# 基于声波感知的移动设备实时防窃方法研究

卢 立 俞嘉地 李明禄
(上海交通大学计算机科学与工程系 上海 200240)

**摘 要** 近年来已经见证了移动终端在人们日常生活和工作中逐步流行和广泛使用.移动终端由于其便携性可能被用户携带前往任意地点,而其并非总是物理安全的.因此,移动终端极易成为偷窃者的目标.移动终端的失窃不仅会带来经济损失,还会引起隐私信息的泄漏从而带来进一步的损失.现有移动终端防护方法主要研究通过建立安全的用户认证系统规避移动终端失窃之后带来的进一步隐私信息泄漏风险.然而,这种滞后的防护手段并不能从根本上阻止用户的移动终端失窃.为给目前广泛使用的移动终端提供实时的安全保护,本文提出了一种基于声波感知的移动终端实时防窃方法 EchoScan,其利用移动终端内建传感器(包括音频设备和运动传感器)对移动终端所处的上下文环境进行实时感知建模以检测偷窃行为.本方法的核心是利用声波感知方法对移动终端所处的环境或状态(即上下文)变化情况实时感知建模.具体而言,EchoScan 利用调频连续波技术感知移动终端所在上下文的变化情况以检测移动终端所处状态.随后,EchoScan 进一步利用降噪自编码器网络和支持向量数据域描述来分别提取移动终端所处上下文的有效特征并实现在无偷窃者训练数据前提下的偷窃行为识别.为提高该方法的能耗效率,EchoScan 还利用运动传感器实时检测移动终端接触行为.验证实验进一步证明 EchoScan 能取得 87% 的平均偷窃检测准确率.且平均检测延迟在 500 ms 以内.

**关键词** 移动设备防窃;实时监测;声波感知;上下文感知;运动传感器
**中图法分类号** TP391 **DOI 号** 10.11897/SP.J.1016.2020.02002

# Towards a Real-Time Anti-Theft Method for Mobile Devices Leveraging Acoustic Sensing

LU Li YU Jia-Di LI Ming-Lu
(*Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai* 200240)

**Abstract** Recent years have witnessed the gradual penetration and wide employment of mobile devices in people daily life and work. Due to the portability of mobile devices, they can be carried by users to any spaces, which are not always physically secure areas. Hence, mobile devices easily become the primary target of thieves. The theft of mobile devices not only leads to direct economic loss, but also further induces the probable leakage of confidential information. Existing studies about mobile device protection concentrated on implementing the user authentication system to avoid the risk of confidential information leakage induced by the theft of mobile devices. However, such hysteretic protections cannot essentially prevent users' mobile devices from theft. To provide the real-time protection for prevalent mobile devices, in this paper, we propose an acoustic-based mobile device anti-theft system, EchoScan, which employs built-in sensors (including audio devices and motion sensors) to sense the context of mobile devices in real time so as to detect the theft. The core of the proposed system is to utilize acoustic sensing for real-time sensing and modeling the environment or state (i.e., the context) of mobile devices. Specifically,

收稿日期:2019-10-21;在线发布日期:2020-02-17.本课题得到国家自然科学基金(61772338)资助.**卢 立**,博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为移动感知与计算、网络安全与隐私、人机交互.E-mail:luli_jtu@sjtu.edu.cn.**俞嘉地**(通信作者),博士,高级工程师,中国计算机学会(CCF)专业会员,主要研究方向为物联网、移动计算与感知、网络安全与隐私、车联网.E-mail:jiadiyu@sjtu.edu.cn.**李明禄**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为云计算、车辆自组织网络、无线传感器网络、大数据分析.

EchoScan utilizes FMCW technique to capture the variation of mobile device context to detect the state of mobile devices. After that, EchoScan further adopts denoising autoencoder network and SVDD to extract efficient features of mobile device context and detect theft without the training data from thieves respectively. To improve the energy efficiency, EchoScan employs motion sensors to detect the touch of mobile devices in real time. Extensive experiments validate that EchoScan can achieve an average theft detection accuracy of 87% and the theft detection delay is within 500 ms on average.

**Keywords**　anti-theft of mobile devices, real-time detection, acoustic sensing, context sensing, motion sensor

## 1　引　言

随着高度集成的电子基础设施和无线通讯方式的快速发展,移动终端在人们日常生活和工作中扮演着愈发重要的角色.因此移动终端逐步成为人们隐私信息的存储介质,如个人身份证、银行卡号及安全码等等.然而,伴随移动终端的便捷性而来的则是移动终端失窃的高风险.赛门铁克的报告①显示被调查人群中 41.2%的用户曾丢失过其移动终端,并随后遭受隐私信息泄露的情况.因此,为移动终端提供足够的安全保障不仅能够减少用户的经济损失,还能够防止其隐私数据泄露造成的进一步风险.

已有工作大部分集中于实现用户认证系统为移动终端的信息保障措施,例如密码、指纹(如苹果触控 ID②)、脸部识别(如支付宝脸部识别登录③)、声纹(如微信声纹锁④)、甚至于说话时嘴部运动[1-2]等方法.然而,对于移动终端失窃而言,用户认证仅能保护移动终端内存储的信息,即实现失窃后的隐私防护功能,而无法提前警告用户以阻止移动终端的失窃.因此,为这些广泛使用的移动终端设计一种准确实时的防窃系统是很有必要的.

音频设备作为被广泛部署在移动终端中的基础传感器之一,因其低廉的价格但强大的感知能力受到广泛关注.基于此,本研究设计了一种移动终端实时防窃方法 EchoScan,其利用声波感知移动终端所处的上下文环境来检测识别偷窃行为.上下文环境是指移动终端所处的环境或被使用的状态,如静止于密闭空间时终端的状态、用户使用时终端的状态、被偷窃者窃取时终端的状态等.

EchoScan 的核心是利用移动终端内建的音频设备发出不易被人耳感知的近超声信号,并通过调频连续波技术来实现检测其所在上下文环境的变化情况以感知上下文的功能.为进一步提取可靠的上下文变化特征,EchoScan 利用降噪自编码器网络以无监督学习的方式提取降维而可靠的特征.随后,EchoScan 利用支持向量数据域描述方法实现无偷窃者训练数据前提下的偷窃行为检测功能.为解决实时声波感知引起的能耗问题,EchoScan 还实现了基于运动传感器的实时移动终端接触检测方法,其利用偏自相关的方法捕捉用户的非周期性运动从而检测接触动作.结合基于声波的上下文感知,EchoScan 最终实现实时的移动终端防窃方法.

本工作的主要贡献如下:

(1)本文提出一种实时移动终端防窃方法,其结合移动终端内建的音频与运动传感器来实时检测偷窃行为;

(2)本文调研利用移动终端内建音频设备实现声波感知其上下文环境并识别偷窃行为的可行性;

(3)本文结合移动终端内建低功耗运动传感器实现了实时检测识别移动终端被接触时刻的功能;

(4)本文执行包含多位志愿者的验证实验来证明提出方法的有效性,实验结果表明提出的方法能够取得 87%的偷窃检测准确率.

本文第 2 节回顾基于声波上下文感知与基于运动传感器人体活动检测相关工作;第 3 节描述基于声波感知的可行性初步研究和基于运动传感器的能耗效率提升初步研究;第 4 节介绍本文提出的系统

① New norton anti-theft to protect lost or stolen smartphones. https://www.symantec.com/about/newsroom/press-releases/2011/symantec_1004_05, 2011.10.4

② 使用 iPhone 和 iPad 上的触控 ID. https://support.apple.com/zh-cn/HT201371, 2019.9.11

③ 支付宝人脸识别新品发布会. https://mrchportalweb.alipay.com/user/logout/news/detail.htm?detailId=dt06, 2019.9.11

④ Voiceprint: The New WeChat Password. https://blog.wechat.com/2015/05/21/voiceprint-the-new-wechat-password/, 2019.9.11

的框架和具体实现细节;第 5 节展示验证实验的评估结果;第 6 节给出本文结论.

## 2 相关研究

本工作结合声波感知与运动传感器来对移动终端所处上下文环境与人体活动进行监测和感知.因此,本节简要讨论已有的基于声波的上下文感知以及基于运动传感器的人体活动监测相关工作.

### 2.1 基于声波的上下文感知

由于音频设备(即扬声器与麦克风)在移动终端上的广泛部署,声波感知技术被广泛用于监测终端所在环境的变化情况.本节将该类工作进一步细化为两小类,即环境感知与活动感知.

(1)环境感知.利用声波感知环境的已有工作主要集中在室内定位和地图构建领域上.早期工作[3]捕捉背景声频谱的声波指纹信息来实现室内定位.同时期的工作 Swadloon[4]则利用预部署的蜂鸣设备提供相对指向位置以实现定位功能.然而,这些工作对环境噪声很敏感且需要很高的能耗代价.为解决该问题,Guoguo[5]测量信号到达时间(Time of Arrival,ToA)以实现定位.但该工作对额外设备(即射频 RF 设备)的需求限制了其广泛部署的能力.WalkieLokie[6]测量手机与目标上附有的扬声器间的距离来定位该目标,其仅需商用手机与低成本的扬声器.其它工作 EchoTag[7]仅自定位物体的静态位置,即仅记忆手机曾放置过的位置.但该工作无法实现实时的室内定位.BatMapper[8]仅利用商用手机以测量信号的 ToA 以估计手机至阻碍物的距离来构建地图,但其需复杂的训练以得到估计参数.SAMS[9]通过手持手机随机移动估计室内环境的边沿位置并构建室内地图以解决该问题.

(2)活动感知.基于声波的技术还被广泛应用于活动感知上.BodyScope[10]设计了一种可穿戴声波设备来提取声波的特征以分类嘴部相关活动,但其需要额外可穿戴设备,故而很难在现实应用场景部署.后续工作[11-12]利用商用音频设备来识别人体活动.Ubicoustic[11]开发一种即插即用式的声波活动识别系统,其无需在使用前进行任何预训练.而 EI[12]则利用周围物体反射的声波信号并基于对抗网络的方式实现了一种与环境无关的活动识别方法.除了日常活动的识别,还有一些工作利用声波感知技术识别车内驾驶员行为以保证驾驶安全.早期工作[13]利用车内扬声器检测驾驶员的手机使用行为,其测量了声波信号的到达时间差(Time Difference of Arrival,TDoA)从而确定使用手机的用户是否为驾驶员.ER[14]则进一步扩展声波感知来检测一些危险驾驶动作,其利用声波信号的多普勒效应来区分危险动作与正常行为以对危险动作发出警告从而保障驾驶安全.

### 2.2 基于运动传感器的人体活动监测

运动传感器(即加速度计与陀螺仪)因其低廉的价格目前被广泛部署在移动终端当中.目前基于运动传感器的人体活动监测工作可以按照其使用方法分为基于特征的方法和基于相似度的方法.

