



计算机科学

COMPUTER SCIENCE

基于点云轨迹和压缩多普勒的跨场景手势识别

张宏旺, 周瑞, 程宇, 刘辰旭

引用本文

张宏旺, 周瑞, 程宇, 刘辰旭. 基于点云轨迹和压缩多普勒的跨场景手势识别[J]. 计算机科学, 2024, 51(2): 182-188.

ZHANG Hongwang, ZHOU Rui, CHENG Yu, LIU Chenxu. [Cross-scene Gesture Recognition Based on Point Cloud Trajectories and Compressed Doppler](#) [J]. Computer Science, 2024, 51(2): 182-188.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于迁移学习的跨对象手语手势识别方法](#)

Transfer Learning Based Cross-object Sign Language Gesture Recognition Method
计算机科学, 2023, 50(6A): 220300232-5. <https://doi.org/10.11896/jsjx.220300232>

[基于LSTM的多维度特征手势实时识别](#)

Real-time LSTM-based Multi-dimensional Features Gesture Recognition
计算机科学, 2021, 48(8): 328-333. <https://doi.org/10.11896/jsjx.210300079>

[基于3D卷积神经网络的CSI跨场景手势识别方法](#)

CSI Cross-domain Gesture Recognition Method Based on 3D Convolutional Neural Network
计算机科学, 2021, 48(8): 322-327. <https://doi.org/10.11896/jsjx.200600122>

[基于UNITY3D的水下机器人视景仿真方法](#)

Underwater Robot Visual Simulation Based on UNITY3D
计算机科学, 2021, 48(6A): 281-284. <https://doi.org/10.11896/jsjx.200700131>

[基于自适应多分类器融合的手势识别](#)

Hand Gesture Recognition Based on Self-adaptive Multi-classifiers Fusion
计算机科学, 2020, 47(7): 103-110. <https://doi.org/10.11896/jsjx.200100073>

基于点云轨迹和压缩多普勒的跨场景手势识别

张宏旺 周瑞 程宇 刘辰旭

电子科技大学信息与软件工程学院 成都 610054

(zhw6770@qq.com)

摘要 毫米波雷达能够用于各种感知任务,如活动识别、手势识别、心率感知等。手势识别作为其中的研究热点,可实现无接触人机交互。目前大多数手势识别研究使用点云或距离多普勒图通过神经网络进行识别感知,但是这些方法存在一些问题。首先,这些方法鲁棒性较差,被感知人员或其位置发生变化都会影响接收到的毫米波信号,降低感知精度。其次,这些方法将完整的距离多普勒图输入神经网络进行识别,由于图中存在较多与感知任务无关的区域,模型复杂且难以专注于感知任务。为解决这些问题,首先从连续多帧点云数据中建立手势轨迹,然后将连续多帧距离多普勒图进行局部切割并压缩获得二维局部多普勒图,最后将点云轨迹和二维局部多普勒图分别经过神经网络特征提取后,对特征进行拼接,通过全连接神经网络进行分类。实验结果表明,所提方法专注于手势,能够达到98%的识别准确率,在人员变化和位置变化情况下对新用户和在新位置的识别准确率分别能够达到93%和92%,高于现有方法。

关键词:毫米波雷达;手势识别;跨场景;位置无关;人无关

中图分类号 TP391

Cross-scene Gesture Recognition Based on Point Cloud Trajectories and Compressed Doppler

ZHANG Hongwang, ZHOU Rui, CHENG Yu and LIU Chenxu

University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract Millimeter wave radar can be used for various sensing tasks, such as activity recognition, gesture recognition, heart rate perception. Among them, gesture recognition is a research hotspot, which can realize contactless human-computer interaction. Most existing studies on gesture recognition make use of point cloud or range-Doppler for pattern recognition through neural networks to achieve sensing. However, there are some problems. Firstly, the robustness of these methods is poor. The changes of the user and his/her location affect the received millimeter wave signals, causing the accuracy of the sensing model to reduce. Secondly, these methods input the complete range-Doppler map into the neural network, which makes the model complicated and makes it difficult for the model to focus on the sensing task, because there are many unrelated regions to the sensing task. To solve these problems, this paper first builds the gesture trajectory from multiple continuous frames of point cloud, and then cuts and compresses the multiple continuous range-Doppler maps to obtain the two-dimensional local Doppler map. Finally, the features are extracted from the point cloud trajectory and the two-dimensional local Doppler map respectively by the neural networks, concatenated and classified by a fully-connected neural network. Experiments show that the proposed method focuses on gestures and can achieve a recognition accuracy of 98%, and can achieve a recognition accuracy of 93% for new users and 92% for new locations in the cases of user changes and location changes, better than the state of the art.

