ICS 35.040
CCS L 71

团　　体　　标　　准

T/AI 113—2021

生物特征识别服务中的隐私保护技术指南

Privacy protection technical guide for biometric recognition services

2021－10－21 发布　　2021－10－21 实施

中关村视听产业技术创新联盟　发布

目　次

前　言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分:标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由新一代人工智能产业技术创新战略联盟 AI 标准工作组提出并归口。

本文件起草单位：海信集团控股股份有限公司、西安电子科技大学、云从科技集团股份有限公司、中国电子标准化研究院、北京大学、上海依图网络科技有限公司。

本文件主要起草人: 陈维强、高雪松、朱辉、胡伟凤、李祁、张淯易、张紫铃、孟祥奇、孙菁、何晨迪、温浩、姚志强、李军、曹霖、李婧欣、田永鸿、赵春昊、张辰宇、寇笑语。

引　言

本文件用于参考和指导使用生物特征识别服务中的隐私保护技术，以配合文件《生物特征模板的安全使用要求》的实施。

本文件从生物特征识别中生物特征信息数据出发，使用技术性手段保护生物特征识别服务中的特征隐私。生物特征信息是指生物特征识别中，系统直接采集到的用户生物特征和利用用户生物特征数据生成的包含用户生物特征的生物特征模板。在对隐私保护提出建议的同时，本文件也兼顾考虑了生物特征识别的性能和准确度。

本文件的目的是针对生物特征识别服务中的用户隐私数据的保护进行一定程度的指导和提出一些参考建议。生物特征识别服务中的隐私保护是非常必要的，但截至目前也没有通用、统一的方法论。

生物特征识别的种类繁多，常见的有指纹、人脸、虹膜、声纹、静脉等。每一个所产生的生物特征样本的形式、内容，以及对应的生物特征识别模型都不同。各个生物特征特性以及目前的研究应用现状也导致准确度与性能差异很大。这些导致本文件无法产生准确统一的安全性与可用性约束条件。因此，本文件针对识别服务中不同的阶段，提出了不同的隐私保护技术使用建议以供参考。

生物特征识别服务中的隐私保护技术指南

1 范围

本文件提供了生物特征识别服务中隐私保护技术的分类，开展全服务生存周期的生物特征隐私保护的实施、评估方法和技术建议。

本文件适用于指导相关技术人员有针对性地使用生物特征隐私保护技术，也适用于网络安全相关主管部门、第三方评估机构等组织开展企业生物特征识别模板数据的安全监督管理、评估等工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版（包括所有的修改单）适用于本文件。

GB/T 29268.1 信息技术 生物特征识别性能测试和报告 第1部分:原则与框架

3 术语和定义

下列术语和定义适用于本文件。

3.1

隐私保护技术 privacy protection technology

通过访问控制、加密、混淆等技术对用户的敏感数据进行处理，对用户的隐私信息进行保护的技术。

3.2

隐私信息 private information

用户所拥有的不想被他人获知的信息，一旦泄露会威胁其财产或人身安全。

3.3

隐私保护算法 privacy protection algorithm

通过访问控制、加密、混淆等技术对用户的敏感数据进行处理，以保护用户隐私信息的技术。

3.4

隐私保护数据 private data

用户所拥有的不想被他人获知的数据，隐私信息的载体。

注:本文件认为数据库中存储的生物特征模板数据即隐私保护数据。

3.5

可逆性 reversibility

隐私保护算法执行前后，隐私信息被还原的能力。

3.6

延伸授权性 extended controllability

跨系统交换过程中的接收方隐私信息保护效果与发送方的保护要求的匹配程度。

3.7

偏差性　deviation

隐私保护算法执行前后，隐私信息分量和隐私保护后发布/攻击者或第三方可观测到的隐私信息分量之间的偏差。

3.8

复杂性　complexity

执行隐私保护算法所需要的代价。

3.9

有用性　usefulness

存在价值可供后续使用的数据的特性，可供后续挖掘、识别、分类、匹配等操作。

3.10

信息损失性　information loss

信息因其他操作所损失的有用性。

注：本文件中其他操作指扰乱、混淆等不可逆的隐私保护算法作用。

4 生物特征识别服务

一般生物特征识别系统的信息流以及组成系统按 GB/T 29268.1-2012 描述的分为：数据采集子系统、数据处理子系统、存储子系统、比对子系统和决策子系统，该分类按照数据的生存周期分类，可对应第 7 章的数据采集、数据计算、数据发布三个阶段。在生物特征识别系统中，这些概念的组件可能不存在或可能不直接对应物理组件，各个组件也可能分布式部署。