(1)特征法识别.基于特征的活动识别方法主要利用不同行为对运动传感器产生影响以生成不同的特征来识别[15-17].这些工作通常利用行为引起的特征作为关键输入来识别分类.PBN[15]基于附着在用户身上的运动传感器采集的数据以及 AdaBoost 分类方法来实现活动监测系统.RisQ[16]则设计一种吸烟活动探测系统,其利用吸烟行为对运动传感器产生的普遍影响以提取出一些普适性的特征,接着利用随机森林分类法识别吸烟行为的一系列连续动作.MoodScope[17]则提出一种基于手机的活动监测方法,其通过一系列与手机有关的人机交互中提取的特征来推测用户的心情.

(2)相似度法识别.不同于基于特征的方法,基于相似度的方法通常会维护一组预构建的活动对应信号模式(profile)[18-20],其通过计算实时采样的信号样本与预构建的信号模式间的相似度来最终识别活动.目前最广为利用的相似度衡量指标有动态时间规整(DTW)距离[21]、推土机距离(EMD)、欧拉(Euclidean)距离等.

### 2.3 本文与已有工作的不同

已有活动识别的工作[14,22-23]利用各类传感器(如运动传感器、声波、无线信号等)捕捉用户动作引起的信号变化,并结合包括机器学习在内的多种方法实现用户动作的识别.直观上已有这些方法也可用于实现偷窃检测识别系统.然而,由于已有工作识别动作要求用户完成整个动作之后以提供完整的信号模式才能实现识别,因此仅能提供较为滞后的识别功能.而对于移动终端的失窃,仅能识别完整的偷窃动作意味着移动终端已经处于偷窃者的控制下,即失窃已经发生.因此,实际中,这些方法并不适用于偷窃检测识别场景.

近来已有与本工作最相似的工作仅有 iGuard[24],其利用手机内置运动传感器判断手机从口袋中取出

是由用户执行还是由偷窃者完成.但该工作基于十分严格的假设,即用户从口袋取出手机的过程中存在步伐速度逐渐减缓的现象而偷窃者则不会.然而在现实中,由于不同个体的习惯差异,该假设极有可能无法成立,从而导致该方法失效.同时,移动终端在失窃出现时所处的空间也未必在口袋中,因此该方法受到极大的条件限制.

不同于已有专注于用户行为的工作,本文转而利用声波感知去探测移动终端所处的上下文环境(其既可以是口袋,也可以是其它半封闭空间)判断其使用是否在用户的控制下.本工作实现的系统在任意情景下均能准确检测识别偷窃行为,从而为移动终端用户提供实时的安全保障.

## 3　初步研究及分析

本节首先分析了通用移动终端失窃场景模型.随后针对该模型进一步研究基于声波感知失窃检测及基于运动传感器的能耗效率提升的可行性.

### 3.1　移动终端失窃场景模型

相比于传统固定桌面设备,移动终端因其易携带的便利性在目前生活工作中被人们广泛使用.许多隐私敏感信息会被存储于这些终端中,如个人身份证号、信用卡信息等.但亦由于移动终端的便携性,该终端可能被用户携带出现在任意地点,其中也包括无物理安全保证的地点空间(如人群密集的广场或大街上、或其它公共场所等).在这些空间中,移动终端有很高的风险出现失窃或丢失情况.

本文面向的移动终端失窃场景具有以下假设:

(1)偷窃者出于利益原因主动发起针对用户移动终端的偷窃行为,且在偷窃过程中需尽量避免被用户察觉.因此其动作幅度应尽可能小,对用户除移动终端以外的物品造成的影响尽可能小;

(2)移动终端在被偷窃前处在用户放置的相对密闭空间当中,如口袋、拉链未关合的包等空间,而非被用户遗忘或丢弃在公共空间当中;

(3)用户使用移动终端时无任何特定限制条件(即用户应按照其自身习惯使用移动终端),且用户对动作幅度较小的偷窃行为感知极弱(如利用长镊子从用户口袋中偷取手机等情况);

(4)偷窃行为可能在任意时刻出现,既可能发生在用户行走过程中,也可能出现在用户坐在某公共场所休息的时刻.

具有以上假设的移动终端偷窃场景在日常生活中实际上是极其常见的①,如用户在进出超市时手机位于随身携带的包内,但包的拉链却未关合,从而导致偷窃者有机会靠近用户并窃取手机.而由于移动终端所存储信息的隐私性和敏感性,这样的失窃对用户而言很可能会造成巨大的损失.

### 3.2　基于声波感知的失窃检测可行性研究

如前所述,移动终端极易面临失窃,并进一步带来经济或敏感信息上的损失.目前最广为采用和部署的阻止移动终端信息泄露的方法即为移动终端实现用户认证系统.然而,对于移动终端失窃而言,用户认证仅能保护移动终端内存储的信息即实现失窃后的防护功能,而无法提前警告用户以阻止移动终端的失窃.因此,为如今广泛使用的移动终端设计准确实时的防窃系统是很有必要的.音频设备作为广泛部署于移动终端的传感器之一因其低廉的价格但强大的感知能力受到广泛关注.基于此,本研究利用基于音频设备的声波感知技术来探索实现失窃检测功能的可行性.如第 3.1 节所述,偷窃者在窃取用户移动终端时为保证不引起用户警惕,其动作幅度通常相对于用户从密闭空间中取出手机小得多.因此,可利用声波感知移动终端所处上下文来判断其所处状态,即合法用户取出或偷窃者窃取.

#### 3.2.1　基于声波 FMCW 的移动终端上下文感知

移动终端所处上下文定义为其所处的环境或被使用的状态,如静止于密闭空间中、被用户手持进行使用、被偷窃者窃取等环境和状态.由于该上下文既可能是动态的也可以是静态的,因此本研究采用声波的调频连续波(Frequency Modulated Continuous Wave,FMCW)技术来感知移动终端所处的上下文.FMCW 技术是一种为雷达设计且广为使用的测距方法.FMCW 的基本原理通过测量发射信号与接收信号的频差来实现声源与目标物体间的距离测量.图 1 为 FMCW 的基本原理图解.首先,移动终端的扬声器发出调制的锯齿波信号,其在周期 $\tau$(如图中 $\tau=0.02$ s)内扫频预定义的频段 $B$(如图中 $B=20\,\text{kHz}-17\,\text{kHz}=3\,\text{kHz}$).在信号从物体上反射后,该锯齿波信号被移动终端的麦克风所接收,如图 1 中的虚线所示.由于声波信号的传播,接收信号相对于发射信号存在时延,即声波信号的飞行时间(Time Of Flight,TOF).通常情况下,TOF 由于太微小而很难被直接测量到.为测量该 TOF,FMCW

① 盘点 9 大最容易遭遇偷窃的地方. https://xw.qq.com/cmsid/FJC2015102304489503,2019.9.6

《计算机学报》

解调接收的声波信号，即在发射信号与接收信号上执行去调频（dechirp）操作[25]测量发射信号与接收信号间的频差 $\Delta f$ 来估测 TOF. 基于三角形的几何相似性原理，飞行时间 $T$ 可被推导为

$$T=\frac{\Delta f\times\tau}{B} \tag{1}$$

由于扬声器与麦克风均集成于同一移动终端，这两者可被近似视作一个质点（即移动终端所在位置）. 故而，移动终端与物体的间距 $d$ 可推导为

$$d=\frac{c\times T}{2} \tag{2}$$

其中，$c$ 为声波信号的传播速度. 通过 FMCW，可得到移动终端与其周围物体间的大致距离，进而可利用该距离对其所在的上下文进行建模.

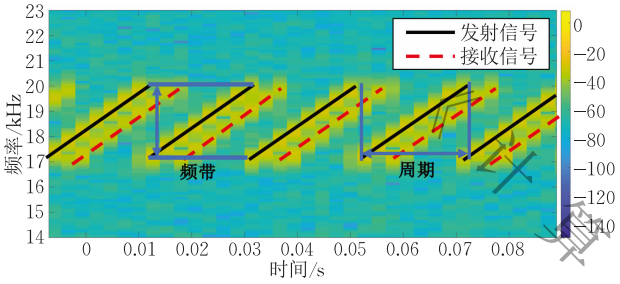
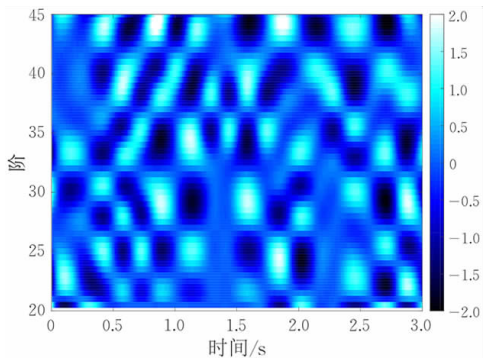



图 1 FMCW 基本原理图解

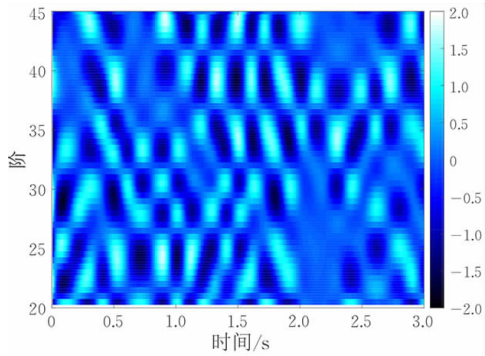
### 3.2.2 上下文感知可行性的实验验证

为验证以上方法感知移动终端所在上下文以实现失窃检测的可行性，本研究招募了 10 位志愿者执行验证实验. 在实验前，基于声波 FMCW 感知的系统被实现在一台三星 Galaxy S6 手机中. 实验将 10 位志愿者分为两组，分别扮演合法用户与偷窃者. 详细来说，7 位志愿者（合法用户）分别将上述的 Galaxy S6 放置于其腰部附近的口袋当中，然后按照其自身的习惯将手机从口袋中取出并进行正常的使用. 每位用户重复以上操作 10 次以避免偶然性因素的影响. 其余 3 位志愿者（偷窃者）则尝试从合法用户的口袋中窃取手机，并按其习惯离开现场. 每位偷窃者对每位用户实施 3 次窃取. 同时，每位合法用户还分别在坐着和站着的状态下重复以上实验. 共完成 266 组实验.

