电 信 终 端 产 业 协 会 标 准

TAF-WG4-AS0026-V1.0.0:2018

移动终端基于 TEE 的人脸识别安全评估方法

Security Evaluation Method on TEE Based Mobile Device Face Recognition

2018 - 09 - 03 发布　　2018 - 09 - 03 实施

电信终端产业协会　　发 布

前　　言

本标准按照 GB/T 1.1-2009给出的规则起草。

本标准由电信终端产业协会提出并归口。

本标准起草单位：中国信息通信研究院、华为技术有限公司、北京小米科技有限责任公司、OPPO广东移动通信有限公司、维沃移动通信有限公司、蚂蚁金服信息技术有限公司、北京豆荚科技有限公司、安谋国际科技股份有限公司

本标准主要起草人：国炜、傅山、常新苗、王思善、周珏嘉、王江胜、郭子青

引　　言

近年来,随着人工智能领域和光学成像器件的突飞猛进,在移动终端上实现的本地人脸识别技术基于这些先进的算法和硬件基础快速发展,移动智能终端厂商纷纷在产品中加入本地人脸识别功能，进一步推动了人脸识别技术在移动金融、安防、移动互联网领域中的应用。与此同时，涉及移动智能终端的本地人脸识别的安全性问题也日渐凸显，如，使用相片或者视频进行欺骗，在终端上的人脸数据被恶意劫持等问题。目前行业中尚未有针对移动终端本地人脸识别功能的安全性的要求，对此没有统一的标准，基于上述考虑，提出移动智能终端人脸识别安全技术要求的标准，规范移动智能终端本地人脸识别的安全性要求，为国内该领域的相关产品的测评提供依据，来促进产业的健康稳定发展。

移动终端基于 TEE 的人脸识别安全评估方法

1 范围

本标准规定了对基于TEE的人脸识别系统的安全评估方法，包括对安全目标分析、安全威胁分析、安全功能要求、评估样本设定、评估流程和方法的设定等。

本标准适用于在本地进行特征比对的移动终端人脸识别系统。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GBT 29268.1-2012 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

TAF-WG4-AS0008-V1.0.0:2017 移动终端安全环境安全评估内容和方法

3 术语、定义和缩略语

3.1 术语定义

模板 template

系统中存储的注册用户的人脸特征及相关信息。

匹配得分 matching score

比对的输出结果，代表参与比对的两个人脸特征的相似程度；匹配或不匹配由该得分是否超过判定阈值来决定。

错误拒绝率 false reject rate

在验证识别过程中，真实者被错误判为拒绝的比例。

错误接受率 false accept rate

在验证识别过程中，冒充者被错误判为接受的比例。

欺骗接受率 spoof accept rate

在验证识别过程中，生物识别模型接受事先制作的已知良好样本的比例。

3.2 缩略语

TEE	Trusted Execution Environment	可信执行环境
TOE	Target of Evaluation	评估对象
RTC	Realtime Clock	实时时钟
ST	Security Target	安全目标
FAR	False Accept Rate	错误接受率

FRR	False Reject Rate	错误拒绝率
SAR	Spoof Accept Rate	欺骗接受率
SE	Secure Element	安全单元
RPMB	Replay Protected Memory Block	重放保护内存块

