

面向毫米波动作识别的视觉辅助信道仿真技术

任振裕^① 吉辰卿^① 余潮^① 陈万里^{*②} 王锐^{*①}

^①(南方科技大学电子与电气工程系 深圳 518055)

^②(深圳技术大学 深圳 518118)

摘要: 该文提出了一种利用计算机视觉技术辅助实现包含运动人体散射特征的毫米波无线信道仿真方法。该方法旨在为无线人体动作识别快速、低成本地生成训练数据集，避免当前实测采集数据集的巨大开销。首先利用基元模型将人体建模为35个相互连接的椭球，并从包含人体动作的视频中提取出人体在进行对应动作时各个椭球的运动数据；其次利用简化的射线追踪方法，针对动作中基元模型的每一帧计算对应的信道响应；最后对信道响应进行多普勒分析，获得对应动作的微多普勒时频谱。上述仿真获得的微多普勒时频谱数据集可以用于训练无线动作识别的深度神经网络。该文针对“步行”“跑步”“跌倒”“坐下”等4种常见的人体动作在60 GHz频段上进行了信道仿真及动作识别的测试。实验结果表明，通过仿真训练的深度神经网络在实际无线动作识别中平均识别准确率可以达到73.0%。此外，借助无标签迁移学习，通过少量无标签实测数据的微调，上述准确率可以进一步提高到93.75%。

关键词: 无线信道建模；无线动作识别；无标签迁移学习；毫米波；计算机视觉

中图分类号：TN957.52

文献标识码：A

文章编号：2095-283X(2024)x-0001-12

DOI: [10.12000/JR24101](https://doi.org/10.12000/JR24101)

引用格式: 任振裕, 吉辰卿, 余潮, 等. 面向毫米波动作识别的视觉辅助信道仿真技术[J]. 雷达学报(中英文), 待出版. doi: [10.12000/JR24101](https://doi.org/10.12000/JR24101).

Reference format: REN Zhenyu, JI Chenqing, YU Chao, et al. Computer vision-assisted wireless channel simulation for millimeter-wave human motion recognition[J]. *Journal of Radars*, in press. doi: [10.12000/JR24101](https://doi.org/10.12000/JR24101).

Computer Vision-assisted Wireless Channel Simulation for millimeter-wave Human Motion Recognition

REN Zhenyu^① JI Chenqing^① YU Chao^① CHEN Wanli^{*②} WANG Rui^{*①}

^①(Department of Electronic and Electrical Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China)

^②(Shenzhen Technology University, Shenzhen 518118, China)

Abstract: This study proposes a computer vision-assisted millimeter wave wireless channel simulation method incorporating the scattering characteristics of human motions. The aim is to rapidly and cost-effectively generate a training dataset for wireless human motion recognition, thereby avoiding the laborious and cost-intensive efforts associated with physical measurements. Specifically, the simulation process includes the following steps. First, the human body is modeled as 35 interconnected ellipsoids using a primitive-based model,

收稿日期：2024-05-25；改回日期：2024-08-26；网络出版：2024-xx-xx

*通信作者：陈万里 chenwanli@sztu.edu.cn；王锐 wang.r@sustech.edu.cn

*Corresponding Authors: CHEN Wanli, chenwanli@sztu.edu.cn; WANG Rui, wang.r@sustech.edu.cn

基金项目：国家自然科学基金(62171213)，高水平专项资金(G030230001, G03034K004)

Foundation Items: The National Natural Science Foundation of China (62171213), High Level of Special Funds (G030230001, G03034K004)

责任编辑：陈彦 Corresponding Editor: CHEN Yan

©The Author(s) 2024. This is an open access article under the CC-BY 4.0 License

(<https://creativecommons.org/licenses/by/4.0/>)

and motion data of these ellipsoids are extracted from videos of human motion. A simplified ray tracing method is then used to obtain the channel response for each snapshot of the primitive model during the motion process. Finally, Doppler analysis is performed on the channel responses of the snapshots to obtain the Doppler spectrograms. The Doppler spectrograms obtained from the simulation can be used to train deep neural network for real wireless human motion recognition. This study examines the channel simulation and action recognition results for four common human actions (“walking,” “running,” “falling,” and “sitting down”) in the 60 GHz band. Experimental results indicate that the deep neural network trained with the simulated dataset achieves an average recognition accuracy of 73.0% in real-world wireless motion recognition. Furthermore, the recognition accuracy can be increased to 93.75% via unlabeled transfer learning and fine-tuning with a small amount of actual data.

Key words: Wireless channel simulation; Wireless human motion recognition; Unlabeled transfer learning; Millimeter wave; Computer vision

1 引言

泛在无线感知由于其在非接触性、感知通信融合性、隐私保护等方面得天独厚的优势，已经逐渐成为下一代无线通信系统的一项核心功能^[1]。在这个背景下，无线人体动作识别(Human Motion Recognition, HMR)的研究得到了广泛关注。该技术旨在利用人体运动带来的信号传播特性变化，推测人体的动作或状态。大量的现有工作致力于基于信道状态信息(Channel State Information, CSI)^[2,3]、调频连续波(Frequency Modulated Continuous Wave, FMCW)雷达^[4,5]或被动式感知^[6,7]来实现人体动作识别。在这些研究中，绝大多数的工作需要采集大量的带标签的实测信道数据或接收信号，并训练人体动作识别的人工智能模型。训练数据的采集往往需要消耗大量的人力、物力和时间；实测数据的采集往往很难覆盖到所有的人体动作识别的场景。为了弥补现有工作的不足，本文提出了一种计算机视觉辅助、面向人体动作识别的无线信道仿真方法。该方法旨在利用信道的计算机仿真替代上述训练数据的实测采集。

近年来，学术界和工业界纷纷致力于拓展现有无线信道模型，添加对被感知目标的建模，为面向人体动作识别的无线信道仿真提供了可能性。以IEEE 802.11bf^[8]国际标准为例，该标准提出了一种数据驱动的混合信道模型(Data driven Hybrid Channel model, DAHC)^[9]。即将无线信道划分为目标无关信道和目标相关信道两部分：前者反映传输环境对无线信道的影响；后者反映被感知对象对无线信道的影响。现有的面向通信的无线信道模型可用于对前者进行建模，而人体基元模型^[10]则可用于对后者进行建模。美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)在毫米波WiFi标准IEEE 802.11 ay^[11]基于射线追踪的准确定性信道模型的基础之上，添加了

额外的目标相关射线(Target-related Rays, T-Rays)，并开发出了相应的软件WiGig Tools^[12]。在学术界，有研究者使用类似的信道模型来优化通感一体系统下的通信感知的综合性能^[13]。另有研究者基于FMCW雷达以及预设人体运动模型开发了微多普勒谱仿真器SimHumalator^[14]，旨在解决实际微多普勒数据集不均衡的问题。上述研究工作均基于特定的人体基元模型^[15]实现了对动作导致的微多普勒效应的仿真；然而这些工作并没有解决多样化动作参数的来源问题。为了让深度神经网络能够广泛适用于不同感知对象的动作识别，对深度神经网络进行训练的动作需要具备多样性；而收集大量人体动作的运动参数将花费不小的成本。针对上述动作参数的采集问题，有研究者使用人体网格运动模型，以参数化生成各种人体动作的轨迹，并开发出用于人体跌倒检测的FMCW回波仿真器SIM-FALL^[16]。还有研究者采用时间-频率扩散模型RF-Diffusion^[17]对Wi-Fi及FMCW信号进行建模和生成，模拟产生多样性的动作参数。但是，这种动作参数的生成方法并不是来自实际的人体动作。另有研究人员针对手势识别的应用场景，利用计算机视觉工具采集人体手部21个三维关键点的运动参数，并以此初步完成了各种常见手势的信道仿真^[18]。

本文在文献^[18]所提出基元法信道仿真方法的基础上，进一步研究利用计算机视觉技术和计算机仿真方法进行人体整体动作识别的技术。具体而言，本文首先介绍了对包含完整人体的无线信道进行建模和仿真的方法，包括如下步骤：首先利用基元模型将人体建模为35个相互连接的椭球，并从包含人体动作的视频中提取出人体在进行对应动作时各个椭球的运动数据；其次利用简化的射线追踪方法，针对动作中基元模型的每一帧计算对应的信道响应；最后对信道响应进行多普勒分析，获得对应动作的微多普勒时频谱。由于视频资源的极大丰富，采集

