增强稀疏点云数据处理以提高隐私感知的人类动作识别

Maimunatu Tunau[1], Vincent Gbouna Zakka[1], and Zhuangzhuang Dai[1]

Department of Applied Artificial Intelligence and Robotics, Aston University, Birmingham, UK

Abstract. 人体动作识别（HAR）在医疗保健、健身跟踪和环境辅助生活技术中起着关键作用。虽然传统的基于视觉的 HAR 系统有效，但它们引发了隐私问题。毫米波雷达传感器提供了一种保护隐私的替代方案，但由于其点云数据的稀疏和噪声性质，提出了挑战。在文献中，三种主要的数据处理方法——基于密度的噪声应用空间聚类（DBSCAN）、匈牙利算法和卡尔曼滤波——已经被广泛用于改善雷达数据的质量和连续性。然而，对这些方法进行全面评估，无论是单独评估还是组合评估，仍然存在不足。本文通过对使用 MiliPoint 数据集的三个方法进行详细性能分析来填补这一空白。我们分别评估每种方法、所有可能的成对组合及所有三种方法的组合，评估识别准确性和计算成本。此外，我们提出针对性方法的改进，旨在提高准确性。我们的结果提供了关于每种方法及其整合的优势和权衡的重要见解，为未来毫米波基于 HAR 系统的研究提供指导。我们的源代码在 [1]

Keywords: Human action recognition · Point cloud data processing · Privacy-aware human action recognition 上公开提供。

在智能环境和普适计算时代，系统能够实时感知并响应人类动作的能力已经成为智能技术发展的基础。从智能家居中的自适应照明和自动安全系统到老年护理的跌倒检测系统，人体动作识别（HAR）在实现响应性、上下文感知的应用中发挥着关键作用。传统上，HAR 依赖于基于摄像机的系统，因为它们具有丰富的视觉信息。然而，这些系统存在一些限制，包括隐私问题、对良好光照条件的依赖以及高计算要求 [?] 。

毫米波（mmWave）雷达作为人体动作识别（HAR）的一个有前途的替代方法出现，提供了保护隐私、照明不变的感知能力。在高频段操作时，毫米波雷达即使在视觉障碍环境中也能实现强大的运动检测 [?] 。当与深度学习模型结合时，这种感知形式在识别人体动作方面取得了相当的成功 [?] 。然而，挑战依然存在——尤其是在处理雷达传感器生成的稀疏、噪声的点云数据时，这可能会妨碍特征提取和准确分类 [?] 。

为了应对这些挑战，最近的研究引入了几种预处理方法，重点关注聚类、跟踪和时间关联。例如，[?] 引入了 mID 管道，结合了 DBSCAN 聚类、匈牙利算法和卡尔曼滤波器，但他们的工作缺乏系统调优，并且在输入稀疏时效果变差。其他研究探讨了对卡尔曼滤波的深度学习增强 [?] ，以及 [?] 强调快速聚类和关联策略。虽然这些方法各自提供了有价值的见解，但它们通常是孤立评估的。这些方法及其组合的统一和系统评估尚未被探索。

[1] https://github.com/Maimunatunau/Human-Action-Recognition-HAR-using-mmWave-Radar 。

www.xueshuxiangzi.com

本文通过对三种广泛使用的雷达点云预处理技术进行全面评估来填补这一空白：DBSCAN、匈牙利算法和卡尔曼滤波。我们的重点是分析每种方法分别执行时的表现、它们成对互动时的表现，以及它们结合使用时的行为——所有这些都以 MiliPoint 数据集为案例研究。我们评估了识别准确性和计算效率，以更好地理解使用每种方法或组合时涉及的权衡。此外，我们提出了对单个技术的实用性增强，以提升在噪声环境中的表现。通过进行这种系统性的评估，我们旨在为设计更有效的基于毫米波雷达的 HAR 系统提供信息，并为选择或结合预处理方法提供指导。HAR 在过去二十年中经过了长足发展，这得益于人工智能、传感器技术和计算能力的进步。早期的 HAR 系统主要依赖于基于视觉的技术，如光流和方向梯度直方图（HOG），这些技术从视频序列中提取运动和形状描述符。尽管这些方法在受控环境中取得了适度的成功，但它们对光照变化、遮挡高度敏感，并引发了对用户隐私的担忧。