4 TOE 概述

4.1 TOE 范围

TOE 所针对的目标产品为依托硬件和软件实现人脸识别功能的终端设备。评估的范围包括目标产品用来提供人脸识别安全功能必需的所有相关的硬件、固件和软件部分。

4.2 系统整体框架

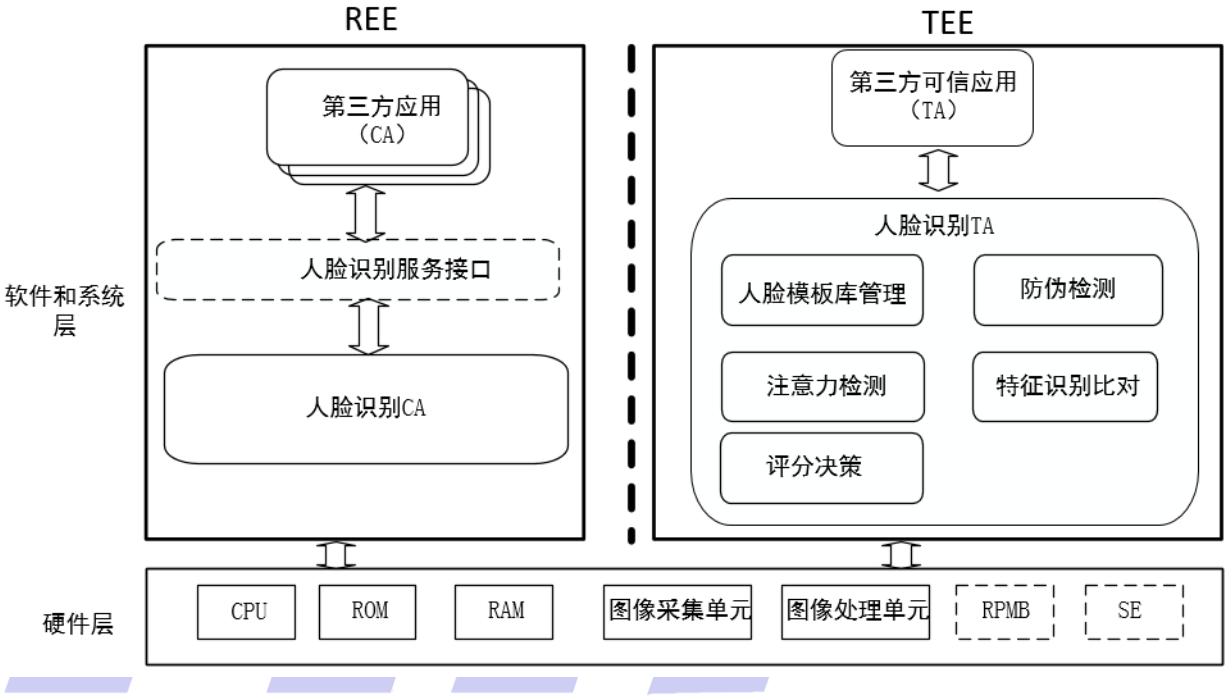

图 1 人脸识别系统整体框架

基于 TEE 的人脸识别系统整体框架如图 1，包括 REE 和 TEE 两部分执行环境，根据安全保护能力的不同，可以通过 SE 和 RPMB 来提升对数据的安全保护能力；其中，REE 侧提供采集传感器和人脸识别服务框架接口供第三方应用来调用，TEE 侧为人脸识别 TA、数据存储和数据通信等提供安全保护。

4.3 系统功能架构

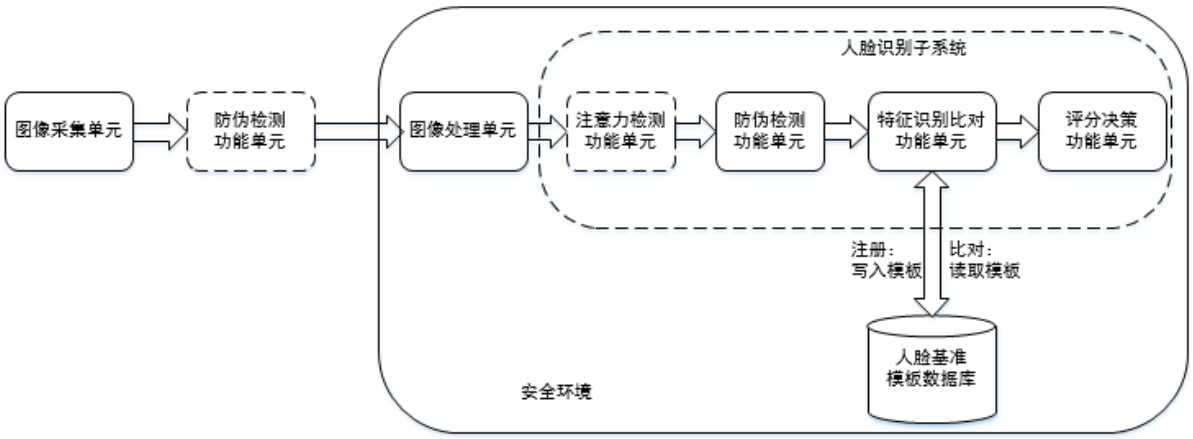

图2 人脸识别系统功能架构图

图像采集单元：负责执行用户人脸数据的光学信号采集工作，并转化为数字图像信息；

图像处理单元：负责将图像采集单元传递过来的数字图像信息进行如白平衡、色彩校正、编解码以及深度信息计算等信息处理的功能单元；

在图像采集单元和图像处理单元之间可以对原始数字图像信息进行防伪检测；

人脸识别子系统：包括注意力检测、防伪检测、特征识别比对和评分决策四个功能单元，负责对成像数据进行特征提取、防伪检测、比对打分，并最终给出决策结果；其中注意力检测功能单元用来检测被识别的对象的眼睛是否睁开且注视摄像头，注意力检测功能可以设置启动或者不启动。

人脸基准模板数据库：负责保存用户做人脸注册时生成的人脸模板数据。

5 TOE 保护资产

1) 人脸实时图像数据

人脸实时图像数据是人脸图像采集和识别系统在运行时产生的中间数据。

属性: 一致性，机密性

2) 人脸基准模板

人脸基准模板是在用户人脸注册过程中创建的，它们是持久化数据。它们在人脸特征识别比对过程中使用。人脸特征参考值与它们唯一的标识符应一同存储。

属性：完整性、回滚保护、真实性、机密性、设备绑定

3) 人脸匹配得分

人脸匹配得分是一种实时数据.在人脸特征值识别比对的过程中，识别比对功能单元基于人脸特征参考值和人脸特征取样进行对比来计算匹配得分.

属性：一致性，机密性。

4) 人脸识别结果

这是一种运行时数据，由决策功能单元根据从识别比对功能单元接收到的人脸特征匹配得分给出人脸识别结果（二进制答案，即匹配/不匹配）

属性：一致性。

5) 人脸识别系统代码

与人脸识别系统相关的代码，至少包括人脸采集设备的驱动程序和与人脸识别功能相关的算法，包括用于人脸特征采集和人脸特征验证的代码等。在 TEE 中运行，通过 TEE 进行保护。这些数据是持久化数据。

属性：完整性，一致性，真实性，防回滚保护。

6) 采集功能代码

采集设备相关的代码，是指采集人脸特征样本的传感器的代码。与实际使用的传感器相关。

属性：一致性，真实性。

7）算法配置数据

这种数据是持久化的,包括比对策略和阈值, 用于识别比对时人脸特征识别参考和取样的匹配过程.

