 RGB-RGB 彩色双目系统，能够同时采集这两种数据构建多光谱数据集。常规 RGB-RGB 立体匹配方法已经取得了惊艳的效果，为了解决真值标签获取困难的问题，本课题充分利用系统特性，提出一种基于视差分布一致性的伪标签生成方法，利用现有立体匹配方法从 RGB-RGB 立体对中估计出可靠的视差伪标签用于后续模型训练。

#### **（2）多光谱立体匹配模型构建**

针对多光谱立体对存在的问题，本课题提出一种高效的多光谱立体匹配模型，基于无监督学习训练模型，具体分为三个部分：(i) 针对多光谱立体对存在显著外观差异问题，提出一种基于特征空间的特征度量一致性损失，约束特征提取器，从多光谱图像中学习具有一致性的特征，在特征空间解决外观差异问题；(ii) 针对多光谱立体对的分辨率差异问题，提出一种基于特征插值的连续特征生成方法，生成任意指定分辨率的特征，同时融合高分辨率 RGB 图像的全局特征恢复全局信息；(iii) 针对视差优化，设计两个带残差结构的多层感知机，进行视差回归得到任意分辨率的视差结果。

**主要研究内容调整的理由与依据：**为了突出研究重点，提升研究方法的通用性，本课题的研究对象不再局限于某个硬件系统，而是多光谱立体匹配问题本身，多光谱图像应用如今已出现在许多领域，而关于多光谱立体匹配的研究还相对落后，基于此对主要研究内容做出相应改动。具体调整有，(i) 将开题报告中的研究内容一与研究内容二合并作为一个研究内容，针对多光谱数据集少，真值难以获取的问题，对原有硬件结构进行改进，并使用新型硬件结构采集数据集，同时提出一种基于 RGB-RGB 立体对的伪标签生成方法，用于后续模型训练。在方法层面上，RGB-RGB 立体匹配是一个经典的计算机视觉问题，经过数十年发展已经非常成熟，如今 SOTA 的深度学习方法在精度和泛化性都取得了惊艳的表现，在思想层面参考蒸馏学习的思想，将基于 RGB-RGB 的立体匹配方法作为教师模型，其生成的伪标签用于训练学生模型，即本课题提出的多光谱立体匹配模型，因此该内容调整是合理的；(ii) 研究内容二与开题报告中研究内容三对应，都是构建一种高效的多光谱模型，有一些具体内容有所调整。在实际的研究过程中发现，多光谱图像往往会出现图像分辨率不一致的情况，因此为了提升模型的通用性，增加了针对分辨率不一致问题的研究，扩大了模型的应用范围。在视差优化方面，由于硬件结构发生改变，原有视差优化路线不再适用，因此对视差优化环节做出了调整。

### 3. 目前论文主要研究内容的总体完成情况

#### (1) 多光谱硬件系统与数据集构建完成情况

现阶段已完成研究内容一：已搭建多光谱硬件系统，系统由两个 RGB 相机和一个 NIR 相机组成，型号分别为 OMNIVISION 公司生产的 AR0135 和 OV2770，

完成系统标定和数据采集；基于视差分布一致性的伪标签生成方法，首先利用现有成熟的 RGB-RGB 立体匹配方法(如 CREStereo)生成含有不可靠视差的伪标签，然后通过视差值、视差梯度和重投影一致性估计视差置信度，剔除不可靠视差结果，得到最终可靠的视差伪标签。

## (2) 多光谱立体匹配模型构建完成情况

现阶段已完成研究内容二主体部分，正在优化相关模块和补充实验：整体模型分为三个部分，分别是特征提取、特征插值和视差回归。(i) 针对多光谱图像的外观差异问题，在特征提取阶段，利用两个不对称特征提取器（即结构相同但不共享参数）提取多光谱特征，为了提升特征提取器从多光谱图像中学习具有一致性特征的能力，提出基于特征空间的特征度量一致性损失进行约束，构建具有一致性的特征空间。(ii) 针对多光谱图像的分辨率差异问题，提出一种基于特征插值的连续特征生成方法，同时为了恢复高分辨率全局信息，利用额外的特征提取器获得 RGB 图像的全局特征，与多光谱特征进行特征插值后拼接，用于后续视差回归，生成任意指定分辨率的视差结果；(iii) 在视差回归部分，设计两个带有残差结构的多层感知机进行视差回归，分别估计视差的整数部分和小数部分，最终生成具有亚像素精度的视差结果。