5 基于密码学框架中融入生物特征识别的隐私保护技术

5.1 概述

本章给出了基于密码学框架中融入生物特征识别的隐私保护技术，该类技术通常需要使用密钥对生物特征进行加密运算，以此来保护生物特征的隐私性。常用的基于密码学框架中融入生物特征识别的隐私保护技术与特性见附录 A。

5.2 特征变换法

特征变换法的核心思想是把原始生物特征变换到另一个域中，生成安全模板，数据库仅存储变换后的安全模板，身份鉴别也通过计算安全模板之间的相似性来实现。该类方法的设计关键要素为映射函数。根据映射函数的特性，可以分为生物哈希法和不可逆变换法。生物哈希法的基本思想是使用特定密钥定义一个正交随机变换函数，并用该函数对生物特征进行变换，变换后得到的数组再进一步二值化得到哈希序列。生物哈希法的特点是一个用户对应一个密钥，认证时只有同时提供正确的生物特征和密钥才能认证通过。不可逆变换法是对生物特征进行不可逆变换，使得从变换后的特征无法恢复出原始特征，以确保生物特征的安全性，这样即使密钥被盗，攻击者也不能获得生物特征。

5.3 密钥绑定法

密钥绑定的基本思想是将外部输入的密钥绑定在生物特征上，并由两者产生中间数据，数据库保存中间数据用于认证。中间数据不显示密钥或生物特征的信息，即在生物特征未知时，从中间数据很难计算出特征或者密钥信息。身份鉴别时，由生物特征数据从中间数据中解出密钥，通过检查密钥的有效性实现身份鉴别。代表性方法有 bioscrypt 算法、模糊承诺方案和模糊保险箱算法。密钥绑定通过用户密

钥和生物模板相融合生成中间数据，具有防止特征泄漏，隐私性较好的优点；同时引入纠错码，克服噪声问题，包容了同一个体的生物特征差异，容错性更强。

但是如果使用不同密钥对同一生物特征生成的多个模板在被同时被获取时，通过交叉比较可能得到原始模板信息;另外，如果攻击者把一部分杂凑点换成自己的特征点对，能够隐蔽地通过系统验证。

5.4 密钥生成法

密钥生成法的基本思想是直接由用户生物特征生成密钥，并将密钥保存于数据库中；在认证阶段，用户生物特征采用同样的方法生成密钥，通过比对密钥来实现身份鉴别。密钥生成算法的优点是不需要保存模板，而生物特征模板和从中提取的关键信息为单向映射关系，即可以由生物特征恢复出关键信息，但通过关键信息并不能恢复出用户生物特征。

但该算法通常面临鉴别力低的问题，如果产生的密钥与输入模板无关，则此方案密钥可重复利用率较高且密钥相关性较低，但同时将导致较高伪接受率。另一方面，如果方案对输入用户模板产生不同密钥，此方案密钥相关性较高且密钥可重复利用率较低，会导致高伪拒绝率。

6 基于生物特征识别框架中融入加密算法的隐私保护技术

6.1 概述

本章给出了基于生物特征识别框架中融入加密算法的隐私保护技术，主要分为两类，有损失的隐私保护技术和无损失的隐私保护技术。上述两类的区别在于信息是否有损失，有损失的隐私保护技术会使数据部分失真，一般具有更高效率；无损失的隐私保护技术可以保留完整数据，一般效率相比前者更低。常用的基于生物特征识别框架中融入加密算法的隐私保护技术与特性见附录 A。

