文章编号:1671-4598(2022)06-0113-07 DOI:10.16526/j.cnki.11-4762/tp.2022.06.017 中图分类号:TP39 文献标识码:A

基于音视信息融合的桌面机械臂技能获取及控制系统

孙 昊,马兴录,丰 艳,李晓旭
(青岛科技大学 信息科学技术学院,山东 青岛 266061)

摘要:针对化学实验机械臂控制编程门槛高且技能获取准确率低等发展制约因素,设计一种基于音视信息融合算法的桌面型实验机械臂的控制系统,实验员通过边做边说的示教方式教给机械臂运动技能,进而代替实验员完成一些繁琐、具有危险性的实验工作;系统分为技能获取以及运动控制两部分;其技能获取部分使用改进的双流卷积网络实现动作检测;使用语音 AI 和正则表达式实现语音提取;再通过音视动作信息融合算法将动作检测和语音部分的识别信息相融合得出高重合度的运动技能,技能获取准确度可达 81%以上;运动控制部分使用电机控制和抓取位姿识别,可实现更精细的控制和抓取;系统可用于具有特定流程化学实验的示教控制工作,在代替实验员来完成化学实验工作的同时大大降低了编程门槛,提高了效率。

关键词:实验机械臂;技能获取;运动控制;动作检测;语音识别;信息融合;位姿识别

Research on Acquiring Teaching Skills of Desktop Manipulator Based on Audio-visual Information Fusion

SUN Hao, MA Xinglu, FENG Yan, LI Xiaoxu
(Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Aiming at the developmental constraints such as the high threshold of control programming for chemical experiment manipulators and the low accuracy of skill acquisition, a control system for desktop experimental manipulators based on audio-visual information fusion algorithm is designed. Experimenter teachs the mechanical arm movement skills, and then the control system replaces the experimenter to complete some tedious and dangerous experimental work. The system is divided into two parts: skill acquisition and movement control. Its skill acquisition part uses an improved dual-stream convolutional network to achieve motion detection; uses voice AI and regular expressions to achieve voice extraction; and then uses audio-visual motion information fusion algorithms to integrate motion detection and voice recognition information to obtain a high degree of coincidence. The accuracy of the skill acquisition can reach more than 81%. The motion control part uses motor control and grasping pose recognition, which can achieve more precise control and grasping. The system can be used for the teaching and control work of chemical experiments with specific processes. It can replace the experimenter to complete the chemical experiment work while greatly reducing the programming threshold and improving the efficiency.

Keywords: experimental robotic arm; skill acquisition; motion control; motion detection; speech recognition; information fusion; pose recognition

0 引言

当前,人工智能科技发展如火如荼,智能控制作为当今关键核心技术,其有效促进了新型控制体系的高速发展[1-3]。化工实验是在化工产业进行研究、学习和生产过程中的一个不可或缺的环节[4]。目前来看,化学实验的智能化水平不高,利用机器人机械臂等智能化的机械设备去进行化学实验所能够完成的实验过程较为单一而且编程较为复杂。并且化工实验中会使用各种各样的化学试剂,这些试剂在实验过程中相互作用会产生各种有害物质,甚至当实验出现失误时还可能出现不可预知的危险。所以,利用更加智能化的机械臂等设备代替实验员去完成相关的化学实验是非常有必要的。

机器学习和机器视觉技术的发展,使得机械臂演绎编程为人机交互提供了新的解决办法,是降低机械臂技能获取难度的重要途径[5-7]。机械臂示教编程是通过对人的示教

收稿日期:2021-12-14; **修回日期**:2022-01-07。
基金项目:国家重点研发计划子课题(2017YFB1400903)
作者简介:孙 昊(1998-),男,山东临沂人,硕士研究生,主要从事嵌入式系统与机器人方向的研究。
丰 艳(1977-),女,山东青岛人,博士,副教授,主要从事图像处理、虚拟现实方向的研究。
通讯作者:马兴录(1970-),男,山东青岛人,硕士,副教授,主要从事嵌入式系统与智能硬件方向的研究。
引用格式:孙 昊,马兴录,丰 艳,等.基于音视信息融合的桌面机械臂技能获取及控制系统[J].计算机测量与控制,2022,30(6):113-119.

