

基于 MediaPipe 的手势识别用于 挖掘机遥操作控制

王如斌¹ 窦全礼^{1,2} 张淇¹ 周诚¹

(1. 华中科技大学 土木与水利工程学院, 武汉 430074;
2. 潍柴动力股份有限公司, 潍坊 261061)

【摘要】 挖掘机有十分广泛的应用场景,但在某些危险工况下,对其采用遥操作是更好的选择。同时,近年来随着机器视觉和深度学习的发展,涌现出一系列手势识别的算法和框架。为了探究基于 MediaPipe 的手势识别算法在挖掘机遥操作中的应用,本文将手部不同手势与挖掘机的不同动作相对应,提出一种利用手势实现挖掘机遥操作的新型控制方式。在实验室环境下,对实验用挖掘机机器人进行运动学分析,采用 MediaPipe 对 15 种手势类型进行实时静态识别,生成指令并利用 Arduino 单片机实现对挖掘机的控制。结果显示,该系统具有良好的性能,可以用于挖掘机的遥操作控制,为挖掘机的远程控制提供了一种新型的人机交互形式。

【关键词】 手势识别; MediaPipe; 挖掘机; 单片机; 遥操作

【中图分类号】 TU17; TU6 **【文献标识码】** A **【文章编号】** 1674-7461(2022)04-0009-08

【DOI】 10.16670/j.cnki.cn11-5823/tu.2022.04.02

引言

挖掘机是普遍应用于土石方工程施工的工程机械之一,传统的人工操作挖掘机的方式有其局限性:一方面,挖掘机操作员在执行重复任务时处在噪声与振动环境中,需要时刻保持专注,以便进行高效的工作^[6-7];另一方面,挖掘机通常在比较恶劣的环境下作业,有时被应用于污染土壤修复、抢险救灾等危险工况,此类情况下如采用人工现场操作挖掘机的作业方式,会对操作者的人身安全构成威胁^[3],因此需要找寻一种远程、无人的作业方式。挖掘机的机器人化是现在急需解决的问题,在全自动智能挖掘机研制出来之前,通过遥操作控制挖掘机在未来相当长的一段时间内都会占有重要的地位,其中基于人机交互的遥操作方式的设计与研究也受到了广泛的关注^[4]。

手势交互是一种重要的自然交互方式,因而手势识别成为人机交互的一个重要分支领域^[5]。近些年来,国内外的研究机构曾成功开发了大量能够实现遥操作的设备和算法,目前存在一些将手势识别和遥操作结合的相关研究,大致可分为基于穿戴设备的手势识别^[1-2]和基于视觉的手势识别^[8]两大类。后者又可分为静态手势识别和动态手势识别两类,静态研究时间节点上的图像,动态识别时序图像^[9]。

本文针对传统的人工现场操作挖掘机的方式与现存的挖掘机遥控方式中存在的控制精度低、复合作业程度低等缺点,提出一种基于 MediaPipe 机器视觉的手势识别用于挖掘机遥操作控制的方案。该遥操作方式能较好地满足挖掘机作业的需求,为挖掘机的远程控制提供了一种更人性化的人机交互形式,具有实际的应用价值和研究价值。

【基金项目】 国家自然科学基金面上项目“工程安全风险预警致因机理与智能预警方法研究”(编号:72171092);湖北省自然科学基金杰出青年基金项目“盾构施工安全风险智能控制”(编号:2021CFA091);湖北省重大科技专项“智能建造关键技术及应用”(编号:2020ACA006)

