申请上海交通大学硕士学位论文

基于声波的手写签名认证研究

论文作者：

学　　号：

导　　师：

专　　业：软件工程

答辩日期：2020 年 1 月 9 日

Submitted in total fulfillment of the requirements for the degree of Master in Software Engineering

Researches on Acoustic-based Handwritten Signature Verification

School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University
Shanghai, P.R.China

Jan. 9th, 2020

基于声波的手写签名认证研究

摘 要

随着信息技术的兴起和应用，许多计算机应用需要用户进行登录操作来身份验证，如密码登录等。为了提升用户体验和信息的安全性，大量基于人体特有生物特征的身份认证技术被不断研究并得到广泛应用，例如指纹识别、人脸识别等。手写签名作为个体的一种重要的行为特征，在金融、法律、政府等领域广泛应用于用户身份的识别，具有极其悠久的历史，然而一个仿造签名可能会造成巨大损失甚至动荡，因此手写签名的正确甄别显得尤为重要。直至目前，手写签名识别主要依靠人工进行，签名验证者根据用户的历史签名来判断当前的签名是否为真实签名，该方式具有人力成本高、依赖验证者经验和误判率高等缺点。由于自动手写签名认证顺应用户对用户体验和安全提升上的诉求，将人力资源从签名识别中解放出来，提高签名识别的准确率和稳定性，因而关于自动签名认证系统的研究和应用逐步开展起来。

目前国内外对手写签名认证的研究大多基于开放数据集，这些研究工作在普适性和用户体验上有待提高。另一方面，探索不同的签名捕捉方式，可以降低签名认证的成本，让签名认证系统在现实中快速普及。本文针对现有研究存在的问题，使用声波记录签名信息，参考传统的签名认证框架，探索出一种新的手写签名认证方案。本文的主要研究成果包括：

（1）提出利用声波感知技术来记录签名运动。使用主流智能手机上扬声器和麦克风进行声波的发射和接收操作，针对商用硬件的特性设计用户友好的声波信号，采用声波相位相关信息实现了对手写签名运动中微小动作的跟踪。

（2）针对声波相位相关信息设计了合适的特征提取和用户独立的识别模型。去除相位相关信息中的环境噪声，将手写签名动作的信息提取出来，结合主流的自动化签名认证系统框架和分类技术，设计出适合声波的手写签名判别流程和模型。

（3）进行实验评估和原型系统实现。对本文方案进行了全面的评估，包括验证精度、交叉用户可用性、硬件补偿的效果、系统鲁棒性、微基准实验、对比实验和重放攻击等。设计并实现了基于声波的签名认证原型系统，对其运行效率进行评估，该方案系统准确度性能达到：AUC（ROC 曲线下方面积）=98.7% 和 EER（等错误率）=5.5%，实验结果表明该方案是个非侵入式、鲁棒的、安全的、低延

迟的手写签名认证方案。

本文首先介绍手写签名认证系统的背景和研究意义，分析当前研究的现状和不足，然后对手写签名认证系统进行简要综述，提出本文的技术路线。在此基础上研究并设计了一种包含四个模块的基于声波的手写签名认证方案，对声波相位相关信号感知技术、特征提取技术、相似性度量技术和深度学习模型等关键技术进行了深入研究，详细介绍了方案中各模块的设计思想和技术细节。接着对该方案设计了多组实验进行评估，设计并实现了一个原型系统。最后总结本文研究内容并展望未来的研究工作。

关键词：声波感知, 智能手机, 手写签名认证

RESEARCHES ON ACOUSTIC-BASED HANDWRITTEN SIGNATURE VERIFICATION

ABSTRACT

With the rise and application of information technology, many computer applications require users to perform login operations to authenticate, such as password login. In order to improve user experience and information security, a large number of identity authentication technologies based on human-specific biometrics have been continuously researched and widely used, such as fingerprint recognition and face recognition. Handwritten signatures, as an important behavioral feature of individuals, have been widely used in the identification of user identities in the fields of finance, law, government, etc., and have a long history. However, a forged signature may cause huge losses and even turbulence. Correct screening is particularly important. Until now, handwritten signature recognition has mainly been performed manually, and the signature verifier determines whether the current signature is a real signature based on the user's historical signature. This method has the disadvantages of high labor costs, reliance on verifier experience, and high rate of false positives. Since the automatic handwritten signature authentication complies with users' demands for user experience and security improvement, it frees human resources from signature recognition and improves the accuracy and stability of signature recognition. Therefore, the research and application of automatic signature authentication systems have been gradually carried out.

At present, researches on handwritten signature authentication at home and abroad are mostly based on open datasets. These researches need to be improved in terms of universality and user experience. On the other hand, exploring different signature capture methods can reduce the cost of signature authentication and make signature authentication systems popular in practice. Aiming at the problems existing in the existing research, this paper uses sonic to record signature information and refers to the traditional signature authentication framework to explore a new handwritten signature authentication scheme. The main research results of this article include:

(1) It is proposed to use sonic sensing technology to record signature movement. Use

the speakers and microphones on mainstream smartphones for sound wave transmission and reception operations, design user-friendly sound wave signals for the characteristics of commercial hardware, and use sound wave phase-related information to track small movements in handwritten signature movements.

(2) A suitable feature extraction and user-independent recognition model is designed for the acoustic wave phase related information. The environmental noise in the phase-related information is removed, and the information of the handwritten signature action is extracted. Combining with the mainstream automated signature authentication system framework and classification technology, a handwritten signature discrimination process and model suitable for acoustic waves are designed.

(3) Perform experimental evaluation and prototype system implementation. Comprehensive evaluation of the scheme in this paper, including verification accuracy, cross-user availability, effects of hardware compensation, system robustness, micro-benchmarking Experiments, comparative experiments, and replay attacks. Designed and implemented a sonic-based signature authentication prototype system and evaluated its operating efficiency. The system's accuracy performance reached: AUC (area under the ROC curve) = 98.7% and EER (equal error rate) = 5.5%. Experimental results shows that the scheme is a non-intrusive, robust, secure, low-latency handwritten signature authentication scheme.

This article first introduces the background and research significance of the handwritten signature authentication system, analyzes the current status and shortcomings of the current research, and then briefly summarizes the handwritten signature authentication system and proposes the technical route of this article. Based on this, a sonic-based handwritten signature authentication scheme with four modules is researched and designed, and key technologies such as sonic phase-dependent signal sensing technology, feature extraction technology, similarity measurement technology, and deep learning models are studied in depth. The design ideas and technical details of each module in the scheme are introduced in detail. Then, several schemes of experiments were designed to evaluate the scheme, and a prototype system was designed and implemented. Finally, this article summarizes the research content and looks forward to future research work.

KEY WORDS: sonic sensing, smartphone, handwritten signature authentication

目 录

插图索引

表格索引

算法索引

第一章 绪论

1.1 研究背景与意义

随着信息技术的兴起和广泛应用，许多应用需要用户首先进行登录操作，这其实就是身份识别方式，只有通过了身份认证的用户才会被系统认为是授权用户。用户的身份认证方式可以为分为三种[1]，如图 1–1所示：

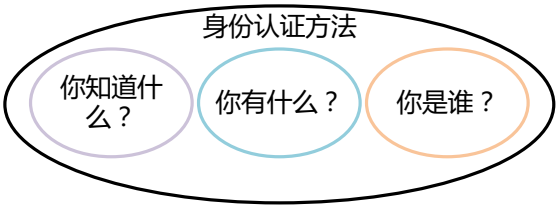


图 1–1 身份认证方式

Figure 1–1 Identification methods

（**1**）**根据用户知道的信息**。（**你知道什么**?）日常生活中，最常用的是密码登录，这是一种依赖用户知道什么来识别是否为授权用户的方法，但一旦用户的密码被泄露，恶意用户输入相同的密码也会被系统识别为授权用户。

（**2**）**根据用户拥有的东西来**。（**你有什么**?）你回家会使用钥匙开门，钥匙就是你所拥有的东西，比起用户知道的信息，此处则体现为用户拥有的物理实体。

（**3**）**根据用户独特的身体特征**。（**你是谁**?）随着智能手机技术进步和普及，目前人脸和指纹识别已经被用于手机系统登录和移动支付。人脸[2]、指纹[3]、虹膜[4]等是人体的固有特征，在随着时间变化上呈稳定状态，这样的认证方式比密码认证更加安全。除此之外，个人独特的行为特征也可以被用于身份认证，比如：唇语识别[5]、手写签名识别[6]、语音识别[7]、步态识别[8] 等。

手写签名作为个体的一种重要的行为特征，在金融、法律、政府等领域广泛应用于用户身份的识别。然而，这种认证方式也会受到恶意用户的攻击，举个例子，攻击者可以仿造存款用户的签名在一张取款支票上签名去银行取款，会造成实际用户的财产损失，降低银行存款安全性，大则可引发金融灾难。而在政府或军事领域，仿造签名会造成更加难以想象的结果和混乱。因此，在这些使用签名进行授权的领域，签名的识别变得尤其重要。在很久以前甚至现在，签名的识别主要依靠人工进行，签名验证者根据用户的历史签名来判断当前的签名是否为真

实签名。这种方式，需要耗费昂贵的人力资源，且依赖于签名认证员的个人能力，容易造成误判。

信息技术使签名认证的自动化成为可能，自动签名认证系统得到研究和应用。自动签名认证系统根据捕捉签名的方式不同，可以被分为两大类：离线签名认证系统和在线签名认证系统[9]。离线签名认证系统依靠签名图像的静态数据进行签名的比较来识别，而在线签名认证系统则依靠用户书写过程中记录下的时间序列数据（如书写速度、笔的压力等）来进行比较和识别。由于外加的时间维度信息，通常在线签名认证系统的表现优于离线签名认证系统。

无论是人工的签名识别还是自动签名认证系统，签名的识别过程都符合图 1–2 所示的签名认证过程。在注册过程，用户需要输入多个他/她的签名存入数据库

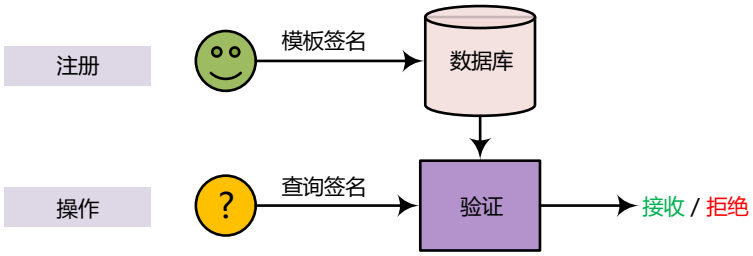


图 1–2 签名认证过程

Figure 1–2 Signature verification architecture

作为模板签名；在操作过程，用户签名作为查询签名提交给系统，系统将这个查询签名和模板签名进行比较，判断查询签名是否为真实签名。

一个签名认证系统的关键问题在于二方面：*a)* 需要能精准跟踪手写签名过程的行为信息。由于手写签名过程，笔记变化多样，而且运动幅度较小，对传感器的敏感度有较高要求。*b)* 用户体验和判别精度。用户体验体现在设备对用户的影响，比如是否需要特殊设备、是否需要用户佩戴设备等，还体现在模板签名数量。用户体验可能会和判别精度发生冲突，比如减少模板签名数量会导致判别精度下降，复杂的设备则能更细粒度跟踪签名运动轨迹，有利于判别精度的提升。

开展自动手写签名认证研究才能顺应用户对用户体验和安全提升上的诉求，将人力资源从签名识别中解放出来，提高签名识别的准确率和稳定性。目前国内外对手写签名认证的研究有很多基于开放数据集进行研究，存在一些问题，在普适性和用户体验上有待提高，探索签名的捕捉方式，可以降低签名认证的成本，让签名认证系统在现实中快速普及。

1.2 国内外研究的现状与存在的问题

这里从数据来源和手写签名认证建模两个方面，对近年来国内外学者在手写签名认证方面的研究进行分析。

(1) 数据来源

在数据来源方面，现有的手写签名认证系统的研究主要依赖于惯性传感器、手写平板设备和扫描图像，其中前二者可用于实现在线签名认证系统，而离线签名认证系统则依赖于扫描图像。

基于惯性传感器的方法[10-12] 利用穿戴在手上的惯性传感器 (如加速度计、陀螺仪等) 跟踪手部运动，提取个人签名时手部独特的运动特征。这种方法要求用户佩戴可穿戴设备（如定制设备、腕表和装有惯性传感器的笔等），会给用户造成额外的负担或困扰，而且设备穿戴的位置不宜固定，这会对签名认证的精度造成较大影响。

基于手写平板设备的方法[13-15] 利用手写笔或者手指在平板书写留下的动态轨迹来识别签名。这样的轨迹在静态轨迹基础上，还能获得书写时对平板压力和轨迹变化速度等时间序列数据，相比仅仅用静态图像的认证方式可以大大提高认证精度。但是这种认证方式需要一个手写平板，而且限制字迹只能写在平板而不是纸上，这是与日常生活中的书写场景是不同的。

基于扫描图像的方法[16-19] 使用扫描仪扫描并切割出签名图像，直接比较存在数据库中的静态图像，与人工判别的方式在数据来源上有相似之处。由于没有书写过程的数据，这种仅仅比较写出结果的方法，往往无法判断出在字迹上看起来很相似的伪造签名。但这种签名认证方式的优势是具有通用性和较好的用户体验，不会给用户签名过程造成干扰，用户可以依照平时的签名习惯来签名。

(2) 手写签名认证建模

手写签名认证建模方面的方法有多种，离线和在线签名认证系统需要采取不同的建模方法。

离线签名认证系统通常采用图像匹配的方法[20]，比如人工神经网络 (Artificial Neural Network, ANN)[21]、像素匹配 (Pixel Matching Technique, PMT)[22] 等。

在线签名认证系统[23] 采用的方法可以分为两类：*a)* 基于函数的方法，使用长短时记忆网络 (Long Short-Term Memory, LSTM)[24]、隐马尔科夫模型 (Hidden Markov Model, HMM)[25, 26]、动态时间规整 (Dynamic Time Warping, DTW)[9] 等这些函数型方法直接在时间序列上进行处理并识别；*b)* 基于特征的方法，这种方法将时间序列从时域数据转化为频域数据，可以用到的变换方法有离散傅里叶变换 (Discrete Fourier Transform, DFT)、离散余弦变换 (Discrete Cosine Transform, DCT)[27] 等，进

而通过频域特征上进行特征选择实现数据降维，再利用提取到的特征进行比较或者分类。匹配方法也可以使用 DTW 或者欧式距离，而分类方法可以使用一些机器学习方法，比如支持向量机 (Support Vector Machine, SVM)、随机森林 (Random Forest)、朴素贝叶斯 (Naive Bayes) 等[28]。这些方法一般是在公开数据集上进行实现和评估，当有新数据源时，需要设计新的合适的模型。而且，由于使用的是公开数据集，这些研究很少评估系统的鲁棒性，比如不同时间段对签名的影响。

1.3 本文研究内容与创新点

1.3.1 研究内容与所做的工作

针对现有存在研究的主要问题，本文研究设计一种非侵入式的，具有鲁棒性、安全性、低时延的在线手写签名认证方案，主要的研究内容和所做的工作如下：

(1) 手写签名认证数据源的研究

本文提出利用声波感知技术来跟踪手写签名时的手部运动和笔的运动状态，进而推断是否为本人签名。声波对不是陌生的信号，人们用声带发射声波，用耳膜接收声波，以此来实现对这个世界的感知。本文采用声波相位相关信息，利用计算弦的方法避开了求相位时去直流的问题，实现了微小动作的跟踪。

(2) 手写签名认证模型的研究

针对声波相位相关信息设计了合适的特征提取和用户独立的识别模型。手写签名动作所产生的影响在相位信号中体现为低频信号，而高频信号则是由环境或者硬件造成的噪声所产生，因此通过将时域数据转换到频域数据，然后提取低频系数的方法，将手写签名动作的信息提取出来。当判别一个查询签名是否为真的时候，需要比较查询签名的特征与模板签名特征的之间的相似度，本文对多个模板签名的特征进行融合操作，获得几个固定数目的特征矩阵，查询签名的特征矩阵和这些矩阵作差获得相似矩阵。最后采用比传统分类器有更好表现的深度神经网络 (Deep Neural Network, DNN)[29] 作为二分类器来分类相似矩阵。

(3) 性能评估分析和原型系统实现

对该方案进行了全面的评估，包括验证精度、交叉用户可用性、硬件补偿的效果、系统鲁棒性、微基准实验、对比实验和重放攻击等。准确性测试以 AUC (Area Under Curve) 和 EER (Equal Error Rate) 作为度量指标；在鲁棒性测试中，改变环境、距离、日期等来对系统进行测试；在安全性测试中，测试系统对重放攻击的抵抗能力；在微基准测试中，改变一些超参数观察系统的变化；在性能测试中，通过度量系统时延。利用三星 Galaxy S6 智能手机设计并实现基于声波的手写签名认证原型系统，对其运行效率进行了评估。

1.3.2 研究创新点

本文主要有如下两个方面的创新点：

（1）第一次提出使用声波感知技术来实现一个在线签名认证，并基于主流智能手机实现了一个非侵入式、用户友好、安全、低延迟、准确的在线签名认证系统。相比于之前的声波感知技术研究，在避免去直流的前提下提取到了相位相关信息，用以跟踪手写签名时所产生的微小的动作。

（2）使用 DCT 提取信号中低频部分作为手写签名行为的特征，使用深度学习进行分类，相比传统的分类方法获得了更好的表现。

1.4 论文结构

本文共分五章，每章的内容安排如下：

第一章为绪论，主要阐述了手写签名认证研究的背景和研究意义、国内外相关技术研究现状、本文的研究内容以及主要的创新点。

第二章是手写签名认证相关技术的分析，首先分别从手写签名认证中的感知技术和手写签名认证中的建模方法两个方面对现有的研究进行了综述，然后分析总结了现有研究存在的问题。基于存在的问题，提出了本文的技术路线，最后介绍了本文涉及到的相关理论和技术。

第三章是基于声波的手写签名认证的关键技术研究，首先对要解决的问题进行了描述，然后提出一种包含四个模块的基于声波的手写签名认证方案，然后对声波相位相关信号感知技术、特征提取技术、相似性度量技术和深度学习模型等关键技术进行了深入研究，详细介绍了方案中各模块的设计思想和技术细节。

第四章对基于声波的手写签名认证方案进行了实验验证，首先描述了实验方法和采集的数据集，提出验证指标和用于对比的方案，之后实验验证本文方案的识别性能，最后探究影响识别性能的主要因素。

第五章实现了一个基于安卓智能手机的手写签名认证系统，首先对系统的用户需求进行分析，然后完成系统架构的设计，最后评估系统的性能效果。

第六章是总结和展望，对本文的研究内容进行了总结，对未来的工作进行了展望。

第二章　手写签名认证相关技术分析

目前手写签名认证的任务呈现普适性、便利性、模型用户无关的发展趋势，对实施技术的要求越来越高，本章首先介绍手写签名认证研究综述，分析总结现有研究存在的问题，然后针对存在的问题提出本文的技术路线，最后对本文涉及到的相关理论与技术进行介绍。

2.1　手写签名认证研究综述

手写签名认证研究所涉及到的技术领域比较广泛，如传感器技术、图像技术、通信技术、机器学习等等。应用这些技术需要解决的最核心问题主要有两点，即“签名行为的数据怎么采集”和“采集到的数据怎么处理”，本文从这两方面对现有研究进行综述。

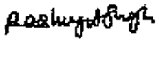
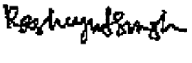
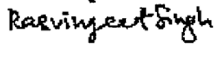
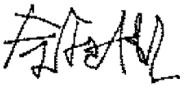
2.1.1　手写签名认证中的感知技术

