

无人机体感交互控制系统的设计与实现

天津大学 程水齐，丁闻哲

- (1. 天津大学电气自动化与信息工程学院，天津市南开区 300072；
2. 天津大学电气自动化与信息工程学院，天津市南开区 300072)¹

李宏跃 工程师

中文摘要：在无人机应用日渐成熟，控制技术趋于多样化的当下，无人机体感交互控制提供了一种便捷灵活的操控方式。传统的基于可穿戴设备的控制方式操作复杂，交互体验较差，针对这一问题，本文提出一种基于人体姿态识别的无人机控制方式。首先根据设计的十二种动作指令收集不同角度拍摄的模板照片；之后设计模型，利用 OpenPose 人体姿态模型提取人体关键点坐标输入到 KNN 模型中与模板进行匹配；设计用户交互界面实现多种控制功能并优化无人机控制；最后将整个模型部署在 Jetson Nano 上。实验结果表明，该控制方式可以达到 91%左右的准确率且在一定程度上解决了推理模型带来的控制延时问题。

英文摘要：With the development of UAVs and the diversification of control technologies, the somatosensory interactive control of UAVs provides a convenient and flexible control method. The traditional control method based on wearable devices has complicated operation and poor interactive experience. To solve this problem, this paper proposes a drone control method which bases on human posture recognition. First, template photos are taken from different angles according to the designed twelve kinds of action instructions; then, this paper designs a model, which uses the OpenPose human body detection model to extract the coordinates of key points of the human body and input them into the KNN model to match the template; what's more, the user interface to achieve multiple control functions and optimize the drone control is added; finally we deploy the model on Jetson Nano. Experimental results show that the control method can reach an accuracy of about 91% and overcome the operation delay problem caused by the inference model to a certain extent.

关键词：人体姿态识别；体感交互；无人机控制；边缘计算

¹ 国家级大学生创新创业训练计划支持项目（202010056110）

作者简介：程水齐（2000-），男，山西太原人，天津大学电气自动化与信息工程学院电气工程及其自动化专业 2018 级本科生；丁闻哲（1999-），女，山西大同人，天津大学电气自动化与信息工程学院通信工程专业 2018 级本科生。

一、引言

近年来，无人机技术不断发展，随着一些小型化、低成本无人机的出现，其应用从最初的军事领域逐渐扩展到遥感航拍、抗震救灾、故障检测以及人们的生活娱乐中^[1]。为了迎合市场需求，降低操作门槛，提高用户体验^[2]，无人机的控制方式也随之不断革新，从最初需要硬件和技术水平要求较高的计算机程序控制以及遥控器控制发展到现在流行的可穿戴设备、体感交互等更加智能简便的控制方式。然而可穿戴设备存在使用不便，时延较长等问题。刘春等通过采用 Leap Motion 作为传感器实现对飞机结构的手势交互控制^[3]；2017 年塔尔萨大学和坦迪大学通过识别 MPII 人体姿势数据库定义的 16 个关键点的位置，实现了基于手势的自主无人机控制^[4]；同年大疆推出了“晓”Spark 无人机，可以实现手势控制飞行功能。可以看出，在传统基于“人的回路”的控制方式以外，各类基于视觉、手势、复合体感的人机交互方式是未来无人机指挥控制的发展趋势^[5]。

本项目基于 OpenPose^[6] 人体姿态识别模型使用自底向上的方法提取出人体 25 点结构；之后将提取出的人体姿态划分为不同动作指令后输入到 KNN 模型进行分类；最后与自动控制技术结合，在无人机平台上实现体感交互姿态控制飞行姿态的目的。模型实现在嵌入式平台（NVIDIA Jetson Nano）的部署，并且准确率在 90%以上；此外还设计了用户图形界面（GUI），形成用户可用的软件，允许用户对无人机进行多模态的控制。

二、人体姿态识别和分类技术

人体姿态识别和分类技术是通过对人体肩部，肘部，头部，膝盖等关键点的定位，把当前姿态进行分类，识别为某一类已知姿态的过程。在本文中，我们用 OpenPose 作为人体姿态识别的关键点提取模型，KNN 作为姿态分类器。

（一）人体姿态识别技术简介

人体姿态识别技术根据输出关键点位置维度分为 2D 人体姿态估计和 3D 人体姿态估计。2D 人体姿态估计可以采用普通摄像头输入，但存在关键点的遮挡，人体姿态通过图像显示存在变形等问题。而 3D 人体姿态估计对摄像头布置要求较高，投入费用较大。二维空间下的姿态识别主要有自底向上和自顶向下两种方法^[7]。