动作进行观看学习，从而自动习得运动轨迹的过程，具有难度低、人机交互便捷、操作灵活等优点[8]。本文将手臂动作检测与语音识别结合，同时借助信息融合和位姿识别，设计了一个机械臂的示教控制系统[9]，让机械臂能够看懂、听懂实验工作人员教授的技能，并能够更准确、快捷地完成人机交互，进而代替实验员完成实验工作。

1 系统整体设计

本设计中的示教控制系统是以桌面型实验机械臂为物理载体，以树莓派操作系统为平台。主要针对在化学分析实验场景下，通过实验员的示教去教会实验机械臂学习模仿人手臂的实验动作，再结合一定的实验仪器的抓取位姿识别算法来辅助控制，示教与运动控制两者相结合来完成一套实验流程组合。

1.1 系统逻辑架构

如图 1 所示，本设计的整体逻辑架构可以分为 3 个层面：自下而上来看分别是硬件组成、设备驱动以及应用软件层：

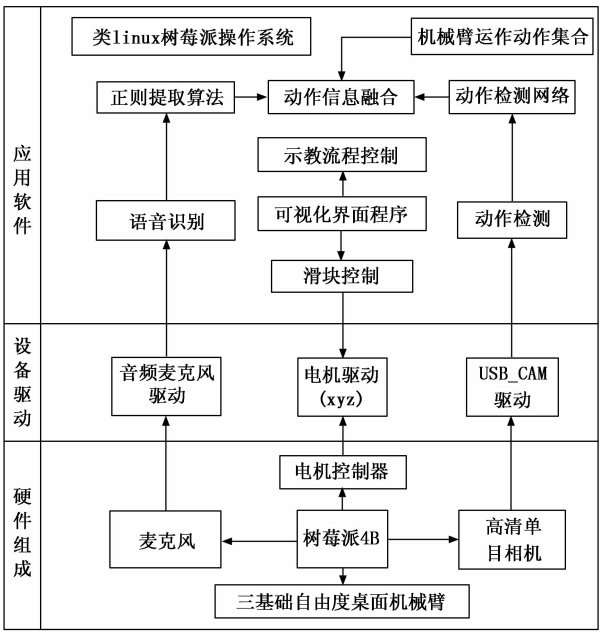

图 1 系统逻辑架构

其中硬件组成为桌面型机械臂（拥有 3 个基础自由度，类似人类手臂）、树莓派 4B 核心板、电机及电机控制板、高清单目相机、麦克风以及机械爪。

设备驱动部分包括对麦克风的音频驱动、USB _ Cam 相机驱动、电机驱动以及其他相关的可编程逻辑等。

应用软件层为在树莓派桌面型操作系统中运行的软件程序和构建的机械臂动作参数集合，该操作系统与 Linux 操作系统类似。其中软件程序部分包括：一个可视化的界面程序用于电机控制和示教流程把握、语音识别[10]功能模块、动作检测模块以及动作信息的融合匹配部分。

1.2 系统功能流程

示教控制系统的整体工作业务流程如图 2 所示。在这一整套实验流程的示教过程中，一套完整实验流程是由许多个简单实验动作组成，叫做动作基元。实验员需要在一边做动作的过程中一边口述其动作，机械臂通过听和看获取信息并进行融合验证，进而能够理解需要去完成的动作。即语音识别模块和动作识别模块分别将对该一整套实验流程动作识别出来的动作基元按顺序组合放入集合中，然后与本地动作库中存储的机械臂动作集中的动作组进行匹配，得出动作信息。过程中的识别信息通过信息融合算法得出最终确定的一套动作组，保存或交给机械臂去运行复现。

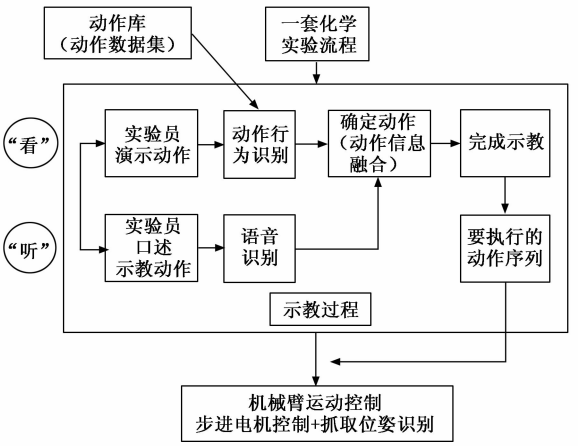

图 2 系统功能流程

系统可视化界面如图 3 所示，可兼容 Windows 和 Linux 操作系统。如图中右侧所示，其中动作参数使用 xml 文件存储，其格式规整简单，便于读取和存储动作参数。可视化界面中设置 3 个 Scale 和 Radiobutton 分别控制 3 个电机 $\{x、y、z\}$ 的步数和方向，通过滑动、添加动作以此来定义动作组并存储。点击开始现场示教开启摄像头，则开始执行示教功能。

图 3 系统界面

2 机械臂组成及性能参数

如图 4 所示，本设计中的桌面型机械臂拥有 3 个基础自由度，可在桌面上方的空间内转动。机械臂以树莓派 4B 为核心控制板，电机驱动扩展板、步进电机驱动模块、3 个步进电机、一个 12 V 电源适配器以及带麦克风的摄像头等组成。抓取与放置操作，使用机械臂通用夹具、夹爪等。