属性:完整性,真实性.

8）加密密钥

这个密钥用以保护人脸基准模板，通常是持久化数据。

属性：完整性，机密性，可用性。

6 安全威胁和安全目标

6.1 安全威胁分析

1） 采集环节

采集的人脸实时数据被拦截或者被篡改；

采集传感器的固件完整性和可用性被破坏。

2） 传输环节

人脸数据在传输过程中被窃取或者被篡改；

恶意样本在传输环节被注入。

3） 存储环节

存储的模板数据密钥被破解或者被窃取

存储的模板数据在更新时被破坏；

存储的模板数据被篡改；

存储的人脸处理中间数据被篡改；

人脸识别算法相关的 AI 模型的哈希值被篡改

4） 特征比对环节

伪造的人脸未能被识别，骗过识别系统；

大量的样本数据攻击，导致误识率提升；

特征比对结果在 REE 侧被篡改；

5） 数据销毁环节

在注册的人脸模板被用户删除后，关联数据未彻底清除或未有防回滚的防护，导致数据被窃取后作为假冒数据或者攻击样本。

6.2 安全目标

人脸识别系统的安全目标是人脸系统所具有的安全功能可以防御 6.1 节描述的安全威胁，保证在 TEE 上运行的人脸识别系统的应保护的资产的完整性、机密性和可用性。

人脸识别系统的总体安全目标如下：

采集设备传感器固件及驱动的完整性、可用性和功能接口的授权访问；

图像处理单元的固件及驱动的完整性、可用性和功能接口的授权访问

人脸数据（包括人脸实时图像数据、人脸基准模板等）应能防窃取、防篡改；

人脸算法配置数据（策略、阈值）应能防篡改；

通信信道（图像采集单元和图像处理单元之间，以及图像处理单元与应用处理器之间）应能防止数据窃取、防止恶意样本注入；

硬件和软件接口应实施访问控制，防止非法应用调用；

人脸识别子系统软件代码、人脸模板等应防止回滚、非法篡改，避免关键防伪检测、验证功能被旁路。

7 TOE 安全功能要求

7.1 采集环节

基础安全功能要求：

a) 传感器固件应有完整性和可用性保护；

b) 传感器驱动程序应有完整性和可用性保护；

c) 传感器采集功能应实施访问权限控制，防止人脸源数据被窃取；

d) 传感器无残留的人脸源数据；

e) 图像处理单元的固件应有完整性和可用性保护

f) 图像处理单元应保证启动镜像安全，工作在安全模式下；

g) 图像处理单元所使用的内存应受 TEE 安全保护或者使用单独物理隔离的内存，不应使用非安全环境的内存；

h) 在用户人脸模板注册时，应先对用户进行身份验证才允许进行人脸采集，且应针对采集的次数限制、超时处理等情况制订管理策略。

7.2 传输环节

基础安全功能要求：

a) 图像处理单元和应用处理器应通过安全信道保护数据通信；

b) 图像处理单元和应用处理器之间的通信应具备有效的安全机制，确保不被重放攻击；

增强安全功能要求：

a) 从图像处理单元开始对图像信息进行处理开始，所有人脸系统的数据传输都应通过安全信道传输，不应经过 REE。

b) 图像处理单元生成的人脸数据应使用从 RTC 设备获取的安全时间戳。

7.3 存储环节

基础安全功能要求：

a) 人脸模板应加密保存，加密密钥应确保每个设备唯一；

b) 人脸模板应采用授权访问读取；

c) 人脸数据应采用有效安全机制，防止被篡改；

d）人脸数据不能在安全内存残留；

e) 人脸模板数据的更新应在 TEE 中进行；

增强安全功能要求：

a) 人脸加密密钥应在安全单元 SE 中存储；

b) 人脸模板数据应具备有效的安全机制，防止数据回滚，如：RPMB

7.4 识别比对环节

识别比对环节的安全功能要求依赖于 8.2 节所设定的测试集。本章节对识别比对环节中影响人脸识别系统安全的识别率（FAR,FRR）指标和防伪率（SAR）指标提出如下要求：

表 1 识别比对环节安全指标要求

要求分级	FRR	FAR	2D SAR	3D SAR			
				A 测试集	B 测试集	C 测试集	D 测试集
基础功能	小于 10%	小于万分之一	小于 15%	N/A	N/A	N/A	N/A
增强功能一	小于 7%	小于万分之一	小于 5%	小于 20%	小于 40%	N/A	N/A
增强功能二	小于 5%	小于五万分之一	0	小于 7%	小于 15%	小于 30%	N/A
增强功能三	小于 5%	小于百万分之一	0	小于 3%	小于 7%	小于 15%	小于 30%