自顶向下是先获取图像中所有人体的位置，再对每个人体进行截取，按照单人识别的方法定位关键点，判断其姿态。这种方法在设计思路上比较直观自然，在可靠的人体检测框之内进行关节点检测可以得到非常高的精度，然而框的数量会随着图像中行人的增多而显著增加，大量重叠的区域会导致重复多次的卷积操作，增加了计算开销，并且会在一定程度上受到个体较大差异性等问题影响。

自底向上是近年来提出的一种新方法，与自顶向下不同，该方法先提取出图像的所有关键点，然后通过加入亲和力场等方式将关键点组合成人体框架。自底向上模型只需要提取一次图像特征，效率较高，且在不同环境下抗干扰能力较强，本文中将使用这类方式作为人体姿态识别模型^[8]。

（二）OpenPose 人体姿态识别原理和应用

OpenPose 是一种自底向上的人体 2D 姿势估计方法，加入了亲和力场（PAFs）用于关联节点，相较于使用传统机器学习与光学处理结合的方式，此方法提供关键点检测，受着装、自遮挡、复杂背景和相机视角

等影响较小^[9]，在本实验中用于提取人体的关键点实现指令判断以及无人机跟随功能。模型整体运行过程如图 1 所示：

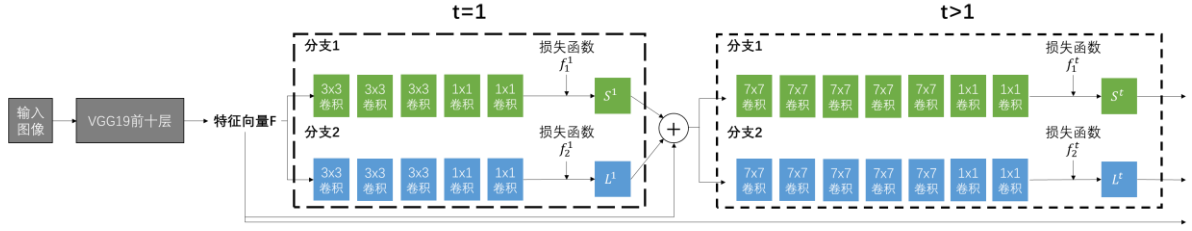


图 1 OpenPose 模型提取关键点流程

将 RGB 图像输入 OpenPose 模型先经过卷积神经网络 VGG19 的前十层用于分析图像并生成特征映射 F 输入第一阶段；第一个阶段内部采用分支结构，其中第一分支利用卷积网络产生检测置信度 S^1 ，第二分支采用亲和力场 L^1 ；之后继续向后进行多个阶段用于加深网络，以前面两个分支输出 S^{t-1} 和 L^{t-1} 以及特征映射 F 作为输入产生 S^t 和 L^t ，如下列公式所示，式中 ρ 、 ϕ 分别表示第一、第二分支的卷积神经网络：

$$S^1 = \rho(F) \quad (1)$$

$$L^1 = \phi(F) \quad (2)$$

$$S^t = \rho^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \quad (3)$$

$$L^t = \phi^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \quad (4)$$

第 t 次迭代后的第一、二分支损失函数分别为：

$$f_s^t = \sum_{j=1}^J \sum_p W(p) \|S_j^t(p) - S_j^*(p)\|_2^2 \quad (5)$$

$$f_L^t = \sum_{c=1}^c \sum_p W(p) \|L_c^t(p) - L_c^*(p)\|_2^2 \quad (6)$$

其中当关节 p 没有在图像显示时 $W(p)$ 取 0，否则取 1； $S_j^t(p)$ 与 $S_j^*(p)$ 分别表示第一分支中第 j 次迭代置信度预测值和准确值， $L_c^t(p)$ 和 $L_c^*(p)$ 分别表示第二分支第 j 次迭代关节连接向量的预测值和准确值。模型总损失函数为：

$$f = \sum_{t=1}^{T_p} f_L^t + \sum_{t=T_p+1}^{T_p+T_c} f_s^t \quad (7)$$

在实际训练中，我们将 $t > 1$ 情况下的 7×7 卷积替换为 3 个 3×3 卷积，在相同感受视野下减少网络参数。利用训练好的模型，我们可以得到人体的 25 个关键节点坐标。

