生物特征数据安全保护技术的发展*)

周玲丽[1] 赖剑煌[2]
(中山大学数学与计算科学学院 广州 510275)[1] (中山大学信息科学与技术学院 广州 510275)[2]

摘 要 生物特征识别相对于传统的身份识别更安全和便捷。随着生物特征识别系统的广泛应用,生物特征数据的安全性和隐私性日益得到重视。生物特征数据的安全保护技术,主要包括生物特征加密(Biometric Salting)、生物特征密钥生成(Biometric Key Generation)、Fuzzy Schemes 等几大类。通过重点分析这几类方法中的具有代表性的算法,来讨论生物特征数据的安全保护技术的研究及其发展,并进一步指出进行生物特征安全保护技术理论与应用研究的发展方向。

关键词 生物特征识别,安全性,生物特征加密,生物特征密钥生成,Fuzzy schemes

Security Technology of Biometric Data: A Survey

ZHOU Lin-li[1] LAI Jian-huang[2]
(School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou 510275, China)[1]
(School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275, China)[2]

Abstract Biometric offers greater security and convenience than traditional methods of personal recognition. With the growing use of biometrics, there are growing concern about the security and privacy of the biometrics data. Security technologies of biometric data mainly include biometric salting approach, biometric key generation approach, fuzzy schemes approach. According to lay emphasis on analysis of several algorithms of these approaches, this paper summarized the research development of security technologies of biometric data and pointed out future research orientations on the theories and applications research of security technology of biometric data.

Keywords Biometric recognition, Security, Biometric salting, Biometric key generation, Fuzzy schemes

1 引言

1.1 生物特征识别技术简介

现今网络信息化时代,如何准确识别一个人的身份是一个关键性的社会问题。在日常生活中有很多需要进行身份鉴别的应用,例如网上银行、电子商务、互联网以及门禁系统等等。目前,通用的管理手段有使用磁卡、IC 卡和密码等,但是这些手段无法充分满足现实的需要。证件、磁卡等容易遗失或伪造,密码容易遗忘或被窃取,这些都给实际的应用带来很大的不便和诸多安全隐患[1]。

生物特征识别技术不仅可以克服上述不足与缺陷,还能可靠而准确地进行身份识别。因此,生物特征识别技术已经成为身份识别领域的研究热点[1-3]。生物特征识别技术是通过计算机与光学、声学、生物传感器等密切结合,利用人体固有的生理特征和行为特征进行个人身份识别的一种较为可靠、鲁棒和便捷的身份识别方法[1-4]。生物特征识别技术识别的是人本身,由于每个人的生物特征具有与其他人不同的惟一性和在一定时期内不变的稳定性,不易伪造和假冒,因此利用生物识别技术进行身份识别,具有安全、可靠、准确的特点,所以,在国家重要机关及社会安防领域具有广泛而特殊的用途。

目前,可用于生物特征识别的生理特征有手形、指纹、人脸、虹膜、视网膜、脉搏、耳廓等,行为特征有签字、声音、步态等。基于这些特征,发展了手形识别、指纹识别、人脸识别、发音识别、虹膜识别、签名识别等多种生物特征识别技术[4]。

1.2 针对生物特征识别系统的攻击问题

一个生物特征识别系统的整体结构通常由 4 部分组成[1]:特征采集仪、特征提取、生物特征模板数据库和匹配器。特征采集仪是用来采集个体的生物特征数据,如指纹扫描仪、视频采集仪等;特征提取指的是提取生物特征中一些主要和关键性的特征数据;生物特征模板数据库是指存储在中央数据库或智能卡中,注册用户已提取的生物特征数据;匹配器是将待识别的特征数据与生物特征模板数据库中的特征数据进行匹配,根据相应的匹配分值,来判定识别者的身份。

