引文格式：孟金龙, 王忠康, 李霏, 等. 面向复杂产品装配的人机协作交互与安全监测方法[J]. 航空制造技术, 2026, 69(20): 26010004.

MENG Jinlong, WANG Zhongkang, LI Fei, et al. An interaction and safety monitoring method under human-robot collaborative assembly of complex products[J]. Aeronautical Manufacturing Technology, 2026, 69(20): 26010004.

面向复杂产品装配的人机协作交互与安全监测方法

孟金龙[1]，王忠康[2]，李　霏[3]，曹志宏[3]，胡　蓝[2]，李鹏程[1]，田　威[1]

（1. 南京航空航天大学，南京 210016；
2. 上海航天设备制造总厂有限公司，上海 200040；
3. 北京电子工程总体研究所，北京 100039）

[摘要] 随着先进制造技术的发展，航空航天产品的制造装配逐渐从人工装配向人机协同装配与自动化装配的方向发展，有效的人机交互方式与人机安全监测方法成为提高人机协同装配效率、保障人机安全的关键技术。本文以航天产品为人机协同装配对象，建立了航天产品人机协同装配下的人机交互与安全监测系统与方法。提出了一种融合Openpose与PointPillars的人体3D碰撞体构建方法，提高了装配现场遮挡下的人员3D位姿估计准确性；研究了非安全区域碰撞体构建方法，并基于点云配准与机器人实时关节信息构建了机器人碰撞体，从而实现了装配过程中人机安全监控；针对传统交互方式的交互效率低、学习成本高的问题，设计了基于Mediapipe的手势识别与分类方法，实现了外部设备的手势控制。通过试验证明，提出的系统与方法能够满足现场装配人员实际装配需求，提高了装配效率与安全性。

关键词：复杂产品装配；人机协作；人机交互；安全监测；手势识别

中图分类号：V2；V19　　**文献标志码：**A　　**DOI：**10.16080/j.issn1671-833x.26010004



孟金龙

博士研究生，研究员，研究方向为机器人装配应用。

航空航天复杂产品具有品种多、小批量、结构复杂、装配空间狭小等特点，导致装配与调整的工作量大，通常占制造过程总时长的50%~80%[1]，因此成为制约航天产品生产效率提高的主要环节。随着智能化水平的提高，航天产品从以手工和经验为主的离散化装配模式，逐步向人机协作装配的智能化模式转变，改善了以往装配人员劳动强度过大，装配效率过慢的问题，但由于航天产品部分结构过于复杂，现有智能化水平无法完全取代人工，因此人机协作装配成为目前航天产品的主要装配模式[2]。

人机协作装配将人与机器人置于同一时空，装配过程中不可避免地给人带来安全隐患。传统人机安全策略主要分为预碰撞策略与后碰撞策略两类。后碰撞策略通过工业机器人本体传感器（如力矩传感器、电流环监测）实现碰撞后的紧急制动或

收稿日期：2025-12-26；**退修日期：**2026-04-10；**录用日期：**2026-05-10

基金项目：国防基础科研项目（JCKY2023205A003）

通信作者：李鹏程，副教授，博士，主要研究方向为飞行器先进装配技术、机器人视觉伺服控制，非线性系统控制，机器人校准与辨识，并联机器人。

田威，教授，博士，长期从事航空航天机器人智能装配技术与装备研究。

停机[3]，Cirillo[4]和Phan[5]等则将柔性电子皮肤技术被应用于机械臂表层覆盖，通过实时感知接触力变化构建动态反馈系统。Haddadins等[6]通过嵌入式扭矩检测模块实时采集各关节传动数据，并基于动力学模型的力矩偏差分析算法判别外部碰撞。Xu等[7]对机器人各个关节电流进行实时监测，将由电流转换的关节力矩与加速度得到的驱动力矩对比判定是否发生碰撞，虽然上述研究均可检测机器人外部碰撞，但其依赖接触触发保护机制，存在响应滞后性，难以主动规避风险。预碰撞策略则通过视觉、激光雷达等感知技术构建虚拟安全区域，当人员侵入时触发减速或制动，Vogel等[8]提出了一种利用传统投影仪和相机结合的人机安全保障方法，该系统利用安置在天花板上的投影仪所发出的光线在机器人工作空间周围形成光障作为安全空间，配套相机用于监控投影图形完整性，用于判断人员是否侵入机器人周围空间。Pilz公司发布的SafetyEYE[9]是世界上第一台安全照相系统，被安装在被监控区域的上方，能够对下方的区域进行三维立体的监控。但上述方法本质上与安全围栏类似，在空间上隔离了工业机器人与人，限制了人机共融的灵活性。

因此，现有的人机安全策略在预碰撞策略层面依然采用人机分离的安全策略，在空间上隔离了机器人与人，限制了人机协作灵活性，不适用于人机协作场景；而后碰撞策略依赖接触触发保护机制，存在响应滞后性，难以主动规避风险，不适合同时保证航天复杂产品装配场景的安全与效率，因此需要探索一种能适用于人机共融环境并能够同时兼顾安全与效率的安全规划策略。

为动态获取人机协作装配过程中人体的位姿变化，人体3D关键点估计成为建立人体碰撞体、监控人机安全的关键方法。从主流算法方面分析，目前可以根据有无深度信息的参与分为基于红绿蓝与深度（Red, Green, Blue, Depth，RGBD）的深度映射与基于红绿蓝（Red, Green, Blue，RGB）的3D关键点估计两种类型。目前基于RGBD的深度映射方法会受到人体表面服饰、皮肤与障碍物遮挡的影响，在结构复杂、遮挡较多的场景下存在较大的误差。基于RGB的3D关键点估计方法会得到多组前后颠倒的人体3D关键点，识别精度不稳定，且数据集获取难度极大，数据集的场景特定性强，无法迁移至其他场景的识别任务中。由于航天复杂产品装配场景下人机遮挡、障碍物遮挡频繁，且人机安全性要求严格，要求人体3D关键点估计方法具有较高的可靠性，所以上述算法均不能实现航天复杂产品装配场景下的人体状态估计。

为提高航天产品人机协作的效率与安全性，本文提出了一种人机交互与安全监测方法，通过在现场布置深度相机与激光雷达等传感器，对装配过程进行安全监控与交互识别，通过对人体、机器人与危险区域的位姿识别，对当前装配现场状态进行碰撞预警，提高装配安全性与交互效率。在实现过程中，针对装配人员3D位姿估计困难、装配现场遮挡频发的问题，提出了一种融合Openpose与PointPillars的人体3D碰撞体构建方法；为实现装配过程中人机安全监控，提出了非安全区域概念及碰撞体构建方法，并基于点云配准与机器人实时关节信息实现了机器人碰撞体的构建；针对传统交互方式的交互效率低，学习成本高的问题，提出了基于Mediapipe的手势识别与分类方法，实现了外部设备的手势控制。通过试验证明，提出的系统与方法能够满足现场装配人员实际装配需求，提高了装配效率与安全性。

1 航天产品装配场景与方法框架

1.1 航天产品装配场景

以飞行器类产品装配为例，场景如图1所示，共有两台协作型机器人、两台深度相机、一台激光雷达，两套旋转工装等配套设备，其中，两台机器人负责在上料点夹取物料并悬停于旋转工装上方，工人先后在入舱装配工位与桁条装配工位分别人工引导机器人装配与人工手动作业，两台深度相机与激光雷达负责保障工作人员在人工操作区、桁条装配工位与入舱装配工位的安全，并负责感知手势用于控制工装操作，在上述过程中，两台机器人可能会与人发生干涉碰撞，入舱装配工位与桁条装配工位上的旋转工装在工作时可能与人体发生干涉碰撞，为预防工人操作失误或设备失控导致的人机安全隐患，需要通过激光雷达与深度相机实时感知人机位姿状态，在人与危险区域或机器人的最小距离小于阈值时发出警报信号。

