**清华大学-腾讯互联网创新技术联合实验室**

**科研合作项目建议书**

**填写要求：简洁、清晰、明了**

1. 项目名称： 基于唇动的活体检测和安全人脸认证系统

研究领域： 网络安全

1. 项目概述（*请****简要****介绍项目的背景、内容、意义及相关工作的调研或基础*）
2. 项目背景：

借助普通的摄像头和专业软件，人脸识别认证为移动设备（例如智能手机）的身份验证提供了一种简单而有效的方式。随着AI芯片（例如海思麒麟970）的推出和移动设备计算能力的不断提升，移动端的人脸识别无论在识别准确率还是识别速度上都达到了非常好的效果。相较于其他生物特征（例如指纹，虹膜），攻击者更容易获取和复制人脸，然后伪造包含受害者人脸的图片和视频，欺骗人脸识别认证系统。特别是随着社交软件（例如微信，微博，抖音）的流行，攻击者很容易从社交网络上获取受害者大量照片和面部视频，这给移动端的人脸识别认证带来了潜在的安全威胁。

为了应对这些基于伪造人脸的安全威胁，各种活体人脸检测方法被提出来用于区分合法用户的真实人脸和攻击者伪造的虚假人脸。然而随着人工智能（AI）技术的快速发展，攻击者可以利用AI技术进行高效的人脸合成，实时模仿各种面部表情并将其应用于受害者的3D人脸模型中，从而通过现有大多数人脸认证活性检测方法。另一方面，近年来研究者发现机器学习算法（特别是神经网络）容易受到对抗样本的攻击，即攻击者向原始图像中添加少量精心设计的人眼难以察觉的微小扰动生成对抗样本图像，能够使得机器学习算法产生误判。利用对抗样本方法，攻击者精心设计不易引起警觉的脸部佩饰物（例如眼镜）或者妆容，从而使得人脸认证系统（通过机器学习算法进行分类）将攻击者识别为目标用户或者逃避人脸识别系统的检测。由于这类攻击拍摄到的人脸图像本身就来自攻击者的真实人脸，所以自然能通过活体人脸检测，这给现有的活体人脸检测方案带来了巨大挑战。

现有的活体人脸检测方案从原理上说主要分为三类，均无法防御逃避活体的攻击。

1. 基于交互的检测方法。认证者根据要求做出眨眼，转头，微笑等各种表情或者根据指示说出一段特定话语。为了进一步抵御基于视频重放的攻击方法，这种方法可以随机要求认证者做出各种表情的顺序，或者随机指示话语内容。这种方法的优点是成本较低，可以在大多数移动设备上实现，能够抵御基于照片的攻击方法和一般的视频重放攻击。然而随着人工智能（AI）技术的快速发展，攻击者可以利用AI技术实时模仿各种面部表情并将其应用于受害者的3D人脸模型中，伪造人脸认证视频，或者根据指示实时合成伪造具有认证者声纹特征的特定话语，从而通过活性检测。
2. 基于3D脸型的检测方法。由于真实人脸具有3D脸型结构，基于照片和视频的人脸攻击方法所呈现的人脸是平面结构的，缺乏立体感。认证设备利用光流法，传感器辅助等方法检测认证人脸是否具有立体结构。这种方法优点是对基于照片和视频的人脸攻击方法都具有很好的抵御效果，然而光流法需要在光照条件很好情况下拍摄高质量照片，传感器辅助的防御已经被基于VR设备的3D攻击攻破。而且基于3D脸型的检测方法无法防御3D人脸攻击，例如硅胶人脸面具攻击，基于对抗样本的假冒攻击。
3. 基于纹理和反射率分析的检测方法。由于照相机、打印机、屏幕显示器以及硅胶面具等人脸攻击工具工艺上的不完美，它们在获取或者展现人脸时难免丢失一些微观纹理细节，或者存在色彩和反射率的细微差异。这种方法分析人脸的色彩和微观纹理特征，以及提取反射率细节，从而判断是否为真实人脸。这种方法优点是对基于照片、视频、3D定制模型的人脸攻击方法都具有一定的抵御效果，但是随着高保真度的照相机，高分辨率的屏幕显示器和3D打印机的发展，利用这种方法判断是否为真实人脸的难度就会越来越大了。
4. 项目内容：

