

# 双目立体视觉技术及其在智慧农业中的应用研究进展

杨学军, 钟金彪, 林开颜\*, 吴军辉, 陈杰, 司慧萍

(同济大学现代农业科学与工程研究院, 上海 201804)

**摘要:** 双目立体视觉技术作为一种获取现实世界立体感知的重要方法, 可以实现农业场景目标的三维定位和点云重建, 提取立体三维信息, 具有很大的应用潜力。该文介绍了双目立体视觉技术的应用流程, 包括双目视觉标定、极线校正、立体匹配等方面, 以及相关研究进展; 综合近期文献, 探讨了双目立体视觉技术在果实定位采摘与地图导航、生长参数测量和病害识别与施药等农业领域最新的应用。综述结果表明, 双目立体视觉技术在农业中定位、测量和识别均具有较高精度, 但仍面临模型复杂、场景受限、数据集少和立体匹配缺少评价标准等问题。展望该技术在农业领域的未来发展, 应着重从算法设计与优化、智能辅助平台搭建、数据集构建和评价体系完善等方面开展研究探索。

**关键词:** 双目视觉; 智慧农业; 立体匹配; 作物参数测量; 三维目标定位; 病害识别

doi: 10.11975/j.issn.1002-6819.202410177

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2025)-01-0027-13

杨学军, 钟金彪, 林开颜, 等. 双目立体视觉技术及其在智慧农业中的应用研究进展[J]. 农业工程学报, 2025, 41(1): 27-39. doi: 10.11975/j.issn.1002-6819.202410177 <http://www.tcsae.org>

YANG Xuejun, ZHONG Jinbiao, LIN Kaiyan, et al. Research progress on binocular stereo vision technology and its applications in smart agriculture[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(1): 27-39. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202410177 <http://www.tcsae.org>

## 0 引言

随着人工智能、机器视觉等先进技术的融入, 农业正逐步迈向智能化、精准化的新时代。近年来, 机器视觉特别是立体视觉技术在农业领域的广泛应用, 极大地促进了智慧农业的发展。立体视觉技术在获取目标三维信息后, 进行分析和建模可以解决农业领域目标定位、测量、识别等难题。获取目标三维信息的方法有主动式和被动式两种。主动式方面, 有结构光、三维激光扫描和飞行时间等方法。被动式方面, 有单目视觉、双目视觉和多目视觉等方法<sup>[1]</sup>。随着深度学习技术的发展, 端到端的深度学习方式也可直接获取目标的三维信息。综合考虑成本、精度和技术复杂性, 双目视觉法已成为目前获取三维信息的重要方式。

双目立体视觉技术通过模仿人类视觉系统对环境进行三维感知<sup>[2]</sup>, 能够拍摄景深图像, 获取目标深度信息和转换处理后的点云数据, 为农业生产的各个环节提供精准的数据支持<sup>[3]</sup>。基于双目立体视觉技术进行农作物场景的三维模型重建, 能提升立体环境感知能力, 结合误差调整, 可完成智慧农业中常见的机器视觉任务, 如作物生长参数的无损检测<sup>[4]</sup>、病害目标识别以及采集机器人对果实的精准定位<sup>[5]</sup>等。双目立体视觉技术的应用将有助于提高智慧农业生产效率, 节省生产成本, 对智

慧农业的发展具有重要意义。本文通过总结与归纳双目立体视觉技术的研究现状, 对双目视觉的原理、关键技术特别是基于深度学习的立体匹配进行了重点介绍, 之后阐述了其在农业领域的应用, 并针对存在的问题与面临的挑战, 对双目立体视觉技术未来研究趋势进行展望。

## 1 双目立体视觉技术研究进展

双目立体视觉技术基本原理如图1所示, 首先由标定后的两台相机对某一区域内的物体进行图像采集, 然后经过一系列图像处理步骤如极线校正、立体匹配得到视差图, 其中立体匹配包括匹配代价计算、代价聚合、视差计算和视差优化4个步骤<sup>[6]</sup>。视差图通过三角测量原理可计算目标的深度图, 进而根据点云变换构造丰富的三维点云<sup>[7]</sup>。

### 1.1 双目相机标定

利用双目相机获取的左右图像, 计算空间点三维坐标的三角测量原理如图2所示。

依据双目立体视觉的透视变化, 可得双目相机采集的像素点  $p_2(x_2, y_2)$ 、 $p_1(x_1, y_1)$  与三维空间点  $P(X, Y, Z)$  之间相互转换的关系, 如式(1)所示:

$$\begin{cases} y_1 = y_2 = f \frac{Y}{Z} \\ x_1 = f \frac{X}{Z} \\ x_2 = f \frac{X-D}{Z} \end{cases} \quad (1)$$

利用式(1)的等式关系转换即可计算  $P$  的三维坐标, 如式(2)所示:

收稿日期: 2024-10-24 修订日期: 2024-12-07

基金项目: 上海市科委科技创新行动计划课题 (23N21900400)

作者简介: 杨学军, 博士, 副教授, 研究方向为农业生态、城市绿化等应用生态研究。Email: [espi@tongji.edu.cn](mailto:espi@tongji.edu.cn)

\*通信作者: 林开颜, 博士, 副教授, 研究方向为农业计算机视觉。

Email: [linkaiyan@tongji.edu.cn](mailto:linkaiyan@tongji.edu.cn)

$$\begin{cases} X = \frac{Dx_1}{x_1 - x_2} \\ Y = \frac{Dy_1}{x_1 - x_2} \\ Z = \frac{Df}{x_1 - x_2} \end{cases} \quad (2)$$

式中  $x_1 - x_2$  为左右图像的视差, 根据立体匹配所得的视差图即可计算出目标的三维坐标, 大量三维坐标点集合即构成目标三维点云。

由于相机畸变会影响成像的准确性, 必须事先对相机完成标定和校正, 求解出相机内外参数, 使其能够进行坐标系转换。双目相机采集的目标从三维世界坐标系经过相机坐标系映射至二维平面图像坐标系并平移至像素坐标系, 如式 (3) 所示:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \xi \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ = \xi \mathbf{M} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

式中  $\xi$  为从世界坐标系到相机坐标系的缩放比例; 由于世界坐标系与相机坐标系之间是一个刚体变换, 通过旋转与平移即可实现, 用  $\mathbf{R}$  表示  $3 \times 3$  的旋转变换矩阵,  $\mathbf{t}$  表示  $3 \times 1$  的平移变换向量;  $c_x$ 、 $c_y$  分别表示点在图像坐标系平移至像素坐标系横、纵方向的平移因子;  $f_x$ 、 $f_y$  分别表示焦距  $f$  在横、纵方向上的像素单位值;  $\mathbf{M}$  为双目相机标定的内参矩阵,  $\mathbf{R}$  和  $\mathbf{t}$  为外参矩阵。相机标定就是通过多组采集图像来求解内参矩阵、外参矩阵以及畸变系数, 建立点在像素坐标与世界坐标之间映射关系的过程。一般标定方法可分为参考物标定法、主动视觉标定法和自标定法, 近年来一些研究人员也提出了神经网络标定方法, 4 种方法的对比如表 1 所示。

表 1 标定方法对比

Table 1 Comparison of calibration methods

标定方法 Calibration methods	关键技术 Key technologies	优点 Advantages	缺点 Disadvantages
参考物标定 Reference calibrations	用平面棋盘格标定板 <sup>[8]</sup> 将图案投影到纯面板上, 基于相位标定 <sup>[9]</sup> 考虑相机和标定物体之间存在介质 <sup>[10]</sup> 用标定杆扩大标定视野 <sup>[11]</sup> 控制双目相机平行移动 <sup>[12]</sup>	高精度, 稳定性高	依赖标定物, 场景受限
主动视觉标定 Active visual calibrations	配备双目相机的六自由度机器人运动 <sup>[13]</sup> 设计外部电路控制相机标定系统旋转 <sup>[14]</sup>	灵活性高, 鲁棒性强	成本高, 设备要求高
自标定 Self-calibrations	结合通用多项式和等距投影模型 <sup>[15]</sup> 对不同姿势拍摄的图像进行非线性优化 <sup>[16]</sup> 共形几何代数原理推断相机位置 <sup>[17]</sup>	需要的图像数量少	稳定性和鲁棒性弱
神经网络标定 Neural network calibrations	双神经网络结合粒子群优化 <sup>[18]</sup> 基于径向基函数神经网络 <sup>[19]</sup> 基于深度强化学习拟合 <sup>[20]</sup>	自适应能力好, 无需建立复杂模型	可解释性差, 依赖训练数据

参考物标定法采用已知形状和尺寸的参考物体 (如棋盘格) 作为相机的目标, 通过图像处理和数学变换, 可以得到相机的内外参数。参考物标定具有高精度但需要易于识别和提取特征的参考对象; 主动视觉标定法通过精确控制相机工作台的运动 (如平移、旋转), 相机

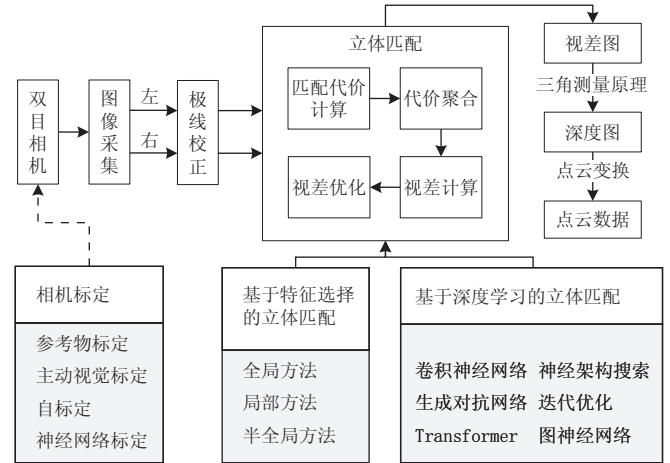
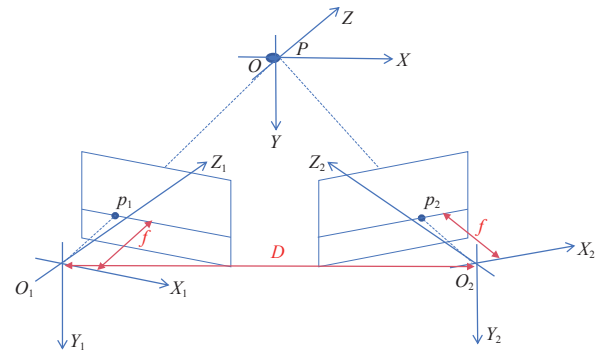


图 1 双目立体视觉技术基本原理

Fig. 1 Basic principle of binocular stereo vision technology



注:  $O-XYZ$ 、 $O_2-X_2Y_2Z_2$ 、 $O_1-X_1Y_1Z_1$  分别为世界坐标系、右相机坐标系、左相机坐标系;  $f$  为双目相机的焦距,  $m$ ;  $D = O_2 - O_1$  为双目相机的基线,  $m$ ;  $P(X, Y, Z)$ 、 $p_2(x_2, y_2)$ 、 $p_1(x_1, y_1)$  分别为  $O-XYZ$ 、 $O_2-X_2Y_2Z_2$ 、 $O_1-X_1Y_1Z_1$  中的同一点。

Note:  $O-XYZ$ 、 $O_2-X_2Y_2Z_2$  and  $O_1-X_1Y_1Z_1$  represent the world coordinate system, the right camera coordinate system, and the left camera coordinate system, respectively;  $f$  is the focal length of the stereo camera,  $m$ ;  $D = O_2 - O_1$  is the baseline of the stereo camera,  $m$ ;  $P(X, Y, Z)$ 、 $p_2(x_2, y_2)$  and  $p_1(x_1, y_1)$  represent the same point in  $O-XYZ$ 、 $O_2-X_2Y_2Z_2$  and  $O_1-X_1Y_1Z_1$ , respectively.