1.2 整体方法框架

为保障装配过程中的人机安全性，本文设计如图2所示的整体方法框架，首先激光雷达与深度相机空间标定部分利用点云配准方法建立全局性的环境点云地图。之后在手势识别方面，若判断手势具有与已有手势库中的特征吻合，则发送手势控制信号控制外部设备；其次在碰撞体构建方面，人体碰撞体构建、机器人碰撞体构建与非安全区域碰撞体构建同步进行，将人体碰撞体、机器人碰撞体、非安全区域碰撞体进行实时碰撞检测，若人体处于危险状态，则发送报警信息，整体过程循环进行，实时监控环境中的交互与人体安全。

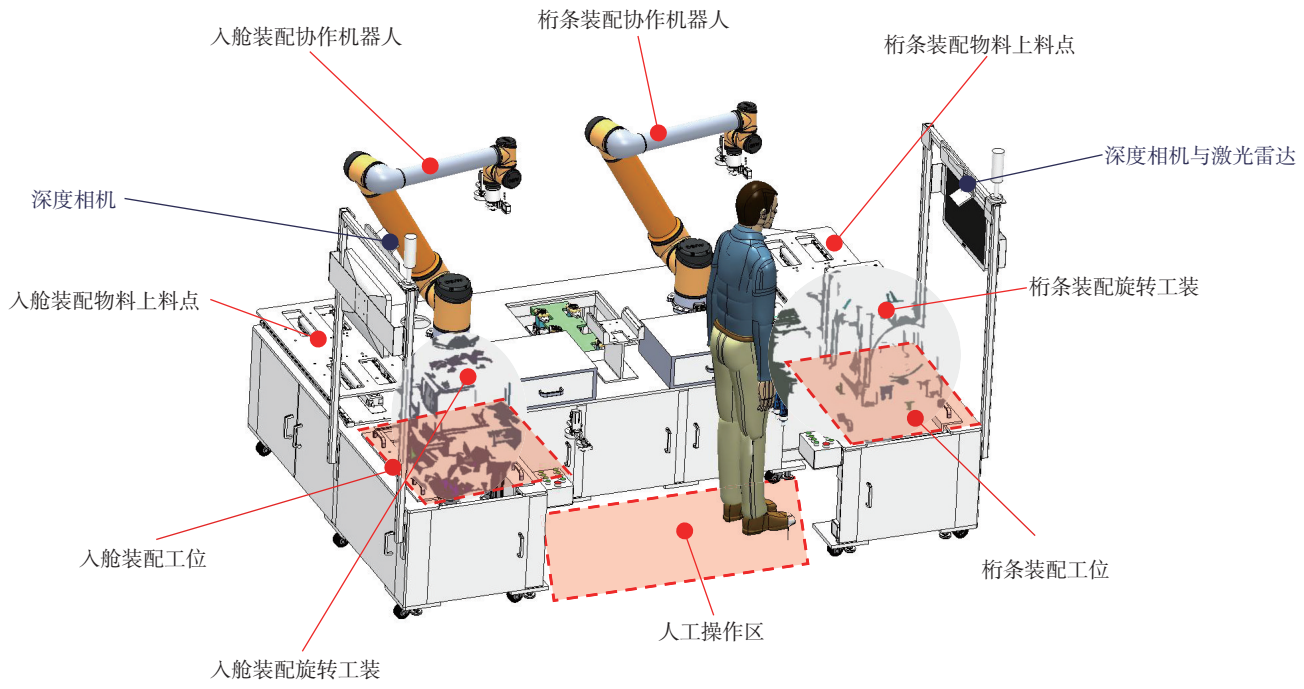


图1 航天产品装配场景

Fig.1 Aerospace product assembly scenario

2 融合Openpose与PointPillars的人体碰撞体构建方法

为实时监控人体位姿，本文提出融合Openpose与PointPillars的人体碰撞体构建方法，通过深度相机与激光雷达的图像与点云信息估计人体碰撞体，为后续的碰撞检测提供数据基础。OpenPose[10]是卡内基梅隆大学提出的一种高效且开源的姿态估计网络，能够同时检测图像中单人或多人场景中的人体关键点，并具有较强的实时性和鲁棒性，本节使用OpenPose算法估计人体2D关键点；PointPillars[11]是一种基于点云的高效三维目标检测网络，其核心设计通过将点云编码为伪图像并利用二维卷积网络实现快速推理。本文选用PointPillars网络，使用自定义仿真数据集对其进行训练。本文方法框架如图3所示。

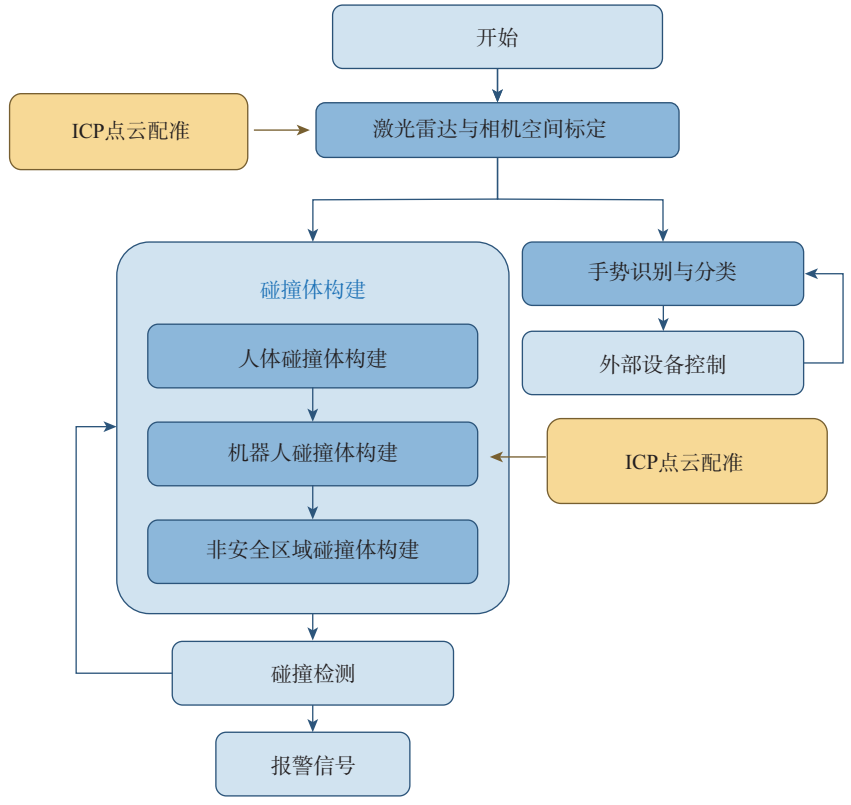


图2 人机交互与安全检测整体方法框架

Fig.2 Overall method framework for human-robot interaction and security detection

首先，将深度相机的彩色图输入到Openpose网络中，得到人体的2D图像关键点；其次，建立仿真数据集生成机制，自定义仿真数据集对PointPillars网络进行训练，网络收敛后将激光雷达点云输入到PointPillars网络中，得到人体包围盒，之后，根据深度相机的深度图与人体包围盒，计算人体3D关键点，并完成人体碰撞体构建。上述过程中，仿真数据集生成、人体3D关键点计