【第一作者】 王如斌(1998-),男,在读硕士研究生,主要研究方向:数字建造与工程安全。

【通讯作者】 窦全礼(1983-),男,助理工程师,战略规划部副部长,主要研究方向:交通工程、智能施工装备。

1 基于 MediaPipe 的手势识别用于挖掘机遥操作的方法框架

反铲挖掘机机器人遥操作系统由单目相机、挖掘机、远程操作硬件系统和远程操作软件系统构成,如图 1 所示。在整个人机交互系统当中,单目相机是主要的外界信息输入设备,其通过采集数据图像并实时传输给手势识别模块,通过 MediaPipe 算法生成控制指令;挖掘机作为被控制的对象,操作人员可以在工作场地或者通过外设监视器观察挖掘机的工作情况,操作挖掘机,进而对挖掘机的运动状态进行调整。

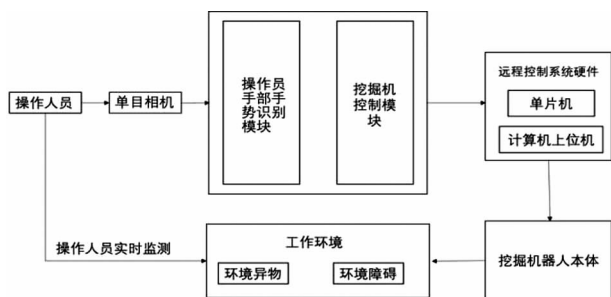


图 1 交互系统示意图

主端(操作人员 + 单目相机 + 计算机)通过改变手势,能够实现控制从端(挖掘机)完成各种工作所需的动作。从端做出与预期不符的动作时,操作者能够及时调整。在熟悉数据手套操作后,挖掘机操作人员能够通过相对省力、简便的动作实现与使用操作杆同样的控制水平。

2 实验用挖掘机遥操作控制系统

2.1 工作装置结构

反铲液压挖掘机的工作装置由铲斗、铲斗液压油缸、斗杆、斗杆液压油缸、动臂、动臂液压油缸和连杆机构组成,且均被铰接,通过缸体的膨胀和收缩来驱动杆,以执行挖掘过程中所需的各种动作,如图 2 所示。动臂、斗杆和铲斗的动作可分别进行,也可同时进行,挖掘时动作灵活、效率较高。正常挖掘机液压挖掘机在正常排气运动工作时,机体通常通过依靠三组液压油缸的动力支持而分别移动。

2.2 运动学分析

(1) 基本原理

采用 D-H(Denavit-Hartenberg)坐标变换法^[10]对实验用挖掘机模型进行运动学分析,通过对工作

装置的齐次坐标变换建立工作装置运动方程。如图 3,平行于关节轴建立两个标准坐标系 $x_{i-1}y_{i-1}z_{i-1}$ 和 $x_iy_iz_i$,在运用 D-H 法建模时,有严格的固定的定义变量、参数和表示字符的方式。

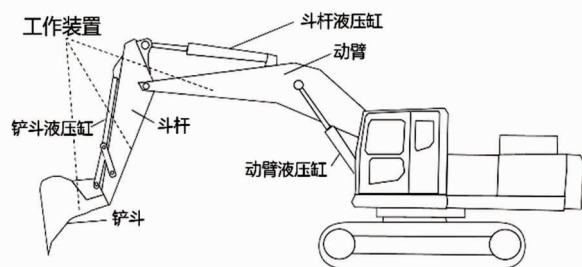


图 2 挖掘机整体结构简化图

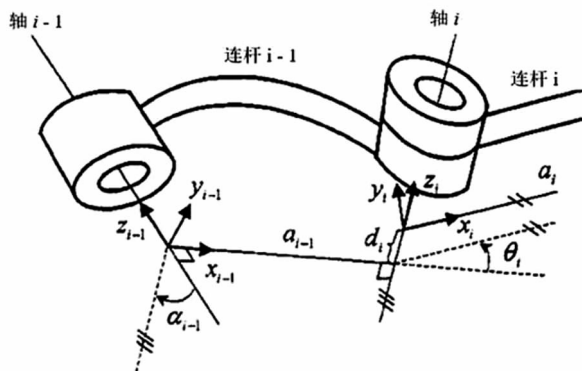


图 3 标准 D-H 法连杆坐标系

连杆参数作定义如下:

a_{i-1} :沿着 x_{i-1} 轴测量的从 z_{i-1} 轴到 z_i 轴的距离

α_{i-1} : z_{i-1} 轴和 z_i 轴绕 x_{i-1} 轴的角

d_i : z_i 轴测量的从 x_{i-1} 轴到 x_i 轴的距离

θ_i : x_{i-1} 轴和 x_i 轴绕 z_i 轴的角

两相邻坐标变换时矩阵为:

$$M = \begin{bmatrix} \cos\theta_i & -\cos\alpha_i \sin\theta_i & \sin\alpha_i \sin\theta_i & a_i \cos\theta_i \\ \sin\theta_i & \cos\alpha_i \cos\theta_i & -\cos\alpha_i \sin\theta_i & a_i \sin\theta_i \\ 0 & \sin\theta_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2-1)$$

右乘坐标变换矩阵的几何意义即是相邻坐标系之间的转换,故重复右乘的操作可完成多个相邻坐标系之间的转换。

(2) 连杆坐标系

根据 D-H 坐标变换法建立挖掘机模型工作装置连杆坐标系如图 4 所示:

其中, O_0 为旋转中心, O_1 为动臂铰接点, O_2 为

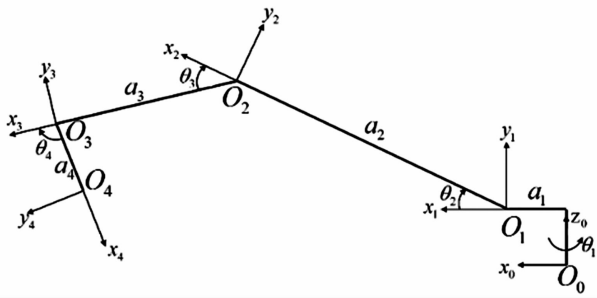


图4 挖掘机工作装置连杆坐标系简图

斗杆铰接点, O_3 为铲斗铰接点, O_4 为工作装置末端。

由 D-H 法确立各个杆件的参数如表 1 所示。

表 1 工作装置 D-H 参数表

杆件编号 i	杆件长度 a_i	杆件扭角 α_i	关节平移量 d_i	关节回转量 θ_i	关节变量
1	0	0	d_1	θ_1	θ_1
2	a_2	0	0	θ_2	θ_2
3	a_3	0	0	θ_3	θ_3
4	a_4	0	0	θ_4	θ_4

(3) 运动学正解

根据坐标系变换原则, 可以求出由旋转中心 O_0 的基础坐标系表示对工作装置末端 O_4 的位姿矩阵:

$${}^0M_4 = \begin{bmatrix} c_1 c_{234} & -c_1 s_{234} & s_1 & c_1 (a_1 + a_2 c_2 + a_3 c_{23} + a_4 c_{234}) \\ s_1 c_{234} & -s_1 s_{234} & -c_1 & s_1 (a_1 + a_2 c_2 + a_3 s_{23} + a_4 s_{234}) \\ s_{234} & c_{234} & 0 & a_1 + a_2 c_2 + a_3 s_{23} + a_4 s_{234} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2-2)$$

其中,

$$s_i = \sin\theta_i, s_{ij} = \sin(\theta_i + \theta_j), s_{ijk} = \sin(\theta_i + \theta_j + \theta_k),$$

$$c_i = \cos\theta_i, c_{ij} = \cos(\theta_i + \theta_j), c_{ijk} = \cos(\theta_i + \theta_j + \theta_k)$$

3 基于 MediaPipe 的手势识别

Google 基于 MediaPipe 框架实现了一种全新的手部感知方法, 该方法利用机器学习 (ML) 推断出单帧内的 21 个 3D 手部关键点, 如图 5 所示, 从而提供手部与手指的高保真追踪轨迹^[11]。