6.2 有损失的生物特征识别隐私保护技术

6.2.1 概述

有损失的生物特征识别隐私保护技术是通过特定策略修改原始生物特征模板数据，使模板数据存在部分失真，进而达到保护效果。本文件所提供的有损失的生物特征识别隐私保护技术主要分为两大类：基于混淆的保护技术和基于差分的保护技术。

6.2.2 基于混淆的保护技术

基于混淆的保护技术包括 k-匿名、l-多样性、和 t-接近性等。基于混淆的保护技术具有计算难度低、计算开销小以及保护效果良好等优点，能够在不需要可信第三方参与的前提下实现保护。与此同时，基于混淆的保护也存在缺点，其中，k-匿名、l-多样性、和 t-接近性技术难以抵御背景知识攻击和同质化攻击。

6.2.3 基于差分的保护技术

基于差分的保护技术是不依赖于攻击者背景知识的保护技术，即差分隐私技术。相比于传统的密码学技术，差分隐私技术的部署成本更低，算法更加轻量，并且可以对用户的隐私信息提供语义安全。对于差分隐私的研究目前主要集中在两个方向：集中式数据模型和本地化数据模型。集中式数据模型是由管理者收集数据并对数据进行统一的差分隐私保护处理。而本地化数据模型是指由用户执行差分隐私保护算法，在客户端完成差分隐私数据保护。本地差分隐私保护技术不需要依赖可信第三方，用户数据的收集只涉及加噪数据版本，原始真实数据完全被保护在本地设备，这既解决了用户对个人隐私数据不

能自主控制的问题，也降低了大量隐私数据在非可信第三方存储的隐私泄露风险。需注意差分会带来数据的损失，因此该技术比较适合统计整体数据特性的场合。

6.3 无损失的生物特征识别隐私保护技术

6.3.1 概述

无损失的生物特征识别隐私保护技术主要是通过数据加解密和陷门函数，基于密钥的管理与保护实现对生物特征识别的隐私保护，本文件所提供的无损失的生物特征识别隐私保护技术包括同态加密、安全多方计算以及矩阵计算三种方法。

6.3.2 同态加密

同态加密是用于解决将敏感数据外包到不受信任的计算环境中的隐私和安全问题的一种加密原语。其主要涉及数学中数论等领域的知识，在安全云计算和委托计算、远程文件存储，密文检索和电子选举等方面得到了广泛的应用。同态加密分为全同态加密和半同态加密两类，全同态加密是指同时满足加同态和乘同态性质，可以进行任意多次加和乘运算的加密函数；半同态加密是指仅满足加同态或乘同态的加密函数。但由于同态加密算法目前普遍存在公钥尺寸过大，并且只支持单比特数据的加密，故需要大量的计算资源成为其缺陷。基于全同态的方案速度较慢，但是支持的运算多，可用性强。基于半同态的方案相比较全同态局限较多例如一个方案只能满足一种运算，但是速度比全同态快。

6.3.3 安全多方计算

安全多方计算致力于解决一组互不信任的参与方之间协同计算的隐私保护问题，是现代密码学研究中的一个重要分支，也是信息安全领域一个重要的研究内容。安全多方计算可广泛应用在数据挖掘、数据库查询、科学计算、几何或者几何关系判断、统计分析等诸多计算领域的数据安全保护中。但在实现时普遍存在资源消耗过大、处理速度慢的缺陷。同时，安全多方计算需要多方参与通信，随着交互次数更多，三方或者多方参与时耗时和通信显著增加。

6.3.4 矩阵计算

矩阵计算，又称随机投影扰动，其方式相比较于同态密码体制，计算效率高，性能开销也相对较小，可以实现数据安全的两方或三方计算，也可将其归为特征变换方法中。矩阵计算的方案实际应用存在一定问题，例如存在加密矩阵泄露风险，导致模板数据和用户请求数据的泄露；对于模板长度过长的生物特征数据，其矩阵计算开销也会显著增加。因此，该方法适用于生物特征模板长度适中或者较小的情况，并能在模板统一使用相同密钥的外包计算中发挥优势。