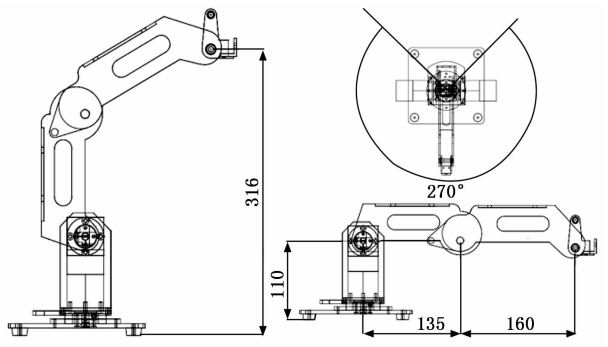

图 4　机械臂结构图

2.1　核心控制板

如图 5 所示，树莓派 4B 实质上是一台微小的嵌入式 PC，类似于身份证大小，其系统基于 Linux 操作系统（本文烧录桌面型的 RasBian System），拥有 64 位、1.5 GHz 的四核 CPU，内存可选择使用 1 GB，2 GB 或 4 GB 的 RAM，带有全吞吐量千兆以太网，还可实现双频 802.11ac 的无线网络，蓝牙为 5.0 版本，且拥有两个 USB 3.0 和两个 USB 2.0 通信端口[11]。

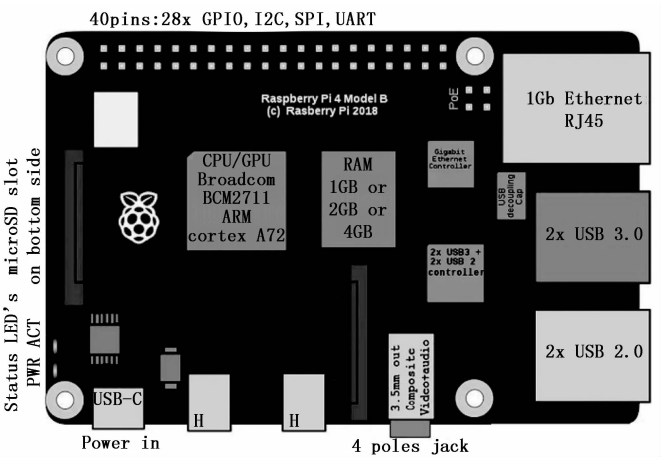

图 5　树莓派 4B

2.2　步进电机及驱动模块

控制机械臂进行运动的关键部位就是电机，本机械臂使用的步进电机有 3 个，实现了 3 个基础自由度的运动控制。机械臂采用大扭矩 42 行星减速式的步进电机，其水平运动的电机其理论减速比为 1：5.18，实际测量为 11：57 接近理论值；两个手臂部分的步进伸缩电机的减速比均为 1：19（实际测量得到 187：3591）。电机的电流为 1.7 A，步距角度为 1.8°，其步距精度为 5%。

表 1　MS 接口对应参数

MS1	MS2	MS3	Microstep Resolution
Low	Low	Low	Full step
High	Low	Low	Half step
Low	High	Low	Quarter step
High	High	Low	Eighth step
High	High	High	Sixteenth step

如图 6 所示，电机驱动器采用 A4988 驱动器，可驱动电机电压 8～35 V。本机械臂采用 12 V 电源给电机供电。其中，每个步进电机驱动模块输出 2 个控制信号，分别为 STEP 和 DIR，与树莓派引脚相接来分别实现步进脉冲和方向的控制；其中 MS1-3 引脚在本设计中均接高电平，接口电平及步数对应参数如表 1 所示，将电机每一步（1.8°）细分为 16 步，实现了更加精细的控制。

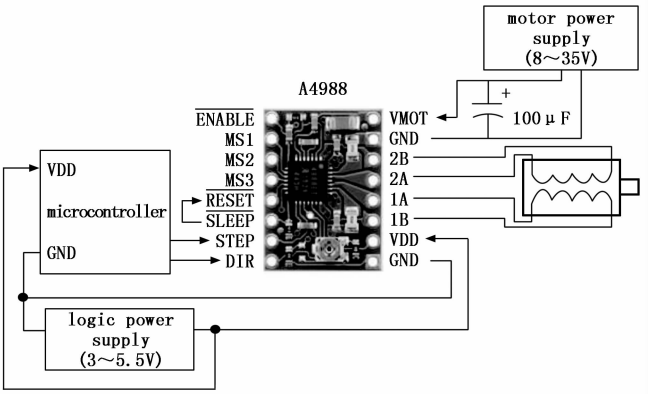

图 6　步进电机驱动模块

3　技能获取模块设计

3.1　动作检测

本文动作检测部分主要基于改进的双流卷积网络，该部分将预处理过的图像信息输入网络，分别使用 Efficient-Netv2[12] 算法计算 RGB 图像和光流图像特征，然后将提取得到的特征信息使用线性分类器 SVM[13-16] 进行行为分类，得到动作的识别信息。