(1) 基于扫描图像的手写签名认证

从图像上签名通常可以分为三类[30]：简单的、草书的、图形的签名，如表 2–1所示。简单的签名便是平时的普通签名；草书签名是写的过程有比划连在一起的签

表 2–1 真实签名与仿造签名图像[30]

Table 2–1 Genuine and forged signatures

类型	真实签名	熟练仿造签名	不熟练仿造签名
简单			
草书			
图形			

名；图形的签名则是用草书方式描述几何模式的签名。从图像上看真实签名、熟

练仿造签名、不熟练仿造签名之间有很大相似之处，但是细看不同之处也有很多，离线签名认证利用这些从纸上扫描出来的图像进行识别。

Madasu[30] 等让签名者使用黑笔在白纸上签名，之后使用扫描仪以 200 dpi 的分辨率扫描获得图像，并经过 50% 的重采样将像素点数量减少到原来的一半，40 名志愿者参与数据集构建，每名志愿者提供 15 个真实签名和 15 个仿造签名，总共 1200 个签名图像。Subhash[21] 等人邀请了 18 名志愿者，每名志愿者提供 15 个真实签名和 15 个仿造签名，总共 540 个尺寸为 $850\times360px$ 的签名图像。除了自己手机签名数据集外，目前公开的签名图像数据集有：MCYT[1]、CEDAR[2]、Brazilian PUC-PR[31]。签名的公开数据库文字上以英文居多，Brazilian PUC-PR 则提供巴西葡萄牙语的签名数据。采集签名的最传统方式是直接让志愿者写在纸上，然后扫描，这种方式耗费人力。而最近，可以让签名者将签名写在平板上，仅仅使用静态轨迹数据就可以作为离线签名的数据集，同时兼顾离线签名认证研究和在线签名认证研究对数据集的需求。

(2) 基于惯性传感器的手写签名认证

在手写签名的过程中，签名者通过手和笔的运动来留下签名字迹，运动过程中手和笔的运动加速度和方向会发生变化，因此可以通过给手或者笔配备惯性传感器（可以将惯性传感器佩戴在手上，如智能腕表，也可以将惯性传感器装在笔上）实现对此运动过程的记录。常见的惯性传感器包括加速度计和陀螺仪，这两个惯性传感器均可产生三维的时间序列数据，分别记录物体线性加速度和角运动，用于描述物体在三维空间中的运动状态。

Alona[32] 等人提出一种可穿戴的手写签名认证系统如图 2–1所示，它使用一个戴在签名手上的智能腕表上的惯性传感器和加速度计实现对签名动作的记录，虽然腕表目前的普及率还远不如智能手机，但腕表逐渐被大众所接受，所以这种方法还是具有普适性的。Isaac[33] 等人使用腕表实现了在多个场景下的手写身份识别，并对多种不同强度的攻击方式进行评估。Bunke[34] 等人在普通笔的笔尖附近贴上了一个带线的加速度计，使用如图 2–2所示，通过细带电缆将加速度时间序列传输到电脑端，据此分析签名动作。

(3) 基于手写平板设备的手写签名认证

如图 2–3所示，签名使用手写笔在平板上签名，由平板记录签名轨迹，手写屏除了能记录轨迹还能记录压力值，如果使用智能笔则还能记录笔的倾斜角，因此采集到的时间序列数据可以包括：采样点的二维坐标、笔尖压力值、笔的水平偏角和垂直偏角[35]。Alona[32] 等人在用腕表上惯性传感器跟踪签名动作的同时，记

[1]http://atvs.ii.uam.es/atvs/mcyt75so.html

[2]https://cedar.buffalo.edu/Databases/CDROM1/

图 2–1 用腕表上惯性传感器

Figure 2–1 Using inertial sensors on swatches

图 2–2 用笔上的惯性传感器

Figure 2–2 Using inertial sensors on pens

图 2–3 在平板上签名

Figure 2–3 Signing on a tablet

录签名轨迹，轨迹数据除了可以给仿造者提供仿造学习材料，还可以用于和两个经典系统[13, 14] 做对比实验。

目前，可用于在线签名认证研究的公开数据集有：MCYT-100[1]、SUSIG[36]、SVC2004[37]、SCUT-MMSIG[38] 等。

MCYT-100 是 MCYT 数据库的一个子集，包含 100 个签名者的数据，每个签名者有 25 个真实签名和 25 个仿造签名，总共 5000 个签名样本。至于 SUSIG 数据库，Kholmatov[36] 等人使用一个分辨率为 300 dpi、具有 128 层垂直压力感知、100 Hz 采样率的压力感知触摸平板，邀请 110 位志愿者参与真实签名的收集，其中包括年龄在 21 岁到 52 岁之间的 29 位女性和 81 位男性，每位志愿者分两个时间段提供 20 个真实签名，而仿造签名由仿造者观看签名过程并练习后为每个真实签名者提供 5 个仿造签名，实验结果证明签名是一个复杂度取决于签名者生物特征。SV2004 是 2004 年香港科技大学举办在线认证比赛提供的数据库，针对比赛中的两个任务，该数据库提供两个数据集，每个数据集都包含 100 个签名集合，每个签名集合包含 20 真实签名和 20 熟练仿造签名，不同的是其中一个数据集只包含

[1]http://atvs.ii.uam.es/atvs/mcyt100s.html

坐标时间序列，而另外数据集则还包含笔的方向和压力，两个数据集中前 40 个签名集合完全不同，而后 60 个集合仅仅是笔的方向和压力不同。比赛中，成绩最好的第一个任务 EER=2.84%, 第二个任务 EER=2.89%。

SVC2004 包含中文签名和英文签名，SCUT-MMSIG 则是由华南理工大学采集的纯中文签名，所以如果做针对中文签名认证研究的研究者可以考虑这两个签名数据作为评估数据集。

本领域的研究者基于公开数据集做了很多工作。Kholmatov[14] 赢得 SVC2004 的冠军。2015 年，Fischer[13] 在两个数据集上进行了评估，在 MCYT 数据集上使用 5 个模板签名时在随机仿造 (random forger) 和熟练仿造 (skilled forger) 情况下的 EER 分别为 1.06% 和 3.94%，在 SUSIG 数据集使用 5 个模板签名时在随机仿造和熟练仿造情况下的 EER 分别为 1.34% 和 3.09%。

2.1.2 手写签名认证中的建模方法

已经有一些文章对签名认证研究进行了综述，如上世纪 80 年代 Rejean[39] 等人的综述，那时研究者开始用人工神经网络实现签名认证，90 年代的 Franck[40] 和 2000 年代的 Donato 等人[41] 的综述, 在 2012 年 Donate 等人对之前的综述进行了补充[42]。近年来，Luiz G.[43] 等人对离线签名认证进行了综述，并加入基于深度学习的相关研究内容，Prathiba[44] 等人对在线签名认证进行了综述。

Kai[45] 提取了签名图像的多种粒度的局部对比几何特征，为每种粒度的特征使用一个适配的多层感知机，多个感知机的输出作为一个决策感知机的输入，决策感知机用于输出最后二分类的结果，在一个超过 3000 个样本的数据集中达到 90% 的分类准确率。

为所有用户训练的模型称为写者独立模型 (writer-independent model, WI model)，需要为每个用户训练一个的模型称为写者依赖模型 (writer-dependent model, WD model)。有些系统会使用熟练仿造签名进行用户独立模型的训练[46, 47]，有些系统使用熟练仿造签名进行用户独立模型训练[16, 48, 49] 后会再用另外一个不同的数据集进行测试。有些系统混合使用用户独立模型和用户依赖模型，使用用户独立模型用于特征提取，利用提取到的特征为每个用户训练一个小型的用户依赖模型，综合了两种模型的优点。随着深度学习在图像领域的广泛应用，现在可以用 DNN 直接从图像中学习得到特征。Hafemann[50] 使用一个卷积神经网络 (Convolutional Neural Network, CNN) 作为特征提取器，再用 SVM 为每个用户训练一个用户依赖模型，之后进一步提出了多任务的卷积神经网络框架[16]，可同时提取到区分签名真实性和用户间差异的特征。

来自签名的动态特征提供了某个时间的笔画数目和顺序、速度、笔的压力等相关信息，可以使签名唯一性得到更好的保障。Alisher[14] 的在线签名认证系统将签名的识别视为一个二分类的模式识别问题，DTW 被用于建立给定签名的合法性：给定一个签名与被申明用户的参考签名计算 DTW 距离，计算出与最近、最远、模板参考签名的 DTW 距离，生成 3 维的特征向量用于后续的分类，在使用熟练仿造者的测试中，EER 达到 2.8%。S.A. Daramolo[51] 为了建立签名的特征序列之间的一致关系，将 DTW 用于训练和分类。Alona[10] 等人结合使用 DCT 和 DTW，得到一个特征向量，最后输入到一个用户独立的二分类器进行判别。

HMM 被证明可以有效用于签名认证，因为它可以高度适应个体间差异，Mohammad M. Shafli 和 Hamid R. Rabiee[52] 介绍了使用变长分段和 HMM 实现的在线签名认证系统，实现了错误接受率 (False Accept Rate, FAR) 和错误拒绝率 (False Reject Rate) 分别为 4% 和 12% 的性能。

Syed Khaleel Ahmed[53] 等人设计的签名认证系统由 4 个模块组成，分别是特征提取模块、参考模块、样本模块、智能决策模块：特征提取模块用于捕捉二维坐标和笔的压力的时间序列；参考模块用于存储训练数据；样本模块包含用于验证的数据；智能决策模块则是一个自组织映射神经网络，用于对数据进行聚类，将高维数据映射为一维或二维数据。

2.2 现有研究存在的问题

根据上文对于手写签名认证的研究综述，对现有研究存在的问题进行分析总结。

(1) 手写签名认证的感知技术中存在的问题

目前，手写签名认证中通常采用的感知技术方案主要有三种：基于惯性传感器、基于扫描图像、基于手写平板设备的解决方案。基于惯性传感器的主要问题是需要在用户或者笔上佩戴惯性传感器：如果是笔上安装惯性传感器，则要求定制的笔，不便于推广；如果是要求用户佩戴腕表之类的设备，则该设备必须在用户写字的手上，在可穿戴设备还没大量普及的前提下该要求过于苛刻，而且腕表在手上的位置变化也会影响到判别的错误率。基于扫描图像的方法，比较符合用户的习惯，但是由于动态特征的缺失，熟练的仿造者可以花足够的时间去仿造签名以达到在静态形状上尽可能像，这给离线签名认证系统带来了巨大挑战。基于手写平板设备的方法，需要手写平板，如果用手指写则不符合用户平时的签名习惯，如果也是用笔在平板上写并且给予书写轨迹的反馈，由于需要特殊的设备而提高了普及化的难度。

(2) 手写签名认证建模技术中存在的问题

DTW 是一种非常经典用于计算两个不同长度序列之间距离的算法，其在模式识别中的应用十分广泛。然而，DTW 技术具有两个明显的缺点：1) 计算开销大；2) 对仿造签名进行规整使得验证更加困难。

而 HMM 模型是无记忆性的，它的当前状态只与其前一个状态有关，无法利用更多复杂信息。

使用深度学习进行用户无关的模型训练会引发巨大的训练开销，而且越是复杂的网络要求有足够大的数据集防止其过拟合。目前的系统大多在公开数据集进行测试，过于追求精度的提升，很少有度量在识别时间上的开销，因此复杂的方法在现实中不一定可行。

2.3 本文的技术路线

本文针对目前手写签名认证中的感知技术和手写签名认证中的建模方法的研究中存在的问题，提出了本文的技术路线，如图 2–4所示，主要是具有递进关系的四种技术研究，下面将分别介绍这四种技术。

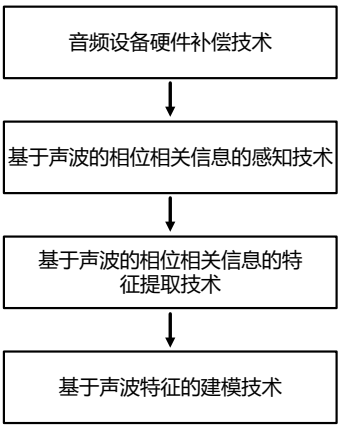


图 2–4 本文的技术路线

Figure 2–4 The technology roadmap

(1) 音频设备硬件补偿技术

本文的数据采集和原型系统实现是基于智能手机上的音频设备。然而，智能手机上的音频设备主要用于娱乐、通信等，其主要设计用途并不包括声波感知，因此当利用智能手机发送高频的声波信号 (17 kHz 以上)，尤其是同时发送多个高频声波信号时，每个频率上声波实际发送能量差异可能会很大。本文针对智能手机

上音频设备的硬件不足，进行了硬件补偿，以便目标频率声波的发射能量不至于太小。

(2) 基于声波的相位相关信息的感知技术

从麦克风设备中获得的是一个原始的音频数据，其中包括了人交谈的声音、环境噪音等人耳可听见和不可听见的声音。这些声波信号并不是本文所需要的，需要对原始的接收信号进行转换以获得两个经降采样的正交信号，这两个正交信号可用于计算相位。相位信息与声波的传播路径长度直接相关，环境中一些周期性运动例如笔记本散热器叶片的转动也会对相位信息产生影响，因此需要对两个正交信号进行去噪，然后提取相位相关的信息。

(3) 基于声波的相位相关信息的特征提取技术

声波的相位信息和声波传播路径直接相关，签名时手的运动引起声波传播路径长度的变化，从而导致相位的波动，而手的运动对相位的影响呈现为低频信号，高频信号则为噪声。本文使用 DCT 将相位相关信号从时域转换到频域，取低频系数作为特征提取和选择的结果。

(4) 基于声波特征的建模技术

本文认为签名真实性的判别是个二分类问题。对一个用户，在收集模板签名之后，根据所有模板签名的特征矩阵，计算出三个用于和查询签名比较的矩阵（最小值、最大值、平均值），和查询签名的特征矩阵通过做差得出距离矩阵。本文设计一个多层 CNN 模型，以距离矩阵作为输入进行二分类。

2.4　相关理论与技术简介

2.4.1　智能手机音频设备的位置和配置

三星 Galaxy S6 是三星公司在 2015 年 5 月推出的一款智能手机，是本文原型系统 ASSV 使用的智能手机，如图 2–5所示。该智能手机搭载了 2 个扬声器和 2

图 2–5 智能手机上的音频设备

Figure 2–5 Audio devices on smartphones

个麦克风：a) 一个降噪麦克风 (Microphone-2) 位于机身的上壁，对高频信号较为敏感，且原来用户打电话时的发生点，可以用于记录背景噪声；b) 一个主麦克风

(Microphone-1) 位于机身的下壁，对人声比较敏感 (通常低于 8 kHz)，用于记录通话语音；c) 一个通话扬声器 (Speaker-2) 位于屏幕上部，用于通话时对准耳朵；d) 一个主扬声器 (Speaker-1) 位于机身的下壁，位于主麦克风的旁边。

表 2–2 三星 Galaxy S6 的硬件配置

Table 2–2 Hardware settings of Samsung Galaxy S6

Body	Dimensions	143.4 x 70.5 x 6.8 mm (5.65 x 2.78 x 0.27 in)
	Weight	138 g (4.87 oz)
Platform	OS	Android 5.0.2 (Lollipop), upgradable to Android 8.0 (Oreo); TouchWiz UI
	Chipset	Exynos 7420 Octa (14 nm)
	CPU	Octa-core (4x2.1 GHz Cortex-A57 & 4x1.5 GHz Cortex-A53)
	GPU	Mali-T760MP8
MEMORY	Card slot	No
	Internal	32/64/128 GB, 3 GB RAM
Sound	Loudspeaker	Yes
	3.5mm jack	Yes
		24-bit/192kHz audio
		Active noise cancellation with dedicated mic
Battery		Non-removable Li-Ion 2550 mAh battery
	Charging	Fast battery charging 15W
		Qi/PMA wireless charging (market dependent)
	Talk Time	Up to 17 h (3G)
	Music play	Up to 49 h

该型号智能手机的硬件配置如表 2–2 [1]所示，其具有较好的计算性能和较大的内存空间，支持 24-bit/192kHz 的音频设备，可以满足本研究对设备的要求。

2.4.2 声波信号下转化

将智能手机上麦克风的采样率设置为 48 kHz，由于接收和发射声波的设备在同一智能手机共享一个时钟频率，因此接收端和发送端之间不存在载频偏移 (Carrier Frequency Offset, CFO)。因此，可使用图 2–6 所示的相干解调器结构将接收到的声波信号下转化为基带信号[54]，接收信号被分成两份副本，分别乘以发射信号 $-cos2\pi ft$ 和它的相位偏移信号 $-sin2\pi ft$，再经过一个低通滤波器获得同相 (In-phase) 和正交 (Quadrature) 分量。

LLAP[55] 中对信号下转化的过程进行了解释。为了更好地理解数字下转化过程，假设一个信号路径 p 的路径长度随时间变化的函数为 $d_p(t)$。来自路径 p 的信

[1]https://www.gsmarena.com/samsung_galaxy_s6-6849.php

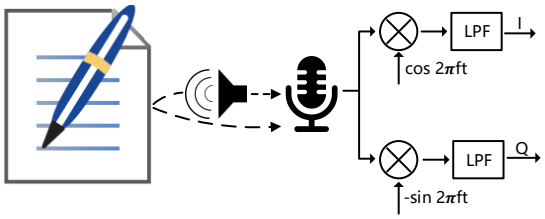


图 2–6 声波信号下转化

Figure 2–6 Sound signal down conversion

号可以表示为:

$$R_p(t) = 2A_p^{'} cos\left(2\pi ft - 2\pi f d_p(t)/c - \theta_p\right). \tag{2–1}$$

其中，$2A_p^{'}$ 是接收信号的强度，$2\pi f d_p(t)/c$ 是产生于传播延迟 $\tau = d_p(t)/c$ 的相位间隔，c 是声音传播速度。初始相位 θ_p 是硬件延迟和由于反射而发生相位反转的结果。按照图 2–6中的结构，将接收信号乘以 $cos(2\pi ft)$，得到：

$$\begin{aligned} & 2A_p^{'} cos\left(2\pi ft - 2\pi f d_p(t)/c - \theta_p\right) \times cos(2\pi ft) \\ = \quad & A_p^{'}\left(cos\left(-2\pi f d_p(t)/c - \theta_p\right) + cos\left(4\pi ft - 2\pi f d_p(t)/c - \theta_p\right)\right). \end{aligned} \tag{2–2}$$

第二个子项的频率为 $2f$，属于高频，可以被低通滤波器去除。所以可以获得基带信号的同相分量 (I-component) 为：

$$I_p(t) = A_p^{'} cos\left(-2\pi f d_p(t)/c - \theta_p\right). \tag{2–3}$$

相似地，可以获得正交分量 (Q-component) 为：

$$Q_p(t) = A_p^{'} sin\left(-2\pi f d_p(t)/c - \theta_p\right). \tag{2–4}$$

结合这两个分量，分别作为复数的实部和虚部，可以得到复数形式的基带信号 ($j^2 = -1$):

$$B_p(t) = A_p^{'} e^{-j\left(2\pi f d_p(t)/c + \theta_p\right)}. \tag{2–5}$$

因此，在路径 p 上的相位为：

$$\phi_p(t) = -\left(2\pi f d_p(t)/c + \theta_p\right). \tag{2–6}$$

当 $d_p(t)$ 变化一个声波波长的长度 $\lambda = c/f$ 时，相位 $\phi_p(t)$ 变化 2π。

2.4.3　离散余弦变换

正交变换和逆变换在维纳滤波器中充当重要角色，DCT[56] 是一种正交变换，可以用于实现一个维纳滤波器和模式识别中的特征选择。在模式识别中 DCT 可以用于特征选择，对原始数据进行降维，以便于进行分类。一个序列 $X(m), m = 0, 1, \cdots, (M-1)$ 的 DCT 可以为定义为:

$$\begin{aligned} G_x(0) &= \frac{\sqrt{2}}{M} \sum_{m=0}^{M-1} X(m), \\ G_x(k) &= \frac{2}{M} \sum_{m=0}^{M-1} X(m) cos \frac{(2m+1)k\pi}{2M}, k = 1, 2, \cdots, (M-1). \end{aligned} \tag{2–7}$$