7.5 销毁环节

基础安全功能要求：

a) 还原所有设置时，要销毁人脸注册模板数据；

b) 恢复出厂设置时，要销毁人脸注册模板数据；

c) 当移动终端被设置为无密码锁屏时，要销毁人脸注册模板数据；

增强安全功能要求：

a) 被销毁的数据应采用有效的安全机制，防止数据回滚

8 TOE 评估流程和方法

8.1 评估流程

人脸识别系统安全评估将遵循以下流程对目标产品进行完整的安全评估。

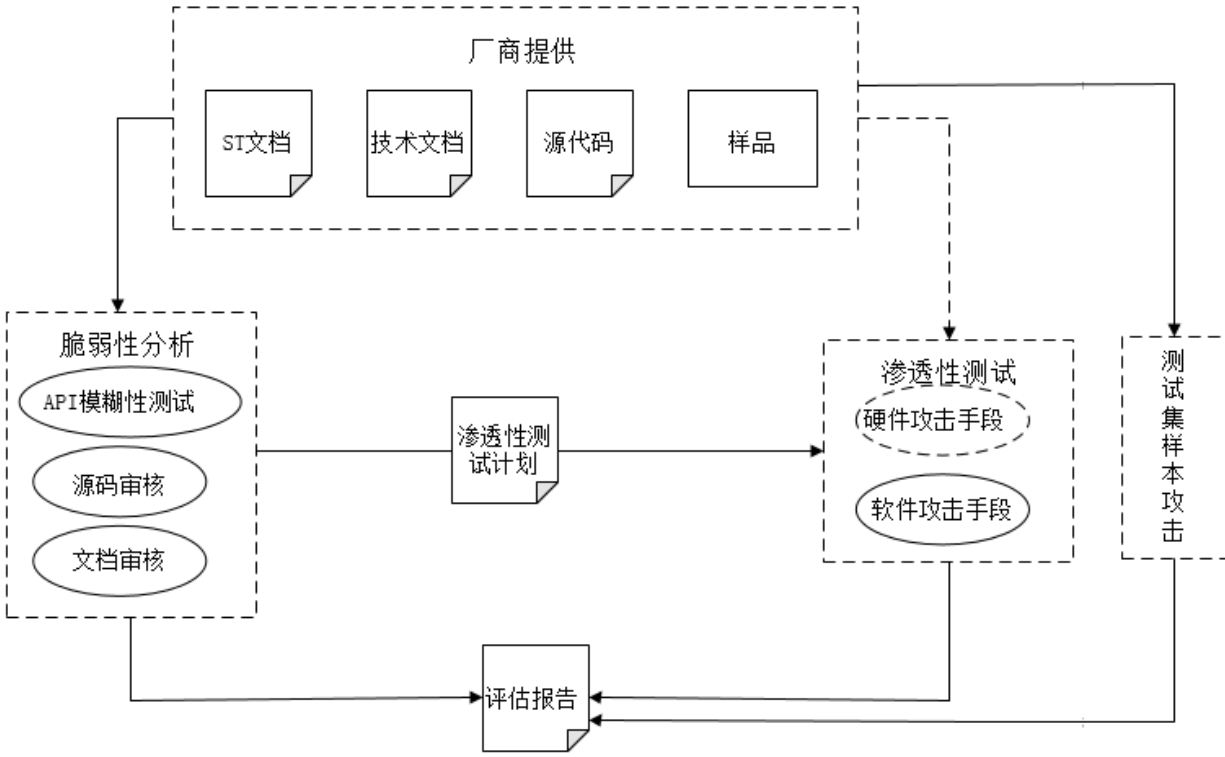

图3 安全评估流程

厂商需要提交包括ST文档、技术文档、相关源代码和被测样品。

脆弱性分析的目标是实验室评估人员根据厂商提交的ST文档，并按照安全功能要求所规定的，对厂商提交的设计文档和相关源码进行初步的方案审核，发现可能的脆弱点和其影响，并由此定义后续渗透性测试的计划和测试例。脆弱性分析的输出作为最终评估报告的一部分。

渗透性测试是实验室根据前期所发现产品的脆弱性，在一定时间内开展对这些脆弱点进行攻击的测试。实验室根据脆弱性分析的结果制定渗透性测试路径以及测试例并实施渗透性测试。在安全评估报告中，要提供实施渗透性测试的环境、产品目标的配置和每一个渗透性测试的结果，同时还要具体地描述实施过程中的步骤。

根据上述安全评估流程，对人脸识别安全系统的安全能力进行评估分级。

8.2 评估方法

8.2.1 人脸数据处理安全评估方法

对人脸识别系统中人脸数据处理（包括采集、传输、存储、比对和销毁）环节的安全性评估，参考TAF-WG4-AS0008-V1.0.0:2017移动终端安全环境安全评估内容和方法，对第7章描述的相关安全功能进行测试评估。

8.2.2 人脸系统识别率和防伪测试集要求

测试集是用来对人脸识别系统防伪能力和防误识能力进行评估的重要工具。

（一）识别率测试集要求

本章节定义的识别率测试集是用来对7.4节中的FAR和FRR指标进行评测，以2D相片构成，；

测试集应包括四种肤色人种的相片：黄色、白色、黑色和棕色；

测试集中相片数量应符合性别男：女=1:1；

测试集可按照肤色和性别的组合构成8个测试子集，每个测试子集中至少应采集500个人的相片，500个人的年龄应该从青少年、中年到老年广泛覆盖，每人采集不少于一张相片，且每张相片中只有一张人脸；