如图 7 所示，双流卷积网络是将输入的视频分为两路来进行处理，其中一个是由卷积神经网络提取 RGB 图像中任务手臂和场景相关信息，另外一部分是处理光流图像信息，最后由 Softmax 函数分别进行归一化融合处理。

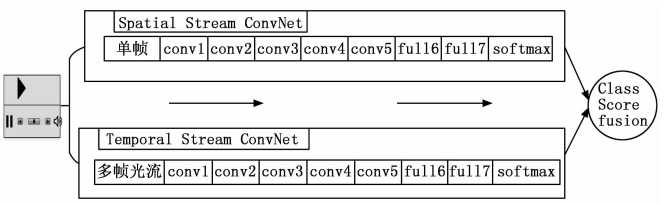

图 7　双流卷积网络结构

网络中光流图像的提取（动作视频预处理）部分是由基于梯度的运算得到，算法关键原理如下：首先设定图像数列 $I(x, y, t)$，向量 $\boldsymbol{X}=[x, y]$，数列是由一段演示视频中的前后帧提取得到，即当视频局部的光流图像基本恒定时，对于任意的 $Y \in N(x)$，有：

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla I(X,t)=\frac{\partial \nabla I}{\partial X}\frac{\partial \boldsymbol{X}}{\partial t}+\frac{\partial \nabla I}{\partial t}=$$

$$\boldsymbol{H}(I)\cdot d+(\nabla I)_t=0 \tag{1}$$

其中：$\boldsymbol{X}$ 为 x 矢量，$\boldsymbol{H}(I)$ 为图像数列 I 的 Hesse 矩

阵，引入 $\boldsymbol{X}$ 与偏移量 d 的关系：

$$E(\boldsymbol{X},d)=\left\|(\boldsymbol{H}(I)\cdot d+(\nabla I)t)\right\|^2 \tag{2}$$

令导数等于 0 可以求得：

$$d=-(\boldsymbol{H}^T(I)\boldsymbol{H}(I))^{-1}(\boldsymbol{H}^T(I)(\nabla I)_t) \tag{3}$$

上述过程，可以总结为依据视频图像中的像素点在时间线上的变化和相邻帧图像之间的关联性来进行分析，找到上一帧和当前一帧的对应关系，计算出运动信息（偏移量即为一种运动信息），进而绘制出光流图像。

在计算特征部分，如图 8 所示，EfficientNetv2 相较于前期的 EfficientNet[17] 算法部分使用 Fused-MBConv 替换 MBConv 结构，即使用常规的 3 * 3 卷积替换 MBConv 中的 3 * 3 深度卷积和 1 * 1 卷积，提高网络的计算速度。

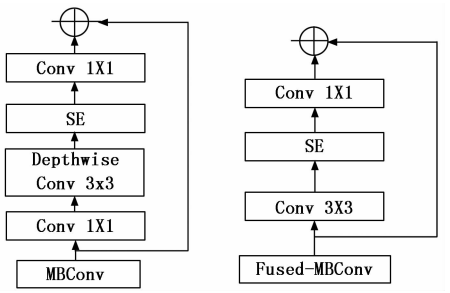

图 8 EfficientNetv2 结构改进示意

得到 RGB 图像和光流图像的特征信息，需要对其进行分类验证。支持向量机 SVM 是一种二分类模型，是用来求解能够正确划分训练数据集并且几何间隔最大的分离超平面。如图 9 所示，$w*x+b=0$ 即分离超平面，这样的超平面一般数量很多，但是几个间隔最大的分离超平面确是仅有唯一的一个。对于其中的最优值，其求解的公式如下：

$$max_{w,b}(min_{x1}\frac{y_i(w^Tx_i+b_i)}{|w|}) \tag{4}$$

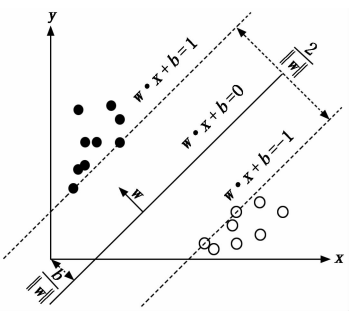

图 9 SVM 原理

整体结构表示如图 10 所示，即将双流卷积网络中的 RGB 和光流特征提取部分使用轻量级的 Efficientnetv2 分别进行卷积和池化处理，再结合 SVM 分类器对两个分支给出的动作信息进行分类，最终给出一个识别到的动作信息。