Keywords Millimeter wave radar, Gesture recognition, Across scenarios, Location independent, User independent

1 引言

无线感知作为一种新兴感知技术,能够在不侵犯用户

隐私的情况下实现无接触感知,也因此受到越来越多研究者的关注。无线信号主要包括毫米波、超声波、WiFi信号等。毫米波是一种波长在1~10 mm的电磁波,由于其具有高频率、

到稿日期:2023-04-27 返修日期:2023-11-16

基金项目:四川省科技服务业(2021GFW027);四川省科技计划(2022YFSY0006);四川省科技计划(2023YFSY0007);泸州市科技计划项目(2022-XDY-192)

This work was supported by the Sichuan Province Science and Technology Service(2021GFW027), Sichuan Province Science and Technology Plan(2022YFSY0006), Sichuan Province Science and Technology Plan(2023YFSY0007) and Science and Technology Plan Project of Luzhou City, Sichuan Province(2022-XDY-192).

通信作者:周瑞(ruizhou@uestc.edu.cn)

高带宽的特点,能够实现高精度感知,因此在无线感知领域具有得天独厚的优势。研究人员采用毫米波技术实现了各种感知任务,如眨眼识别^[1]、心率检测^[2-3]、身份识别^[4]、手势识别^[5-7]、轨迹跟踪^[8]和血压监测^[9]等。手势识别作为其中的研究热点,通过空中手势来控制家电、灯光等,可实现无接触人机交互。文献[5]设计了一种使用毫米波雷达进行手势采集的方法,然后使用深度神经网络进行手势分类。文献[6]致力于减小不同人对手势识别的影响,实现与人无关的手势识别。文献[7]致力于手势识别的应用化,如控制音响和灯光,实现了实时手势识别。尽管使用毫米波雷达进行手势识别的研究日益增多,仍存在一些问题:

(1)用户依赖性。现有手势识别系统大多只适用于参与训练的用户。当增加新用户时,需要重新采集新数据来进一步训练适应,这不利于系统的实际部署。文献[6]从毫米波中提取了一种新的信号伪代表模型(Pseudo Representative Model,PRM)信号作为新输入数据,以该信号为基础,去除人物干扰,实现了音乐播放器的手势控制。但其中模型精度和训练人数呈正相关,要达到高精度,至少需要20~30人的数据量。

(2)位置依赖性。当用户和雷达之间的相对位置发生变化时,多普勒信息也会发生变化。这是多普勒强度感知减弱造成的,体现在多普勒数据上便是动作信息部分丢失,导致模型在训练位置上精度高,在新位置上精度下降。文献[7]提出了基于不同距离的手势感知,建立了多个模型对应多个距离,根据点云信息计算检测对象的距离,根据距离选择相应模型。这种方法需要实现针对距离的建模,通用性不强。另外,当用户和雷达之间的相对角度发生偏转时,接收到的多普勒信息也会发生改变,导致识别准确率大幅下降。这个问题目前还没有好的解决方案。

(3)深度神经网络复杂,难以实现实时检测。从毫米波雷达中采集的点云数据和距离多普勒均为多帧时序数据,从这些数据中提取特征需要复杂的神经网络,且难以专注于感知任务,导致计算时间较长,不利于实时检测。

针对以上两个问题,本文提出从毫米波信号中提取两种新的信息,即点云轨迹和二维局部多普勒图,融合后进行手势识别。点云轨迹从连续多帧点云数据中提取,通过对点云聚类,找到当前帧的中心点,逐帧聚类并连贯,获得点云轨迹信息。二维局部多普勒图从连续多帧距离多普勒图中经过切割和压缩进行提取。点云轨迹通过长短期记忆(Long Short-Term Memory,LSTM)网络提取特征,二维局部多普勒图通过卷积神经网络(Convolutional Neural Network,CNN)提取特征。两种特征经过拼接后,通过全连接神经网络实现手势分类。实验验证了本文方法具有良好的手势识别效果和鲁棒性,能够消除用户和位置变化的影响,同时能够简化神经网络,减少训练时间。本文的主要贡献如下:

(1)提出从连续多帧点云数据中提取轨迹信息进行手势识别。轨迹信息对于位置变化具有鲁棒性,当用户更换位置时,仍可以实现高精度手势识别。同时实现了一种记忆体制,弥补了毫米波雷达某些时刻不能捕获数据的缺陷。

(2)提出从连续多帧距离多普勒图中提取二维局部多普

勒信息进行手势识别。以点云轨迹信息为基础,分离出局部多普勒信息,在速度维度进行压缩获得二维局部多普勒信息,达到专注于手势本身并降低模型复杂度的目的。

(3)融合轨迹信息与二维局部多普勒信息进行手势识别。轨迹信息在位置改变时鲁棒性高,二维局部多普勒信息在用户改变时鲁棒性高,本文方法兼顾两种数据的优点,使得手势识别获得高精度与高鲁棒性。

2 相关工作

2.1 毫米波感知

毫米波作为一种重要的无线信号,在无线感知研究中得到了高度关注,如眨眼识别^[1]、心率检测^[2-3]、身份识别^[4]、手势识别^[5-7]、轨迹跟踪^[8]和血压监测^[9]等。在文献[1]中,研究人员使用毫米波雷达检测人物的眨眼次数,从而有效分析人物的疲劳状态。文献[2]使用毫米波进行人物的心率检测,从而对人物的健康数据进行记录,当检测到异常时及时报警。文献[5-7]使用毫米波对人物手势进行识别,以此为基础实现无接触人机交互。文献[4]使用毫米波雷达进行人物跟踪和身份识别,在不侵犯隐私的情况下,实现入侵检测和居家安全保障。文献[9]使用毫米波进行血压监测,将其应用于医学领域。毫米波的应用方兴未艾,正以一种泛在的方式改变着人们生活。