针对生物特征识别系统的 4 个组成部分,存在可能的 8 个攻击和安全性问题(图 1)[1]。

1)伪造生物特征。利用伪造的生物特征在特征采集仪上进行特征采集,进而侵入系统。例如,利用窃取的指纹仿造橡胶手指或利用面具等手段。

2)重复使用生物特征数据。利用曾经使用过的生物特征数据直接作为特征提取器的输入,绕过攻击①中的特征采集仪。例如,利用一个已经使用过的指纹照片或一段录音数据。

3)越过特征提取器。利用木马程序入侵特征提取器,并使特征提取器提取出的特征是被预先选定的。

4)篡改提取后的生物特征数据。生物特征被提取后,其特征数据被具有伪装性的特征数据置换。通常特征提取和匹

*)本项目得到国家自然科学基金(0675016,0633030),广东省自然科学基金(06023194)的支持。周玲丽 工程师,在职博士生;赖剑煌 教授,博导。

配这两部分是合并的，所以在这之间进行攻击的难度很大。但是，如果提取后的生物特征数据必须通过网络在异地进行匹配，则上述的危险是存在的。

5)侵蚀匹配器。匹配器被破坏以至于能够产生预先期望的匹配分值。

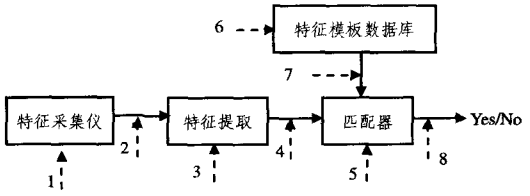

图 1　生物特征识别系统的攻击

6)篡改模板数据库。存储生物特征的数据库和匹配器可以是合并或是分离的。攻击者可以通过修改一个或多个服务器中存储的数据，利用伪造的特征数据来进行正确的匹配，或者破坏数据，以至于正确的特征被系统拒绝。

7)攻击模板数据库和匹配器之间的联接渠道。这种情况发生在模板数据库和匹配器分离的情形下，攻击者可以通过攻击传输渠道进行信息包的篡改。

8)控制输出。如果最后输出的匹配结果可以被攻击者修改，那么整个系统可以说完全失效了。

1.3　生物特征识别技术的安全性和隐私性问题

随着生物特征识别技术的广泛应用，生物特征的安全性日益显得重要和紧迫。尽管生物特征识别技术相比传统的身份识别技术具有本质的优势，但确保生物特征数据的安全性和隐私性是一个严峻问题。

在生物特征识别系统可能遭受到的 8 个攻击中，由于特征提取总是和特征采集仪结合在一起，因此攻击②和攻击③可以忽略。对于使用伪造生物特征的攻击①，可以用生命迹象检测来预防。而对于攻击模板数据库和匹配器之间联接渠道的攻击⑦和控制输出的攻击⑧，不属于生物特征安全性考虑的范畴。因此，一个生物特征识别系统的安全性主要是考虑攻击④、攻击⑤和攻击⑥。其中，针对特征模板数据库的安全性技术[1-3,5]是一个重要的课题。

在具体的应用中，生物特征模板的安全性与隐私性主要面临以下 3 个威胁：

1)特征伪造。如果一个攻击者从特征模板数据库中窃取了生物特征数据，则可以伪造一个生物特征以通过各种识别[6-8]。

2)无法撤销性。由于生物特征的唯一性，导致个人生物特征的丢失就意味着个人身份的丢失。例如，生物特征被破坏或窃取，不能像密码和 IC 卡那样撤销和重新更新[9]，且每个人的生物特征都是有限的，例如一个人只有一个人脸和 10 个手指指纹。另外，研究表明运用“Hill Climbing Attacks”[10]技术，能够获得生物特征模板数据库中的特征数据。因此，生物特征模板数据库的安全性尤显重要。

3)个人隐私的泄露。生物特征包含一些个人的敏感信息。例如，指纹包含某些基因遗传信息，视网膜能够反映糖尿病、高血压等疾病情况等[11-13]。

1.4　生物特征数据安全保护技术的发展概况

鉴于特征模板数据在整个生物特征识别系统中的重要性，本文主要是从特征模板数据的安全性角度出发，探讨生物特征数据的安全保护技术。

由于生物特征的惟一性(Invariable)，以及由此导致的无法撤销性(Irrevocable)和隐私性(Privacy)，Ratha 等[1]于 2001 年提出了可删除生物特征(Cancelable Biometrics)的概念。对生物特征进行人为的、不可逆的变换，使得特征模板数据库中保存的不再是原始的生物特征，而是生物特征的变换形式(图 2)。由于变换的不可逆性，即使攻击者获取了特征模板数据库中的数据，也无法得到原始的生物特征。如果特征模板数据丢失或被窃，则可以通过修改不可逆变换的参数得到新的特征模板数据，使得特征模板数据库中的特征数据具有可删除性，进而克服了生物特征的不可删除性所带来的不足。

在可删除生物特征方法中，由于采用不可逆变换，匹配是在变换域中进行的。而不可逆变换对差异特别敏感，生物特征的微小差异会导致变换后的数据有较大的差距。因此，针对特征模板数据的安全保护性技术需要重点考虑以下的两个问题。

1)噪声。特征数据在采集过程中不可避免地受到一些条件的影响。例如，指纹扫描仪被玷污，人脸采集时光照的影响等，从而导致采集到的特征数据具有一定的噪声[4]。

2)类内差异(Intra Vriance)。由于光照、姿势、表情、化妆等的影响，同一个人在识别时提供的生物特征与注册时的生物特征有一定的差异。而对特征数据进行的变换，能否保持甚至改善这种类内差异[4]，会对生物识别的可辨识性产生很大的影响。