## 二、论文具体研究进展与所取得的阶段性成果

### 1. 多光谱硬件系统与数据集构建

#### (1) 硬件系统搭建

系统由两个 RGB 相机和一个 NIR 相机组成，型号分别为 OMNIVISION 公司生产的 AR0135 和 OV2770，RGB 相机分辨率为 1280\*720、焦距 3.6mm、成像光谱范围 380-800nm；NIR 相机分辨率为 640\*480、焦距 3.6mm、成像光谱 800-1000nm 实物如图 1 所示。



图 1: 硬件系统

## （2）系统标定与校正

系统标定即三个相机的内外参数标定，标定方法采用张正友标定法，使用本课题搭建的硬件系统同时拍摄棋盘格标定板，采集 18 组标定图像，以左相机坐标系为参考坐标系进行系统标定，随后利用 OpenCV 编写标定程序，获得相机内外参数，包括相机内参： $K_1$ 、 $K_2$ 、 $K_3$ ；相机之间旋转矩阵： $R_{1 \rightarrow 3}$ 、 $R_{2 \rightarrow 3}$ ；相机之间平移矩阵： $T_{1 \rightarrow 3}$ 、 $T_{2 \rightarrow 3}$ ；相机畸变参数： $D_1$ 、 $D_2$ 、 $D_3$ 。其中下标 1 表示右侧 RGB 相机，2 表示 NIR 相机，3 表示左侧 RGB 相机。标定完成后进行图像校正，首先利用标定获得的畸变参数进行图像去畸变，然后利用 Bouguet 立体校正方法将 RGB-RGB 立体对进行极线校正，使得对应点在同一水平线上，然后利用改进的 Fusiello 矫正方法将 NIR 的光心矫正到两个 RGB 光心所在直线，x 轴与 RGB 相机坐标系 x 轴平行，像素平面与 RGB 像素平面共面，流程图如图 2 所示。

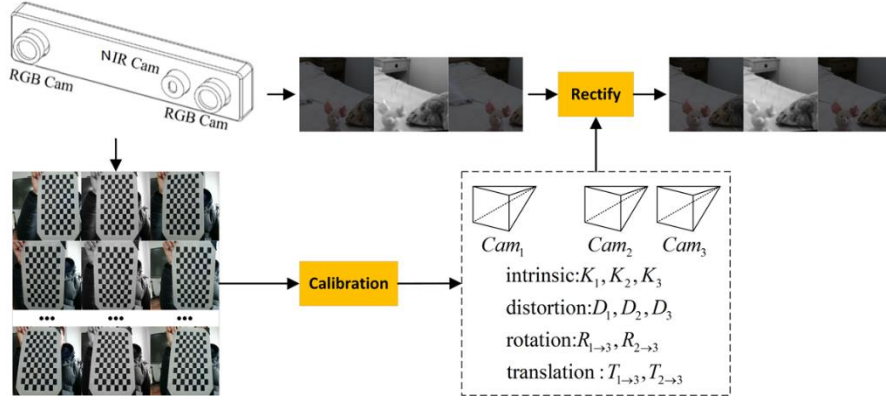


图 2：系统标定与校正

## （3）伪标签生成

现有多光谱数据集少，且视差真值的获取往往需要价格高昂的高精度传感器（如激光雷达）或者比较耗时的传统高精度方法（如时空编码结构光方法），真值获取代价高昂，针对这些问题，本课题提出一种基于视差分布一致性的伪标签生成方法，充分利用现有成熟的 RGB-RGB 立体匹配方法（如 CREStereo）估计出相对可靠但带有一定噪声的视差结果，然后经过重投影误差一致性、视差值以及视差梯度计算置信度掩码，剔除不可靠视差值，生成最终可靠的视差伪标签，如图 3 所示。

