**人脸-手机关联数据描述**

设备：

（1）摄像头，用于采集人脸数据，采集范围为一般两个方向，30-40米范围内的人物头像。

（2）手机雷达侦码设备，用于捕捉手机IMSI码，采集范围为设备的周围200米范围内。

**数据特性分析介绍**

1. 手机IMSI数据

IMSI采集原理遵循移动终端小区重选原则，我们部署的侦码设备小区优先级设置为最高，当移动终端接近时会检测到信号强度好且优先级高的侦码设备并接入、上报IMSI等信息，完成信息采集后，侦码设备会下发拒绝信息，为了降低重复采集对用户的影响，通常情况下同一侦码设备对同一终端（手机）半小时只采集一次。针对行人步行等场景，实现步行/电单车/慢速通过车辆目标的IMSI采集，覆盖范围200m左右；侦码设备的采集率约80%，用户有业务发生时不会采集，比如打电话/看视频等。

1. 人脸数据

每一张采集的人脸照片都会通过人脸算法进行识别归档，实现人脸聚类，建立一人一档。

人脸数据也有一定漏采率，真实采集率预计90%，同时由于人脸算法误识率，建立的一人一档混入他人图片概率预计3%左右。最终数据中，舍弃人脸的基础特征，只保留人脸的ID号，比如同一个人出现在不同的场地，数据中会用ID号体现，而不是图片数据。

**数据采集描述**

根据场景需要会在一个城市内部署不同数量的手机雷达侦码设备，一般情况下100到300台的规模，有部署在人脸抓拍相机同一位置的，也有单独部署的，如下图所示，红圈代表同一位置部署。即：摄像头的位置可能与手机雷达侦码设备相同，也可能不同，需要通过GPS进行判断。考虑到相机拍摄的角度限制，根据车流/人流方向问题同一位置可能部署多台相机（例如两台摄像头的GPS相同，而ID号不同），人脸的抓拍相机范围扇形：距离3-10米，宽是3米。

****

**数据格式**

三个数据表：

人物头像表：

FID（脸ID）、UPTIME（采集时间）、LON（经度）、LAT（纬度）、PLACEID（近似采集地点）

手机IMSI表：

IMSI（手机ID）、UPTIME（采集时间）、LON（经度）、LAT（纬度）、PLACEID（近似采集地点）

人物头像-手机IMSI 关联表：

FID（脸ID）、IMSI（手机ID）、label（值为0-1000的数字，表示是同一个人的程度）

**目标**

预测提供的数据中所有为同一个人的FID（脸ID）、IMSI（手机ID）。

例如，

#FID，IMSI，label

782213608087068672，460008552700204，995