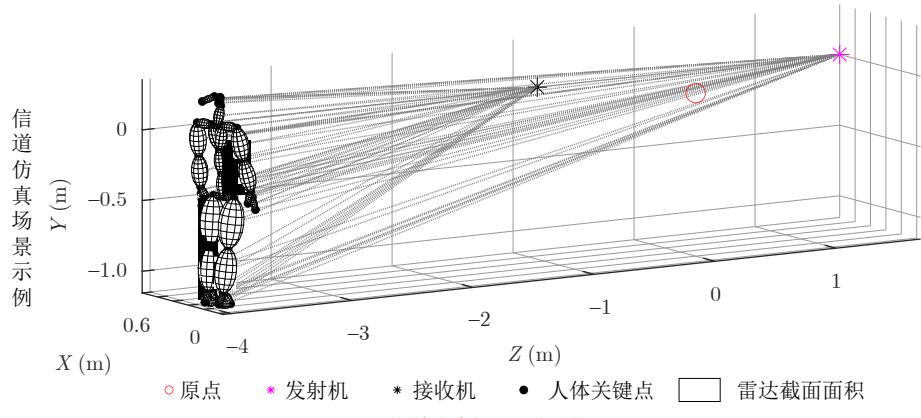
人体运动参数的开销极小。从视频中仿真获得的微多普勒时频谱用于动作识别神经网络的预训练。由于仿真获得的微多普勒时频谱与实测动作的微多普勒时频谱可能存在细节特征上的差异,我们在实际动作识别之前还利用了迁移学习对预训练后的神经网络进行调整。在实验中,我们针对“步行”“跑步”“跌倒”“坐下”等4种常见的人体动作在60 GHz频段上进行了信道仿真及动作识别的测试。实验结果表明,通过仿真数据集预训练的深度神经网络在实测数据集中平均识别准确率可以达到73.0%。此外,借助无标签迁移学习,通过少量实测数据的微调准确率可以进一步提高到93.75%。

2 仿真器框架

本文所提出的无线信道仿真器是基于人体基元模型构建的。其中,信道仿真器可以分为两个部

分:(1)无线信道生成器;(2)人体动作捕捉器。前者利用人体三维关键点的运动数据构建人体基元,计算雷达截面面积模型(Radar Cross Section, RCS);并在给定收发机的位置的情形下,通过射线追踪的方式仿真人基元及静态环境所产生的信道冲激响应。后者则基于现有的计算机视觉领域的人体关键点提取工具,从海量的人体动作视频中提取人体三维关键点运动数据,将其作为无线信道生成器的输入。

如图1(a)所示,在无线信道仿真器中,发射机、接收机以及人体目标可以被设置在一个全局坐标系中任意的位置。一个完整的人体动作可以分解为一系列以 Δt_s 为时间间隔的时间戳上的姿势。在每一个时间戳 t 上,我们假设无线信道冲激响应(Channel Impulse Response, CIR)是准静态的,并通过基元法^[10]进行仿真信道生成。具体来说,我们通过使用 I 个人体三维关键点(即人体关节)和 N 个连



(a) 信道仿真场景示例
(a) Illustration of channel simulation scenario

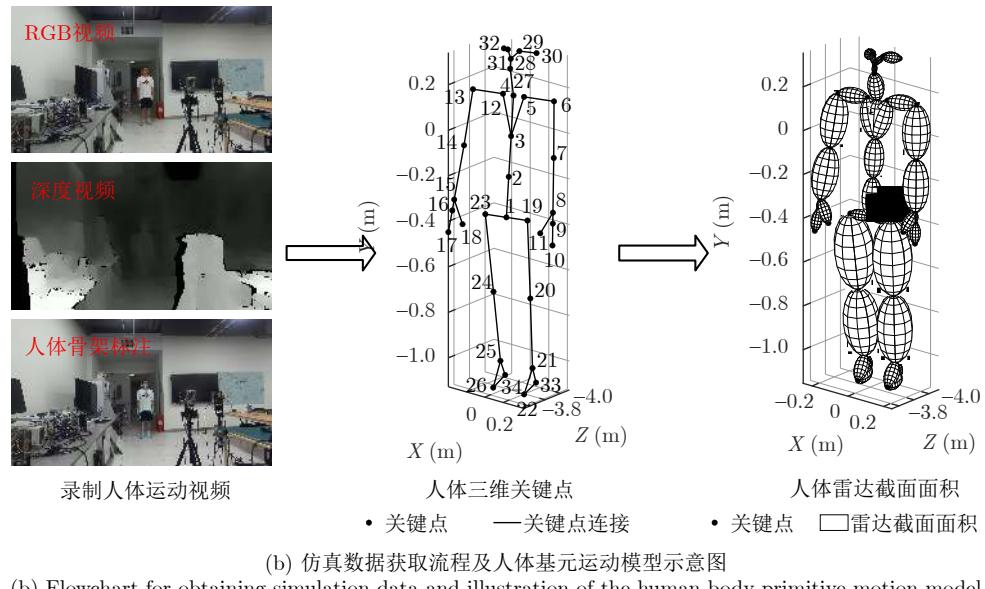


图1 信道仿真场景及具有34个关键点, 35个基元的人体模型示意图

Fig. 1 Illustration of channel simulation scenario and primitive-based human model with 34 keypoints and 35 primitives

接相邻两个关键点的椭球基元来建模人体运动目标，并利用每个椭球中心的散射径来对非视距(Non-Line-of-Sight, NLoS)信道进行建模。在本文中， I 个人体三维关键点的位置坐标被定义为 $\mathbf{p}_i(t)$ ， $i = 1, 2, \dots, I$ 。而 N 个椭球基元的散射中心的位置坐标则被定义为 $\mathbf{p}_n^c(t)$ ， $n = 1, 2, \dots, N$ 。基于此，每个时间戳 t 上的无线信道冲激响应可以被建模为从发射机到接收机的视距(Line-of-Sight, LoS)径，由人体目标散射的 N 条NLoS径，以及由静态环境产生的NLoS径的集合。

本文提出的无线信道仿真方案可以根据任何人体关键点运动数据源以及任意数量的三维关键点和连接椭球基元进行调整。常见的动作捕捉方式可以基于计算机视觉的单目/双目相机、惯性传感器等方法，生成具有不同关键点数量和连接方式的多种人体关键点模型^[19]。在本文的仿真和实验设置中，人体动作捕捉器通过一个ZED 2i深度相机^[20]开发实现。如图1(b)所示，根据ZED 2i软件开发套件的设置， $I = 34$ 和 $N = 35$ ^[20]。本文所采用的34个关键点包含了头部、躯干、四肢等产生微多普勒效应的人体几个核心部位，可以体现人体运动对无线信道的主要影响。该34个关键点由ZED 2i深度相机直接导出，其关键点标注并不是随机选取的，而是通过其内置的AI模型选取^[20]。事实上，基于计算机视觉提取人体骨架运动信息一直是计算机视觉领域的热点话题之一，衣服的褶皱、人体照片的轻微遮挡都不会对关键点的提取产生过大的误差^[21]。另外，由于深度相机录制视频的时间间隔 Δt_v 往往远大于 Δt_s ，在ZED 2i捕捉到的相邻两帧人体动作数据之间，需要进行插值处理，以确保足够的动作采样频率，避免频率混叠现象的发生。