图 2 双目相机成像示意图

Fig. 2 Binocular camera imaging diagram

从多个角度捕获图像, 然后利用相机运动与捕获图像之间的约束关系求解相机参数。主动视觉标定考虑了相机的径向畸变、切向畸变和薄透镜畸变, 但对设备有一定的要求, 需要严格控制相机移动; 自标定法无需标定物, 仅通过分析相机在不同位置和姿态下拍摄的图像序列估

算相机的内外参数，图像数量需求少，但算法稳定性和鲁棒性弱，依赖于非线性优化的初始条件；神经网络标定法利用神经元之间的非线性映射来建模相机坐标、世界坐标和像素坐标之间的函数关系，无需求解相机的内外参数，通过训练确定神经元的权重完成标定，具有良好的自适应能力，但依赖大量的高质量训练数据，而训练数据的采集也需要高精度的仪器。

上述标定方法适用于双目和单目相机。对于双目相机，需要确定左右相机的相对位置和方向，并在使用中保持不变，同时目标图像需同步采集；单目相机则仅需标定单个相机。双目相机完成标定后，即可开始采集左右目标图像。

## 1.2 极线校正

极线校正的目的是利用极线约束将匹配点的搜索范围由二维降为一维，以节省计算量，排除虚假匹配点。满足极线约束的匹配点是指三维空间中的一点，与其在左右图像上的两个投影点及左右相机的中心点这五点处于同一平面内。因此对于左图像的某个特征点，与其在右图像上的对应点必然处在对极线上。然而，由于相机畸变的存在，直接拍摄的图像很难满足极线约束。需要对图像进行畸变校正，使得左图上的点及其在右图的对应点必然具有相同的行号，这样在搜索某一点的对应点时，只在行内进行一维搜索即可。

Bouguet 校正方法<sup>[21]</sup>是一种经典的校正算法，也有研究者在外极几何理论上提出了一些新的极线校正方法。如 DARMON 等提出了一种以外极为中心的极变换校正

方法<sup>[22]</sup>，它对任何一个极点或两个极点处于图像域内都适用，但是该方法会导致生成的图像失真，即原始图像中的直线弯曲。ELHASHASH 等提出了用于外延极线校正的球面模型<sup>[23]</sup>，以最大限度地减少由于误差引起的失真。

## 1.3 立体匹配

立体匹配是指寻找两图像中相同物体或场景的对应点，从而计算出双目图像所对应的视差图。视差为同一物体在两个图像中投影点之间的水平距离，由式(2)可知它与物体到相机的距离(深度  $Z$ )成反比，因此可以用视差来估计物体的深度。传统基于 SURF (speeded up robust features)<sup>[24]</sup> 和 ORB (oriented fast and rotated brief)<sup>[25]</sup> 特征的匹配算法均先提取密集特征点，后采用暴力搜索、最近邻搜索或 KD 树 (k-dimension tree) 等方法寻找匹配点，并经过筛选剔除得到最佳匹配点对，这类方法匹配像素的搜索具有盲目性。后续大多数立体匹配算法的设计<sup>[26-27]</sup> 均基于 SCHARSTEIN 等<sup>[6]</sup> 提出的 4 个步骤：代价计算、代价聚合、视差计算(估计)、视差优化，其中匹配代价通常采用绝对差、平方差等方式来衡量，代价聚合和视差优化通常被看作是 2D 图分割问题而使用图分割或置信传播算法来优化。目前立体匹配算法较多，按原理主要可分为两大类，基于特征选择的匹配算法和基于深度学习的匹配算法。

### 1.3.1 基于特征选择的匹配算法

立体匹配时，基于特征选择的匹配算法通过像素或特征点之间的相似性来衡量是否匹配，根据匹配像素的范围，可分为局部方法、全局方法和半全局方法，如表 2 所示。

表 2 基于特征选择的立体匹配方法  
Table 2 Stereo matching method based on feature selection

立体匹配方法 Stereo matching methods	关键技术 Key technologies	优点 Advantages	缺点 Disadvantages
局部方法 Local methods	基于区域匹配，根据自适应跨区域的结构特征并计算正交权重 <sup>[28]</sup> 基于特征匹配，根据潜在的特征对应关系进行相似性度量 <sup>[29]</sup> ，或使用非相似性度量减少系数组的操作数量来提升速度 <sup>[30]</sup>	速度快，实时性较高	弱纹理区域精度较低，难以处理大范围视差
全局方法 Global methods	基于相位匹配，将匹配成本在固定的窗口汇总 <sup>[31]</sup> 基于图像分割匹配，在 Census 变换后将汉明距离作为能量函数数据项中的相似性度量 <sup>[32]</sup> 基于置信度匹配，建立估计置信度的网络结构 <sup>[33]</sup> 基于梯度匹配，用分割树的方法进行代价聚合 <sup>[34]</sup>	对弱纹理区域友好	速度慢，细节易丢失
半全局方法 Semi-global methods	以自适应窗口大小改变匹配代价计算的方式 <sup>[35]</sup> 结合 Census 变换和颜色信息来计算初始匹配代价，后将多路径代价引入多尺度的代价聚合中 <sup>[36]</sup> 引入最小池化代价聚合，捕获更多全局信息并减少噪声来增强视差图的连续性 <sup>[37]</sup>	精度和速度平衡	相比前两种复杂

局部方法通过在周边局部区域查找相似度最高的点，实现左右视图的匹配，常见的有基于区域、特征和相位的算法。由于只考虑局部信息，局部匹配方法速度较快，但是难以处理大范围视差；全局方法基于全局能量函数，综合数据项(测量像素之间的相似性)和平滑项(平滑像素之间的视差关系)逐步细化求解，常见的有基于图像分割、置信度和梯度的算法。全局匹配方法优缺点正好和局部方法相反，速度较慢，但是其考虑了全局信息，细节处理能力有限；半全局方法通过在多个方向上优化能量函数的路径形式，在每个方向上按照动态规划<sup>[38]</sup>的思想进行能量累积，然后将各个方向上的匹配代价相加得到总的匹配代价。相对前两种方法来说，其实现了匹配结果质量和计算复杂性之间的平衡。

### 1.3.2 基于深度学习的匹配算法

研究表明，基于特征选择的立体匹配方法在弱纹理

或重复纹理以及遮挡区域表现不佳。近年来，随着机器学习的发展，立体匹配算法已经成为一项深度学习任务，基于深度学习的立体匹配方法近年来取得了显著的进展<sup>[39]</sup>，根据其网络框架可以分为：基于卷积神经网络 (convolutional neural network, CNN) 的方法、基于生成对抗网络 (generative adversarial network, GAN)<sup>[40]</sup> 的方法、基于 Transformer<sup>[41]</sup> 的方法及其他方法，方法对比如表 3 所示，其中精度和复杂度比较来源于现有立体匹配公开数据集 Scene Flow<sup>[42]</sup> 和 KITTI2015<sup>[43]</sup> 的 State-of-the-Art 模型结果。

1) 基于 CNN 的方法。由于 CNN 具有自动学习图像局部特征的能力，通过一系列卷积层对双目图像进行卷积运算可计算其匹配代价，相比于基于特征选择的方法计算像素之间的相似度来衡量匹配代价，其可计算更复杂的图像特征信息。如图 3 所示，基于 CNN 的方法流程大致为：双目图像分别提取多维度特征后构建代价体，

代价体通常是一个三维或四维矩阵,代表某一像素对在不同视差值下的匹配代价,且依赖于预定义的视差范围;再经过正则化卷积神经网络实现代价聚合,正则化网络的卷积维度取决于代价体的维度,三维代价体对应2D卷积,四维代价体对应3D卷积;进而使用soft-argmin等方式回归初始视差,最后根据左图的上下文网络实现视差细化得到最终视差。ŽBONTAR等<sup>[44]</sup>首次从二分类的角度将CNN用于匹配代价计算,以激光雷达获得的视差图为标签,将左右图像切分为 $9 \times 9$ 像素的图像块输入卷积网络,经softmax层输出后预测双目图像是否匹配,并以负类的输出值作为匹配代价,在交叉邻域内进行代价聚合,以使得代价最小的视差作为视差结果。经典的方法MC-CNN<sup>[45]</sup>、GC-Net<sup>[46]</sup>分别以2D和3D卷积来计算匹配代价,比使用特征选择方法准确度大幅提升,GC-Net网络中首次引入了代价体的概念,并构造3D卷积编码器和卷积解码器<sup>[47]</sup>从高度、宽度、视差3个维度来学习几何结构和上下文信息<sup>[48]</sup>以产生最终的代价体,最后

以回归的角度在代价体视差维度上实现视差估计。后续XU等<sup>[49]</sup>提出的代价体上采样模块可以无缝嵌入到许多现有的立体匹配网络中,例如GC-Net、PSM-Net<sup>[50]</sup>和GA-Net<sup>[51]</sup>,使得网络在保持精度的同时,速度提升数倍。但在无纹理或弱纹理区域,卷积编码器和卷积解码器的结构在恢复图像细节方面存在局限,进行立体匹配实现视差计算仍然很困难。一些研究者设计了金字塔型的分层网络结构<sup>[52]</sup>,以粗到精的方式从特征中获取全局和局部信息计算匹配代价,通过组合每层包含不同信息的视差图输出最终的视差图。也有研究者从设计更高维度代价的角度来提升算法性能,KOU等提出使用四维代价体<sup>[53]</sup>,WANG等提出构建多尺度代价体<sup>[54]</sup>,以实现更稳健的匹配,并结合分层3D卷积分支来提高实时性能,以便在训练过程中更好地聚合代价。基于CNN的方法虽然大幅度提升精度,但是随着卷积维度的提升,卷积计算成本也会随之增加;并且卷积运算感受野有限,无法获得广泛的上下文信息。