$G_x(k)$ 是第 k 个 DCT 系数。值得注意的是，基向量集合 $\{1/\sqrt{2}, cos((2m+1)k\pi)/(2M)\}$ 实际上是一组离线切比雪夫多项式。DCT 的逆变换 (Inverse Discrete Consine Transform, IDCT) 可以被定义为：

$$X(m) = \frac{1}{\sqrt{2}} G_x(0) + \sum_{k=1}^{M-1} G_x(k) cos \frac{(2m+1)k\pi}{2M}, m = 0, 1, \cdots, (M-1). \tag{2–8}$$

在性能上，DCT 好于离散傅里叶变换，接近于最优的 Karhunen-Loeve Transform (KLT)。DCT 的算法最简单的可以根据上述公式进行计算，原文[56] 中提出了可以使用快速傅里叶变换进行高效计算，之后研究者们提出了一些高效算法[57-59]。

2.4.4　卷积神经网络

CNN[60] 是一种用于处理网格状数据的特殊 ANN，通常包括卷积和池化两类操作，在图像处理领域得到广泛应用。通常被实现为交叉相关函数的卷积操作，在一个输入二维图像 (也可以是更高维的张量) 和一个卷积核上生成一个如下的特征图：

$$S(i, j) = (K \times I)(i, j) = \sum_m \sum_n I(i+m, j+n) K(m, n). \tag{2–9}$$

其中，I、K、S 分别是输入图像、卷积核和特征图，图 2–7 演示了一个在二维张量上的卷积运算过程。稀疏交互、参数共享和等变表示是卷积运算中的三个思想，三个思想之间互相影响，卷积神经网络只需要较少的模型参数便能实现可靠地特征提取。池化函数使用某个位置的相邻输出的总体统计特征来代替网络在该位置的输出，可对输入进行降维，为神经网络提供了某种程度平移不变性，常用的池化函数有最大池化函数、平均池化函数等。

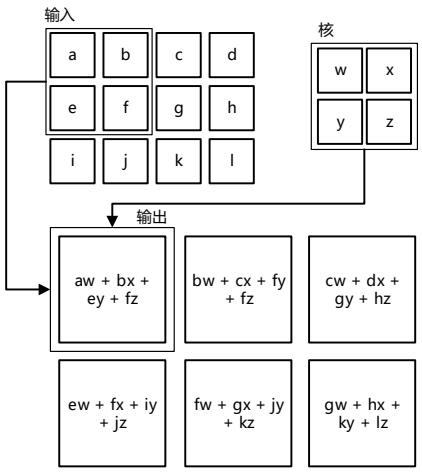


图 2–7 卷积过程

Figure 2–7 Convolution process

2.5 本章小结

本章是手写签名认证相关技术的分析，首先分别从手写签名认证中的感知技术和手写签名认证中的建模方法两个方面对现有的研究进行了综述，然后分析总结了现有研究存在的问题。基于存在的问题，提出了本文的技术路线，最后介绍了本文中所涉及到的相关理论和技术。

第三章　基于声波的签名认证方案的关键技术研究

3.1　问题描述与方案设计

3.1.1　问题描述

本文要解决的问题是利用声波的相位相关信息实现手写签名认证，提出了一种基于声波的一个非侵入式、用户友好、安全、低延迟、准确的在线签名认证方案。其中需要考虑的几个关键的技术问题是音频设备硬件补偿技术技术问题、基于声波的相位相关信息的感知技术问题、基于声波的相位相关信息的特征提取技术问题以及基于声波特征的建模技术问题。基于智能手机产品设计实现了一个手写签名认证系统，并在实际场景下部署测试了该系统。

本文列举两个签名认证的例子来描述应用场景，场景一描述生活中在银行柜台取款的常见场景便于理解手写签名认证系统的实际作用，场景二则描述了本文所设计系统的可应用场景：

（1）场景一

一位叫波波的小伙子去银行取存款，申明他是某个存款账户的拥有者，银行职员要求他签名以便进行下一步操作。根据银行的实际情况，小伙子可能会在一张普通的纸上签字也可能在一个手写平板上签字。接着手写签名认证系统运行：首先它从数据库中查询了之前该小伙子留下的参考签名；接着它使用设定的算法比较查询签名和参考签名，计算出查询签名和参考签名之间的相似度。如果系统实时运行的话，银行职员根据签名认证系统的输出结果进行下一步操作。显然，对于基于普通纸张的签名，一个静态的图像将会从纸张上扫描所得，此时签名认证系统为离线签名认证系统；然而，基于手写平板的签名，一个额外的时间信息通常被用于提交认证的进度，此时的签名认证系统为在线认证系统。

（2）场景二

本文的签名认证方案可以作为一种需要和其他认证方案相结合的辅助认证方案，例如可以和现有的离线签证认证系统或者人工签名认证相结合，最终的系统利用多模的优势提高签名认证的精度。如图 3–1 所示，当一个人在一张现金支票上的签名区域签名时候，将他/她的智能手机放在签名区域旁边记录手和笔在签名动作过程中的模式，该模式将会发送到银行进行识别；纸上的签名将被扫描作为离线签名认证系统的输入；两种签名认证方案使用不同的权重，进行融合。可在支票打印一个二维码，通过二维码可以把智能手机获得签名模式和这张支票联系

图 3–1 签名场景

Figure 3–1 Senario of signing

起来。

在两个场景中，默认参考签名已经预先被系统记录下。场景二是本文系统所适用的场景，签名者来的操作包括：获取支票、寻找签名区域、拿笔准备写、开启声波跟踪、开始手写、结束手写、结束声波跟踪、获得验证结果，如图 3–2所示。智能手机可以在场景中使用声波跟踪签名过程中手和笔的运动模式。然而使用智能手机收发声波跟踪手和笔运动实现在线认证的方案仍然面临着诸多挑战：首先，个人的签名仍然可以被观察和练习，这是目前所有签名认证系统所面临和需要克服的挑战；另外，声波可以被其他恶意的音频接收设备所接收，之后再播放出来实行重放攻击。

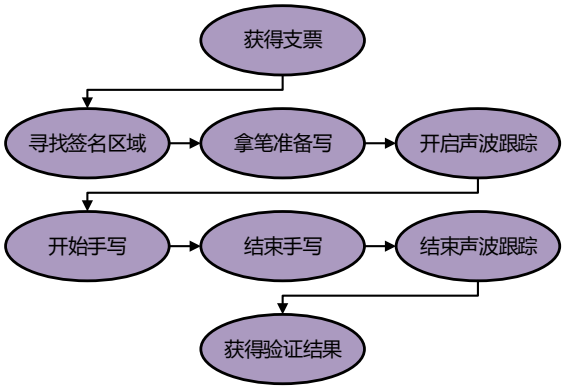


图 3–2 场景二签名步骤

Figure 3–2 Signing steps in Senario 2

3.1.2　方案设计

针对上述问题，本文设计提出了一种基于声波的签名认证方案，如图 3–3所示。该方案主要将声波感知技术引入到手写签名认证的研究中，包含四个关键步骤：声波收发、声波相位相关信息采集、特征提取、分类模型。

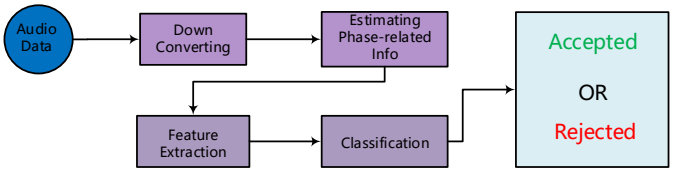


图 3–3 基于声波的手写签名认证方案

Figure 3–3 Acoustic-based handwritten signature verification method

(1) 声波收发

在发射声波之前，首先需要设计使用扬声器发射的音频文件，这个音频文件的设计需要考虑用户体验、系统可用性、系统鲁棒性这些特点，而且还需要适配扬声器的硬件特性，进行相应的硬件补偿。

(2) 声波相位相关信息采集

在收到声波反射信号之后，对声波信号进行向下转换从而获得基带信号的两个正交分量。通常在求相位之前，需要对这两个正交分量去直流，然后计算出相位信号，但是在这两个正交信号变化过于多样时，去直流变得异常困难。本文提出使用求弦长的方法，估算了相关信息 (速度和加速度)。除了解决去直流问题，还需要预先对两个正交信号进行去噪。

(3) 特征提取

尽管声波的相关信息是经过降采样的，它的数据量仍然较大，不宜直接进行分类。本步骤对信号进行频域分析并做余弦变换，选择低频信息作为特征，并计算查询签名与参考签名之间的距离矩阵。

(4) 分类模型

本文采用 CNN 作为二分类的分类模型。模型在调整好参数和使用一些防止过拟合的方法后，使用上一步骤中距离矩阵作为输入进行二分类，输出一个范围为 0 到 1 的概率值，输出值越大表明输入签名是真实签名的概率越大。

3.2　音频设备硬件补偿技术的研究

本节先介绍原型系统 ASSV 中的声波信号的设计理念，然后针对设计好的声波信号结合智能手机采取硬件补偿的措施。

3.2.1　声波信号的设计

用于在智能手机上发射的声波实际上是使用扬声器播放一个按特定目的设计的音频信号，因此关于声波信号的设计实际就是关于音频文件中二进制数据序列的设计。为了使麦克风上接收的信号可分析，本文设计一种特定的声音让扬声器发射，在此设计中主要考虑三个方面：用户体验、系统可用性、系统鲁棒性。

***(a)* 声波信号设计需要考虑用户体验**。为了让系统提供用户友好型的交互和良好的用户体验，由智能手机发送的声波不应该是人耳可听见的。根据 Rodríguez Valiente[61] 的综述，当声波的频率高于 17 kHz 的时候，声音对人便变得不可听。因此，本文生成的声波的频率均大于 17 kHz，而且这些声波均可以被商用音频设备 (包括本文所用到的智能手机上的扬声器) 发射。

***(b)* 声波信号设计需要考虑系统可用性**。与 LLAP[55] 中所发射的声波信号相似，为了测量随着传播路径变化的声波信号的相位，本研究生成声波信号的时候使用连续波 (Continuous Wave, CW) 信号:

$$Acos2\pi ft,$$

其中 A 代表信号强度，而 f 代表信号频率。再则，ASSV 使用同一个智能手机上的扬声器和麦克风来发射和接收声波，这两个音频设备在同一个智能手机上使用同一个时钟频率，所以不存在载频偏移问题。

***(c)* 声波信号设计需要系统鲁棒性**。本系统需要具有鲁棒性，即使多径效应丰富的环境下也能保持可用。如图 3–4所示，将声波的传播路径分成 4 种类型：手和笔的反射路径、直接路径、静止路径和动态多径。第一种类型，如图 3–4a所示，信号从扬声器发出，经过笔和手的反射，传播回到麦克风中。该路径长度随着手写动作的进行而发生变化，实际上，路径是跟踪签名活动中人行为特征的真正所需要的路径；第二种类型，本文称这种路径为直接路径。如图 3–4b所示，信号直接从扬声器传播到麦克风，没有经任何物体反射，因为直接传播路径使所有可能传播路径中最短的一条，所以从直接路径上传播的信号具有最大的能量，换而言之，具有最大的信号波振幅；第三种类型，本文称这种路径为静止路径，如图 3–4c所示，声波仅仅被障碍物反射了一次，图中的障碍物可以是桌子、人的身体等一些在智能手机附近可以阻挡声音的障碍物，直接路径和静止路径传播的声波信号在本文被看作背景信号，手写签名过程中手和笔的运动并不能对背景信号产生丝毫影响；第四种类型如图 3–4d所示，声波信号从扬声器出发，经过手写过程中的手和笔的反射，再次经过障碍物的反射，最后才传播回到麦克风中，因为障碍物是不确定的，通常是由形状不规则的非结构化的物体组成，所以经过这种路径传播的声波信号总体上被认为是不稳定和不确定的。

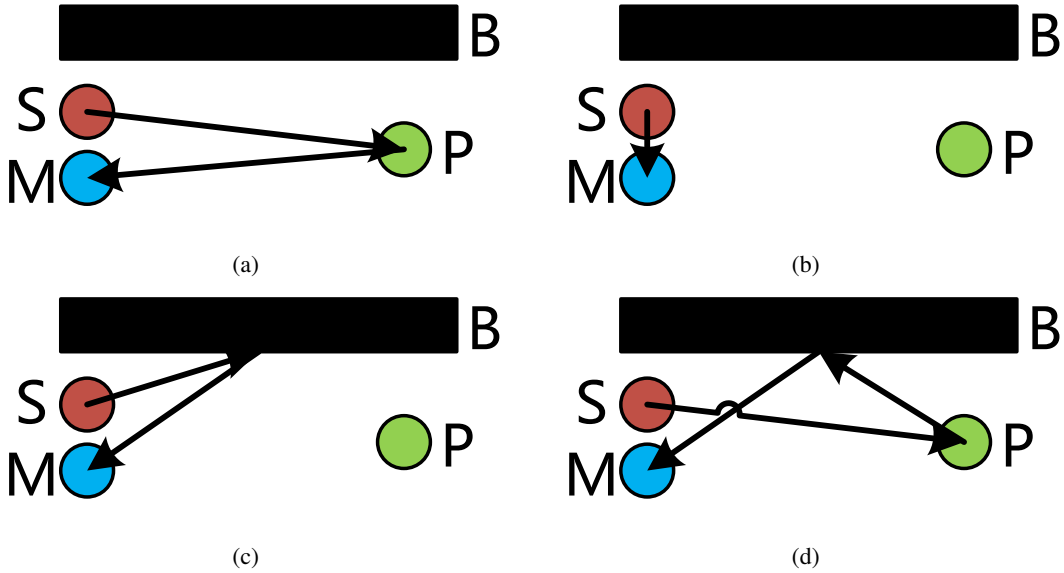


图 3–4 信号传播路径 - (a) 手和笔的反射路径; (b) 直接路径; (c) 静止路径; (d) 动态多径。**S**, **M**, **P**, **B** 分别代表扬声器 (**S**peaker), 麦克风 (**M**icrophone), 笔 (**P**en), 障碍物 (**B**lock)。

Figure 3–4 Signal traveling paths - (a) Hand and pen reflected path; (b) Direct path; (c) Static multipath; (d) Dynamic multipath. **S**, **M**, **P**, **B** represent **S**peaker, **M**icrophone, **P**en, **B**lock respectively.

本文使用信号的指数形式来解释最终的接受信号。假设 $x_i^{'}(t) = A_i^{'}e^{-a_i^{'}t}, x_j^{''}(t) = A_j^{''}e^{-a_j^{''}t}, x_m^{'''}(t) = A_m^{'''}e^{-a_m^{'''}t}, x_n^{''''}(t) = A_n^{''''}e^{-a_n^{''''}t}$ 分别表示第一种类型、第二种类型、第三种类型、第四种类型路径的信号，其中 i, j, m, n 是正整数用于表示“其中一条路径”的意思。假设 $X^{'}(t), X^{''}(t), X^{'''}(t), X^{''''}(t)$ 分别表示四种类型路径上信号的和，I, J, M, N 分别代表四种类型路径集合中各自的路径数量，可以得到：

$$X^{'}(t) = \sum_{i=1}^{I} x_i^{'}(t) = \sum_{i=1}^{I} A_i^{'}e^{-a_i^{'}t}, \tag{3–1}$$

$$X^{''}(t) = \sum_{j=1}^{J} x_j^{''}(t) = \sum_{j=1}^{J} A_j^{''}e^{-a_j^{''}t}, \tag{3–2}$$

$$X^{'''}(t) = \sum_{m=1}^{M} x_m^{'''}(t) = \sum_{m=1}^{M} A_m^{'''}e^{-a_m^{'''}t}, \tag{3–3}$$

$$X^{''''}(t) = \sum_{n=1}^{N} x_n^{''''}(t) = \sum_{n=1}^{N} A_n^{''''}e^{-a_n^{''''}t}. \tag{3–4}$$

$X(t)$ 表示麦克风接收的信号，则可以得到：

$$X(t) = X^{'}(t) + X^{''}(t) + X^{'''}(t) + X^{''''}(t). \tag{3–5}$$

将式 3–2、式 3–3、式 3–4和式 3–4代入式 3–5中，可以得到：

$$X(t) = \sum_{i=1}^{I} A_i^{'} e^{-a_i^{'} t} + \sum_{j=1}^{J} A_j^{''} e^{-a_j^{''} t} + \sum_{m=1}^{M} A_m^{'''} e^{-a_m^{'''} t} + \sum_{n=1}^{N} A_n^{''''} e^{-a_n^{''''} t}. \tag{3–6}$$

式 3–6中第一项是本研究所关心可用于跟踪手写行为的变量，第二项与第三项可以认为是常量，第四项是不稳定变化的。第二项和第三项可以通过作差法去除。第四项是在设计声波信号时需要着重考虑的。通常，结合来自不同频率上的信号的结果可以缓解第四种类型的传播路径最后对结果的影响，这种方法可以被用于提高轨迹跟踪或者活动识别的精度。

本文选择多个频率高于 17 kHz 的声波信号组成最后的发射信号。根据香浓采样定理，麦克风在采样频率为 48 kHz 时，可接收的声波频率不应高于 24 kHz，实际由于硬件原因，这个上限会更小。看似所选用的不同频率的信号越多则越有利于缓解动态多径所产生的影响，但是实际上并不能这样做，这是因为频率越多，每个频率上的信号分得的能量便越小，能量减小导致信噪比下降。本文通过实验来解释这个现象，分两次发射不同的声波，第一次发生 18 kHz 声波，第二次发射 18 kHz 和 18.6 kHz 的声波，每次发射 3 秒，取最后一秒利用快速傅里叶变换 (Fast Fourier Transform, FFT) 进行频域分析。如图 3–5和图 3–6所示，双频率声波的 18

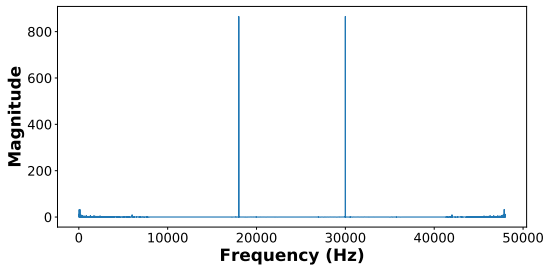


图 3–5 单频率声波频域分析

Figure 3–5 Frequency analysis on one-frequency sound

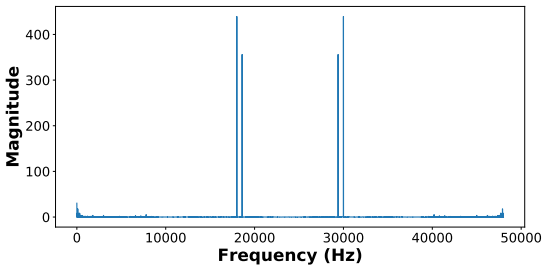


图 3–6 双频率声波频域分析

Figure 3–6 Frequency analysis on two-frequency sound

kHz 和 18.6 kHz 频率上的能量相加才是单频声波的 18 kHz 频率上的能量。本文采用 8 个频率合成发射声波，在研究中所选的频率集合 S_{freq} 可以表示为：

$$S_{freq} = \{f | f = 17350 + 700i, i \in [0, 7] \cap \mathbb{N}\}. \tag{3–7}$$

其中，8 个频率开始频率为 17350 Hz，每个频率所占频带和抗干扰保护频带分别为 200 Hz 和 500 Hz。8 个不同频率的信号被计算并标准化到取值范围 $[-1, 1]$, 如