3 无线信道生成器

本节将具体阐述信道生成器生成 t 时刻的信道冲激响应的方法。如图1(a)所示，从发射机到接收机的多径信道可以被划分为两部分：目标相关信道以及目标无关信道。具体来说，分别定义 $h(\tau, t)$ 和 $u(\tau, t)$ 为 t 时刻整体的信道冲激响应和目标相关的信道冲激响应(τ 代表信道冲激的时延)，并定义 $v(\tau)$ 为目标无关的不随时间变化的信道冲激响应。根据文献[8]所定义的信道模型，有

$$h(\tau, t) = u(\tau, t) + v(\tau) \quad (1)$$

其中， $u(\tau, t)$ 和 $v(\tau)$ 的建模分别在3.1节以及3.2节中详细叙述。

3.1 目标相关信道建模

令 \mathbf{p}_t 和 \mathbf{p}_r 分别代表发射机和接收机的三维位置

坐标， $\mathbf{p}_i(t)$ 和 $\mathbf{p}_j(t)$ 分别代表在 t 时刻连接着第 n 个椭球基元的两个关键点坐标($\forall n, t$)。基于此，第 n 个椭球基元的散射中心坐标可以记为 $\mathbf{p}_n^c(t) = [\mathbf{p}_i(t) + \mathbf{p}_j(t)]/2$ 。同时，记连接两个关键点 $\mathbf{p}_i(t)$ 和 $\mathbf{p}_j(t)$ 的椭球长轴的长度为 $2l_n(t)$ ，则

$$l_n(t) = |\mathbf{p}_i(t) - \mathbf{p}_j(t)|/2 \quad (2)$$

椭球基元的剩余两个短轴的长度相等，记为 $2r_n(t)$ 。文中， $r_n(t) = l_n(t)/2$ 。由于人体运动的非刚体特性，椭球基元的尺寸($r_n(t)$ 与 $l_n(t)$)被建模为与 t 相关，可能随着时间的推移轻微变化。

令 $R_t^n(t) = |\mathbf{p}_t - \mathbf{p}_n^c(t)|$ 代表发射机到第 n 个椭球基元散射中心的距离， $R_r^n(t) = |\mathbf{p}_r - \mathbf{p}_n^c(t)|$ 代表接收机到第 n 个椭球基元散射中心的距离， $G_t^n(t)$ 和 $G_r^n(t)$ 分别代表发射天线和接收天线在入射方向 $\mathbf{p}_t - \mathbf{p}_n^c(t)$ 和散射方向 $\mathbf{p}_n^c(t) - \mathbf{p}_r$ 上的天线增益， $\sigma_n(t)$ 代表第 n 个椭球基元在 t 时刻的双基地雷达截面面积， c 代表光速， f_c 和 λ 分别代表信号的载波频率和波长。由第 n 个椭球基元散射所生成的信道冲激响应可以表示为

$$u_n(\tau, t) = \lambda \sqrt{\frac{\sigma_n(t) G_t^n(t) G_r^n(t)}{(4\pi)^3 (R_t^n(t) R_r^n(t))^2}} e^{-j\phi_n(t)} \delta(\tau - \tau_n(t)) \quad (3)$$

其中， $\delta(a)$ 是单位冲激函数，即满足当 $a = 0$ 时， $\delta(a) = 1$ ，反之则 $\delta(a) = 0$ 。 $\tau_n(t) = [R_t^n(t) + R_r^n(t)]/c$ 和 $\phi_n(t) = 2\pi f_c \tau_n(t)$ 分别代表该散射路径的传输时间(时延)和相位偏移。

双基地雷达截面面积 $\sigma_n(t)$ 可以通过文献[22]中的方法计算。如图2所示，令 $\theta_t^n(t)$ 和 $\theta_r^n(t)$ 分别代表入射和散射的俯仰角(Elevation)， $\phi_t^n(t)$ 和 $\phi_r^n(t)$ 分别代表入射和散射的方位角(Azimuth)， $\mathbf{v}_n(t) =$

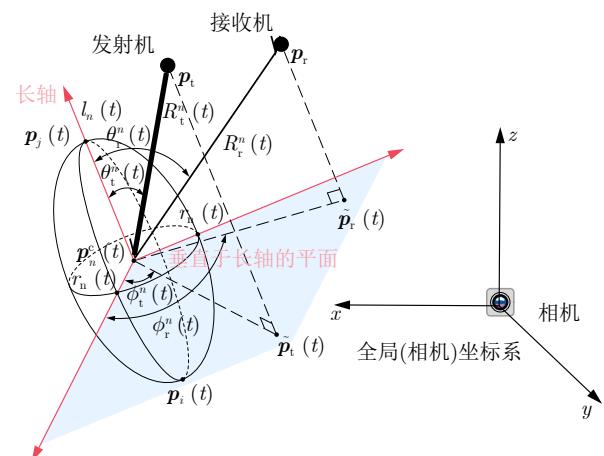


图 2 对第 n 个椭球的双基地雷达截面面积的计算参数示意图

Fig. 2 Illustration of bistatic radar cross section (RCS) calculation parameters for n -th ellipsoid

$[\mathbf{p}_i(t) - \mathbf{p}_j(t)]/(2l_n(t))$ 代表沿着椭球长轴方向的单位向量, 有

$$\theta_t^n(t) = \arccos \left((\mathbf{p}_n^c(t) - \mathbf{p}_t)^T \mathbf{v}_n(t) / R_t^n(t) \right) \quad (4)$$

$$\theta_r^n(t) = \arccos \left((\mathbf{p}_n^c(t) - \mathbf{p}_r)^T \mathbf{v}_n(t) / R_r^n(t) \right) \quad (5)$$

和

$$\begin{aligned} & |\phi_r^n(t) - \phi_t^n(t)| \\ &= \arccos \left(\frac{(\mathbf{p}_n^c(t) - \tilde{\mathbf{p}}_t(t))^T (\mathbf{p}_n^c(t) - \tilde{\mathbf{p}}_r(t))}{|\mathbf{p}_n^c(t) - \tilde{\mathbf{p}}_t(t)| |\mathbf{p}_n^c(t) - \tilde{\mathbf{p}}_r(t)|} \right) \end{aligned} \quad (6)$$

其中, $\tilde{\mathbf{p}}_t(t) = \mathbf{p}_t - \mathbf{v}_n(t)(\mathbf{p}_t - \mathbf{p}_n^c(t))^T \mathbf{v}_n(t)$ 和 $\tilde{\mathbf{p}}_r(t) = \mathbf{p}_r - \mathbf{v}_n(t)(\mathbf{p}_r - \mathbf{p}_n^c(t))^T \mathbf{v}_n(t)$ 分别代表发射机和接收机在 t 时刻于垂直于椭球长轴 $\mathbf{v}_n(t)$ 且包含椭球基元散射中心 $\mathbf{p}_n^c(t)$ 的平面上的投影。基于此, 在 t 时刻第 n 个椭球基元的双基地雷达截面面积可以表示为

$$\sigma_n(t) = \frac{4\pi\alpha(t)r_n^4(t)l_n^2(t)}{[\beta(t)r_n^2(t) + \gamma(t)l_n^2(t)]^2} \quad (7)$$

其中,

$$\alpha(t) = [(1 + \cos\theta_t^n(t)\cos\theta_r^n(t))\cos(\phi_r^n(t) - \phi_t^n(t)) + \sin\theta_t^n(t)\sin\theta_r^n(t)]^2$$

$$\begin{aligned} \beta(t) &= \sin^2\theta_t^n(t) + \sin^2\theta_r^n(t) + 2\sin\theta_t^n(t) \\ &\cdot \sin\theta_r^n(t)\cos(\phi_r^n(t) - \phi_t^n(t)) \end{aligned}$$

$$\gamma(t) = [\cos\theta_t^n(t) + \cos\theta_r^n(t)]^2$$

将所有的由椭球基元散射产生的NLoS信道冲激响应相加, 目标相关信道冲激响应可以表示为

$$u(\tau, t) = \sum_{n=1}^N u_n(\tau, t) \quad (8)$$

3.2 目标无关信道建模

如前文所述, 目标无关信道可以借助现有的无线信道模型。本文将室内环境建模为 K 个静止的散射点。令第 k 个散射点的双基地雷达截面面积, 发射机天线增益, 接收机天线增益, 到发射机的距离, 到接收机的距离分别表示为 σ_k , G_t^k , G_r^k , R_t^k , R_r^k ; 目标无关信道的NLoS信道冲激响应可以表示为

$$v_{\text{NLoS}}(\tau) = \sum_{k=1}^K \lambda \sqrt{\frac{\sigma_k G_t^k G_r^k}{(4\pi)^3 (R_t^k R_r^k)^2}} e^{-j\phi_k} \delta(\tau - \tau_k) \quad (9)$$

其中, $\tau_k = (R_t^k + R_r^k)/c$ 和 $\phi_k = 2\pi f_c \tau_k$ 分别代表第 k 条NLoS径的传输时延及相位。