表3 基于深度学习的立体匹配算法比较  
Table 3 Comparison of stereo matching algorithms based on deep learning

方法 Methods	优点 Advantages	缺点 Disadvantages	适用农业场景 Agricultural applicability
卷积神经网络 Convolutional neural network	精度较好、鲁棒性较好,模型大小取决于网络的深度和卷积核的大小	难以处理大范围视差,需要大量真实视差图	复杂背景和光照变化的场景
生成对抗网络 Generative adversarial network	可生成视差图	精度较差,模型较复杂,结果可解释性较差	数据集难获取的场景,用于前期数据扩充
Transformer	精度高、鲁棒性好,能结合上下文和长依赖信息	模型复杂	大场景、大范围视差和复杂纹理场景
神经架构搜索 Neural architecture search	可自动设计网络架构,适应不同资源限制,精度较好	搜索成本高,扩展性有限	需要不同资源配置、自动化建模场景
迭代优化 Iterative optimization	资源节省,适用于高分辨率图像,可处理大范围视差	计算资源消耗与迭代次数成正比,可能导致处理时间较长	大视差范围和高分辨率图像的精细匹配场景
图神经网络 Graph neural network	精度高,全局特征表示能力强	图结构复杂,需大量资源存储特征间关系	高遮挡、复杂纹理的场景

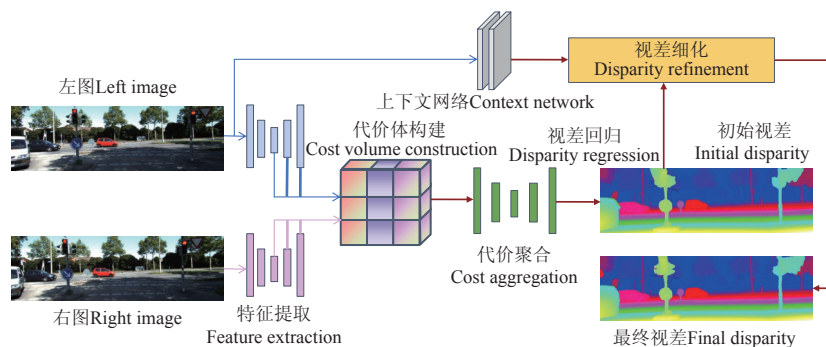
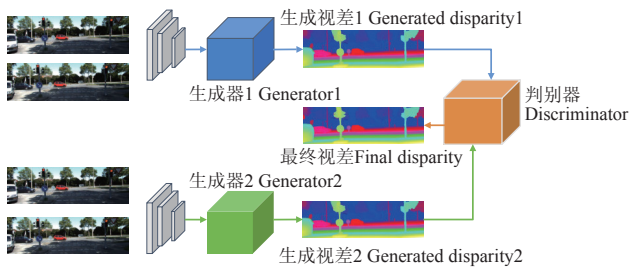


图3 基于卷积神经网络的立体匹配框架  
Fig.3 Stereo matching framework based on convolutional neural network

2) 基于GAN的方法。基于CNN的双目立体匹配方法取得了显著进展,但仍受限于双目数据集中真实视差图难以获取。为了缓解真实视差图数据不足的问题,研究者利用GAN方法能生成“伪”数据的特点,提出循环生成对抗网络(cycle-GAN)<sup>[55]</sup>和条件生成对抗网络(conditional-GAN)<sup>[56]</sup>等方法。如图4a所示,基于cycle-GAN的立体匹配方法由两组对称的部分组成,每部分的生成模型先从不同视角生成视差图,上半部分和下半部分生成的视差图通过判别模型相互约束输出最终精准的视差图,因未使用真实视差图标签,所以其从无监督学习的角度解决数据集不足的问题。如图4b所示,

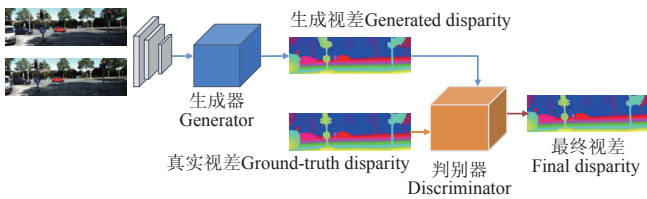
基于conditional-GAN的立体匹配方法仅由一个生成模型和判别模型组成,双目图像经提取特征后输入生成模型,生成模型利用其生成尽可能与视差图标签相近的视差图,判别模型以其为条件区分生成视差图和真实视差图标签,以极小数据集使得生成模型达到以假乱真的效果。从扩充数据集的角度,一些研究者在立体匹配生成的视差图后采用GAN网络,根据估计的视差和右侧图像生成新的合成左侧图像,判别模型区分合成图像和真实图像,并将其差异作为损失函数引入立体匹配网络以估计更准确的视差图<sup>[57]</sup>。上述研究者提出的使用合成数据进行训练的方法,不可避免地引入合成域与真实域之间的域鸿

沟问题，许多研究采用图像翻译作为域适应的解决方案，由于图像翻译并没有考虑图像的几何约束，使得翻译后的图像纹理缺少一致性，进而引起视差歧义。为解决这一问题，LIU 等提出了 StereoGAN 网络<sup>[58]</sup>，在域翻译模块和立体匹配模块之间建立双向连接，并通过两者联合优化来保持转换过程的极线约束关系不变。厉行等<sup>[59]</sup>在 StereoGAN 的基础上提出在域翻译模块融入边缘信息使生成图像更加真实，在立体匹配模块引入法线损失来捕获更多局部视差变化细节。基于 GAN 的方法通常较为复杂，涉及生成模型和判别模型的设计与优化，可解释性也较差。



a. 基于循环生成对抗网络的立体匹配框架

a. Stereo matching framework based on cycle generative adversarial network



b. 基于条件生成对抗网络的立体匹配框架

b. Stereo matching framework based on conditional generative adversarial network

图 4 基于生成对抗网络的立体匹配框架

Fig.4 Stereo matching framework based on generative adversarial network

3) 基于 Transformer 的方法。Transformer 的自注意力机制<sup>[60]</sup>能够捕捉双目图像之间的中长距离依赖，特征图能够结合全局信息并根据彼此之间的相似性进行加权计算匹配代价，可解决 CNN 在中长距离像素匹配感受野受限的问题。如图 5 所示，与 CNN 网络结构相近，其以 Transformer 替换了代价体构建和代价聚合模块<sup>[61]</sup>，将立体匹配表述为序列到序列问题，自注意力机制 (self-attention) 在特征内部学习全局上下文信息<sup>[62]</sup>，交叉注意力机制 (cross-attention) 整合自注意力机制的信息并建立匹配关系。因其不使用代价体，所以无需初始设定视差范围约束，对于难以事先估计所有可能视差值的动态或复杂场景，可以自动调整匹配策略实现大范围视差的估计。典型算法 STTR (stereo transformer)<sup>[63]</sup>使双目图像经提取特征后输入 Transformer，通过反复叠加自注意力机制和交叉注意力机制以增强像素间的差异，并在另一幅图像上寻找与其最接近的像素<sup>[64]</sup>，以实现更灵活的匹配产生原始视差估计，进而将原始图像的上下文信息与视差图融合，并通过卷积对结果进行修正生成最终视差图。也有研究者结合 CNN 和 Transformer 各自的优点，使用 Transformer 进行特征学习、CNN 进行代价聚合<sup>[65]</sup>，加快模型收敛速度、提高准确性和泛化能力。如 ELFNet<sup>[66]</sup>融合了以 CNN 为代表的 PCWNet 框架<sup>[67]</sup>和以 Transformer

为代表的 STTR 框架，融合模块有效地集成了来自 PCWNet 的局部信息和来自 STTR 的全局信息，实现预测结果的互补。遮挡区域视差估计的逐步细化是基于 Transformer 方法的另一个重要应用，GOAT 网络引入了并行视差和遮挡估计模块<sup>[68]</sup>，可互不干扰地估计视差和遮挡，并且利用遮挡区域对视差估计结果进一步细化。然而 Transformer 模型由于自注意力计算需要较大参数数量，导致计算成本更大，简化 Transformer<sup>[69]</sup>以实现更快的速度和有竞争力的性能也成为目前的一个方向。

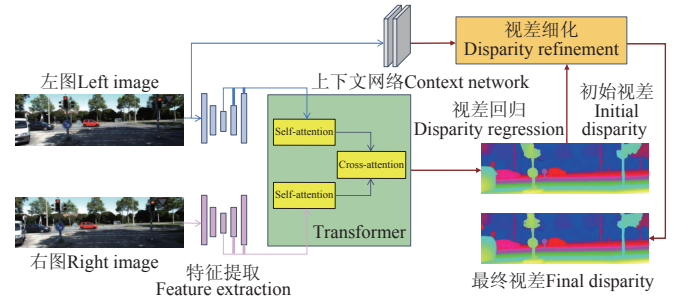


图 5 基于 Transformer 的立体匹配框架

Fig.5 Stereo matching framework based on Transformer

4) 其他方法。研究较为广泛的还有基于神经架构搜索 (neural architecture search, NAS)、基于迭代优化 (iterative optimization, IO) 和基于图神经网络 (graph neural network, GNN) 等方法。因 NAS 在分类和检测等视觉任务自动化设计神经架构取得了成功，并且深度立体网络的搜索空间巨大，研究者开始将其应用于立体匹配中。LEAStereo 是第一个端到端的立体分层 NAS 框架<sup>[70]</sup>，它由 2D 特征提取网络和 3D 代价聚合网络组成，并且融合了特定任务的人类知识，实现了在不依赖人工设计架构的情况下产生准确的立体匹配模型。但 LEAStereo 计算成本高并且难以扩展，EASNet 在其基础上将架构搜索空间覆盖得更加全面，包括深度、宽度和卷积核大小等信息，使网络能够适应不同的资源限制，同时保持高精度<sup>[71]</sup>。基于 IO 的方法不经过代价聚合步骤而直接对代价体迭代更新获取视差结果，节省了巨大的计算资源，可适用高分辨率图像，但其资源消耗与迭代次数成正比；此外，也能通过避免预定义的视差范围，在保证精度的同时处理大范围视差。RAFT-Stereo 是其代表性方法，通过构建轻量级代价体和引入卷积门控循环单元 (convolutional gated recurrent units, ConvGRU) 的更新算子迭代细化视差图得到最终视差<sup>[72]</sup>。但 RAFT-Stereo 是通过大量的迭代来获取精准视差，并没有解决算法处理时间长的问题，IGEVStereo 对几何、上下文信息和局部特征进行编码构建组合几何编码体<sup>[73]</sup>，能够准确回归 ConvGRU 迭代的初始点，将迭代次数由 RAFT-Stereo 的 32 次减少为 3 次，大幅减少迭代所消耗的计算资源。因 GNN 在建模特征间复杂关系以及全局特征表示方面的出色表现，研究者将其引入立体匹配领域以解决 CNN 仅依靠局部卷积来提取特征而感受野受限的问题，其可通过边将特征信息直接联系起来，提高模型的全局信息学习能力。如图 6 所示，基于 GNN 的方法与 CNN 网络结构相近，其以 GNN 替换特征提取模块，将像素特征传输到图上节点，特征之间的联系通过图的边存储，边