下所示：

$$Signal = \frac{1}{|S_{freq}|} \sum_{f \in S_{freq}} cos2\pi ft. \tag{3–8}$$

3.2.2　音频设备硬件补偿

由于智能手机的扬声器不是专为声波感知而设计的，能量在不同频率的信号上不是均匀分布的，如图 3–7所示，通过对 1 秒时间的接收信号进行傅里叶变换，

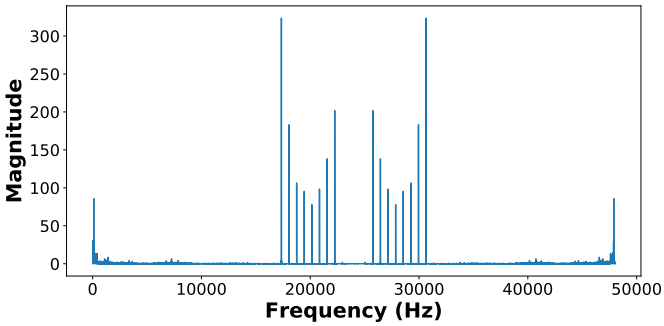


图 3–7 在硬件补偿之前的频域效果

Figure 3–7 Frequency-domain effect without dardware compensation

观察到有些频率上的信号能量十分小，而有些频率上的信号能量比较大。由于低能量的信号会导致低信噪比 (signal-to-noise ratio, SNR)，本文需要使用针对音频设备硬件的补偿技术：

（1）发射声波信号 $Signal$ 5 秒钟，挑选出第 3 秒这一秒的数据序列，这个数据序列在采样率为 48 kHz 的情况下一秒的数据有 48000 个数据点。

（2）对所选的数据序列进行傅里叶变换，从而获得所选频率上的幅度。假设 $f_i = 17350 + 700i$ 代表第 i 个频率，A_i 代表对傅里叶变换所得复数系数进行 abs 操作，abs 是一个 Python 中计算绝对值的函数。

（3）将发射声波 $Signal$ 替换为:

$$Signal_{compensated} = \alpha \sum_{i=0}^{|S_{freq}|-1} \frac{1}{A_i} cos2\pi f_i t. \tag{3–9}$$

其中 α 是幅度缩放系数以保证 $\alpha \sum_{i=0}^{|S_{freq}|-1} \frac{1}{A_i} = 1$。

3.3　基于声波的相位相关信息的感知技术研究

为了获得接收声波中的相位相关信息，需要先对接收信号进行下转化 (signal down conversion)，然后根据获得的两个正交分量计算相位相关信息。

3.3.1　信号下转化

在计算相位相关信息之前，首先需要对接收信号进行信号下转化，将信号从通带信号 (passband signal) 转化为基带信号 (baseband signal)，这个过程在 LLAP[55] 中有所描述，本文重点描述与其的不同之处。

为了易于实现 ASSV，本文使用巴特沃斯低通滤波器 (low-pass Butterworth filter) 作为所需的低通滤波器，而不是 Cascaded Integrator Comb (CIC) 滤波器。不失一般性，本文首先描述单一频率声波信号的信号下转化过程，如图 2–6所示。在此过程中，接收信号分两个分支进行操作：其中一个分支中，接收信号乘以 $cos2\pi ft$ 并紧跟一个低通滤波操作，获得基带信号的同相分量 (In-phase, I)；另外一个分支中，接收信号乘以 $-sin2\pi ft$ 并紧跟一个低通滤波操作，获得基带信号的正交分量 (Quadrature, Q)。对两个分支经低通滤波获得的输出进行降采样，将数据量缩小到原来的 1/300。既然在扬声器上的声波发射频率和麦克风上的声波接收频率均为 48 kHz，所以最后数据的采样率是 160 Hz，这其实是计算开销和判别准确性之间的一个折衷。

在实验的同时，有人在说话，有空调在运行，有个人电脑的散热器在转动，所以接收信号受到繁杂的环境噪声影响，如图 3–8所示，可以看到充满噪声的同相分

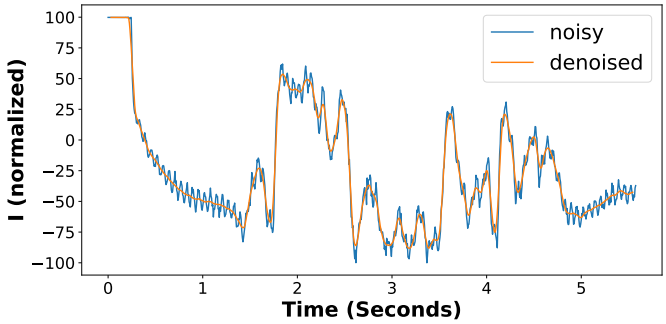


图 3–8 有噪声和去噪后的同相分量

Figure 3–8 Noisy and denoised In-phase components

量的曲线。虽然大多数噪声（比如人说话的声音）已经被低通滤波器所滤去，但是一些噪声仍然存在于基带信号中。在同相和正交分量中，可以发现一个周期性的噪声信号，此信号可以通过如下三个步骤进行测量:

（1）在没有手写动作影响的情况下，获得一段主要受到环境噪声影响的同相或者正交分量；

（2）使用一个基于函数的方法找出这段曲线中所有的波峰；

（3）计算两个波峰之间时间间隔的平均值，用于估算噪声的波动周期。

获得噪声信号的周期之后，本文直接在时域信号上使用另外一个基于函数的趋势季节分解 (Trend-Seasonal Decomposition, STD)[62] 方法，该方法使用滑动平均的方式实现，目前有乘法 STD 和加法 STD。在乘法 STD 中，原始序列为趋势序列、季节序列和残差序列之积，适用于季节的幅度随着时间变大而变大的情形；在加法 STD 中，原始序列为趋势序列、季节序列和残差序列之和，适用于季节的幅度不随着时间的变化而变化。STD 函数要求输入一个周期，即使用波峰之间时间间隔估算出的周期，分解后之后可以获得三个序列：

a. 趋势序列。趋势序列包含低频信号，本文中该低频信号随着手写动作的变化而变化；

b. 季节序列。季节是一个周期性序列，其周期是调用 STD 时所输入的值；

c. 残差序列。残差序列是一个这样的一个序列，为了满足相加模型：$Y[t] = T[t] + S[t] + e[t]$，其中，$Y[t]$ 代表原始序列，$T[t]$ 代表趋势序列，$S[t]$ 代表季节序列，$e[t]$ 为残差序列。

图 3–9展示分解效果，从上到下的子图分别是原始序列、趋势序列、季节序

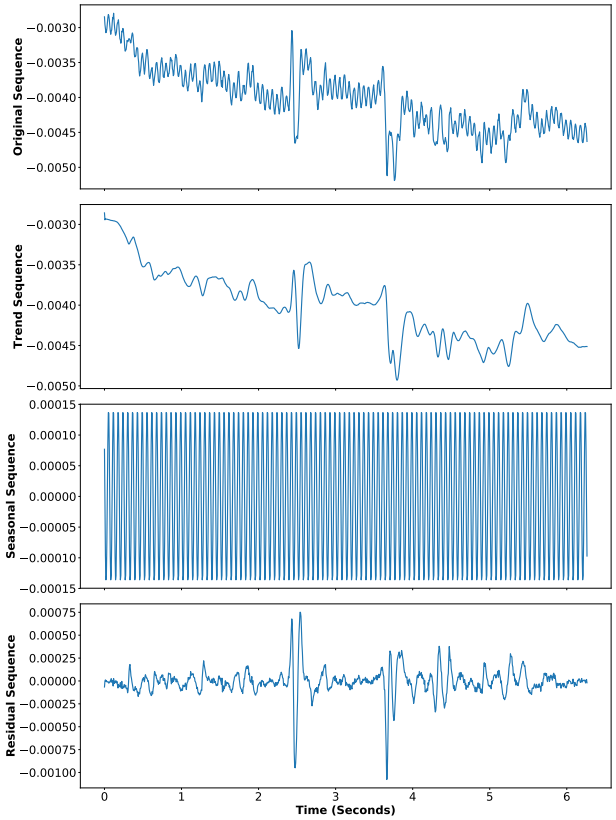


图 3–9 趋势季节分解效果

Figure 3–9 Trend-Seasonal Decomposition Effect

列和残差序列。在 STD 的输出序列中，本文只采用趋势序列作为结果序列，如图 3–8所示，充满噪声的是原始序列，而相对平滑的是趋势序列，该图展现 STD 在本文中良好的去噪声效果。

本文使用的是离线版本的 STD，只有当声波信号被全部接收和下转化之后，STD 才能对整个信号使用。既然该基于函数的方法使用滑动平均的方法，STD 实际上可以被实现成在线版本 STD2[63], 在线版本更适用于实际应用，可以减少延迟。

3.3.2　计算相位相关信息

为了更好地理解估计相位相关信息的方法，首先展示 2 个关键发现，然后使用它们的数学原理。

第一个关键发现。研究三种运动模式下的相位变化：(a) 在每个方向都有中等幅度移动的运动；(b) 在每个方向都有较小移动的运动；(c) 在每个方向只有较小幅度的移动。当智能手机上扬声器和麦克风启动，同相和正交分量的幅度从零开始持续增加到一个稳定的值，这个阶段在 ASSV 中大约持续 1 秒时间，正如图 3–10、图 3–11和图 3–12中路径 $O \rightarrow S$ 所示。这三幅图展示了三种运动模式下的 I/Q 变化轨迹，其中点 C 是运动轨迹的中心，而 x 轴和 y 轴分别代表同相分量 (I-component) 和正交分量 (Q-component)。深入地分析这三幅图，如果点 C 为圆心，可以发现圆弧的度数就是相位的大小，为了计算这个度数，$C(x_0, y_0)$ 的位置需要从同相分量和正交分量的序列估计出来。

如图 3–10所示，对于每个方向上都有大幅度移动的运动模式，这种运动模式会造成相位在同一个方向 (顺时针方向) 上运动较长时间，在这种情况下 x_0 和 y_0 通过分别计算同相分量和正交分量的平均值而得到。

如图 3–11所示，对于每个方向上都有中等幅度移动的运动模式，相位变化的方向变化较快，以致于同相分量和正交分量不是均匀分布，而是集中在 *Area 1*，但是同相分量和正交分量的最小值和最大值还是达到了，对于这种情况，第一种情况中所使用的方法不再适用，但是仅仅依赖于最小值和最大值的方法 LEVD[55] 仍然可以工作得很好。

如图 3–12所示，对于每个方向上的移动都是比较小的运动，相位的改变方向过于频繁，这种情况是手写运动中相位变化最常见的一种模式，在之前情况中所使用的方法将不适用，在这种情况下计算 x_0 和 y_0 变得很有挑战性。

第二个关键发现。研究了相位和弦长度变化之间的关系，对于半径一定的圆，越大的弦长意味着越大的圆心角，而圆心角在本文中就是相位 (图 3–10、图 3–11和图 3–12)。基于这个观察，一个衡量弦长变化的方法被提出用于估量相位相关的信

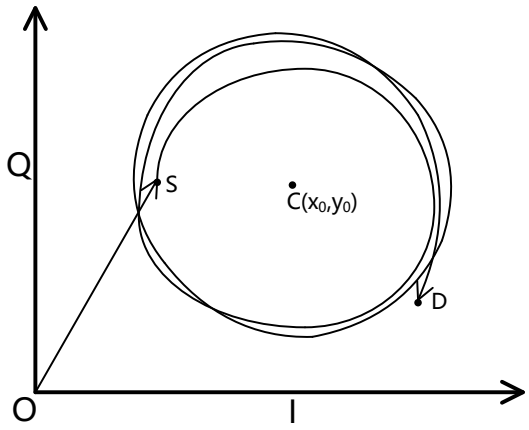


图 3–10 大幅度运动

Figure 3–10 Large-magnitude movements

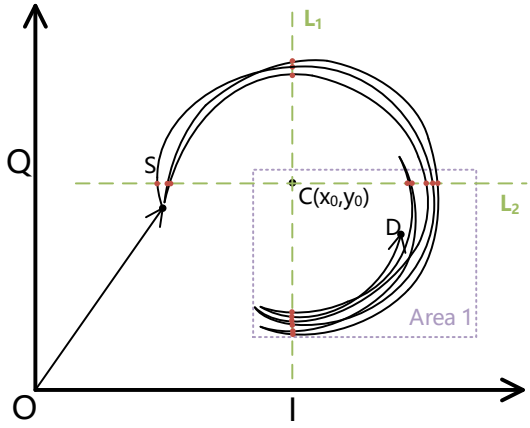


图 3–11 中等幅度运动

Figure 3–11 Medium magnitude movements

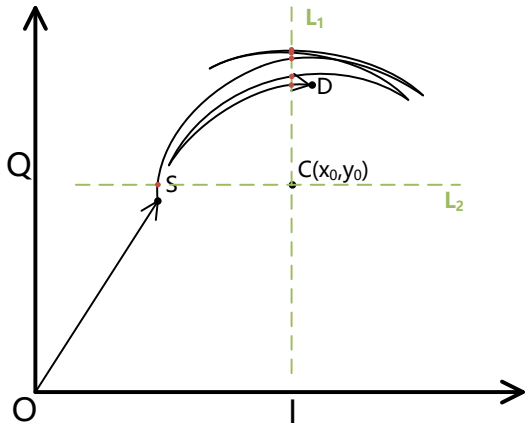


图 3–12 小幅度运动

Figure 3–12 Small magnitude movements

息，而非直接计算相位。

假设 $Q[t]$ 和 $I[t]$ 分别是正交分量和同相分量序列，其中 $t = 1, 2, 3, ..., n$。图 3–13展示 了 $I[t]$ 和 $Q[t]$ 的二维形状。P_k 是一个在 I/Q 轨迹上的离散点，它的坐标

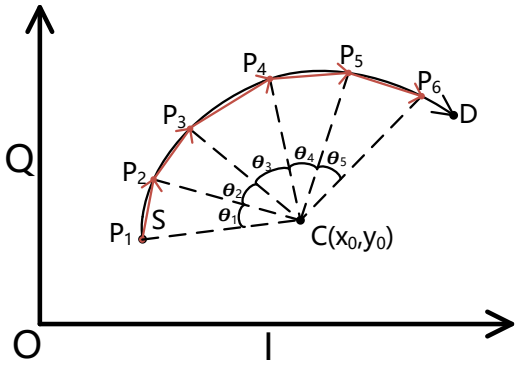


图 3–13 I/Q 估计的上离散点

Figure 3–13 Discrete points on the I/Q trace

是 $(I[t], Q[t])$, 其中 $k = 1, 2, ..., n$。θ_k 是对应于弦 P_kP_{k+1} 的圆心角。既然半径在很短的时间变化很小 (在 ASSV 中为 6 ms)，可以得出：

$$\begin{cases} \theta_k \le \theta_{k+1}, & \text{for } P_kP_{k+1} \le P_{k+1}P_{k+2} \\ \theta_k > \theta_{k+1}, & \text{for } P_kP_{k+1} > P_{k+1}P_{k+2} \end{cases}.$$

所以弦长的变化可以用于计算相位的变化。进一步，假设半径为 r，使用数学方法

证明这个推理，使其更有说服力：

$$\Delta Chord_{k,k+1} = \|P_{k+1}P_{k+2}\| - \|P_kP_{k+1}\| \tag{3–10}$$

$$= 2rsin\frac{\theta_{k+1}}{2} - 2rsin\frac{\theta_k}{2}$$

$$= 2r\left(sin\frac{\theta_{k+1}}{2} - sin\frac{\theta_k}{2}\right). \tag{3–11}$$

利用 sin 的泰勒展式的第一项，可以得到：

$$sin\frac{\theta_{k+1}}{2} \approx \frac{\theta_{k+1}}{2}, \tag{3–12}$$

$$sin\frac{\theta_k}{2} \approx \frac{\theta_k}{2}. \tag{3–13}$$

将式 (3–12)和式 (3–13)代入式 (3–11)，可以得到：

$$\Delta Chord_{k,k+1} \approx r(\theta_{k+1} - \theta_k) = r\Delta\theta_{k,k+1} \Rightarrow \Delta\theta_{k,k+1} \approx \frac{1}{r}\Delta Chord_{k,k+1}. \tag{3–14}$$

虽然半径 r 仍然是未知的，但是，考虑到当 r 保持相同的时候，$\Delta\theta_{k,k+1}$ 和 $\Delta Chord_{k,k+1}$ 成正相关变化。因为 $\Delta Chord$ 序列和相位变化的速度相关，所以相位变化相关的加速度可以通过计算 $\Delta Chord$ 的导数而进一步获得：

$$Acceleration_k = \frac{\left(\Delta Chord_{k,k+1} - \Delta Chord_{k-1,k}\right) + \left(\Delta Chord_{k+1,k+2} - \Delta Chord_{k-1,k}\right)/2}{2}. \tag{3–15}$$

相比只用两个点的计算方法，这种估计方式更具有鲁棒性。既然 $\|P_kP_{k+1}\|$ 在 ASSV 中可以被计算为：

$$\|P_kP_{k+1}\| = \sqrt{(I[k+1] - I[k])^2 + (Q[k+1] - Q[k])^2} \tag{3–16}$$

相似地，计算 $\|P_{k+1}P_{k+2}\|$，然后将 $\|P_kP_{k+1}\|$ 和 $\|P_{k+1}P_{k+2}\|$ 代入式(3–14)中，使用估计所得的 $\Delta Chord$ 和式 (3–15)，$Acceleration$ 可以被计算得到。

3.4　基于声波的相位相关信息的特征提取技术研究

3.4.1　使用 DCT 的动机

经过上一步，可以获得声波信号的基带信号，里面包含了签名运动的信息。此基带信号的采样率为 160 Hz，如果按一次签名 6 秒时间算，则一个频率上会获得 960 个数据点，8 个频率上总共上 7680 个数据点，直接使用这 7680 个数据点进行分类，会导致分类器难以收敛，分类效果不佳。本文采取对时域信号进行频域分析，

签名运动的信息通常被认为是包含在低频信号里，而噪声信息则存在于高频信号中。时频转换的常见方法有：离散小波变换 (Discrete Wavelet Transform, DWT)、离散傅里叶变换 (Discrete Fourier Transform, DFT) 和离散余弦变换 (Discrete Cosine Transform, DCT)[64]。

小波变换[65] 和短时傅里叶变换 (Short-time Fourier Transform, STFT) 具有相似之处，都可以获得时域和频域信息。如图 3–14所示，STFT 使用一个滑动窗口，对窗

图 3–14 短时傅里叶变换

Figure 3–14 Short-time Fourier transform

口内的信号做快速傅里叶变换 (Fast Fourier Transform, FFT)，从而获得时域信息和频域信息用于绘制频谱图，了解频率随时间的变化情况，其中滑动窗口大小和窗口移动步长可以进行设置。然而窗口的大小限制了离散傅里叶变换的频率分辨率，离散小波变换能更好地解决这个问题。可以通过以下方式估计一个 $f[n] \in l_2(Z)$：

$$f[n] = \frac{1}{\sqrt{M}} \sum_{k} W_\phi[j_0, k]\phi_{j_0,k}[n] + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_{k} W_\phi[j, k]\phi_{j,k}[n]. \tag{3–17}$$