为了增强生物特征识别系统的安全性，Uludag 等[2]于 2004 年提出了生物特征加密系统(Biometric Cryptosystem)的概念，将生物特征识别系统与密码学结合起来。在这个系统中，特征模板数据库中保存的不再是原始的生物特征，而是其在密码学架构下的一种变换形式，从而保护了特征模板数据，并增强了生物特征的隐私性。从密码学的角度来看，生物特征加密系统与可删除生物特征方法是等价的。

Juels 等[27]在 1999 年运用纠错码技术提出了 Fuzzy Commitment 的算法，在此基础上发展了一大类相关算法[14,26,29]。

目前，针对特征模板数据的安全保护技术的研究，主要有以下几大类[14]：

1)生物特征加密(Biometric salting)。

2)生物特征密钥生成(Biometric key generation)。

3)Fuzzy schemes。

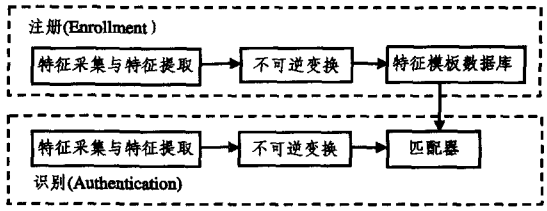

图 2　可删除生物特征方法(Cancelable Biometrics)

生物特征数据的安全性问题日益凸现，而在这方面，国内外没有比较全面的综述，因此本文主要是从特征模板数据的角度出发，对生物特征数据的安全保护技术进行较为全面的论述，重点分析了其中一些具有代表性的算法，并对 Bio-Hashing 算法进行了实验分析，进一步指出了生物特征数据的安全保护技术在理论与应用研究上的方向和趋势。

2 生物特征加密技术(Biometric salting)

在这类方法中,主要是将一个用户特定的随机信息 S 加入到用户的生物特征数据 P 中,然后用哈希算法 H 对 $P+S$ 进行加密。因此,特征模板数据库中保存的是 $H(P+S)$,以此提高了特征数据的安全性,具体的算法见文献[15-19]。

下面主要分析两个生物特征加密(Biometric salting)算法。一个是利用随机卷积核对特征模板数据进行加密的算法,其中用户特定的随机信息 S 是 PIN 码,以 PIN 码为种子通过伪随机生成器来生成随机卷积核。另一个是利用随机映射的 BioHashing 算法,其中用户特定的随机信息 S 是每个用户特定的密钥 K,通过密钥生成随机投影空间。最后通过实验分析该算法的几个关键点。

2.1 随机卷积核加密算法

M. Savvides[16]等提出了利用随机卷积核对生物特征模板进行加密的算法。其中相关滤波器采用的是最小平均相关能量滤波器(MACE,minimum average correlation energy filters)。最小平均相关能量滤波器是一种频域合成的复合滤波器,它克服了一般线性组合相关滤波器旁瓣效应较大的缺点,具有相关峰尖锐、易于识别和定位的特点[32]。

整个算法的过程如下:

1)注册阶段。将训练图像与随机生成的卷积核进行卷积运算,其中卷积核是将用户的 PIN 码作为种子通过伪随机数生成器生成,利用经过卷积运算的训练图像来生成最小平均相关能量滤波器。由于对训练图像做了卷积运算,即使是对生成的最小平均相关能量滤波器进行傅立叶反变换也无法得到原始的训练图像,从而达到了对图像进行保护的目的。具体过程如图 3(其中"*"表示卷积运算)。

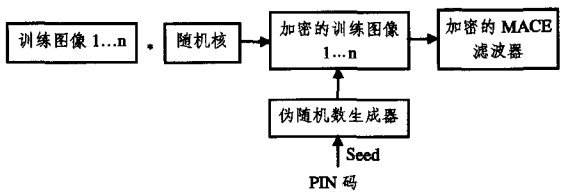

图 3 随机卷积核加密算法的注册阶段

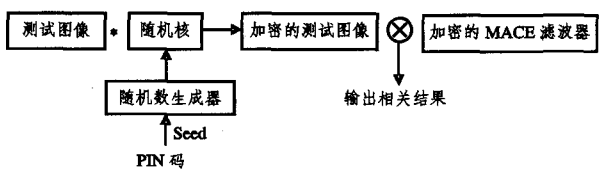

图 4 随机卷积核加密算法的识别阶段

2)识别阶段。测试图像与通过 PIN 码生成的卷积核进行卷积运算。将加密的测试图像与在训练阶段生成的加密 MACE 滤波器进行相关运算,来判断测试的结果。具体过程如图 4。

加密的 MACE 滤波器可以保存在用户的智能卡或 Token 中,以进行用户身份的识别。一旦卡丢失或被窃,则可以重新生成一个新的加密 MACE 滤波器,进而具有可重复生成的性质。攻击者由于不知 PIN 码从而无法得到随机核,即使对加密的 MACE 滤波器进行反卷积运算,也无法得到原始的图像。