当获得一套完整动作的识别信息之后，需要去确定该模块机械臂动作的执行序列，获得机械臂的可执行动作。

首先设计机械臂的动作基元并命名，将所有设计动作基元存储到库。上文中检测到的动作信息按顺序存储、匹

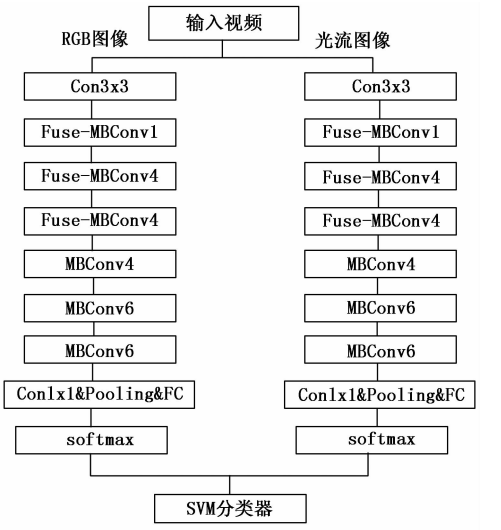

图 10 动作检测网络结构

配目录下的动作基元用来确定当前时间内的动作基元序列。程序在每次的示教过程中都会匹配出由多个动作基元组成的顺序组合，然后在机械臂动作组库中搜索关联动作。识别到的动作编号串联得到一个序列，用这个序列去找到一个完整匹配或者最接近的一个动作组（动作基元及其编号情况如图 11 所示），则认定其为识别到的实验动作，并且给出一个覆盖度（Coverage）。

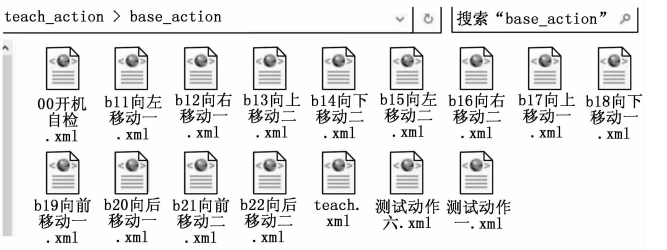

图 11 部分动作基元

当程序找到最匹配的一个动作组之后，继续与库中动作组进行匹配，存储其匹配过程中动作组覆盖度高于 50% 的相关信息。

3.2 语音识别及关键字提取

示教是一个边做边说的过程，实验员在手臂运动的过程中口述当前动作，这就需要语音技术的加成。语音技术的关键是对自然语言进行处理，对语音进行识别及文本生成，从而使机器具备能听会说、能理解会思考的能力[18]。本文使用百度的实时语音识别，其基于 Deep Peak2 端到端建模，将接收到的音频流实时识别成文字字符，进而使用正则表达式去提取动作关键字。该部分伪代码如下：

Function：openvoc(blist)线程用于开启语音模块功能。

Function1(Function)：GetAIvocworld 用于实时获取语音接口返回的识别结果。

Input1：APP_ID，API_KEY，SECERT_KEY，VOC 控制台注册的应用 ID、Key 和输入的音频信息。

Process1：首先与控制台建立 websocket 协议连接。

If：websocket_is_ok()，无错误码。

Do：Process2：边上传音频边获取识别结果。

Output1：result = client.asr(get_voccontent)识别结果还包括 voctxt=result['result'][0]、begin time result['result'][1]和 end time result['result'][2]。

Process3：Function：getRegularword(OutPut1(result))使用正则表达式提取上述输出的文字信息，按照顺序进行存储。

OutPut2：action 关键词列表。

在这个过程中，由于是在化学实验中的语音识别，所以需要将针对化学实验中的许多术语以及当前实验流程所需要的语音数据对训练集文本语料进行补充，语料数据包括对语音文件的格式命名以及文本信息。语料数据集总结了约 55 min 的相关识别内容，其音频文件经过模数转换形成直接的二进制序列 PCM 文件格式存储，实现声音数字化的同时删减了不同于其他文件格式的文件头和结束标志，便于文件的串接。经过对该语料数据集的补充，能有效提升实验业务场景下语音识别准确率 7%～15%。

在上文中语音识别过程中，程序需提前在文本中读取前期人为设定好的动作文件关键字，写入到内存变量中。在每次示教过程中将语音识别到的关键字进行匹配文本，如正确匹配则立即记录其标志编号 IdentNumber 并进行存储。

将一整个示教过程中的所有匹配到的语音关键字编号按顺序汇总，同动作检测与匹配部分相似，查找机械臂的动作组，搜索到一个完整匹配或者最接近的一个动作组，暂认定其为语音识别到动作组，保留覆盖度信息，并记录其他覆盖度高于 50%的相关动作组信息。