机器学习算法的引入，例如支持向量机（SVM），通过实现复杂运动模式的监督分类，为 HAR（人体动作识别）带来了进步 [?] 。然而，这些方法需要大量的特征工程，并且在多类动作识别环境中往往面临可扩展性的问题。最近，深度学习，尤其是卷积神经网络（CNNs）和递归神经网络（RNNs），大大替代了传统模型，通过自动化特征提取和提高在不同数据集上的识别性能 [?,?] 。

尽管有这些改进，基于视觉的系统依然受到对环境光照依赖性和可能的侵入性所限制。结果是，最近的方法越来越多地转向其他传感模式，如惯性传感器、红外系统，最近则是毫米波 (mmWave) 雷达。基于雷达的人体行为识别 (HAR) 具有几个优势，包括光照不变性、低功耗操作和个人隐私保护。然而，使用毫米波雷达进行 HAR 面临独特的挑战，这些挑战与数据稀疏性和噪声有关 [?] 。

毫米波雷达，工作在 30-300 GHz 范围内，通过宽频带和精细的多普勒测量实现高分辨率空间感知。这使其即使在光线条件不佳的情况下，也适合检测细微的人体动作。然而，雷达信号通常包含大量噪声和非目标反射，需要进行数据预处理流程以确保可靠的动作识别。

为减少噪声并提高跟踪效果，最近的方法采用了各种数据处理技术。基于噪声应用的密度聚类（DBSCAN）经常被使用，因为它能识别任意形状的聚类并过滤掉离群值，而不需要预先知道聚类的数量 [?] 。与传统的聚类方法如 K-means 相比，DBSCAN 对噪声和不规则的雷达点云数据更具韧性 [?] 。

为了时间一致性和目标跟踪，匈牙利算法（HA）被广泛采用用于聚类或检测到的质心的帧间最佳关联 [?] 。HA 为指派问题提供了计算效率高的解决方案，并在大规模场景中表现出优于拍卖算法等替代方案的性能 [?] 。其他方法使用卡尔曼滤波（KF）来预测和更新物体状态，有效地平滑轨迹并处理动态场景中的部分观察 [?] 。在涉及近线性人类运动的 HAR 应用中，标准卡尔曼滤波器已被证明既有效又计算轻量 [?] 。

尽管已经应用了这些单独的技术，但对它们的单独性能以及在不同组合中如何表现却很少进行系统的分析。因此，本研究评估了这些方法，评估了它们的识别准确率和计算成本，并进一步提出了每种技术的改进以提高性能。

1 方法论

提出的数据处理流程框架旨在逐步优化原始点云数据，从噪声过滤和分割到聚类、时间关联和轨迹预测，确保每种方法的优点都能为一个稳健的整体系统做出贡献，在 1 中可以看到这一流程。

www.xueshuxiangzi.com

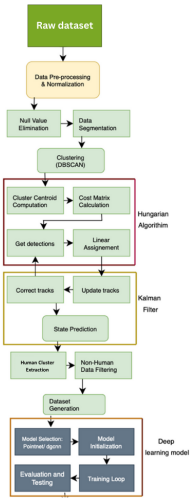

Fig. 1. 所提议的人类动作识别流程的流程图。

1.1 数据集和准备

我们使用了 MiliPoint 数据集 [?]，这是一个利用毫米波雷达传感器记录的结构化点云数据集，用于人体活动识别（HAR）。它包含 545,000 帧数据，收集自 11 位参与者执行 49 种不同动作。每帧由 (x, y, z) 坐标组成。按照 [?] 中的方法，时间数据被堆叠到每帧 1100 个点。数据集被分为训练（80 %）、验证（10 %）和测试（10 %）。