整个识别过程均是在加密域中进行的,其中 MACE 滤波器的相关结果是用峰值旁瓣比(PSR)来度量。这是因为采用加密的 MACE 滤波器与不加密的 MACE 滤波器得到的峰值旁瓣比几乎相等,这也正是利用随机卷积核对生物特征数据进行加密的理论基础。但是该算法对系统的可辨识性、识别率等没有进行分析,也没有后续的研究报道。

2.2 BioHashing 算法

为了解决生物特征识别过程中错误拒绝率(FRR)较高的问题,Goh[15,19]和 Teoh 等[33,34]提出了一种基于两因素的生物特征识别算法 BioHahsing,使得系统的等错率 EER(Equal Error Rate)能够为零。这种算法是将提取的生物特征向量与 Token(或智能卡)中存储的随机序列进行迭代内积运算,从而得到一些基于用户的编码(BioHashCode)。从某种意义上来说,BioHahsing 算法引入了外部因素,近似于不可逆变换和加密算法,从而达到对生物特征进行保护的目的。

BioHahsing 算法的主要过程(图 5)由两部分组成:

1)随机投影(Random mapping)。将特征向量投影到每个用户各自不同的子空间(子空间是通过用户 Token 中存储的密钥 K 来生成)。

2)阈值化(Thresholding)。通过阈值处理将投影结果二值化,从而将实值的特征向量转化成一个二值的串,特征模板数据库中存储的是二值化的特征数据,而不再是原始的特征数据。

由 Johnson-Lindenstrauss 引理[37]表明随机投影能保证欧氏距离的不变性,从而满足了可辨识的要求。而 BioHashCode 之间的匹配则是通过汉明距离(Hamming distance)来进行。

B. Kong 等[36]提出 BioHashing 算法中零等错率的获得是建立在一个不实际的假设之上的,就是假设密钥 K 没有被窃。因此,Alessandra Lumini 等[37]提出了改进的 BioHahsing 算法,通过将特征向量进行循环置换、投影空间增大等方法,使得 BioHashCode 的长度增加,从而提高算法的效果。

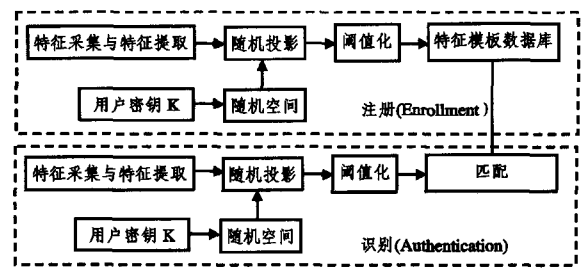

图 5 BioHashing 算法

2.3 BioHashing 算法分析与探讨

由于 BioHahsing 算法提出的框架引入了外部因素(密钥 K),从而达到对生物特征进行保护的目的。目前,越来越多的研究基于这种算法框架来展开,因此针对 BioHashing 算法,我们运用实验来具体分析影响算法的一些关键性因素。实验中用 ROC(Receiver Operating Characteristic)曲线来衡量 BioHashing 算法的优劣,使用的人脸库是 ORL(Olivetti Face database),其中有 40 个人,每个人有 10 幅人脸图像,其中 5 幅用来训练,5 幅用来测试。

BioHashing 算法主要包括两部分:特征提取和特征离散化(阈值化)。在特征提取部分,特征提取算法的不同和提取的特征向量维数的大小对算法有很大的影响。图 6 和图 7 分别比较了采用 PCA (Principal component Analysis)和 LDA

(Linear Discriminant Analysis)算法提取特征，特征向量维数不同的情形，图中 PCA-和 LDA-后面的数字分别表示提取的特征向量维数。从图中可以看出，特征维数越大，算法越好。图 8 中比较了采用 PCA 算法和 LDA 算法提取特征对整个算法的影响，可以看到 LDA 算法提取特征效果好些。对特征向量二值化的过程，采用了一个阈值，这个阈值的选取对算法也有一定影响。如图 9 所示，阈值选取得越大，效果越好。

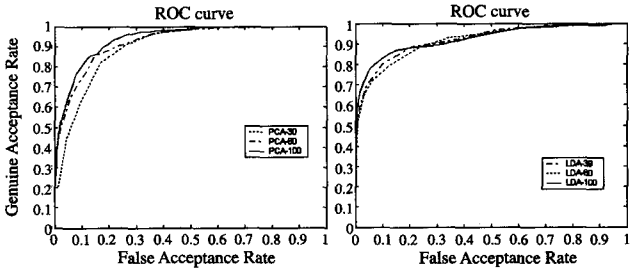

图 6　特征维数的比较(PCA)

图 7　特征维数的比较(LDA)