实验结果首先展示若干声波感知上下文的模式样例以验证用户和偷窃者从口袋中取出手机时的不同之处. 图 2 和图 3 分别展示了两位用户从口袋中取出手机时的声波感知上下文示例和两个偷窃者从口袋中取出手机时的 20 阶～45 阶频差示例（图中的灰度表示频差值）. 通过对比图 2 和图 3 可以发现，用户从口袋中取出手机时的频差绝对值显著大于偷窃者从口袋中取出手机时的频差绝对值，即用户从口袋中取出手机时，手机所处上下文环境出现明显的变化从而导致手机与其周围物体的距离明显

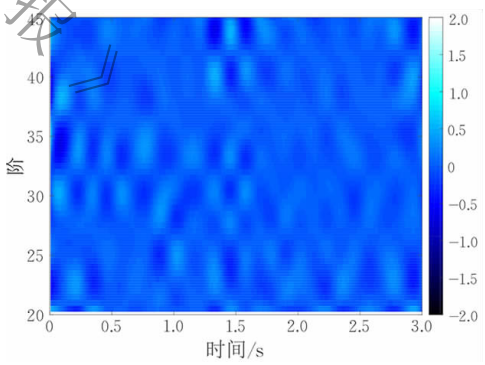



（a）用户1

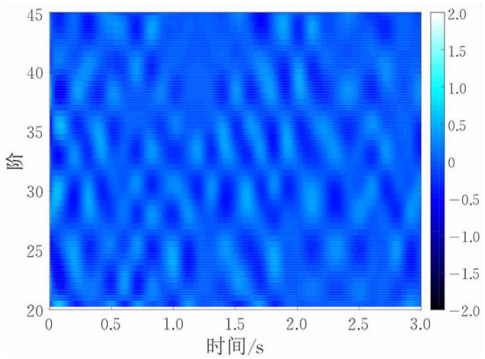



（b）用户2

图 2 用户从口袋中取出手机时声波感知的上下文模式




（a）偷窃者1




（b）偷窃者2

图 3 偷窃者从口袋中取出手机时声波感知的上下文模式

《计算机学报》

增大；相反在偷窃者取出手机时则没有出现这样显著的频差变化. 这是因为用户会遵从其自身习惯较为随意地取出手机，因此口袋(即此时移动终端所处的上下文环境)会出现明显的形变；而偷窃者在取出手机时为避免被用户感知会尽量避免口袋产生明显形变. 因此用户和偷窃者取出手机时其感知的上下文变化情况会有明显差别. 这些信号模式示例验证了本文在第 3.1 节中所做假设.

图 4 进一步展示了实验中 10 位志愿者实验采集信号模式间的平均动态时间规整(Dynamic Time Warping，DTW)距离矩阵. 从图中可以发现用户从口袋中取出手机时所得的上下文感知信号模式与偷窃者从口袋中取出手机时所得的信号模式差距明显. 而用户与用户之间的信号模式差距，以及偷窃者与偷窃者之间的信号模式差距则小的多. 该结果进一步验证在从口袋取出手机这一相同动作下，用户和偷窃者对手机上下文会造成显著的差异. 利用该差异，本文提出一种基于声波感知上下文的移动终端失窃检测方法，从而为广泛使用的移动终端提供先置的安全保障，而非滞后的信息保护手段.

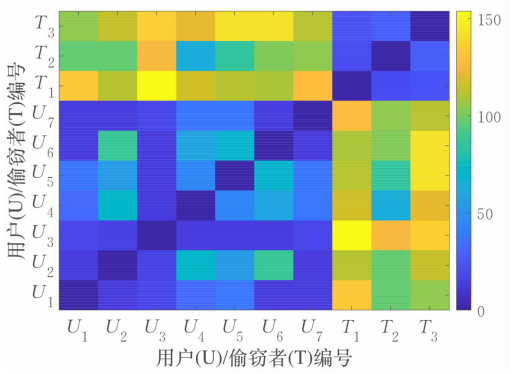



图 4　10 位志愿者采集信号模式间的平均 DTW 距离矩阵

### 3.3　利用运动传感器提高能耗效率

由于基于声波的上下文感知需要移动终端不间断发出近超声信号实现实时偷窃检测，该方法极易引起移动终端能耗的大幅增加. 故而，本研究寻求移动终端其他内置低功耗传感器配合声波感知以实现低功耗且实时的偷窃行为检测.

在正常使用/偷窃过程当中，实际上只有用户/偷窃者接触到移动终端后，移动终端所处的上下文环境才会出现明显变化. 根据该观察，可利用低功耗的传感器实时检测移动终端是否被触碰的状态来决定是否启用声波感知上下文检测其失窃状态，从而实现实时而低功耗的移动终端防窃目标. 目前移动终端广泛部署的传感器包括运动传感器(即加速度计和陀螺仪)、距离传感器、光线传感器、音频设备、摄像头等. 而既能检测用户/偷窃者接触手机的行为，又能实时低功耗检测的传感器只有运动传感器. 因此，本文探索利用运动传感器辅助声波感知实现实时低功耗的移动终端防窃功能.

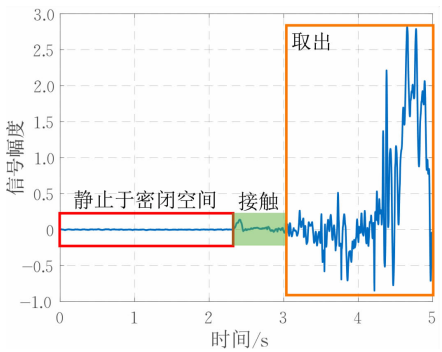
运动传感器能够实时提供移动终端所处状态六维信息，即 $X$、$Y$、$Z$ 三轴的线性加速度信息(加速度计)和角速度信息(陀螺仪). 虽然加速度计在较长时间的测量值是正确的，但其在较短时间内由于存在信号噪声导致误差. 而陀螺仪在较短时间内则比较准确而较长时间则会因漂移而存有误差. 因此，结合以上两者共六维信息可以实现在短期和长期内的精确检测[26].

为验证运动传感器实时检测移动终端接触行为的可行性，本研究招募 10 位志愿者执行验证实验. 实验将基于运动传感器的数据收集系统实现在一台三星 Galaxy S6 手机上，其采样率设置为 200 Hz. 为考虑取出移动终端前用户不同行为的影响，实验将 10 位志愿者平均分为两组，分别执行静止站立和步行两种状态. 两组用户都将手机放置于其口袋中，并在随后的一段时间内的静止站立/步行，最后从口袋中将手机取出. 由于本实验仅探究移动终端在被人接触时能否被运动传感器捕捉到，因此该实验中并不区分志愿者取出的方式(即以用户或偷窃者的方式取出手机). 每位志愿者重复该实验 5 次，故而共收集到 50 组实验数据.

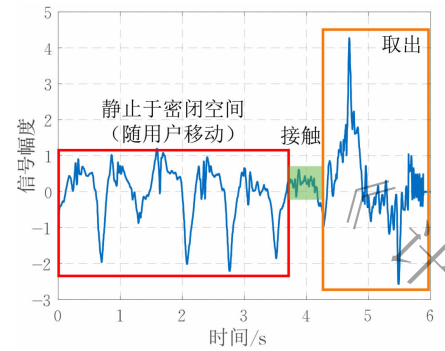
图 5 展示了用户处于站立/步行状态下移动终端被取出时运动传感器记录的数据. 从图 5(a)可以看出运动传感器首先表现出稳定的状态，其表达了用户此时处于静止站立状态，且手机随用户同样处于静止状态. 接着，运动传感器数据出现细微的波动，其表示手机受到触碰，很有可能即将被取出. 随后数据出现明显的峰值，其表达手机被志愿者从口袋中取出的大幅度运动状态. 图 5(b)则显示了用户处于步行状态下运动传感器捕捉的数据. 不同于图 5(a)，由于用户初始状态是运动的，因此运动传感器数据在全程均呈现不稳定趋势. 但由于人的步行是周期性行为，即一定长度的时间窗口中移动终端随人的运动是重复的，故运动传感器在人步行时的数据模式亦呈现周期性. 因此，仍可从数据中捕捉手机被接触时的时刻. 可以发现大约在时刻 4 s 附近，运动传感器数据突变为较嘈杂的模式，该时刻下即表示手机即将脱离随人体周期性运动的状态. 随后与图 5(a)相似，数据开始呈现明显峰值变化，其

《计算机学报》

表征手机被取出的过程. 该实验结果表明运动传感器可被应用于实时检测移动终端被接触行为.




(a) 用户处于静止状态




(b) 用户处于步行状态

图 5 用户处于不同状态下移动终端被取出时运动传感器数据示例

综合以上分析,结合移动终端内建音频和运动传感器来实现移动终端实时防窃方法是可行的.

## 4 移动终端实时防窃系统设计

为给移动终端提供实时先置的安全保护,本研究提出了一种移动终端实时防窃系统 EchoScan,其通过声波感知移动终端所处的上下文环境进行偷窃行为的检测,并进一步利用低功耗的运动传感器实现实时的防窃方法.

### 4.1 系统概述

图 6 展示了 EchoScan 的系统框架图.

EchoScan 实时利用低功耗运动传感器检测移动终端所处状态. 运动传感器实时采集移动终端的移动信号,并经过"移动终端接触检测模块"处理. 在该模块中,EchoScan 首先利用截止频率为 10 Hz 的 Equiripple 低通滤波器对运动传感器信号预处理以消除硬件产生的高频噪声. 接着,EchoScan 通过计算固定长度窗口内预处理采样信号的偏自相关值,实时解析移动终端相对上下文环境静止的状态,即

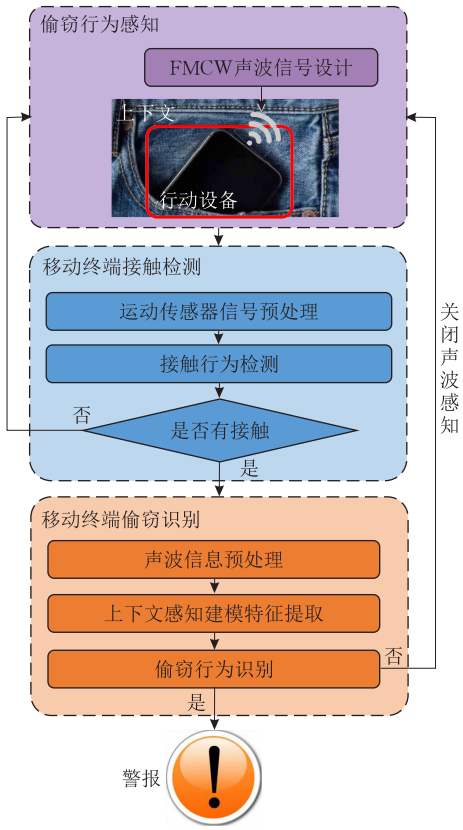



图 6 EchoScan 系统概述图

移动终端位于相对密闭空间中的状态. 详细而言,EchoScan 求解 Yule-Walker 方程来得到采样信号的偏自相关值. 当信号的偏自相关值到达极低点即不再保持周期性时,系统视其为移动终端被人接触的时刻标志. 同时,为使 EchoScan 在移动终端任意摆放的前提下仍能准确识别该接触时刻,系统进一步利用加速度计及陀螺仪的共六维信息,并采用 DBSCAN 聚类算法进行多数票决,最终得到接触时刻. 随后,基于声波的移动终端偷窃识别模块被激活.

"移动终端偷窃识别模块"预先设计了用于声波感知的锯齿波信号(即锯齿波的周期、初始频率、带宽),其综合考虑了声波感知偷窃行为所需的时间解析度和频率解析度. 该模块被激活时,首先通过移动终端的扬声器发射该锯齿波信号,并通过麦克风不断接收反射的声波信号. 被接收的声波信号通过采用截止频率为 17 kHz 和 20 kHz 的 Equiripple 带通滤波器进行预处理以消除环境带来的噪声干扰. 随后,EchoScan 利用如式(1)和式(2)所示的 FMCW 技术解析出移动终端与上下文间的距离,以描述上下文的形变情况. 为消除接收信号中强度较大的视距信号干扰,本工作提出第 $n$ 大频率响应的频差搜索方法以提取出接收信号中隐含的反射信号成分.