1.2 分割

为了确保稳健的聚类并保持时间连续性，在进行任何聚类或跟踪之前，每个原始帧被细分为多个连续的片段。我们评估了三种分割策略，即每帧 2、5 和 10 个片段，并发现将每帧分为五个相等的片段在噪声去除和数据保留之间提供了最佳平衡。具体而言，每帧的原始雷达数据被划分为五个片段，每个片段包含 220 个连续的 (x, y, z) 点。通过将帧分成较小的片段，我们减少了片段内的方差并提升了后续聚类的一致性。

空值移除 在每个片段中，应用 L2 范数阈值丢弃接近原点（零填充）的点：这一步去除了零填充，防止 DBSCAN 误将填充点解释为有意义的数据。在分割和基于 L2 范数的空值去除之后，我们观察到五段配置始终保留每段用于有效聚类的最佳数量的有效点。平均来说，五段中的每段在阈值处理后保留约 63 个非零点，而两段分割约为 132 个点，十段分割约为 32 个点。两段配置保留了较多的点，但引入了显著的段内变化，负面影响了 DBSCAN 形成紧密、连贯聚类的能力。相反，十段配置过度分裂数据，经常导致段内点数不足以形成有意义的聚类。五段方法在保持时间粒度的同时，维持了足够的点密度，达到了最佳平衡。这种配置减少了来自零填充的噪声，避免了数据稀疏性，导致更稳定的聚类和改进的下游性能。因此，五段分割被选为管道所有后续阶段的配置。

www.xueshuxiangzi.com

我们进行了参数调优以选择 DBSCAN 的最有效参数。参数调优后，选择了以下参数：$\epsilon = 0.4$ 和 min_samples $= 6$ 。此外，我们在欧几里德度量中引入了一个垂直加权因子 $\alpha = 0.25$ 。此调整降低了 z -轴的重要性，使聚类过程与雷达捕捉到的主要水平人类运动对齐。DBSCAN 检测到的聚类提供了细化的空间表示，这不仅过滤掉环境噪声，还作为后续数据关联的基础。

为了构建连续的运动轨迹，我们如下计算聚类的质心：

$$Centroid = \left(\frac{1}{n} \sum_{i=1}^{n} x_i, \frac{1}{n} \sum_{i=1}^{n} y_i, \frac{1}{n} \sum_{i=1}^{n} z_i \right). \tag{1}$$

随后基于相邻段间质心的欧几里得距离形成一个代价矩阵。HA 使用这个矩阵在各段之间优化匹配聚类，确保在帧间进行一致的轨迹分配。这一步很重要，因为准确的段间关联为可靠的轨迹预测准备了数据。

使用卡尔曼滤波的轨迹预测（KF） 通过关联的聚类，然后应用卡尔曼滤波器使用状态向量方程 2 和方程 and observation vector 来预测和修正关联聚类的轨迹。状态向量被定义为

$$X = [\, x,\ y,\ v_x,\ v_y \,]^T, \tag{2}$$

而观测向量为

$$z = [x,\ y]^T. \tag{3}$$

为了优化滤波器的性能，我们通过贝叶斯优化（BO） [?] 在一组样本帧上调整噪声协方差参数 Q 、R ，以及状态协方差 P ，以防止在帧间产生偏差。BO 过程迭代评估多个样本帧的参数设置以最小化预测误差。经过 15 次迭代后，获得了一个最佳配置 $Q = 29.41$ 、$R = 0.081$ 和 $P = 14.64$ ，实现了最小的预测误差 -0.001001 。另外，一个 2.0 的欧几里得距离阈值确保只有紧密匹配的聚类被跟踪，从而减少误关联。这个滤波阶段弥补了剩余的噪声，并提供了平滑的轨迹估计，利用了 HA 形成的可靠关联。

1.3 人工聚类选择和最终评估

在卡尔曼滤波更新后，预测轨迹的质心与人体关键点的真实中位数进行比较。预测准确性通过均方根误差（RMSE）进行定量评估：

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{predicted position}_i - \text{actual position}_i)^2}. \tag{4}$$