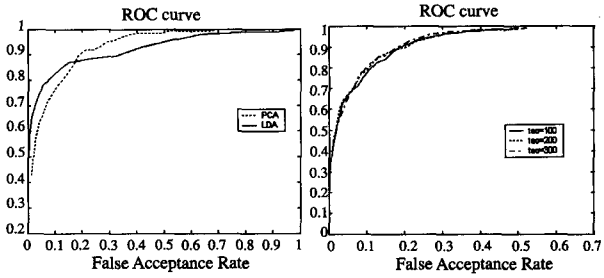

图 8　特征提取算法的比较

图 9　不同阈值的比较

BioHashing 算法基于两个因素来考虑，一个是用户特有的密钥 K，另一个则是用户的生物特征。该算法及其后续的研究中，只考虑了其中一个因素的丢失对整个算法安全性的影响。但是，如果用户的密钥和生物特征同时被窃取或篡改的话，算法的安全性则彻底得不到保证，这正是该算法最大的缺陷。

3　生物特征密钥生成技术(Biometric key generation)

这类方法是直接从生物特征信号中产生一个密钥。模板数据库中存储的是生物特征 B 的密钥 $K(B)$，通过 $K(B')$ 与 $K(B)$ 的比较来识别另一个生物特征 B'。这类方法主要面对的问题是，如何从带有噪声的生物特征数据中得到鲁棒的密钥，以及密钥的容错性能，主要算法见文献[20-23]。

在这类方法中，主要介绍加强口令(Hardened password)算法[21]，该算法在概念上与生物特征加密有一定的相似性又有所不同。

传统的用户名-口令认证一直是对访问计算机系统的用户进行身份认证的主要方式。但这种用户身份认证机制的最大弊端是口令容易被盗，存在着严重的安全隐患。尤其是口令如果设得较简单，那么通过基于字典的方式很容易被他人破解。

鉴于目前口令保护机制的脆弱性，Monrose 等[21]提出利用用户的击键特征(keystroke)和口令(pwd)生成加强口令(hpwd)以提高用户身份认证的安全性。击键特征是指由于人们对键盘的熟悉程度，以及击键习惯等不同，使得每个人在输入自己口令时形成了自己独特的击键模式。例如，按键的持续时间(Key hold times)以及连续两次击键的时间间隔(keystroke latency)等。算法的具体过程如下：

1)注册阶段。在模板数据库中存储如下信息：

① 长为 K 比特的随机数 R。

② 由口令加密的指示表(Instruction table)。将用户的击键特征通过阈值化处理得到二值的特征表示，结合随机数 R，运用香农密钥共享原理(Shamir's secret sharing scheme)[38]生成指示表。

③ 利用 hpwd 加密的历史文件(History file)。

2)识别阶段。利用模板数据库中的随机数 R、指示表及认证口令 pwd'，计算出 $hpwd'$，然后利用 $hpwd'$ 去解密历史文件。如果解密成功，则通过认证，且在模板数据库中对用户的随机数 R 和历史文件进行相应的修改。如果在规定次数内，解密没有成功，则认证失败。

这个算法的不足之处在于：一方面，口令的熵值只是增加了 15 比特，因此安全性并没有得到较大的改进。在其后续研究中，Monrose 等[22,39]对算法进行了一些修改，并运用到声音特征(Voice)中。另一方面，由于用户的击键特征具有较大的不稳定性，因此该算法的具体实用性值得进一步的探讨。

4　Fuzzy schemes 技术

这类方法引入了辅助信息 P(又称为帮助数据 helper data、屏蔽函数 shielding functions 或模糊提取器 fuzzy extractors 等)，辅助信息 P 并不会对生物特征数据造成泄漏。主要的思想是：用户特有的信息 K 和生物特征 B，产生一个公共的信息 P 和秘密的信息 S，$Gen(B,K)\rightarrow\langle S,P\rangle$。如果另一个用户的生物特征 B' 与特征 B，在某种度量下足够相似的话(B 与 B' 并不完全相等)，则通过特征 B' 和公共信息 P，可以得到秘密信息 S，$Rep(B',P)\rightarrow S$。主要的算法见文献[24-29]。

由于生物特征数据中噪声的存在，因而考虑引入纠错码(Error Correct Code)对噪声导致的误差进行处理。Juels 等[25]提出了 Fuzzy Commitment 的算法，在此基础上 Juels 等[26]进行改进，提出了 Fuzzy Vault 的算法，这类算法的大致框架如图 10 所示。

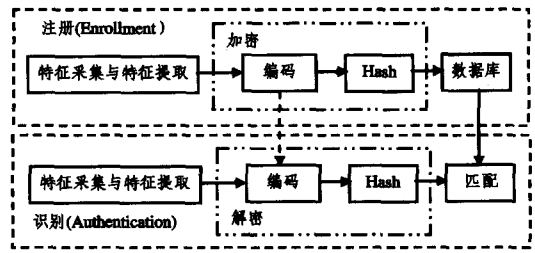

图 10　Fuzzy schemes 框架

4.1　Fuzzy Commitment schemes 算法

在 Fuzzy Commitment schemes 算法中，纠错码与原始特征向量的长度相等，均为 n，且其最小距离为 $2t+1$，t 是纠错码能纠正的错误数。f 是投影函数，能将长度为 n 的特征向量投影到对应汉明距离最小的码上。如果一个向量 v 与一个码字 C 的距离不超过 t 的话，则有 $f(v)=c$。