算、人体碰撞体构建3个部分分别在2.1至2.3小节中进行详细描述。

2.1 仿真数据集生成

由于面向航空航天领域的专用公开数据集较少，在现实世界采集真实数据集需要较多的精密设备，对时间成本与经济成本要求较高，因此本文基于UnReal游戏引擎，建立了人机装配环境下人体的仿真环境与数据集收集系统，从而达到便捷收集数据集的目的。本文建立了仿真数据集生成系统，使用仿Kitti数据集的格式收集数据，采集目标包含激光雷达点云（*xyz*格式）与标签信息，其中标签信息包含：类别、3D目标的高度，宽度，长度、3D目标框的中心坐标(x,y,z)以及目标朝向角。其中，由于只估计人体包围盒，因此类别只有人体一类，其他数据均由虚幻引擎计算得到。

设置仿真场景如图4所示。整体仿真环境由模拟激光雷达、人体以及相关设备工装组成，模拟激光雷达通过虚幻引擎中的射线检测机制获取射线在物体表面的坐标，通过射线检测模拟真实激光雷达的激光束，进而得到模拟点云。

仿真数据集的生成过程如图5所示。首先，对仿真场景中的人体模型的坐标进行随机采样，随机改变人体模型的位置、姿势（空闲，奔跑等），但始终保持人体模型始终位于地面上方；其次，触发模拟激光雷达的点云获取功能，得到场景点云。由于模拟激光雷达的位姿与地面不平行，为满足后续网络训练的数据集要求，将整体点云进行坐标转换，将激光雷达坐标系与地面平行，变换为kitti格式；之后，将人体包围框进行同样的位姿变换，得到最终点云中人体包围框的位姿；最后，将人体高度，宽度，长度和人体中心坐标(x,y,z)，以及目标朝向角度等信息进行计算组合得到仿真数据集。图6为部分仿真数据集，其中点云中红色包围框内部

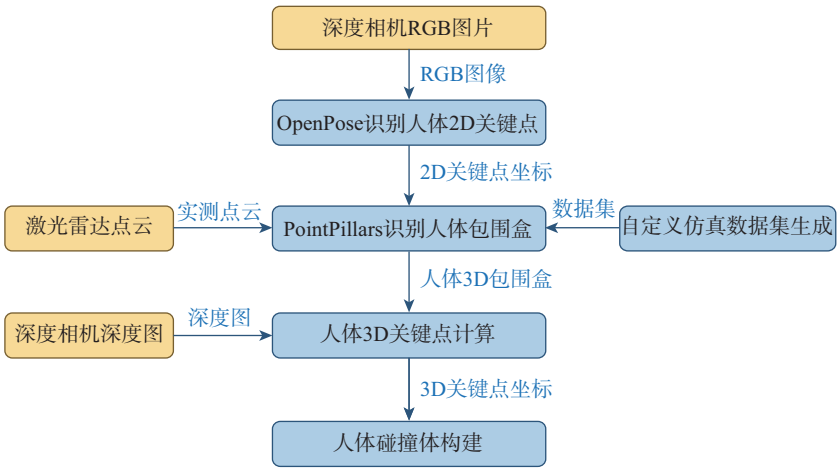
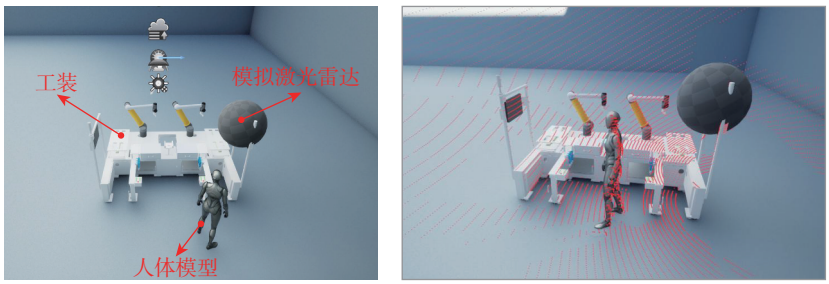


图3 融合Openpose与PointPillars的人体碰撞体构建方法框架

Fig.3 Framework for constructing human collision body by integrating OpenPose and PointPillars



（a）场景设置 （b）模拟激光雷达散发射线采集点云

图4 仿真场景设置

Fig.4 Simulation scenario setting

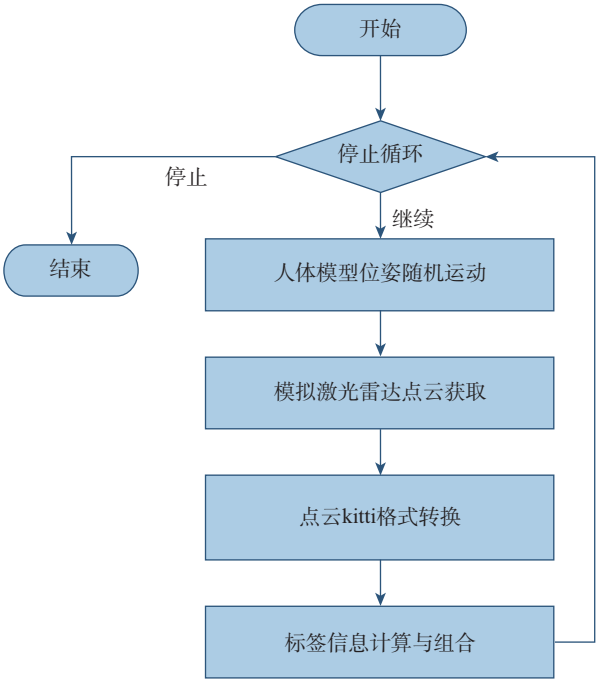
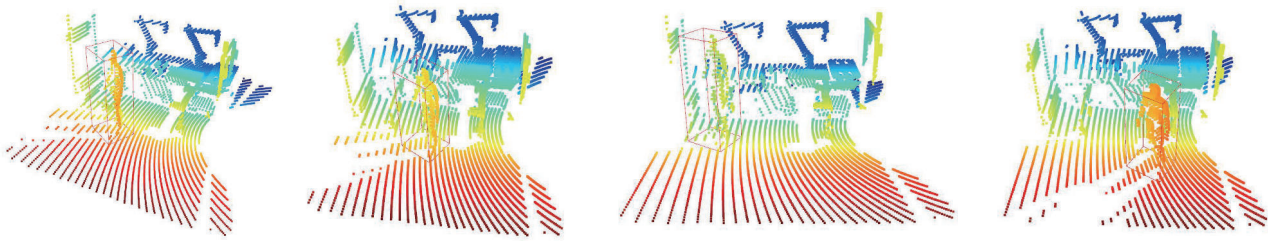


图5 仿真数据集生成流程

Fig.5 Simulation dataset generation process



（a）示例1　（b）示例2　（c）示例3　（d）示例4

图6　部分仿真数据集

Fig.6　Partial simulation datasets

为人体模型的点云。

2.2　人体3D关键点计算

人机装配场景中难免存在人体被障碍物遮挡的情况，在遮挡情况下人体姿态估计虽然能够预测出大致像素位置，但是其对应的2D关键点位于障碍物表面，无法通过深度映射至人体表面，由于实际环境较为复杂，在实际场景中存在物体遮挡人体的情况，如图7所示，当人站在工装旁时，腿部关键点会被工装遮挡，3D关键点的获取更为困难，本节的目的是依据人体2D关键点与人体包络体，计算人体3D关键点。

