**Towards 3D Human Pose Construction Using WiFi**

Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su

*State University of New York at Buffalo, Buffalo, NY USA*

1. **概述**

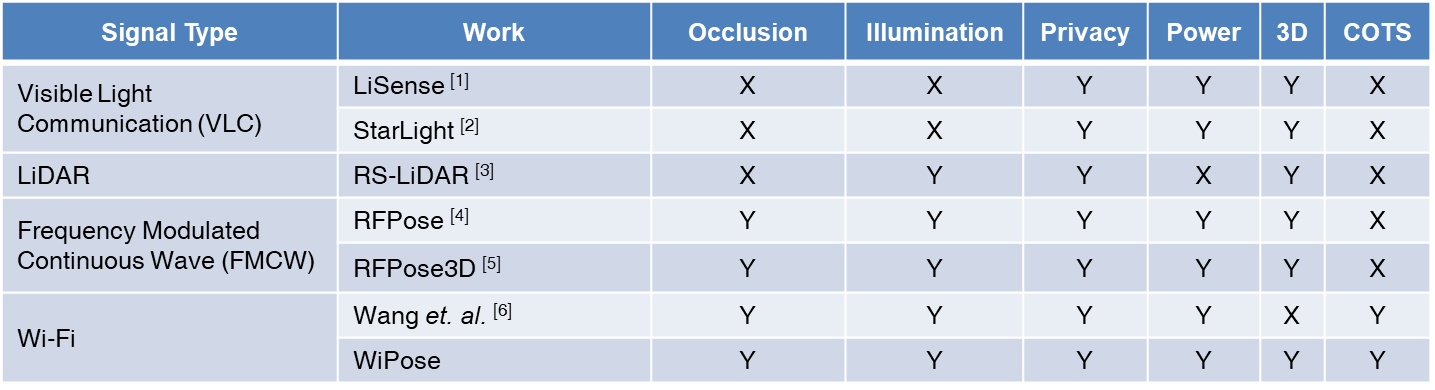
与传统的基于监控摄像头的姿态估计技术相比，无线信号受环境亮度的影响更小，可以探查到被遮挡的目标，并且具有较低的隐私泄露的风险。因此，在某些需要对目标人员进行大范围连续透明监控的应用场景中，比如健康监测、智能家居以及虚拟现实技术，基于无线信号的姿态估计技术拥有更加广阔的应用前景。当前，基于无线信号的姿态估计的精度已接近与传统的基于监控摄像头的方法，然而高昂的设备成本却限制了其实用性。

因此，本文提出了一种利用普通商用的WiFi设备进行实时人体姿态估计的方法，WiPose。无线信号具有隐晦的语义特征，无法像传统方法从RGB图像中直接识别人体的关键点构建姿态模型。另外，受环境、设备和其他活动人员的影响，估计结果中的噪声存在降低估计精度的风险。针对这些问题，WiPose将BVP模型扩展至3D空间以提取细粒度的动作特征，并基于时空合理性限制姿态估计空间，使结果趋于真实结果。