7 不同服务阶段的隐私保护

7.1 数据采集阶段的隐私保护

7.1.1 概述

在注册阶段服务系统需要从用户方采集生物特征数据，生成生物特征模板；在验证阶段，服务系统需要从用户方采集生物特征数据，并将其与存储的生物特征模板进行匹配。

7.1.2 需求分析

在数据采集阶段，生物特征数据从用户客户端上传到了服务系统，该过程是存在安全隐患和隐私泄露风险的。客户端获取了用户最原始的生物特征数据，如果不经处理直接发送便很容易造成用户隐私泄

露，因此需要采取一定的隐私保护手段。

7.1.3 技术建议

针对采集过程中的隐私泄露风险，可以使用同态加密、矩阵计算的技术对用户生物特征模板进行加密，以抵抗重构还原等攻击并保护用户的生物特征模板信息。也可以使用本地化差分隐私技术在客户端对生物特征数据加定量的噪声使原本数据得到保护但却不影响生物特征模板生成和使用。

7.2 数据计算阶段的隐私保护

7.2.1 概述

在验证匹配阶段，用户的验证用生物特征模板和注册用生物特征模板在系统服务器中进行模板匹配计算。

7.2.2 需求分析

数据计算阶段一般在系统服务器中进行，这一阶段的隐私保护手段取决于服务器是否可信。若在可信服务器上进行计算则可采用轻量级的隐私保护技术；若在不可信服务器上进行计算则需要采用完全的隐私保护手段，否则容易产生生物特征模板被窃取、丢失和泄漏的风险。

7.2.3 技术建议

计算阶段建议使用同态加密和矩阵计算技术来实现隐私保护。同态加密和矩阵计算技术可以实现数据在密文状态下的计算，以防止服务器端的数据被窃取和泄漏后造成的用户生物特征模板的曝光。

7.3 数据发布阶段的隐私保护

7.3.1 概述

在一般的生物特征识别服务中，服务器在匹配计算完成后系统会返回匹配的相似度。

7.3.2 需求分析

匹配计算发布的相似度结果如果被攻击者获取，攻击者则能够进行多次有方向性地尝试攻击，或者进行差分攻击。

7.3.3 技术建议

宜对匹配计算结果采用差分隐私保护技术，可以有效地防止攻击者通过对系统进行输入输出的比对来窃取用户的生物特征信息。

8 针对生物特征识别服务隐私保护效果的评估

宜从可逆性、复杂性、敏感度不变性、延伸授权性、偏差性、信息损失性 6 个评估指标来对隐私保护效果进行评估。对于一个隐私保护算法宜具有以下性质：

a) 不可逆性：攻击者/第三方从所观测到的脱敏后或是加密后的模板数据应无法推断出用户原始的生物特征模板，即隐私保护算法应不具有可逆性；
b) 低复杂性：对于生物特征模板安全使用，其包括但不限于通信开销、计算开销、交互次数、存储开销等内容，对于各种各样的生物特征识别隐私保护算法，都应不影响用户的正常使用体验，即要有较低的复杂性；

c) 敏感度不变性：隐私保护算法对由噪声、表达向量的长度、是否有序和是否对齐等因素变化有一定的容忍度，容忍度越高，敏感度不变性越高；
d) 延伸授权性：在跨系统交换过程中，接收方的隐私信息保护效果应与发送方的保护要求一致；
e) 大偏差性：攻击者观测到的模板数据应与原始模板数据完全不同，即隐私保护算法加密应具有很大的偏差性；
f) 信息损失性：隐私保护算法应具有可控且适当的信息损失性，保证识别精度与结果的可用性。

对于类似指纹基于细节点的生物特征识别方案，本文件建议使用基于在密码学框架中融入生物特征识别的隐私保护方法，例如密钥绑定方法，该方法具有高精度与高效率。对于基于嵌入表示深度学习的生物特征识别方案，对于具有高精度识别需求的服务，本文件建议使用全同态、矩阵计算等隐私保护方法；对于具有快速识别需求的服务，本文件建议推荐使用安全两方计算、半同态等技术方法对生物模板信息进行隐私保护。对于外包计算类的典型场景，推荐使用矩阵计算等隐私保护方法。