3.3　音视信息融合

由于本文机械臂示教程序主要在嵌入式设备中运行，某些场景中移动端、嵌入式设备相比于服务器或 PC 机，其配置、运算能力以及设备性能等会逊色许多[19]。在如此限制条件下，想要实现高速、准确的识别算法、示教方式难度较高。而且，示教是一个动态的过程，需要连续性的进行语音和动作的识别，这个过程难免会出现识别失效等情况影响准确率。所以本文借鉴传感器的信息融合[20-22]，编写具有强针对性的算法将动作检测以及语音识别两个单独模块得到的关键信息进行结合，以此来提高技能获取准确度，同时节约性能。

机械臂示教过程中得到的视频识别信息和语音识别信息为一个动作组以及其覆盖率。在运行效果最理想的情况下，视频和语音给出的动作组长度是相等的，其数据如表 2 为例。表 2 中两个模块列出的是当前示教过程产生的最高覆盖度的信息 ACT _ G0 以及其他大于 50%覆盖度的动作覆盖信息。

如果在这一次的示教过程中，两个模块最大的覆盖度动作组相同，即视频和语音部分都选定了匹配度最高的相同的动作组，那么可以认定该动作为示教动作，不需要经历融合的算法过程，而在这种低概率的情况之外则需要。

表 2　匹配度分布表

	ACT_G0	ACT_G1	ACT_G2	ACT_G3	ACT_G4
Video module	0.8	0.75	0.7	0.6	0.5
Audio module	0.8	0.7	0.7	0.6	0.6

算法思路：定义 Video Module 部分的 ACT _ G0 为 $VG0$，后续依次定义为 VGn，Audio Module 部分同理定义为 $AG0$、$AG1\cdots\cdots AGn$。用 $AG0$ 同 VG 部分的 $1-n$ 进行比较，按照正序找到一个 VGx 同 $AG0$ 的相似度为 100%，即 VG 存在 50% 以上部分同 $AG0$ 匹配，则记录存储 $AG0$ 与 VGx 的覆盖度积（$AG0 * VGx$，$0<x\leqslant n$），AG 部分同理。将上述两部分结果比输出最优匹配值的动作组（权重平分），如公式（5）所示：

$$FG = Mag[(AG0 \times VGx),(VG0 \times AGy)]x,y \in [0,n] \tag{5}$$

其中：Mag 为求最大值动作组的函数，FG 为覆盖率相乘之后取得最大值的机械臂动作组。

当上述过程中与 ACT _ G0 相似度为 100%的动作组为空时，则进入以下搜寻算法：任意两套动作之间均有相似度（<100%），用原始的覆盖度之积同这两者之间的相似度，得出两者的关联值，将所有的关联值汇总输出最大值动作，即这个最大值是由两个不同覆盖度的动作融合得出的，然后在这两个动作组中选择与原始的一套动作基元序列最接近（覆盖度最大）的动作组，即为最终确定的示教动作，上述过程如公式（6）、（7）所示：

$$FQ(x,y) = AGx \times VGy \times Fit(AGx,VGy)x,y \in [0,n] \tag{6}$$

$$FFG = MAX\_G\{MAX[FQ(x,y)]\}x,y \in [0,n] \tag{7}$$

其中：Fit 为两套动作组之间的相似度，FQ 为一对覆盖度之积再乘上两者相似度的动作相关值，式（7）中将两个模块所有的动作信息融合，输出拥有最优相关值的一对动作，然后再选出与原始动作基元序列覆盖度最高的动作组，即 FFG。以此，完成对视频模块以及音频模块相关动作信息的验证融合，机械臂确定最终的运动技能。

4　机械臂运动控制

示教完成之后，程序加载最终确定的 XML 动作参数文件，其内容大致如图 3 右侧所示。程序按照顺序读取其中 X、Y、Z 电机的对应的方向以及步数参数，依次执行，控制树莓派相关引脚的高低电平输出，来控制电机的运转。其中，步数参数 1 024 对应的机械臂转动角度为 18°，最高为 10 240。

机械臂在运动中，安装在机械臂前端夹爪附近的相机可以接收到实时图像数据并通过 USB 端口传输到树莓派，系统将其结合抓取位姿识别[23]算法进行抓取，算法使用了

5 个变量：$\{x, y, \theta, h, w\}$ 来描述机械臂抓取物体时夹爪的抓取位置和夹爪方向。如图 12 中的矩形框所示，其中 (x, y) 被用来表示矩形框的中心位置，用 θ 来表示图像中的水平横轴与当前矩形框倾斜位置的夹角，h 即 Heigh 表示高度，w 即 Width 用来表示宽度。

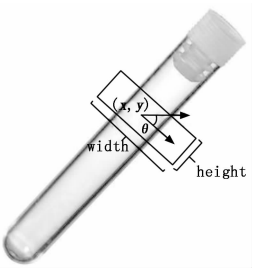

图 12 抓取位姿

当机械臂到达物体附近，需根据矩形框的位置对机械臂的前端位置进行细微的调整。即设置了 6 个基础调整动作：左移和右移、上移和下移、顺/逆时针旋转移动，利用上述动作对机械臂调整以到达所预设的位置实现物体抓取。

