# 走跑数据及标签说明

By 张琛

## 版本控制

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 编辑说明 | 编辑人 | 编辑时间 |
| V1.0 | 创建文档 | 张琛 | 2013-10-18 20:19 |
| V1.1 | Part 6 更改问题处理方式为：多行标签 | 张琛 | 2013-10-18 20:48 |

## 采集位置

考虑到身体左右对称性，我们认为设备放置在身体一侧采集数据是足够的。本次采集要求手机分别置于右手，右上衣侧口袋，右裤口袋。

由于部分采集人员上衣两侧没有口袋，采集时设备置于左胸口袋中，此部分数据未做标记与区分。

## 文件命名

数据文件命名为“人名[run]\_动作ID\_采集次数.xml”。人名后面跟“run”，为跑步数据，没有“run”则为走路数据；动作ID对应设备位于人体的位置；采集次数（0,1,2…），即同一采集人员将设备放置在同一身体位置下采集数据次数，本次采集，同一位置大多只采集一次，因此大多为0。

其中，动作ID对应关系如下（标红字体为本次采集设备要求位置）：

|  |  |
| --- | --- |
| 左手 | a4 |
| 右手 | a5 |
| 左上衣口袋 | a6 |
| 右上衣口袋 | a7 |
| 左裤口袋 | a8 |
| 右裤口袋 | a9 |

例：文件名“ZC\_a5\_1.xml”，表示采集人员张琛（ZC人名缩写），右手手持设备，第二次采集，走路数据。

**注意**：由于采集人员可能使用不同机器进行同一动作、同一位置的采样，因此存在不同文件夹下出现同名文件的可能

## 数据采集人员

人员及注册名对应表

|  |  |
| --- | --- |
| 采集人员（共14人） | 数据文件名缩写 |
| 冯玲珑 | FLL |
| 胡钰 | HY |
| 耿老师 | GWD |
| 师母 | SM |
| 张琛 | ZC |
| 雷昊 | LH |
| 李霞 | LX |
| 金秉文 | JBW |
| 李萌坚 | LMJ |
| 张娜 | ZN |
| 卫文韬 | WWT |
| 王庚 | WG |
| 任光阔 | RGK |
| 王镇 | WZ |

## 数据有效性

本次采集数据要求时长1min 30s 左右，部分数据由于采样人员没有将设备锁屏，中途­­­­­手指、身体误触碰手机屏幕上的停止按钮，导致采样提前结束，没有达到要求时间。但考虑到实际运动可能处于复杂的场景，这些数据仍然作为有效数据。

## 数据标签

|  |  |
| --- | --- |
| 行为类型 | 标签值 |
| 走路 | 1 |
| 跑步 | 2 |

## 目前问题处理

部分数据在采集过程中，出现某一段时间内传感器读书线性变化的情况，不符合实际场景，如图1红圈所示。推测原因是运动过程中，设备震动剧烈，或发热情况导致传感器读数失真。