2.2 无接触手势识别

手势识别是毫米波感知的一个重要方向,是实现人机交互的重要接口,在各种信号领域均有研究。在声波领域,由于信号频率低,大部分工作都围绕多普勒频移开展。文献[10-11]研究基于声波的手势识别和手跟踪。在WiFi领域,有大量有关手势识别的研究。文献[12-13]基于WiFi进行手势识别,但是WiFi信号受时间、环境和人物等的影响较大,难以实现广泛应用。在视频领域,文献[14-15]实现了高精度手势识别,但是视频方法有隐私泄露风险,在无光^[16]和遮挡场景下会失效。

作为无线感知的重要内容,毫米波手势识别研究也没有停下脚步。文献[5]提出了一种使用小型毫米波雷达的手势动作采集方法,并使用深度神经网络进行识别,但是该方法没有实现与人无关性。为了实现与人无关的手势检测,文献[6]提出伪代表模型PRM信号进行模式识别,实现了人变化时的鲁棒性。文献[7]建立了一个实时在线手势检测系统,使用马尔可夫链进行手势推测,可以在手势动作完成之前进行手势推断。文献[17-18]研究无接触的手势识别,在实时及应用性方面进行讨论。

上述毫米波手势识别方法尽管能够取得较好的识别结果,但是都没有达到真正的鲁棒性。文献[7]为了解决距离变化的影响,根据距离从多个模型中选取其一。文献[6-7]使用的方法信息源主要是多普勒信息,然而多普勒信息仅仅反映靠近和远离雷达的径向速度信息,因而在位置改变时,在该方向上的分量表现不尽相同,导致在位置上鲁棒性不足;此外,不同人物的习惯不同,同一动作,多普勒速度会有差异。文献[6]从每一帧提取出十维特征作为信号源,其实质反映的是动作的流体性,是对手势肢体的各部分描述,但是仍然离不开对

速度信息的依赖,当位置改变时,多普勒效应不明显时,分类结果准确率会降低。

3 手势识别算法

本文的目标是从现有的毫米波雷达数据中提取高鲁棒性的信息,使得识别过程更专注于手势本身,忽略其他无关信息。以此信息为基础,使用神经网络进行特征的提取与融合,进而实现手势识别。本文方法的整体框架如图 1 所示,包括雷达数据采集、鲁棒信息提取和手势识别三大部分。采集手势动作时,毫米波雷达可以采集到点云时序数据和距离多普勒

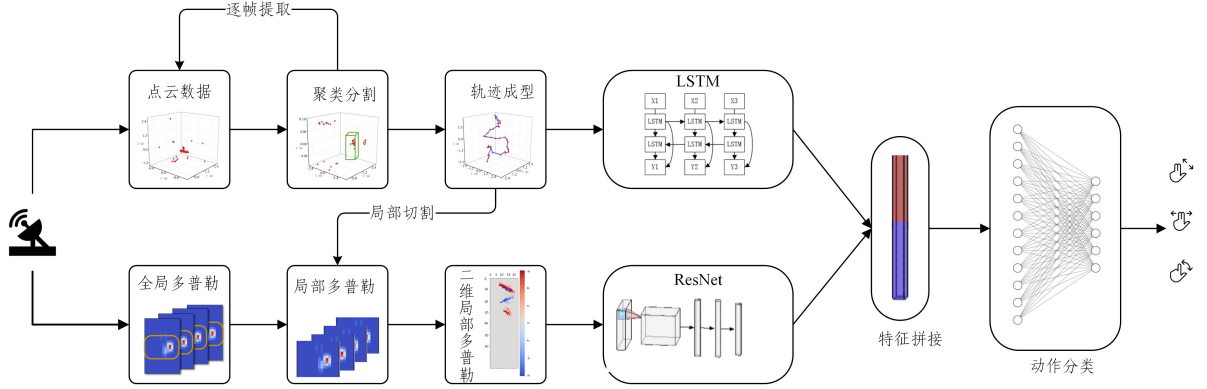


图 1 方法总体框架

Fig. 1 General framework of the proposed method

3.1 毫米波雷达数据

本文使用 TI IWR1443 boost 雷达板采集数据,其工作在 76~81 GHz 频段,拥有 4 GHz 连续线性调频信号,包括内置了 PLL 和 A2D 转换器的 3 根发射天线和 4 根接收天线。通过配套程序 mmWave_Visualizer,可以从该雷达板中获取点云数据和距离多普勒数据。