的值即代表特征交互的权重。WANG 等<sup>[74]</sup>首次将 GNN 应用于立体匹配领域, 提出使用多尺度图结构来提取多尺度图特征, 特征不仅可以在每个尺度内进行全局推理, 而且可以跨不同尺度进行交互, 获取具有多尺度感受野的全局上下文信息, 打破了现有深度模型只能提取单一

尺度的限制, 但存储大量特征间信息带来了资源上的消耗。SHI 等<sup>[75]</sup>提出使用基于集群的稀疏图来提高匹配效率, 将具有高相关性的关键点聚集到相同的子图中, 使得每个关键点仅与同一子图中的其他关键点交互, 有效减少了冗余特征间信息的存储, 提高了计算与存储的效率。

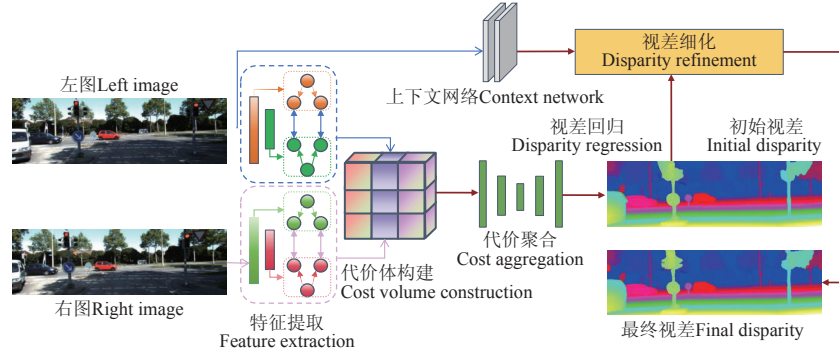


图6 基于图神经网络的立体匹配框架

Fig.6 Stereo matching framework based on graph neural network

## 2 双目立体视觉技术在农业领域中的应用

随着双目立体视觉技术在农业应用研究中的深入, 目前不少研究者从多个应用领域对其研究, 其中大部分集中在果实定位采摘与地图导航、生长参数测量和病害识别与施药等方面。如图7所示, 于2024年10月通过国内及国外数据库对双目立体视觉、定位、采摘、测量、病害等关键词进行检索, 并筛选近20年(2004—2024年)的数据进行统计分析, 设置相应条件如下: 数据库: 国内数据来源于中国知网、万方数据库, 国外数据来源于Web of Science、EI (Engineering Village)、Elsevier Science Direct; 技术主题关键词: 双目视觉 (binocular vision)、立体视觉 (stereo vision)、双目立体视觉 (binocular stereo vision); 应用主题关键词: 定位 (location)、采摘 (picking)、测量 (measurement)、病害 (pests) 等; 研究领域: 农业工程学科 (agriculture)。

从图7文献数量来看, 近5年国内外的果实定位采摘和作物参数测量的研究呈现出显著增长的趋势, 国内外的病害识别施药的研究则发展相对平稳。这表明在现阶段, 双目立体视觉技术在农业领域, 特别是在果实定位采摘与作物参数测量方面, 正逐渐成为研究热点, 同时, 在病害识别与施药方面也具有广阔的研究空间。

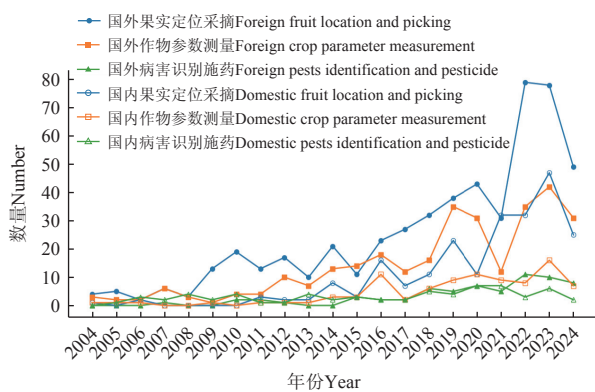


图7 出版数量折线图

Fig.7 Line chart of publication number

### 2.1 果实定位采摘与地图导航

果实采摘是一项机械且重复的工作, 耗时耗力, 对水果采摘机器人的研究在近年来广泛兴起。研究主要从两个技术方面展开, 其一是果实目标的三维定位, 其二是采摘机器人在复杂的自然果园的路径导航。通过双目视觉系统获取果园环境中的视觉信息, 能够为机器设备的定位与导航提供技术支持。

采摘机器人配备双目相机, 在采摘工作中采集双目图像, 一些研究者通过对双目图像中果实目标进行识别和分割去除背景信息得到感兴趣区域 (region of interest, ROI) <sup>[76-78]</sup>, 仅针对 ROI 部分利用双目立体视觉技术获取视差图, 然后利用三角测量原理计算出果实的三维坐标<sup>[79-80]</sup>, 这种方法的优势在于减少了非目标区域的干扰, 从而提高计算效率。顾宝兴等<sup>[76]</sup>使机器人在果园中随机移动, 通过双目相机对图像进行采集, 分别传输至改进的YOLOv3目标识别模型中进行果树树干检测输出ROI, 再对输出的左右ROI果树树干进行匹配, 最后通过双目相机三角测量原理对果树树干进行定位, 得到的横向和纵向定位平均绝对误差分别为39.0和266.0 mm。为研究双目立体视觉技术的定位稳定性和对光照的鲁棒性, TANG等<sup>[77]</sup>在阳光和阴影条件下对山茶花果实进行定位, 中位误差分别为23.568和23.524 mm。HOU等<sup>[78]</sup>在距离柑橘类水果200~500 mm的位置, 得到采摘点定位的平均绝对误差为8.63 mm, 平均相对误差为2.76%。然而, 由于只针对ROI部分, 可能会忽略果实周围环境信息, 在复杂场景中具有一定局限性。也有一些研究者将目标识别任务与视差图计算任务两部分工作同步进行<sup>[81-82]</sup>, 用果实的像素位置通过视差图间接得到其三维坐标, 这种方式能在保持实时性和精度的前提下, 更好地应对动态场景中的果实定位问题。LIU等<sup>[81]</sup>先将左侧图像输入到改进的YOLOv3目标识别模型中以获得图像中菠萝的位置信息, 同时通过立体匹配算法完成左右图像的匹配和视差计算, 最后对齐两者的结果计算出左图中菠萝的三维坐标, 在1.7~2.7 m距离处, 定位的平均绝对误差

为 24.414 mm，平均相对误差为 1.17%；HU 等<sup>[82]</sup>研究了不同识别算法对草莓定位的影响，其中 YOLOv3 识别算法后续的定位在 X、Y 和 Z 轴的综合平均绝对误差为 3.1 mm，而 Mask R-CNN 为 3.9 mm，均达到高精度的定位效果。也有研究者通过视差图生成点云，将点云与双目图像进行对齐和配准<sup>[83]</sup>，通过一系列操作，如去噪、滤波、聚类、轮廓拟合获取果实精准的三维坐标<sup>[84]</sup>。文朝武等<sup>[83]</sup>在温室环境中对花束进行定位并成功完成授粉；鲍秀兰等<sup>[84]</sup>在 1.5 m 的距离下进行柑橘果实定位试验，在低光照或正常光照条件平均绝对误差在 -20~20 mm 范围占 82.5%，在高光照条件平均绝对误差在 -20~20 mm 范围占 72%。研究均表明采摘机器人能够根据目标果实的相对位置确定采摘点，实现自动采摘。

果实采摘机器人在工作过程中，根据视差图生成所采集图像的局部稀疏点云和相机位姿运动轨迹<sup>[85]</sup>，完成体素滤波和点云拼接，进而生成密集点云构建果园三维稠密地图，实现高精度即时定位与地图构建（simultaneous localization and mapping, SLAM）<sup>[86]</sup>。为提高 SLAM 精度和实时性，陈劲等<sup>[85]</sup>根据全局图优化后的相机位姿和运动轨迹，通过点云拼接生成全局彩色稠密点云地图。在构建好的三维地图上，区分环境和可行路径进行点云聚类<sup>[87]</sup>实现障碍物定位规避<sup>[88]</sup>，根据采摘目标的位置和机器人的当前位置，在果园中可采用作物行识别寻找采摘路径<sup>[89]</sup>。张振乾等<sup>[87]</sup>采用改进的 K-means 聚类算法将道路两侧香蕉树分离，基于最小包围矩形提取导航路径，其导航路径提取平均横向偏差为 142.7 mm，平均航向偏差为 4.83°。近年来研究者开始使用进化算法<sup>[90]</sup>，如粒子群优化算法、A\*算法、蚁群算法和模拟退火算法等规划出最优的采摘路径实现导航。LI 等<sup>[90]</sup>采用改进的 A\*和 DWA 算法相结合实现密植枣园的机器人全局和实时局部路径规划，其在“U”型地图的导航平均误差为 26.9 mm、直线型地图平均误差为 2.47 mm 和“L”形地图平均误差为 26.8 mm，大幅提高导航精度。

### 2.2 作物生长参数测量

农作物生长过程中形态参数的变化可以直接反应其生长情况，对农作物的产量和品质预测有非常重要的指导意义<sup>[91]</sup>。植物工厂需要实时对作物生长进行监测，人工测量方法效率低下，甚至可能对农作物的正常生长造成影响，使用双目立体视觉技术进行监测已成为趋势。

