# 毕业设计报告 肖鹏 131220124

毕业设计题目：基于K匿名及差分隐私的定域功能性影像隐私保护

摘要：

近些年来，随着机器学习及数据挖掘研究的深入与进步，大数据产业变得十分火热，无论是学术机构或者是社会部门，都十分关注大数据分析带来的收益与相关产业的发展。就目前的数据交换现状来看，数据的交互过程中存在严重泄露的风险，使得拥有大量数据的部门如刑警部门在需求相关学术机构对其所持有的数据进行分析时会有对数据泄露风险的担忧。

本文的主要内容是针对影像数据（照片，视频资料等），在大数据交互分析中会产生泄漏风险的部分，人脸信息，寻求了一种较为安全的隐私保护构想。通过借助差分隐私模型，通过对人脸数据添加干扰噪音，从而达到即便攻击者已经掌握了除必要数据以外的辅助数据，该攻击者仍然无法从影像资料中推测出具体到个体的行为。同时为了更好的满足数据可用性，保护手段中同时保障了数据发布的K匿名性。为此我们将该算法的实现分成三部分，第一部分利用人脸识别算法，将影像资料中的人脸数据提取到标准尺寸，并按规则划分。预先提取需要处理的资料。第二部分是本保护算法的重点，依据不同规则划分出的人脸单位进行内部面部相似度比较，每个个体提取出前K个与之最相似的人脸。将每个个体与之最相似的K－1个个体进行加权线性叠加，建立新的对应个体。第三部分，将新生成的个体与原数据影响进行面部相似度比较，重现提取前K个分析是否达到保护要求，并将新生成的个体重新覆盖到原始的影像资料中。

为了验证算法的有效性，收集356个不同的个体数据划分后进行交叉测试。实验结果证明经过标准化提取以及规则划分后加权线性重叠处理后，新生个体与原始个体平均相似度只有。。％,在所有数据个体最相似排名中基本都大于K，满足K匿名的要求。同时，整个算法执行过程在单规则分类下，167个个体处理完成及输出比对分析结果需要15min，这些结果说明本文提出的算法是高效可行的。

  

**原图片 提取人脸后 添加噪声后**