其中，$f[n]$，$\phi_{j_0,k}[n]$ 和 $\phi_{j,k}[n]$ 是定义域为 $[0, M-1]$，总共 M 个点的离散函数。集合 $\{\phi_{j_0,k}[n]\}$ 和集合 $\{\phi_{j,k}[n]\}_{(j,k)\in Z^2, j\geq j_0}$ 相互正交，所以可以通过求内积的方式获得小波系数：

$$W_\phi[j_0, k] = \frac{1}{\sqrt{M}} \sum_{n} f[n]\phi_{j_0,k}[n]. \tag{3–18}$$

$$W_\psi[j, k] = \frac{1}{\sqrt{M}} \sum_{n} f[n]\psi_{j,k}[n], j \geq j_0. \tag{3–19}$$

式 3–18是近似系数，而式 3–19是细节系数。近似系数中包含低频信息，而细节系数包含高频信息，如图 3–15所示，原始信号经过小波变换可以得到近似系数和细节系数，而对近似系数和细节系数不断进行小波变换，从而得到想要的那段频率范围内的信息。

离散傅里叶变换是常用的将时频变换的方法，假设 X_k 是信号的振幅序列，则傅里叶系数可通过以下的方式求得：

$$X_k = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi k n}{N}} \quad n = 0, ..., N-1. \tag{3–20}$$

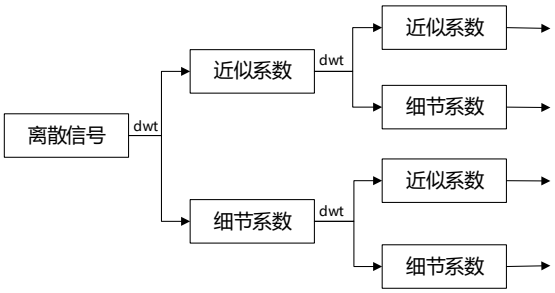


图 3–15 小波变换

Figure 3–15 Discrete wavelet transform

而离散傅里叶逆变换则是以下形式：

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k \cdot e^{j\frac{2\pi}{N}kn} \quad n = 0, ..., N-1. \tag{3–21}$$

从式 3–21可以看出，计算 DFT 的时间复杂度为 $O(n^2)$，计算过程中存在很多重复的过程数据，快速傅里叶变换 (Fast Fourier Transform, FFT) 是离散傅里叶变幻的快速算法，将时间复杂度降为 $O(nlogn)$。N 个点的傅里叶变换的频谱分辨率为 $\frac{2\pi}{N}$，这就解释了短时傅里叶变换的局限性：窗口越大傅里叶变换粒度变细，但是窗口数量减少，需要在时域信息和频域信息的多少之间进行权衡。傅里叶系数是复数系数，从式子 3–20可知，第一个系数实际为所有数据点的和，傅里叶系数存在冗余。

离散余弦变换与傅里叶变换相关，最初的算法就提出了使用傅里叶变换实现余弦变换，余弦变换是一种无冗余系数的实数变换，详细参考 2.4.3节。DCT 的特点是前面系数包含低频信息，后面的系数包含高频信息，可以通过获取前几个系数来获得低频信息，不像离散小波变换需要经过多次变换才能获得所需的信息。因此，本文采用余弦变换用于特征提取和选择。

3.4.2　特征矩阵和距离矩阵

在描述特征提取和选择过程之前，首先介绍两种矩阵：特征矩阵和距离矩阵。对于特征矩阵来说，每个手写签名都具有一个特征矩阵，这个矩阵包含所有提取和选择到的特征。对距离矩阵来说，它度量了查询签名和参考签名之间的相似度。通常认为，如果一个查询签名和相对应的参考签名具有较高的相似度，那么这个查询签名更有可能是一个真实签名，如果一个查询签名和相对应的参考签名具有较高的不相似度，那么这个查询签名更有可能是一个假签名。

对于每个接收的声波信号序列，因为在发送端有 8 个频率被使用到，因此可以获得 8 对 $\Delta Chord$ 和 $Acceleration$ 序列。先将每个 $\Delta Chord$ 或者 $Acceleration$ 序列正规化到 0-1 的范围，然后采用 DCT 来获得信号的频域特征。DCT 旨在实现数据的降维，它在信息包装和计算复杂度上有很好的权衡。不相关的 DCT 系数被产生，其中 DCT 系数的数量为输入序列的数据点数，前几个 DCT 系数通常被选取，而通常包含噪声信息的高阶系数通常被删除。结果，每个不同长度的时间序列都可以被转化为固定维度的特征向量，对于每个速度序列和加速度序列，本研究经验性地选择前 10 个 DCT 系数。

（1）特征矩阵

假设 W 是对于每个用户的参考签名数量。考虑 8 个频率，则特征向量的数量为 $16*W$, 其中包括 $8*W$ 速度特征向量和 $8*W$ 个加速度特征向量。对于每个参考签名，有 8 个速度特征矩阵，一个由这 8 个速度矩阵组成的形状为 (8,10) 的特征矩阵可以表示为：

$$M_i^{vel} = \left(m_{i,j,k}^{vel}\right)_{8\times 10} \quad i = 1, 2, .., W. \tag{3–22}$$

接下来计算：

$$a_{j,k}^{vel} = min\{m_{i,j,k}^{vel} | i = 1, .., W\} \tag{3–23}$$

$$b_{j,k}^{vel} = max\{m_{i,j,k}^{vel} | i = 1, .., W\} \tag{3–24}$$

$$c_{j,k}^{vel} = \frac{1}{W}\sum_{1}^{W} m_{i,j,k}^{vel}, \tag{3–25}$$

进而得到三个矩阵：

$$A^{vel} = \left(a_{j,k}^{vel}\right)_{8\times 10} \tag{3–26}$$

$$B^{vel} = \left(b_{j,k}^{vel}\right)_{8\times 10} \tag{3–27}$$

$$C^{vel} = \left(c_{j,k}^{vel}\right)_{8\times 10}, \tag{3–28}$$

其中，矩阵 3–26是最小值矩阵，矩阵 3–27是最大值矩阵，矩阵 3–28是平均值矩阵。

对于加速度序列有:

$$M_i^{acce} = \left(m_{i,j,k}^{acce}\right)_{8\times 10} \quad i = 1, 2, .., W. \tag{3–29}$$

进而得到三个矩阵：

$$A^{acce} = \left(a_{j,k}^{acce}\right)_{8\times 10} \tag{3–30}$$

$$B^{acce} = \left(b_{j,k}^{acce}\right)_{8\times 10} \tag{3–31}$$

$$C^{acce} = \left(c_{j,k}^{acce}\right)_{8\times 10} \tag{3–32}$$

其中，矩阵 3–30是最小值矩阵，矩阵 3–31是最大值矩阵，矩阵 3–32是平均值矩阵。

结果，每个注册用户都有一个保存在本地文件系统中的参考矩阵:

$$S_{reference} = \left\{a_{j,k}^{vel}, b_{j,k}^{vel}, c_{j,k}^{vel}, a_{j,k}^{acce}, b_{j,k}^{acce}, c_{j,k}^{acce}\right\}. \tag{3–33}$$

（1）距离矩阵

当收到一个需要进行验证的查询签名的时候，ASSV 首先计算该签名的两个特征矩阵：M^{vel} 和 M^{acce}，然后计算 6 个距离矩阵：

$$M^{vel} - A^{vel} \tag{3–34}$$

$$M^{vel} - B^{vel} \tag{3–35}$$

$$M^{vel} - C^{vel} \tag{3–36}$$

$$M^{acce} - A^{acce} \tag{3–37}$$

$$M^{acce} - B^{acce} \tag{3–38}$$

$$M^{acce} - C^{acce} \tag{3–39}$$

结合式 3–34、3–35、3–36、3–37、3–38和 3–39，获得一个三维的形状为 $(8, 10, 6)$ 的矩阵，该矩阵将被作为分类模型的输入，用于训练和测试。

总而言之，特征矩阵用于进一步产生距离矩阵，而距离矩阵则被作为分类模型的输入。

3.4.3　算法

到目前为止，结合上文音频设备硬件补偿技术的研究 3.2和基于声波的相位相关信息的感知技术的研究 3.3，计算特征矩阵、参考矩阵和距离矩阵。算法 A–2（参见附录）用于计算特征向量。基于算法 A–1（参见附录）, 可得出计算特征矩阵的算法。有了特征矩阵之后，可以计算参考矩阵，根据参考矩阵和查询签名的特征矩阵计算距离矩阵。

3.5　基于声波特征的建模技术研究

本文采用深度卷积神经网络作为二分类模型，相比传统机器学习分类器具有更优的分类性能。在距离矩阵经由计算得到后，一个二分类模型被训练用以判断给定的签名是否为真实签名，本文使用一个深度卷积神经网络以达到鲁棒的分类效果，该分类模型是一个用户独立的分类模型，当有新用户注册时，不需要重新训练模型。CNN 是一种特殊的人工神经网络，旨在处理网格状的数据，它通常包含卷积操作和池化两类操作，图 3–16展示了本文设计的 CNN 模型。

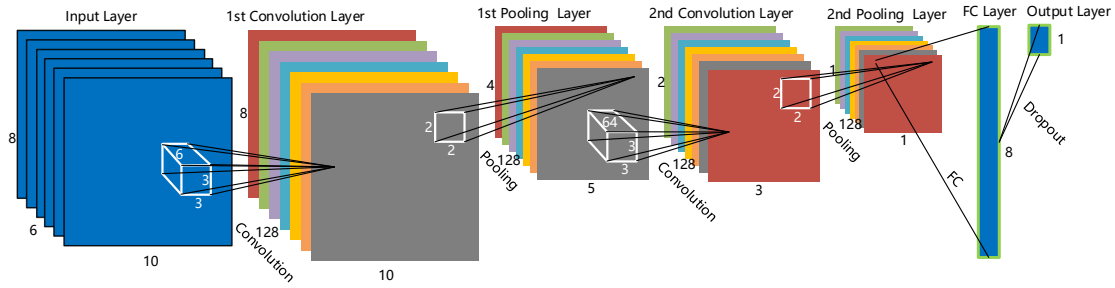


图 3–16 CNN 模型架构

Figure 3–16 CNN achitecture

首先，距离矩阵作为一个三维张量输入第一层卷积层，该卷积层的激活函数是表达式为 $max(0, x)$ 的 *RELU*。填充 (padding) 方式为 *SAME*, 用以保持输入尺寸与输出尺寸一样。第一层卷积层的输出被作为输入，输入到第一层池化层中，用以降采样操作，减小模型容量。池化层的输出作为下一层卷积池化层的输入。值得注意的是，每层卷积层的卷积核大小和数量被经验性地设为 3 和 128，本文采用的池化类型为 *max pooling*, 这是一种最常用的池化操作，池化层的池化大小为 2。虽然两重卷积池化层具有相似的配置，但是它们在模型中扮演着不同的角色：第一层卷积层旨在收集低层的特征信息，而第二层卷积层旨在收集较高层的特征信息。在选择卷积核数量的时候，本文在特征数量和计算开销之间进行了折衷，此外过多的卷积核更容易导致过拟合。

其次，最后一层卷积池化层的输出被铺平成一个向量。为了让 CNN 模型有更好的收敛速度，添加了一个批正规化层 (Batch Normalization, BN)[66]，它不仅解耦了最后一层卷积池化层，而且保留了数据的表示能力。并且 BN 具有正则化的效果，不同小批之间差异性造成的随机噪声可以增加 CNN 模型的泛化性能。BN 层的输出流入具有 8 个激活函数为 *RELU* 的神经元的全连接层。为了进一步增加模型的泛化性能，采用了正则化的方法对核权重矩阵施加了一个 $\alpha = 0.001$ 的 L^2 的正则化，对全连接输出层施加了一个 $\alpha = 0.001$ 的 L^1 的正则化。此外，全连接层

被施加了另外一个正则化技术，dropout[67]，它通过以一定概率丢弃神经元来防止过拟合，本文经验性设置 dropout rate 为 0.2。

最后，施加了 dropout 的输出被输入到输出层里，输出层只有一个激活函数为 *Sigmoid* 的神经元，输出值的取值范围在 0 到 1，代表了查询签名为真实签名的概率。与输出层的结构相对应地，二元交叉熵损失 (binary cross entropy loss) 被用于训练 CNN 模型。

3.6　本章小结

本章是基于声波的签名认证的关键技术研究，首先对要解决的问题进行了描述，然后提出一种基于声波的签名认证方案，该方案主要包含四个模块，对应了四种关键技术，依次对涉及到的音频设备硬件补偿技术、基于声波的相位相关信息的感知技术、基于声波的相位相关信息的特征提取技术和基于声波特征的建模技术进行深入研究，详细描述了方案中各模块的设计思想和技术细节。

第四章　基于声波的签名认证方案的实验

本章将对基于声波的签名认证方案进行实验验证以及分析。首先介绍在一定实验配置下的数据采集过程，然后展示评估实验及其相应的结果，这些评估实验包括：验证精度、交叉用户可用性、硬件补偿的效果、系统鲁棒性、微基准实验、对比实验和重放攻击。

4.1　实验准备

使用 Kotlin 语言开发了一个运行在安卓平台上的数据采集应用程序，将其安装在三星 Galaxy S6 上，使用主扬声器播放生成的 WAV 格式的音频文件，使用主麦克风接收声波。为了高效地采集数据，接收的声波数据首先保存在智能手机的本地存储空间里，等一个或者多个用户完成试验后，通过局域网传输到电脑上。所有分析过程和生成 WAV 文件的函数都使用 Python 语言实现，运行在电脑上用于分析和评估。CNN 模型的训练和评估得到一台笔记本上的 NVIDIA GeForce GTX 860M GPU 的加速，这台笔记本还搭载了 16GB RAM 和一颗 Intel Core i7-4710MQ 2.5GHz CPU。本研究实现了一个原型系统——ASSV，该原型系统由两部分组成 1）一个安卓应用程序 (apk 大小:38.9 MB)，执行的任务包括：数据采集、预处理和特征提取；2）一个运行在服务器上的分类器，它接收来自安卓应用程序的距离矩阵并返回结果。

志愿者受邀请到配合做实验，实验环境中有正在运行的空调、正在工作的电脑散热器和正在走路或者讲话的人员，实验环境，充满噪声，给系统带来了较大挑战。在做实验之前，每个志愿者会得到 10 分钟的训练，以熟悉采集程序并知道如何使用它。图 4–1展示应用程序的用户界面，它支持的操作包括：开始采集，停止采集，输入保存接收的声波数据的文件名和路径上的两个目录名，删除所有的声音文件，以表格形式列出已经保存的声音文件，显示播放时间等。要求志愿者将智能手机放在一张经过精心设计的 A4 的纸上，该 A4 纸的每一面上包含 28 个签名区域，如图 4–2所示，每个签名区域包含两条线段：一条 9 毫米长的分隔线段 (separation line) 和一条 38 毫米长的写线段 (writing line)，分隔线段的一端有一个十字，用于指引将智能手机的麦克风放在十字上方，并且它的底部边界与分隔线段和写线段所在的直线重合。在正确放置智能手机之后，志愿者按四个步骤在写线段上签名，如图所示 4–3：

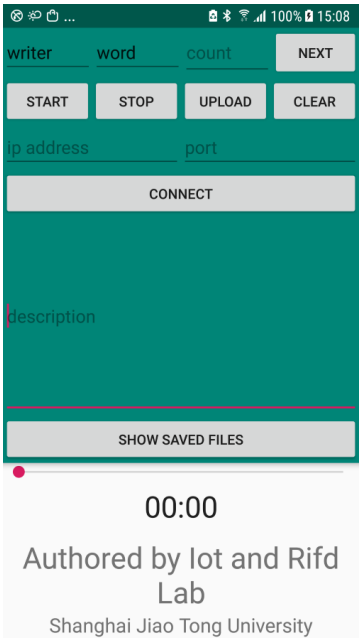


图 4–1 应用程序界面

Figure 4–1 App UI

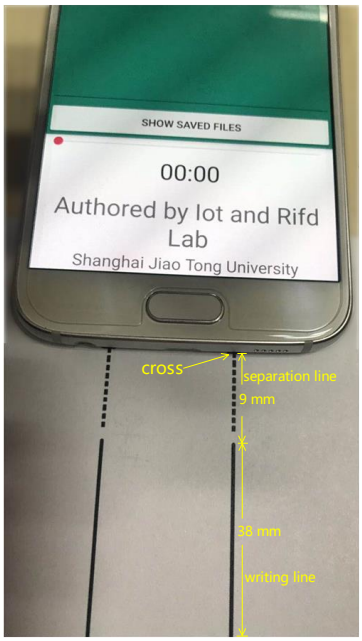


图 4–2 智能手机设置

Figure 4–2 Smartphone setting

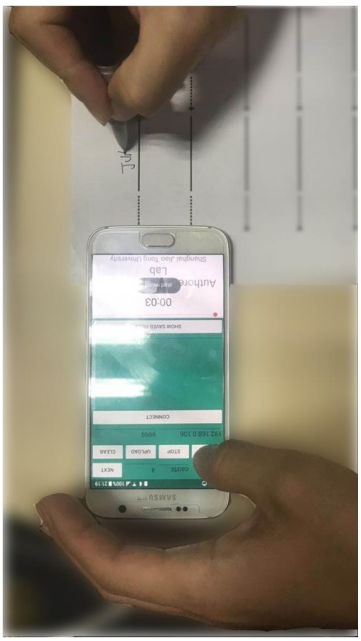


图 4–3 数据采集

Figure 4–3 Data collection

1) 右手握住笔，做好准备在写线段上开始写的姿势；

2) 使用左手点击 *start* 按钮；

3) 根据手机界面上的时间显示，在 1 秒钟之后开始写；

4) 在书写动作结束之后，使用左手点击 *stop* 按钮。注意，每写完一个签名，手机界面上的计数器会自增一次，计数器上的数字组成了保存声波数据文件的文件名前缀。

14 名研究生志愿者被邀请参加两阶段的签名实验，14 位志愿者包含 3 名女生和 11 名男生，所有的志愿者习惯使用右手写字。两阶段的签名实验过程如下：

（1）在第一阶段，每个志愿者被邀请书写他/她的签名。每个志愿者会写 112 个签名，大概耗费 30 分钟。

（2）在第二阶段，随机地打乱志愿者的名单以获得一个循环队列。向志愿者展示循环队列中在他/她之后的五个志愿者的签名记录。为了成为一个熟练的模仿者，志愿者必须照着被模仿者的签名记录练习模仿签名，以致于模仿的签名和被模仿者的足够相似，然后为每个选中的被模仿者提供 12 个模仿签名。

总之，每个志愿者有 112 个真实签名和 60 个熟练仿造签名。

两星期后又另外邀请了 22 位本科生志愿者帮助做签名实验，由 4 名女生和 18 名男生组成，他们都是习惯使用右手写字，限于每位学生只有 15 分钟的实验

时间，每位本科生被要求写 32 个真实签名和 16 个熟练模仿签名，后 18 位学生志愿者，每人需要模仿他/她前面 4 个同学的签名，为每个被模仿者提供 4 个熟练模仿签名，因此对于前 18 个志愿者，他们每人有 32 个真实签名和 16 个熟练模仿签名。最后，没有使用后 4 位同学提供的真实签名。由于每次实验的时间限制，这个本科生的数据集比研究生的数据集质量差，但是这个数据集可以用作交叉用户可用性测试。