该抓取位姿算法基于 Cornell Grasping Dataset 数据集，并在该数据集的基础上继续补充训练了化学器材相关的抓取位置，补偿性的提高了抓取位姿识别的准确度。

5 系统性能测试与分析

5.1 测试实验及结果

在信息融合部分对两个模块是否产生同样数量的动作基元会有要求，因此针对该要求设置实验来记录每个模块的动作基元个数。

测试实验：实验过程为在固定的试验台上，实验员模拟整套实验动作，并且同时口述动作。由于该过程不需要让机械臂运动，只是观察其示教过程，所以实验程序在 Windows10 操作系统中运行，旨在记录其产生的动作组中动作基元的个数，过程共进行了 12 次不同种实验的示教测试，其每次示教过程中得到的语音部分和动作部分检测的一整套动作的动作基元个数如表 3 所示。

表 3 动作基元个数

	task1	task2	task3	task4	task5	task6
Video module	4	5	5	6	8	8
Audio module	4	5	5	6	6	8
	task7	task8	task9	task10	task11	task12
Video module	6	7	5	8	8	3
Audio module	7	7	5	7	9	3

可以看出，在动作基于少的动作组示教过程中，两个模块产生的动作基元个数基本一致，随着动作基元个数增多，会出现少数不一致情况，总体结果效果良好。

在上述一致情况的基础下，其最终动作组的覆盖率如表 4 所示。

表 4 动作组覆盖率 %

task1	task2	task3	task4	task6	task8	task9	task12
100	80	80	83.3	75	71.4	80	100

记录表中所示结果，再次补充了 28 次示教实验：本次在树莓派 Rasbian 桌面操作系统中运行示教程序，分为 4 个实验员完成 4 种不同的化学实验，每人重复 7 次实验过程，记录其过程中的识别覆盖率。根据测试任务中动作基元的个数以及其动作组覆盖率，最终计算出其整体示教准确率约为 81.4%。该结果在化学机械臂不使用支撑设备的无接触技能获取领域取得了较好的成绩。

5.2 测试效果及问题分析

目前系统对于出现不一致情况下还没有明确的解决办法，后续的研究内容还需要针对不一致情况下的相关算法进行完善和修改。

在准确率方面，该机械臂示教系统的稳定性与准确性取决于其动作行为识别以及语音识别部分的准确率。由于是在嵌入式设备中运行，其动作识别匹配部分算法需要保证轻量级，所以在提升识别速度的情况下牺牲了部分准确率。语音识别部分的近场中文普通话识别准确率达 95%，其进一步进行的关键字匹配正确率相比简单识别来说要求更高。两个模块在示教的过程中连续识别，难免会存在丢失识别的情况，这样准确率就会降低，但基于信息融合之后，将两者的优势结合起来，其得出的最终动作覆盖率可基本稳定在 81%以上，这个结果目前来看是令人满意的。

性能速度方面，在 2 GB RAM 的树莓派系统中运行程序，其单个动作基元的击中时间（识别出所耗费的时间）两个模块均在 1 s 以内（动作检测程序段跳出时间约0.59 s，语音识别约 0.82 s）。另外语音部分其后期需要执行多次循环验证时时，可以改为使用机器码编译的语言来专门运行循环部分，创建针对该部分的 dll 动态链接库，利用外部函数库 Ctypes 去调用，可显著提高循环速度。

对于位姿抓取部分，因试管等器材为透明材料，对识别效果有较大影响，尝试改用标签识别夹取（如图 13 所示），或者将标签的识别作为弥补性的措施，以此来提高识别效率。

图 13 标签位置抓取

6　结束语

实验员在化学实验室中需要进行各种各样的化学实验，这些实验繁琐且有些实验对人体有害。所以要利用机械臂代替实验员进行实验。但是目前所用到的实验机械臂需要专业的机械臂控制人员去编程实现，这对于化学实验员来说是有难度的。所以为了解决上述问题，本文基于动作检测和语音识别提出了一个编程简单高效的机械臂示教系统。

实验员在边做实验的过程中边口述其动作，系统利用动作检测检测实验员手臂运动、搭配语音识别、使用信息融合相关的算法进行机械臂示教控制，教给机械臂应该执行的运动技能，机械臂在执行过程中结合位姿识别进行物体的抓取，完成一套实验任务。该系统在运动技能获取部分的准确率能达到 81%以上，在运动控制部分将步进电机的每一步（1.8°）细分成 16 步的同时结合物体抓取位姿识别，实现了更精细的控制和抓取。

参考文献：

[1] 罗　磊，赵　宁．人工智能在物流行业的应用综述与发展趋势[J]．物流技术与应用，2021，26（7）：116－121.