计算机学报

进一步为提取冗余度低且鲁棒的特征，EchoScan 利用降噪自编码器网络从第 $n$ 大频率响应频差信号模式中提取降维特征模式. 最后，EchoScan 基于该特征模式利用支持向量数据域描述(Support Vector Domain Description，SVDD)以实现在无偷窃者数据用于训练的前提下对移动终端所处当前状态进行识别以判断移动终端是否处于被偷窃的风险中.

### 4.2　FMCW 声波信号设计

在 FMCW 技术中，锯齿波信号设计包括对信号周期与信号频带(即锯齿波的初始频率和带宽)的选择，其对声波感知的响应时间和解析度均有影响. 锯齿波周期的设计与人体运动有关. 由于移动终端通常被用户随身携带，故而其所在上下文的感知建模需要匹配人体运动的相对静止时间. 通常情况下，人的一个行为动作会在较短时间内完成. 为了捕捉到这种短时窗口内的时不变运动，本系统将锯齿波信号的周期设置为 50 ms.

此外，对于锯齿波信号的带宽设计，则需保证发射信号能提供足够解析度感知移动终端所处上下文环境的变化，同时还需对用户保持不可感知性以保证良好的用户体验. 根据傅立叶变换理论[27]，由移动终端所处上下文环境反射的两个连续锯齿波信号能被解析当且仅当其频差满足以下条件：

$$\Delta f > \frac{1}{\tau} \tag{3}$$

其中，$\Delta f$ 为两个锯齿波信号的频差，$\tau$ 为锯齿波信号的周期. 基于式(1)和式(2)，可推导出

$$\Delta f = \frac{2dB}{c\tau} \tag{4}$$

其中，$d$ 为手机至其附近上下文环境(如口袋)间的距离，$B$ 为锯齿波信号的带宽，$c$ 为声波信号的传播速度. 结合式(3)和式(4)，可推导出

$$d > \frac{c}{2B} \tag{5}$$

这意味着 FMCW 的解析度会随着锯齿波信号带宽的增加而增加. 因此带宽的选择应当越大越好.

结合人耳不可听的范围(即≥17 kHz)及移动终端有限的录音能力(即≤20 kHz)[28]，本工作选择[17～20]kHz 作为发射信号的频带(即锯齿波初始频率为 17 kHz，带宽为 3 kHz).

利用该锯齿波信号，EchoScan 能以足够响应时间和解析度感知移动终端所处上下文环境的变化.

### 4.3　移动终端接触检测

如第 3.3 节所述，本研究利用低功耗的运动传感器检测移动终端被接触的时刻，从而为声波感知的激活提供实时支撑且保证能耗有效. 由于运动传感器在日常活动中亦被激活用于人体活动的追踪(如记步等)，因此利用该传感器进行实时移动终端接触检测不会带来明显的额外能耗.

#### 4.3.1　运动传感器信号预处理

由于移动终端硬件出厂时并非完美，因此运动传感器采集数据中常伴有噪声，其会对移动终端接触时的信号模式产生明显干扰. 因此，去除运动传感器数据噪声以保证准确的接触时刻检测是很有必要的. 在移动终端被接触前，其所处环境是跟随用户行为变化的，尤其是用户在运动状态下产生的变化. 但无论移动终端是静止的还是运动的，运动传感器感知的运动模式均为低频变化. 例如，人走路步频通常在 1.33 步/s[29]左右. 相比于步频，硬件产生的运动传感器噪声基本为高频噪声. 基于该观察，EchoScan 利用截止频率为 10 Hz 的 Equiripple 低通滤波器对原始运动传感器数据进行预处理，以消除高频噪声对后续接触行为检测的影响.

#### 4.3.2　接触行为检测

如第 3.3 节所述，用户的移动终端被取出时未必处于静止状态，因此运动传感器在被接触前采集的数据亦会出现明显波动. 为从这些数据波动中准确提取接触时的信号模式，EchoScan 利用自相关方法去识别接触行为. 由于移动终端被接触前，用户行为通常呈现周期性(如走路、跑步等)，可利用这些强相关运动引起的相似信号模式来区别接触前运动和接触时动作. 详细而言，运动传感器任意一维数据的自相关 $R(\theta)$ 可被推导为

$$R(\theta) = \sum_{t=\theta+1}^{N} m(t)m(t-\theta) \tag{6}$$

其中，$m(t)$ 为运动传感器任意一维在 $t$ 时刻的感知采样数据，$\theta$ 为自相关延迟参数，$N$ 为信号的采样点总数目. 由于运动传感器实时采集数据，因此在求解自相关的过程中很难得到采样点总数目 $N$. EchoScan 利用固定时间窗的偏自相关[30]来实时求解当前时间段与之前数据点间的相关性. 首先 EchoScan 利用固定长度的时间窗口采集运动传感器数据 $m(t)$，$t \in [1, n]$($n$ 为窗口宽度). 当数据长度达到窗口宽度时，EchoScan 计算该窗口内数据的偏相关值. 详细而言，$t$ 时刻的运动传感器数据可表示为 $m(t) = R_p^1(t)m(t-1) + \cdots + R_p^k(t)m(t-k) + \varepsilon_t$，其中 $k$ 为偏自相关延迟参数，$R_p^k(t)$ 为数据 $m(t)$ 的偏自相关. 通过求解该公式的 Yule-Walker 方程，可求得各个时间窗口内运动传感器数据的偏自相关值 $R_p^k(t)$.

《计算机学报》

随后在计算出的偏自相关值中找到最小偏自相关对应的时间段编号，即 $i=\arg_t \min R_p^k(t)$. 最后通过匹配时间段对应的原始运动传感器数据中的时刻点，即可搜索到接触行为出现的时刻. 图7显示了基于偏自相关的接触动作检测的示例.

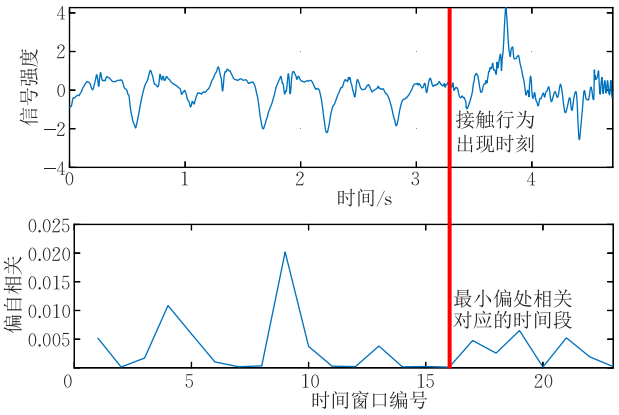
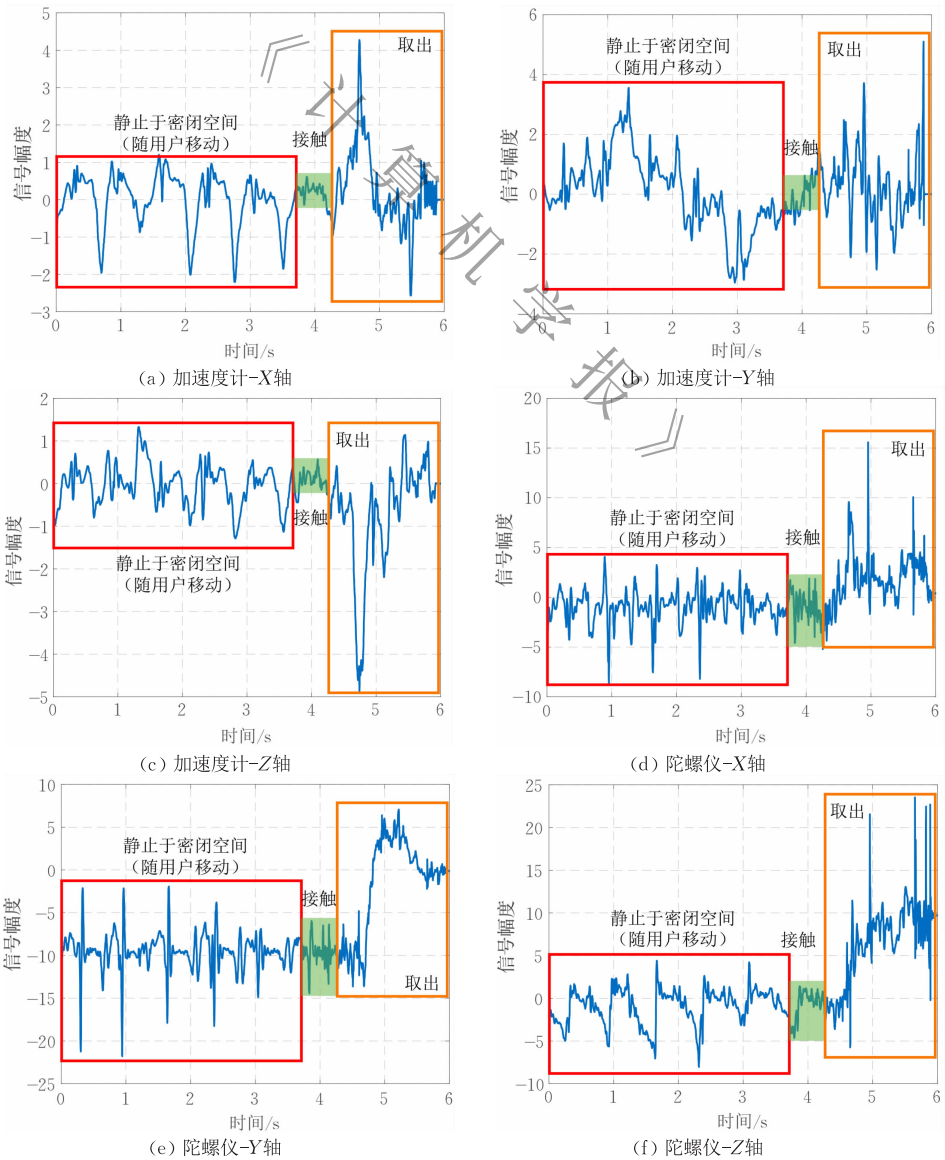



图 7 基于偏自相关的接触动作检测示例

在现实情况中，由于在相对密闭空间时移动终端可能处在任意一种放置状态(即重力加速度可能会影响任意多个维度的运动传感器数值)，因此单独利用某个维度的信息来检测接触状态很可能导致错误的结果. 图 8 展示了用户步行状态下移动终端被取出时运动传感器六维数据示例. 可以发现虽然加速度计的 $X$ 轴和 $Z$ 轴能较好表现出可区分的用户正常步行时和手机被取出时的状态(如图 8(a)和图 8(c)所示)，但加速度计的 $Y$ 轴数据则在接触行为之前就出现了明显的非周期性行为(如图 8(b)所示). 这是因为该实验中手机在口袋中处于侧放的状态，即 $Y$ 轴近乎与重力轴重合，从而导致了该轴数据无法真实反应出人行走的周期性规律. 图 8(d)、图 8(e)和图8(f)的陀螺仪数据示例反映了相似的




(a) 加速度计-$X$轴 (b) 加速度计-$Y$轴

(c) 加速度计-$Z$轴 (d) 陀螺仪-$X$轴

(e) 陀螺仪-$Y$轴 (f) 陀螺仪-$Z$轴

图 8 用户步行状态下移动终端被取出时运动传感器六维数据示例

问题. 因此，EchoScan 综合利用运动传感器六个维度的信息来检测识别接触. 通常情况下，重力加速度只会对少数维度的数据造成明显的影响，如图 8(b)的加速度计 $Y$ 轴数据. 基于此观察，EchoScan 采用多数票决的方式来确定具体的接触行为时刻点. 详细而言，系统利用偏自相关法从运动传感器六维信息中得到六个不同的接触时刻即 $t_{AX}$、$t_{AY}$、$t_{AZ}$、$t_{GX}$、$t_{GY}$、$t_{GZ}$. 随后利用 DBSCAN 聚类方法得到若干类，其中时刻数最多的类即为多数票决结果. 接触时刻通过计算该类中的时刻平均值得到.