如果一个轨迹在 RMSE 和空间接近度测量中占据主导地位，额外的逻辑考虑第二好的轨迹以确保检测到的人类聚类的完整性。最终，未被选中的聚类被归零，只留下最佳表示人类聚类的主要轨迹。不同算法的集成使每种方法能够在前面的步骤中建立：DBSCAN 有效地降低噪声并分割数据，HA 提供跨段的可靠时间关联，KF 通过预测和校正轨迹来优化运动估计。

www.xueshuxiangzi.com

2 结果

在本节中，我们评估了所提数据处理流程的性能，并讨论其在人类活动识别（HAR）中的意义。我们评估了从分割、空值移除到聚类、帧间关联及轨迹预测的流程每个阶段。其次，我们比较了每种方法的性能、它们的成对组合以及所有三种方法的组合，评估动作识别准确性和计算成本。最后，我们将性能与基线模型架构进行比较。

DBSCAN 独立地应用于如第 1.2 节描述的五个分段中的每一个，以识别对应于人类运动的空间聚类。我们的实验表明，一些分段产生了定义明确、噪声较低的聚类，而另一些则显示出较高的变异性。图 **??** 展示了聚类结果，其中不同的颜色表示不同的聚类，黑点表示噪声。

为了在不同分段间建立时间上的连续性，我们计算了连续分段中聚类中心之间的欧几里得距离的成本矩阵。然后 HA 最优地链接了这些聚类，以形成贯穿整个框架的轨迹。我们的结果产生了四条轨迹，其中轨迹 0 和轨迹 3 在多个分段中显示出强劲的一致性，而轨迹 1 和 2 尽管覆盖的分段较少，仍然保持了连贯的运动模式。图 **??** 展示了代表性的成本矩阵，图 **??** 展示了形成的轨迹。

在应用如第 1.2 节描述的卡尔曼滤波器后，所有轨迹稳定，其中轨迹 0 在第 2 次迭代时显示约 1.19 的状态变化，到第 4 次迭代时变化接近于零（约 0.02）。相比之下，轨迹 1 和轨迹 2 经历了较大的转变，表明噪声较高，而轨迹 3 尽管点密度较高，仍快速稳定。图 **??** 和图 **??** 分别展示了状态变化和预测误差。

我们将最终聚类与真实值关键点进行比较，利用均方根误差（RMSE）和中位距离来量化对齐程度。表 **??** 总结了每条轨迹在使用一个样本时的均方根误差和中位距离。轨迹 3 显示了最低的均方根误差（0.7158），而轨迹 0 展示了到关键点最短的中位距离（0.0964 m）。合并轨迹 0 和轨迹 3 结合了精准的预测和与真实值的紧密对齐。

图 2 展示了从原始点云到最终精炼簇的管道进展。结合最佳指标将每个样本从 1100 个点减少到 60 到 300 的范围，保留了运动数据并去除了噪声。

2.1 比较：数据处理流水线与基线模型性能

表 **??** 的结果提供了对使用毫米波雷达数据进行人体动作识别（HAR）的各种数据预处理策略的详细比较，特别关注它们在四种深度学习模型（DGCNN、Pointformer、PointNet++ 和 PointMLP）上的准确性以及相应的计算成本。基线行显示了每个模型在没有任何预处理情况下的性能，作为评估不同处理流程有效性的参考点。

个体方法 在单独的预处理方法中，Kalman 过滤（KM）和 DBSCAN（DS）与基线相比，均显著提高了模型的准确性。特别是 DBSCAN，在所有四种模型上的准确率均超过 98 %，凸显出其在滤除噪声和分离有意义的数据点簇方面的优势。Kalman 过滤也表现出良好的性能，但在 DGCNN 和 Pointformer 上的表现略低于 DBSCAN。相比之下，作为单独方法的匈牙利算法（HG）表现不一致。虽然它在 Pointformer 和 PointNet++ 上有明显改进，但在 DGCNN 上却表现不佳，表明其作为独立方法的效用有限，可能依赖于具体模型。而其相对较高的处理时间也降低了其作为独立预处理步骤的吸引力。当结合两种方法时，最佳的整体性能通过匈牙利算法和 DBSCAN（HG + DS）的组合实现。该组合