令沿着LoS径方向发射和接收的天线增益分别为 $G_{t, \text{LoS}}$ 和 $G_{r, \text{LoS}}$, 发射机与接收机之间的距离为 R_{LoS} 。基于此, 目标无关信道的LoS信道冲激响应可以通过自由空间传输模型表示为

$$v_{\text{LoS}}(\tau) = \frac{\lambda \sqrt{G_{t, \text{LoS}} G_{r, \text{LoS}}}}{4\pi R_{\text{LoS}}} e^{-j\phi_{\text{LoS}}} \delta(\tau - \tau_{\text{LoS}}) \quad (10)$$

其中, $\tau_{\text{LoS}} = R_{\text{LoS}}/c$ 和 $\phi_{\text{LoS}} = 2\pi f_c \tau_{\text{LoS}}$ 分别表示LoS径的传输时延及相位。根据文献[8]所定义的信道模型, 则有

$$v(\tau) = v_{\text{LoS}}(\tau) + v_{\text{NLoS}}(\tau) \quad (11)$$

4 人体动作捕捉器

我们可以把一个完整的人体动作表示成为一系列等间距(Δt_s)时间戳上人体三维关键点的位置。人体动作捕捉器利用ZED 2i深度相机[20]获取上述关键点位置。ZED 2i深度相机录制视频的帧率一般处于每秒30~120帧。该相机内置工具或其他的计算机视觉处理软件可以直接识别并提取每个视频帧中各个人体三维关键点的位置。然而, 上述帧率不足以用于信道的微多普勒仿真。例如, 在60 GHz频段上, 4 m/s 的人体运动最大速率产生的多普勒频率最大可以达到800 Hz。这意味着每秒至少需要仿真1600个等间隔时间戳上的信道(即帧率大于1600), 大大高于ZED 2i深度相机所能够提供的帧率。

为解决上述问题, 本文在相邻的两个视频帧之间插入 $\Delta t_v/\Delta t_s - 1$ 个时间戳(Δt_v 代表视频帧的时间间隔), 并利用3次样条插值方法[23]计算新插入时间戳上各个人体三维关键点的位置。以此为基础, 可以依据上文介绍的方法计算每个时间戳上的信道冲激响应。

5 “仿真到真实”的无监督迁移推理

基于上文介绍的信道仿真器和人体动作捕捉器, 我们可以构建一个人体动作微多普勒谱的仿真数据集。该数据集可用于训练人体动作识别的机器学习模型, 并在对实际测量的人体动作微多普勒谱进行识别, 即实现“仿真到真实”的推理(Simulation-to-reality inference)。然而, 由于仿真和实测的微多普勒谱在图像特征上有一定的差异, 直接进行“仿真到真实”的推理可能导致较大的误判概率。迁移学习工具可以用来解决这一类问题[24]。本节将提出一种无监督迁移学习方法, 以提高“仿真到真实”推理的准确率。

记带有动作标签的仿真微多普勒谱的集合为源域。以 \mathcal{X}_s 和 \mathcal{Y}_s 分别代表源域中仿真微多普勒谱的集合与其一一对应的标签集合, 以 N_c 代表人体动作的类别数。记少量无标签实测微多普勒谱的集合 \mathcal{X}_t 为目标域。如图3所示, 本文所提出的无监督迁移学习可以分为3个阶段:

(1) 基于残差神经网络(即ResNet18)^[25]在源域上预训练人体动作识别模型。ResNet18可以划分为特征提取器 $M_s(\cdot)$ 及分类器 $C(\cdot)$ 两部分。特征提取器 M_s 包含ResNet18的17个卷积层, 其输入为微多普勒谱 $\mathbf{x}_s \in \mathcal{X}_s$, 输出为512维特征向量。分类器 C 包含ResNet18的全连接层部分, 其输入则为512维特征向量, 输出为预测的标量标签($n = 1, 2, \dots, N_c$)。ResNet18预训练部分可总结为优化以下交叉熵损失函数:

$$\min_{M_s, C} \mathcal{L}_{\text{cls}}(\mathcal{X}_s, \mathcal{Y}_s) = -\mathbb{E}_{(\mathbf{x}_s, y_s) \in (\mathcal{X}_s, \mathcal{Y}_s)} \sum_{n=1}^{N_c} 1_{[n=y_s]} \cdot \log C(M_s(\mathbf{x}_s)), \quad (12)$$

其中, \mathbb{E} 代表数值期望操作, $1_{[n=y_s]}$ 是一个指示函数(在 $n = y_s$ 时值为1, 反之为0)。

(2) 使用无标签实测数据 \mathcal{X}_t 结合对抗性判别领域自适应(Adversarial Discriminative Domain Adaptation, ADDA)方法^[26]对第一阶段的预训练ResNet18模型进行微调。在这里, 一个目标域特征提取器 M_t 被初始化为与 M_s 相同的特征, 并配合无标签实测数据的输入来产生目标域特征向量。在每一轮的训练中, 目标域特征向量与源域特征向量被一同输入一个由多层全连接层组成的判别器(discriminator)进行训练, 用以区分输入特征向量的域标签(源域为1, 目标域为0)。接着, 目标域特征提取器 M_t 不断微调其参数, 以混淆判别器, 将其判断为源域特征。最终, M_t 的输出可模仿源域特征的分布, 从而适配源域分类器 C 。上述过程可总结为优化以下两个损失函数:

$$\begin{aligned} \min_D \mathcal{L}_{\text{adv}_D}(\mathcal{X}_s, \mathcal{X}_t, M_s, M_t) \\ = -\mathbb{E}_{\mathbf{x}_s \in \mathcal{X}_s} [\log D(M_s(\mathbf{x}_s))] \\ - \mathbb{E}_{\mathbf{x}_t \in \mathcal{X}_t} [\log(1 - D(M_t(\mathbf{x}_t)))] \end{aligned} \quad (13)$$

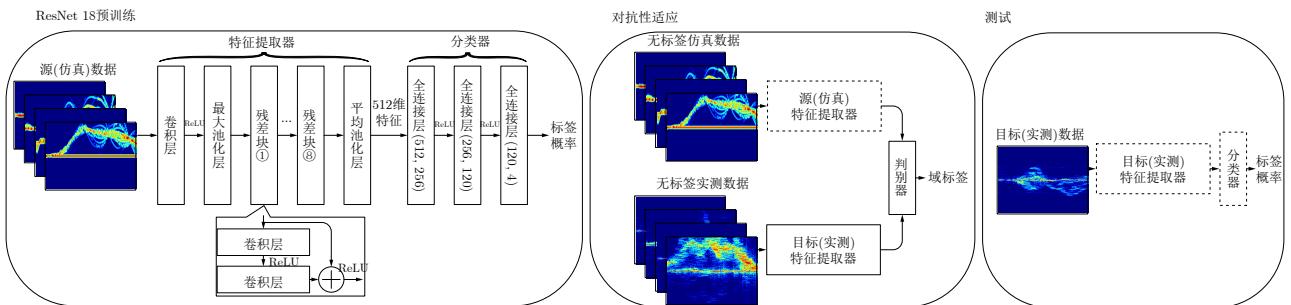


图3 无标签迁移学习框架(虚线框格代表着训练或测试阶段神经网络的参数保持不变, 而实线框格表示神经网络的参数随着训练的过程不断更新)

Fig. 3 An overview of unsupervised transfer learning (Dashed boxes represent the neural network parameters that remain unchanged during training or testing phases, while solid boxes indicate neural network parameters that are continuously updated throughout the training process)

$$\begin{aligned} & \min_M \mathcal{L}_{\text{adv}_M}(\mathcal{X}_s, \mathcal{X}_t, D) \\ & = -\mathbb{E}_{\mathbf{x}_t \in \mathcal{X}_t} [\log D(M_t(\mathbf{x}_t))] \end{aligned} \quad (14)$$

(3) 将目标域特征提取器 M_t 与源域分类器 C 结合为新的ResNet18模型, 用于实测人体微多普勒谱的动作识别。

6 人体动作识别实验验证

为了验证本文提出的基于仿真训练的人体动作识别技术, 本节首先分别介绍仿真与实测人体动作微多普勒数据集的构建。后者用于动作识别的测试和对比方案的性能验证。其次, 本文定性及定量论证了所构建微多普勒数据集的高保真度。最后, 在仿真与实测数据集上验证了前文提出的“仿真到真实”推理的准确率。