3.1.1 点云数据

手势样本的点云数据为时序数据,每一时间点包含一帧点云数据,每一帧点云数据 P 包括各个点的基本信息:

$$P = (p_1, p_2, \dots, p_n) \quad (1)$$

其中, n 代表点云中点的数目,该雷达板能检测到的最大点数为 64。 p_i 表示点 i 的数据,为一个六维向量:

$$p_i = (r, v, s, x, y, z) \quad (2)$$

其中, r 代表距离, v 代表速度, s 代表强度, (x, y, z) 代表该点位置坐标。图 2 展示了一个 O 型手势中的两帧点云数据,表示该手势的当前空间分布。

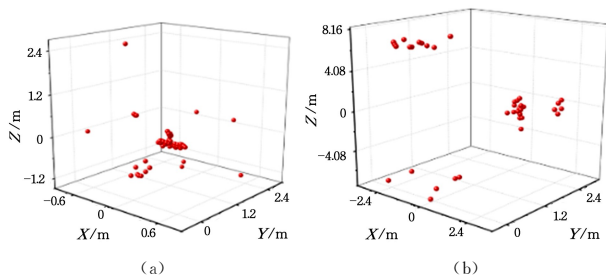


图 2 多帧点云数据

Fig. 2 Multi-frame point cloud data

3.1.2 距离多普勒图

雷达采集的多普勒频移和速度计算公式可表示为:

勒时序数据。为了获取高鲁棒信息,本文方法首先对点云数据逐帧聚类,求取聚类中心,而后将动作帧的各个聚类中心连接起来,实现轨迹的提取。当获取轨迹之后,从现有轨迹出发,可以获得手势发生时的距离信息。以此距离信息为基础,从距离多普勒图中截取与当前手势相对应的局部多普勒图,并在速度维度进行压缩,从而获得二维局部多普勒数据。最后以两种信息作为网络模型的输入数据进行训练,针对这两种信息,本文分别采用预训练模型 ResNet 神经网络和 LSTM 神经网络作为特征提取器,使用全连接作为分类器。训练完成后,可以将实时测试数据输入网络中,进行手势识别。

$$F = \frac{2V}{\lambda}, F = \frac{K \times f_c}{N} \quad (3)$$

其中, F 是求取的多普勒频移, v 是求取的速度, λ 为波长, f_c 为 chirp 的频率, k 表示求取快速傅里叶后的最大值的取样的频率值, N 为每秒采样的次数。

采集的手势的距离多普勒数据 $D_{t,r,v}$ 为多帧距离多普勒图,可表示为:

$$D_{t,r,v} = (S_{r,v}^1, S_{r,v}^2, \dots, S_{r,v}^t) \quad (4)$$

其中, t 是时间索引,一帧距离多普勒数据 $S_{r,v}$ 则是一个二维矩阵:

$$S_{r,v}^t = \begin{pmatrix} s_{1,1}^t & \dots & s_{1,v}^t \\ \vdots & \ddots & \vdots \\ s_{r,1}^t & \dots & s_{r,v}^t \end{pmatrix} \quad (5)$$

其中, r 表示距离索引, v 表示速度索引。在该雷达中 v 的数目为 16, r 的数目为 128。一帧距离多普勒图反映的是不同距离上人物的速度分布。图 3 给出了一个 O 型手势的距离多普勒序列数据。

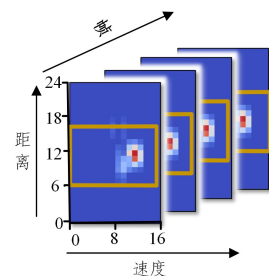


图 3 多帧距离多普勒图

Fig. 3 Multi-frame range Doppler map

3.2 基于记忆体机制的点云轨迹提取

将采集的多帧点云数据转化为轨迹数据。首先针对手势数据中的每一帧点云,使用基于密度的聚类算法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)实现点云聚类 and 离散点去除。通过观察雷达采集的手势数据,可以发现点云的排列在三维空间中的Z轴排列稀疏,而在X和Y轴排列紧密。因此本文定义DBSCAN算法的距离计算公式为:

$$d = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + 0.001 * (z_a - z_b)^2} \quad (6)$$

其中, (x_a, y_a, z_a) 和 (x_b, y_b, z_b) 分别为两个点的位置坐标。设置DBSCAN算法的最大半径 $R = 150$, 最小点数 = 3。DBSCAN对点云聚类后,该点簇代表了检测对象。对每一帧点云进行处理,进而获得每一帧中检测对象的中心位置。将检测对象的各帧连贯即可获得完整的手势轨迹。经过DBSCAN聚类后,得到数据:

$$\mathbf{K} = (\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_n) \quad (7)$$

其中, \mathbf{k}_i 为第 i 帧的聚类中心,表示当时检测对象的位置。当雷达捕获点云时,某些帧可能未获得点云,即 \mathbf{k}_i 没有数据,这会导致将手势分割为两个或以上的不相关轨迹。为解决此问题,本文提出了一种记忆机制进行轨迹的连贯。当发现当前帧没有捕捉到检测对象时,轨迹绘制并不会立即结束,而是将上一帧的检测对象保存在记忆体中,而后继续检测后续多帧。如果后续重新出现该检测对象,则将其与前面帧进行连贯,否则认为这是检测对象的一个新的手势。通过记忆体机制,可以将检测对象的各帧连贯,从而获得完整的手势轨迹。图4展示了两个手势的轨迹。