如图 8 所示，使用左右相机同时采集作物左右图像，通过双目立体视觉技术和相机标定的参数获取作物点云，能够实现作物三维重建和生长参数测量。

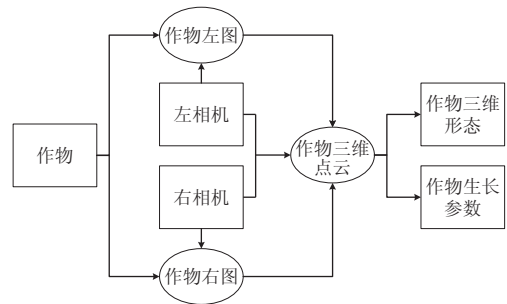


图 8 作物参数测量流程图

Fig.8 Flow chart of crop parameter measurement

根据点云与二维图像间的映射关系将像素点逐一对应完成匹配<sup>[92]</sup>，然后进一步拟合处理能够实现作物形态的重建<sup>[93]</sup>，如包围盒模型对作物进行包裹<sup>[94]</sup>、近似作物的多边形拟合<sup>[95]</sup>、平面方程拟合<sup>[96]</sup>等方法，3 种拟合方法概括如表 4 所示。并且对三维重建后的点云数据分析和计算能够实现生长参数测量，如通过高点到平面的距离测量高度<sup>[97-98]</sup>、采用拟合的三维形态用三角剖分测量体积<sup>[99]</sup>、利用骨架模型测量叶倾角<sup>[100]</sup>。魏新华等<sup>[97]</sup>将点云的各点到平面的距离结合密度峰聚类和 K-means 聚类的方法对田间作物高度数据分类，实现作物高度信息的提取，其高度检测平均绝对误差为 43.0 mm。但其适用的仅为大面积、噪声较多的田间场景。YUAN 等<sup>[98]</sup>根据 Ghostnet-YOLOv4 目标检测模型获取的 ROI 顶点和基点找到双目立体视觉技术获取的 3D 点云顶点和基点，最终得到树苗高度，总体准确率为 92.2%，实现了复杂环境下的精确高度测量。罗陆锋等<sup>[99]</sup>通过求解葡萄的最大截面，再将该截面绕中心轴旋转 360°得到葡萄空间包围体，当深度距离在 1 000 mm 以内时，最大直径误差小于 5.64%。这些方法处理的是整颗植株甚至是整个场景的点云，难以处理复杂形态和局部细节，精度可能受限，也有研究者通过对作物进行单器官分割实现更精准的测量，通过对点云进行配准、骨架提取<sup>[101]</sup>以及茎叶分割<sup>[102]</sup>获取需要分析的作物目标区域。对植株器官部位进行测量，如寻找最小外接矩形测量茎粗和 Delaunay 三角网格剖分法结合海伦公式测量叶面积<sup>[103]</sup>等生长参数<sup>[104]</sup>。如 MIAO 等<sup>[101]</sup>实现了准确自动测量株高、冠径、茎高、直径、叶宽和长度 6 个表型参数，并且拟合的可决系数  $R^2$  值均在 0.92 以上。这些研究涵盖了作物测量的不同场景，使得双目视觉能够更好地应用于复杂的自然环境中监测作物生长参数信息，提高作物的产量。

表 4 三维重建的拟合方法

Table 4 Fitting methods for 3D reconstruction

方法 Methods	优点 Advantages	缺点 Disadvantages	三维重建案例 3D reconstruction cases
包围盒模型 Bounding box model	算法简单，适合快速、粗略估计作物尺寸	精度较低，无法准确描述复杂形态	用多个轴对齐包围盒获取火龙果不规则枝条的三维空间位置信息，枝条最外轮廓尺寸偏差在 10 mm 以内 <sup>[94]</sup>
多边形拟合 Polygon fitting	能更好地近似作物的形状，适合不规则作物	计算复杂度较高，结果易受噪声点影响	用多个不规则三棱柱拟合粮堆，体积测量平均相对误差小于 0.5% <sup>[95]</sup>
平面方程拟合 Plane equation fitting	可描述平面形态，适用于规则表面或简化计算	仅适用于平坦区域，无法反映复杂形态	通过基于辅助平面拟合对点云坐标进行旋转，使三维重建保持尺度一致性 <sup>[96]</sup>

### 2.3 病害识别与施药

传统的虫害治理方法主要依靠人工,效率低下,不能获取精确的运动轨迹和参数,而引入视觉技术能够做到高精度和无损识别,并且农药施用能够从粗放型喷洒发展到精细化剂量作业。

双目立体视觉技术对作物病害识别主要有两方面,其一是识别作物自身的病害。针对自然环境存在光照、作物间生长遮挡等问题,光照分布不均会导致二维图像病害区域的特征提取不稳定,影响模型的检测准确率,与之相比,双目立体视觉技术对获取的视差图进一步处理后得到深度图,其像素值反映的是目标深度信息,不依赖于光照条件的变化,因此在光照不均时,深度图仍能保持对作物和背景的分隔;在深度图中叠加掩码可以消除图像中的远距离背景和非目标区域等信息,保留病害相关区域,实现对复杂背景的有效抑制,并在病害检测模型中完成对作物病害的识别<sup>[105]</sup>;通过采集不同视角的图像,利用立体视觉技术重建作物病害区域,可减少单一视角遮挡带来的信息缺失。其二是识别并定位作物中的害虫,先在作物图像中检测出害虫的像素位置,后通过双目相机获取的深度信息进行害虫<sup>[106-107]</sup>三维定位,进而喷药或激光射杀。LI等<sup>[106]</sup>在400~600 mm之间进行试验,其X、Y和Z轴的平均定位误差分别为0.40、0.30、0.51 mm,做到精准定位并射杀害虫。此外,研究者还将双目立体视觉技术运用于装备喷洒系统的无人机<sup>[108]</sup>、机器人<sup>[109]</sup>上,经测量作物的冠状体积来判断药物具体的喷洒量<sup>[110-111]</sup>。MOU等<sup>[109]</sup>提出了一种搭载了双目视觉定位的小型农药喷洒机器人,实现对单株玉米秧苗的精确位置喷药,当相机检测到喷药范围内的玉米秧苗时便启动喷洒系统,但没有做到真正意义上的精细化喷洒剂量作业;闫成功等<sup>[110]</sup>实现实时探测葡萄叶幕的深度并计算其冠层体积来控制药液剂量,相比于定量喷洒,用药量节省了约55.27%。这些研究从不同的方面进行农作物病虫害识别,为实现精准病害控制提供技术支持。

## 3 问题分析与对策

由于目前双目立体视觉技术受成像条件影响,精度逊于激光雷达,该技术在农业领域应用还存在一些挑战,主要有以下几个方面。

1) 实时性与精度难以平衡。双目相机的标定过程涉及相机内参和外参的求解,尤其在现场环境多变的情况下,需要频繁重新标定;图像校正需对每个像素重新映射,涉及插值计算;对目标三维信息的计算涉及大量图像数据的处理和匹配,该过程通常使用高维度卷积或注意力机制,计算量与图像大小和视差范围成正比。这些均会导致算法运算量较高,尤其在处理复杂场景和高分辨率图像时。而在智慧农业生产中,实时处理能力是性能的关键,它直接影响到系统的响应速度和决策效率。然而,目前的方法大多侧重于提高精度,高精度算法往往模型计算复杂,导致延迟增加,难以满足实际作业对实时性的需求。为解决这一问题,可深入研究轻量级模型,减少计算量;结合边缘计算与云计算,分散处理任务,提升系统响应速度;设计自适应标定算法避免频繁标定的时间消耗。

2) 应用场景受限。大多数应用研究主要在试验环境或理想的晴天条件下进行,缺乏对恶劣天气和复杂环境的验证,限制了该技术在实际应用中的推广。采摘机器人在自然环境中运行时,会受到外部环境的干扰,如地面不平整、障碍物阻挡以及光线不足等问题。因此,双目视觉成像还需克服由于抖动引起的图像模糊,以及太阳光强度变化造成的曝光不均等干扰。为提高技术适应性,需开发适用于复杂环境的算法,采用鲁棒性好的图像增强和去噪技术,在弱光或强光条件下提升成像质量。此外,设计具备抗干扰能力的双目视觉系统,如通过多模态融合(激光雷达、惯性测量单元(inertial measurement unit, IMU)等)提高数据可靠性。

3) 数据样本获取困难。缺少大规模、高质量的双目标注公开数据集,是当前农业领域双目视觉目标定位面临的主要挑战之一。有效的双目目标检测数据集不仅需要目标物体进行详细标注,还需标注视差、相机参数等信息。高质量的数据集对算法的训练和验证至关重要,它直接影响模型的准确性和鲁棒性。然而,农作物生长周期较长,获取大量真实有效的数据标签比较困难。此外,标注过程较为耗时且需要专业知识,进一步增加了数据集构建的难度。为了应对这一问题,可以利用自动化标注工具和半监督学习方法,降低数据标注成本,探索数据生成技术以补充真实数据。

4) 立体匹配缺乏统一评价标准。在立体匹配算法方面,尽管已有大量研究,但实现方法仍依赖于图像特征处理。这种方法通常侧重于局部特征的提取和匹配,导致在复杂场景下定位精度的提高受到限制。相比之下,深度学习方法能够通过端到端的训练优化立体匹配算法,利用自学习能力,提高算法的适应性和精确性。然而,众多立体匹配研究仍缺乏统一的标准来评价其优劣,不同算法之间的比较十分困难,阻碍了该技术的发展和应用。为推动算法发展,可构建农业场景特定的立体匹配算法评价体系,覆盖视差准确率、鲁棒性、实时性等指标。

## 4 总结与展望

双目立体视觉技术作为一种智能识别、定位和监测技术,将越来越广泛地应用在智慧农业生产之中。目前该技术并没有广泛覆盖农业领域,未来的发展需要解决当前存在的问题。展望未来,可以重点围绕以下几个方面进行研究。

1) 算法设计与优化。结合最新的深度学习技术路线,尤其是在复杂背景和纹理缺失的条件下对双目立体视觉技术进行研究。在提升物体识别和深度估计准确性的同时优化网络结构来降低模型的复杂性,如以深度条带卷积代替标准卷积捕获图像水平和垂直条带特征<sup>[112]</sup>、深度卷积和逐点卷积结合代替标准卷积<sup>[113]</sup>等角度优化网络以获得更轻量级的立体匹配模型,提高模型的实时处理能力和效率。为增强实际应用能力,开发能够适应不同环境条件(如光照、天气变化)的自适应算法尤为重要,使系统在实际农业场景下仍能保持高效的性能。此外,探索与激光雷达等形成互补的多源数据融合算法,进一步提高数据的准确性和可靠性,通过整合来自不同传感器的信息,提升对复杂场景的理解和处理能力,弥补单