Fuzzy Commitment schemes 算法的主要过程如下：

1)注册阶段。C 是随机产生的码字，即为秘密信息，T 为特征向量，数据库中存储的是 C 的 Hash 形式 $hash(C)$，以及 T 与 C 的差值 $T-C$，整个过程表示为(加密)：$E_n(T)=[hash$

$(C), T-C]$。

2)识别阶段。待识别的特征为 T',则有 $C'=f(T'-(T-C))=f(C+(T'-T))$

如果 T 与 T' 的距离不大于阈值 t 的话,由 f 的性质,则 C 与 C' 相等,通过比较 $hash(C')$ 与 $hash(C)$ 得到匹配结果。

Fuzzy Commitment schemes 算法中数据库中存储的是 $hash(C)$ 与 $T-C$,从而避免了特征数据 T 的泄露。但是该算法的不足在于:要求特征数据 T 和 T' 的表达形式是有序的;并且算法假设特征数据的分布是服从均匀分布的。但是特征数据的分布是否服从或近似于服从均匀分布并不是很明确,算法只提出了一个框架,并没有采用具体的纠错码来进行讨论。

4.2 Fuzzy Vault Schemes 算法

Fuzzy Commitment schemes 算法的一个不足是要求特征数据的表示形式必须是有序的,为此 Juels 等[26,29] 提出了 Fuzzy Vault Scheme 算法。该算法主要是基于集合的差(set difference),利用集合 A 中的无序元素将秘密信息 S 加密到(fuzzy vault)V 中。如果想从 V 中将秘密信息 S 解密出来,必须具有另一个与集合 A 足够相似的集合 B。如果集合 A 和集合 B 只有较少的元素不相等,则秘密信息 S 可以被成功地解密。

Fuzzy Vault Schemes 算法的具体过程:

1)利用 Reed-Solomon 编码[40],用集合 A 将秘密信息 S 隐藏起来。将秘密信息 S 作为多项式 $p(x)$ 的系数,对于集合 A 中的元素 $x_1, x_2, \cdots, x_n$,分别计算 $p(x_1), p(x_2), \cdots, p(x_n)$,再加入一些随机产生的不依赖于多项式 $p(x)$ 的 Chaff 点形成了点集 R,Chaff 点的加入增强了秘密信息 S 的安全性。

2)若希望利用集合 B 获取秘密信息 S,如果集合 B 与集合 A 足够相似的话,能够通过计算多项式 $p(x)$ 对于集合 B 中各个元素的相应取值,获得点集 R 中依赖于多项式 $p(x)$ 的大多数点。这样,依赖于多项式 $p(x)$ 的大多数点被找到,只有少部分有误差,再利用 Reed-Solomon 解码来重构多项式 $p(x)$,这样隐藏在多项式 $p(x)$ 中的秘密信息 S 就被提取出来。

虽然 Fuzzy Vault Scheme 算法解决了 Fuzzy Commitment schemes 算法中的有序问题,但是其针对生物特征信号中特有的差异性(噪声)算法的鲁棒性不是很明确,并且没有提及如何解决特征数据表达中的对齐问题。

5 未来的发展方向与展望

尽管针对生物特征数据的安全保护技术的研究才刚刚起步,理论研究还不是很完善,但由于生物特征识别的广泛应用,生物特征数据的安全性日益显得重要,从而进行这方面的理论研究还是很有应用价值的。

本文讨论了针对生物特征模板数据的三类安全性技术算法。其中,在不可逆变换算法中,如何找到既具有不可逆性质又保持可判别性(Discriminability-preserving)的变换,是这类方法需要解决的主要问题,也是这类算法的难点。在生物特征密钥生成算法中,面临一个主要问题,就是从具有差异或带有噪声的生物特征中如何获取具有鲁棒性的特征密钥。因此,针对生物特征模板数据的安全性技术,目前的研究工作主要是侧重于生物特征加密(Biometric salting)和 Fuzzy schemes 这两类算法。

生物特征加密算法主要是建立在可删除生物特征(Cancelable Biometrics)的思想上,在该类算法中,最值得关注的是下面几个方面:

1)二值化问题。在这类算法中需要对特征数据进行变换,其中主要是进行二值化处理,因此面临这样一个二值化的转换问题。需要将实值的特征数据转化为二值的数据,这种变换不仅不可逆而且必须要满足保持类的分布,如何构造这种二值变换是一个值得研究的问题。