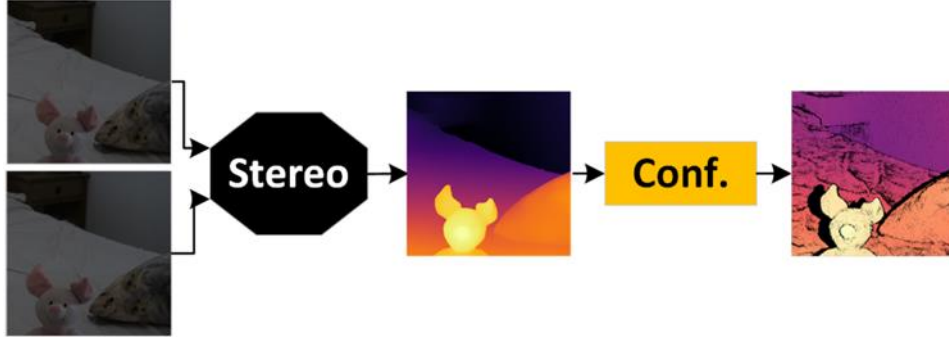


图 3：伪标签生成

图中 Stereo 模块为 RGB-RGB 立体匹配方法（如 CREStereo），Conf.模块为基于视差分布一致性的置信度计算方法，主要根据重投影误差一致性、视差值以及视差梯度计算置信度掩码：

重投影误差一致性置信度计算是基于光度一致性假设，即对应点的亮度  $I$  是一样的，假设以左图为参考图像，将右图按照估计的视差进行图像变换，生成左图像再与真正的左图像逐像素比较亮度误差值，误差越大则认为该点的视差值不可靠，反之视差值可靠，设重投影误差置信度为  $\omega_r$ ，重投影误差为  $\varepsilon = |I_L - I_{R \rightarrow L}|$ ，超参数为  $\delta_r$ ，其计算公式如式（1）：

$$\omega_r = e^{\frac{-\varepsilon^2}{2\delta_r^2}} \quad (1)$$

视差值置信度计算，根据一些先验知识，如室内场景的最大视差和最小视差有一个大概的区间，超过区间的视差值都认为是不可靠的，设视差值置信度为  $\omega_d$ ，估计的视差值为  $d$ ，可靠的视差区间为  $(d_{min}, d_{max})$ ，超参数为  $\delta_{d_1}, \delta_{d_2}$ ，计算公式如式（2）：

$$\omega_d = \begin{cases} e^{\frac{-|d-d_{min}|}{2\delta_{d_1}^2}} & , d < d_{min} \\ 1 & , d_{min} \leq d \leq d_{max} \\ e^{\frac{-|d-d_{max}|}{2\delta_{d_2}^2}} & , d > d_{max} \end{cases} \quad (2)$$

视差梯度置信度计算，一般某个物体的视差是平滑变化的，视差突变的区域往往发生在边缘区域，图像的大部分区域视差都是平滑变化的，而且现有方法在边缘视差估计上还存在一些问题，因此本课题认为视差梯度较大的部分不可靠。设视差梯度置信度为  $\omega_g$ ，视差梯度和逆视差梯度为  $\nabla d$ ， $\nabla(1/d)$ ，超参数为  $\delta_g$ ，



使用 sobel 边缘算子计算视差梯度，计算公式如式（3）：

$$\omega_g = e^{\frac{-\|\nabla d\|^2}{2\sigma_g^2}} \cdot e^{\frac{-\|\nabla(1/d)\|^2}{2\sigma_g^2}}, \quad (3)$$

$$\nabla d = \sqrt{sobel_x'^2 + sobel_y'^2}, \nabla(1/d) = \sqrt{sobel'_x'^2 + sobel'_y'^2}$$