6.1 仿真与实测数据集

本文所构建的实测与仿真数据集分别包含由两名不同体型的志愿者所完成的4种人体动作, 每种人体动作则分别包含100个样本。4种动作分别为:(1)步行靠近或远离收发机链路; (2)跑步靠近或远离收发机链路; (3)跌倒; (4)坐下。

如图4所示, 为了采集实测微多普勒谱(时间-频率谱)的数据集, 本文构建了一个工作在60 GHz毫米波频段的被动感知与通信系统^[7]。在发射机端, 本文使用软件无线电平台NI USRP-2954R^[27]产生500 MHz的中频信号, 随后通过毫米波相控阵Sivers将其上变频到60 GHz进行传输。在接收机端, 我们使用两个Sivers相控阵^[28]连接单个USRP设备, 分别用于接收参考道及监视信道的信号, 并通过互相关模糊函数(Cross-Ambiguity Function, CAF)^[7]计算人体动作微多普勒谱。传输的毫米波信号通过正交频分复用(Orthogonal Frequency-Division Multiplexing, OFDM)的方式进行调制, 其载波频率为60 GHz, 带宽为2 MHz。

在以ZED 2i深度相机为原点的相机坐标系之中, 发射机被放置在了[0.8 m 0 m 1.7 m]的位置, 接收机则被放置在[0.8 m 0 m -0.8 m]的位置, 而人体则在距离接收机1~4 m的范围之内的不同位

置随机活动。实测人体微多普勒谱的例子如图5所示, 其中微多普勒谱的频率范围为±800 Hz, 持续时间为2 s。

另外, 本文使用ZED 2i相机对每种动作分别

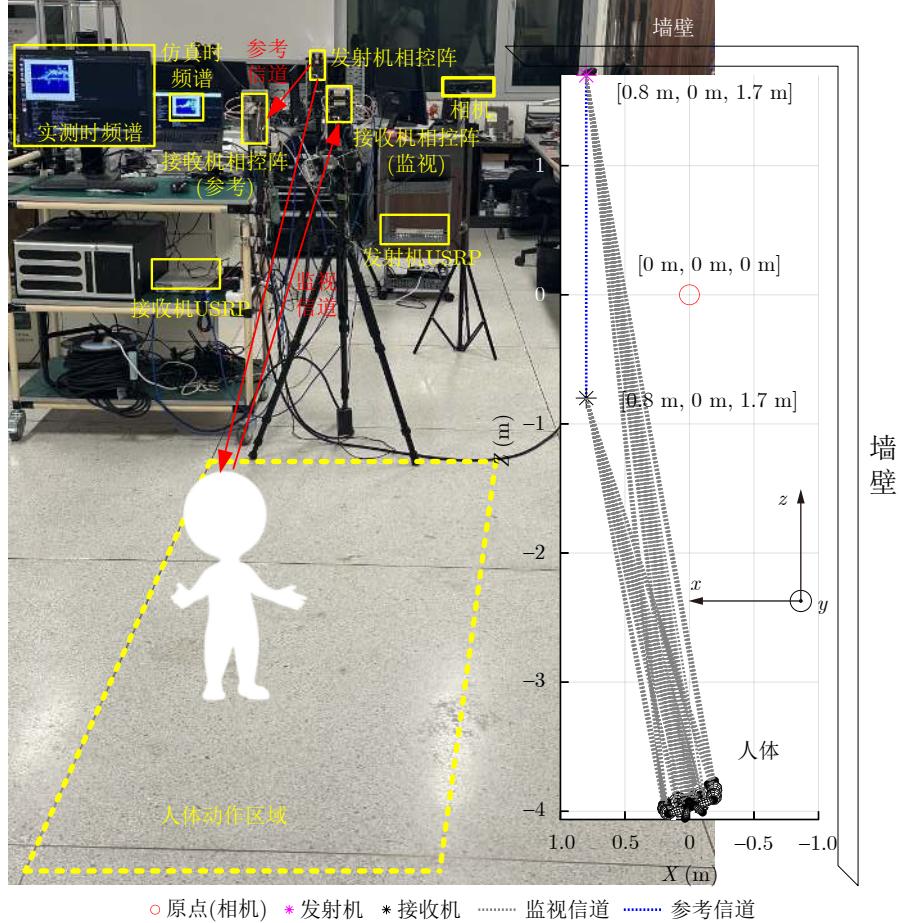


图4 实验设备及实验场景示意图

Fig. 4 Illustration of facilities and scenario of experiment

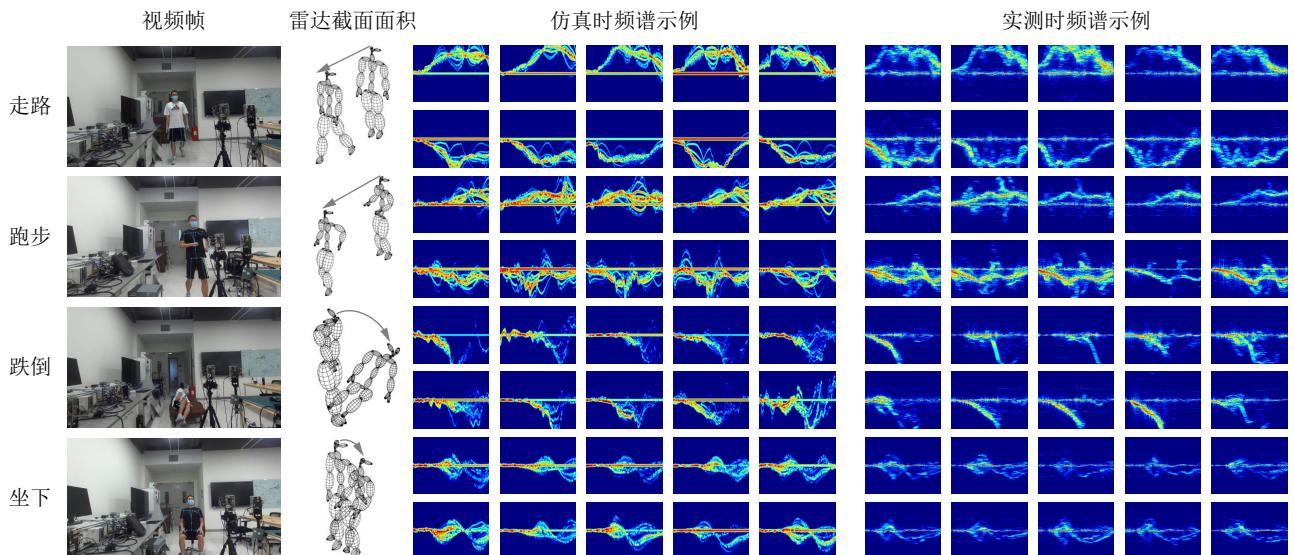


图5 仿真与实测数据集微多普勒谱示意图

Fig. 5 Illustration of the simulated and experimental spectrogram datasets

以30帧/秒的帧率录制了100组长度为2 s的视频。导出人体三维关键点并进一步插值到2000帧/秒后，可以得到100组该动作对应的信道冲激响应。基于此，仿真微多普勒谱可以通过短时傅里叶变换(Short Time Fourier Transform, STFT)计算获得。在本文的仿真中，STFT窗口长度为0.125 s(约250个连续的信道冲激响应)。为了模拟目标无关的信道，本文随机生成了 $K = 10$ 个静止的散射点，其雷达截面面积服从均值为 0.005 m^2 ，方差为 0.001 m^2 的正态分布，而其位置则随机分布在以接收机为中心的 $5 \text{ m} \times 5 \text{ m} \times 5 \text{ m}$ 的立方体之内。这些散射体的位置可以用来计算式(9)中的 G_t^k , G_r^k , R_t^k , R_r^k 等参数。另外，为了准确模拟目标无关信道，仿真考虑了与实测相同的收发天线增益设置。根据Sivers相控阵的使用手册，收发天线的增益被估计为在方位角(Azimuth)上保持 $\pm 60^\circ$ 的准全向增益，在俯仰角(Elevation)上保持 $\pm 22.5^\circ$ 的准全向增益。这意味着仿真器同样考虑了不同时刻某些人体基元并未被覆盖或完全覆盖在收发天线波束之下的情形，即式(3)中的发射与接收天线增益 $G_t^n(t)$ 和 $G_r^n(t)$ 随着时间 t 及不同基元 n 变化。最终获得的仿真微多普勒谱数据集如图5所示。