9.隐私保护过程

9.1 概述

生物特征识别服务中的隐私保护通常可分为确定目标、处理生物特征数据以及验证效果等步骤。

9.2 确定目标

9.2.1 概述

确定目标步骤包括确定生物特征对象、建立生物特征目标和制定工作计划等内容。

9.2.2 确定生物特征对象

在进行隐私保护的数据集范围内，宜根据以下要素确定哪些数据属于隐私保护的生物特征对象：

a) 法规标准。了解国家、地区或行业的相关政策、法律、法规和标准，待采集或发布数据是否涉及生物特征相关要求。
b) 组织策略。了解数据是否属于组织列入的重要数据或敏感数据范畴，数据应用时是否存在隐私保护的要求。
c) 数据来源。了解生物特征数据采集时是否做过隐私保护相关承诺。
d) 业务背景。了解数据来源相关信息系统的业务特性，了解业务内容和业务流程，即生物特征识别服务的相关算法细节。
e) 数据用途。了解待发布数据的用途，即是否用于生物特征识别服务。
f) 关联情况。了解数据披露历史和隐私保护历史情况，待披露数据是否和历史数据存在关联情况。

9.2.3 建立生物特征隐私保护目标

建立生物特征隐私保护目标，具体包括确定隐私保护程度以及数据有用性最低要求。

需要考虑的因素包括：

a) 数据用途。了解数据隐私保护后的用途，涉及业务系统的功能和特性，即生物特征识别服务中用到的算法特性。考虑数据隐私保护的影响，确定数据有用性的最低要求；
b) 数据来源。了解数据获取时的相关承诺，以及涉及的生物特征；
c) 风险级别。了解数据属性和业务特性，拟采用的攻击模型及设定的隐私保护级别；
d) 隐私保护模型和技术。了解生物特征数据适用的保护标准，以及可能采用的隐私保护模型和技术。

9.2.4 制定工作计划

制定生物特征识别服务中的隐私保护实施计划，包括隐私保护的目的、目标、数据对象、实施团队、实施方案、利益相关方、应急措施以及进度安排等。

9.3 处理生物特征数据

9.3.1 概述

处理生物特征数据步骤分为预处理、选择模型技术、实施隐私保护三个阶段工作。

9.3.2 预处理

预处理是在对生物特征数据集正式实施隐私保护前的准备过程，一般地，预处理是对数据集施加某种变化，使其有利于后期进行处理。

预处理阶段工作可参考如下方法进行：

a) 形成规范化，或满足特定格式要求的数据，例如矩阵类数据要扩展或压缩成适合隐私保护技术的维度；
b) 对数据抽样，减小数据集的规模；
c) 增加或扰乱数据，改变数据集的真实性等方法。

9.3.3 选择模型技术

应依据数据的类型和服务特性，考虑隐私保护的影响，选择合适的隐私保护模型和技术，在可接受的隐私泄露风险范围内满足数据可用性的最低要求。选择的参考因素包括但不限于如下方面：

a) 是否需要对隐私泄露风险进行量化；
b) 聚合数据是否够用；
c) 数据是否可删除；
d) 是否需要满足可逆性；
e) 服务对效率的要求；
f) 服务对精度的要求；
g) 是否需要满足敏感度不变性；
h) 是否需要满足延伸授权性；
i) 是否可以对数据实施随机噪声添加；
j) 识别服务的算法特性；

9.3.4 实施隐私保护

根据选择的隐私保护模型和技术，对生物特征数据集实施隐私保护。主要工作包括：

a) 若存在多个需要隐私保护的生物特征数据集，则根据数据特点和业务特性设定隐私保护的顺序；
b) 依次选择相应的工具或程序；
c) 设置工具或程序的属性和参数，如设置源数据、用户名/口令、算法参数等；
d) 依次执行隐私保护工具或程序，获得结果数据集。