基于手势识别系统的搭建需要经过三步: 手势检测、手势跟踪、手势识别^[12], 基于 MediaPipe 的手势识别同样包含手掌检测模型、手部关键部位模型、手势识别器三个部分。其中, 手掌检测模型, 用

于从图像中识别手掌, 定义手部边界区域; 手部关键部位模型, 用于对手部边界区域检测并返回 3D 手部关键点; 手势识别器, 则用根据手部关键点定义并识别各种手势。手势识别系统的架构示意图如图 6 所示。

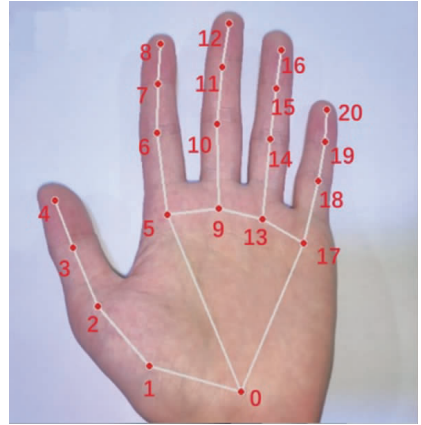


图 5 手部关键点

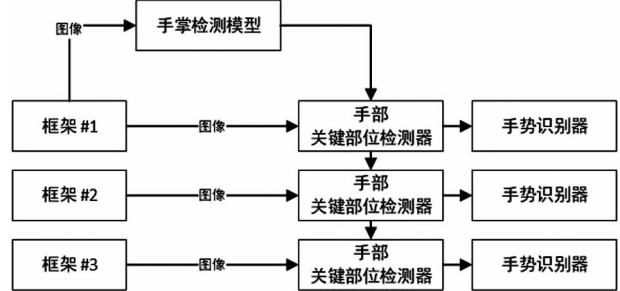


图 6 交互系统示意图

4 手势识别和遥操作实验

4.1 实验装置与环境

手势识别和遥操作实验是在实验室中进行的, 对镭速 AC360 型号的 1:14 仿真合金液压挖掘机进行了模拟沙土挖掘的实验。挖掘机模型的几何尺寸如图 7 所示 (单位: 毫米), 其可通过直流电源驱动电机与油泵, 实现对实际反铲挖掘机的高度仿真。

本文提出的方法是在 AMD Ryzen 5 3600X 6-Core 六核 CPU 和 Nvidia GeForce GTX 760 (2 GB) 的台式电脑上利用 Python 实现的, 当然也可以在任何其他安装了 Python 的 PC 上执行, 需要注意的是要有一个录像规格在 1080 × 720、25FPS 以上的摄像头。此外为定义操作指令、控制具体运行程序从而实现了对挖掘机的控制, 如图 8、图 9 所示, 选用主芯