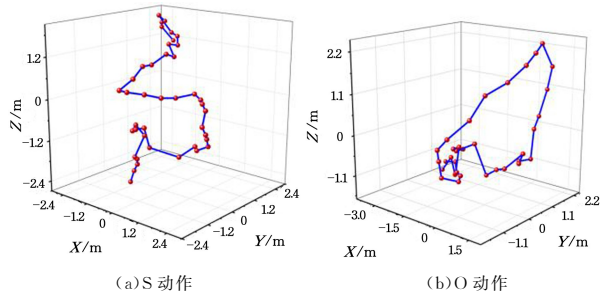


图4 手势轨迹图

Fig. 4 Gesture trajectories

3.3 局部二维多普勒信息的提取

从雷达获取多普勒信息,涉及全部距离上的速度变化,对于手势识别来说,这并不是必要的。同时使用多帧时序数据,需要复杂神经网络,不利于实时部署。因此,本文将多帧距离多普勒图进行切割,使用局部多普勒信息将视野局限于检测对象的手势,排除其他信息干扰,并降低神经网络复杂度。其中,使用局部信息隐性包含了只着眼于重要反射点的信息,如果使用完整的距离多普勒,则无法确认捕获的运动是否集中于目标身上,这就要求神经网络具有一定的排除干扰能力,使得网络复杂性增加。局部多普勒信息的提取方法如下:以轨迹信息为基准,获得检测对象的位置,根据位置将检测对象限制在一定空间范围内,从全局多普勒数据中分离出该检测对象的局部多普勒数据。该数据仍是三维数据,三维分别为时间-距离-速度。令 R_i 表示检测对象第 i 帧的位置,根据此位置

对该帧的全局多普勒数据进行切割,获取检测对象该帧的局部多普勒数据。对于第 i 帧全局多普勒数据 $S_{r,v}^i$,局部多普勒切割为:

$$S_{r',v'}^i = S_{r,v}^i [0:v, R_i - 12:R_i + 13] \quad (8)$$

即截取距离索引长度为 25 的矩形,作为检测对象第 i 帧的局部多普勒,其中 25 为经验长度,足以包括一个人物的动作范围。对时间序列中每一帧全局多普勒进行切割,最后获得检测对象的手势的多帧局部多普勒信息:

$$\mathbf{D}'_{t,r',v'} = (S_{r',v'}^1, S_{r',v'}^2, \dots, S_{r',v'}^t) \quad (9)$$

其中, t 表示时间维度, $S_{r',v'}^i$ 代表其中一帧,其格式为:

$$S_{r',v'}^i = \begin{pmatrix} s_{r',v'}^{i,1} & \dots & s_{r',v'}^{i,v'} \\ \vdots & \ddots & \vdots \\ s_{r',v'}^{i,1} & \dots & s_{r',v'}^{i,v'} \end{pmatrix} \quad (10)$$

其中, r' 和 v' 分别表示截取后的距离和速度索引维度。

对于一个局部多普勒来说,每一帧多普勒信息代表了该帧的检测对象在各距离范围上的速度可能性。毫米波雷达使用快速傅里叶(Fast Fourier Transform, FFT)进行速度求取,反映在多普勒图像上,便是当检测对象的距离确定后,速度可能性反映在求取 FFT 之后各个速度索引的强度上,强度越大表示检测对象在该位置上的速度越接近这个速度索引。对于手势识别来说,更需要关注的是一帧中各个距离上的实际速度和随时间各个帧的变化情况。因而对于速度索引维度上的概率反映,不仅会影响手势判断,还会增大神经网络的数据输入量。故本文在速度索引维度上压缩,将局部多普勒数据从三维压缩成二维。依据强度最大的速度索引值,计算出检测对象的真实速度,将二维的 $S_{r',v'}^i$ 压缩为一维的 $S_{r'}^i$,即:

$$S_{r'}^i = \arg \max_{v'} S_{r',v'}^i = (v_1, v_2, \dots, v_{r'}) \quad (11)$$

表示各个距离索引上的瞬时速度。对动作序列中所有帧进行压缩,从而将 $\mathbf{D}'_{t,r',v'}$ 压缩为 $\mathbf{D}'_{t,r'}$,得到完整的多帧手势数据:

$$\mathbf{D}'_{t,r'} = (\arg \max_{v'} S_{r',v'}^1, \arg \max_{v'} S_{r',v'}^2, \dots, \arg \max_{v'} S_{r',v'}^t) \quad (12)$$

本文将三维局部多普勒数据压缩为二维局部多普勒数据,作为后续神经网络的输入。二维多普勒信息中一个维度代表时间维度,另一个维度代表距离维度,其中的数值代表速度大小。该信息反映了检测对象随着时间变化在距离上的速度变化。多普勒数据切割和压缩过程如图5所示。