### 4.4　移动终端偷窃识别

在 EchoScan 检测到移动终端被用户/偷窃者接触后，基于声波的上下文感知模块被激活判断此次接触是否为偷窃行为. 声波感知模块激活后，移动终端通过扬声器发出如第 4.2 节所设计的锯齿波信号；声波信号经移动终端周围的上下文环境反射后，被移动终端的麦克风接收. EchoScan 随后利用该信号对移动终端所处上下文环境进行感知建模.

#### 4.4.1　声波信号预处理

声波信号感知上下文时会经过多种传播路径反射、折射、衍射等物理过程(即多径效应)，随后被麦克风接收. 经过多径传播的信号除包含有移动终端所在上下文的特征，还带来周围环境中隐含的噪声，其会干扰 EchoScan 准确识别上下文的能力. 因此 EchoScan 首先对接收的声波信号进行预处理以消除噪声影响.

如第 3.2 节所述，EchoScan 利用 17 kHz～20 kHz 的近超声信号实现 FMCW 的上下文感知能力. 而环境中常见噪声频段通常在 8 kHz 以下(如机器噪声、人交谈说话声等). 因此，EchoScan 采用截止频率为 17 kHz 和 20 kHz 的 Equiripple 带通滤波器来滤除与声波 FMCW 无关的噪声. 如第 3.1 节所述，本文面向的移动终端失窃场景主要指移动终端从相对封闭的空间中被取出的情景. 由于移动终端的音频设备功率有限，穿透密闭空间后的声波能量衰减十分显著. 而经过用户周围物体反射并进一步穿透密闭空间被移动终端接收的信号则会出现更加明显的能量衰减，其远远小于移动终端所处上下文环境(即该相对密闭空间)反射的信号强度. 因此，即使在人员密集场所，面向如 3.1 节所述的失窃场景的 EchoScan 接收的信号也不会受到明显的来自用户周围密集物体的多径问题影响.

在实际情况中，由于人体自身运动，可能会导致移动终端与其上下文环境发生相对运动. 该相对运动会在接收声波信号引入多普勒频移，其会影响基于 FMCW 的距离测量结果. 然而，该距离测量结果的变化不会影响基于声波的上下文感知建模以及后续的偷窃行为识别. 这是因为多普勒效应表征的物体运动(即移动速度)与 FMCW 频差均能反映出用户与偷窃者在取出移动终端行为的不同. 如第 3.1 节所述，用户取出移动终端时的动作通常较为粗放即对终端所在上下文的施加的形变较大，同时使得移动终端与其上下文间的相对移动也较快. 因此，在该情况下，接收信号的多普勒频移较大. 相反地，偷窃者取出移动终端时较为小心，其导致的终端与其上下文的相对移动较慢，故多普勒频移较小. 同时，在基于 FMCW 的距离测量中，多普勒频移会直接影响 FMCW 频差，从而增大用户和偷窃者对应信号模式的差别，进一步提高识别能力. 因此，EchoScan 并不消除接收信号中的多普勒频移，而是直接利用其与 FMCW 频差的结合用于提取上下文感知建模特征及后续的偷窃行为识别.

#### 4.4.2　上下文感知建模特征提取

经信号预处理后，EchoScan 接着利用 FMCW 技术感知建模移动终端所处上下文环境，如第 3.2 节所述. 但是，由于移动终端集成的扬声器与麦克风间的距离很近，故而接收信号中的视距(Line-Of-Sight，LOS)信号会淹没其它信号的特征，尤其是移动终端上下文(如口袋、包等环境物体)反射的声波信号特征，如图 9 所示. 该物理现象直接导致基于声波的移动终端所在上下文感知能力的弱化. 因此，很有必要对接收信号进行预处理以消除 LOS 信号的影响.

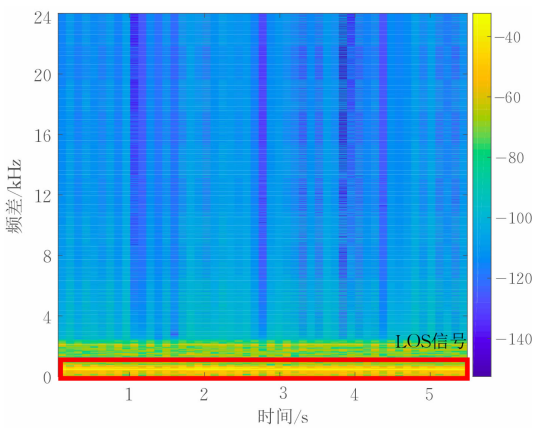



图 9　接收信号中 LOS 信号淹没问题的图解

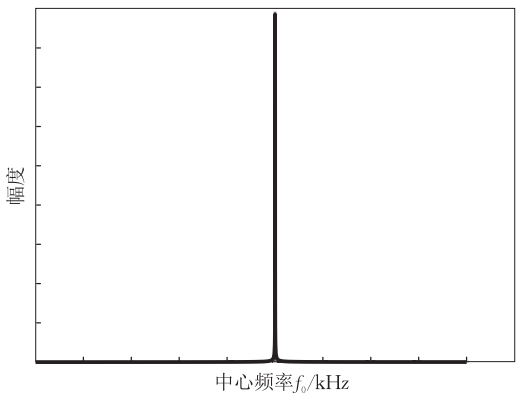
由于移动终端的扬声器与麦克风的相对位置总是保持不变的，因此 LOS 信号在接收信号中总能表现出稳定的频率响应. 故而，可以通过比较该频

率响应来区分 LOS 信号与上下文反射的信号. 本研究设计基于短时傅里叶变换(Short Time Fourier Transformation,STFT)的启发式算法,其将长时间的信号分割为多个短时滑动窗口,并分别在每个滑动窗口中应用傅里叶变换推导出频率响应. 详细而言,首先在解调信号(即 $s_d=s_t\times\overline{s_r}$,其中 $s_t$ 与 $s_r$ 分别为发射信号与接收信号)上执行 STFT 计算,然后在每个滑动窗口中搜索具有第 $n$ 大频率响应的频差 $\Delta f$,即

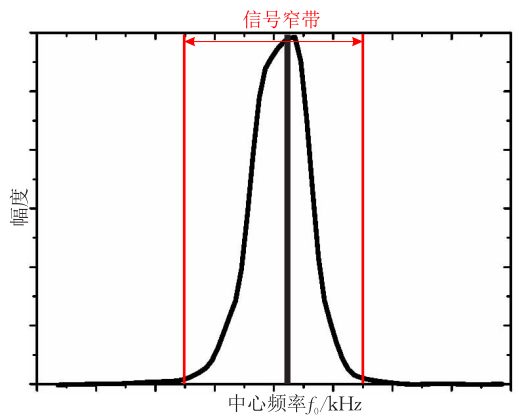
$$\Delta f_n(t)=\arg_f \max^n FFT_f(s_d(t)) \tag{7}$$

其中,$FFT_f(\cdot)$ 为快速傅里叶变换(Fast Fourier Transformation,FFT),$t$ 为滑动窗口编号,$\max^n(\cdot)$ 为搜索第 $n$ 大值的操作. 此后,可以提取出具有第 $n$ 大频率响应的频差序列 $\Delta f_n(t)(t=1,\cdots,N)$(即第 $n$ 阶频差),其表征接收信号中的特定信号分量(如 LOS 信号和反射信号等). 因为 LOS 信号的信号强度显著大于其它反射信号的信号强度,故而低阶(即 $n$ 值较小)频差表征 LOS 信号而高阶(即 $n$ 值较大)频差可表征上下文反射信号.

理论上直接从扬声器传播到麦克风的声波信号仅有一条(如图 10(a)所示),故而只需将第 1 阶频差消除即可得到上下文反射信号. 但在实际情况中,由于移动终端的扬声器能力限制及麦克风有限的采样率,得到的频峰信号会覆盖中心频率附近的一段窄带,如图 10(b)所示. 因此,仅消除第 1 阶频差无法完全解决 LOS 信号的淹没问题. 经验性研究发现阶数 $n>20$ 时 LOS 基本被消除. 经过该方法得到的特征表达如图 2 和图 3 所示. 由于该特征的维度较高,直接利用机器学习方法识别分类极易遭遇维数灾难.



(a) 理论频域信号



(b) 实际频域信号

图 10 理论上与实际中经 FFT 变换得到的频域信号对比

为解决该问题,本研究设计了基于降噪自编码器[31]的模型从以上 FMCW 频差特征中提取有效且可靠的降维特征. 该模型包含一个降噪层和一个自编码器层. 首先,降噪层向输入的频差特征中随机添加噪声数据以降低原始数据隐含的白噪声. 详细而言,$N\times n^r$ 维的 FMCW 频差特征首先被重构为 $Nn^r\times 1$ 维的列向量,其中 $N$ 为采样的总点数,$n^r$ 为采用特征阶的数目. 降噪层接着以均匀分布的概率将该向量中的数据部分清零从而为原始数据引入噪声数据. 随后,自编码器层以无监督方式将降噪后的特征抽象为压缩表达. 该压缩表达能够建模移动终端所处的上下文环境. 详细而言,降噪层首先对输入的频差特征 $\Delta f$ 降噪以得到特征 $\Delta f^r$. 随后,自编码器将该特征映射为压缩表达 $u=\sigma(w\Delta f^r+b)$,其中 $\sigma(\cdot)$ 为逻辑斯谛(logistic)函数(即 $\sigma(x)=\frac{1}{1+e^{-x}}$),$w$ 和 $b$ 分别为自编码器的权重和偏倚. 该编码器层可通过优化如下目标训练:

$$\min DIF(\Delta f^r,\Delta\overline{f^r})=\min\frac{1}{N}(\Delta f^r_{(i)}-\Delta\overline{f^r_{(i)}})^2+\lambda\Omega_{\text{weights}}+\beta\Omega_{\text{sparsity}} \tag{8}$$

其中,$N$ 为训练样本的数目,$\Delta f^r_{(i)}$ 和 $\Delta\overline{f^r_{(i)}}$ 分别为降噪输入 $\Delta f^r$ 和重构输入 $\Delta\overline{f^r}$ 中的第 $i$ 个元素,$\Omega_{\text{weights}}$ 和 $\Omega_{\text{sparsity}}$ 分别为参数和稀疏度的 $L_2$ 正则子,$\lambda$ 和 $\beta$ 分别为这两个正则子的系数. 本目标最小化了原始输入 $\Delta f^r$ 与重构输入 $\Delta\overline{f^r}$(其中 $\Delta\overline{f^r}=\sigma(w^{\mathrm{T}}u+b')$)间的差距. 该目标保证压缩表达 $u$ 能抽象出原始输入 $\Delta f^r$ 的大部分关键信息. 通过该降噪自编码器网络,EchoScan 能提取出有效的上下文环境特征.