最终的置信度为三个置信度的乘积，设置信度为  $\omega$ ，如公式（4）：

$$\omega = \omega_r \cdot \omega_d \cdot \omega_g \quad (4)$$

当置信度小于阈值  $T=0.5$  时则剔除该点视差，剩下的视差伪标签即为可靠的视差值。

#### （4） 实验结果

采用本课题搭建的多光谱硬件系统采集室内场景图像 11920 对图像，划分为训练集 10000 张，测试集 1920 张，为了验证 RGB-RGB 伪标签生成方法的有效性，使用时空编码结构光立体匹配方法获得 12 个场景的真值标签。部分数据集图像如图 4 所示，中间一列为 NIR 图像，方便展示将其缩放到相同分辨率，可以观察到 NIR 图像比 RGB 图像模糊，这是分辨率低造成的：



图 4：数据集图像

经过不断实验获得伪标签置信度估计的超参数如表 1 所示。伪标签生成需要先用 RGB-RGB 立体匹配方法生成初始伪标签，本课题尝试用不同的方法生成伪标签包括传统方法（如 SGM）和深度学习方法，带真值的 12 个场景评估各个方法的可靠程度，获得实验结果如下表 2，表中误差评估指标为视差平均终点误差

(Disparity Average End-Point-Error, D-AEPE).由表中结果可以看出 CRE-Stereo 精度较高, 且能达到伪标签的要求, 因此本课题后续使用 CREStereo 方法作为伪标签生成模型:

表 1: 置信度估计超参数

模块	参数	值
视差值	$d_{min}$	0.5
	$d_{max}$	64
	$\delta_{d1}$	0.1
	$\delta_{d2}$	0.5
视差梯度	$\delta_g$	0.5
重投影误差	$\delta_r$	0.1

表 2: 伪标签生成对比实验

方法	误差 (D-AEPE)
SGM	1.55
PSMNet	0.98
IGEV-Stereo	0.45
CRE-Stereo	<b>0.29</b>

伪标签可视化结果如图 5 所示, 图中黑色部分为被剔除的不可靠视差:

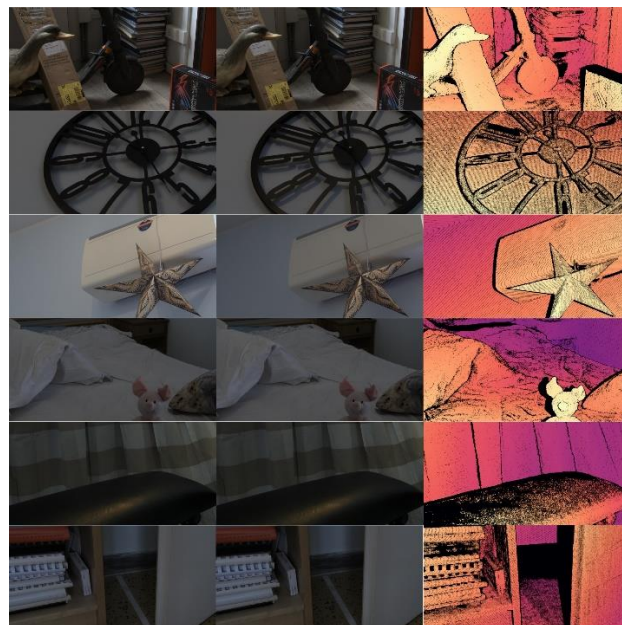


图 5: RGB-RGB 伪标签结果

## 2. 多光谱立体匹配模型构建

### (1) 特征提取阶段

光度一致性假设是多数常规多光谱立体匹配的前提，而多光谱立体对因成像光波长不同，导致图像外观差异大，即对应点亮度值差异大，因此不满足光度一致性假设，进而使得现有立体匹配方法失效。观察到，人的视觉系统有能力从多光谱图像中分辨出相同的物体，这说明多光谱图像中存在具有一致性的语义特征。基于此，为解决多光谱图像外观差异大的问题，本课题构建一致性特征空间，提出一种基于特征空间的特征度量一致性损失，用于约束特征提取器，提升其从多光谱立体对中学习出具有一致性的语义特征的能力，在特征空间克服多光谱外观差异问题，具体实现过程如图 6 所示。