(三) KNN 姿态识别原理和方法

KNN 算法是一种用于分类和回归的非参数统计方法，利用计算待归类样本与已有标签样本的距离来将其分类。其中， k 是一个用户定义常数，样本最终被归为最接近该点的 k 个样本点中最频繁使用的一类^[10]。

我们可以利用这一方法进行姿态识别。首先利用 OpenPose 提取收集 12 种姿态的关键点坐标作为模板，包括起飞、降落等 11 种无人机动作控制姿态和 1 种“无动作”姿态，如图 2 所示。

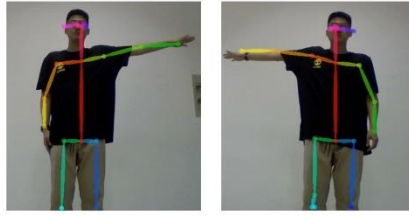


图 2 起飞与降落指令的镜像姿态

每种姿态收集相等数量的图片，且在 25 点坐标中仅需要保留 0-8 号节点所表示的上肢姿态信息。 $\vec{v}_0$ 表示方向向量， $\vec{a}$ 、 $\vec{b}$ 为空间任意两向量，由公式：

$$\vec{v}_0 = \frac{\vec{a} - \vec{b}}{|\vec{a} - \vec{b}|} \quad (8)$$

可得右肩、左肩、身体、脖子、右上臂、右下臂、左上臂、左下臂的单位方向向量^[11]。用相同方式求出无标签姿态的方向向量，计算其与有标签向量的欧氏距离 L_2 为：

$$L_2 = \|\vec{v}_{IN} - \vec{v}_m\|_2 \quad (9)$$

式中， $\vec{v}_{IN}$ 表示待判断姿态， $\vec{v}_m$ 为模板姿态。统计距离最近的 k 种情况中出现次数最多的标签 1，若标签 1 出现的频率大于设定阈值，则将该动作归为指令 1。实验中，我们取 $k=10$ 效果最优。

三、人体感交互控制系统设计

(一) 系统整体设计

总体来讲，系统通过对人体姿态的提取识别判断动作指令，对无人机发出动作信号进行控制。

如图 3 所示为整体设计结构图，其中硬件设备有摄像头、显示器、嵌入式平台以及无人机，其中 Jetson Nano^[12] 与无人机两部分摄像头均可用来将数据传入嵌入式平台；软件设备分为交互界面、操作系统、推理模型及加速器、识别模型、控制系统五大部分。软件系统中，操作系统我们使用的使 Jetson Nano 上安装的乌班图 18.04 系统；推理模型利用 OpenPose 来提取人体关键点信息；识别系统采用 KNN 模型判断人体姿态所对应的指令；控制系统用来根据不同工作模式以及动作指令对无人机采取不同控制策略。

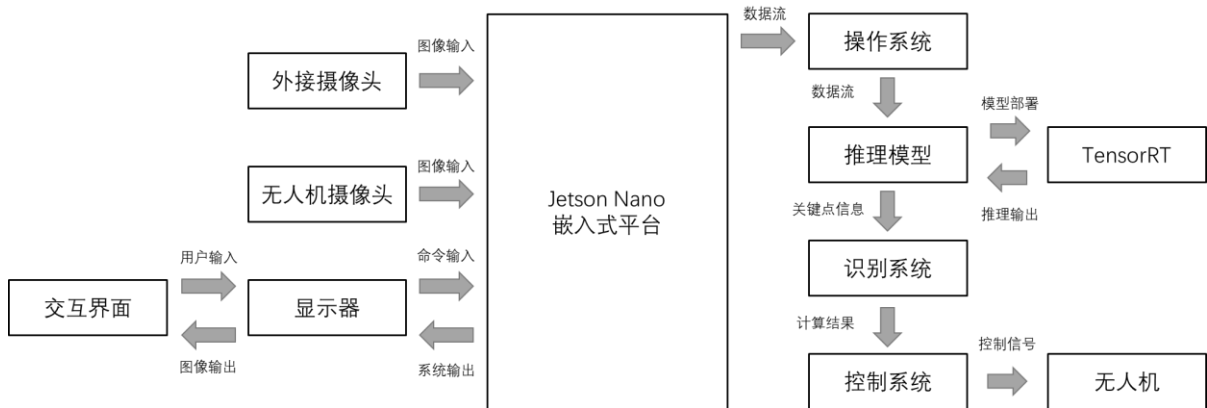


图 3 系统整体设计示意图

(二) 无人机交互指令设计

本文中共设计 12 种姿态识别交互指令，分别控制无人机的起飞降落、空翻、顺时针逆时针旋转、前后左右上下动作以及无指令状态。此外，在图形交互界面设计无人机链接指令使无人机与嵌入式平台通过路由器连接；通过设计开启渲染以及关闭背景可以达到仅显示操作者人体框架的效果，使姿态识别过程更为直观；摄像头切换指令允许使用者在固定外接摄像头和无人机自带摄像头之间切换，同时控制无人机的不同操作模式，在软件系统设计介绍中详细展开。