一双目相机传感器的不足，为农业自动化和智能监测系统提供更全面的数据支持。

2) 智能辅助平台。建立基于双目视觉数据的智能分析平台，通过双目立体视觉技术对作物的三维结构特征进行精确重建，同时整合温度、湿度和光照等实时环境数据分析，提供可视化的决策支持工具，及时掌握作物生长状况。结合作物生理特性和生长规律进行形态测量方向研究，基于双目立体视觉技术提取的株高、叶面积、茎粗等关键生长指标构建作物表型特征数据库。通过机器学习方法分析历史生长数据，建立作物生长预测模型，实现作物生长异常的快速识别与诊断，为农业生产提供科学的决策依据。

3) 数据集构建。构建和共享农业领域高质量数据集，解决作物生长周期长、双目视差真值难获取和数据集质量低的问题。通过搭建双目数据采集平台，系统性地采集包含不同作物品种、不同生长时期、不同环境条件下的图像数据，并结合激光雷达等高精度设备测量可靠的视差真值，有效降低重复采集标注的时间和人力成本。同时建立严格的数据质量控制体系，对图像清晰度、标注准确性和场景多样性进行把控，确保数据集的可行价值，进而验证和优化深度学习立体匹配模型在农业场景下的性能表现，为双目立体视觉技术发展提供基础性支撑。由于数据集的限制，探索基于弱监督或自监督的学习策略也是双目立体视觉领域的重要突破方向，以提升模型在未标注数据上的泛化能力。

4) 完善评价体系。需从多维度构建和完善适用于农业场景的立体匹配算法评价体系，涵盖不同作物、不同生长阶段及多样环境条件下的视差准确率、鲁棒性、实时性和模型参数量等评价指标。针对作物和生长阶段的变化，评价应聚焦于算法的视差估计准确性和一致性。为提高鲁棒性，需纳入多种环境条件试验。考虑农业场景的实时性需求，应评估算法处理时延。精度分析应全面覆盖整体误差和关键区域误差（如边缘、遮挡和纹理稀疏区），通过专门指标衡量，便于寻找算法针对某特定区域的不足并优化性能。

综上所述，农业领域应用双目立体视觉技术仍需深入研究，解决目前存在的问题将是其改进的关键，国内农业正处于从传统农业向现代化智慧农业发展的关键时期，随着双目立体视觉技术的不断发展，将为智慧农业提供强大的技术支持，促进农业现代化、智能化发展。

#### [参 考 文 献]

- [1] 张铭桂, 琚俊, 刘厚诚. 基于机器视觉的三维重建技术在蔬菜上的应用综述[J]. 江苏农业科学, 2024, 52(9): 27-35.
- [2] 杨晓立, 徐玉华, 叶乐佳, 等. 双目立体视觉研究进展与应用[J]. 激光与光电子学进展, 2023, 60(8): 180-196. YANG Xiaoli, XU Yuhua, YE Lejia, et al. Research progress on binocular stereo vision applications[J]. Laser & Optoelectronics Progress, 2023, 60(8): 180-196. (in Chinese with English abstract)
- [3] 孙静, 刘晓燕. 基于双目立体视觉和轻量化神经网络的交通标志分割和识别[J]. 微型电脑应用, 2024, 40(6): 38-41. SUN Jing, LIU Xiaoyan. Traffic sign segmentation and recognition based on binocular stereo vision and lightweight neural network[J]. Microcomputer Applications, 2024, 40(6): 38-41. (in Chinese with English abstract)
- [4] 王瑞萍, 刘东风, 王先琳, 等. 基于多视图几何的白菜薹分割与关键表型测量[J]. 农业工程学报, 2022, 38(16): 243-251. WANG Ruiping, LIU Dongfeng, WANG Xianlin, et al. Segmentation and measurement of key phenotype for Chinese cabbage sprout using multi-view geometry[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(16): 243-251. (in Chinese with English abstract)
- [5] 陈青, 殷程凯, 郭自良, 等. 苹果采摘机器人关键技术研究现状与发展趋势[J]. 农业工程学报, 2023, 39(4): 1-15. CHEN Qing, YIN Chengkai, GUO Ziliang, et al. Current status and future development of the key technologies for apple picking robots[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 1-15. (in Chinese with English abstract)
- [6] SCHARSTEIN D, SZELISKI R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International Journal of Computer Vision, 2002, 47(1-3): 7-42.
- [7] ZHAI G, ZHANG W, HU W, et al. Coal mine rescue robots based on binocular vision: A review of the state of the art[J]. IEEE Access, 2020, 8: 130561-130575.
- [8] ZHANG Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [9] SHAO M, WANG P, WANG Y. Phase-based calibration method for a binocular vision sensor[J]. IEEE Access, 2021, 9: 44354-44362.
- [10] HU Y, RAO W, QI L, et al. A refractive stereo structured-light 3-D measurement system for immersed object[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-13.
- [11] WANG D, SUN H, LU W, et al. A novel binocular vision system for accurate 3-D reconstruction in large-scale scene based on improved calibration and stereo matching methods[J]. Multimedia Tools and Applications, 2022, 81(18): 26265-26281.
- [12] MENG L, LI Y, ZHOU H, et al. A hybrid calibration method for the binocular omnidirectional vision system[J]. IEEE Sensors Journal, 2022, 22(8): 8059-8070.
- [13] FU G M, GUO W J, GENG T, et al. Kinematics parameter calibration method of industrial robot based on binocular vision and distance error model[J]. Machine Tool and Hydraulics, 2019, 49(15): 10-16.
- [14] YE J J, LIU X J. Automatic calibration system for high speed camera based on virtual binocular vision[J]. Electric Measurement Technology, 2021, 42(3): 74-79.
- [15] JI S, QIN Z, SHAN J, et al. Panoramic SLAM from a multiple fisheye camera rig[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 169-183.
- [16] CHEN L, ZHONG G, WAN Z, et al. A novel binocular vision-robot hand-eye calibration method using dual nonlinear optimization and sample screening[J]. Mechatronics, 2023, 96: 103083.
- [17] STODOLA M, FROLÍK S. Self-calibration method of binocular vision based on conformal geometric algebra[J]. Mathematical Methods in the Applied Sciences, 2024, 47(3): 1289-1304.
- [18] CHEN W Y, XU J, YANG H. Camera calibration method based on double neural network[J]. Infrared and Laser Engineering, 2021, 50(11): 20210071.1-20210071.9.
- [19] ZHU H, WANG M, XU W. Research on calibration method of binocular vision system based on neural network[J]. Security and Communication Networks, 2021, 2021(1): 5542993.
- [20] REN J, GUAN F, WANG T, et al. High precision calibration algorithm for binocular stereo vision camera using deep reinforcement learning[J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 6596868.

- [21] BOUGUET J Y, PERONA P. Camera calibration from points and lines in dual-space geometry[C]// Proceedings of the 5th European Conference on Computer Vision. Freiburg, Germany: Springer, 1998: 2-6.
- [22] DARMON F, MONASSE P. The polar epipolar rectification[J]. *Image Processing on Line*, 2021, 11: 56-75.
- [23] ELHASHASH M, QIN R. Investigating spherical epipolar rectification for multi-view stereo 3d reconstruction[EB/OL]. (2022-04-08) [2024-10-01]. <https://arxiv.org/abs/2204.04141>.
- [24] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [25] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]// Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2564-2571.
- [26] 程德强, 李海翔, 寇旗旗, 等. 融合边缘保持与改进代价聚合的立体匹配算法[J]. *中国图象图形学报*, 2021, 26(2): 438-451.  
CHENG Deqiang, LI Haixiang, KOU Qiqi, et al. Stereo matching algorithm based on edge preservation and improved cost aggregation[J]. *Journal of Image and Graphics*, 2021, 26(2): 438-451. (in Chinese with English abstract)
- [27] 吕倪祺, 宋广华, 杨波威. 特征融合的双目半全局匹配算法及其并行加速实现[J]. *中国图象图形学报*, 2018, 23(6): 874-886.  
LYU Niqi, SONG Guanghua, YANG Bowei. Semi-global stereo matching algorithm based on feature fusion and its CUDA implementation[J]. *Journal of Image and Graphics*, 2018, 23(6): 874-886. (in Chinese with English abstract)
- [28] KONG L, ZHU J, YING S. Local stereo matching using adaptive cross-region-based guided image filtering with orthogonal weights[J]. *Mathematical Problems in Engineering*, 2021, 2021(1): 5556990.
- [29] XIAO B, YE S, LI X, et al. A stereo matching method for three-dimensional eye localization of autostereoscopic display[C]// Image and Graphics: 11th International Conference, ICIG 2021, Haikou, China: Springer, 2021: 28-43.
- [30] REAL-MORENO O, RODRÍGUEZ-QUIÑONEZ J C, SERGIYENKO O, et al. Fast template match algorithm for spatial object detection using a stereo vision system for autonomous navigation[J]. *Measurement*, 2023, 220: 113299.
- [31] ZHANG S, WU M, WU Y, et al. Fixed window aggregation AD-census algorithm for phase-based stereo matching[J]. *Applied Optics*, 2019, 58(32): 8950-8958.
- [32] LU B, SUN L, YU L, et al. An improved graph cut algorithm in stereo matching[J]. *Displays*, 2021, 69: 102052.
- [33] KIM S, MIN D, KIM S, et al. Unified confidence estimation networks for robust stereo matching[J]. *IEEE Transactions on Image Processing*, 2018, 28(3): 1299-1313.
- [34] 洪梓嘉, 李彦明, 林洪振, 等. 基于双目视觉的种植前期农田边界距离检测方法[J]. *农业机械学报*, 2022, 53(5): 27-33,56.  
HONG Zijia, LI Yanming, LIN Hongzhen, et al. Field boundary distance detection method in early stage of planting based on binocular vision[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(5): 27-33,56. (in Chinese with English abstract)
- [35] 黄超, 赵华治. 根据灰度值信息自适应窗口的半全局匹配[J]. *中国图象图形学报*, 2019, 24(8): 1381-1390.  
HUANG Chao, ZHAO Huazhi. Semi-global stereo matching with adaptive window based on grayscale value[J]. *Journal of Image and Graphics*, 2019, 24(8): 1381-1390. (in Chinese with English abstract)
- [36] DENG C, LIU D, ZHANG H, et al. Semi-global stereo matching algorithm based on multi-scale information fusion[J]. *Applied Sciences*, 2023, 13(2): 1027.
- [37] ZHANG W, DONG P, CHEN L, et al. Min-pooling cost aggregation for semi-global matching of stereo vision processor[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2024.
- [38] HIRSCHMULLER H. Stereo processing by semiglobal matching and mutual information[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 30(2): 328-341.
- [39] 龙霄潇, 程新景, 朱昊, 等. 三维视觉前沿进展[J]. *中国图象图形学报*, 2021, 26(6): 1389-1428.  
LONG Xiaoxiao, CHENG Xinjing, ZHU Hao, et al. Recent progress in 3D vision[J]. *Journal of Image and Graphics*, 2021, 26(6): 1389-1428. (in Chinese with English abstract)
- [40] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014: 2672-2680.
- [41] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017: 6000-6010.
- [42] 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. Las Vegas, NV, USA: IEEE, 2016: 4040-4048.
- [43] MENZE M, GEIGER A. Object scene flow for autonomous vehicles[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 3061-3070.
- [44] ŽBONTAR J, LECUN Y. Computing the stereo matching cost with a convolutional neural network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 1592-1599.
- [45] ŽBONTAR J, LECUN Y. Stereo matching by training a convolutional neural network to compare image patches[J]. *Journal of Machine Learning Research*, 2016, 17(1): 2287-2318.
- [46] 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. Venice, Italy: IEEE, 2017: 66-75.
- [47] HUANG Z, NORRIS T B, WANG P. ES-Net: An efficient stereo matching network[EB/OL]. (2021-03-05) [2024-10-01]. <https://arxiv.org/abs/2103.03922>.
- [48] BADKI A, TROCCOLI A, KIM K, et al. Bi3d: Stereo depth estimation via binary classifications[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 1597-1605.
- [49] XU B, XU Y, YANG X, et al. Bilateral grid learning for stereo matching networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 12492-12501.
- [50] CHANG J R, CHEN Y S. Pyramid stereo matching network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 5410-5418.
- [51] ZHANG F, PRISACARIU V, YANG R, et al. Ga-net: Guided aggregation net for end-to-end stereo matching[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 185-194.
- [52] EMLEK A, PEKER M. P3SNet: Parallel pyramid pooling stereo network[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(10): 10433-10444.
- [53] KOU L, YANG K, LUO L, et al. Binocular stereo matching of real scenes based on a convolutional neural network and computer graphics[J]. *Optics Express*, 2021, 29(17): 26876-