6.2 仿真微多普勒数据集保真度分析

图6对比了本文所考虑的4种人体动作的仿真与实测微多普勒谱。可以看出，仿真和实测的人

体运动微多普勒谱在总体幅度上基本一致，但在小尺度特征形状和背景噪声上存在一定差异。造成这些差异的原因有3点：(1)通过椭球雷达截面面积来模拟人体基元是一种对复杂人体散射模型的简化表达，这种简化能维持人体关键点运动带来的核心多普勒特征，但忽略了人体衣物褶皱、肌肉收缩和呼吸等生命体征的影响。(2)信道仿真难以精确模拟室内反射、散射、衍射等电磁波传播现象带来的多径效应和小尺度衰落。(3)另外，由于收发机硬件设备不完美导致的物理噪声也同样难以精确模拟。

为了量化仿真与实测时频谱的差异及进一步刻画本文所提出的信道仿真方法的高保真度，我们使用结构相似性指数(Structural Similarity Index Measure, SSIM)^[29]来描述任意两张仿真与实测微多普勒谱的相似性。SSIM的取值范围是0~1，其中1表示两幅图像完全相同，0表示没有相似性。值越接近1，表示图像相似度越高。结构相似性指数SSIM的计算公式在文献[29]中详细叙述，在此不再赘述。为了保证SSIM验证的一般性，对于“步行”“跑步”“跌倒”“坐下”这4种人体动作，我们将100张仿真数据集中的微多普勒时频谱与100张实测数据集中的微多普勒时频谱一一对应并计算SSIM，最终得到每个动作SSIM的累积分布函数(Cumulative Distribution Function, CDF)曲线。如图7所示，约90%的步行动作样本组合

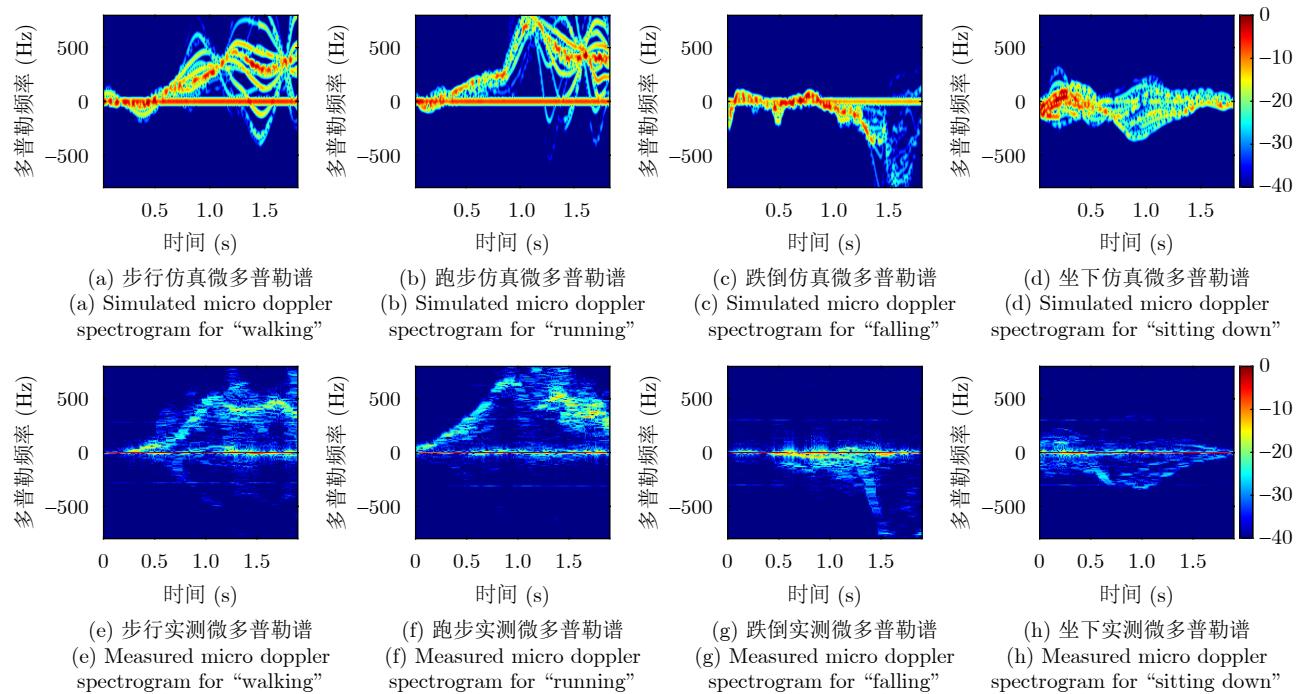


图 6 4个人体动作的仿真与实测微多普勒谱对比

Fig. 6 Spectrogram comparison of four human motions generated by simulation and experiment

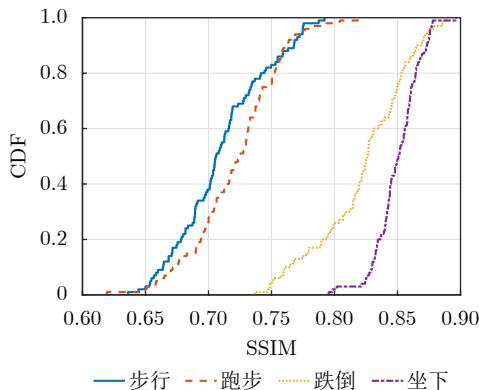


图 7 不同人体动作仿真与实测样本间SSIM值的CDF曲线

Fig. 7 The CDF curves of SSIM values between simulated and measured samples for different human motions

SSIM大于0.6638；约90%的跑步动作样本SSIM大于0.6716；约90%的跌倒样本组合SSIM大于0.7602；约90%的坐下动作样本SSIM大于0.8296。结果显示，本文提出的信道仿真方法对4种不同人体动作的保真度有所不同，其中原地动作“跌倒”和“坐下”具有较高的保真度，而位移动作“步行”和“跑步”的保真度相对较低。与文献RF-Diffusion^[17]提出的基于生成式人工智能模型的微多普勒时频谱仿真方法相比，本文所提出仿真方法的SSIM处于相近的较高水平。然而，生成式人工智能可能会生成约5%的SSIM值在0~0.6的微多普勒时频谱，而本文采用的仿真方法则将SSIM值集中在0.6~0.9，从而确保了信道仿真在保真度下界上处于较高的水平。

6.3 人体动作识别

6.2节论证了仿真微多普勒数据集的高保真度，以及基于仿真数据集训练人体动作识别模型并实现“仿真到真实”推理的可行性。为了进一步验证本文所提出的仿真方法的高保真性以及“仿真到真实”推理效果，本文测试了以下4个人体动作识别的训练和测试方案：

方案1：仿真数据集训练，实测数据集测试。训练集中每个动作60个仿真微多普勒谱，测试集中每个动作40个实测微多普勒谱。

方案2：仿真数据集训练+迁移学习，实测数据集测试。以方案一中的模型为源域预训练模型，利用40个额外的无标签实测微多普勒谱进行微调，再次对方案1中的测试集进行测试。

方案3：实测数据集域内训练及测试。训练集中每个动作60个实测微多普勒谱，测试集中每个动作40个实测微多普勒谱。

方案4：仿真数据集域内训练及测试。训练集

中每个动作60个仿真微多普勒谱，测试集中每个动作40个仿真微多普勒谱。

4个人体动作识别方案的准确率分别为73%，93.75%，100%，100%。它们的识别混淆矩阵如图8所示。可以看出，对于方案3和方案4(仿真数据集或实测数据集域内训练及测试)而言，经过ResNet18训练之后的人体动作识别准确率可达到100%。这说明4种人体动作可以通过微多普勒谱得到有效的区分。另外，方案1(仿真数据集训练，实测数据集测试)的人体动作识别准确率为73%，其中步行与坐下，跑步与跌倒存在着40%与17.5%的较大混淆概率。仿真与实测微多普勒谱的固有特征差异，可能是造成这种动作模糊的原因之一。最后，方案2的人体动作识别准确率可达到93.75%，且各个动作之间的混淆较之方案1都有不同程度的减小。这体现了迁移学习在本文所提出的方案中的必要性。