主要贡献：这是第一篇利用普通商用WiFi设备进行3D人体姿态估计的工作，展现了WiFi信号更广泛的应用潜力。

关键字：Human Pose Construction，WiFi Sensing， Deep Learning

1. **相关工作**

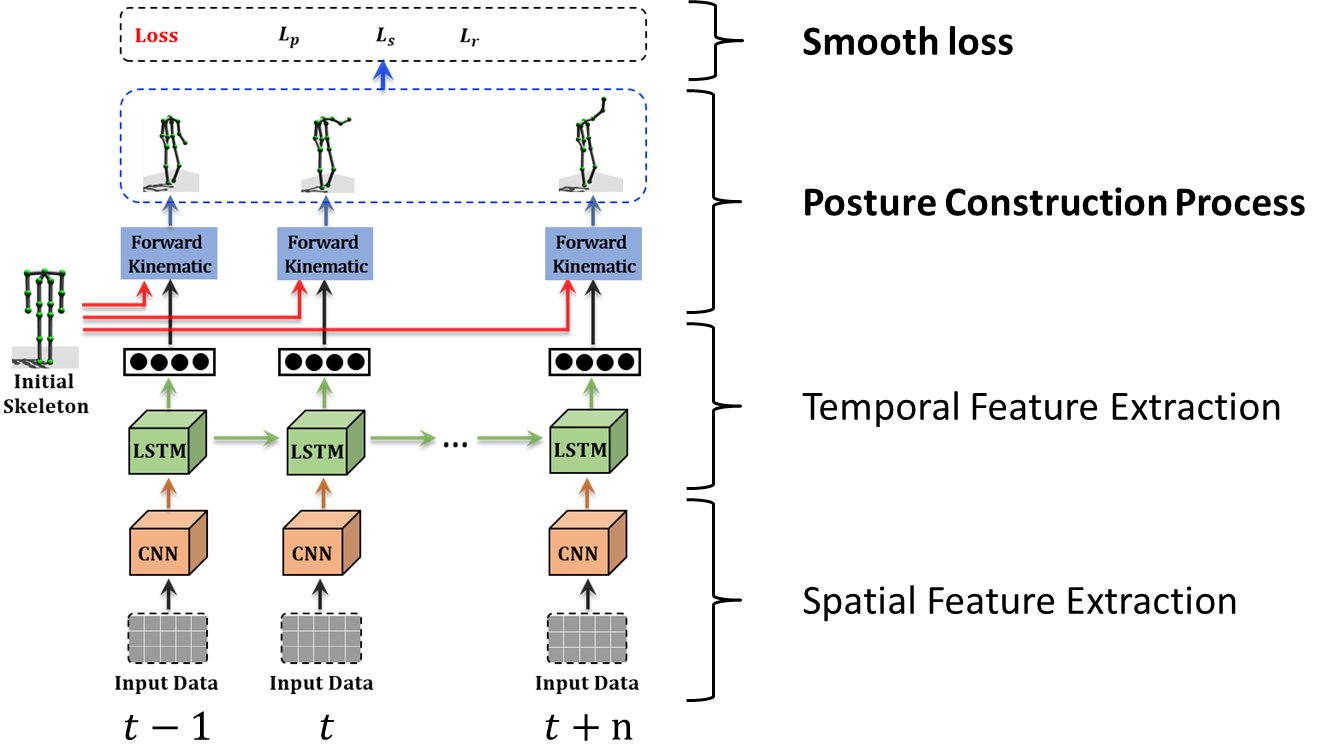
****

1. **挑战**

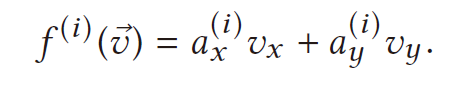
* 真实性（Realistic-looking）：从空间的角度保证估计姿态的合理性，比如双臂的长度区别不应太大。
* 连续性（Continuous and smooth）：从时间的角度保证估计姿态的合理性。现实中，人类的活动应是连续的，前后时刻变化不应过。
* 可行性（Substantial information）：接受到的无线信号是经多重路径传播之后，在某时刻由天线接收到的叠加值。其中，难以直接从中得到目标动作的信息。

1. **主要方法**

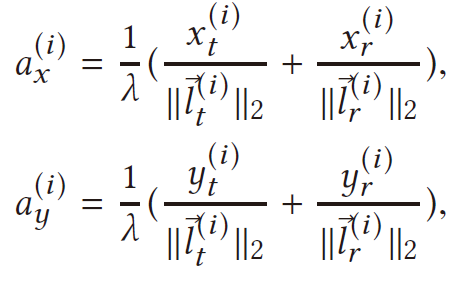
针对以上三个挑战，WiPose提出一种End-to-end的RNN回归模型进行姿态估计，如下图：



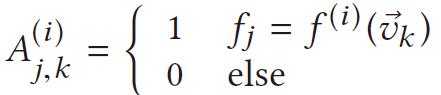
* WiPose将BVP模型扩展至可表示3D特征的形式，并依此转化接收到的WiFi信号作为此End-to-end模型的输入（Input Data）。BVP模型由Zheng et al. 等人在WiDar3.0[7]这个工作中提出，该模型根据目标各个身体部位在X、Y两个方向上的速度分量推测其真实的运动情况。公式如下：



为移动目标在频率分量上造成的频率便宜，、 为与天线对i 有关的相关系数。根据菲涅耳区的理论， 可表示为：



与 分别为发送、接收天线在X轴上的坐标， 为信号的波长。因此，可以确定相关系数 、 仅与天线的摆放位置有关，WiDar3.0 用一个二进制矩阵A来表示这些相关系数：