- 26893.
- [54] WANG Q, XING H, YING Y, et al. CGFNet: 3D convolution guided and multi-scale volume fusion network for fast and robust stereo matching[J]. *Pattern Recognition Letters*, 2023, 173: 38-44.
- [55] PILZER A, XU D, PUSCAS M, et al. Unsupervised adversarial depth estimation using cycled generative networks[C]//2018 International Conference on 3D Vision(3DV). Verona, Italy: IEEE, 2018: 587-595.
- [56] WEI D, LIU H. Stereo matching algorithm based on CGAN[C]//2021 China Automation Congress (CAC). Beijing, China: IEEE, 2021: 5607-5612.
- [57] WANG C, BAI X, WANG X, et al. Self-supervised multiscale adversarial regression network for stereo disparity estimation[J]. *IEEE Transactions on Cybernetics*, 2021, 51(10): 4770-4783.
- [58] LIU R, YANG C, SUN W, et al. Stereogan: Bridging synthetic-to-real domain gap by joint optimization of domain translation and stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020:12754-12763.
- [59] 厉行, 樊养余, 郭哲, 等. 基于边缘领域自适应的立体匹配算法[J]. *电子与信息学报*, 2024, 46(7): 2970-2980.  
LI Xing, FAN Yangyu, GUO Zhe, et al. Edge domain adaptation for stereo matching[J]. *Journal of Electronics & Information Technology*, 2024, 46(7): 2970-2980. (in Chinese with English abstract)
- [60] 徐磊, 宋慧慧, 刘青山. 多层次融合注意力网络的双目图像超分辨率重建[J]. *中国图象图形学报*, 2023, 28(4): 1079-1090.  
XU Lei, SONG Huihui, LIU Qingshan. Super-resolution reconstruction of binocular image based on multi-level fusion attention network[J]. *Journal of Image and Graphics*, 2023, 28(4): 1079-1090. (in Chinese with English abstract)
- [61] DOSOVITSKIY A. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2020-10-22) [2024-10-01]. <https://arxiv.org/abs/2010.11929>.
- [62] 贾迪, 蔡鹏, 吴思, 等. 面向弱纹理目标立体匹配的Transformer网络[J]. *中国图象图形学报*, 2024, 29(8): 2413-2425.  
JIA Di, CAI Peng, WU Si, et al. Transformer network for stereo matching of weak texture objects[J]. *Journal of Image and Graphics*, 2024, 29(8): 2413-2425. (in Chinese with English abstract)
- [63] LI Z, LIU X, DRENKOW N, et al. Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 6177-6186.
- [64] 晏敏, 王军政, 李静. 多尺度相似性迭代查找的可靠双目视差估计[J]. *中国图象图形学报*, 2022, 27(2): 447-460.  
YAN Min, WANG Junzheng, LI Jing. Reliable binocular disparity estimation based on multi-scale similarity recursive search[J]. *Journal of Image and Graphics*, 2022, 27(2): 447-460. (in Chinese with English abstract)
- [65] CHENG X, ZHONG Y, HARANDI M, et al. Deep laparoscopic stereo matching with transformers[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Singapore: MICCAI, 2022: 464-474.
- [66] LOU J, LIU W, CHEN Z, et al. Elfnet: Evidential local-global fusion for stereo matching[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 2023: 17738-17747.
- [67] SHEN Z, DAI Y, SONG X, et al. Pcw-net: Pyramid combination and warping cost volume for stereo matching[C]//European Conference on Computer Vision. Tel Aviv, Israel: Springer, 2022: 280-297.
- [68] LIU Z, LI Y, OKUTOMI M. Global occlusion-aware transformer for robust stereo matching[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE, 2024: 3523-3532.
- [69] RAO Z, HE M, DAI Y, et al. Sliding space-disparity transformer for stereo matching[J]. *Neural Computing and Applications*, 2022, 34(24): 21863-21876.
- [70] CHENG X, ZHONG Y, HARANDI M, et al. Hierarchical neural architecture search for deep stereo matching[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 22158-22169.
- [71] WANG Q, SHI S, ZHAO K, et al. EASNet: Searching elastic and accurate network architecture for stereo matching[C]//European Conference on Computer Vision. Tel Aviv, Israel: Springer, 2022: 437-453.
- [72] LIPSON L, TEED Z, DENG J. Raft-stereo: Multilevel recurrent field transforms for stereo matching[C]//2021 International Conference on 3D Vision (3DV). London, United Kingdom: IEEE, 2021: 218-227.
- [73] XU G, WANG X, DING X, et al. Iterative geometry encoding volume for stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 21919-21928.
- [74] WANG X, YU J, SUN Z, et al. Multi-scale graph neural network for global stereo matching[J]. *Signal Processing: Image Communication*, 2023, 118: 117026.
- [75] SHI Y, CAI J X, FAN M, et al. Learning to match features with discriminative sparse graph neural network[J]. *Pattern Recognition*, 2024, 156: 110784.
- [76] 顾宝兴, 刘钦, 田光兆, 等. 基于改进YOLOv3的果树树干识别和定位[J]. *农业工程学报*, 2022, 38(6): 122-129.  
GU Baoxing, LIU Qin, TIAN Guangzhao, et al. Recognizing and locating the trunk of a fruit tree using improved YOLOv3[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(6): 122-129. (in Chinese with English abstract)
- [77] TANG Y, ZHOU H, WANG H, et al. Fruit detection and positioning technology for a camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision[J]. *Expert Systems with Applications*, 2023, 211: 118573.
- [78] HOU C, XU J, TANG Y, et al. Detection and localization of citrus picking points based on binocular vision[J]. *Precision Agriculture*, 2024, 25(5): 2321-2355.
- [79] LEI X, WU M, LI Y, et al. Detection and positioning of camellia oleifera fruit based on LBP image texture matching and binocular stereo vision[J]. *Agronomy*, 2023, 13(8): 2153.
- [80] ZHU C, WU C, LI Y, et al. Spatial location of sugarcane node for binocular vision-based harvesting robots based on improved YOLOv4[J]. *Applied Sciences*, 2022, 12(6): 3088.
- [81] LIU T H, NIE X N, WU J M, et al. Pineapple (Ananas comosus) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model[J]. *Precision Agriculture*, 2023, 24(1): 139-160.
- [82] HU H, KAIJU Y, ZHANG H, et al. Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN[J]. *International Journal of Agricultural and Biological Engineering*, 2022, 15(6): 175-179.
- [83] 文朝武, 龙洁花, 张宇, 等. 基于3D视觉的番茄授粉花朵定位方法[J]. *农业机械学报*, 2022, 53(8): 320-328.  
WEN Chaowu, LONG Jiehua, ZHANG Yu, et al. Positioning method of tomato pollination flowers based on 3D vision[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(8): 320-328. (in Chinese with English abstract)
- [84] 鲍秀兰, 马志涛, 马萧杰, 等. 丘陵果园自然环境下柑橘采摘机器人设计与试验[J]. *农业机械学报*, 2024, 55(4):