4.2　硬件补偿评估实验

为了评估启发式方法的硬件补偿方案，本研究生成了两个 WAV 格式的音频文件：一个使用了硬件补偿，另外一个没有使用硬件补偿，每个音频文件以 48 kHz 的频率播放 6 秒钟。在每种情况中，截取第 3 秒记录的声波数据，对这些数据做 FFT，获得 48000 个输出系数，考虑在 8 个频率上的 FFT 系数幅度。如图 4–4所示，当没有使用硬件补偿方法时，最小值、最大值、极差和标准差分别是 77.8、323.5、245.7 和 76.4。相反地，如图 4–5所示，当使用硬件补偿方法时，最小值、最大值、极差和标准差分别是 137.2、164.4、27.1 和 8.5。很明显，硬件补偿方法大大减少了极差和标准差。进一步，由于过小的幅度导致过小的信噪比，因此该硬件补偿方案通过增加最小值的幅度还可以避免一种情况：在任何一个频率上的信噪比都不至于过小。

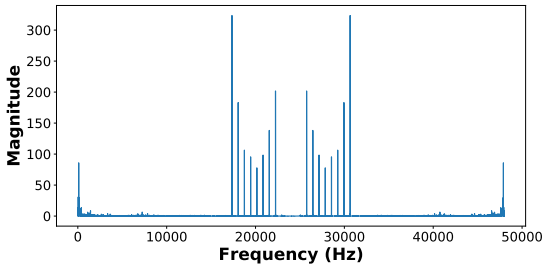


图 4–4 没有硬件补偿时的 FFT 系数幅度

Figure 4–4 FFT coefficient magnitudes w/o compensation

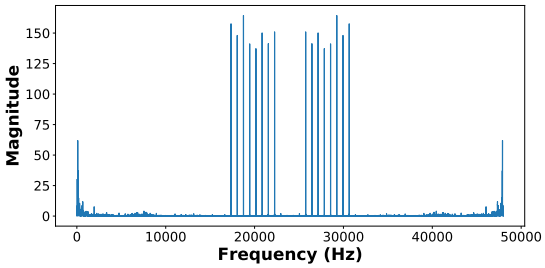


图 4–5 有硬件补偿时的 FFT 系数幅度

Figure 4–5 Smartphone setting

4.3　精度评估实验

与相关研究[10, 13] 类似，本研究使用两个度量指标衡量验证准确度：ROC 曲线下方的面积 AUC 和等错误率 EER。EER 是 ROC 曲线中 FAR(false acceptance rate) 和 FRR(false rejection rate) 相等的点，一个出色的手写签名认证系统应该具备较高

的 AUC 和较低的 EER。在本节评估中，使用研究生数据集进行分析和评估，展示在熟练仿冒者、随机仿冒者和两者结合这三种情况下的两个度量指标的值。给定一个志愿者，通过在其余志愿者的真实签名集中随机挑选 4 个签名，来组成该志愿者的随机仿造签名的集合，所以每个志愿者具有 112 个真实签名，60 个熟练仿造签名和 65 个随机仿造签名。

打乱每个研究生的真实签名，前 20 个真实签名被用作参考签名，它们被用于产生参考矩阵 $S_{reference}$(§3.4.2)，接下来的 90 个签名被用作查询签名，和参考矩阵相结合，用于产生距离矩阵。随机地从熟练模仿签名中挑选 45 个，从随机模仿签名中挑选 45 个，这样每名研究生志愿者总共有 180 个签名，包括：90 个真实签名、45 个熟练模仿签名和 45 个随机模仿签名，将总共拥有 $180 \times 14 = 2520$ 个带标签记录。再次将这些记录打乱，将它们分为：训练集和测试集两个部分，训练集有 1764 条带标签记录，占总记录数的 70%，测试集有 756 条带标签记录，占总记录数的 30%。为了绘制 ROC 曲线，通过改变 CNN 模型输出值 (0 到 1) 的阈值，获得一系列的 TPR(true positive rate) 和相对应的 FPR(false positive rate)。给定一个查询签名，如果 CNN 模型的预测值大于选中的阈值，它将被判别为真实签名，否则将被判定为仿造签名。图 4–6展示 了三条 ROC 曲线，包括：熟练仿造签名情况

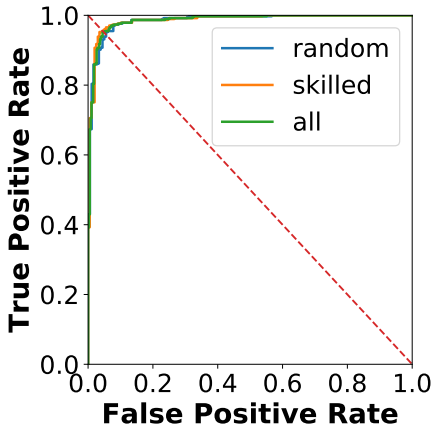


图 4–6 三种仿造类型的 ROC 曲线

Figure 4–6 ROC curves for the three types of forgers

下的 ROC 曲线、随机仿造签名情况下的 ROC 曲线和两种都有的情况下的 ROC 曲线。为了计算三种情况下的 AUC 值，计算三条 ROC 曲线下方的面积，为了获得三种情况下的 EER 值，在单元正方形中绘制一条主对角线，从该对角线与 ROC 曲线的交点可以获得 EER 值。

将以上过程重复了 20 次，从而可以求得 AUC 和 EER 值的平均值。如图 4–7所

示，在随机仿冒的情况下 AUC=0.987，在熟练仿冒的情况下 AUC=0.987，在两者

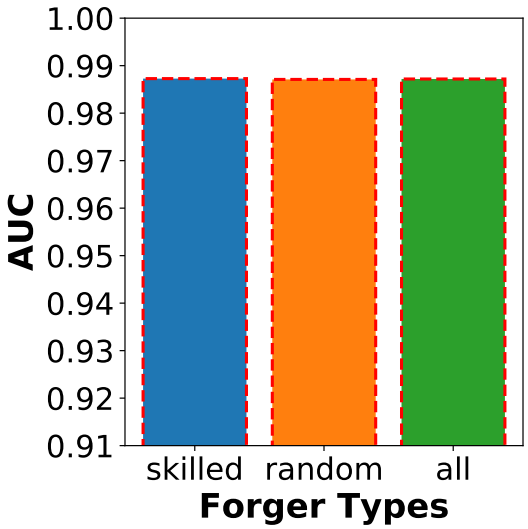


图 4–7 三种仿造类型的平均 AUC 值

Figure 4–7 Mean AUCs for the three types of forgers

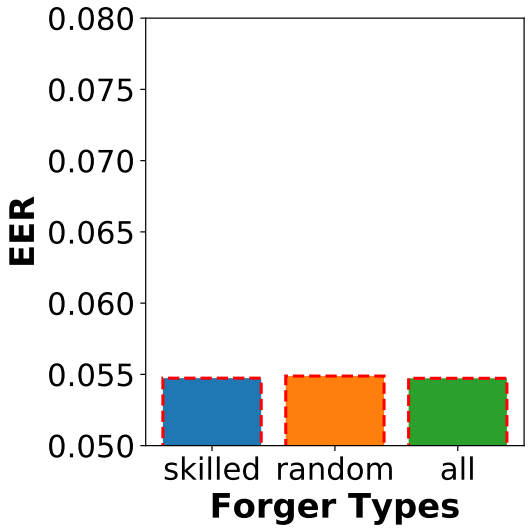


图 4–8 三种仿造类型的平均 EER 值

Figure 4–8 Mean ERRs for the three types of forgers

结合的情况下 AUC=0.987。如图 4–8所示，在随机仿冒的情况下 EER=0.055，在熟练仿冒的情况下 EER=0.055，在两者结合的情况下 EER=0.055。通常地，因为熟练仿造的签名往往与真实签名在形状上非常相似，所以当使用扫描图像识别熟练仿造签名的离线签名认证系统会遭受巨大挑战，然而 ASSV 在熟练仿造签名识别中表现出了良好的性能。

4.4　交叉用户可用性

关于交叉用户可用性的实验中使用的 CNN 模型，使用研究生志愿者的数据集作为训练集，使用本科生志愿者的数据集作为测试集。这种类型的实验经常被现有的研究工作所忽略，这些研究的测试集和训练集中的用户有重合。本研究的系统模型是一个用户独立的模型，当有新用户注册的时候，可以不需要额外的高开销的训练过程，所以进行一个交互用户可用性测试是非常有必要的。在实验中采用 10 个参考签名通过重复切分数据集和训练操作，得到 20 个模型，然后建立测试数据集，对于每个测试用户，随机挑选 10+20 个真实签名 (10 个参考签名和 20 个真实签名)，10 个熟练仿造签名，10 个随机仿造签名 (挑选五个其他用户，每用户提供 2 个真实签名)，最终拥有 $22 \times (20 + 10 + 10) = 880$ 个距离矩阵用于测试。重复以上生成距离矩阵集合的过程 10 次，获得 10 个不同距离矩阵集合。

接着实施评估，$20 \times 10 = 200$ 个 AUC 和 EER 对被获得而用于计算平均 AUC 和 EER 对，如表 4–1和表 4–2所示， 相比于非交叉用户评估，交叉用户试验的性

表 4–1 交叉用户可用性 - EER

Table 4–1 Cross-user usability - EER

Model	Random(%)	Skilled(%)	All(%)
No-cross-user	6.1	5.8	5.8
Cross-user	14.1	19.9	17.1
Retrained	4.4	9.4	7.9

表 4–2 交叉用户可用性 - AUC

Table 4–2 Crooss-user usability - AUC

Model	Random(%)	Skilled(%)	All(%)
No-cross-user	98.3	98.5	98.4
Cross-user	92.8	88.3	90.5
Retrained	99.0	97.6	98.2

能有所下降，但是系统仍然保持不错的交叉用户可用性。

将测试模型用作一个预训练模型，测试集中 20% 的数据用于再次训练模型从而获得一个新的模型，这个模型使用剩下的 80% 的数据进行测试，结果发现性能得到了很大的提升：第一种情况中，EER 达到 4.4%，AUC 达到 99.0%；第二种情况中，EER 达到 9.4%，AUC 达到 97.6%；第三种情况中，EER 达到 7.9%，AUC 达到 98.2%。

4.5 鲁棒性评估实验

为了测试系统的鲁棒性，本研究做了三种实验：不同环境下的实验、不同距离下的实验和不同日期下的实验，邀请了 4 位志愿者参与鲁棒性实验，包括 2 位男性和 2 位女性。在每种实验的每种情况下，每位志愿者被要求写下 20 个他/她的签名。在同一个鲁棒性实验中的所有情况中，选择其中一个情况中所得的签名用作参考签名，人工地选择那个用作参考签名的情况。实际上，一个适应算法可以在未来被提出用于为每个用户选择合适的签名，这个超出了本文的研究范围。记录分类器的每个输出 (0-1)，对于每个查询情况输出值的中位数规整为 0-100 的度

量值，越高的度量值意味着查询签名和参考签名之间的相似度越高，进一步意味着系统更具鲁棒性。三种鲁棒性实验的具体情况如下：

（**1**）**不同环境**: 每名志愿者被要求在三个不同的环境下进行签名，这三个环境包括：安静的环境，有播放音乐的环境，有人在旁边行走的环境。在安静的环境中，没有人在房间内说话，也没有人走路；在有播放音乐的环境中，另外一个智能手机放在是实验智能手机的旁边播放音乐；在有人在旁边行走的环境中，有人在实验者签名的时候在其旁边行走。

（**2**）**不同距离**: 每个志愿者被要求在改变智能手机到签名线的距离时，进行签名。这些距离包括：3cm, 6cm,9cm。

（**3**）**不同日期**: 每个志愿者被要求在不同的日期下进行签名，总共有 3 个不同的日期。

表 4–3显示 了鲁棒性实验结果。在改变环境或日期的情况下，本文中的系统

表 4–3 系统鲁棒性 - **RS** 意思是 *reference signatures*(参考签名)

Table 4–3 System robustness - **RS** means *reference signatures*

Volunteer	**Environment**			**Distance**			**Day**		
	Quiet	*Music*	*Walk*	*3 cm*	*6 cm*	*9 cm*	*Day 1*	*Day 2*	*Day 3*
No. 1	RS	93.3	86.1	**4.8**	91.4	RS	RS	**67.1**	93.4
No. 2	RS	81.2	93.3	**53.3**	93.6	RS	RS	94.0	95.4
No. 3	RS	94.4	97.3	**57.3**	89.8	RS	RS	94.4	96.7
No. 4	RS	83.2	84.8	**4.4**	84.9	RS	RS	84.6	90.1

保持了鲁棒性。然而，当将智能手机和签名线之间距离设置为 9cm 的情况下所得的签名作为参考签名，发现在 3cm 的情况下所得的签名作为查询签名时效果较差，而在 6cm 的情况下所得的签名作为查询签名时的效果还算可以，由此可知，当查询签名和参考签名相距为 6cm 时，系统性能变差。因此，建议查询签名和参考签名的距离最好控制在 3cm 以内。

4.6　微基准测试

ASSV 在通过改变一些内部参数的情况下被评估，这些参数包括：参考签名的数量，CNN 卷积核的数量，训练集大小，DCT 系数数量，声波频率数量和分类器类型。

4.6.1　参考签名数量

为了评估在签名认证时，参考签名数量对 AUC 和 EER 结果的影响，选择在参考签名数量在 1 到 22 时测试系统。对于每个选择的参考签名数量，通过重复实验 (参考 §4.3) 来得到平均的 AUC 和 EER。图 4–9 和图 4–10显示 了改变参考签名

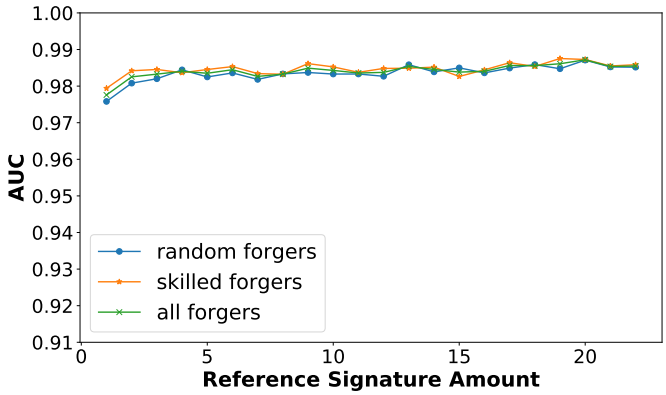


图 4–9 参考标签数量影响 - AUC

Figure 4–9 Reference amount effect - AUC

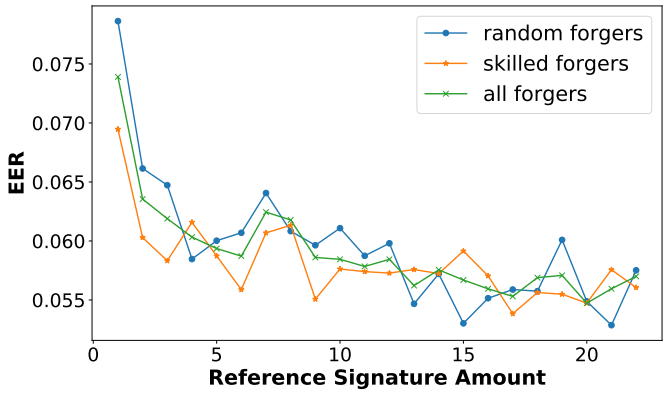


图 4–10 参考标签数量影响 - EER

Figure 4–10 Reference amount effect - EER

数量的试验结果。总体上，当参考签名数量增加，系统性能变好。当仅仅使用一个参考签名时，AUC=0.978 和 EER=0.074，当使用 6 个参考签名时，AUC=0.984 和 EER=0.059，可见当参考签名从 1 个增加到 6 个时，系统性能快速改善。在那之后，随着参考签名数量的增加，系统改善幅度变化不明显，意味着：通过增加参考签名所得到的益处开始收敛。所以本系统只需要很少的参考签名数量，就能

达到良好的性能，好处是在用户注册阶段，新用户只需要花费较少的时间去写参考签名，这样使得本文的系统更加用户友好。

4.6.2　CNN 卷积核的数量

选择 20 作为参考签名的数量，将 CNN 卷积核数目设置为 8、16、32、64、128 和 256，在不同的卷积核数量情况下测试系统性能。图 4–11显示 了在第三种情况

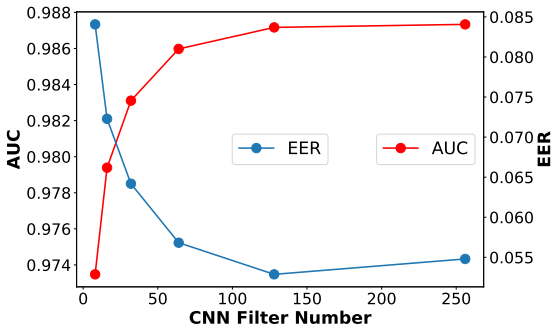


图 4–11 CNN 卷积核数量的影响

Figure 4–11 CNN filter number effect

下 (*all forgers*) 的 AUC 和 EER 的结果：当卷积核数量从 8 增加到 128 时，系统性能变得更好；当卷积核数量为 128 时，系统性能达到最佳：AUC=0.987 和 EER=0.053；即使当卷积核数量为最小的 8 时，结果仍然可以达到 AUC=0.973 和 EER=0.084，显示出了特征提取的成功。

4.6.3　训练集大小

选择 20 作为参考签名的数量，128 作为卷积核数量。为了评估训练集大小对试验结果的影响，保持测试集的大小不变，将训练集的大小改变成 100，300，...，1700。如图 4–12所示，当训练集大小从 100 增加到 1700 时，AUC 从 0.814 增加到 0.980，EER 从 0.263 增加到 0.071。就如预期的那样，当训练集大小增加到一定程度后，AUC 和 EER 的曲线开始收敛。

4.6.4　DCT 系数数量

选择 20 作为参考签名的数量，为了评估 DCT 系数数量对实验结果的影响，将 DCT 系数数量设置为 8、9、10、15、20、25、30、35 和 40。对前面的几个 DCT 系数数量设置间隔较小，实现较为精细的评估。在三种情况下测试 DCT 系数的数量

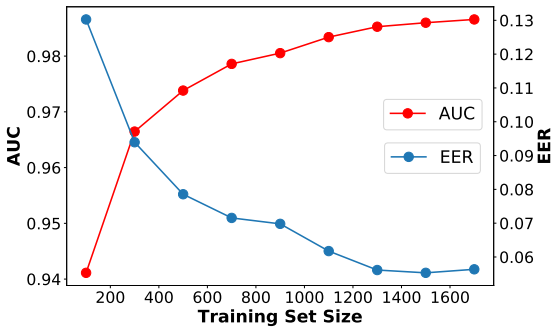


图 4–12 训练集大小的影响

Figure 4–12 Train set size effect

的影响：第一种情况下，只有相位变化的速度信息被使用；第二种情况下，只有相位变化的加速度信息被使用；第三种情况下，相位变化的速度信息和加速度信息都被使用。假设 C 表示被设置的 DCT 系数数量，所以在前两种情况下，距离矩阵的形状为 $(C, 8, 3)$，而在第三种情况中，距离矩阵的形状为 $(C, 8, 6)$。图 4–13和图 4–14显示了评估结果， 当 DCT 系数数量是从 8 到 40 的时候，只使用相位变化