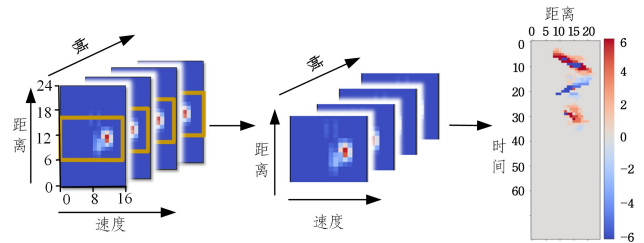


图5 多普勒数据的切割和压缩

Fig. 5 Cutting and compression of Doppler data

3.4 手势识别

手势识别的模型基本结构如图1所示,包括 LSTM 网络特征提取器(隐藏层参数为 10)和 ResNet(ResNet18)网络特征提取器,最后是全连接神经网络分类器(256-64-6 为三层的特征输出)。轨迹信息是一种时序数据,反映了手势动作作

时间在三维空间中的位置变化,因此本文采用 LSTM 网络作为点云轨迹的特征提取器。二维局部多普勒信息在不同时间、不同位置展现出不同速度,是一个有高光区的图像,因此使用预训练卷积神经网络 Resnet18 进行空间特征提取。本文对于二维局部多普勒使用三通道数据,主要根据速度的正负性来构建三通道输入数据。从 $D'_{t,r}$ 中分离出两个二维子矩阵 $D'_{t,r,+}$ 和 $D'_{t,r,-}$, 其中 $D'_{t,r,+}$ 中速度数值均小于 0, 大于 0 的数值用 0 填充; $D'_{t,r,-}$ 中速度数值均大于 0, 小于 0 的数值用 0 填充。分类器采用三层全连接神经网络。手势识别时,将轨迹数据和二维局部多普勒数据分别输入对应的特征提取器中,提取相关的特征,在进行特征拼接后,输入全连接神经网络,实现手势识别。相关的损失函数为:

$$Loss = - \sum_{i=1}^n y_i \log \hat{y}_i \quad (13)$$

本文方法中使用了轨迹信息和二维局部多普勒作为神经网络的输入。轨迹信息反映了手势的位置变化,用户位置的改变并不会在实质上影响手势轨迹的变化方式。二维局部多普勒可以有效检测人物手势,通过切割和压缩,使其专注于手势动作,忽略其他小的肢体习惯特征,从而实现鲁棒性。经实验验证,单独使用轨迹信息,通过 LSTM 神经网络进行特征提取,在换人、换房间和换位置的情况下都有良好的鲁棒性,但整体准确率低于二维局部多普勒信息。单独使用二维局部多普勒,在换人和换房间时准确率较高,但是在换位置时准确率会下降,因为随着位置改变,接收到的多普勒信息也会减弱,而轨迹则不受此影响。融合两种信息进行特征拼接可以有效弥补两者的不足,增强网络的鲁棒性,进而使得网络可以在跨场景情况下使用。

4 实验验证

4.1 实验设置

为了评估本文方法,我们使用 TI IWR1443 boost 毫米波雷达板进行数据采集和手势识别。设置该雷达的采样频率为 15 帧/秒,每一帧 chirp 数目为 128 个,采样时间为 4 s,因此,一个手势持续时间内可以采集到 60 帧数据。采集得到的点云数据的维度是 (60, 64, 6), 其中 64 表示最多采集到 64 个数据点, 6 是描述点云数据的一维向量长度。采集到的多普勒数据的维度是 (60, 128, 16), 其中 16 是速度索引个数, 128 是距离索引个数。实验模型训练的机器采用 RTX3090 显卡 (12GB 显存), 设备内存大小为 16 GB。手势识别在 3 个场景

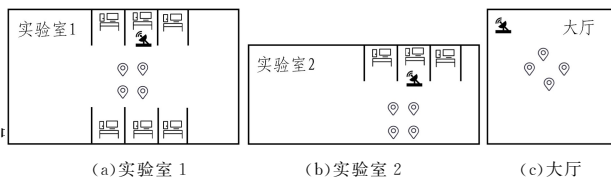


图 6 实验场景

Fig. 6 Experiment scenes

实验室 1 采样范围大小为 $8\text{m} \times 3\text{m}$, 发射器和检测对象的距离为 1 m 和 1.5 m; 实验室 2 采样范围大小为 $8\text{m} \times 2\text{m}$, 发射器距离检测对象的距离为 1 m 和 1.5 m; 大厅的采样范围大小为 $5\text{m} \times 5\text{m}$, 发射器和检测对象的距离为 1 m 和 1.5 m。每个实验场景 4 个位置, 其中前两个位置在中线上, 与发射器

的距离分别为 1 m 和 1.5 m, 后两个位置在 1 m 和 1.5 m 的基础上偏离中线 0.5 m。每个场景中 4 人参与实验。手势包括 6 个字母: L, O, V, S, W, Z, 每个手势重复 20 次。