人体2D关键点可以准确估计出遮挡下的关键点坐标，但是由于相机采集的深度图是可见范围内物体表面的深度，因此在人体被遮挡的情况下，使用传统的深度图映射的方式，只能得到2D关键点在障碍物上的投影，图7中P_2^{2d}关键点对应的3D关键点投影$P_2^{3d-obscale}$，这种偏差会导致人体碰撞体出现畸形，因此亟须解决人体3D点的正确估计问题。

图8展示了人体3D关键点估计的算法框架，首先，对所有人体2D关键点进行深度映射，可以得到人体表面与障碍物上的3D关键点坐标，深度映射的公式为

$$\frac{x}{z}f_x+c_x=\mu \tag{1}$$

$$\frac{y}{z}f_y+c_y=\upsilon \tag{2}$$

式中，(μ,υ)为二维关键点坐标；z为对应像素上深度图的取值；f_x,f_y,c_x,c_y为相机对应的内参，针对每个2D关键点(μ,υ)，都可计算得到对应的3D关键点坐标(x,y,z)。

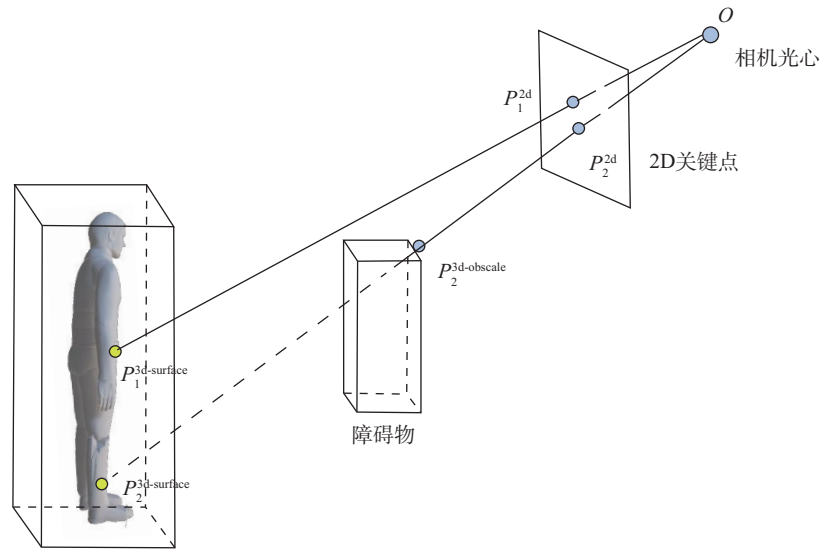
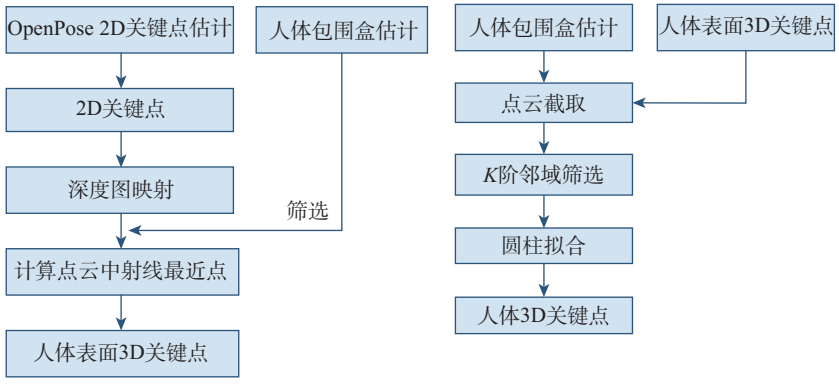


图7　障碍物遮挡下的深度映射

Fig.7　Depth mapping under obstacle obstruction



（a）人体表面3D关键点计算　（b）人体3D关键点计算

图8　人体3D关键点计算过程

Fig.8　Calculation process of 3D key points in the human body

其次，对每个3D关键点坐标(x,y,z)进行筛选，根据3D人体包围盒，判断3D点是否在包围盒内部，若3D点坐标在包围盒内部，则证明此点无障碍物遮挡，记为人体表面3D关键点；若3D点坐标在包围盒外部，则此点受到障碍物遮挡，坐标处于障碍物表面，如图7中$P_2^{3d-obscale}$。针对受到障碍物遮挡的3D关键点，通过摄像机模型建立相机光心到2D关键

点的射线，如图 8 中的 $O-P_2^{2d-obscale}$ 所示。计算公式为

$$\boldsymbol{K}=\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

式中，$\boldsymbol{K}$ 为相机内参矩阵，从像素坐标转换到归一化相机坐标（去畸变后），相机坐标系下的方向向量 $\boldsymbol{p}_c$ 可表示为

$$\boldsymbol{p}_c=\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix}=K^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}=\begin{bmatrix} \dfrac{u-c_x}{f_x} \\ \dfrac{v-c_y}{f_y} \\ 1 \end{bmatrix} \tag{4}$$

式中，X_c 为相机坐标系下 p 点的 x 坐标；y_c 为相机坐标系下 P 点的 y 坐标；u 为像素坐标系下 p 点的 x 坐标；v 为像素坐标系下 p 点的 y 坐标。相机坐标系下的射线可表示为

$$r(\lambda)=\lambda\begin{bmatrix} \dfrac{u-c_x}{f_x} \\ \dfrac{v-c_y}{f_y} \\ 1 \end{bmatrix} \tag{5}$$

式中，$\lambda>0$ 为射线参数，决定了点在射线上的具体位置，射线起点是相机光心（0,0,0）方向向量为 $\left((u-c_x)/f_x,(v-c_y)/f_y,1\right)$。计算完毕射线公式后，计算包围盒内点云到射线的最短距离，取点云中距离射线最近的点为新的人体表面 3D 关键点。

通过上述算法过程，完成图 8（a）中人体表面 3D 关键点计算，去除了外部遮挡影响的关键点畸变的情况，但是人体 3D 关键点投影位于人体表面，存在于人体衣服上，无法深入骨骼以进行碰撞体建立，图 9 中展示了人体 3D 关键点的计算流程。

图 9 左侧展示了人体表面 3D 关键点在点云上的分布情况，本文将人体每个骨骼连杆单独处理，进行如图 8（b）所示的点云截取、k 阶邻域筛选、圆柱拟合等操作，例如，对于骨骼连杆 10–11，已知人体表面 3D 关键点 $P_{10}^{3d-surface}$，$P_{11}^{3d-surface}$，在人体包络体点云内，建立垂直于 $P_{10}^{3d-surface}P_{11}^{3d-surface}$ 连线，分别经过 $P_{10}^{3d-surface}$，$P_{11}^{3d-surface}$ 的两平面，如图 10 中的平面 1 与平面 2，使用点云滤波保留两平面内部的点云，建立两个人体关键点投影的中点 P_{middle}，以中点 P_{middle} 为中心，进行 k 阶邻域搜索，通过设定阈值范围将骨骼连杆 10–11 附近的点云提取出来，最后进行最终的点云圆柱拟合，得到圆柱体的上下两底面中心，作为人体 3D 关键点 $P_{10}^{3d-bone}$、$P_{11}^{3d-bone}$。

除了人体表面 3D 关键点 $P_0^{3d-surface}$，其他骨骼连杆同理也可以计算出 3D 关键点。人体表面 3D 关键点 $P_0^{3d-surface}$ 代表了人体结构中的头部，点云结构与球体类似，因此本文以 $P_0^{3d-surface}$ 为中心，对包围框内的点云进行球体截取，之后将截取出的点云进行点云球体拟合得到人体 3D 关键点 $P_0^{3d-bone}$。