(三) 软件系统设计

1. 姿态识别系统

OpenPose 是一种自底向上提取人体关键点的模型，在 2D 摄像头下可以在较复杂的多人背景下有良好的表现。KNN 是一种传统的聚类方法，寻找最接近的模式可以完成归类任务。

利用 OpenPose 可以区分多人模式与单人模式，其中单人模式下可以直接提取人体关键节点，随后计算肢体向量与动作模板比较，利用 KNN 判断指令。在多人模式下需要首先判断操作者，可以通过指令判断、以及人脸识别等方式实现。本文中使用最近距离锁定与模板匹配相结合的方式锁定目标。具体方法为在无人机开始时令操作者站在所有画面中人物最前面，通过判断脸部面积最大者实现目标确认，并提取其面部图像；之后调用 face recognition 库函数检测操作者面部区域，落在该区域内的鼻子关键点所对应的人体框架即为待判断的人体框架。找到操作者后将其关键点信息提取并分类。

2. 无人机控制系统

操作模式：在 Jetson Nano 摄像头模式下，硬件系统读取外接摄像头读取图像并判断操作者姿态，对无人机进行单步控制，即一种指令动作控制无人机动作固定距离或角度。操作者可以在控制界面切换到使用无人机自带摄像头，进入无人机摄像头模式，视频图像传入硬件系统进行解码后作为模型输入。操作者可能在无人机动作过程中处于摄像头感知范围以外失去对无人机的控制，因此在此模式下我们加入跟踪功能^[13]。具体实现为控制无人机上下飞行以及轴向旋转，使某一关键点 $K(x, y)$ （如鼻子）处于画面参照点 $O'(x_0, y_0)$ 上。为了使系统稳定运行，我们引入闭环 PID 控制方式，以 y 方向为例， err 表示当前误差量， err_0 表示上一次误差量， K_p 、 K_I 、 K_D 分别表示 PID 中比例、积分、微分系数

$$err = y_0 - y \quad (10)$$

$$I = I_0 + err \times dt \quad (11)$$

$$D = \frac{(err - err_0)}{dt} \quad (12)$$

$$OUTPUT = K_p err + K_I I + K_D D \quad (13)$$

经过实验，我们取 $y_0 = 0.3h$ ， h 为画面高度， K_p 、 K_I 、 K_D 分别为 0.5，0.1，0.05，输出值控制无人机动作以达到目标高度 y_0 。

指令识别优化：在指令识别设计中，一方面要关注指令本身的识别效果，另一方面能否正确完成我们的预期操作也是要考虑到情况。如图 4 所示，为控制无人机下降的镜像动作，但是在动作过程中会短暂经过无人机降落指令可能被触发的区间。

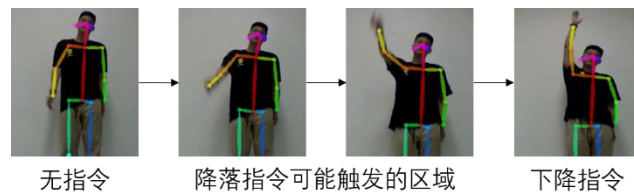


图 4 引发误动作示意

为解决该问题，我们引入指令识别空窗期的方法，在第一次识别出某指令后延时 0.2s 再次判断，若一致则执行该动作指令，否则无人机不会动作。

3. 嵌入式计算机部署与优化

硬件配置方面，在 Jetson Nano 上外接摄像头获取人体姿态视频信息；加入显示器允许用户在图形界面操作；利用 WIFI 模块与无人机通信，获取图像信息以及发送控制指令。软件设置上，在 Jetson Nano 安装的乌班图操作系统下配置环境将 OpenPose 的 TensorFlow 模型部署到设备中。

四、实验方案

为了测试系统性能，我们对操作指令识别的准确性和帧率进行了测试。本实验中用到的硬件设备型号为 Jetson Nano B01，软件系统中使用 python3.7 作为编程语言，在乌班图 18.04 系统下，调整输入图像分辨率为 480x640 进行测试。