4.2 实验结果

4.2.1 手势识别结果

实验首先验证不变场景下的手势识别结果, 作为后续的正确率上限。实验结果如表 1 所列, 3 个房间指 3 个房间同一位置同一个人的数据, 其中 80% 数据用于训练, 另外 20% 数据用于测试; 4 个用户指同一个房间同一个位置 4 个人的数据, 其中 80% 数据用于训练, 另外 20% 数据用于测试; 4 个位置指同一房间同一个人的 4 个位置, 其中 80% 数据用于训练, 另外 20% 数据用于测试。从表 1 中可以看出, 在不变场景下, 点云轨迹和二维局部多普勒可以获得较高的识别准确率。在 3 种情况下, 采用点云轨迹的准确率分别可以达到 90%, 93%, 93%, 采用二维局部多普勒的准确率分别可以达到 98%, 99%, 97%, 两种信息融合之后分别可以达到 98%, 99%, 99% 的准确率。

表 1 不变场景下手势识别结果

Table 1 Gesture recognition results in an invariant scene

	点云轨迹	二维多普勒	特征拼接
3 个房间	0.90	0.98	0.98
3 个用户	0.93	0.99	0.99
4 个位置	0.93	0.97	0.99

4.2.2 新用户手势识别

为验证本文方法对新用户的鲁棒性, 在同一房间同一位置进行了实验, 采用 3 人数据进行训练, 另外 1 人作为新用户进行测试。依次将每人作为新用户, 新用户数据完全没有参与训练, 仅作为测试使用。使用点云轨迹、局部二维多普勒和融合 3 种方法进行实验, 实验结果如表 2 所列。可以看出, 在用户变化时, 采用轨迹数据识别的效果不如采用二维多普勒数据的效果, 融合后识别精度明显提升。

表 2 新用户手势识别结果

Table 2 Gesture recognition results for new users

	点云轨迹	二维多普勒	特征拼接
用户 1	0.85	0.94	0.96
用户 2	0.86	0.88	0.90
用户 3	0.91	0.95	0.98
用户 4	0.84	0.89	0.90
平均	0.87	0.92	0.94

4.2.3 新房间手势识别

为验证本文方法对房间变化的鲁棒性, 实验选取 2 个房间数据作为训练数据, 另外 1 个房间数据作为测试数据, 一个用户参与实验, 在房间中位置固定。依次选取每个房间进行测试, 分别使用点云轨迹、局部二维多普勒和融合 3 种方法。手势识别结果如表 3 所列, 局部二维多普勒对房间变化具有较高鲁棒性, 融合后进一步提高了手势识别精度。

表 3 新房间手势识别结果

Table 3 Gesture recognition results in new rooms

	点云轨迹	二维多普勒	特征拼接
大厅	0.89	0.96	0.99
实验室 2	0.89	0.97	0.97
实验室 1	0.88	0.94	0.96
平均	0.89	0.96	0.97

4.2.4 新位置手势识别

为验证本文方法对新位置的鲁棒性,选取同一房间中 3 个位置作为训练数据,另 1 个位置作为测试数据,同一个人进行实验。依次选取每个位置进行测试,分别采用点云轨迹、局部二维多普勒和融合 3 种方法。手势识别结果如表 4 所列。当位置变化时,局部二维多普勒数据的识别精度明显下降,主要是因为多普勒接收减弱,导致一部分信号丢失,但点云轨迹对位置变化具有鲁棒性。融合之后识别结果较为稳定。

表 4 新位置手势识别结果

Table 4 Gesture recognition results for new positions

	点云轨迹	二维多普勒	特征拼接
位置 1	0.900	0.8600	0.920
位置 2	0.910	0.8800	0.950
位置 3	0.900	0.8500	0.91
位置 40	0.915	0.8800	0.920
平均	0.900	0.8675	0.925