- 124-135.  
BAO Xiulan, MA Zhitao, MA Xiaojie, et al. Design and experiment of citrus picking robot in hilly orchard natural environment[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(4): 124-135. (in Chinese with English abstract)
- [85] 陈劭, 郭宇翔, 高天啸, 等. 移动机器人 RGB-D 视觉 SLAM 算法[J]. *农业机械学报*, 2018, 49(10): 38-45.  
CHEN Shao, GUO Yuxiang, GAO Tianxiao, et al. RGB-D visual SLAM algorithm for mobile robots[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(10): 38-45. (in Chinese with English abstract)
- [86] XUE Z. Analysis of simultaneous localization and mapping technology for mobile robot based on binocular vision[C]//2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). Changchun, China: IEEE, 2023: 1735-1739.
- [87] 张振乾, 李世超, 李晨阳, 等. 基于双目视觉的香蕉园巡检机器人导航路径提取方法[J]. *农业工程学报*, 2021, 37(21): 9-15.  
ZHANG Zhenqian, LI Shichao, LI Chenyang, et al. Navigation path detection method for a banana orchard inspection robot based on binocular vision[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(21): 9-15. (in Chinese with English abstract)
- [88] 魏建胜, 潘树国, 田光兆, 等. 农业车辆双目视觉障碍物感知系统设计与试验[J]. *农业工程学报*, 2021, 37(9): 55-63.  
WEI Jiansheng, PAN Shuguo, TIAN Guangzhao, et al. Design and experiments of the binocular visual obstacle perception system for agricultural vehicles[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(9): 55-63. (in Chinese with English abstract)
- [89] 翟志强, 熊坤, 王亮, 等. 采用双目视觉和自适应 Kalman 滤波的作物行识别与跟踪[J]. *农业工程学报*, 2022, 38(8): 143-151.  
ZHAI Zhiqiang, XIONG Kun, WANG Liang, et al. Crop row detection and tracking based on binocular vision and adaptive Kalman filter[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(8): 143-151. (in Chinese with English abstract)
- [90] LI Y, LI J, ZHOU W, et al. Robot path planning navigation for dense planting red jujube orchards based on the joint improved A\* and DWA algorithms under laser SLAM[J]. *Agriculture*, 2022, 12(9): 1445.
- [91] 鞠执一, 邓干然, 李旭荣, 等. 双目视觉技术在农业监测中的应用现状及展望[J]. *现代农业装备*, 2021, 42(6): 2-7,14.  
JU Zhiyi, DENG Ganran, LI Xurong, et al. The application status and prospect of stereo vision in agriculture[J]. *Modern Agricultural Equipment*, 2021, 42(6): 2-7,14. (in Chinese with English abstract)
- [92] GAO Y, WANG Q, RAO X, et al. OrangeStereo: A navel orange stereo matching network for 3D surface reconstruction[J]. *Computers and Electronics in Agriculture*, 2024, 217: 108626.
- [93] 殷悦, 张慧春, 郑加强. 基于双目立体视觉的植物三维重建系统[J]. *中国农机化学报*, 2021, 42(3): 129-135.  
YIN Yue, ZHANG Huichun, ZHENG Jiaqiang. Three dimensional reconstruction system of plant based on binocular stereo vision[J]. *Journal of Chinese Agricultural Mechanization*, 2021, 42(3): 129-135. (in Chinese with English abstract)
- [94] 高嘉正, 李文涛, 罗陈迪, 等. 基于语义分割和几何分析的火龙果果实与枝条的三维重建[J]. *农业工程学报*, 2024, 40(12): 157-164.  
GAO Jiazheng, LI Wentao, LUO Chendi, et al. Three-dimensional reconstruction for dragon fruits and branches using semantic segmentation and geometric analysis[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(12): 157-164. (in Chinese with English abstract)
- [95] ZHAO Z, CHANG H, WU C. Study on the measurement method of wheat volume based on binocular Structured Light[J]. *Sustainability*, 2023, 15(18): 13814.
- [96] 朱磊, 江伟, 孙伯颜, 等. 基于神经辐射场的苗期作物三维建模和表型参数获取[J]. *农业机械学报*, 2024, 55(4): 184-192,230.  
ZHU Lei, JIANG Wei, SUN Boyan, et al. Three-dimensional reconstruction and phenotype parameters acquisition of seeding vegetables based on neural radiance fields[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(4): 184-192,230. (in Chinese with English abstract)
- [97] 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(3): 225-233.  
WEI Xinhua, ZHANG Min, LIU Qingshan, et al. Extraction of crop height and cut-edge information based on binocular vision[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(3): 225-233. (in Chinese with English abstract)
- [98] YUAN X, LI D, SUN P, et al. Real-time counting and height measurement of nursery seedlings based on ghostnet-YoloV4 network and binocular vision technology[J]. *Forests*, 2022, 13(9): 1459.
- [99] 罗陆锋, 邹湘军, 叶敏, 等. 基于双目立体视觉的葡萄采摘防撞空间包围体求解与定位[J]. *农业工程学报*, 2016, 32(8): 41-47.  
LUO Lufeng, ZOU Xiangjun, YE Min, et al. Calculation and localization of bounding volume of grape for undamaged fruit picking based on binocular stereo vision[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2016, 32(8): 41-47. (in Chinese with English abstract)
- [100] WEN J, YIN Y, ZHANG Y, et al. Detection of wheat lodging by binocular cameras during harvesting operation[J]. *Agriculture*, 2023, 13(1): 120-133.
- [101] MIAO T, ZHU C, XU T, et al. Automatic stem-leaf segmentation of maize shoots using three-dimensional point cloud[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106310.
- [102] JIN S, SU Y, WU F, et al. Stem-leaf segmentation and phenotypic trait extraction of individual maize using terrestrial LiDAR data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(3): 1336-1346.
- [103] 彭程, 苗艳龙, 汪刘洋, 等. 基于三维点云的田间香蕉吸芽形态参数获取[J]. *农业工程学报*, 2022, 38(Z): 193-200.  
PENG Cheng, MIAO Yanlong, WANG Liuyang, et al. Morphological parameters extraction of banana sucker in the field based on three-dimensional point cloud[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(Z): 193-200. (in Chinese with English abstract)
- [104] 彭程, 李帅, 苗艳龙, 等. 基于三维点云的番茄植株茎叶分割与表型特征提取[J]. *农业工程学报*, 2022, 38(9): 187-194.  
PENG Cheng, LI Shuai, MIAO Yanlong, et al. Stem-leaf segmentation and phenotypic trait extraction of tomatoes using three-dimensional point cloud[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(9): 187-194. (in Chinese with English abstract)
- [105] SUN H, XUE J, SONG Y, et al. Detection of fruit tree

- diseases in natural environments: A novel approach based on stereo camera and deep learning[J]. *Engineering Applications of Artificial Intelligence*, 2024, 137: 109148.
- [106] LI Y, FENG Q, LIN J, et al. 3D locating system for pests' laser control based on multi-constraint stereo matching[J]. *Agriculture*, 2022, 12(6): 766.
- [107] CHEN C J, HUANG Y Y, Li Y S, et al. Identification of fruit tree pests with deep learning on embedded drone to achieve accurate pesticide spraying[J]. *IEEE Access*, 2021, 9: 21986-21997.
- [108] HAN Y, JIANG W, AHMED A, et al. Study on unmanned aerial vehicle with neural network stereovision perception for reducing pesticide and water saving in precision agriculture[C]//2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST). Guangzhou, China: IEEE, 2021: 651-654.
- [109] MOU Z, YEFENG P U, WEN J, et al. Identification system of pesticide spraying robot based on binocular vision[J]. *Agricultural Engineering*, 2023, 13(10): 27-31.
- [110] 闫成功, 徐丽明, 袁全春, 等. 基于双目视觉的葡萄园变量喷雾控制系统设计与试验[J]. *农业工程学报*, 2021, 37(11): 13-22.
- YAN Chenggong, XU Liming, YUAN Quanchun, et al. Design and experiments of vineyard variable spraying control system based on binocular vision[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(11): 13-22. (in Chinese with English abstract)
- [111] RAYAMAJHI A, JAHANIFAR H, ASIF M, et al. Measuring ornamental 3D canopy volume and trunk diameter using stereo vision for precision spraying and assessing tree maturity[C]//2024 ASABE Annual International Meeting. Anaheim, CA, USA: American Society of Agricultural and Biological Engineers, 2024: 1.
- [112] GUO X, ZHANG C, NIE D, et al. Lightstereo: Channel boost is all your need for efficient 2d cost aggregation[EB/OL]. (2024-06-28) [2024-10-01]. <https://arxiv.org/abs/2406.19833>.
- [113] SHAMSAFAR F, WOERZ S, RAHIM R, et al. Mobilestereonet: Towards lightweight deep networks for stereo matching[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE, 2022: 2417-2426.

## Research progress on binocular stereo vision technology and its applications in smart agriculture

YANG Xuejun, ZHONG Jinbiao, LIN Kaiyan<sup>\*</sup>, WU Junhui, CHEN Jie, SI Huiping

(Modern Agricultural Science and Engineering Institute, Tongji University, Shanghai 201804, China)

**Abstract:** To achieve precision operations, it is essential to accurately locate targets within agricultural scenes. As an important method for obtaining three-dimensional (3D) perception of the real world, binocular stereo vision technology can facilitate the 3D localization and point cloud reconstruction of targets in agricultural environments, thereby showcasing considerable application potential. This paper conducts an in-depth study of binocular stereo vision technology and its applications in the agricultural field. Firstly, we summarize the pipeline of binocular stereo vision technology, reviewing its latest research advancements along the technical threads of binocular camera calibration, epipolar rectification, and stereo matching. The binocular camera calculates the depth information of targets based on disparity results. The objective of stereo vision calibration is to determine the intrinsic and extrinsic parameters of the camera, establishing a mapping between points in pixel coordinates and world coordinates, which includes reference calibration, active vision calibration, self-calibration, and neural network calibration methods. Epipolar rectification employs constraints to reduce the search space for matching points from two dimensions to one. Stereo matching calculates disparity by matching left and right images in both feature-based and deep learning methods. Feature-based methods can be further categorized into local, global, and semi-global methods, depending on the search range of matching pixels. Local methods search for matching points within surrounding areas, global methods minimize the global energy function, and semi-global methods aggregate costs from various directions. In contrast, deep learning methods can learn more complex features to enhance stereo matching results, further categorized by network frameworks such as Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Transformer methods. In addition, three prominent methods of Neural Architecture Search (NAS), Iterative Optimization (IO) and Graph Neural Network (GNN) are also introduced. CNNs perform extensive convolution operations to compute matching costs, with notable research directions aim at improving accuracy, including convolutional encoders and decoders, hierarchical pyramids, and complex cost volumes. GANs synthesize data through adversarial generation, alleviating the challenges of acquiring realistic disparity in binocular datasets. Transformer methods utilize self-attention mechanisms to capture contextual information, addressing the limitations of CNN receptive fields. NAS can automatically construct stereo matching network architectures by incorporating human prior knowledge, eliminating the need for manual design. IO methods do not require the construction of cost volumes and cost aggregation, saving significant resources and enabling the processing of large disparity ranges. GNN can model complex relationships between features and extract global information. Furthermore, the trend in the number of publications in recent years are analysed to examine the widespread applications of binocular stereo vision technology in agricultural research. Recent literature is synthesized to explore the latest applications. This technology enables 3D localization of fruit targets and facilitates map navigation in practical operations based on point cloud processing. It also supports 3D reconstruction of crops or segmentation of individual organs for growth parameter measurement. Additionally, it aids in the identification of crop diseases or pests, combined with precision spraying by agricultural machinery. Ultimately, we summarize the challenges of applying binocular stereo vision technology in agriculture. Although it demonstrates high precision in localization, measurement, and identification, it still faces issues such as model complexity, scene limitations, scarcity of datasets, and a lack of evaluation standards for stereo matching. Looking ahead, future research in this technology for agricultural applications should focus on algorithm design and optimization, the establishment of intelligent assistance platforms, the construction of comprehensive datasets, and the improvement of evaluation system to further enhance the practicality and efficiency of binocular stereo vision systems in precision agriculture.

**Keywords:** binocular vision; smart agriculture; stereo matching; crop parameter measurement; three-dimensional target positioning; disease identification