因此，BVP和DFS的对应关系可由如下公式计算：



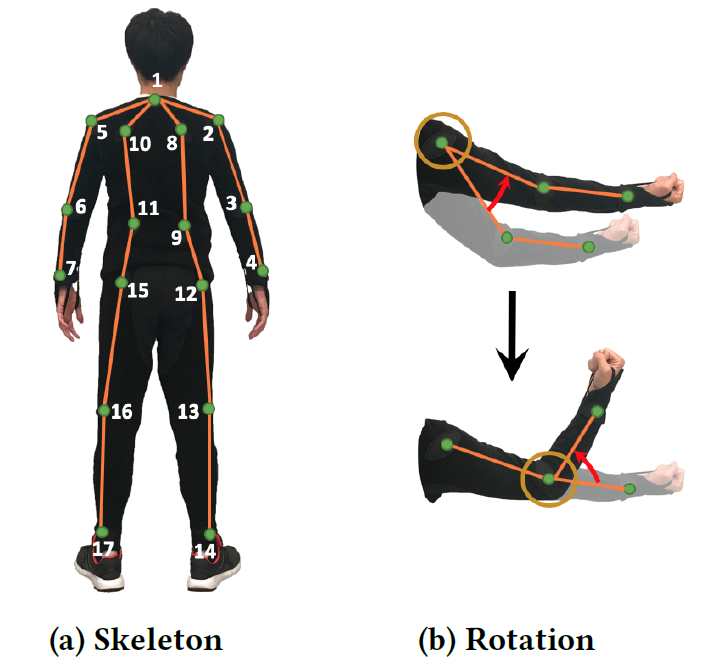
最后，通过求得使D与A\*V最相似的c最优解即可推算出目标活动的速度矩阵。

WiDar3.0 提出的BVP针对的使2D的动作检测，其包含的信息还不足以用于构建3D的姿态模型。因此，WiPose对BVP进行如下扩展：

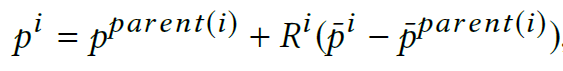


并利用相同的方式进行计算，将接收到的CSI信号转化为一个K\*K\*K的速度张量，K为在某个方向上可能存在的速度个数。

* 每一个时刻RNN输出一个用以表示姿态的向量矩阵，并通过Posture construction process来保证结果的空间合理性。Posture construction process 首先初始化个人目标姿态，其中各个关键点之间的连线长度是固定不变的，如下图所示：



这些关键点组成了一个树形结构，其中每一个关键点仅受其父节点影响。因此，关键点i 的状态可表示为：

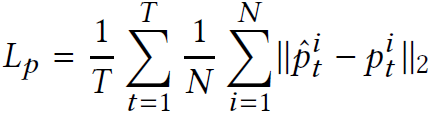


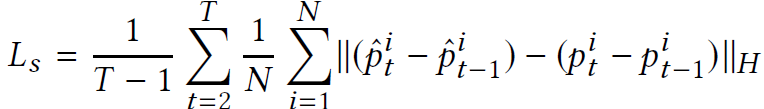
、 分别为关键点i及其父节点的极坐标，由于是对3D姿态进行估计，因此极坐标需要包括X、Y、Z三个方向上的旋转角度。是旋转矩阵，表示关键点i绕其父节点的旋转状态。当得到每一个关键点的旋转状态之后，即可根据初始化的人体姿态估算目标当前的姿态。

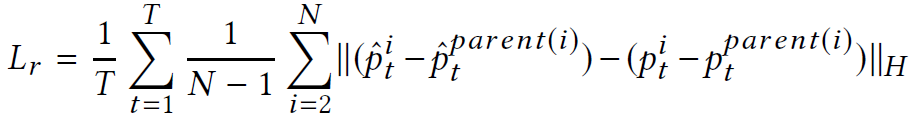
为保证预测结果在时空上的连续性，WiPose利用一种平滑损失函数（Smooth Loss）来避免模型训练时的梯度爆炸问题。Loss主要由三个部分组成，公式如下：



、、 分别为估计姿态与监督姿态的相似度，估计关键点的位置变化与监督系统提供的关键点位置变化的相似度，以及估计关键点相较于其父节点的位置变化与监督系统提供的关键点相较于其父节点的位置变化的相似度。公式如下：