本项目提出一种智能手机上的基于唇动的安全人脸认证活性检测系统—LipFace，在用户进行人脸认证的同时，要求用户默念预先设定的密语（不限定内容）。手机的扬声器发出超声波信号，麦克风同时接收唇部反射的含有运动信息的超声波信号，用于进行活性检测。由于基于视频或者虚拟现实技术的人脸伪造方法都不包含用户物理的唇部运动，因此麦克风接收的超声波信号不包含用户唇部运动信息，无法通过活性检测。另外，研究表明每个人的唇部运动具有独特性，佩戴基于对抗样本精心设计的脸部佩饰物的攻击者也无法模拟用户独特的唇部运动，因此也不能通过唇动检测因而无法通过人脸认证。

此外，本项目提出了一种自动安全的人脸认证活性检测机制，能检测存在对抗攻击性的恶意人脸图像。不需要用户额外提供多余的信息，认证系统能区别是恶意的人脸图像还是正常的人脸，并对有意欺骗认证系统的恶意图像进行修复处理，从而能准确识别攻击者的身份，增强人脸认证系统的安全性。在保证合法用户利益不受侵犯的同时，能及时定位有作恶倾向的攻击者。

1. 项目目标（*请填写项目的目标、产出及其考核指标与方式、对腾讯/业界/社会的价值*）
2. 项目目标：

借助智能手机配备的普通摄像头、扬声器和麦克风，为智能手机提供一种简单实用，安全有效的人脸活性检测方案，并且可以准确认证相应的用户。该方案可适用于普通智能手机，不需要配备额外硬件设备，能够抵抗现有各种先进的人脸伪造攻击。

项目总体目标是输出以下三种活体检测的基础能力：

(1) 能够有效地区分平面和立体（平面攻击和真人形状）;

(2) 能够有效地区分嘴唇是否运动;

(3) 能够通过嘴唇运动判断用户是否读指定数字。

1. 项目产出及其考核指标与方式：
2. 交付内容：

a. 超声波生成，发射及采集SDK， 使用sdk的Android app demo及源码

b. 信号处理API，源码及使用说明（信号处理负责将原始数据处理为神经网络的输入数据）

c. 模型结构，参数以及训练测试设计的源码

1. 设备兼容性：
   1. 兼容Android 4.4（KitKat）及以上操作系统（可覆盖市面上96.2% 安卓系统 https://developer.android.com/about/dashboards），适配华为，OPPO，VIVO，小米等主流高低端机型；
   2. 系统设计可兼容iOS操作系统；
2. 交互方式：
   1. 兼容优图人脸核身活体检测产品（包括静默活体，四字活体，动作活体，反光活体）的用户交互方式；
   2. 交互时间（从启动活体检测到显示检测结果，包括与服务器交互时间）不超过2秒；
3. 在优图协助下建立LipFace活体检测的评测数据集：
   1. 正样本10000组
   2. 平面攻击（屏幕翻拍，打印纸片，冲印照片，包装盒等）20000组：攻击方式包含扭曲，抖动等干扰）
   3. 面具攻击（纸片和硅胶面具）10000组：攻击方式包含裁剪和扣洞面具，露出五官
   4. 3D头模10000组
4. 算法指标：
   1. 用户静默：正样本通过率不低于99%，平面攻击的误通过率低于0.1%，面具攻击的误通过率低于1%；
   2. 用户读数字：正样本通过率不低于97%，平面攻击和3D头模的误通过率低于0.1%，面具攻击的误通过率低于1%；
5. 鲁棒性：
   1. 消除用户晃动，或者手机抖动的干扰；
   2. 兼容用户手持手机的距离和角度的影响（支持用户正常交互下使用，约16cm~24cm）；
   3. 兼容手机发射的载波频率不理想的影响；
   4. 兼容嘴唇形状，方言等造成的个体差异；
6. 完成一篇顶级国际会议论文投稿以及一项专利；
7. 2-3名学生在腾讯实习完成相关技术的开发与验证。
8. 项目价值：