测试集中相片采集时的光线条件应包括：逆光、正面光、90度侧光，建议采集张数的比例为：1:2:1；

测试集中按照脸部的偏航角角度分为A，B两个子类，对测试集中样本脸部的偏航角角度的说明参见附录1。A子类中，相片中人脸的偏航角角度在-30度到+30度之间的样本占比98%，用于当注意力检测功能开启时的识别率测试；另应采集偏航角角度在-90度到-31度和31度到90度的样本占比2%，用于FAR测试。 B子类中，相片中人脸的偏航角角度在-45度到+45度之间的样本占比98%，用于当注意力检测功能关闭时的识别率测试；另应采集偏航角角度在-90度到-46度之间和46度到90度之间的样本共占比2%，用于FAR测试。

测试集中按照脸部的俯仰角角度分为A，B两个子类，对测试集中样本脸部的俯仰角角度的说明参见附录A。A子类中，相片中人脸的俯仰角角度在-30度到+15度之间的样本占比98%，用于当注意力检测功能开启时的识别率测试；另应采集俯仰角角度在-90度到-31度和16度到60度的样本占比2%，用于FAR测试。 B子类中，相片中人脸的俯仰角角度在-45度到+15度之间的样本占比98%，用于当注意力检测功能关闭时的识别率测试；另应采集俯仰角角度在-90度到-46度之间和16度到90度之间的样本共占比2%，用于FAR测试。

测试集中相片的分辨率应不低于640*480像素；

测试集中相片采集时应包括有遮挡（墨镜、口罩、帽子等）和浓妆的相片，遮挡和浓妆的总数据量应在测试集中至少有2%的比例。

（二）2D防伪测试集要求

2D防伪测试集以相片和视频库构成，用来对7.4节中的2D SAR指标进行评测。

相片和视频样本的共性要求如下：

测试集应包括四种肤色人种：黄色、白色、黑色和棕色；

测试集可按照肤色和性别的组合构成8个测试子集，

测试集中样本采集数量应符合性别男：女=1:1；

测试集中样本采集时的光线条件应包括：逆光、正面光、90度侧光，建议采集样本中不同光线条件的比例为：1:2:1；

测试集中样本采集时的脸部偏航角角度分为A，B两个子类，对测试集中样本脸部的偏航角角度的说明参见附录1。A子类中，样本中人脸的偏航角角度在-30度到+30度之间的样本占比98%，用于当注意力检测功能开启时的情况；另应采集偏航角角度在-90度到-31度和31度到90度的样本占比2%。 B子类中，样本中人脸的偏航角角度在-45度到+45度之间的样本占比98%，用于当注意力检测功能关闭时的情况；另应采集偏航角角度在-90度到-46度之间和46度到90度之间的样本共占比2%。

测试集中样本采集时的脸部俯仰角角度分为A，B两个子类，对测试集中样本脸部的俯仰角角度的说明参见附录A。A子类中，样本中人脸的俯仰角角度在-30度到+15度之间的样本占比98%，用于当注意力检测功能开启时的识别率测试；另应采集俯仰角角度在-90度到-31度和16度到60度的样本占比2%，用于FAR测试。 B子类中，样本中人脸的俯仰角角度在-45度到

+15度之间的样本占比98%，用于当注意力检测功能关闭时的识别率测试；另应采集俯仰角角度在-90度到-46度之间和16度到90度之间的样本共占比2%，用于FAR测试。

测试集中样本采集时应包括有遮挡（墨镜、口罩等）和浓妆的视频，遮挡和浓妆的总数据量应在测试集中至少有2%的比例。

对相片样本，除上述共性要求外，还应遵循如下要求：

每个测试子集应至少包括十个人；

相片若打印，应以不小于A3纸尺寸大小打印，打印材质包括相片纸和打印纸；

相片打印设备应包括激光和喷墨，比例为1：1；

相片应包括黑白和彩色两种，比例为1:1；

相片采集设备应包括可见光和近红外两种；

打印出的相片，除完整纸张外，还应包括沿头型剪裁的相片，还应包括抠掉鼻子、眼睛或者嘴的相片；

相片若不打印，应是手机等电子设备展示的高清真人相片；

对视频样本，除上述共性要求外，还应遵循如下要求：

每个测试子集中至少应采集100个人的视频，100个人的年龄应该从青少年、中年到老年广泛覆盖，每人采集不少于一段10s的视频；

测试集中视频的分辨率应不低于640p，且应至少有三分之一达到1080p；

测试集中的视频应能在多种设备上播放，包括但不限于：手机、平板、电脑和电视机等；且必须包括播放时人头大小与真人人头大小达到1:1的播放设备。播放设备的分辨率应能支撑上述对视频采集的分辨率要求。

（三）3D防伪测试集

3D防伪测试集以3D面具/假人头构成，用来对7.4章节中的3D SAR指标进行评测；

测试集中的样本制作应综合考虑不同的制作材质、制作数据源和制作精度，综合这三个维度，至少应包括四个分类的测试子集。

每个类别的测试子集中应至少包括20个测试样本。

表 2 3D 防伪测试集分类

分类	材质	数据源	制作精度
A 测试集	纸质	2D 照片（1 张或多张）	小于 10 毫米
B 测试集	石膏、石英砂	2D 照片、3D 重建模型（RGB）	小于 5 毫米
C 测试集	硅胶、树脂	2D 照片、3D 重建模型（RGB）	小于 5 毫米
D 测试集	类肤质	2D 照片、3D 重建模型（RGB +Depth）	小于 1 毫米