2.3 人体碰撞体模型构建

由于人体属于柔性结构，并且不同的人体型大小不同，因此无法使用固定的数模对其碰撞体进行表征，本小节使用以下分体圆柱结构对人体碰撞体建模，如图 10 所示。

与人体 3D 关键点计算环节一致，将人体 3D 关键点拆解为骨骼连杆单独考虑，除了头部以外，将全部骨骼连杆全部建模为圆柱体，圆柱体的上下底面中心是骨骼 3D 点，半径为人体 3D 关键点计算环节（图 8（b））中的点云拟合半径，头部使用球体建模，半径根据人体 3D 关键点计算环节中的点云球体拟合半径确定。为了进一步确保人机安全，人体碰撞体中圆柱体半径、球体半径均可进行倍数扩大，弥补点云拟合过程中不可避免的误差损失，保障人机安全。

3 机器人碰撞体与非安全区碰撞体构建

为保障机器人与人体、环境之间的安全，准确估计机器人运动与建立机器人碰撞体必不可少，由于机器人运动过程中关节角度可以通过控制器实时获取，因此只需完成

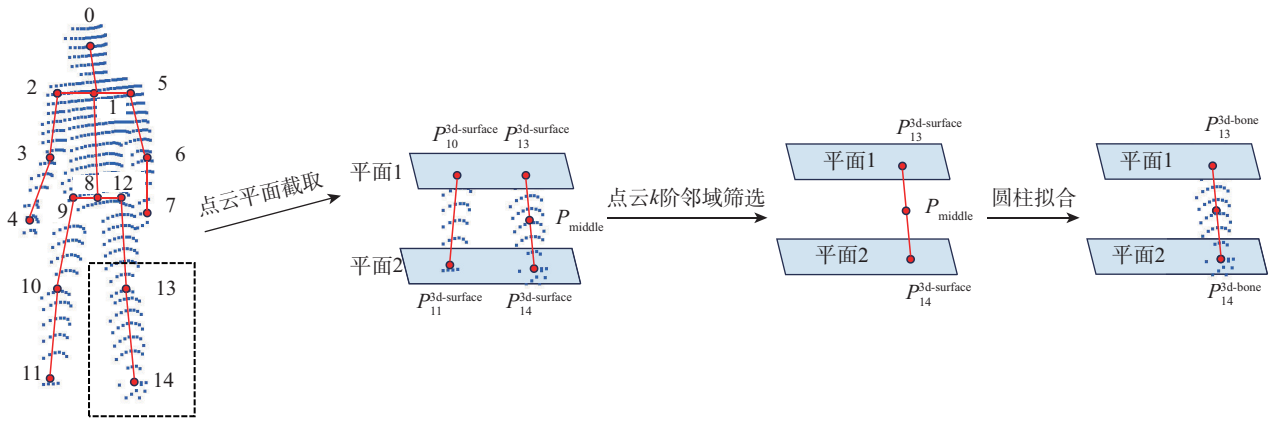


图 9 人体 3D 关键点计算

Fig.9 Calculation of key points in 3D human body

机器人基坐标系与激光雷达坐标系之间的标定即可获取机器人在激光雷达坐标系下的各关节位姿。

针对上述问题，本文采用点云配准的思路解决上述问题，首先通过采样将标准机器人数模转化为点云，其次将数模中机器人基坐标系与激光雷达之间的理论转换关系作为点云配准初始值，之后使用ICP精配准算法将机器人数模点云逐步逼近实际点云，完成机器人基坐标系的标定。

在非安全区碰撞体构建方面，为防止工人在旋转工装运行时受伤，需要对工艺过程中的每个工步规定危险区域，并警示工人远离。为满足上述需求，本文建立了根据工步划分的非安全区碰撞体，通过人体碰撞体与非安全区碰撞体之间的实时碰撞检测监控安全状态，建立了系统内部定义与外部STL导入2种非安全区碰撞体的构建方式。

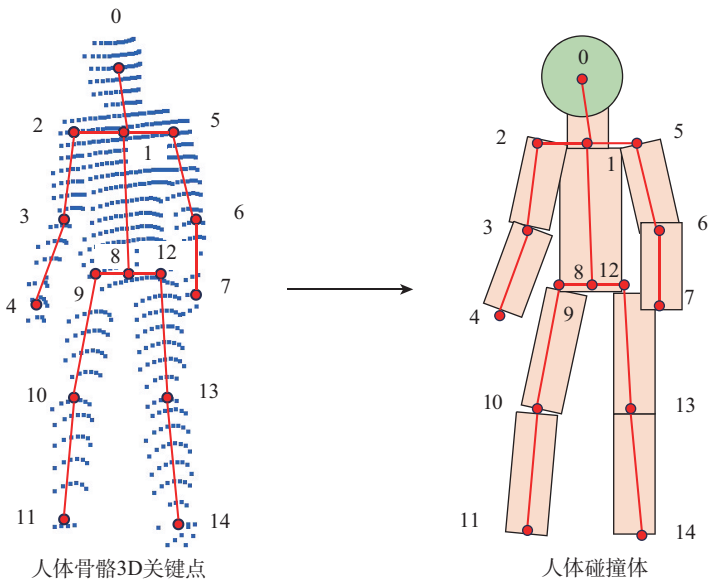


图10 人体碰撞体构建

Fig.10 Construction of human collision body

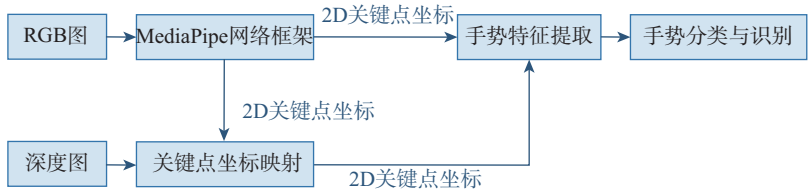


图11 手势识别框架

Fig.11 Hand gesture recognition framework

4 手势识别与分类判定

4.1 手势识别

本节以深度相机作为手势监测传感器，以彩色图像与深度图像为输入，通过基于MediaPipe的手部关键点识别网络，估计手部关键点的2D与3D关键点，其次通过手部特征提取，得到手势特征值，作为手势分类识别的依据。整体识别框架如图11所示。

MediaPipe手部关键点识别网络是一种基于深度学习的实时视觉分析框架[12–13]，能够在低延迟条件下实现高精度的手部姿态估计。本文选用MediaPipe主要原因在于此网络采用轻量级网络结构，适用于移动端及嵌入式设备，能够在实时帧率（30+FPS）下运行。在得到手部关键点二维坐标后，为了提高手势识别的精准性，本文将二维坐标与深度图相映射，得到二维关键点对应的三维坐标，为手势识别分类提供更多判定依据。三维关键点的计算方式参考式（1）。