本项目能够为现有大多数智能手机提供安全可靠的人脸识别认证活性检测方案，不需要任何额外的硬件设备，具有很强的推广价值，特别对腾讯应用等远程认证具有很强的应用价值。

1. 研究内容（*请****简要****填写项目的主要研究内容*）

本项目研究人脸认证过程的安全性问题，测试现有人脸认证活性检测方案对各种人脸伪造技术的抵御能力，并且提出一种智能手机上的基于唇动的安全人脸认证活性检测系统—LipFace。在用户进行人脸认证的同时，我们要求用户默念预先设定的密语（不限定内容），手机的扬声器发出超声波信号，麦克风同时接收唇部反射的含有运动信息的超声波信号，用于进行活性检测。由于基于视频或者虚拟现实技术的人脸伪造方法都不包含用户物理的唇部运动，因此麦克风接收的超声波信号不包含用户唇部运动信息，无法通过活性检测。另外，研究表明每个人的唇部运动具有独特性，佩戴基于对抗样本精心设计的脸部佩饰物的攻击者也无法模拟用户独特的唇部运动，因此也不能通过活性检测。

1. 技术方案（*请****简要****填写项目的主要技术方案及人力投入*）
2. 主要技术方案：

人脸伪造技术，我们重点研究基于对抗样本的假冒攻击。利用对抗样本方法，攻击者精心设计不易引起警觉的脸部佩饰物（例如眼镜）或者妆容，从而使得人脸认证系统（通过AI算法进行分类）将攻击者识别为目标用户或者逃避人脸识别系统的检测。攻击者首先打印一幅添加过扰动的眼镜框，接着佩戴打印后的眼镜框并由人脸认证系统验证，最终佩戴眼镜的攻击者可以冒充受害者，从而通过人脸认证系统的验证。如图1所示，左图表示佩戴眼特制眼镜的攻击者，右图表示对应的攻击目标，即左图的攻击者会被人脸识别系统识别成右图的攻击目标。由于很多人配戴眼镜，因此这种攻击很难被人察觉。

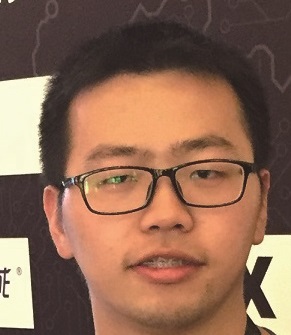
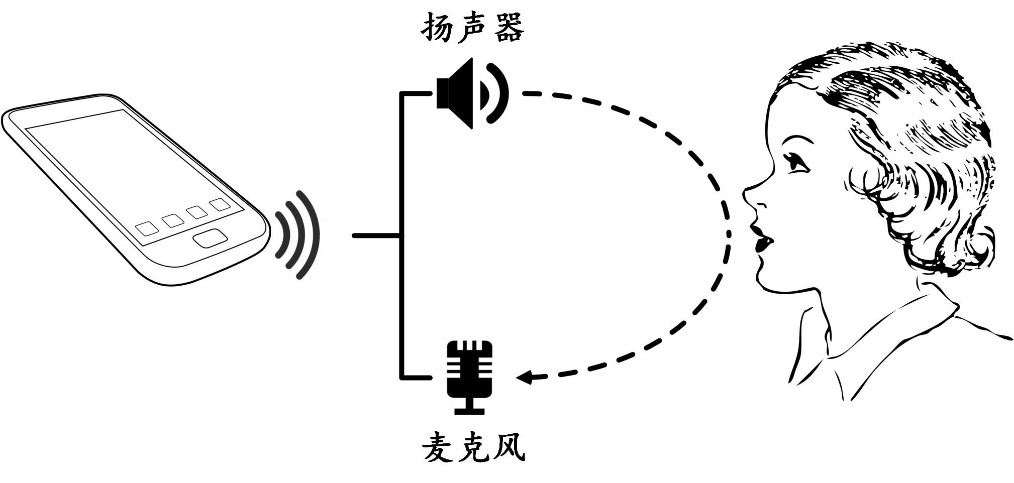
 