9.4 验证效果

9.4.1 验证效果含义

对生物特征数据隐私保护后进行验证，以确保生成的数据集在隐私泄露风险和数据可用性方面均

符合预设的目标。在验证满足目标过程中，需对隐私保护后隐私泄露风险进行评估，计算出实际风险，与预期可接受风险阈值进行比较，如果风险超出阈值，需继续进行调整直到满足要求。由于针对生物特征的攻击技术和能力在迅速演变，需要由内部专业人员或权威的外部组织定期展开验证评估。

9.4.2 验证生物特征数据隐私保护效果

验证生物特征数据进行隐私保护后的保护效果的方法包括：

a) 检查生成的生物特征模板数据，以确保模板数据中不直接显示原始生物特征数据；
b) 检查生成的生物特征模板数据，以确保所得数据符合既定隐私泄露风险要求；
c) 评估隐私保护算法及其参数配置；
d) 进行有针对性的入侵者测试，例如基于身份伪造的攻击，基于窃听的攻击，重放攻击，模板碰撞攻击，重构、爬山等欺骗攻击等攻击方式。

这些方法不能保证隐私保护后的数据满足生物特征信息安全保护的全部要求，但它们可以作为整个组织风险评估的一部分。

9.4.3 验证隐私保护后的生物特征数据可用性和准确性

隐私保护降低了数据质量和生成数据集的可用性。因此，需要考虑隐私保护后的生物特征数据集对于生物特征识别服务仍然有用。

存在一些方法用于验证数据有用性。例如，内部人员可对原始数据集和隐私保护后的数据集执行相同识别算法，并对结果进行比较，以查看隐私保护后是否导致不可接受的更改。一些隐私保护算法与生物特征识别算法可以进行深度融合，例如矩阵计算，安全内积计算等，这样的算法可以构成一个黑盒，具有与原始数据相同的可用性和准确性。

9.4.4 验证隐私保护的生物特征识别服务效率

生物特征识别服务进行隐私保护以后必然会影响一定的效率。因此，需要考虑隐私保护的生物特征识别服务的效率是否满足实际使用需求。

服务效率主要由计算开销，通信开销和交互次数决定。内部人员可以对各个算法的计算开销，通信开销和交互次数分别统计和对比以筛选效率高的算法，也可以直接统计数据从输入到输出经过的时间作为对比效率的依据。

附 录 A
（规范性）
生物特征识别隐私保护技术的选择

常用生物特征识别隐私保护技术的特性见表 A.1。

表 A.1 常用生物特征识别隐私保护技术的特性

类别	子类	输出数据类型	适用数据类型	可逆性	计算开销	通信开销	敏感度不变性	延伸授权性	偏差性	信息损失性
基于密码学框架中融入生物特征识别的隐私保护技术	特征变换法	微数据	所有	不可逆	低	低	中等	有	较高	中等
	密钥绑定法	微数据	所有	不可逆	较低	低	较高	有	较高	中等
	密钥生成法	微数据	所有	不可逆	较低	低	中等	有	较高	较高
基于生物特征识别框架中融入加密算法的隐私保护技术	混淆技术	微数据	分类数据	可逆	低	低	较高	有	较低	中等
	差分隐私	统计数据	数字类数据	不可逆	低	低	较高	有	中等	较高
	同态加密	微数据	所有	不可逆	高	较低	较低	有	高	较低
	安全多方计算	微数据	所有	不可逆	较低	较高	较低	有	较高	较低
	矩阵计算	微数据	所有	不可逆	较低	中等	较低	有	较高	较低

参考文献

[1] GB/T 26238-2010 信息技术 生物特征识别术语

[2] GB/T 29268.2-2012 信息技术 生物特征识别性能测试和报告 第2部分：技术与场景评价的测试方法

[3] GB/T 33767.1-2017 信息技术 生物特征样本质量 第1部分：框架

[4] ISO/IEC 29100, Information technology — Security techniques — Privacy framework