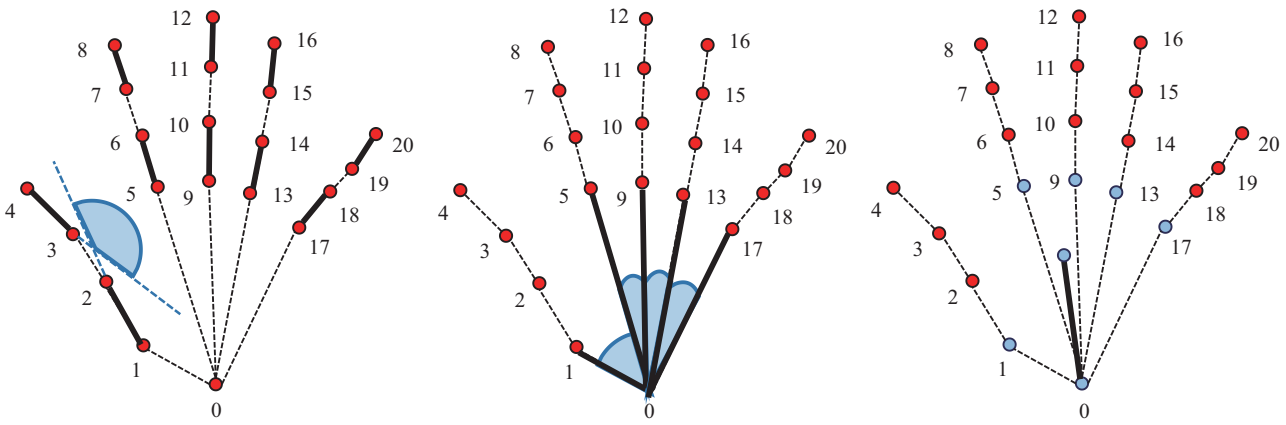
4.2 手势分类判定

在得到手部关键点的二维坐标与三维坐标后，需要依据坐标数据对不同手势进行特征提取与编码，用以区分不同手势。基于手掌特征，本文通过计算每根手指的手指弯曲角度、相邻手指之间的分叉角度以及手掌的绝对方向向量等20个维度特征值，对手势含义进行判断。

手指的手指弯曲角度计算原则为：以图12（a）为例，大拇指的2D手指弯曲角度为关键点连线1–2直线与关键点连线3–4直线的夹角；食指的2D手指弯曲角度为关键点连线5–6直线与关键点连线7–8直线的夹角；其他手指的2D手指弯曲角度以此类推，3D手指的弯曲角度则将对应的关键点计算为3D坐标，求取空间夹角即可。

相邻手指之间的分叉角度计算原则为：以图12（b）为例，大拇指与食指之间的夹角为关键点连线0–1与关键点连线0–5之间的夹角；食指与中指之间的夹角为关键点连线0–5与关键点连线0–9之间的夹角；其他以此类推，3D手指的弯曲角度则将对应的关键点计算为3D坐标，求取空间夹角即可。

上述指标计算的是相对手势，与手部在三维空间中的位置无关，为了更加准确约束与识别特定手势，使用手掌的绝对方向向量计算手掌的绝对方向，保障手势识别的准确性。以图12（c）为例，针对一只手，手掌的绝对方向向量计算原则如下，将关键点1，5，9，13，17的二维坐标加和取平均，得到的2D点与关键点0之间的连线的单位向量作为手掌的2D绝对方向向量，当手掌的朝向发生改变时，绝对方向向量



（a）每根手指弯曲角度　（b）相邻手指之间分叉角度　（c）手掌的绝对方向向量

图12　手势分类识别特征值

Fig.12　Hand gesture classification recognition feature values

也会发生改变；3D手掌的绝对方向向量的计算同理，只是将2D关键点换为3D关键点即可。

在手势分类时，建立如下判定原则对未知手势进行判定：(1)计算未知手势的20个特征值；(2)将未知手势的20个特征值与手势模板库中的20个特征值进行比对，当每个特征值之间的差距小于规定阈值时，判定手势含义为对应模板库中的手势含义。

5　试验与结果

5.1　试验环境

本文建立了如图13所示的某飞行器类航天产品真实试验环境，工装及机器人等硬件装备均用于飞行器类产品的装配过程，本文通过布设相机及激光雷达，在此环境下对上述方法进行验证试验，详细试验过程在5.2~5.4小节中进行描述。

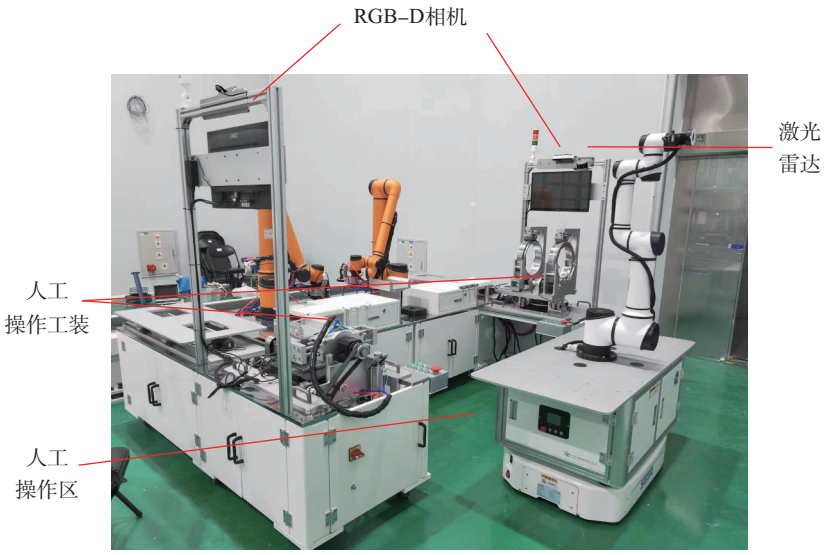


图13　某飞行器类航天复杂产品装配试验环境

Fig.13　Assembly experiment environment for a certain complex aerospace product of aircraft type

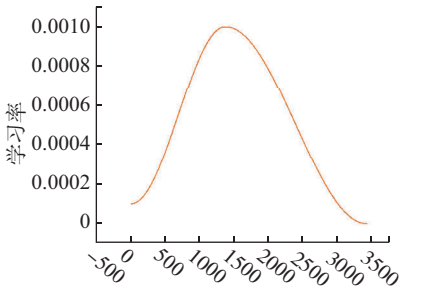
5.2　人体包围盒识别效果分析

本文在完成自定义数据集构建后，使用OpenPCDet对自定义数据集进行训练。网络训练与测试所采用的计算机配置为：CPU：Intel(R) Core(TM) i9 – 14900KF；GPU：RTX4090；RAM：64GB。

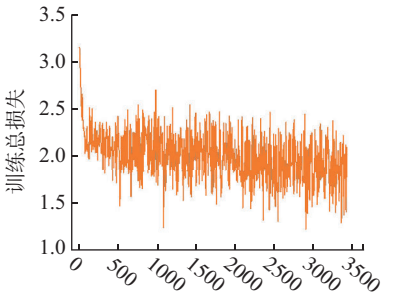
图14展示了基于PointPillars模型对自定义数据集训练过程中指标的变化，其中，图14(a)展示了学习率指标随着训练过程的进行而变化的过程，学习率指标在训练初期逐渐增加，在训练稳定后逐渐开始下降，这种策略有利于避免模型在初期的不稳定，模型在初始阶段需要逐步适应数据特性，可提升训练稳定性；训练稳定后逐渐开始下降是因为在模型接近收敛时，大幅降低学习率可避免参数在最优解附近震荡，从而更精细地调整参数。图14(b)~(d)分别展示了训练总损失、分类损失、位置损失随着训练次数增加产生的变化，其中，分类损失用于衡量模型对物体类别预测的准确性，如每个包围盒内是否包含目标，目标的类别是否正确；位置损失用于衡量模型对物体位置、尺寸和方向的误差，包含包围盒中心坐标、包围盒长宽高尺寸、包围盒方向；训练总损失是分类损失和位置损失的加权和。可以看出，随着训练次数的增加，训练总损失、分类损失以及位置损失都在逐步降低，表明模型在训练过程中学习到了规律，表明了模型对自定义数据集的训练有效性。