8.3 安全能力分级

本标准将人脸识别的评估等级划分为依次递增的四个安全等级，不同的安全等级将采用不同的安全评估手段并覆盖第7章中规定的不同的安全功能要求。

一级：依据基础安全功能要求，对厂商提供的文档和相关源码进行审核，并结合API接口的模糊性测试分析产品安全漏洞，使用识别率测试集和2D防伪测试集进行攻击；最终给出评估报告；

二级：依据基础安全功能要求和部分增强安全功能要求，对厂商提供的文档和相关源码进行审核，并结合API接口的模糊测试分析产品安全漏洞，根据安全漏洞设计渗透性测试攻击路径和测试例实施软件的渗透性攻击，使用识别率测试集、2D防伪测试集进行攻击；最终给出评估报告。

三级：依据基础安全功能要求和部分增强安全功能要求，对厂商提供的文档和相关源码进行审核并结合API接口的模糊测试分析产品安全漏洞，根据安全漏洞设计渗透性测试攻击路径和测试例实施软件的渗透性攻击，使用识别率测试集、2D防伪测试集、3D防伪测试集进行攻击；最终给出评估报告。

四级：依据基础安全功能要求和部分增强安全功能要求，对厂商提供的文档和相关源码进行审核并结合API接口的模糊测试分析产品安全漏洞，根据安全漏洞设计渗透性测试攻击路径和测试例实施软件的渗透性攻击,使用识别率测试集、2D防伪测试集、3D防伪测试集进行攻击；最终给出评估报告。

表 3 人脸识别系统安全评级对照表

安全功能		一级	二级	三级	四级
采集环节	基础功能要求	文档审阅； API模糊性测试	文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试
传输环节	基础功能要求	文档审阅 API模糊性测试	文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试
	增强功能要求			文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试
存储环节	基础功能要求	文档审阅； API模糊性测试	文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试	文档审阅； API模糊性测试； 渗透性测试
	增强功能要求a)				文档审阅； API模糊性测试；

					渗透性测试;
	增强功能要求b)			文档审阅; API模糊性测试; 渗透性测试	文档审阅; API模糊性测试; 渗透性测试
识别比对环节	基础功能	文档审阅; API模糊性测试; 防误识测试集攻击			
	增强功能一		文档审阅; API模糊性测试; 防误识测试集攻击; 2D防伪测试集攻击		
	增强功能二			文档审阅; API模糊性测试; 渗透性测试; 防误识测试集攻击; 2D防伪测试集攻击; 3D防伪测试集攻击	
	增强功能三				文档审阅; API模糊性测试; 渗透性测试; 防误识测试集攻击; 2D防伪测试集攻击; 3D防伪测试集攻击
销毁环	基础功能要求	文档审阅; API模糊性测试	文档审阅; API模糊性测试;	文档审阅; API模糊性测试;	文档审阅; API模糊性测试;

节			渗透性测试	渗透性测试	渗透性测试
	增强功能要求			文档审阅； API 模糊性测试； 渗透性测试	文档审阅； API 模糊性测试； 渗透性测试

附 录 A
（资料性附录）
文档编写记录

修订时间	修订后版本号	修订内容
2018.7.11	V1.0.0	全文格式

附　录　B
（资料性附录）
测试集人脸样本采集角度说明

人脸样本的采集角度是指在相机位置不变的情况下，呈现在相机中的人脸在三维坐标系内相对于坐标轴的旋转角度。

下边将就如何根据相机和人脸的位置构建人脸三维坐标系和三维坐标系内的角度定义做出详细说明。

（一）人脸三维坐标系

被采集人直立于水平地面，双眼平视，正面面向相机。相机垂直于水平地面，确保人脸可以完整清晰在相机中成像，且人脸和相机之间距离宜在30-40cm之间。此时保持相机不动，以被采集人的鼻尖为原点构建一个三维坐标系，其中xz平面与水平地面平行，y轴垂直于水平地面，z轴方向是视线方向，如图4所示。