考虑到 RGB 和 NIR 图像还存在分辨率和图像通道数差异，NIR 图像分辨率低且通道数为 1，因此先将 RGB 图像进行灰度化并缩放到和 NIR 图像分辨率一致，然后将缩放后的低分辨率多光谱立体对作为不对称特征提取器（结构相同，参数不共享）的输入，然后生成的一致性特征计算视差代价体，生成视差代价特征，同时本课题提出的特征度量一致损失用于约束不对称特征提取器，使其从多光谱图像中学习一致性特征，为了保留高分辨率 RGB 图像的全局信息，利用另一个全局特征提取器提取全局特征。设全局特征提取器为  $\Phi_\theta$ ，视差代价特征提取器为  $\Psi_\theta$ ，其中  $\theta$  为可学习参数，RGB 图像分辨率为  $H_\Phi \times W_\Phi$ ，NIR 图像分辨率为  $H_\Psi \times W_\Psi$ ，全局特征  $F_\Phi$ ，视差代价特征  $F_\Psi$  以及特征度量一致性损失为  $\mathcal{L}_{fm}$  的计算如公式 (5-7)：

$$F_\Phi, \mathcal{R}^{H_\Phi \times W_\Phi \times 3} \rightarrow \mathcal{R}^{H_\Phi \times W_\Phi \times F_\Phi} \quad (5)$$

$$F_\Psi, \mathcal{R}^{H_\Psi \times H_\Psi \times 2} \rightarrow \mathcal{R}^{W_\Psi \times H_\Psi \times D_\Psi} \quad (6)$$

$$\mathcal{L}_{fm} = \left\| f_{RGB_{gray\downarrow}} - f_{NIR_{L \rightarrow R}} \right\|_1 + \alpha \left( 1 - SSIM(f_{RGB_{gray\downarrow}}, f_{NIR_{L \rightarrow R}}) \right) \quad (7)$$

### (2) 特征插值阶段

由于多光谱图像分辨率不一致，导致全局特征和视差代价特征的尺度不一致，如图 6 中特征插值部分所示，进而无法进行后续的特征拼接，也无法估计任意指定分辨率的视差值。如图 7 所示，一个真实世界的模拟场景，可以被离散为

任意分辨率的离散数字图像，不同分辨率之间可以通过插值的方式来获得某点的像素值，如图中已知源图像，求目标图像中某点（如 A 点）的像素值，可以根据分辨率的差异获得原像素中 B 点的亚像素坐标，然后利用插值方法（如双线性插值）计算 B 点的像素值即可替代 A 点的像素值，计算过程如公式（8）。基于此，本课题提出一种基于特征插值的连续特征生成方法，生成任意指定分辨率的特征，用于后续视差回归。