图15展示了训练好的模型在真实激光雷达点云中的表现，可以观

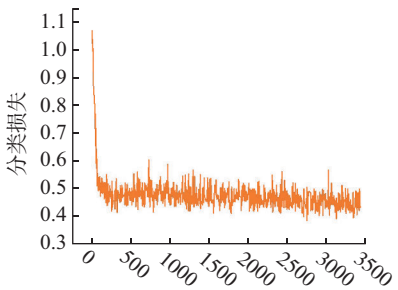
察到模型识别到的绿色包围盒完全将人体点云包围，即使人体在点云中处于不同位置，模型的识别结果依然准确，可以表明基于激光雷达点云的人体包围盒估计方法的有效性。



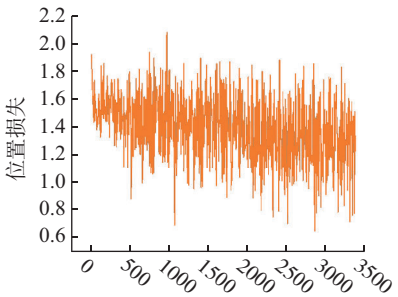
（a）学习率



（b）训练总损失



（c）分类损失



（d）位置损失

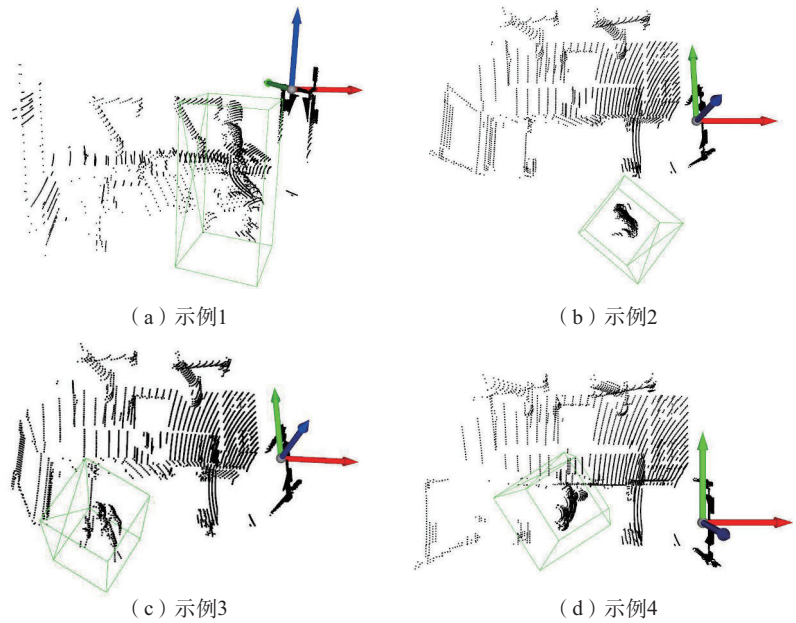
图14 人体3D包围盒估计网络在训练过程中的指标变化

Fig.14 Index changes of the human 3D bounding box estimation network during the training process

5.3 手势识别与分类方法验证

为验证基于手势的人机交互的有效性与鲁棒性，试验过程中制定了如图16所示的6种手势，用于测试基于手势的人机交互的识别正确率。

图17展示了测试时的手势效果，手势识别与判定方法对预先定义的6种手势全部进行了特征提取与手势识别，并绘出了手部关节点的连线，识别结果均与预定义手势含义相同，表明方法的识别稳定性，本文也进行了真实试验环境下的试验验证，试验过程中的手势识别结果成功率为100%，验证了基于手势



（a）示例1　（b）示例2

（c）示例3　（d）示例4

图15 人体包围盒识别结果（绿色包围框为识别到的人体包围盒，黑色为整体点云）

Fig.15 Recognition results of human body bounding box (green bounding box represents the recognized human body bounding box, black represents the overall point cloud)



（a）确定　（b）竖直锁定　（c）竖直解锁

（d）水平锁定　（e）水平解锁　（f）水平自动旋转

图16 6种手势

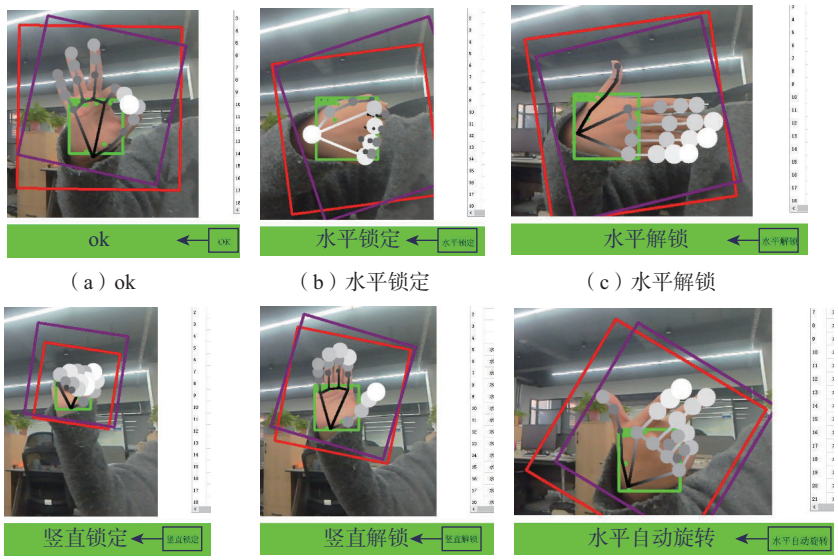
Fig.16 Six hand gestures

识别的人机交互的有效性。

5.4 人机安全监测功能验证

人机安全监测试验分为人与非安全之间的安全监测以及人与机器人之间的安全监测。图18展示了人与非安全区的安全测试过程，首先，如图18(a)所示，在CATIA软件中设定每个工步的每个工序的非安全区，图18(a)使用盒状体表征非安全区；之后，导入到系统软件中(图18(b))，与现场实际标定数据相匹配，生成非安全碰撞体；最后，将人体碰撞体与非安全区碰撞体进行实时碰撞检测，并输出两者之间的实时最小距离，图18(c)展示了非安全与人体之间的最小距离；图18(d)展示了人与非安全区之间存在的碰撞反馈，可以看出已经发生碰撞。通过设定非安全区域、人体姿态估计与实时碰撞检测，经过上述试验过程，可以证明人机共融安全系统的安全有效性。

图19展示了人与机器人之间的安全测试，可以观察到人体碰撞体与机器人碰撞体，软件界面右侧显示了两者之间的实时最小距离，可以表明人与机器人之间存在最小距离的安全碰撞监测。



