

卓越计划——毫米波雷达手势识别终期报告

张紫石 王怡 齐成浩 刘瑾鸿

0 项目背景

手势识别技术是人机交互领域的典型代表。在各类人机交互可传递的信息种类中，手势是最为特殊的一项。手势作为一种非语言类沟通方式，诞生时间早，表意相对直接，可在较大程度上弥补语言沟通的欠缺，在人日常生活中起到举足轻重的作用。手势识别技术，通过使计算机获取并理解人的手势信息，进而做出相应反应，可打破人机接触式交互的诸多限制，增加人机交互方式，拓宽交互场景。手势识别技术的应用领域十分广泛，如车辆驾驶环境中，通过手势对车内各项功能进行控制，可有效提升驾驶安全性；物联网及智能家居环境中，通过手势控制各类家居使用，如电视换台、音响控制等，可使人居生活更具便利性。

现有较为成熟的手势识别技术，大多通过光学传感器或可穿戴设备获取手势信息，进行手势识别，而这两种信息获取手段均存在其局限性。基于光学传感器的手势识别方案，通过摄像头采集手势信息，以二维图像

的数据形式进行呈现。该方案对手势所处环境的光照强度较为敏感，在光照强度变化剧烈时，识别准确率会显著降低。此外，光学图像的获取无法保证有效防止用户隐私泄漏问题，故采用光学传感器的方案存在较大隐忧。对于可穿戴设备的方案而言，其通过速度传感器、陀螺仪等对手势数据进行测量，存在便携性较差，成本较高等问题。基于上述局限性，研究基于其他传感器的手势识别方案具有很大意义。基于毫米波雷达的手势识别作为新兴技术，克服了传统基于光学图像的诸多问题，应用前景广阔，然而由于动态手势具有复杂多样、时空特征多变的特点，给人机交互和手势识别的研究带来了许多新的挑战，因此该技术在具有极大研究价值的同时，还具有一定挑战性。

1 国内外研究现状

手势识别最初属于计算机科学领域的一个课题，是实现人机交互的最直观有效的方法之一。最初的手势识别典型应用有数据手套，主要是利用多个传感器直接检测手、胳膊等各关节的方向、位置等信息，再传输到计算机系统中。1983年由Grimes等人获得的“数据手套”专利。但是数据手套价格昂贵，且舒适感较差。

于是视觉系统的图像识别应运而生，代德武等人基于视觉的手势识别系统，分

析引入的信息熵具备的特点，设计出得到视频关键帧的算法。Li Yunan 和 Miao Qiguang 将 RGB 图像特征和深度图像特征融合后送入支持向量机(Support Vector Machine, SVM)分类器完成分类和识别。Hong Chen 和 Lu Yang 等人将基于深度图像的手势识别研究归为了三类：第一类是静态手势识别，第二类是手势跟踪和识别，第三类是连续手势动作识别。在手势识别领域深度图像传感器主要有 Kinect、Leap motion 以及基于测量飞行时间(Time of Flight, TOF)原理的第 2 代 Kinect。微软推出的 Kinect 是 Xbox 360 游戏机的交互设备，能够利用第三方提供的开发套件和开发库方便地提取出人体的骨骼框架，从而完成手势识别。Vatavu R 基于手势骨架提取研究了控制电视的人机交互应用。此外，深度信息则常与 RGB 图像结合形成 RGB-D(Red-Green-Blue-Depth)数据，再基于图像处理和计算机视觉算法进行研究。这种基于视觉系统手势识别相对于数据手套虽然体感更舒适、表达更自然、识别准确率更高、识别手势更复杂多样，但是需要经过大量的图像处理算法进行计算，并且计算机硬件要求比较高。另外由于基于视觉系统的传感器需要直接的视线，其存在的致命的缺点有：对诸如光线、雾、雨和背景混乱之类的环境条件敏感，恶劣情况下会影响采集的图像的质量甚至致盲。

与基于视觉的系统不同，雷达传感器

可以独立于照明和操作环境来检测和分类目标，并且雷达可以捕获非常小的幅度的运动，能够对非常微妙和精确的手势进行分类。近年来，基于雷达的动态手势识别方法已经引起了工业界和学术界的广泛关注。

近 20 年来，国内外众多学者致力于手势识别应用的研究，微软研究的 Sound Wave 技术，利用多普勒效应，辨别出靠近计算机的手势动作。在 2014 年，有学者利用连续波雷达研究出 PCA 算法完成伸手、举手两种动作的识别。2016 年谷歌展示了 Project Soli 毫米波雷达技术，通过手势控制设备的微型雷达，该团队利用 2 年时间，从雷达的软硬件层面出发，使雷达芯片体积、功耗以及算法效率上均有较大提升，同时在 2019 年发布的 Google Pixel4 系列手机中得到实际应用。。2018 年北京航空大学的学生提出的卷积神经网络算法，提高了手势识别率并保证了良好的实时性。2018 年瑞典 Acconeer 公司经过长达十多年的科研工作与商业化探索，发布了 PCR 雷达传感器，该雷达采用 66GHz 频段，结合 AI 算法，可以准确实现 5 种以上手势检测，并具有低延迟实时的特性，同时为多种设备的应用提供了解决方案，如耳机、手环。