[2] 刘祖杨．工业机器人的发展现状与对策研究——以常州为例[J]．营销界，2019（21）：149.

[3] 王慧艳．广东智能机器人战略性新兴产业集群发展现状和对策研究[J]．广东经济，2020（12）：24－29.

[4] 冯云华．智能化学化工实验室的研究与设计[J]．石化技术，2019，26（3）：310，309.

[5] 张晓颖，李德伟，席裕庚，等．六自由度机械臂约束预测控制系统的设计[J]．控制理论与应用，2014，31（11）：1464－1472.

[6] 黄晓林，周浩锋，曾泳达，等．基于手势识别远程同步控制的智能机械臂研究[J]．计算机测量与控制，2021，29（9）：128－131.

[7] 李炳臻，刘　克，顾佼佼，等．卷积神经网络研究综述[J]．计算机时代，2021（4）：8－12，17.

[8] 花傲阳．基于手势交互的机械臂轨迹示教系统[J]．软件工程，2021，24（8）：55－59.

[9] SUAY H B，TORIS R，CHERNOVA S. A Practical comparison of three robot learning from demonstration algorithm[J]. International Journal of Social Robotics，2012，4（4）：319－330.

[10] 胡钊龙，李栅栅．语音识别技术在智能语音机器人中的应用[J]．电子技术与软件工程，2021（13）：72－73.

[11] CLAUSSEN M，禾　沐．树莓派 4：全新的背后，依然是优秀的表现吗？[J]．单片机与嵌入式系统应用，2020，20（6）：1－4.

[12] TAN M，LE Q V. Efficientnetv2：Smaller models and faster training[J]. arXiv preprint arXiv：2104.00298，2021.

[13] SCHULDT C，LAPTEV I，CAPUTO B. Recognizing human actions：a local SVM approach[C]//Proceedings of the 17th International Conference on Pattern Recognition，2004. ICPR 2004. IEEE，2004，3：32－36.

[14] DUAN Y，ZOU B，XU J，et al. OAA-SVM-MS：A fast and efficient multi-class classification algorithm[J]. Neurocomputing，2021，454（2）：448－460.

[15] 张全贵，王炳超，李　凡，等．基于 SVM 的步态识别方法综述[J]．测控技术，2016，35（8）：1－5.

[16] 乔风娟，郭红利，李　伟，等．基于 SVM 的深度学习分类研究综述[J]．齐鲁工业大学学报，2018，32（5）：39－44.

[17] TAN M，LE Q. Efficientnet：Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning，MLR，2019：6105－6114.

[18] 程　美，王力华．医疗智能语音技术与应用综述[J]．中国数字医学，2021，16（8）：1－7.

[19] 李运寰，闻继伟，彭　力．高帧率的轻量级孪生网络目标跟踪[J]．计算机科学与探索，2021：1－13.

[20] XU R. Path planning of mobile robot based on multi-sensor information fusion[J]. EURASIP Journal on Wireless Communications and Networking，2019，2019（1）：1－8.

[21] 陈　慈，张敬磊，盖姣云，等．融合方法研究进展[J]．科技视界，2019（17）：32－33，6.

[22] 黄战华，陈智林，张晗笑，等．基于音视频信息融合的目标检测与跟踪算法[J]．应用光学，2021，42（5）：867－876.

[23] REDMON J，ANGELOVA A. Real-time grasp detection using convolutional neural networks[C]//2015 IEEE International Conference on Robotics and Automation（ICRA），IEEE，2015：1316－1322.

（上接第 90 页）

[17] PAN B，YU L P，YANG Y Q，et al. Full-field transient 3D deformation measurement of 3D braided composite panels during ballistic impact using single-camera high-speed stereo-digital image correlation[J]. Composite Structures，2016，157：25－32.

[18] XING H Z，ZHANG Q B，RUAN D，et al. Full-field measurement and fracture characterisations of rocks under dynamic loads using high-speed three-dimensional digital image correlation[J]. International Journal of Impact Engineering，2018，113：61－72.

[19] 徐　超，张一凡，韩晓明，等．基于机器视觉的大柔性结构振动位移测量[J]．振动、测试与诊断，2017，37（4）：781－786.

[20] NGOCSAN H，TAILIE J，NAMSEO G. Modal analysis of an artificial wing mimicking an Allomyrina dichotoma beetle's hind wing for flapping-wing micro air vehicles by noncontact measurement techniques[J]. Optics and Lasers Engineering，2013，51（5）：560－570.

[21] HAGARA M，TREBUNA F，HUNADY R，et al. Experimental identification of modal parameters of thin metal sheets by using of DIC[J]. Procedia Engineering，2012，48：180－188.