(一) 测试方法

首先，对指令识别准确性的评估需要在真实环境下进行，以将随机的干扰因素考虑在内，我们选择有人员来往的室外场地作为测试背景；此外还要考虑人体动作过程中每种动作间的连贯动作可能带来的误操作问题，因此我们随机交替进行 12 个指令动作，每种动作累计进行 100 次进行统计，列出动作指令混淆矩阵如表 1 中所示。

表 1 动作指令测试混淆矩阵

真实	预测											
	起飞	降落	顺转	逆转	向上	向下	向左	向右	向前	向后	空翻	保持
起飞	94	0	0	0	0	0	0	3	2	0	0	1
降落	0	97	0	0	0	0	2	0	0	1	0	0
顺转	0	5	86	0	0	0	9	0	0	0	0	0
逆转	0	0	0	89	11	0	0	0	0	0	0	0
向上	4	0	5	0	91	0	0	0	0	0	0	0
向下	0	13	0	0	0	87	0	0	0	0	0	0
向左	0	7	0	0	0	3	90	0	0	0	0	0
向右	3	0	0	0	0	0	9	88	0	0	0	0
向前	8	0	0	0	0	0	0	0	87	5	0	0
向后	6	0	0	0	0	0	0	0	4	90	0	0
空翻	0	0	0	0	0	0	0	0	0	3	97	0
保持	0	0	0	0	0	0	0	0	0	0	0	100
总计	115	122	91	89	102	90	110	91	93	99	97	101

(二) 测试效果

由混淆矩阵可计算得到平均准确率为 91.33%，可见该方法具有一定的可靠性和实用价值，可以进一步推广运用。图形交互界面最终效果图如图 5、图 6 所示。



图 5 硬件设备实现示意



图 6 图形交互界面示意

五、结束语

相较于传统的无人机控制方式，本文提出一种结合姿态识别深度学习模型的无人机控制方法，一定程度上降低系统对硬件设备的要求。通过软件系统的设计和部署，本系统在实际背景下有较好的识别效果和反应速度，实现在嵌入式平台的计算推理，对于推理模型在边缘设备中的部署应用提出一种可行的方式。此外，加入图形交互界面的设计方便操作，使该方式可进一步推广到实际应用中。

参考文献：

- [1]黄泽满, 刘勇, 周星, 陈天伟, 梁毅东. 民用无人机应用发展概述[J]. 赤峰学院学报(自然科学版), 2014(24).
- [2]张凤芝.基于体感控制器的无人机飞行控制[J].信息系统工程,2019(08):127.
- [3]刘春, 李秋雨. Leap Motion 体感控制器及其在飞机结构展示系统中的应用. 计算机应用与软件. 2016 Apr;33(4):227-9.
- [4]Jonathon Bolin, Chad Crawford, William Macke, et al. Gesture-Based Control of Autonomous UAVs[J]. Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, 2017, 8-12:1484-1486.
- [5]李大超,施展. 面向无人机指挥控制的下一代人机交互技术研究[A]. 航空工业测控技术发展中心、中国航空学会测试技术分会、状态监测特种传感技术航空科技重点实验室.第十六届中国航空测控技术年会论文集[C].航空工业测控技术发展中心、中国航空学会测试技术分会、状态监测特种传感技术航空科技重点实验室:《测控技术》杂志社,2019:5.

- [6]Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2929 257.
- [7]熊紫华.基于深度学习的人体姿态估计研究[D].武汉: 华中科技大学, 2019.
- [8]周燕,刘紫琴,曾凡智,周月霞,陈嘉文,罗粤.深度学习的二维人体姿态估计综述[J].计算机科学与探索,2021,15(04):641-657.
- [9]苏超,王国中.基于改进 OpenPose 的学生行为识别研究[J/OL].计算机应用研究:1-8[2021-06-10].
- [10]Min-Ling Zhang, Zhi-Hua Zhou,ML-KNN: A lazy learning approach to multi-label learning,Pattern Recognition,Volume 40, Issue 7,2007,Pages 2038-2048,ISSN 0031-3203.
- [11]魏英姿,史伏雨.一种人体动作分类的局部描述方法[J].电子世界,2020(21):102-104.
- [12]农元君,王俊杰,徐晓东,赵雪冰.面向嵌入式平台的安全帽实时检测方法[J/OL].计算机工程与应用:1-8[2021-06-10].
- [13]唐凯,张海涛,郑建杰,孔维迪.基于立体视觉的无人机动态目标实时跟踪[J].信息技术与网络安全,2020,39(12):30-36.