$$A = \text{Bilinear}(B) \quad (8)$$

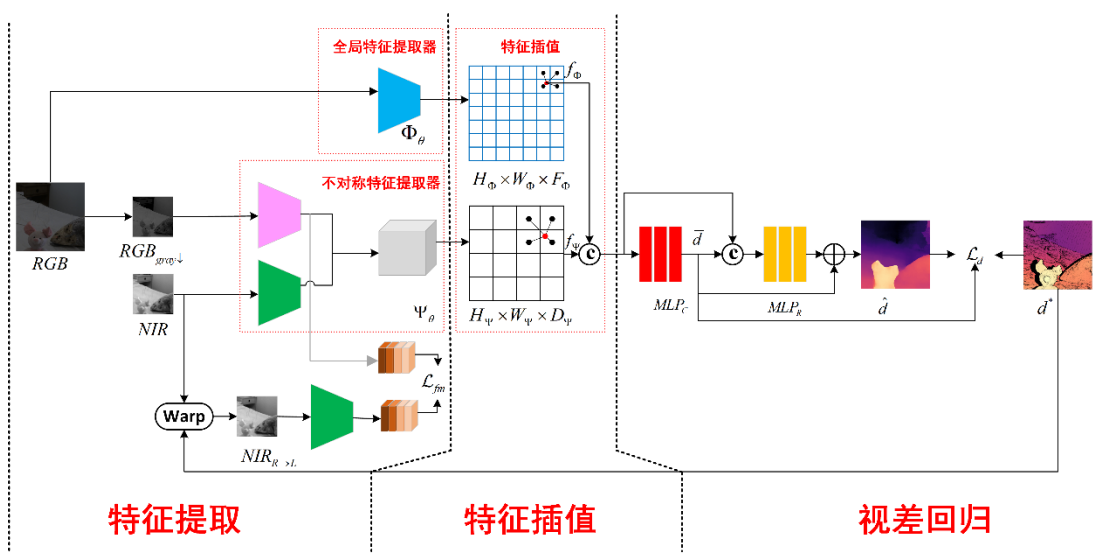


图 6: 多光谱立体匹配模型

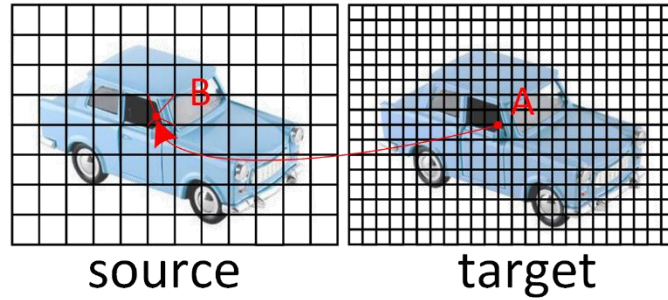


图 7: 特征插值示意图

### (3) 视差回归

如图 6 所示，经过特征插值后的全局特征  $f_\phi$  和视差代价特征  $f_\psi$  进行特征拼接形成新特征  $f_t$ ，其尺寸为  $H_t \times W_t \times (F_\phi + D_\psi)$ ，现有的视差回归方法普遍采用  $L_1$  损失进行约束，然而这可能会在视差不连续区域引起严重的过平滑效应，导致

在视差不连续区域生成的点云存在拖尾噪声。为了估计视差值准确且具有锐利的边缘，现有方法针对前景和背景分别建模，在边缘处区分前景和背景，这种方法效果显著但建模复杂。本课题利用其思想，采用一种更简单但高效的策略来缓解过平滑问题，如图 6 所示，本课题利用两个多层感知机来完成视差回归：第一个多层感知机  $MLP_C$  用于估计视差类别分布，即估计整数部分视差，将视差估计作为分类任务从视差范围  $(0, d_{max})$  中进行分类；第二个多层感知机  $MLP_R$  用于估计亚像素偏移部分即取值范围是  $[-1, 1]$ ，两个 MLP 估计的结果之和即为最终亚像素视差结果，两个 MLP 之间具有残差结构，都接受特征  $f_T$  作为输入。设整数视差分布为  $\bar{d}$ ，亚像素偏差为  $\dot{d}$ ，最终视差值为  $\hat{d}$ ，计算过程如公式 (9-11)：

$$\bar{d} = \arg \max(MLP_C(f_\Psi, f_\Phi)) \quad (9)$$

$$\dot{d} = MLP_R(f_\Psi, f_\Phi, \bar{d}) \quad (10)$$

$$\hat{d} = \bar{d} + \dot{d} \quad (11)$$

$MLP_C$  作为分类任务以交叉熵损失进行监督，而  $MLP_R$  则使用  $L_1$  损失进行监督，设视差损失函数为  $\mathcal{L}_d$ ，视差标签为  $d^*$  计算如公式 (12)：

$$\mathcal{L}_d = -\mathcal{N}(d^*, \sigma) \cdot \log(MLP_C(f_\Psi, f_\Phi)) + |MLP_R(f_\Psi, f_\Phi, \bar{d}) + \bar{d} - d^*| \quad (12)$$