图1：人脸识别系统的攻击示意图

针对人脸伪造的防御方案，我们利用唇部运动的独特性设计一种智能手机上的基于唇动的安全人脸认证活性检测系统—LipFace。本系统利用扬声器发射人耳无法感知的超声波信号，并用麦克风接收经过嘴唇反射回来的信号（如图2所示）。由于反射信号与原信号相比存在相位偏移，这个相位偏移是由传播延迟造成，当嘴唇运动时，会改变反射路径的长度，也就是改变传播延迟，这种改变进而会引起相位偏移的变化，提取出这个随嘴唇运动而变化的相位偏移就可以提取出嘴唇的运动信息，将这个信息与数据库中的用户预先设定密语的唇部运动信息进行对比就可以进行身份认证。麦克风除了接收到嘴唇反射回来的超声波信号之外，还会接收环境噪音。由于环境噪音通常不超过8 kHz，因此环境噪音很容易利用频率的差异性消除掉。

图2：利用智能手机上的扬声器和麦克风感知嘴唇运动



一般来说，每一个音节（汉语是单音节语言，每一个字就是一个音节）都涉及到一种嘴唇形状，但一种嘴唇形状可能和多个音节对应，常见的嘴唇形状如图3所示。嘴唇运动依靠的是许多骨骼以及肌肉的协同运动，由于不同的人有不同的骨骼结构和肌肉机构，这就导致了即使发同一个音，不同用户所呈现出的嘴唇形状也会有细微差异，而且不同的人有不同的说话习惯和说话方式。这些因素使得嘴唇运动具有个体差异性，尽管人眼很难发现这种个体差异性，但是利用机器学习通过高精度的数据分析可以发现这种差异性。

利用声纹认证或者用户在进行人脸认证时根据指示说出一段特定话语进行语音识别也是一种防御攻击的手段，然而这种方法同样具有被攻击的巨大风险。首先由于语音信道的开放性，现有的语音识别系统存在很大的安全隐患——容易遭受重放攻击和妥协攻击；同时，由于环境噪声的干扰，在低信噪比的环境中，系统识别准确率会降低。重放攻击中，攻击者利用提前录制的受害者语音重放通过语音识别和声纹认证，我们可以通过引入随机性，每次指定用户说话内容来抵御重放攻击。但是同样由于社交网络的发展，攻击者更加容易获取受害者的语音数据，妥协攻击通过分割，分析受害者之前的语音数据，根据指示实时合成伪造具有受害者声纹特征的特定话语，从而通过检测。本项目利用主动声呐系统通过超声波信号获取用户唇动信息，首先由于超声波信号和环境噪音频率分布不同，因此可以避免 环境噪声的干扰，我们通过每次改变超声波信号（唇动信号的载波）频率引入随机性防御重放攻击。我们要求用户认证时默念预先设定的密语，因此密语内容很难被窃听泄露，由于唇动信号的载波是超声波信号，不会在社交网络上传播，因此攻击者很难获取受害者之前的唇动信号数据，也就无法实时合成伪造具有受害者唇动特征的信号，因此我们的系统也能抵御妥协攻击。

基于活体信息构架可以实现安全的认证机制，通过初始阶段预设唇动可实现有效的用户认证。基于唇动的唯一性特性即可实现比传统基于图像更安全的人脸认证机制。活体识别和认证过程可以部署在本地或者云端服务器上。声音信号与图像信号相比，无论是数据量还是数据精度都存在数量级的差距，这使得对声音信号的处理不需要如图像处理那样巨大的计算开销，而且随着各种移动设备计算能力的提升，识别过程的本地部署是可行的。另外，本系统用户体验好，用户通过说话即可同时完成语音信息和嘴唇运动信息的输入，不需要用户进行繁琐的操作，也不需要用户穿戴特殊的设备。

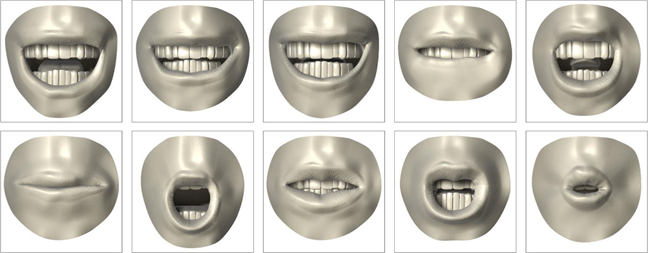


图3：嘴唇形状

此外，为了实现更高安全级的人脸认证机制，我们将实现基于唇动和人脸的双因子认证。人脸因子目前最严峻的问题就是类似图1所示对抗眼镜的人脸噪声。为了解决这个问题，我们分析人脸纹理信息并通过基于频谱残差的人脸纹理分析可以有效检测人脸上的恶意噪声信息。通过计算捕获图像数据的log频谱，对于自然的正常图像，把当前频谱曲线减去基准频谱可以得到我们需要移除的噪声数据特别是恶意的对抗噪声，这样可以过滤掉大部分干扰信息，从而实现准确的人脸信息识别。基于此，我们实现基于唇动和可对抗攻击的人脸信息的强双因子认证。