片为 AT Mega328P (Uno) 单片机的 Arduino 舵机控制器, 其有六个可以用于 PWM 信号输出的接口。

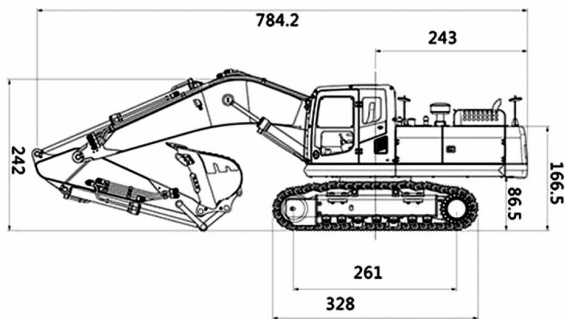


图7 LESU AC360 挖掘机模型几何尺寸

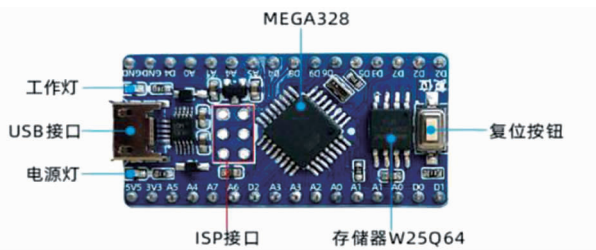


图8 ATmega328P 单片机

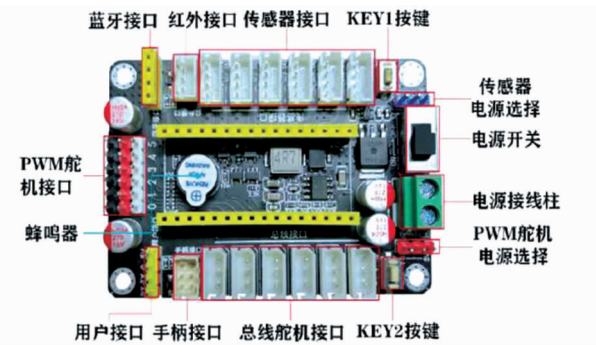


图9 Arduino 舵机控制器接口

实验在两个钢制沙盘中进行, 挖掘机放入右侧沙盘中间, 其前方设置高于沙土面的沙土堆, 以供挖掘机挖掘, 示意图和实物图如图 10 所示。实验过程分为两个主要部分: 手势识别和挖掘机控制器。

4.2 静态手势识别实验

(1) 手势交互设计

本系统手势交互由一只手和单目相机完成, 通过单手手势的变化, 可以对挖掘机器人的行走装置、工作装置进行控制。比如呈竖起大拇指状时, 油泵启动; 呈大拇指和食指张开的手势“八”状时, 车体前进。具体的手势设计如图 11 所示。

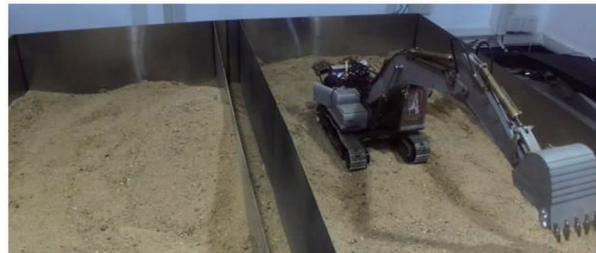
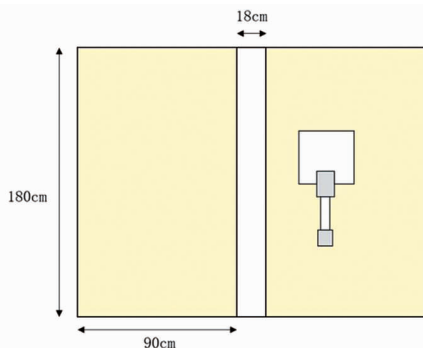


图10 沙盘示意图和实物图



图11 手势交互设计

表3 各手势预测模型的精确率、召回率以及

F1 综合评价指标

	精确率 (%)	召回率 (%)	F1 综合评价指标 (%)
STOP	94.94	100.00	97.40
P-ON	93.93	87.67	90.69
P-OFF	92.38	93.00	92.69
CW	90.38	94.00	92.16
ACW	93.09	94.33	93.71
BO-UP	94.30	93.67	93.98
BO-DOWN	86.38	99.33	92.40
ARM-UP	93.58	92.33	92.95
ARM-DOWN	91.54	97.33	94.35
BU-UP	99.28	91.33	95.14
BU-DOWN	95.61	94.33	94.97
FORWARD	90.51	89.00	89.75
BACK	99.64	92.00	95.67
LEFT	100.00	100.00	100.00
RIGHT	100.00	91.00	95.29

整手指动作对应数据的上下边界,还需要对视频流下手势识别的准确度进行试验测试。在实验中依次做出图 11 所示的 15 个手势,记录计算机中的上位机显示的数据,判断是否对应了正确的识别手势名称,实验过程中静态手势识别图如图 14 所示。