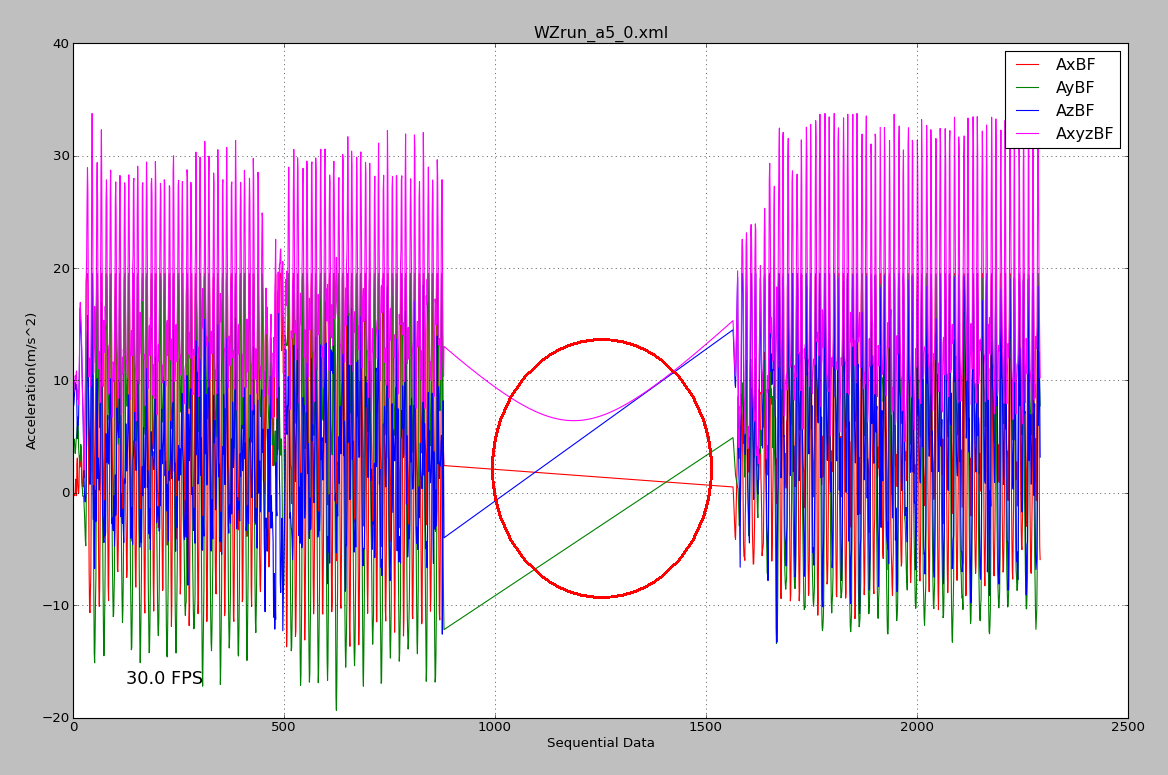


图1 出现问题的数据

~~目前处理方式为，如果此问题发生在数据采集中间时刻，无视这段错误数据，仍将整段数据标记为走路（或跑步）状态，如图2蓝色框选区域所示；如果此问题发生在采集开端或结束时刻，则抛弃这段数据。~~

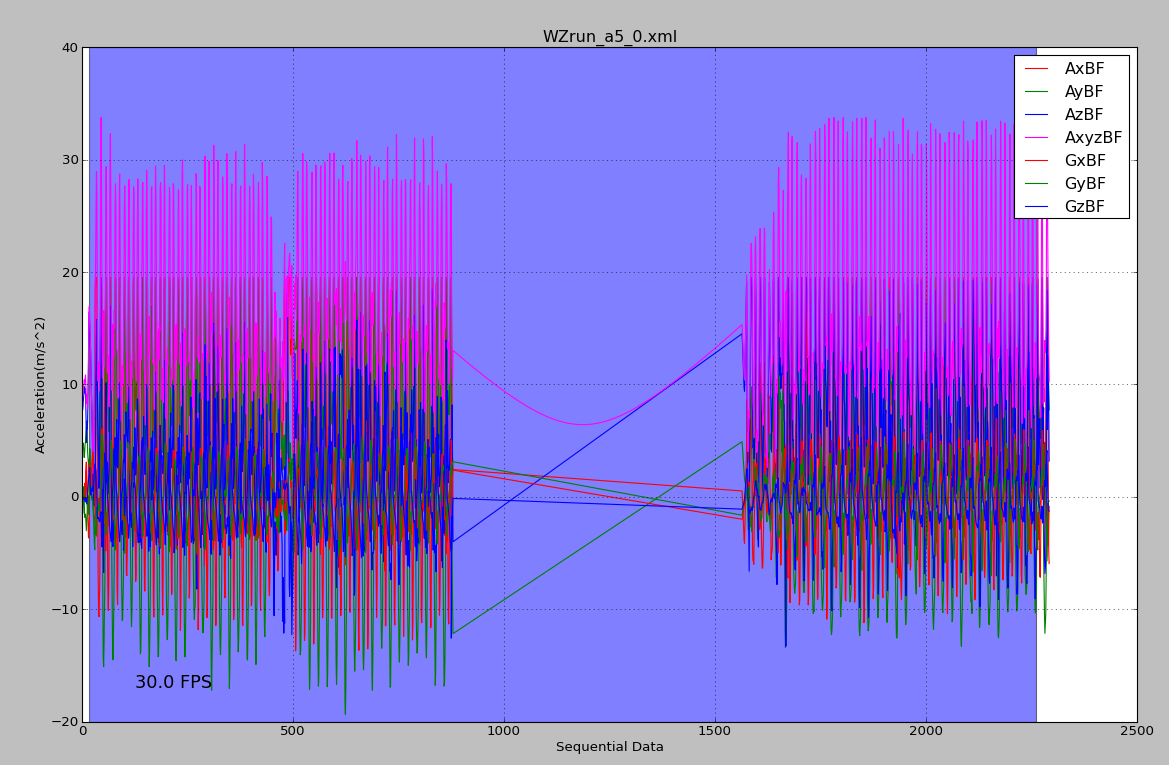
~~~~

图2 标记整段区间为动作区间

目前处理方式为，对可信的数据子区间进行标记，将标签txt文件分为多行，每行为一段可信子区间。重新标记的文件有：

WZrun\_a5\_0.xml 多行

HYrun\_a5\_0.xml 重划单行

HYrun\_a7\_0.xml 重划单行

\a780L-huaweiproj\LMJrun\_a5\_0.xml 重划单行

\s4-huaweiproj 所有文件 舍掉

出现此问题的数据文件列表如下：

|  |
| --- |
| \wz-walk-run-data\WZ\_a5\_0.xml |
| \wz-walk-run-data\WZrun\_a5\_0.xml |
| \MX-huaweiproj\LMJrun\_a9\_0.xml |
| \s4-huaweiproj 所有文件 |

由于采样用到的 Samsung S4 是隔壁借来的，目前无法用这台机器进行问题重现。

出问题文件示例图：