另一个导致动作之间混淆的原因可能是微多普勒谱观察时间窗口的长度过短。本文现有的仿真及实测数据集的动作时间长度仅有2 s，这导致部分动作在一个较短的时间窗口内观察时显得过于相似。如图9所示，延长时间窗口长度至5 s可更有效区分长期动作(如步行和跑步)与瞬时动作(如跌倒和坐下)。另外，延长时间窗口后的“步行”“跑步”“跌倒”“坐下”4组动作的SSIM分别为0.7408, 0.8000, 0.8972, 0.8511；相较于图7 SSIM平均值(0.7104, 0.7215, 0.8226, 0.8501)，显示出更高的保真度。这表明延长时间窗口有望减少人体动作识别中的混淆。此外，设计一个对微多普勒时间窗口敏感的动作识别神经网络架构有望进一步提升人体动作识别的效果。

7 结语

本文提出了一种利用计算机视觉技术辅助实现包含运动人体散射特征的毫米波无线信道仿真的方法，以及一种无监督迁移学习方案。首先，利用录制的人体运动视频捕捉三维关键点，计算人体动作微多普勒谱的仿真数据集。其次，使用毫米波被动感知平台采集人体动作微多普勒谱的实测数据集。通过使用大量带标签的仿真数据进行动作识别模型的预训练，再用少量无标签的实测数据进行微调，获得最终动作识别模型。实验表明，预训练的人体动作识别模型在“仿真到真实”推理中的准确率为73%，而经过无监督迁移学习微调后，准确率提升至93.75%。

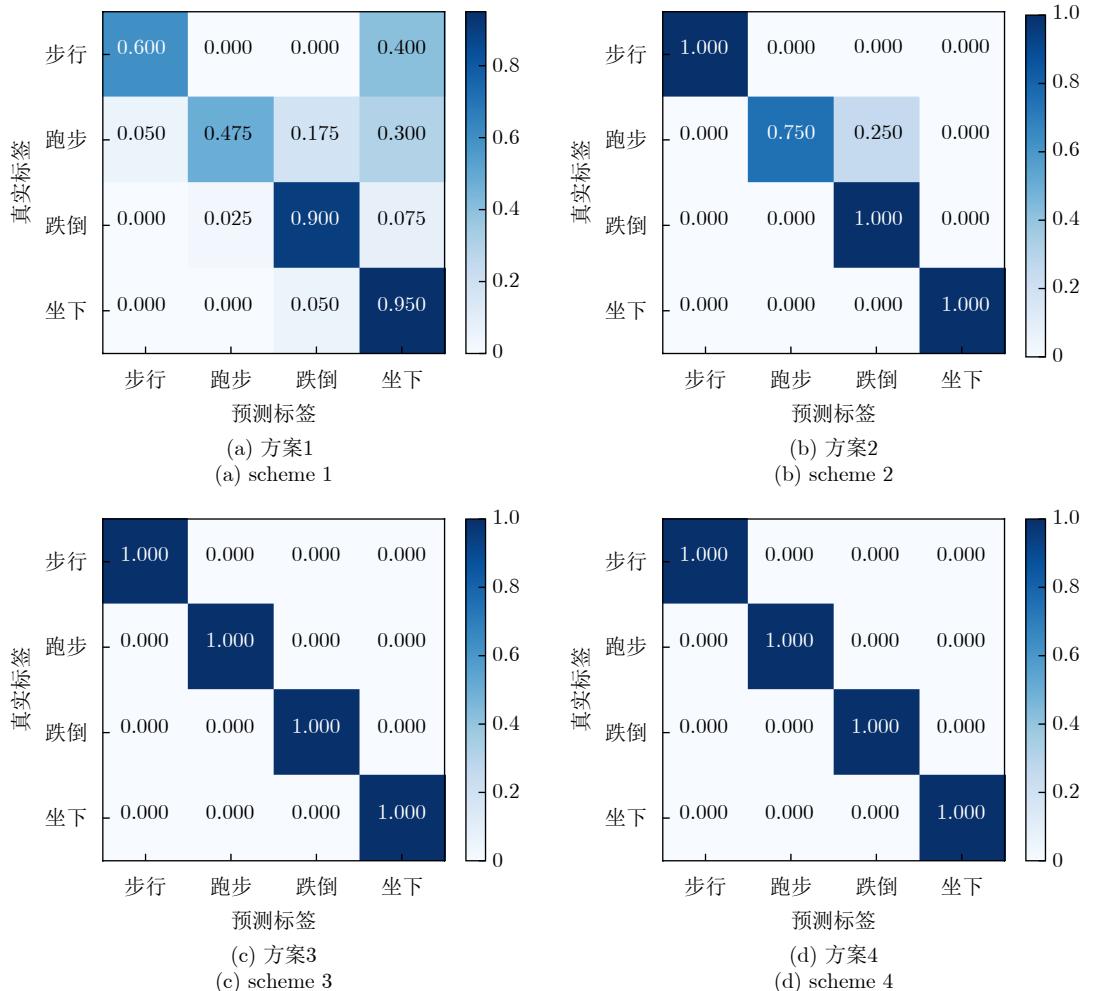


图 8 人体动作识别结果

Fig. 8 Human motion recognition result

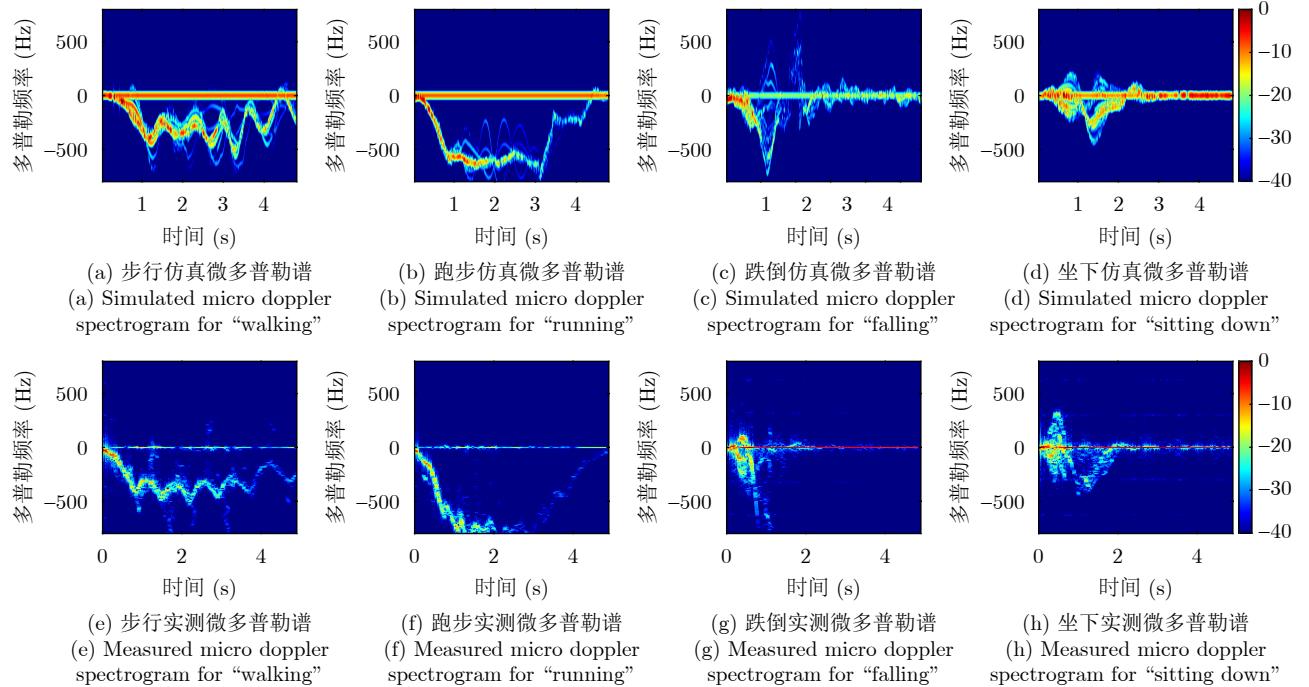


图 9 4个人体动作的仿真与实测微多普勒谱对比(样本时长5 s)

Fig. 9 Spectrogram comparison of four human motions generated by simulation (experiment for 5 seconds)