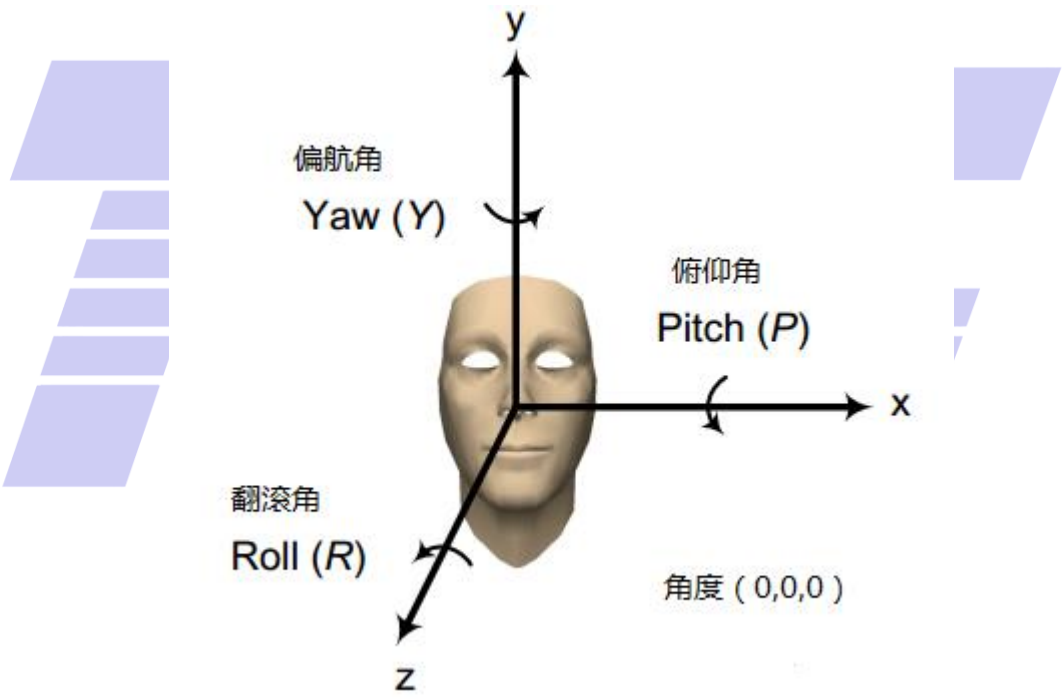

图4　人脸三维坐标系

偏航角角度（Yaw）是指围绕着y轴旋转形成的角度，以人脸向左转动记为正向角度，通俗来说，yaw角度用来描述被采集人相对相机左右转动头部而带来的角度偏转；

俯仰角角度（Pitch）是指围绕着x轴旋转形成的角度，以朝上转动记为正向转动。通俗来说，pitch角度用来描述被采集人相对相机进行低头抬头动作带来的角度偏转。

翻滚角角度（Roll）是指围绕着z轴旋转形成的角度，以向左转动记为正向转动；通俗来说，roll角度用来描述被采集人始终保持正面面对相机时进行头部左右摆动带来的角度偏转。