(a) ok　(b) 水平锁定　(c) 水平解锁

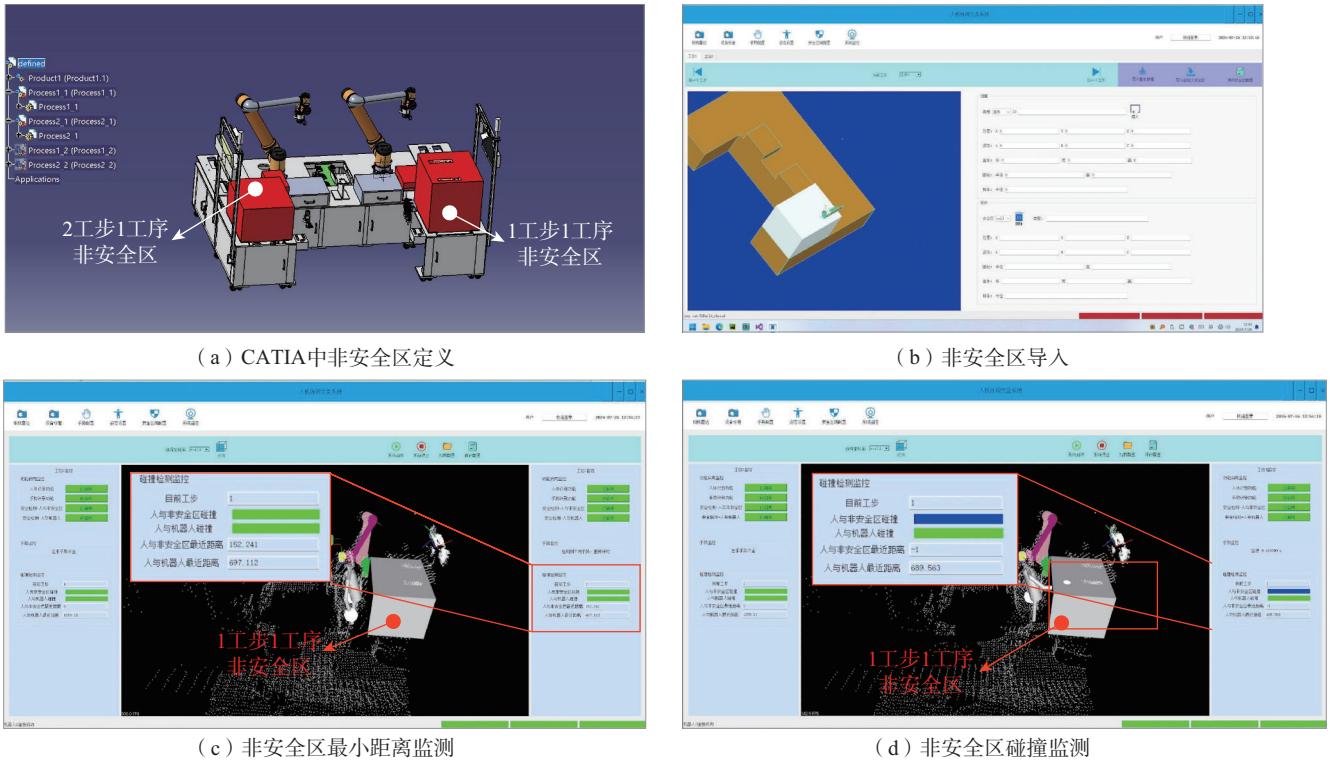
(d) 竖直锁定　(e) 竖直解锁　(f) 水平自动旋转

图17 手势测试效果

Fig.17 Hand gesture test effect

6 结论

本文针对复杂产品人机协作装配过程，开展了人机交互及安全监测方法研究。针对装配人员3D位姿估计困难、装配现场遮挡频发的问题，提出了一种融合Openpose与



(a) CATIA中非安全区定义　(b) 非安全区导入

(c) 非安全区最小距离监测　(d) 非安全区碰撞监测

图18 人与非安全区的安全测试过程

Fig.18 Safety testing process between humans and non safe areas



图19 人与机器人的安全测试

Fig.19 Safety testing of humans and robots

PointPillars的人体3D碰撞体构建方法，提高了装配过程人体识别准确率；提出了非安全区域概念及碰撞体构建方法，基于点云配准与机器人实时关节信息实现了机器人碰撞体的构建，保障了装配人员–设备–产品共融环境下的安全；提出了基于Mediapipe的手势识别与分类方法，优化了人与设备之间的实时效率。通过试验证明，提出的系统与方法能够满足现场装配人员的实际装配需求，提高了装配系统的效率与安全性。

参 考 文 献

[1] Blanco D, Rubio E M, Agustina B, et al. Advanced procedures and scheduling for aircraft assembly processes: A systematic review approach[J]. Procedia CIRP, 2024, 126: 817–822.

[2] Nie Q W, Shen Y P, Ma Y, et al. Probing augmented intelligent human-robot collaborative assembly methods toward industry 5.0[J]. Electronics, 2025, 14(15): 3125.

[3] Li W D, Hu Y D, Zhou Y, et al. Safe human-robot collaboration for industrial settings: A survey[J]. Journal of Intelligent Manufacturing, 2024, 35(5): 2235–2261.

[4] Cirillo A, Ficuciello F, Natale C, et al. A conformable force/tactile skin for physical human-robot interaction[J]. IEEE Robotics and Automation Letters, 2016, 1(1): 41–48.

[5] Phan S, Quek Z F, Shah P, et al. Capacitive skin sensors for robot impact monitoring[C]//2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco: IEEE, 2011: 2992–2997.

[6] Haddadin S, Albu-Schaffer A, De Luca A, et al. Collision detection and reaction: A contribution to safe physical human-robot interaction[C]//2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Nice: IEEE, 2008: 3356–3363.

[7] Xu T, Tuo H, Fang Q Q, et al. A novel collision detection method based on current residuals for robots without joint torque sensors: A case study on UR10 robot[J]. Robotics and Computer-Integrated Manufacturing, 2024, 89: 102777.

[8] Vogel C, Poggendorf M, Walter C, et al. Towards safe physical human-robot collaboration: A projection-based safety system[C]//2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco: IEEE, 2011: 3355–3360.

[9] Pilz. Safe Camera system safetyeye, monitoring and control with a single safe camera system [EB/OL]. [2025−11−02]. https://www.pilz.com.

[10] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 1302–1310.

[11] Lang A H, Vora S, Caesar H, et al. PointPillars: Fast encoders for object detection from point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 12689–12697.

[12] Zhang F, Bazarevsky V, Vakunov A, et al. MediaPipe hands: On−device real−time hand tracking[PP/OL]. arXiv [2025−11−02]. https://arxiv.org/abs/2006.10214.

[13] Lugaresi C, Tang J Q, Nash H, et al. MediaPipe: A framework for building perception pipelines[PP/OL]. arXiv [2025−11−02]. https://arxiv.org/abs/1906.08172

An Interaction and Safety Monitoring Method Under Human-Robot Collaborative Assembly of Complex Products

MENG Jinlong[1], WANG Zhongkang[2], LI Fei[3], CAO Zhihong[3], HU Lan[2], LI Pengcheng[1], TIAN Wei[1]

(1. Nanjing University of Aeromautics and Astronautics,Nanjing 210016,China;
2. Shanghai Aerospace Equipment Manufacturing Co., Ltd, Shanghai 200240, China;
3. Beijing General Institute of Electronic Engineering, Beijing 100039, China)

[ABSTRACT] With the advancement of manufacturing technologies, aerospace assembly has gradually transitioned from manual operations to human–robot collaborative and automated assembly. Efficient human–robot interaction and reliable safety monitoring are therefore essential. This study proposes an integrated human–robot interaction and safety monitoring system for aerospace collaborative assembly. To address challenges in human 3D pose estimation under frequent occlusions, a human 3D collision body construction method integrating OpenPose and PointPillars is introduced. Non-safety zones and corresponding collision models are defined to enable real-time human–robot safety monitoring, while robot collision bodies are constructed using point cloud registration and real-time joint information. In addition, a MediaPipe-based gesture recognition method is developed to improve interaction efficiency and reduce learning costs. Experimental results demonstrate that the proposed system effectively enhances assembly efficiency and operational safety.

Keywords: Assembly of complex products; Human-robot collaboration; Human-robot interaction; Safety monitoring; Gesture recognition

（责编　晓月）