#### 4.4.3 偷窃行为识别

如第 3.2 节所述,用户和偷窃者从密闭空间中取出移动终端时会对其所处上下文产生差异明显的变化影响. 因此,利用声波 FMCW 感知上下文的信号模式识别偷窃行为是可行的. 然而,该问题与传统的二分类问题有所区别. 在实际情况中,系统仅能在使用前要求用户提供足够的训练数据(即正类数据)

《计算机学报》

以构建用于偷窃检测的分类识别模型. 而对于负类数据(即偷窃者对移动终端所处上下文造成变化引起的信号模式),系统则无法要求用户提供. 因此,系统在使用时无法取得完整的两类数据. 传统二分类监督学习方法通常要求用户提供完整的两类数据才能进行模型构建和训练. 为解决该问题,本研究将偷窃行为识别视为单分类问题,并采用一种特殊的单分类支持向量机(Support Vector Machine,SVM),即支持向量数据域描述(SVDD),来实现偷窃行为的识别.

SVDD 的基本原理是为所有训练样本点寻找一个超球面(半径为 $r$,球心为 $b$),使得大部分数据都在该球面的包覆范围内,即

$$\min_{r,\vec{b},\varepsilon} r^2+\frac{p}{n}\sum_{i=1}^{n}\varepsilon_i \tag{9}$$

$$\text{s.t. } \|u_i-b\|^2\le r^2+\varepsilon_i,\ \forall i\in 1,\cdots,n$$

其中,$p$ 为参数,$\varepsilon_i$ 为第 $i$ 个训练样本点对应的松弛系数,$\varepsilon$ 为所有松弛系数的集合,$u_i$ 为第 $i$ 个训练样本点,$n$ 为总的训练样本数目,$\|\cdot\|^2$ 为 $L_2$ 范数计算. 为避免得到的超球面结果欠拟合,EchoScan 进一步引入径向基(Radial Basis Function,RBF)核函数将输入特征首先映射向高维空间,并通过求解优化问题(9)得到高维空间超球面,最后映射回低维特征空间生成最终单分类偷窃行为识别模型.

## 5　性能测试与分析

本节用实验验证 EchoScan 的可行性和有效性.

### 5.1　实验配置与方法

实验以 Android APP 的形式将 EchoScan 实现在三台不同的移动终端上,其分别为三星 Galaxy S6、三星 Galaxy S3、华为 P10. 声波 FMCW 感知的发射近超声信号如第 4.2 节所设计. 三台终端均采用其主扬声器和主麦克风作为声波感知使用的音频设备. 此外,三台终端的运动传感器采样率均设置为 200 Hz,麦克风采样率设置为 48 kHz. 实验在三种不同场景下进行,即场景 1:用户坐着且移动终端放置于其裤子口袋中、场景 2:用户走动且移动终端放置于其裤子口袋中、场景 3:用户走动且移动终端放置于随身背包中. 在每个场景中,实验随机招募 13 位志愿者包括 7 位男性和 6 位女性,且其年龄在 18～55 岁之间. 在这 13 位志愿者中,其中 10 位持有移动终端作为该终端的合法用户,而另外 3 位则扮演偷窃者试图在不引起用户感知的前提下窃取移动终端. 同时,实验在三种不同的环境中即实验室(理想环境)、超市(拥挤嘈杂环境)、地铁(轻微震动环境),重复以上三种场景进行性能评估. 在每次实验中,每位合法用户志愿者随机选择一台移动终端放置于其口袋/背包中,其重复取出的过程 10 次来完成实验. 随后,每位偷窃者随机对其他 10 位合法用户实施 10 次不被感知的偷窃行为,即偷窃者以自身方式实施多次偷窃行为,如被用户感知到则被视为无效样本;反之才计入有效偷窃行为中. 当系统将一次取出行为识别为偷窃行为时,其触发手机的震动功能以提醒用户,并一直持续至用户手动给出取消指令.

实验定义若干衡量标准评估 EchoScan 性能. 假设实验共执行 $N$ 次,其中系统正确检测偷窃行为的次数为 $n_{TP}$,系统正确检测正常行为的次数为 $n_{TN}$,系统误报偷窃行为的次数为 $n_{FP}$,系统漏报偷窃行为的次数为 $n_{FN}$,其中 $N=n_{TP}+n_{TN}+n_{FP}+n_{FN}$. 故

(1)偷窃检测准确率. 偷窃/正常取出行为被准确识别为偷窃/正常取出行为的概率,即 $ACC=(n_{TP}+n_{TN})/N$.

(2)偷窃检测漏报率. 偷窃行为没有被检测识别为偷窃行为的概率,即 $FNR=n_{FN}/(n_{FN}+n_{TP})$.

(3)偷窃检测误报率. 常规移动终端取出行为被误报为偷窃行为的概率,即 $FPR=n_{FP}/(n_{FP}+n_{TN})$.

(4)延迟检测时间. 假设移动终端被接触到的时刻为 $t_b$,系统发出警报/关闭声波感知的时刻为 $t_e$. 延迟检测时间被定义为 $t=t_b-t_e$.

(5)时均耗电量. 假设系统总运行时间为 $T$,移动终端总电量为 $P$. 使用过程中移动终端追踪本系统消耗总电量为 $w$,则系统的时均耗电量为 $p=(P\times w)/T$.

### 5.2　总体性能

(1)偷窃检测性能. 实验首先通过与仅利用运动传感器的 iGuard[24] 及仅利用声波感知的方法比较来评估 EchoScan 的偷窃检测准确率. 图 11 展示了 EchoScan 在不同场景下与其它两种方法的偷窃检测准确率比较. 可以发现在三种不同的场景下,EchoScan 均能取得超过 85% 的准确率. 而 iGuard 在场景 1 下仅能取得 26.5% 的准确率. 这是因为 iGuard 仅能面向用户走动过程中出现的偷窃行为,且对偷窃者与用户取出移动终端的模式有较强的假设. 同时,由于 EchoScan 利用运动传感器检测移动终端接触动作引入些许误差,其准确率略低于全声波感知方法. 但可以看到,在三种场景下两者的准确率差距均在 5% 以下. 此外,EchoScan 在三种不同场景下的偷窃检测准确率平均标准差仅为 2.7%,

其优于 iGuard 的平均标准差 4.3%. 如第 5.1 节所述,实验招募了不同年龄、不同性别的用户,且其取出手机的行为遵从其自身习惯. 因此该实验结果证明 EchoScan 对不同用户的年龄、性别、终端使用习惯等因素均鲁棒.

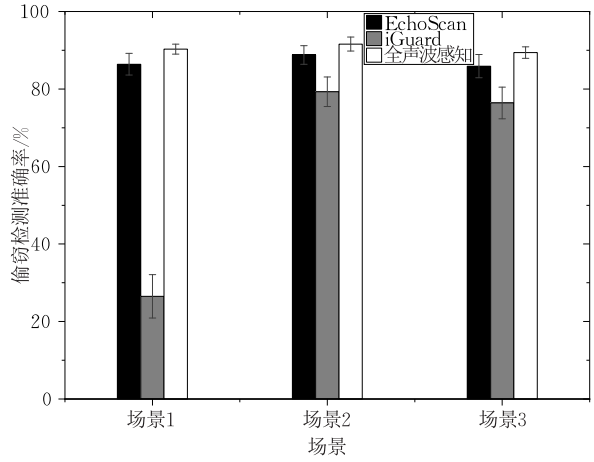



图 11 系统偷窃检测准确率

实验还评估了 EchoScan 与其它两种方法的偷窃检测漏报率. 图 12 展示了 EchoScan 在不同场景下与其它两种方法的偷窃检测漏报率比较. 可以看到在三种场景下,EchoScan 的平均偷窃检测漏报率为 5.9%,其远低于 iGuard 的平均偷窃检测漏报率 26.6%;尤其是在场景 1 下,iGuard 的偷窃检测漏报率高达 50.5%,而 EchoScan 则为 6.4%. 这是因为在场景 1 中,用户的移动终端在取出前后仅出现速度由零变为非零的过程,而未出现 iGuard 假设中速度减缓和变换方向的过程. 因此,在该场景下,iGuard 的性能显著降低. 相比之下,EchoScan 没有对用户和偷窃者行为的强假设,因此能适应各类不同的场景. 此外,全声波感知方法能取得 5.4%的平均偷窃检测漏报率,其仅比 EchoScan 低 0.5%. 该

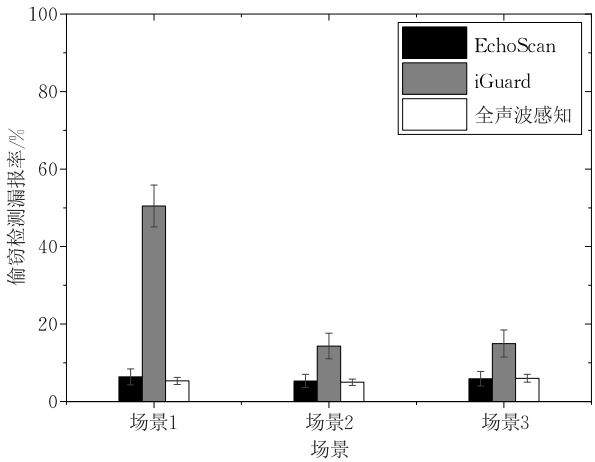



图 12 系统偷窃检测漏报率

结果进一步证明 EchoScan 能够有效地检测不同场景下的偷窃行为. 此外,EchoScan 的偷窃检测漏报率平均标准差仅为 1.9%,其进一步证明 EchoScan 对不同的用户是鲁棒的.