www.xueshuxiangzi.com

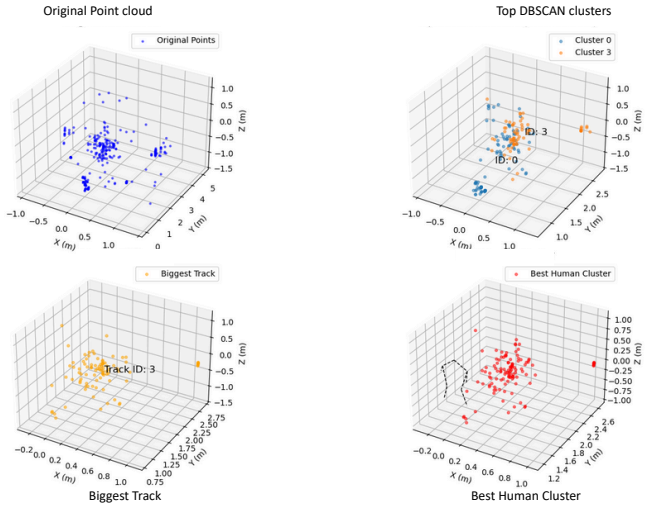

Fig. 2. （左上）原始数据集；（右上）最大的 DBSCAN 簇；（左下）最大的轨迹（轨迹 3）；（右下）最佳人工簇。

在所有模型中保持高准确性，与单独使用 DBSCAN 相比，额外的计算成本非常少。这表明了一种协同效应，其中 DBSCAN 有助于清理数据，而匈牙利算法增强了时间或空间的一致性。有趣的是，结合卡尔曼和匈牙利算法（KM + HG）或 DBSCAN 和卡尔曼（DS + KM）会导致 DGCNN 和 Pointformer 准确性下降—这些模型更依赖于输入的几何结构。这种性能下降可能表明在聚类和跟踪阶段之间的过度平滑或错位可能会干扰模型提取有意义特征的能力。特别是，DGCNN 在基于邻域关系构建动态图时，如果预处理扭曲了这些局部结构，可能会受到影响。

完整的处理流程结合了 DBSCAN、卡尔曼过滤和匈牙利算法（DS + KM + HG），在所有模型中都获得了强劲和平衡的性能。虽然在某些情况下其准确性略微落后于最佳的单一或双方法设置，但它在所有模型上保持了较高的一致性（超过 91 %），并在不同的模型架构中提供了鲁棒性。然而，其计算成本明显高于任何简单的组合（0.2369 秒），这可能限制其在实时或资源限制环境中的应用。

这项研究在使用毫米波雷达数据进行人体活动识别（HAR）的背景下，对三种广泛使用的数据预处理方法——DBSCAN、匈牙利算法（HA）和卡尔曼滤波（KF）进行了全面评估。结果表明，DBSCAN 在从环境噪声中隔离人体运动方面非常有效，显著提高了识别精度。匈牙利算法通过在各帧之间维持准确的聚类关联，增强了时间一致性，而卡尔曼滤波通过其预测建模贡献了更平滑和更可靠的运动跟踪。当结合使用时，这些方法形成了一个强大的预处理管道，大大提高了下游 HAR 模型输入数据的质量。我们的分析显示，结合所有三种方法实现了最高的整体识别精度——使用 PointNet++ 达到了 95.67%，尽管处理时间增加。在实时或资源有限的场景中，仅使用 DBSCAN 或与匈牙利算法结合使用

www.xueshuxiangzi.com

提供了高效和有效的折中方案，以最小的计算成本提供强大的性能。这些发现强调了在 HAR 管道中慎重选择和结合方法的重要性。它们还为在需要仔细平衡精度、隐私和效率的环境中部署基于雷达的 HAR 系统提供了实用指导。

www.xueshuxiangzi.com