2)类内变化问题。生物特征具有较大的类内变化,例如人脸特征。如果采用不可逆变换或其他安全性方面的算法则效果不佳。如何使变换后的特征数据能够保持类内变化,增强类间的差异,也是一个研究的热点。针对这个问题,Feng 和 Yuen[42] 提出了基于人脸特征的保持类分布的变换(CDP, Class Distribution Preserving),通过 CDP 变换能够使得类内的差距在变换前后变化不大。

Fuzzy schemes 算法主要是基于纠错码技术提出的,此类方法不仅能保护特征模板数据,而且能阻止攻击者侵入生物特征识别系统。在该类算法中,以下几个研究方向值得关注。

1)预处理过程。由于纠错码对输入具有一定的要求,例如,汉明码和 BCH 码要求输入是二值的。而原始的特征模板数据不满足这些要求,因此需要进行一个预处理的过程,使特征模板数据符合纠错码的要求。

2)纠错码的选择问题。在 Fuzzy schemes 类方法中,RS 码和 BCH 码的应用较多,但是哪种纠错码的效果是最优的,值得进一步的研究。

目前,我们利用各种人脸数据库,分析了三种主要的外在因素(光照、表情和姿势)对人脸特征的影响所导致的错误分布情况,发现其中光照造成的影响最为突出。因此,在对特征数据进行安全保护技术的研究时,重点要考虑光照带来的噪声问题。

另外,在应用研究中,要根据不同生物特征的特点来具体考虑。在已有的一些研究中,主要集中于对虹膜、指纹等的研究,对人脸特征的研究较少,而人脸特征与虹膜、指纹等在数据的表达形式上有很大的不同。人脸的特征数据是实值向量,而指纹或虹膜的特征数据是二值的串,并且人脸特征数据的维数较大。因此,在应用各种技术对特征数据进行保护时,要考虑人脸的特殊性,及各类算法的适用性。

其次,可以考虑将加密技术(密码学)中更好的方法引入到对生物特征的保护中,例如,将多生物特征的融合技术与加密技术结合起来,既能提高识别率也能达到安全的目的。或考虑引入外部因素(Token 等)来对生物特征进行保护,从而提高安全性。

结束语 随着 911 事件的发生,生物特征识别日益广泛应用于各行各业。因此,生物特征数据的安全保护技术目前正成为生物特征识别领域中较为活跃的研究主题之一。本文描述和分析了生物特征数据安全保护技术的发展和其中几类具有代表性的算法,并通过实验分析了其中一个算法。最后,对未来的研究和发展趋势做了分析,希望能对相关领域的研究人员和技术人员有所裨益。

参 考 文 献

[1] Ratha N K, Connell J H, Bolle R M. Enhancing Security and Privacy in biometrics-based authentication system. IBM Systems Journal, 2001, 40(3)

[2] Uludag U, Pankanti S, Prabhakar S, et al. Biometric cryptosys-

tems: issues and challenges// Proceedings of the IEEE. 2004, 92 (6): 948-960

[3] Jain A K, Ross A, Pankanti S. Biometrics: A Tool for Information Security. IEEE Transactions on Information Forensics and Security, 2006, 1(2): 125-143

[4] Jain A K, Ross A, Prabhakar S. An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(1): 4-20

[5] Ratha N K, Connell J H, Bolle R M, et al. Cancelable biometrics: A case study in Fingerprints// Proceedings of International Conference on Pattern Recognition. 2006

[6] Bolle R M, Connell J, Pankanti S, et al. Biometrics 101. Report RC22481. IBM Research, 2002

[7] Maltoni D, Maio D, Jain A K, et al. Handbook of Fingerprint Recognition. New York: Springer-Verlag, Inc. , 2003

[8] van der Putte T, Keuning J. Biometrical fingerprint recognition: Don't get your fingers burned // IFIP TC8/WG8. 8 Fourth Working Conference on Smart Card Research and Advanced Applications. Kluwer Academic Publishers, 2000: 289-303

[9] Schneier B. Inside risks: The used and abuses of biometrics. Comm. of the ACM, Aug. 1999, 42: 136

[10] Adler A. Images can be regenerated from quantized biometric match score data// Proceedings of Canadian conference of Electrical and Computer Engineering. 2004: 469-472

[11] Babler W J. Embryologic development of epidermal ridges and their configuration. Birth Defects Original Article Series, 1991, 27(2)

[12] Mulvhill J J. The genesis of dermatoglyphics. The Journal of Pediatrics, 1969, 75(4): 579-589

[13] Penrose L S. Dermatoglyphic topology. Nature, 1965, 205: 545-546

[14] Ratha N K, Chikkerur S, Connell J H, et al. Generating cancelable fingerprint templates. IEEE Transactions on Pattern Analysisand Machine Intelligence, 2007, 29(4): 561-572