在达手势识别的技术研究和应用过程中，虽然已经取得了一定的成果，但通过上文的分析，发现仍然存在一些局限：

(1) 硬件设计。手势识别技术系统的应用离不开精度高、速度快、功耗低、体积小的硬件平台，然而目前的硬件设备很少能同时满足上述条件。

(2) 设备精度。识别设备的信息采集有限，如低频段的雷达平台测量精度有限。

(3) 数据预处理。好的分类算法离不开稳定的数据，如何将手势类型与周围环境分离，进而确定手掌位置一直是难点问题之一。

(4) 数据标签。目前大量的分类方法多采用深度学习的方法，而建立带标记的雷达手势样本训练集是一件非常耗时的工作。

(5) 特征的利用。动态手势复杂多变，现有研究中大多利用单一特征进行分类，代表信息有限且存在的一定的不稳定性。

(6) 分类算法。现有方法注重整体识别率的提高，而很少从局部入手，如寻找易混淆手势分类的方法，从而提高整体的识别率。

2 整体方案

2.1 系统指标

此项目的目的是识别 6 种手势，六种手

势分别为向左滑动，向右滑动、前进、后退、顺时针画圆、逆时针画圆，其中分别对应歌曲的上一首、下一首、暂停/播放、打开/关闭、提高音量、降低音量。如下图所示：

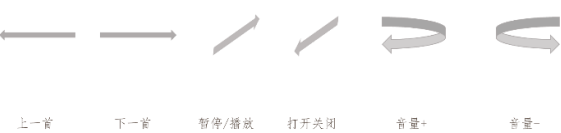


图 1 手势选取以及对应的功能

2.1 系统整体架构

如图 2 所示，系统首先通过 IWR1843+DCA1000 获得 ADC 原始数据，再进行 2DFFT 得到距离多普勒热图，再从中提取六个特征（每一帧对应六个数），分别为瞬时能量，加权距离，加权速度，加权方位角，距离和速度相关性，方位角和速度相关性。之后以十六帧作为一个动作的检测过程，并形成六个长度为 16 的向量，然后通过 ANN 网络训练，最后实现手势命令的识别，并通过手势命令控制音乐的播放。

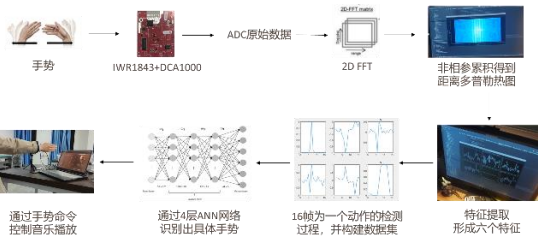


图 2 系统流程图

2.2 硬件系统

硬件系统选取德州仪器的 77G 毫米波雷达开发板 IWR1843 BOOST 作为传感器及数据

采集装置, 该雷达系统如下图所示:

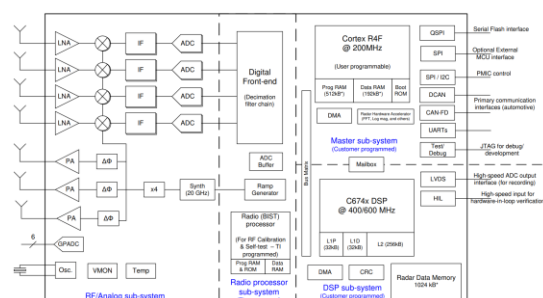


图 3 雷达系统结构框图

该雷达具有三发四收的天线，天线分布如下图所示，可以良好估计目标方位角，

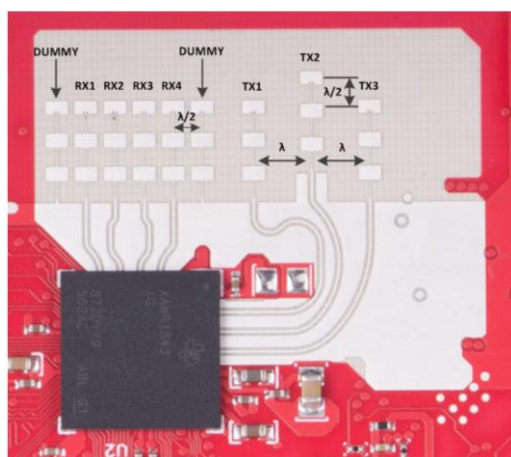


图 4 IWR1843 天线分布图

配合 DCA1000 数据采集卡使用，可以实时获取四个通道接收的原始数据。系统硬件平台如下所示：

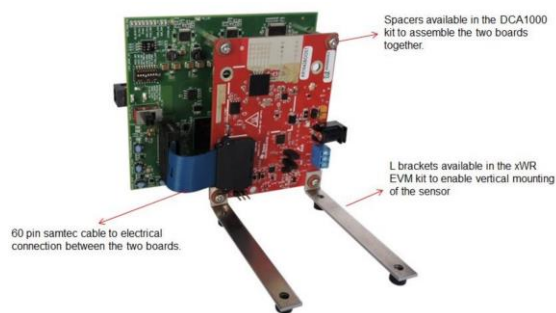


图 5 IWR1843+DCA1000

在本系统中选取的雷达具体参数如下所示:

平台: TWR1843 (三发四收)

距离分辨率: 0.214m

速度分辨率: 0.08m/s

最大不模糊距离: 10.95m

最大径向速度: 5.12m/s

水平角分辨率: 15°

距离维 FFT 点数: 64

速度维 FFT 点数: 128

3 系统实现

3.1 特征提取

上文提到我们在一帧中提取出六个特征分别为瞬时能量, 加权距离, 加权速度, 加权方位角, 距离和速度相关性, 方位角和速度相关, 本小节将具体阐述其提取方法。

瞬时能量

为防止躯体等其他运动物体的抖动对本手势识别系统造成影响,瞬时能量取距离维度的前三维进行求和,并设置阈值作为判断是否存在手势的判断依据,结合上述的距离分辨率为 0.214m 可知该系统的有效识别范围大约在正对雷达天线方向的 0 到 0.6m 之间。

加权距离

在进行距离加权时,首先将距离多普勒热图的速度维累积起来,在本系统中由于距离维FFT点数为64,速度维FFT点数为128,因此得到的距离多普勒热图为 64×128 的矩阵,结合上述方法,我们得到一个 64×1 的一维向量,并将这个一维数组累加得到一个值A,接着初始化一个与这个一维向量大小相同的一维数组,在本系统中选择的一维数组为 $[-32, -31, \dots, 30, 31]$,将这两个数组做点乘运算,之后的到一个 64×1 的一维数组,并将其累加得到一个值B,并用 B/A 得到所需要的加权。

加权重度

与距离加权方式类似,不同的是只选取距离维的前3维进行累积,初始化的一维向

量为 $[-63.5, -62.5, \dots, 62.5, 63.5]$

加权方位角

与距离加权方式类似，不同的是只选取距离维的前 3 维进行累积，且沿角度维做 FFT 时先将其 padding 为 128 维后再做 FFT。

距离和速度相关性

将上述的距离加权和速度加权，取 5 帧求解相关系数。

方位角和速度相关性

将上述的水平角加权和速度加权，取 5 帧求解相关系数。

3.2 形成特征向量

如下图所示将当前帧所提取出的六个特征值，与前 7 帧与后 8 帧的特征值拼接，得到了 6 个长度为 16 的特征向量，并将其送入 ANN 网络，进行训练。

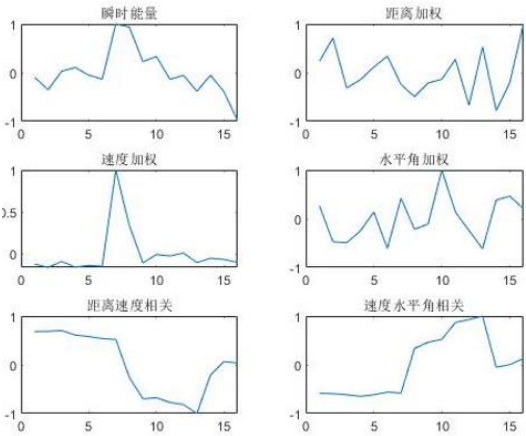


图 6 16 帧组成的特征向量示意图（后退手势）

3.3 ANN 网络结构

该系统使用的简单的 4 层 ANN 网络，如下图所示：

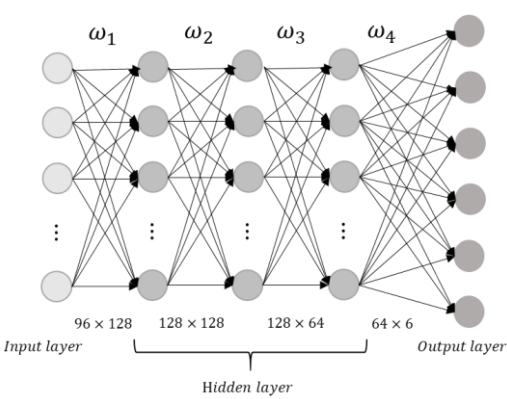


图 6 ANN 网络结构图

每层的激活函数为 LeakRelu 函数，每层正则化因子为 0.25，每一层参数都进行 Xavier 初始化。

训练时，使用了 Adam 优化器、对每个手势的 60 个数据样本进行反复训练。

4 实现效果

对训练数据的识别率能达到 98%，音乐播放能体验感较好。

5 未来展望

尽管系统初步实现了功能但是仍然存在许多待改进的地方，如下所述：

- 1) 提高系统可靠性：提高手势的识别准确率，减少误判的几率；
- 2) 提高系统泛化性：在更广泛的场景下、在手势不够标准的情况下也能精确识别手势；
- 3) 提高系统功能性：提高手势识别的种类，将手势识别应用到更广泛的场景种。