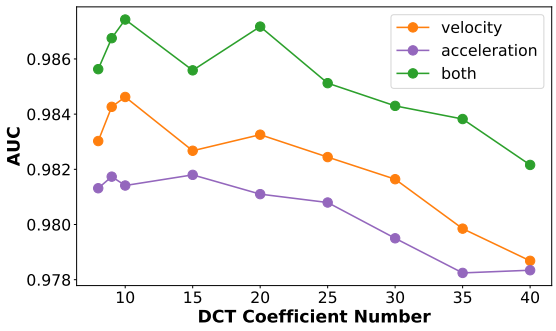


图 4–13 DCT 系数数量 - AUC

Figure 4–13 DCT coefficient number - AUC

的速度信息的系统性能优于只使用加速度信息的系统，总体上相位变化的速度信息和加速度信息的结合可以提升系统性能。接着，将注意力转移到第三种情况中，如图 4–13所示，当时 DCT 系数数量为 10 时，系统的性能达到最佳，当 DCT 系数继续增加的时候，由于噪音的增加，系统性能开始变差。

4.6.5　声波频率数量

选择 20 作为参考签名的数量。为了评估频率数量对试验结果的影响，将频率数量设置为 $FN = 3, 4, ..., 8$。所以，距离矩阵的形状为 $(10, FN, 6)$。由于 *max pooling*

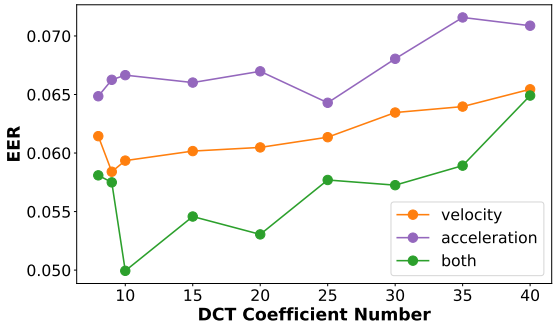


图 4–14 DCT 系数数量 - EER

Figure 4–14 DCT coefficient number - EER

的降维效果，当频率数量小于 8 的时候，两层卷积池化层将无法工作。因此，为了能够顺利训练和测试模型，在原有的 CNN 模型基础上减去了一层卷积池化层，使用修改后的模型。图 4–15报告 了结果。当频率从 3 增加到 8 的时候，AUC 从

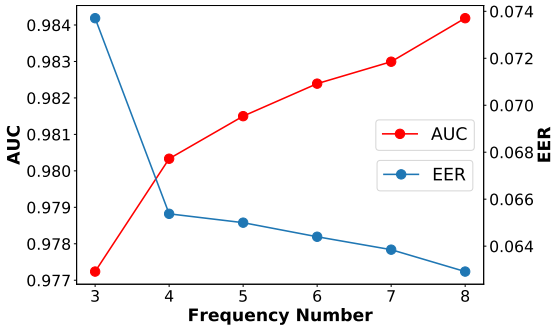


图 4–15 频率数量的影响

Figure 4–15 Frequency number effect

0.977 增加到 0.984，EER 从 0.074 减少到 0.063。所以，多频率的使用可以使得 ASSV 更加鲁棒和准确。

4.6.6　分类器类型

选择 20 作为参考签名的数量，为了比较本研究的 CNN 模型和传统分类器的效果，将 CCN 模型替换为传统的分类器，比如随机森林、支持向量机或者朴素贝叶斯作为二分类器。如表 4–4和表 4–5所示， 随机森林模型也呈现了较好的结果 (第三种情况中，AUC = 93.6 和 EER=7.3)，但是本研究的 CNN 达到最佳性能 (第三种情况中，AUC=98.4 和 EER=5.8)。

表 4–4 分类器 - AUC

Table 4–4 Classifier - AUC

Classifier	Random(%)	Skilled(%)	All(%)
Naive Bayes	86.1	83.5	84.8
SVM	92.5	96.3	94.5
Random Forest	93.8	93.5	93.6
CNN	**98.3**	**98.5**	**98.4**

表 4–5 分类器 - EER

Table 4–5 Classifier - EER

Classifier	Random(%)	Skilled(%)	All(%)
Naive Bayes	14.6	18.2	16.3
SVM	16.7	9.0	12.3
Random Forest	7.2	7.5	7.3
CNN	**6.1**	**5.8**	**5.8**

4.7 经典系统对比实验

选择一个离线手写签名认证系统[16] 来和 ASSV 进行比较。为了获得两个数据集，签名图像 (尺寸: 483×312 *px*) 被扫描和切割。根据[16]，使用第一个数据集作为 *development dataset*, 用于训练一个写者独立的人工神经网络，这个人工神经网络将被用于特征提取，第二个数据集将被用作 *exploitation dataset*，每个用户的 10 个真实签名将被用于训练一个写者依赖的分类器。下载到源代码 [1]用于评估，如表 4–6所示，尽管[16] 达到了 96.5% 的 AUC 和 10.6% 的 EER，使用了二次训练的模型的 ASSV 还是达到了最优性能。

表 4–6 和经典系统的比较

Table 4–6 Comparison with the state-of-the-art

Model	AUC(%)	EER(%)
The state-of-the-art	96.5	10.6
Cross-user	90.5	17.1
Retrained	**98.2**	**7.8**

[1]https://github.com/luizgh/sigver

4.8 重放攻击实验

在重放攻击实验中，2 名自愿者被邀请参加写 20 个参考签名，并且在写的时候，放置了另外一个用于监听的智能手机在目标智能手机的旁边，在参考签名得到处理后，监听智能手机会对每个志愿者实施 20 次重放攻击。本文采用与 §4.5中相似的度量指标，40 个预测值的中位数是 4.1，最大值在 50 到 60 之间。所以，如果将阈值设置为 50, 只有一次重放攻击能够成功；如果将阈值设置为 60，那么所有重放攻击都是失败的。结果显示本文系统可以很好地处理这样的重放攻击。其中一个原因是，系统中所提取的特征与智能手机和手的相对位置具有很大的关联，除非监听器和目标智能手机在同一个位置，否则监听数据所得到的特征与真实值将有很大区别。另外，在攻击的时候，有两个智能手机在发射声波，所以接收的声波信号由两部分组成，这也会导致重放攻击的失败。

4.9 本章小结

本章对基于声波的手写签名认证方案进行了实验验证和评估。在线签名认证和离线签名认证已经被研究很长时间了，在这个领域有很多先进的方法被提出，它们大多数在已有的数据集上被评估和测试。本研究是第一次将声波信号应用到手写签名认证，之前的研究使用了多模信号，实现了较高精度，本研究仅仅使用声波信号便实现了较高精度。根据评估结果，ASSV 已经达到满意的结果，并且只需要使用很少的参考签名。除了准确度外，还评估系统的交叉用户可用性、鲁棒性、重放攻击，实验显示出 ASSV 系统良好的可用性。

第五章　基于声波的签名认证方案的系统设计与实现

本章针首先分析基于声波的签名认证系统的用户需求，然后完成系统架构设计，接着实现基于声波的签名认证系统，最后进行一系列实验来评估系统的运行性能。

5.1　系统需求分析

本节将设计并实现基于声波的签名认证的原型系统——ASSV。本原型系统的主要目的是让用户可以在自己的手机上实时判别自己签名是否真实，及时给出反馈，主要针对的场景是需要使用手写签名进行身份认证的场景。系统的主要需求是：用户手写自己签名用于认证，系统给出签名真实性判别结果。通过进一步分析，将面向用户的主要功能展示于图 5–1的系统用例图中。

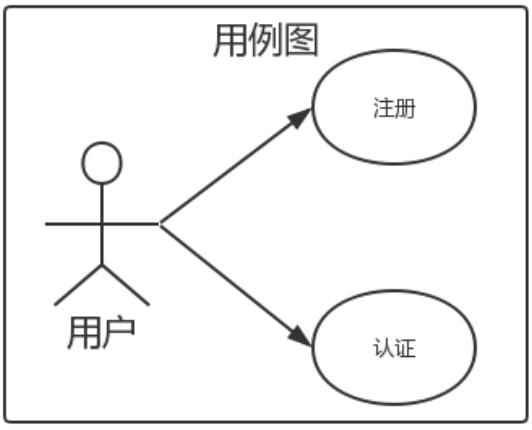


图 5–1 系统用例图

Figure 5–1 The use case diagram

面向用户的主要功能如下：

（**1**）**注册**。用户使用系统之前，需要写一定数量的参考签名完成注册，完成注册之后，才能使用认证功能。

（**2**）**认证**。用户在认证的时候，开始记录签名动作，完成签名，系统给出反馈，告诉用户这个签名的真实性。

5.2　系统设计与实现

5.2.1　系统框架设计

根据对基于声波的签名认证系统的需求分析，为保证原型系统具有良好的可扩展性和可维护性，本研究对原型系统采用分层架构。框架设计如图 5–2, 系统从

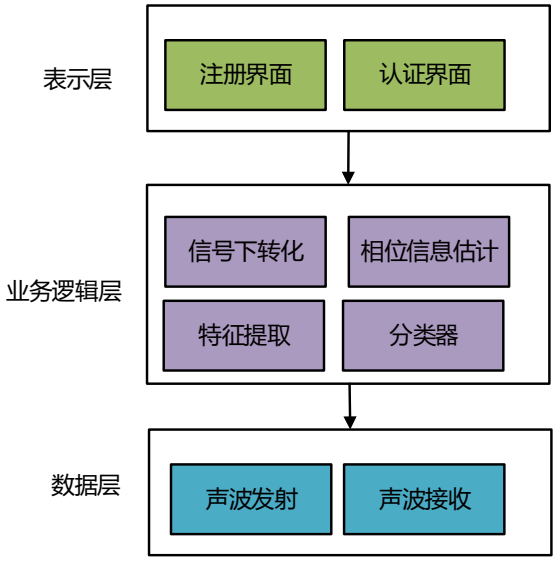


图 5–2 系统架构设计

Figure 5–2 The design of system framework

下到上分为三层：最下层是数据层，主要功能是负责扬声器和麦克风相关系统接口的调用和协调，播放预先生成的音频文件发射声波，接收声波数据，上传给上层；中间层是业务逻辑层，主要负责接收到原始声波数据之后的相关处理，包括：信号下转化、相位信息估计、特征提取和分类；最上层是表示层，主要是提供一个和用户的交互 UI，包括应用程序的注册界面和认证界面，用户从界面发出命令给系统，相应结果也由界面展示。整个系统分为两部分：服务端和安卓端，服务端用 Python 实现，运行 CNN 预测模型，其余部分都在安卓端实现，通过 socket 实现安卓端和服务端的通信。

5.2.2　数据层

数据层主要负责声波的发射和接收，由于需要与外设扬声器和麦克风通信，这部分需要调用相关的安卓系统接口。安卓版本为 android 7，apk 文件的最小 sdk 版本是 21，目标和编译 sdk 版本都是 27。

在发射声波部分，需要调用扬声器相关的 API，android.media.MediaPlayer 类来实现对预先生成音频文件的播放来发射声波。如代码 5–1所示，

代码 5–1 创建 MediaPlayer 对象

```
mediaPlayer = MediaPlayer.create(this, R.raw.sound)    //创建
    MediaPlayer 对象
```

需要创建一个 MediaPlayer 对象，第一个参数是上下文 Context 对象，第二个参数是音频文件的资源编号。代码 5–2展示

代码 5–2 开始播放

```
mediaPlayer.seekTo(0) //到开始位置
mediaPlayer.start()  //开始播放
```

了开始播放的执行过程，代码 5–3展示

代码 5–3 暂停播放

```
mediaPlayer.pause()  //暂停播放
```

了暂停播放的执行过程。在操作过程中，满足 MediaPlayer 的生命周期 [1]，防止出错。

在接收声波部分，需要调用麦克风相关的 API。使用 android.media.AudioRecord 类来实现声波的接收。如代码 5–4所示，

代码 5–4 创建 AudioRecord 对象

```
audioRecord = AudioRecord(
ConfigInfo.audioSource,    //MediaRecorder.AudioSource.MIC
ConfigInfo.sampleRateInHz, //48000
ConfigInfo.channelConfig,  //AudioFormat.CHANNEL_IN_MONO
ConfigInfo.audioFormat,   //AudioFormat.ENCODING_PCM_FLOAT
ConfigInfo.bufferSize    // 48000*2*2
)
```

需要创建 AudioRecord 对象，它的 startRecord 和 stop 方法用于开始和停止声波数据的记录。麦克风产生的是流式数据，需要通过循环读来获取数据，如代码 5–5所示，将读到的数据写到一个输出流中。

[1] https://developer.android.com/reference/android/media/MediaPlayer

代码 5–5 读取数据流

```
fun writeData(outputStream: OutputStream) {
    val buffer: ByteBuffer = ByteBuffer.allocateDirect(
       ConfigInfo.bufferSize!! / 3) //创建 Direct 内存缓存
       区
    val byteArray: ByteArray = ByteArray(ConfigInfo.bufferSize!!
        / 3)   //创建缓存数
       组
    var len: Int = 0
    try {
        while (audioRecord!!.recordingState != AudioRecord.
       RECORDSTATE_RECORDING) {
        }
        do {
            len = audioRecord!!.read(buffer, buffer.capacity())
         //循环读取数据
       流
            buffer.rewind()
            if (len > 0) {
                buffer.get(byteArray, 0, len)
                buffer.clear()
                outputStream.write(byteArray, 0, len)  //写入一个
       输出流中
            }
        } while (len > 0 || audioRecord!!.recordingState ==
       AudioRecord.RECORDSTATE_RECORDING)
    } finally {
        outputStream.flush()
        outputStream.close()
    }
}
```

5.2.3　业务逻辑层

业务逻辑层拥有一条与数据层中的输出流通过管道相连的输入流，可以处理原始的声波数据，处理步骤包括：信号下转化、相位信息估计、特征提取和分类。当用户执行注册操作的时候，不需要分类步骤。安卓端业务逻辑部分使用 Java 实现，打包成 Jar 包，引入到安卓项目中；服务端的分类器使用 Python 实现；两者间，将数据使用 Json 序列化后，利用 Socket 进行 TCP 通信。

信号下转化，采用在线处理的方式，每获得一个数据点，就针对这个点进行信号下转化。低通滤波器采用在线 Butterworth 滤波器，可以实现一个一个数据点处理。运行过程如代码 5–6所示，

代码 5–6 实时信号下转化

```
public boolean offer(double value) {
      if (this.size == BUFFER_SIZE) {
          return false;
      }
      //丢弃前面几个
      if (abandonCount < ABANDON_FORMER_COUNT) {
          ++abandonCount;
          return true;
      }
      //处理
      boolean add2Buffer = false;
      for (int i = 0; i < FREQUENCY_COUNT; ++i) {
          double IValue = getI(value, i, size);
          IValue = IFilters[i].filter(IValue);
          double QValue = getQ(value, i, size);
          QValue = QFilters[i].filter(QValue);
          if (size % 300 == 0) {
              buffer[i][0][bufferSize] = IValue;
              buffer[i][1][bufferSize] = QValue;
              add2Buffer = true;
          }
      }
      if (add2Buffer) {
          ++bufferSize;
      }
      ++size;
      return true;
    }
```

其中参数 value 为声波数据点，getI 与 getQ 函数中进行三角函数的相乘操作，QFilter 和 IFilter 为滤波器，buffer 为缓存区。

相位信息估计，利用保存在缓存区 buffer 里的同相分量和正交分量的数据，来估计相位相关信息。

特征提取，DCT 转化和使用前 10 个系数，返回特征矩阵，如代码 5–7所示。

代码 5–7 特征提取

```
private double[][] extractFeature(List<double[]>
       phaseRelatedInfos) {
    double[][] result = new double[16][10];
    for (int i = 0; i < phaseRelatedInfos.size(); ++i) {
        double[] data = phaseRelatedInfos.get(i);
        //规范化到 0-1
        MinMaxScaler minMaxScaler = new MinMaxScaler();
        minMaxScaler.fit(data);
        data = minMaxScaler.transform(data);
        //提取频域特征
        double[] dctCoefficients = DCT.dct(data);
        int selectedCoefficientNum = 10;
        selectedCoefficientNum = Math.min(selectedCoefficientNum
       , dctCoefficients.length);
        for (int j = 0; j < selectedCoefficientNum; ++j) {
            result[i][j] = dctCoefficients[j];
        }
    }
    return result;
}
```

分类器，由 Python 实现，借助 Keras 框架进行模型的构建，安卓端远程调用模型进行判别。

5.2.4 表示层

表示层提供用户与系统的交互接口。常用的用户操作主要为开始签名和结束签名，无论是在注册阶段还是验证阶段都需要用到这两个操作。当用户第一次使用时，需要进行注册，录入参考签名，本系统设置参考签名的数量为 10 个，用户输入 10 个参考签名后，系统提示注册完成；用户认证时，从业务逻辑层可以获得一个范围为 0-1 的预测值，需要设置一个阈值来判断当前签名是否为真实签名，系统中该阈值设置为 0.6。如图 5–3所示，左边为真实签名，右边为仿造签名，点击右下角的红色按钮可进入注册阶段，点击中间播放按钮可开始签名。

5.3 系统运行效率评估

为了成为一个实时系统,ASSV 的内存使用和延迟必须在一定范围内,对 ASSV 的运行效率进行了评估，包括 CPU 使用、内存使用和系统延迟。让 app 持续发射

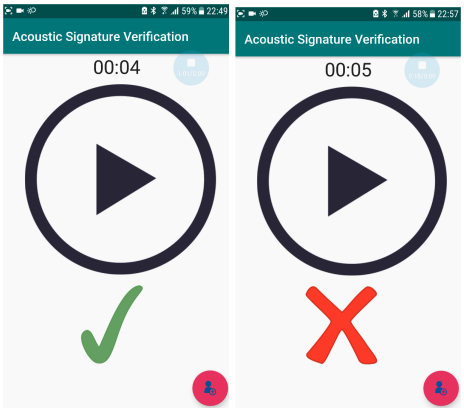


图 5–3 验证结果 - 左边为真实签名，右边为仿造签名

Figure 5–3 Verification result - The left message indicates a genuine signature while the right message indicates a forged signature.