(2) 偷窃检测延迟. 偷窃检测延迟决定了用户能否及时阻止正在发生的偷窃行为,其为偷窃检测系统能否有效实施的重要衡量指标. 图 13 展示了 EchoScan 在不同终端下的延迟检测时间累计概率密度(CDF). 可以发现对三种不同的移动终端,90%测试样例的偷窃检测延迟分别小于 444.8ms、600ms、700ms. 在该时间内,用户能够有充足的反应时间阻止偷窃行为. 由于三种终端不同的计算能力,因此偷窃检测延迟有所不同. 但在实验中,即使是性能最差的 Galaxy S3 也能将检测延迟控制在 1s 之内,其保证用户能够及时发现偷窃行为,并进一步阻止该行为导致的经济或信息损失.

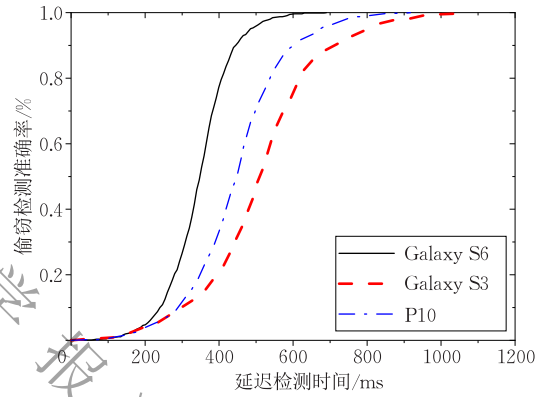



图 13 系统延迟检测时间的累计概率密度

EchoScan 的延迟主要由基于运动传感器的接触检测延迟与基于声波感知的偷窃识别延迟组成. 为评估这两部分延迟对总体延迟时间中的具体影响,本实验除了记录移动终端被接触到的时刻 $t_b$ 与系统发出警报/关闭声波感知的时刻 $t_e$ 以外,还记录了系统激活声波感知模块的起始时刻 $t_i$. 则接触检测延迟为 $t_i-t_b$,而偷窃识别延迟为 $t_e-t_i$. 图 14 展示了 EchoScan 在不同阶段的延迟检测时间. 可以发现在三台不同终端下,接触检测延迟均大于偷窃识别延迟. 尤其在较新的终端下(即 Galaxy S6),接触检测延迟是组成总体延迟时间的主体. 还可以观察到不同手机的接触检测延迟很接近(均在 280ms 附近),而偷窃识别延迟则存在较大的差别. 这是因为接触检测延迟使用的算法较为简单,对硬件计算能力要求不高. 该延迟主要是由模式匹配使用的时间窗口导致的. 相反,偷窃识别利用了较为复杂的特征提取方法和分类方法,对硬件性能有一定要求,因此在不同计算能力的终端上表现为不同的识别延

迟.但可以看到即使在实验中性能最低的手机(即Galaxy S3)上,该部分延迟也能小于250 ms.

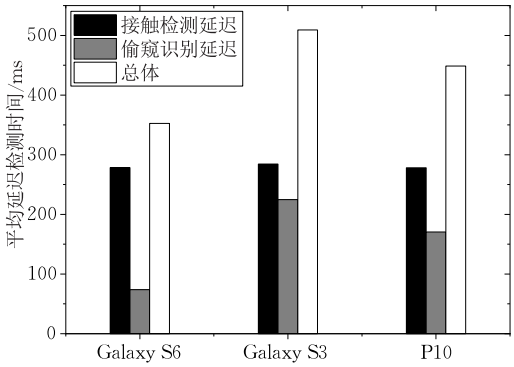



图 14　系统在不同阶段的延迟检测时间

### 5.3　用户体验

(1)检测误报概率.系统误检测偷窃行为并发出警报会极大影响用户的使用体验.图15展示了EchoScan在不同场景下与其他两种方法的偷窃检测误报率比较.可以发现EchoScan在三个场景下的平均偷窃检测误报率为5.9%,其远小于iGuard的偷窃检测误报率32.9%.尤其是在场景1下,iGuard的偷窃检测误报率增长到67.3%,而EchoScan则维持在6.7%左右.这是因为场景1中移动终端被取出前后用户不满足iGuard对其行为的假设,因此偷窃检测误报率快速上升,这与偷窃检测漏报率相关实验结果的分析一致.此外,全声波感知方法的平均偷窃检测误报率为5.1%,其仅比EchoScan低0.8%,证明了EchoScan引入基于运动传感器的接触检测并未对偷窃检测性能造成明显影响.此外,EchoScan的偷窃检测误报率平均标准差仅为1.9%.该结果说明在用户体验上EchoScan仍然对不同用户是鲁棒的.

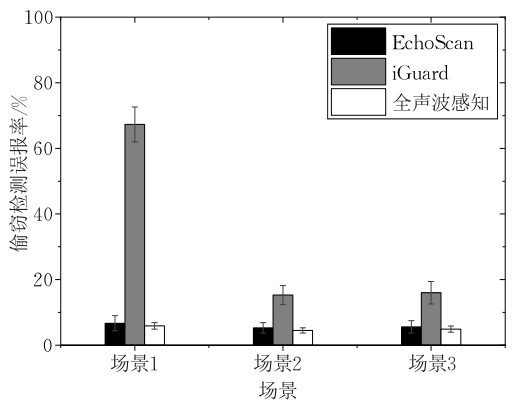



图 15　系统偷窃检测误报率

(2)系统能耗.移动终端的电量总是有限的,因此系统的耗电量会直接影响用户的移动终端使用体验.EchoScan的设计目标是在任意环境中,系统能够在用户不主动激活的情况下以低功耗的形式帮助用户检测移动终端可能被偷窃的情况,从而减少用户的经济和隐私损失.由于运动传感器会持续运行于移动终端后台以检测是否发生接触行为(即EchoScan会持续运行于后台),因此本实验以时均耗电量为标准衡量EchoScan能否以低功耗的形式实现偷窃检测功能.图16展示了EchoScan在不同终端下与全声波感知方法的时均耗电量比较.此处时均耗电量定义为每小时系统的耗电量,并用电池电量mAh作为计量单位.可以发现EchoScan的时均耗电量为123 mAh,其远低于全声波感知方法的时均耗电量408.3 mAh.该结果证明了EchoScan引入的基于运动传感器的接触检测方法能够有效降低系统的能耗,并提高移动终端的用户体验.此外,在不同终端下,EchoScan的耗电量会有些许不同,其标准差为10.2 mAh.这是因为不同手机的计算能力、音频装置功率不同,因此系统实际运行的能耗也有所差别.但也可以看到该差异较小,因此EchoScan对不同终端是鲁棒的.

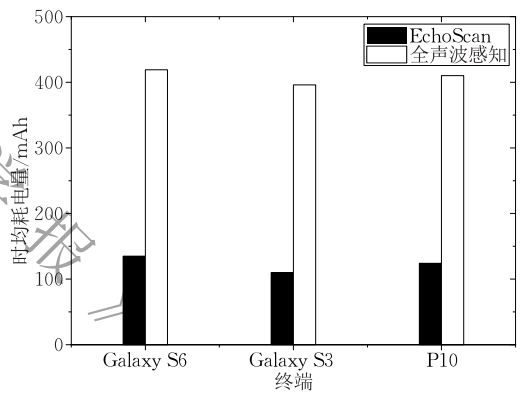



图 16　系统时均耗电量

### 5.4　影响EchoScan性能的因素

(1)移动终端型号.不同移动终端的扬声器与麦克风位置可能有所不同.由于EchoScan利用FMCW声波感知方法对用户取出手机的行为模式进行感知建模,因此扬声器与麦克风的相对位置会对偷窃检测性能有所影响.不同于之前的实验,本实验分别采用三台移动终端完成模型的训练过程,随后在另外两台终端上进行测试.其它实验配置与第5.1节中所描述相同.图17展示了EchoScan使用一台终端训练而在另外两台终端测试的偷窃检测准确度.可以发现Galaxy S6与P10之间的互用性较好,偷窃检测准确率均能超过85%.而在Galaxy S3上训练,在另外两台终端测试的准确度则仅有80.6%和79.8%.这是因为Galaxy S6与P10的扬声器及麦克风相对位置较为相似,即均位于终端底

部且距离接近.不同于这两台终端,Galaxy S3 虽然主麦克风位置仍位于终端底部,但其主扬声器则位于终端背部的摄像头右侧,故而使得扬声器与麦克风相对距离大大增加.正由于此,Galaxy S3 收集的FMCW 频差信号模式与其他两台终端收集的信号模式有所不同,并进而导致较低的互用性.

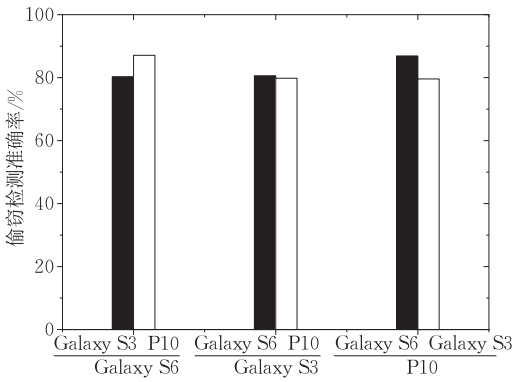



图 17 系统使用一台终端训练并在另外两台终端测试的偷窃检测准确率

(2)环境.实验还评估了不同环境对 EchoScan 偷窃检测性能的影响.图 18 展示了 EchoScan 在不同环境中的偷窃检测准确率.可以观察到在三种场景下,实验室环境中的偷窃检测准确率均高于其它两种环境,这是因为在较为拥挤的超市和地铁环境中,移动终端所处的上下文(即口袋或背包)易受到外界因素干扰从而导致 EchoScan 的性能下降.但也可以看到这种差距并没有很大.在三种场景下,后两种环境中的偷窃检测准确率与实验室环境中准确率的差距均在 6%以下.该结果说明系统对不同的环境是鲁棒的.此外,可以发现超市和地铁环境中的准确率相差不多.不同于超市,地铁环境中存在轻微的环境震动,实际上会影响 EchoScan 基于运动传感器的接触检测功能,即系统会频繁误检测移动设备受到接触,并激活声波识别偷窃行为模块.但幸运的是,由于该轻微震动不会对移动终端所处的上下文环境(即口袋或背包)造成明显形变,其对声波识别偷窃功能的影响十分微弱.