其中  $\mathcal{N}(d^*, \sigma)$  为以  $d^*$  为中心的高斯分布，方差  $\sigma = \sqrt{2}$ 。结合特征度量一致性损失后，设总损失函数为  $\mathcal{L}$ ，有：

$$\mathcal{L} = \mathcal{L}_d + \beta \mathcal{L}_{fm} \quad (13)$$

#### (4) 实验结果

训练数据集采用本课题采集的多光谱数据集，包含训练集 10000 张，测试集 1920 张，RGB 图像输入为 1280\*720，NIR 图像输入为 640\*480，伪标签视差有效值不少于 70%，防止过于稀疏的监督。上述网络架构使用 PyTorch 编码实现，优化器使用 Adam，优化器超参数设置为  $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ 。代价体骨干网络使用 PSMNet，学习率设置为 0.001，批次大小设置为 2，在两张 NVIDIA 2080Ti GPU 训练 70 epochs。在测试集上的预测结果如图 8 所示：

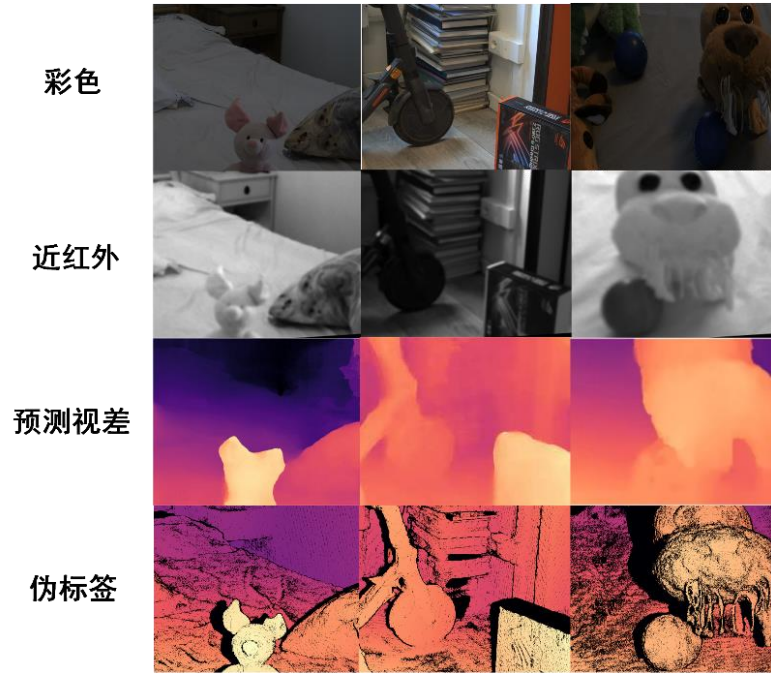


图 8: 预测结果可视化

由图 8 可以观察到预测的视差结果在边缘部分比较锐利, 整体的视差分布正确, 为了进一步评估该模型的有效性, 还需要对预测结果进行定量实验分析, 评估指标为视差平均终点误差 (Disparity Average End-Point-Error, D-AEPE) 和绝对深度误差 (Absolute depth error , ADE)。并且对比了常规 RGB-RGB 方法 PSMNet 和多光谱 RGB-NIR 方法 DMC, 实验结果如图 9 和表 3 所示, 结果表明我们的方法是有效的。