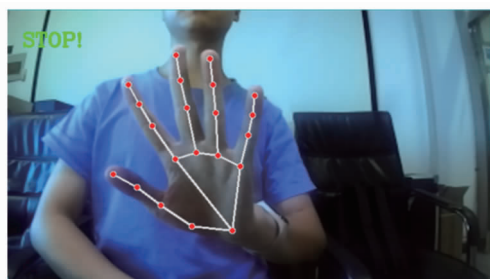


图 14 静态手势识别

4.3 基于手势识别的主从操控实验

挖掘机行走装置与工作装置的运动以及单目相机识别到的手势变化之间的映射关系是本课题的核心内容。为验证本文所设计的基于手势识别控制挖掘机的系统能否正常运行,设计了基于手势识别的主从操控实验。

(1) 主控端设计

在主控端,首先会由单目相机输入视频流,取每秒 25 帧图像进行静态手势识别,将手势信息转换为单片机的串口指令。

在静态手势识别的实验中,可以得到较为稳定的识别结果,但是,仍然存在识别错误的情况。为了消除这种偶现的错误对主从操控实验稳定性的影响,创建了一个手势缓冲器。当手势识别结果达到一定的阈值($\geq 0.5s$),再传输到串口指令生成单元,输出控制指令。

(2) 从控端设计

在挖掘机从控端,利用 Arduino IDE 写入 Arduino 程序,包含 setup 和 loop 两个主函数。在 setup 函数中对系统进行初始化操作后,loop 函数会循环检测串口,接收并解析串口指令,控制舵机等执行对应操作驱动液压缸,实现动臂落下等动作,表 4 给出了部分舵机操作串口指令。

(3) 通信设计

在挖掘机从控端与计算机主控端,采用蓝牙串口的形式实现双端通信。在计算机端,通过蓝牙模拟串口将手势识别得到的控制指令发送到单片机的蓝牙模块,继而控制挖掘机各部件。

(4) 主从操控实验

实验时首先需要调试软硬件系统,随后在沙盘中进行了单动作控制与挖掘沙土的模拟,实验过程部分图像如图 15 所示。根据运动学分析,通过获取

表 4 Arduino 舵机串口指令

指令类型	指令示例	指令解释
常规指令	\$ RST!	复位指令,停止所有动作,重新启动单片机
舵机操作指令	\$ DST!	所有舵机停止在当前位置
	\$ DST;003!	003 号舵机停止在当前位置
	#IndexPwmTime!	单个舵机指令,Index:3 位,000-254; pwm:4 位,0500-2500; time:4 位,0000-9999
	{#000P1500T1000! #001P0900T1000!}	多个舵机指令,用 {} 封装多个单舵机指令
动作序列指令	\$!	清除预存动作序列
	<G0000#000P1500T1000! #001P0900T1000! >	存储动作序列
	\$ DGT;0-10,1!	调用动作 G0000~G0010 组 1 次,若为 0 次则代表循环执行

三个液压缸的行程能解算出铲斗相对于挖掘机车体的位置。因此,可以存储一个动作序列,让液压缸以一定速率伸缩一定的时间,使得铲斗依次达到不同的位置,完成的沙土挖卸动作。将某一手势映射到该动作序列指令,即可实现单手势对挖卸沙土的实现。



图 15 实验过程部分图像

结果显示控制挖掘机的软件与硬件系统都能够正常运行,对于单动作控制手势识别准确、控制响应迅速,也可以采用此方法实现一个完整的沙土挖掘动作序列。

5 结论

本文的主要研究课题是基于 MediaPipe 的手势识别用于反铲挖掘机的遥控操作系统,最终实现的功能是通过识别的手势控制挖掘机,能够通过手势操作使挖掘机实现一个完整的挖掘沙土的动作序列。本文设计了 15 种不同的手势,通过两个实验,使该系统的可行性和鲁棒性得到了验证,能够较准确地识别手势并远程控制挖掘机,为其遥控操作提供了一种新型的人机交互思路。