4.2.5 业界对比

为证明本文方法的有效性,将其和业界其他方法^[6,8,19]进行了实验对比。文献^[6]中提出了一种新的信号量 PRM,

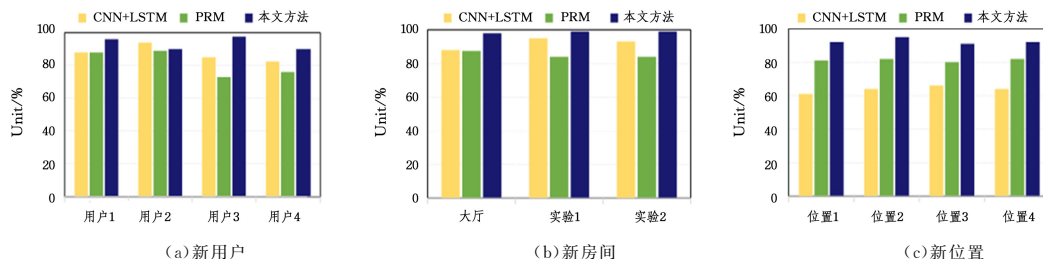


图 7 对比实验

Fig. 7 Comparison of experiment results

结束语 为了解决基于毫米波雷达的手势识别对用户、环境和位置的依赖,增强手势识别精度和鲁棒性,本文提出了两种新的信号特征:点云轨迹和局部二维多普勒。基于轨迹信息的手势识别在新位置的表现良好,而基于二维局部多普勒的手势识别在新人物和新房间中表现良好,因此本文融合两种信息进行特征拼接,使得本文方法能综合两者的优点,实现高精度、高鲁棒性手势识别,降低环境变化、用户多样、位置变化带来的负面影响。实验中使用两种信号特征进行手势识别,在场景不变的情况下平均准确率达到 98.6%,最终在新房间的平均准确率为 97%,新用户的平均准确率为 93%,新位置的平均准确率为 92%,高于现有方法。在后续工作中,将使用本文中提取的两种信号进行多个人物的感知识别,这可能要依赖于后续的对轨迹特征的分离,以及对多普勒信号的分离。与此同时,对移动人物的手势感知也将是接下来的重点研究方向。

参考文献

- [1] SHU Y, WANG Y, YANG X, et al. An improved denoising method for eye blink detection using automotive millimeter wave radar[J/OL]. EURASIP J. Adv. Signal Process. 2022, 9. <https://asp-urasipjournals.springeropen.com/articles/10.1186/s13634-022-00841-y>.
- [2] ZHAO P, LU C, WANG B, et al. Heart Rate Sensing with a Robot Mounted mmWave Radar[C]// IEEE International Conference on Robotics and Automation. 2020:2812-2818.
- [3] LING Z, ZHOU W, REN Y. Non-Contact Heart Rate Monitoring Based on Millimeter Wave Radar[J]. IEEE Access, 2022, 10:74033-74044.
- [4] ZHAO P, LU X, WANG J, et al. Human tracking and identification through a millimeter wave radar[J]. Ad Hoc Networks, 2021, 116(5):102475. 1-102475. 11.
- [5] WANG S, SONG J, LIEN J, et al. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum[C]// Proceedings of 29th Annual ACM Symposium on User Interface Software and Technology. 2016:851-860.
- [6] LIU H, DONG Z, SUN Y, et al. M-Gesture: Person-Independent Real-Time In-Air Gesture Recognition Using Commodity Millimeter Wave Radar[J]. IEEE Internet of Things Journal, 2022, 9(5):3397-3415.
- [7] LIU H, WANG Y, ZHOU A, et al. Real-time Arm Gesture Recognition in Smart Home Scenarios via Millimeter Wave Sensing [C]// Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technology. 2020.
- [8] WEI H, LI Z, GALVAN A, et al. IndexPen: Two-Finger Text

- Input with Millimeter-Wave Radar[C]//Proceedings of ACM Interactive, Mobile, Wearable and Ubiquitous Technology, 2022.
- [9] SINGH L, YOU S, JEONG B J, et al. Remote Estimation of Blood Pressure Using Millimeter-Wave Frequency-Modulated Continuous-Wave Radar[J]. *Sensors*, 2023, 23(14):6517.
- [10] MAOW, HE J, QIU L. CAT: High-precision acoustic motion tracking[C]//Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, 2016:69-81.
- [11] WATANABE H, TERADA T. Improving ultrasound-based gesture recognition using a partially shielded single microphone [C]//Proceedings of the 2018 ACM International Symposium on Wearable Computers, 2018:9-16.
- [12] GAO R, ZHANG M, ZHANG J, et al. Towards Position-Independent Sensing for Gesture Recognition with Wi-Fi[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021, 5(2):1-28.
- [13] XIAO R, LIU J, HAN J, et al. OneFi: One-Shot Recognition for Unseen Gesture via COTS Wi-Fi[C]//Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, 2021:206-219.
- [14] DOAN H G, VU H, TRAN T H. Recognition of hand gestures from cyclic hand movements using spatial-temporal features [C]//Proceedings of the 6th International Symposium on Information and Communication Technology, 2015:260-267.
- [15] PIGOU L, VAN DEN OORD A, DIELEMAN S, et al. Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video[C]//International Journal of Computer Vision, 2018:430-439.
- [16] ZHAO W, WANG T, TAN A. Nighttime Pedestrian Detection Based on a Fusion of Visual Information and Millimeter-Wave Radar[J]. *IEEE Access*, 2023, 11:68439-68451.
- [17] XIA Z, DING G, WANG H. Person Identification With Millimeter-Wave Radar in Realistic Smart Home Scenarios[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19:1-5.
- [18] GAO Z, ALI L, WANG C. Real-Time Non-Contact Millimeter Wave Radar-Based Vital Sign Detection [J]. *Sensors*, 2022, 22(19):7560.
- [19] ZHANG Y, PENG L. Dynamic Gesture Recognition Model Based on Millimeter-Wave Radar With ResNet-18 and LSTM [J]. *Frontiers in Neurorobotics*, 2022, 16:903197.



ZHANG Hongwang, born in 1998, post-graduate. His main research interests include wireless sensing and deep learning.



ZHOU Rui, born in 1974, Ph.D, associate professor, is a member of CCF (No. 52289M). Her main research interests include pervasive computing, Internet of things, and artificial intelligence.

(责任编辑:何杨)