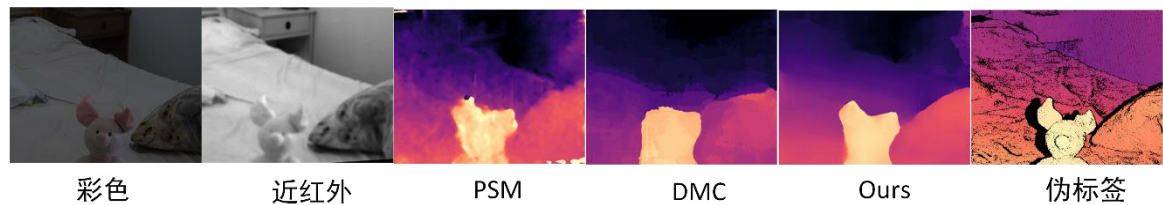


图 9: 对比实验可视化

表 3: 对比实验定量分析

方法	D-AEPE	ADE
PSM	1.451	0.103
DMC	1.945	0.116
<b>Ours</b>	<b>0.731</b>	<b>0.057</b>

### 3. 研究成果

[1] 张佳预, 贾同. 基于 RGB-D 图像序列的大场景三维重建方法研究[J]. 中国传媒大学学报(自然科学版). 2023,30(01):53-65.

[2] 贾同, 张佳预. 多模态监控相机感知监控系统 V1.0 软著 登记号: 2022SR0965725.

### 4. 相符情况

综上所述, 目前已完成研究内容与开题报告所定的研究计划是相符的。

## 三、后续论文研究内容与研究计划

### 1. 后续研究内容

(1) 研究本课题系统于与数据集构建的改进方法。目前系统标定和校正精度还有进一步的改进空间, 同时伪标签生成方法还可以尝试更多方法进行对比;

(2) 研究本课题在现有模型基础上的进一步提升方法, 如在特征提取模块, 可以尝试引入注意力机制, 在光谱差异大的区域和光谱差异小的区域赋予不同的权重, 提升特征提取器学习一致性特征的能力;

(3) 补充实验内容, 全面验证本课题所提方法的有效性, 在对比实验方面需要增加更对对比方法, 包括传统方法(如 SGM)和多光谱深度学习方法, 在消融实验方面还需要增加对各个模块的消融对比, 证明模块的有效性。

### 2. 研究计划

- (1) 2023 年 12 月, 进一步优化为标签生成方法;
- (2) 2024 年 1-2 月, 进一步优化不对称特征提取器;
- (3) 2024 年 2-3 月, 进一步优化视差结果;
- (4) 2024 年 3-5 月, 扩充实验, 整理研究成果, 撰写毕业论文。

## 四、目前存在的问题与解决方法

(1) **问题:** 伪标签生成质量可以进一步优化。

**解决方法:** 在系统层面上, 由于存在三个相机且分辨率也有区别, 系统标定和校正都存在一定误差。针对标定可以使用圆环标定板, 进一步减小标定和校正误差; 在方法层面, 尝试更多 RGB-RGB 立体匹配方法包括传统方法如(PatchMatch,



AD-Census 等)和深度学习方法,包括现有 SOTA 的有监督和无监督方法,综合评估这些方法的生成质量;置信度计算方面考虑更多可能影响因素。

**(2) 问题:** 不对称特征提取器可以进一步优化

解决方法: 尝试引入注意力机制,区分外观差异大和差异小的区域,提升特征提取器的表征能力。

**(3) 问题:** 视差结果有进一步提升空间

解决方法: 尝试使用迭代优化型视差回归方式,引入门控循环单元 (GRU) 优化视差结果。

**(4) 问题:** 实验不充分

解决方法: 补充实验内容,增加消融实验,扩充对比实验。在消融实验方面还需要增加对各个模块的消融对比,证明模块的有效性;在对比实验方面需要增加更对对比方法,包括传统方法 (如 SGM) 和多光谱深度学习方法。

## 五、论文的预期成果

- (1) 实现一套多光谱立体匹配硬件系统;
- (2) 实现实现伪标签生成方法,构建多光谱数据集;
- (3) 实现高效无监督多光谱立体匹配模型算法代码;
- (4) 实现多光谱立体匹配模型验证,并应用于下游应用;
- (5) 发表一篇 SCI 论文。

导师签字: 贾同