（二）人脸样本偏航角角度扭转示例

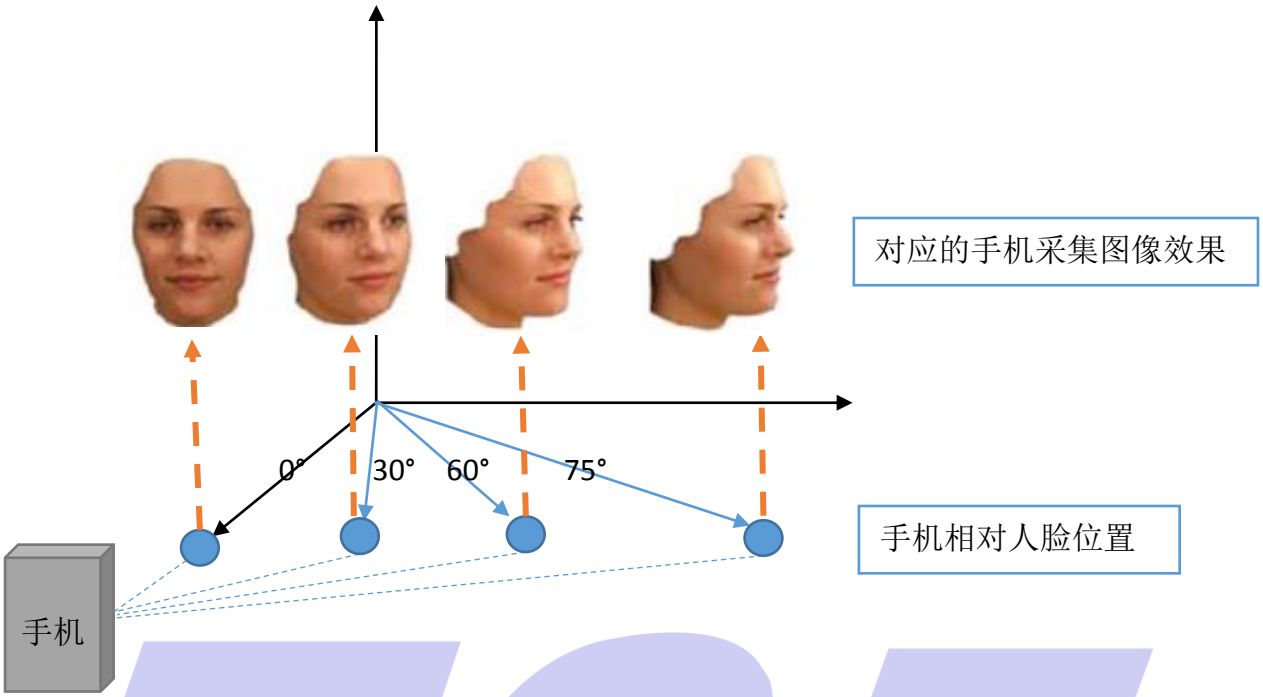

图5 人脸样本偏航角角度转动示例（三）人脸样本俯仰角角度扭转示例

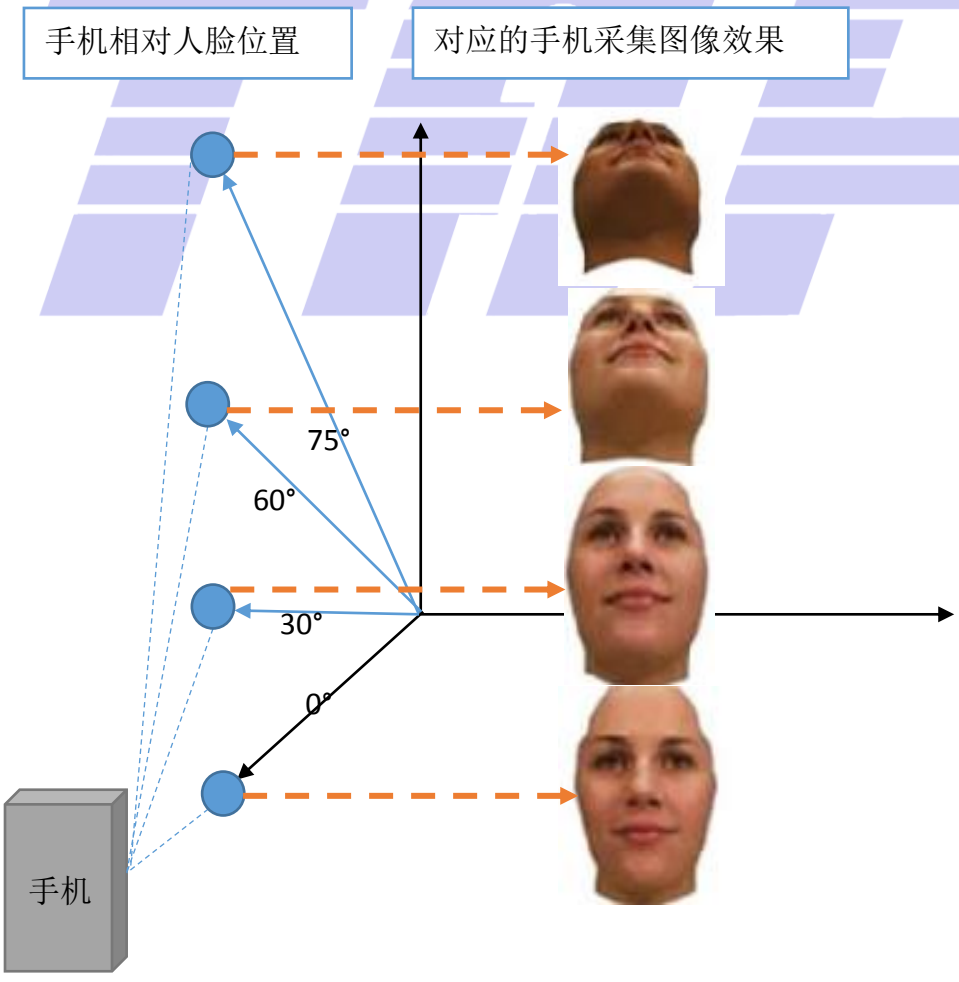

图6 人脸样本俯仰角角度示例

（四） 人脸样本翻滚角角度示例

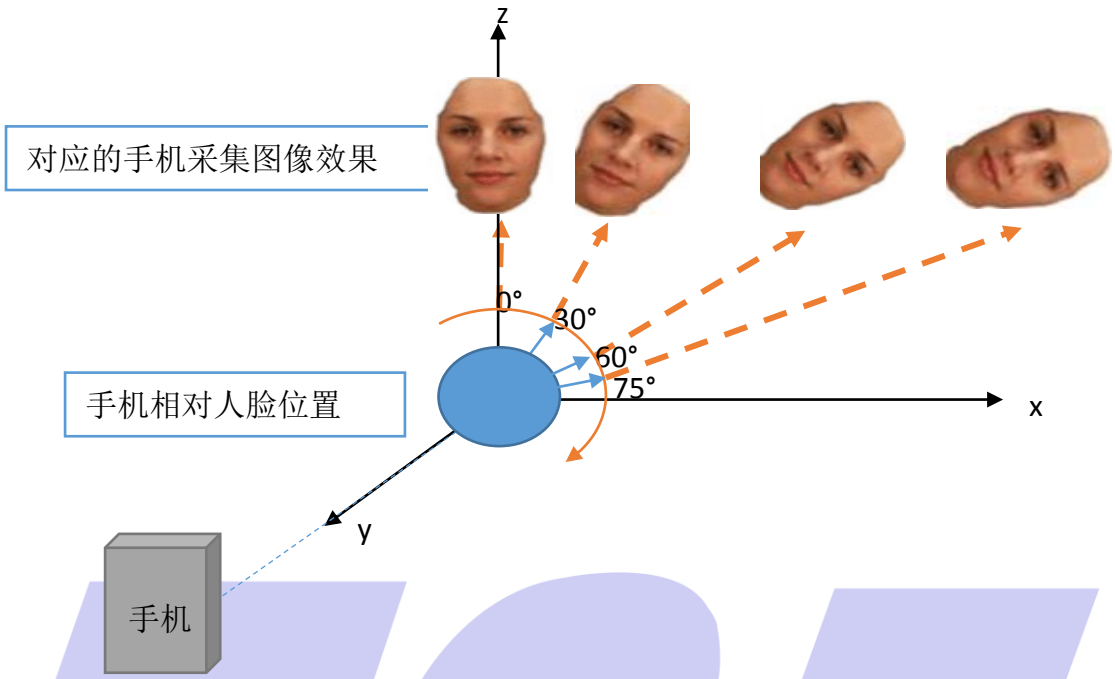

图7 人脸样本翻滚角角度示例

参 考 文 献

[1] GB/T 26237.5-2014 《信息技术 生物特征识别数据交换格式 第 5 部分：人脸图像数据》；