利益冲突 所有作者均声明不存在利益冲突

Conflict of Interests The authors declare that there is no conflict of interests

参 考 文 献

- [1] LI Chenning, CAO Zhichao, and LIU Yunhao. Deep AI enabled ubiquitous wireless sensing: A survey[J]. *ACM Computing Surveys (CSUR)*, 2022, 54(2): 32. doi: [10.1145/3436729](https://doi.org/10.1145/3436729).
- [2] LIU Jian, LIU Hongbo, CHEN Yingying, et al. Wireless sensing for human activity: A survey[J]. *IEEE Communications Surveys & Tutorials*, 2020, 22(3): 1629–1645. doi: [10.1109/COMST.2019.2934489](https://doi.org/10.1109/COMST.2019.2934489).
- [3] ZHANG Yi, ZHENG Yue, QIAN Kun, et al. Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 8671–8688. doi: [10.1109/TPAMI.2021.3105387](https://doi.org/10.1109/TPAMI.2021.3105387).
- [4] 元志安, 周笑宇, 刘心溥, 等. 基于RDSNet的毫米波雷达人体跌倒检测方法[J]. 雷达学报, 2021, 10(4): 656–664. doi: [10.12000/JR21015](https://doi.org/10.12000/JR21015).
YUAN Zhian, ZHOU Xiaoyu, LIU Xinpui, et al. Human fall detection method using millimeter-wave radar based on RDSNet[J]. *Journal of Radars*, 2021, 10(4): 656–664. doi: [10.12000/JR21015](https://doi.org/10.12000/JR21015).
- [5] 金添, 何元, 李新羽, 等. 超宽带雷达人体行为感知研究进展[J]. 电子与信息学报, 2022, 44(4): 1147–1155. doi: [10.11999/JEIT211044](https://doi.org/10.11999/JEIT211044).
JIN Tian, HE Yuan, LI Xinyu, et al. Advances in human activity sensing using ultra-wide band radar[J]. *Journal of Electronics & Information Technology*, 2022, 44(4): 1147–1155. doi: [10.11999/JEIT211044](https://doi.org/10.11999/JEIT211044).
- [6] SUN Hongbo, CHIA L G, and RAZUL S G. Through-wall human sensing with WiFi passive radar[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(4): 2135–2148. doi: [10.1109/TAES.2021.3069767](https://doi.org/10.1109/TAES.2021.3069767).
- [7] LI Jie, YU Chao, LUO Yan, et al. Passive motion detection via mmWave communication system[C]. IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), Helsinki, Finland, 2022: 1–6. doi: [10.1109/VTC2022-Spring54318.2022.9860809](https://doi.org/10.1109/VTC2022-Spring54318.2022.9860809).
- [8] DU Rui, HUA Haocheng, XIE Hailiang, et al. An overview on IEEE 802.11bf: WLAN sensing[J]. *IEEE Communications Surveys & Tutorials*, 2024. doi: [10.1109/COMST.2024.3408899](https://doi.org/10.1109/COMST.2024.3408899).
- [9] ZHANG Meihong, DU Rui, PENG Xiaohui, et al. Channel models for WLAN sensing systems[R]. IEEE 802.11-21/0782r5, 2021.
- [10] LI Guoliang, WANG Shuai, LI Jie, et al. Wireless sensing with deep spectrogram network and primitive based autoregressive hybrid channel model[C]. IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Lucca, Italy, 2021: 481–485. doi: [10.1109/SPAWC51858.2021.9593198](https://doi.org/10.1109/SPAWC51858.2021.9593198).
- [11] ZHOU Pei, CHENG Kaijun, HAN Xiao, et al. IEEE 802.11ay-based mmWave WLANs: Design challenges and solutions[J]. *IEEE Communications Surveys & Tutorials*, 2018, 20(3): 1654–1681. doi: [10.1109/COMST.2018.2816920](https://doi.org/10.1109/COMST.2018.2816920).
- [12] ASSASA Hany, et al. A collection of open-source tools to simulate IEEE 802.11ad/ay WLAN networks in network simulator ns-3[EB/OL]. <https://github.com/wigig-tools>.
- [13] LI Guoliang, WANG Shuai, LI Jie, et al. Integrated sensing and communication from learning perspective: An SDP3 approach[J]. *IEEE Internet of Things Journal*, 2024, 11(4): 5589–5603. doi: [10.1109/JIOT.2023.3309837](https://doi.org/10.1109/JIOT.2023.3309837).
- [14] VISHWAKARMA S, LI Wenda, TANG Chong, et al. SimHumalator: An open-source end-to-end radar simulator for human activity recognition[J]. *IEEE Aerospace and Electronic Systems Magazine*, 2022, 37(3): 6–22. doi: [10.1109/MAES.2021.3138948](https://doi.org/10.1109/MAES.2021.3138948).
- [15] BOULIC R, THALMANN N M, and THALMANN D. A global human walking model with real-time kinematic personification[J]. *The Visual Computer*, 1990, 6(6): 344–358. doi: [10.1007/BF01901021](https://doi.org/10.1007/BF01901021).
- [16] LI Jiamu, ZHANG Dongheng, CHEN Qi, et al. SIMFALL: A data generator for RF-based fall detection[C]. ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, 2024: 8165–8169. doi: [10.1109/ICASSP48485.2024.10446234](https://doi.org/10.1109/ICASSP48485.2024.10446234).
- [17] CHI Guoxuan, YANG Zheng, WU Chenshu, et al. RF-diffusion: Radio signal generation via time-frequency diffusion[C]. The 30th Annual International Conference on Mobile Computing and Networking, Washington, USA, 2024: 77–92. doi: [10.1145/3636534.3649348](https://doi.org/10.1145/3636534.3649348).
- [18] REN Zhenyu, LI Guoliang, JI Chenqing, et al. CASTER: A computer-vision-assisted wireless channel simulator for gesture recognition[J]. *IEEE Open Journal of the Communications Society*, 2024, 5: 3185–3195. doi: [10.1109/OJCOMS.2024.3398016](https://doi.org/10.1109/OJCOMS.2024.3398016).
- [19] WANG Jinbao, TAN Shujie, ZHEN Xiaotong, et al. Deep 3D human pose estimation: A review[J]. *Computer Vision and Image Understanding*, 2021, 210: 103225. doi: [10.1016/j.cviu.2021.103225](https://doi.org/10.1016/j.cviu.2021.103225).
- [20] Stereolabs. Stereolabs docs: API reference, tutorials, and integration[EB/OL]. <https://www.stereolabs.com/docs>.
- [21] ZHENG Ce, WU Wenhan, CHEN Chen, et al. Deep learning-based human pose estimation: A survey[J]. *ACM Computing Surveys*, 2024, 56(1): 11. doi: [10.1145/3603618](https://doi.org/10.1145/3603618).
- [22] TROTT K D. Stationary phase derivation for RCS of an

- ellipsoid[J]. *IEEE Antennas and Wireless Propagation Letters*, 2007, 6: 240–243. doi: [10.1109/LAWP.2007.891521](https://doi.org/10.1109/LAWP.2007.891521).
- [23] DE BOOR C. A practical guide to splines[M]. New York, USA: Springer, 1978: 40–48.
- [24] ZHUANG Fuzhen, QI Zhiyuan, DUAN Keyu, et al. A comprehensive survey on transfer learning[J]. *Proceedings of the IEEE*, 2021, 109(1): 43–76. doi: [10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555).
- [25] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [26] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 2962–2971. doi: [10.1109/CVPR.2017.316](https://doi.org/10.1109/CVPR.2017.316).
- [27] National Instruments. USRP-2954[EB/OL]. <https://www.ni.com/en-us/shop/model/usrp-2954.html>.
- [28] Sivers IMA. EVK 06002/00[EB/OL]. <https://www.siversima.com/product/evk-06002-00/>.
- [29] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600–612. doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861).

作者简介

任振裕, 硕士生, 主要研究方向为毫米波人体动作识别技术。

吉辰卿, 硕士生, 主要研究方向为毫米波轨迹重构技术。

余 潮, 硕士, 主要研究方向为通信感知一体化、毫米波被动感知。

陈万里, 高级实验师, 主要研究方向为软件无线电通信、毫米波雷达技术、通信感知一体化技术。

王 锐, 博士, 副教授, 主要研究方向为通信系统的随机优化与资源调度、近似马尔可夫决策过程、智能无线感知技术。

(责任编辑: 于青)