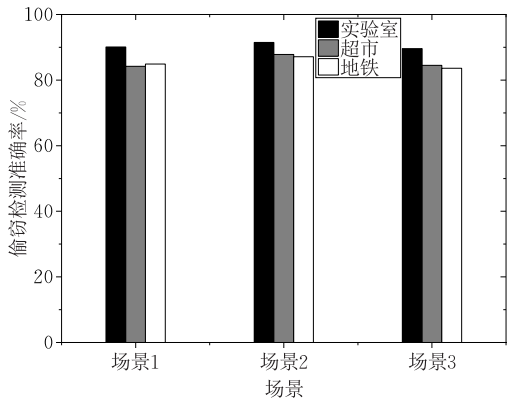



图 18 系统在不同环境中的偷窃检测准确率

(3)用户感知与偷窃检测.实验还对比了用户对偷窃行为感知与 EchoScan 对偷窃行为检测来评估偷窃者通过模仿用户取出移动终端的行为来避免偷窃检测的可能性.不同于第 5.1 节中的实验配置描述,本实验将被用户感知到的偷窃行为也计入结果统计.同时,扮演偷窃者的志愿者被告知在偷窃过程中除了要避免用户感知,还要试图绕过本文提出的系统.在进行该实验前,志愿者均被告知 EchoScan 的基本原理.每位偷窃者向每位用户实施 20 次偷窃行为,且在三种不同环境中各重复一次实验.表 1 展示了系统检测偷窃行为的准确率和用户感知偷窃行为的准确率对比.可以观察到当 EchoScan 漏报偷窃行为时,用户在 12.9%÷13.9%=99.2%的情况下均能通过自身感知发现偷窃行为并进行阻止.通过对实验的分析发现当 EchoScan 漏报偷窃行为时,大部分偷窃者均模仿了用户取出手机的行为.但在模仿该取出行为时,通常会对手机的上下文造成明显的形变(如口袋的剧烈变化),从而引起用户的感知.实验中仅有 2 次偷窃行为同时避过了用户感知和 EchoScan 检测.这是因为这 2 次偷窃均发生于超市中(即拥挤而嘈杂的空间),用户在该环境中被过多的噪声影响了感知能力,从而使得偷窃成功.但也可以发现这种情况是十分罕见的(实验中仅有 0.1%÷13.9%=0.8%),因此偷窃者在大部分情况下很难通过模仿用户取出移动终端的行为以实施成功的偷窃.

**表 1 系统检测偷窃与用户感知偷窃对比**

| 用户感知 | EchoScan | | |
|---|---|---|---|
| | 成功检测/% | 失败检测/% | 合计/% |
| 成功检测 | 23.3 | 12.9 | 46.3 |
| 失败检测 | 63.7 | 0.1 | 63.7 |
| 合计 | 87.0 | 13.0 | 100.0 |

## 6 总 结

本文提出基于声波感知的移动终端实时防窃方法,其利用移动终端内建的传感器(包括音频设备和运动传感器)来感知建模移动终端所处的上下文环境以实时识别偷窃行为.本系统首先利用低能耗的运动传感器来实时检测移动终端接触行为.通过偏自相关方法,系统能够捕获用户的异常非周期行为点,即用户或偷窃者接触手机的时刻.随后,系统利

用调频连续波技术感知移动终端所在上下文环境的变化情况.为了得到可靠的变化特征并实现单分类偷窃行为检测,系统进一步利用降噪自编码器网络和支持向量数据域描述来分别提取有效特征和识别偷窃行为.验证实验招募了多位志愿者进行偷窃行为模拟以证明系统的可行性和有效性,实验结果表明本系统能够取得87%的偷窃检测准确率.

## 参 考 文 献


[1] Lu Li, Yu Jiadi, Chen Yingying, et al. LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals//Proceedings of the IEEE International Conference on Computer Communications. Honolulu, USA, 2018: 1466-1474

[2] Lu Li, Yu Jiadi, Chen Yingying, et al. Lip reading-based user authentication through acoustic sensing on smartphones. IEEE/ACM Transactions on Networking, 2019, 27(1): 447-460

[3] Tarzia S P, Dinda P A, Dick R P, Memik G. Indoor localization without infrastructure using the acoustic background spectrum //Proceedings of the ACM International Conference on Mobile Systems, Applications and Services. Bethesda, USA, 2011: 155-168

[4] Huang Wenchao, Xiong Yan, Li Xiang-Yang, et al. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones//Proceedings of the IEEE International Conference on Computer Communications. Toronto, Canada, 2014: 370-378

[5] Liu Kaikai, Liu Xinxin, Li Xiaolin. Guoguo: Enabling fine-grained indoor localization via smartphone//Proceedings of the ACM International Conference on Mobile Systems, Applications and Services. Taipei, China, 2013: 235-248

[6] Huang Wenchao, Li Xiang-Yang, Xiong Yan, et al. Stride-in-the-loop relative positioning between users and dummy acoustic speakers. IEEE Journal on Selected Areas in Communications, 2017, 35(5): 1104-1117

[7] Tung Yu-Chih, Shin Kang G. EchoTag: Accurate infrastructure-free indoor location tagging with smartphones//Proceedings of the ACM Annual International Conference on Mobile Computing and Networking. Paris, France, 2015: 525-536

[8] Zhou Bing, Elbadry M, Gao Ruipeng, Ye Fan. BatMapper: Acoustic sensing based indoor floor plan construction using smartphones//Proceedings of the ACM International Conference on Mobile Systems, Applications and Services. Niagara Falls, USA, 2017: 42-55

[9] Pradhan S, Baig G, Mao Wenguang, et al. Smartphone-based acoustic indoor space mapping. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(2): 75:1-75:26

[10] Yatani K, Truong K N. BodyScope: A wearable acoustic sensor for activity recognition//Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. Pittsburgh, USA, 2012: 341-350

[11] Laput G, Ahuja K, Goel M, Harrison C. Ubicoustics: Plug-and-play acoustic activity recognition//Proceedings of the ACM Symposium on User Interface Software and Technology. Berlin, Germany, 2018: 213-224

[12] Jiang Wenjun, Miao Chenglin, Ma Fenglong, et al. Towards environment independent device free human activity recognition //Proceedings of the ACM Annual International Conference on Mobile Computing and Networking. New Delhi, India, 2018: 289-304

[13] Yang Jie, Sidhom S, Chandrasekaran G, et al. Detecting driver phone use leveraging car speakers//Proceedings of the ACM Annual International Conference on Mobile Computing and Networking. Las Vegas, USA, 2011: 97-108

[14] Xu Xiangyu, Gao Hang, Yu Jiadi, et al. ER: Early recognition of inattentive driving leveraging audio devices on smartphones //Proceedings of the IEEE International Conference on Computer Communications. Atlanta, USA, 2017: 1-9

[15] Keally M, Zhou G, Xing G, et al. PBN: Towards practical activity recognition using smartphone-based body Sensor networks//Proceedings of the ACM International Conference on Embedded Networked Sensor Systems. Seattle, USA, 2011: 246-259

[16] Parate A, Chiu M, Chadowitz C, et al. RisQ: Recognizing smoking gestures with inertial sensors on a wristband//Proceedings of the ACM International Conference on Mobile Systems, Applications and Services. Bretton Woods, USA, 2014: 149-161

[17] Likamwa R, Liu Y, Lane N D, Zhong L. MoodScope: Building a mood sensor from smartphone usage patterns//Proceedings of the ACM International Conference on Mobile Systems, Applications and Services. Taipei, China, 2013: 389-402

[18] Ranjan J, Whitehouse K. Object hallmarks: Identifying object users using wearable wrist sensors//Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. Osaka, Japan, 2015: 51-61

[19] Akl A, Feng C, Valaee S. A novel accelerometer-based gesture recognition system. IEEE Transactions on Signal Processing, 2011, 59(12): 6197-6205

[20] Akl A, Valaee S. Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, USA, 2010: 2270-2273

[21] Salvador S, Chan P. Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis, 2007, 11(5): 561-580

[22] Wang H, Lai T T-T, Choudhury R R. MoLe: Motion leaks through smartwatch sensors//Proceedings of the ACM Annual International Conference on Mobile Computing and Networking. Paris, France, 2015: 155-166


《计算机学报》

[23] Wang Yan, Liu Jian, Chen Yingying, et al. E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures//Proceedings of the ACM Annual International Conference on Mobile Computing and Networking. Maui, USA, 2014: 617-628

[24] Jin Meng, He Yuan, Fang Dingyi, et al. iGuard: A real-time anti-theft system for smartphones//Proceedings of the IEEE International Conference on Computer Communications. Atlanta, USA, 2017: 1-9

[25] Skolnik M I. Radar Handbook. New York, USA: McGraw-Hill Inc., 1970

[26] Cheng P, Oelmann B. Joint-angle measurement using accelerometers and gyroscopes — A survey. IEEE Transactions on Instrumentation and Measurement, 2010, 59(2): 404-414

[27] Iovescu C, Rao S. The fundamentals of millimeter wave sensors. Dallas, TI, USA: Texas Instruments, Technical Report, 2017

[28] Landau H. Sampling, data transmission, and the Nyquist rate. Proceedings of the IEEE, 1967, 55(10): 1701-1706

[29] Farley C T, Gonzalez O. Leg stiffness and stride frequency in human running. Journal of Biomechanics, 1996, 29(2): 181-186

[30] Ramsey F L. Characterization of the partial autocorrelation function. The Annals of Statistics, 1974, 2(6): 1296-1301

[31] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders//Proceedings of the International Conference on Machine Learning. Helsinki, Finland, 2008: 1096-1103

[32] Metz C E, Herman B A, Shen J-H. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Statistics in Medicine, 1998, 17(9): 1033-1053

[33] Tax D M J, Duin R P W. Support vector domain description. Pattern Recognition Letters, 1999, 20(1113): 1191-1199

**LU Li**, Ph. D. candidate. His research interests include mobile sensing and computing, cyber security and privacy, human-computer interactions.

**YU Jia-Di**, Ph. D., senior engineer. His research interests include Internet of Things (IoT), mobile computing and sensing, cyber security and privacy, connected vehicles.

**LI Ming-Lu**, Ph. D., professor. His research interests include cloud computing, vehicular ad-hoc network, wireless sensor network, big data analysis.

《计算机学报》

**Background**

With the rapid development of highly-integrated electronic infrastructures and convenient wireless communications, mobile devices become more pervasive in our daily life. Such devices gradually become the storage medium of people's privacy, including personal information (e. g., personal ID) and financial information (e. g., banking account and CVV code of credit card) etc. However, along with the convenience brought by mobile devices, people also suffer from the severe risk of mobile device losses. Recent reports exhibit that the loss (especially the theft) of mobile devices not only leads to direct economic losses, but also induces probable privacy leakage. Hence, it is necessary to provide protection for mobile devices to prevent such losses. Existing studies mostly concentrate on developing user authentication systems for mobile devices to prevent the privacy leakage after the theft of mobile devices. But such approaches cannot prevent mobile devices from potential theft.

To implement the anti-theft solution for mobile devices, in this paper, we explore the in-built audio infrastructures in mobile devices to sense and model the context of the devices for detecting the theft behaviors. We further integrate the in-built motion sensors to detect the touch gestures before activating the audio infrastructures to improve the power efficiency and thus the user experience.

Utilizing the audio infrastructures and motion sensors for anti-theft of mobile devices, we first develop Frequency Modulated Continuous Wave (FMCW) on acoustic signals to sense the context of mobile devices. To further extract reliable features for context sensing, we integrate the denoising autoencoder neural network to reduce the dimension of features in an unsupervised manner. After that, we employ the Support Vector Domain Description (SVDD) to realize the theft detection. To further improve the power efficiency, we integrate the motion sensors and develop a partial autocorrelation-based method to capture the non-periodic movements so as to detect the touch gestures. The experiments show that the proposed system can accurately detect the theft of mobile device in real time with limited power consumption.

This work is supported by the National Natural Science Foundation of China (No. 61772338).