其中 与 是对估计结果在时间维度上的限制。为了使估计的结果更加平滑，避免受到由突然出现的噪点导致的梯度爆炸问题，WiPose采用了一种常用的平滑损失函数, Huber Loss。 和 为人工定义的草参数，用于控制各个Loss的重要程度。

1. **实验记录**

|  |  |
| --- | --- |
| **实验设置** | |
| WiFi设备 | * 发送天线1根 * 接收天线3根 * 工作信道164 * 采样率 1000 Hz |
| 监督设备（VICON） | * VICON相机 21台 * Markers 17个 * 采样率 10 Hz |
| 目标活动 | * Lifting left/right hand (45/90/180) * Lifting both hands (90/180) * Sweeping left/right (45/90) * Sweeping both hands (90) * Lifting leg * Waving hand * Walking on the spot |
| 基准方法 | * RFPose3D |
| 测试人数 | * 10人 |
| 模型参数 | * 4-layer CNN (64,128,64,1) * Dropout rate = 0.2 * Activation = Leaky ReLU () * LSTM (Hidden State Number = 544) * Dropout rate = 0.1 * (, ) = (1.0, 1.0) |

|  |  |
| --- | --- |
| **评估内容** | |
| Posture Construction  姿态构建精度 | * 以表格的形式描述关键点的预测结果 * 同时比较了以CSI和3DVP的输入形式 * 与RFPose3D对比 * 平均精度为28.3mm (CSI), 36.7mm (3DVP) |
| 光照条件 | * 用图片描述某个例的结果和测试环境 |
| 训练集大小的影响 | * 调整不同的训练数据比例观察结果 * 以表格的形式记录结果 |
| 采样率的影响 | * 通过设置不同的采样率观察准确度 * 以表格的形式记录结果 |
| 接收天线数的影响 | * 比较不同天线数下的准确度 * 以表格的形式记录结果 |
| 对遮挡目标的估计 | * 用图片描述某个例的结果和测试环境 * 以表格的形式记录结果 |
| 环境影响 | * 不同环境下的估计精度 * 以表格的形式记录结果 |
| 用户多样性 | * 以表格的形式记录结果 |
| 不同信号状态的影响  (强反射，不同房间) | * 不同信号状态下的估计精度 * 以表格的形式记录结果 |
| 运行时间分析 | * 分析模型各个部分的运行时间 * 以表格的形式记录结果 |

1. **总结及启发**

* 该文章利用CNN-RNN网络从普通商用WiFi设备接收到的CSI信号中提取了细粒度的用户动作信息，并依此实时的估计用户姿态。
* 美中不足的地方是，扩展的3DVP模型并没有解决BVP模型对硬件部署位置的需求，无法利用位置的WiFi设备进行姿态估计。另外，从实验设置可以看出，WiPose目前还无法准确的估计移动目标的姿态，目标人员必须处于某个固定的位置。
* 由于WiPose目标是进行细粒度的姿态检测，周围人员的活动可能会对估计结果带来负面的影响，从而导致WiPose进行多人姿态检测的难度较大。从另一角度来看，如果感知目标是粗粒的，比如人员位置，那么BVP或者3DVP模型可能有潜力做到多人的实时定位。
* 3DVP提供了丰富的空间信息可用于姿态的构建。如果结合空间分辨率更高的无线信号，如毫米波和亚毫米波，则感知系统可能实时精确的估计细粒度的动作，如手写识别。这将会真正的使这类应用具备实用性。

1. **参考文献**
2. Tianxing Li, Chuankai An, Zhao Tian, Andrew T Campbell, and Xia Zhou. Human sensing using visible light communication. *In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM, 2015.
3. Tianxing Li, Qiang Liu, and Xia Zhou. Practical human sensing in the light. *In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, ACM, 2016.
4. Zhe Wang, Yang Liu, Qinghai Liao, Haoyang Ye, Ming Liu, and Lujia Wang. Characterization of a RS-LiDAR for 3D Perception. *In Proceedings of the 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE, 2018.
5. Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. ACM, 2018.
6. Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. RF-based 3D skeletons. *In Proceedings of the Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)*. ACM, 2018.
7. Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, and Dong Huang. Person-in-WiFi: Fine-grained Person Perception using WiFi. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2019.
8. Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. *In Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 2019.