声波 24 秒，同时采集智能手机性能数据，最大的内存和 CPU 使用分别是 132.7 MB 和 12.8%，作为分类器的服务端消耗内存 550.8 MB，占用 7% 的 CPU 时间。测量用户在结束签名之后的等待时间，即系统延迟；当 app 持续发射声波 5 秒结束后，测得平均延迟为 0.937 秒；当 app 持续发射声波 10 秒结束后，测得平均延迟为 3.188 秒。测延迟的两种情况中，每个过程重复了 10 次以得到平均值，相对于签名的时间，这个延迟已经相当小。

5.4　本章小结

实现了安卓平台上的基于声波签名认证系统，该原型系统主要提供给用户身份识别功能。该系统采用分层架构，对其中的数据层、业务逻辑层和表示层的具体实现进行了详细介绍。最后从 CPU 使用、内存使用和系统延迟三个方面对系统运行效率进行了评估。

第六章 总结与展望

6.1 工作总结

自动手写签名认证是通过现代计算机技术将手写签名真实性判别这个过程自动化，减少人力开销和人为错误。手写签名从产生至今已经有很长的历史，被广泛应用于政府部门、银行业等领域，然而这种认证方式易于遭受仿造，减少手写签名认证错误的研究很有意义。按照数据来源划分，自动手写签名认证方式可以被分为两种：离线手写签名认证和在线手写签名认证。在线签名认证由于相比离线签名认证多出了时间维度上的信息，具有更高的准确度。先前的工作，如果要使用那些自动手写签名认证系统，需要用户使用特定的设备，或者佩戴可穿戴式设备。将无线感知技术运用到手写签名认证中，利用智能手机发射和接收声波信号感知签名者的签名行为，通过与参考签名比较来判断是否为本人签名，设计了一种非侵入式、鲁棒的、安全的、低延迟的在线手写签名认证方案，并按照该方案实现了一个原型系统——ASSV。本文取得的成果如下：

(1) 基于声波信号的感知方法

声波的发射和接收装置比较容易获得，例如普通智能手机上的扬声器和麦克风便能发射和接收声波，因此近年来声波感知研究工作大量涌现。设置声波信号的频率和传播速度分别是 17000 Hz 和 346 m/s，那么波长为 $346/17000 \approx 0.02m$，这么小的波长意味着声波的相位信息对周围物体的运动会很敏感。本文使用智能手机发射和接收声波，使用声波跟踪签名者签名时的行为特征。设计并生成了多频率的音频数据，以供扬声器发射。针对当前场景中动作微小的这个特征，使用“弦”的方法提取到声波相位相关信息，并成功避免了去直流问题。这种方法，相比其他基于定制设备或者可穿戴式设备，充分利用了无线感知技术的优势，更具普适性，并且有助于提升用户体验。

(2) 基于 DCT 和 CNN 的建模方法

手写签名动作所产生的影响在相位信号中体现为低频信号，而高频信号则是由环境或者硬件造成的噪声所产生，因此可以通过将时域数据转换到频域数据，然后提取低频系数的方法，将手写签名动作的信息提取出来。DFT 是一种常见的获取频域数据的方法，但它的系数是 2 倍冗余的复数系数，所以采用 DCT 作为到频域数据的转换方法，并提取其前几个低频系数作为特征。本文借助于 DCT 将时域信息转化为频域信息，去除高频噪声，保留低频系数，从而实现特征选择和降维，

结合多频获得特征矩阵。利用参考签名的特征矩阵，获得参考矩阵。查询签名的特征矩阵和参考矩阵作差，可以获得距离矩阵，这个矩阵可以衡量查询签名和参考签名之间的相似度。得到距离矩阵之后，对距离矩阵进行而分类，分为真实签名和仿造签名，设计了 CNN 模型作为二分类器，并使用多种模型优化手段。

(3) 方案的实验验证

本文针对基于声波的签名认证方案进行实验验证以及分析。首先介绍在一定实验配置下的数据采集过程，然后展示众多评估实验及其相应的结果，这些评估实验包括：验证精度、交叉用户可用性、硬件补偿的效果、系统鲁棒性、微基准实验、对比实验和重放攻击。系统准确度性能达到：AUC=98.7% 和 EER=5.5%。在鲁棒性测试中，改变环境、距离、日期等来对系统进行测试，观察系统准确性的变化；在安全性测试中，通过模拟重放攻击来测试系统对此类攻击的抵抗能力；在微基准测试中，改变一些超参数，来观察系统的变化，以便设置较合适的超参数；在性能测试中，通过度量系统时延，来判断系统是否具有良好的用户体验和是否对硬件要求过高。

(4) 手写签名认证的原型系统

利用一个三星手机和普通的个人电脑，设计并实现了基于声波的签名认证系统。首先分析系统的用户需求，接着完成系统架构的设计，最后进行一系列评估实验来评估系统的运行效率性能效果。

6.2 研究展望

本文提出的基于声波的签名认证方案具有比较好的准确度和低延迟，有很强的普适性。但仍然存在一些不足之处：在系统的测试和评估中，声音音量都固定设置为 10。不同智能手机的扬声器和麦克风的硬件特性可能有所不同，而扬声器和麦克风在智能手机上的位置也会随着手机自身架构的原因有所不同。因此，如果是其他手机 (不是三星 Galaxy S6)，可能需要进行一些适配。

基于现有的原型系统 ASSV，可以进一步探索如何提升签名认证系统的性能，有两个可行方向可以考虑：一是设计并使用更多特征提取方法，如深度学习，可以被探索；二是将离线签名认证系统 (善于辨认随机仿造签名) 与本系统相结合，实现多模系统。

附录 A 算法

算法 A–1 获得特征矩阵

procedure GET_FEATURE_MATRIX(*data*) ▷ data 是接收到的离散声波信号
    $BaseFrequency \leftarrow 17350, FreqNum \leftarrow 8, dctNum \leftarrow 10$ ▷ 初始化起始频率、频率数目和 DCT 系数数目
    $VelocityFeatureVectors \leftarrow []$ ▷ 初始化速度特征向量列表
    $AcceleratinoFeatureVectors \leftarrow []$ ▷ 初始化加速度特征向量列表
    $freqIndex \leftarrow 0$
    while $freqIndex < FreqNum$ **do** ▷ 遍历数据点
        $freq \leftarrow BaseFrequency + freqIndex \times 700$
        $VelocityFeatureVector, AccelerationFeatureVector \leftarrow GET\_FEATURE\_VECTOR(data, freq, dctNum)$
        add $VelocityFeatureVector$ to list $VelocityFeatureVectors$ ▷ 添加速度特征向量到特征向量列表中
        add $AccelerationFeatureVector$ to list $AcceleratinoFeatureVectors$ ▷ 添加加速度特征向量到特征向量列表中
        $freqIndex \leftarrow freqIndex + 1$
    end while
    $VelocityFeatureMatrix \leftarrow$ construct_feature_matrix_from $VelocityFeatureVectors$ ▷ 形状为 (8,10)
    $AccelerationFeatureMatrix \leftarrow$ construct_feature_matrix_from $AccelerationFeatureVectors$ ▷ 形状为 (8,10)
    return $VelocityFeatureMatrix, AccelerationFeatureMatrix,$ ▷ 返回结果
end procedure

算法 A–2 获得特征向量

procedure GET_FEATURE_VECTOR($data$, f, $dctNum$) ▷ data 是接收到的离散声波信号,f 是声波频率,dctNum 为 dct 索取系数数量
  $FreqRate \leftarrow 48000$ ▷ 采样率为 48 kHz
  $I \leftarrow [], Q \leftarrow []$ ▷ 初始化 I-component 和 Q-component
  $dataLen \leftarrow$ getLengthof $data$ ▷ 获得 data 的数据点数量
  $i \leftarrow 0$
  while $i < dataLen$ **do** ▷ 遍历数据点
    $I[i] = data[i] * cos(\frac{2\pi fi}{FreqRate})$
    $Q[i] = data[i] * (-sin(\frac{2\pi fi}{FreqRate}))$
    $i \leftarrow i + 1$
  end while
  $I \leftarrow$ DownSample $\leftarrow$ LowPassFilter I ▷ 获得 I-component
  $Q \leftarrow$ DownSample $\leftarrow$ LowPassFilter Q ▷ 获得 Q-component
  $I\_seasonal, I\_trend, I\_resid \leftarrow$ seasonal decompose I ▷ 趋势季节分解
  $Q\_seasonal, Q\_trend, Q\_resid \leftarrow$ seasonal decompose Q
  $j \leftarrow 1$
  $\Delta Chord \leftarrow []$
  while $j < getLengthOf(I\_trend)$ **do** ▷ 遍历数据点
    $\Delta Chord[j-1] \leftarrow \sqrt{(I[j] - I[j-1])^2 + (Q[j] - Q[j-1])^2}$
    $j \leftarrow j + 1$
  end while
  $k \leftarrow 1$
  $Acceleration \leftarrow []$
  while $k < getLengthOf(\Delta Chord) - 1$ **do** ▷ 遍历数据点
    $Acceleration[k-1] \leftarrow \frac{(\Delta Chord[k+1] - \Delta Chord[k]) + (\Delta Chord[k+1] - \Delta Chord[k])/2}{2}$
    $k \leftarrow k + 1$
  end while
  $VelocityFeatureVector \leftarrow dct(\Delta Chord)[: dctNum]$
  $AccelerationFeatureVector \leftarrow dct(Acceleration)[: dctNum]$
  return $VelocityFeatureVector, AccelerationFeatureVector$ ▷ 返回结果
end procedure

参考文献

[1] HUANG X, YANG X, CHONKA A, et al. A Generic Framework for Three-Factor Authentication: Preserving Security and Privacy in Distributed Systems[J]. IEEE Transactions on Parallel & Distributed Systems, 2011, 22(8): 1390-1397.

[2] ZHAO W, CHELLAPA R, PHILLIPS P, et al. Face recognition: A literature survey[J]. ACM Comput. Surv., 2003, 35: 399-458. DOI: 10.1145/954339.954342.

[3] ANDREW A M. Handbook of Fingerprint Recognition[J]. Ch Synthetic Fingerprint Generation, 2005, 33(5-6): 1314.

[4] WILDES R P. Iris recognition: An emerging biometrie technology[J]. Proceedings of the IEEE, 1997, 85(9): 1348-1363.

[5] CETINGUL H E, YEMEZ Y, ERZIN E, et al. Discriminative Analysis of Lip Motion Features for Speaker Identification and Speech-Reading[J]. IEEE Transactions on Image Processing, 2006, 15(10): 2879-2891.

[6] PLAMONDONA R, LORETTEB G. Automatic signature verification and writer identification —the state of the art[J]. Pattern Recognition, 1989, 22(2): 107-131.

[7] RASHID R A, MAHALIN N H, SARIJARI M A, et al. Security system using biometric technology: Design and implementation of Voice Recognition System (VRS)[C]//International Conference on Computer & Communication Engineering. 2008.

[8] BOULGOURIS N V, HATZINAKOS D, PLATANIOTIS K N. Gait recognition: a challenging signal processing technology for biometric identification[J]. Signal Processing Magazine IEEE, 2005, 22(6): 78-90.

[9] 鄢晨丹, 杨阳, 程久军, 等. 基于统计模型的 DTW 签名认证系统[J]. 信息网络安全, (7): 70-76.

[10] LEVY A, NASSI B, ELOVICI Y, et al. Handwritten Signature Verification Using Wrist-Worn Devices[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(3): 119.

[11] GRISWOLD-STEINER I, MATOVU R, SERWADDA A. Wearables-Driven Freeform Handwriting Authentication[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2019.

[12] BUNKE H, CSIRIK J, GINGL Z, et al. Online signature verification method based on the acceleration signals of handwriting samples[C]//Iberoamerican Congress on Pattern Recognition. 2011: 499-506.

[13] FISCHER A, DIAZ M, PLAMONDON R, et al. Robust score normalization for DTW-based on-line signature verification[C]//2015 13th international conference on document analysis and recognition (ICDAR). 2015: 241-245.

[14] KHOLMATOV A, YANIKOGLU B. Identity authentication using improved online signature verification method[J]. Pattern recognition letters, 2005, 26(15): 2400-2408.

[15] SAE-BAE N, MEMON N. A simple and effective method for online signature verification[C]//2013 International Conference of the BIOSIG Special Interest Group (BIOSIG). 2013: 1-12.

[16] HAFEMANN L G, SABOURIN R, OLIVEIRA L S. Learning features for offline handwritten signature verification using deep convolutional neural networks[J]. Pattern Recognition, 2017, 70: 163-176.

[17] HAFEMANN L G, OLIVEIRA L S, SABOURIN R. Fixed-sized representation learning from offline handwritten signatures of different sizes[J]. International Journal on Document Analysis and Recognition (IJDAR), 2018, 21: 219-232.

[18] FERRER M A, ALONSO J B, TRAVIESO C M. Offline geometric parameters for automatic signature verification using fixed-point arithmetic[J]. IEEE transactions on pattern analysis and machine intelligence, 2005, 27(6): 993-997.

[19] KALERA M K, SRIHARI S, XU A. Offline signature verification and identification using distance statistics[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2004, 18(07): 1339-1360.

[20] 赵桂敏, 夏利民, 陈爱斌. 手写签名的快速认证[J]. 计算机工程, 2003, 29(7): 56-58.

[21] CHANDRA S, MAHESKAR S. Offline signature verification based on geometric feature extraction using artificial neural network[C]//2016 3rd International Conference on Recent Advances in Information Technology (RAIT). 2016: 410-414.

[22] BHATTACHARYA I, GHOSH P, BISWAS S. Offline signature verification using pixel matching technique[J]. Procedia Technology, 2013, 10: 970-977.

[23] 李成华, 龚良慧, 江小平, 等. 基于 EMD 和 SVD 的在线手写签名特征提取方法[J]. 中南民族大学学报（自然科学版）, v.35;No.118(01): 107-111+117.

[24] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[25] RABINER L R, JUANG B H. An introduction to hidden Markov models[J]. Ieee assp magazine, 1986, 3(1): 4-16.

[26] 雷涛. 隐马尔可夫模型下视频手写签名认证算法研究[J]. 计算机测量与控制, (7).

[27] 李成华, 刘磊, 龚良慧, 等. 基于 DCT 和 SVDD 的在线手写签名认证方法[J]. 计算机系统应用, (7): 196-199, 共 4 页.

[28] 周志华. 机器学习[M]. 中国: Qing hua da xue chu ban she, 2016.

[29] SCHMIDHUBER J. Deep learning in neural networks: an overview.[J]. Neural Netw, 2015, 61: 85-117.

[30] HANMANDLU M, YUSOF M H M, MADASU V K. Off-line signature verification and forgery detection using fuzzy modeling[J]. Pattern Recognition, 2005, 38(3): 341-356.

[31] FREITAS C, OLIVEIRA L S, SABOURIN R, et al. Brazilian forensic letter database[C]//11th International workshop on frontiers on handwriting recognition, Montreal, Canada. 2008.

[32] LEVY A, NASSI B, ELOVICI Y, et al. Handwritten Signature Verification Using Wrist-Worn Devices[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(3): 1-26.

[33] GRISWOLD-STEINER I, MATOVU R, SERWADDA A. Wearables-Driven Freeform Handwriting Authentication[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2019: 1-1. DOI: 10.1109/TBIOM.2019.2912401.

[34] BUNKE H, CSIRIK J, GINGL Z, et al. Online Signature Verification Method Based on the Acceleration Signals of Handwriting Samples[J]., 2015.

[35] 房育勋. 在线签名认证若干关键问题研究[D]. 华南理工大学.

[36] KHOLMATOV A, YANIKOGLU B. SigSA: On-line Handwritten Signature Database[Z]. 2006.

[37] YEUNG D Y, CHANG H, XIONG Y, et al. SVC2004: First International Signature Verification Competition[C]//ZHANG D, JAIN A K. Biometric Authentication. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004: 16-22.

[38] LU X, FANG Y, KANG W, et al. SCUT-MMSIG: A Multimodal Online Signature Database[C]//ZHOU J, WANG Y, SUN Z, et al. Biometric Recognition. Cham: Springer International Publishing, 2017: 729-738.

[39] PLAMONDON R, LORETTE G. Automatic signature verification and writer identification—the state of the art[J]. Pattern recognition, 1989, 22(2): 107-131.

[40] LECLERC F, PLAMONDON R. Automatic signature verification: The state of the art—1989–1993[G]//Progress in Automatic Signature Verification. World Scientific, 1994: 3-20.

[41] IMPEDOVO D, PIRLO G. Automatic signature verification: The state of the art[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2008, 38(5): 609-635.

[42] IMPEDOVO D, PIRLO G, PLAMONDON R. Handwritten signature verification: New advancements and open issues[C]//2012 International Conference on Frontiers in Handwriting Recognition. 2012: 367-372.

[43] HAFEMANN L G, SABOURIN R, OLIVEIRA L S. Offline handwritten signature verification—literature review[C]//2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). 2017: 1-8.

[44] PRATHIBA M, BASAVARAJ L. Online handwritten signature verification system: A Review[J]. International Journal of Emerging Trends & Technology in Computer Science, 2014, 3(2): 263-267.

[45] HUANG K, YAN H. Off-line signature verification based on geometric feature extraction and neural network classification[J]. Pattern Recognition, 1997, 30(1): 9-17.

[46] RIVARD D, GRANGER E, SABOURIN R. Multi-feature extraction and selection in writer-independent off-line signature verification[J]. International Journal on Document Analysis and Recognition (IJDAR), 2013, 16(1): 83-103.

[47] ESKANDER G S, SABOURIN R, GRANGER E. Hybrid writer-independent–writer-dependent offline signature verification system[J]. IET biometrics, 2013, 2(4): 169-181.

[48] YILMAZ M B, YANIKOLU B. Score level fusion of classifiers in off-line signature verification[J]. Information Fusion, 2016, 32: 109-119.

[49] RANTZSCH H, YANG H, MEINEL C. Signature embedding: Writer independent offline signature verification with deep metric learning[C]//International symposium on visual computing. 2016: 616-625.

[50] HAFEMANN L G, SABOURIN R, OLIVEIRA L S. Writer-independent feature learning for offline signature verification using deep convolutional neural networks[C]//2016 International Joint Conference on Neural Networks (IJCNN). 2016: 2576-2583.

[51] DARAMOLA S, IBIYEMI T. Efficient on-line signature verification system[J]. International Journal of Engineering & Technology IJET-IJENS, 2010, 10(4): 48-52.

[52] SHAFIEI M M, RABIEE H R. A new online signature verification algorithm using variable length segmentation and hidden Markov models[C]//Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. 2003: 443-446.

[53] AHMED S K, RAMASAMY A K, KHAIRUDDIN A S M, et al. Automatic online signature verification: A prototype using neural networks[C]//TENCON 2009-2009 IEEE Region 10 Conference. 2009: 1-4.

[54] TSE D, VISWANATH P. Fundamentals of wireless communication[M]. Cambridge university press, 2005.

[55] WANG W, LIU A X, SUN K. Device-free gesture tracking using acoustic signals[C]//Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking. 2016: 82-94.

[56] AHMED N, NATARAJAN T, RAO K R. Discrete cosine transform[J]. IEEE transactions on Computers, 1974, 100(1): 90-93.

[57] WINOGRAD S. On computing the discrete Fourier transform[J]. Mathematics of computation, 1978, 32(141): 175-199.

[58] LEE B. A new algorithm to compute the discrete cosine transform[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(6): 1243-1245.

[59] HOU H. A fast recursive algorithm for computing the discrete cosine transform[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1987, 35(10): 1455-1461.

[60] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. MIT press, 2016.

[61] RODRIGUEZ VALIENTE A, TRINIDAD A, GARCIA BERROCAL J, et al. Extended high-frequency (9–20 kHz) audiometry reference thresholds in 645 healthy subjects[J]. International journal of audiology, 2014, 53(8): 531-545.

[62] CLEVELAND R B, CLEVELAND W S, MCRAE J E, et al. STL: A seasonal-trend decomposition[J]. Journal of official statistics, 1990, 6(1): 3-73.

[63] THORNBURG H. Real-Time Decomposition of Time Series[EB/OL]. 2016. https://developer.ibm.com/streamsdev/docs/real-time-decomposition-of-time-series/.

[64] 安霄霄. 手写签名的多重数字水印及认证算法研究[D]. 哈尔滨工业大学.

[65] CHUN-LIN L. A tutorial of the wavelet transform[J]. NTUEE, Taiwan, 2010.

[66] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. ArXiv preprint arXiv:1502.03167, 2015.

[67] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

攻读学位期间发表的学术论文

[1] First Author. ASSV: handwritten signature verification using acoustic signals[J]//Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2019, 3(3): 80.

[2] First Author. TTBA: An RFID-based Tracking System for Two Basic Actions in Free-Weight Exercises[C]//Proceedings of the 14th ACM International Symposium on QoS and Security for Wireless and Mobile Networks. ACM, 2018: 7-14.

攻读学位期间参与的项目

[1] 上海市科学技术委员会平台建设项目

[2] 上海市经济和信息化委员会、上海市人力资源和社会保障局上海市高技能人才培养基地资助项目