本研究的局限在于没有在手势识别上进行更深层次的研究,手势 P-ON 和 FORWARD 的识别综合评价指标分别为 90.69% 和 89.75%,并不理想。本研究的算法中,主要是通过判别手指的开合来分类识别手势的,在以手背面向摄像头时,就难以实现个别手势的识别。然而,通过这些手势也不能非

常流畅地控制挖掘机,无法在挖掘机实际所处的复杂环境下高效施工。

未来研究的目的在于更流畅便捷地通过手势实现对挖掘机的遥操作,并实现部分自动化。本文中使用的姿势识别算法、数据集、通信协议都需要做出相应的改进。虽然 MediaPipe 可以准确地识别出手部关键点,减小了对数据集的需求,但仍应覆盖较多的人和较复杂的手部识别背景。最好能够实现双手协作和动态的手势识别,将行走装置与工作装置的控制剥离,并对一定时序内的动态手势进行识别。

参考文献

- [1] 梁万福,黄鸣辉,刘浩,等.某型液压挖掘机驾驶室内低频噪声控制方法研究[J].工程机械,2018,49(08):11-15+87.
- [2] 张琳.基于用户模式的挖掘机人机系统研究[D].太原理工大学,2018.
- [3] 甘平,杨乐巍,房增强,等.挥发性有机物污染场地挖掘过程中污染扩散特征[J].环境科学,2013,34(12):4619-4626.
- [4] 李运华,范茹军,杨丽曼,等.智能化挖掘机的研究现状与发展趋势[J].机械工程学报,2020,56(13):165-178.
- [5] 张凤军,戴国忠,彭晓兰.虚拟现实的人机交互综述[J].中国科学:信息科学,2016,46(12):1711-1736.
- [6] 惠丹.基于深度学习算法的可穿戴设备手势识别系统设计[J].微型机与应用,2019,038(009):30-33.
- [7] 刘续丹,侯红霞,杨钰淇,等.基于数据手套的交互系统设计[J].通讯世界,2019,26(04):319-320.
- [8] 易靖国,程江华,库锡树.视觉手势识别综述[J].计算机科学,2016,43(S1):103-108.
- [9] 沙洁,麻建,牟海军,等.基于视觉的动态手势识别综述[J].计算机科学与应用,2020,010(005):P.990-1001.
- [10] Denavit J, Hartenberg R S. A kinematic notation for lower-pair mechanisms based on matrices[J]. 1955.
- [11] Zhang F, Bazarevsky V, Vakunov A, et al. MediaPipe Hands: On-device Real-time Hand Tracking[J]. 2020.
- [12] 张圆圆.基于视觉的手势识别技术及其应用研究[J].计算技术与自动化,2015,34(01):131-135.

Gesture Recognition Based on MediaPipe for Excavator Teleoperation Control

Wang Rubin¹, Dou Quanli^{1,2}, Zhang Qi¹, Zhou Cheng¹

(1. *School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;*
2. *Weichai Power Co., Ltd., Weifang 261061, China*)

Abstract: Excavators have broad application scenarios, but in certain hazardous working conditions, using teleoperation is a better option. Meanwhile, with the development of machine vision and deep learning in recent years, a number of algorithms and frameworks for gesture recognition are becoming available. In order to explore the application of MediaPipe-based gesture recognition algorithms in excavator teleoperation, this paper corresponds different hand gestures to different movements of the excavator, and proposes a new control method using hand gestures to achieve excavator teleoperation. The kinematic analysis of the experimental excavation robot was carried out in a laboratory environment, and MediaPipe was used to perform real-time static recognition of 15 gesture types, generate commands and realise the control of the excavator using an Arduino microcontroller. The results show that the system has good performance and can be used for teleoperation control of the excavator, which provides a new form of human-computer interaction for remote control of the excavators.

Key Words: Gesture Recognition; MediaPipe; Excavating Teleoperation; Microcontroller; Teleoperation