目前申请团队已经在声纳检测唇动的可行性方面进行了大量研究和验证，基于50名自愿者的实验表明，唇动的信号可以有效区分不同的认证用户。接下来讲基于唇动的活体检测机制，并展开基于大量自愿者进行实验验证，同时验证基于唇动的认证机制。

1. 人力投入（团队成员）：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **姓名** | **职称** | **主要职责** | **工作量** | **联系方式** |
| 1 | 李琦 | 副教授 | 活体检测和认证总体设计 | 4个月 | qli01@tsinghua.edu.cn |
| 2 | 王骞 | 教授 | LipFace方案总体设计 | 4个月 | qianwang@whu.edu.cn |
| 3 | 张晓丽 | 博士生 | 人脸边缘检测算法的设计 | 8个月 | zhangxl15@mails.tsinghua.edu.cn |
| 4 | 周满 | 博士生 | LipFace系统设计及实现 | 8个月 | zhouman@whu.edu.cn |
| 5 | 江沛佩 | 博士生 | 基于对抗样本的假冒攻击测试 | 8个月 | ppjiang@whu.edu.cn |
| 6 | 郭蕴哲 | 硕士生 | 算法的有效性验证实验 | 8个月 | guo-yz17@mails.tsinghua.edu.cn |
| 7 | 刘旋恺 | 硕士生 | 人脸显著性检测的设计 | 8个月 | liuxk18@mails.tsinghua.edu.cn |

1. 项目计划（*请****简要****填写项目计划，并给出****里程碑时间节点及产出***）
2. 2019年9月1日~2019年9月30日

测试现有活体识别和认证系统的安全性；

1. 2019年10月1日~2019年12月31日

设计和实现一种基于唇动的活体检测引擎及对抗人脸样本检测引擎；

1. 2020年1月1日~2020年1月30日

完成基于唇动的安全人脸认证算法；

1. 2020年2月1日~2020年3月31日

基于真实用户的活体识别和人脸认证系统测试；

1. 2020年4月1日~2020年4月30日

搭建测试环境并进行原型部署验证。

1. 资源需求（*请****简要****填写在资金、数据、时间或其他方面的需求*）
2. 资金需求：
3. 数据需求：真实人脸数据及恶意攻击数据、人力外包验证包括志愿者的数据集收集和实验验证
4. 项目周期：8个月
5. 其他需求：无
6. 团队简介（*请****简要****填写****负责人及团队****的专业领域、研究方向、成员规模、已有技术成果、获得奖励等信息*）

申请人研究网络安全、移动安全和大数据安全，研究团队由1名正教授、1名副教授、8名研究生以及1名本科生组成。研究团队在相关研究领域，特别是机器学习安全和人脸认证等专业领域有多年的研究基础。目前申请人已经研发多项恶意账号检测的算法并且得到部署验证。此外，团队已经研究基于声纳的手机安全机制，研究利用麦克风和扬声器分析和破解手指解锁图案锁的行为，攻击准确率接近100%，为本研究的顺利进行奠定了基础，该研究成果也发表于网络安全顶级会议CCS 2018[1]。研究团队相关研究成果录用或者发表于CCS 2015/2017-2019、USENIX Security 2019、ASPLOS 2019、ACSAC 2018、CIKM 2017、DSN 2017/2019、RAID 2015/2019和IEEE/ACM会刊等著名期刊和会议。相关研究也年获得2017年北京市科学技术二等奖以及ICDCS、ICNP、SECURECOMM等最佳论文奖等荣誉。

1. Man Zhou, Qian Wang, Jingxiao Yang, Qi Li, Feng Xiao, Zhibo Wang, and Xiaofeng Chen. PatternListener: Cracking Android Pattern Lock Using Acoustic Signals. In the Proceedings of ACM Conference on Computer and Communications Security (CCS), Toronto, Canada, October 15-19, 2018.