[15] Goh A, Ngo D L. Computation of Cyrptographic Keys from Face Biometrics// Proc. IFIP: Int'l Federation for Information Processing. 2003: 1-13

[16] Savvides M, Vijaya Kumar B V K, Khosla P K. Cancelable Biometric Filters for Face Recognition // Proc. Int'l Conf. Pattern Recognition. 2004: 922-925

[17] Soutar C, Roberge D, Stoinav A, et al. Biometric Encryption Using Image Processing// Proc. SPIE. 1998, 3314: 174-188

[18] Connie T, Teoh A B J, Goh M K O, et al. PalmHashing: A Novel Approach for Cancelable Biometrics. Information Processing Letters, 2005, 93(1): 1-5

[19] Jin A T B, Ling D N C, Goh A. Biohashing: two factor authentication featuring fingerprint data and tokenized random number. Pattern Recognition, 2004, 37(11): 2245-2255

[20] Davida G I, Frankel Y, Matt B. On Enabling Secure Applications through Off-Line Biometric Identification// Proc. IEEE Symp. Security and Privacy. 1998: 148-157

[21] Monrose F, Reiter M K, Wetzel S. Password Hardening Based on Key Stroke Dynamics // Proc. ACM Conf. Computer and Comm. Security. 1999: 73-82

[22] Monrose F, Reiter M K, Li Q, et al. Cryptographic Key Generation from Voice // Proc. IEEE Symp. Security and Privacy. May 2001: 202-213

[23] Vielhauer C, Steinmetz R, Mayerhoefer A. Biometric Hash Based on Statistical Features of Online Signatures // Proc. 16th Int'l Conf. Pattern Recognition. 2002, 1: 123-126

[24] Dodis Y, Reyzin L, Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data// Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques. May 2004: 523-540

[25] Juels A, Wattenberg M. A Fuzzy Commitment Scheme // Proc. Sixth ACM Conf. Computer and Comm. Security. 1999: 28-36

[26] Juels A, Sudan M. A Fuzzy Vault Scheme// A. Lapidoth and E. Teletar, eds. Proc. IEEE Int'l Symp. Information Theory. 2002: 408

[27] Linnartz J P, Tuyls P. New Shielding Functions to Enhance Privacy and Prevent Misuse of Biometric Templates// Proc. Fourth Int'l Conf. Audio and Video-Based Biometric Person Authentication. 2003: 393-402

[28] Tuyls P, Akkermans A H M, Kevenaar T A M, et al. Practical Biometric Authentication with Template Protection // Proc. Sixth Int'l Conf. Audio and Video-Based Biometric Person Authentication. 2005: 436-446

[29] Uludag U, Pankanti S, Jain A K. Fuzzy Vault for Fingerprints// Proc. Sixth Int'l Conf. Audio and Video-Based Biometric Person Authentication. 2005: 310-319

[30] Ang R, Safavi-Naini R, McAven L. Cancelable Key-Based Fingerprint Templates // Proc. 10th Australian Conf. Information Security and Privacy. July 2005: 242-252

[31] Sutcu Y, Sencar H T, Nemon N. A Secure Biometric Authentication Scheme Based on Robust Hashing// Proc. Seventh Workshop Multimedia and Security. 2005: 111-116

[32] Mahalanobis A, Vijaya Kumar B V K, Cassset D. Minimum average correlation energy filters[J]. Applied Optics, 1987, 26 (17): 3633-3640

[33] Teoh B J A, Ngo C L D. Cancellable Biometrics Featuring with Tokenised Random Number. Pattern Recognition Letters, 2005, 26(10): 1454-1460

[34] Teoh B J A, Ngo C L D, Goh A. Personalised Cryptographic Key Generation Based on FaceHashing. Computers and Security J. , 2004, 23(7): 606-614

[35] Johnson W B, Lindenstrauss J. Extensions of Lipschitz mappings into a Hilbert space// Conference in modern analysis and probability (New Haven, Conn. , 1982). 1984: 189-206

[36] Kong B, Cheung K, Zhang D, et al. An analysis of Biohashing and its variants. Pattern Recognition, 2006, 39: 1359-1368

[37] Lumini A, Nammi L. An improved BioHashing for human authentication. Pattern Recognition, 2007, 40: 1057-1065

[38] Schneier B. Applied Cryptography: Protocols, Algorithms, and Source Code in C. 2nd ed. New York: Wiley, 1995

[39] Monrose F, Reiter M K, Li Q, et al. Using voice to generate cryptographic keys // Proc. 2001: A Speaker Odyssey, Speaker Recognition Workshop. 2001: 237-242

[40] Lin S , Costello D J . Error Control Coding . Pearson Prentice Hall. Second edition. New Jersey, 2004

[41] Feng Y C, Yuen P C. Class-Distribution Preserving Transform for Face Biometric Data Security// Proceedings of IEEE International Conferences on Acoustics, Speech, and Signal Processing (ICASSP). 2007