**2018年全国大学生信息安全竞赛**

**作品报告**

**作品名称：基于信息负表示的隐私保护虹膜识别系统**

**电子邮箱： 1114787570@qq.com**

**提交日期： 2018年5月30日**

填写说明

1. 所有参赛项目必须为一个基本完整的设计。作品报告书旨在能够清晰准确地阐述（或图示）该参赛队的参赛项目（或方案）。

2. 作品报告采用A4纸撰写。除标题外，所有内容必需为宋体、小四号字、1.5倍行距。

3. 作品报告中各项目说明文字部分仅供参考，作品报告书撰写完毕后，请删除所有说明文字。(本页不删除)

4. 作品报告模板里已经列的内容仅供参考，作者可以在此基础上增加内容或对文档结构进行微调。

5. 为保证网评的公平、公正，作品报告中应避免出现作者所在学校、院系和指导教师等泄露身份的信息。

**目 录**

[摘要 4](#_Toc515961609)

[第一章 作品概述 6](#_Toc515961610)

[第二章 作品设计与实现 12](#_Toc515961615)

[第三章 作品测试与分析 29](#_Toc515961632)

[第四章 创新性说明 31](#_Toc515961633)

[第五章 总结 33](#_Toc515961634)

[参考文献 35](#_Toc515961635)

# 摘要

为了加强虹膜识别系统的安全性，保护用户的虹膜数据，我们选择了负数据库作为隐私保护的手段，开发了“负虹膜识别系统”。该系统可以做到隐藏用户虹膜数据的同时，对用户身份进行识别，是一套安全、高效的虹膜识别系统

负虹膜识别系统，是基于信息负表示、识别负匹配的安全策略所设计出的一套完备的虹膜识别系统。信息的负表示是受人工免疫系统中的负选择机制启发得到，负数据库是信息负表示的主要模型之一。传统的数据库通常直接存储原始数据本身，即是以“正”的形式存储数据，故可称为“正数据库”。负数据库则通过存储原始数据的补集信息来实现数据安全、隐私保护等目的，因其存储的信息是原始数据的补集，故称其为“负数据库”。负数据库通过引入任意匹配符‘\*’对补集信息进行压缩，实现高效存储和操作。Esponda等人证明了对负数据库求逆恢复出原始数据是*NP*难的[1，2]。

本作品的实现利用的是负数据库不可恢复的特性，改进目前的虹膜识别系统，使之达到虹膜信息“抗泄漏”的目的，我们称其为“负虹膜识别机制”。这样即使虹膜信息数据库被攻破，攻击者也无法通过数据库中的“负”的信息恢复出用户的“正”的虹膜信息，这有效的避免了用户生物特征信息数据的泄露，极大地提高了系统的安全性。

本项目的创新点在于：

1. 利用海明距离计算公式的特性，通过改进已有的基于负数据库的海明距离计算公式，对负数据库进行压缩。将原有负数据库中的*m×r*条负数据压缩成了两条，大大提高了系统的计算效率。
2. 压缩的负数据库使得对负数据库求逆恢复出原始数据这个NP难解的问题变得更加难以解决，进一步提升了系统的安全性。
3. 利用了负虹膜识别技术实现了一套负虹膜识别系统，负虹膜识别技术是一种高效的基于负数据库的安全虹膜识别体制，通过扩展虹膜识别中的经典策略（如移位和掩码），提升了负虹膜识别的识别性能。
4. 负虹膜识别系统具有良好的可扩展性，它可用于结合其他的一些信息认证手段，加强系统的安全性。

# 第一章 作品概述

## 1.1 背景分析

随着互联网技术的发展、“互联网+”时代和大数据时代的到来，万物互联、人机交互的快速发展，网络中传播和分享的信息越来越多。海量的数据中往往包含一些敏感信息或者隐私数据，这些数据的泄露会导致个人或机构的权益受到损害。据媒体调查显示，互联网时代，55.8%的受访者认为保护个人隐私“越来越难”，29.3%的人认为，“个人信息被随意公开泄露”。近几年出现了多次用户隐私数据泄露的事件，隐私保护和数据安全已成为人们重点关注的问题之一，加强个人信息保护刻不容缓。

现如今，生物识别技术已经逐渐普及到人们的日常生活中，不仅仅是常见的手机指纹，人脸检测。识别稳定性、准确率都极为优秀的虹膜识别技术同样被广泛应用于军事、银行等高度限制性区域，有广阔的发展前景。由于生物特征的唯一性和在一定时期内不变的稳定性，不易伪造和假冒，所以利用生物识别技术进行身份认定，安全、可靠、准确。但是由于人的一些生物特征的大部分信息在一生中通常都是不变的，生物特征数据一旦泄露也就意味着永久性泄露，攻击者可以利用泄露的生物特征数据伪装他人获得非法的权益，这将不可避免的造成个人和社会利益的损失。

虽然如今已有大量的数据安全和隐私保护技术被提出，这些技术均有各自的优缺点，无法适用于不同的应用环境。进入大数据时代后，隐私保护和数据安全方面面临着更大的挑战，不仅需要保证足够的安全性，更要确保具有足够高的效率，这使得很多传统的安全技术（例如传统的加解密算法）难以适用。因此，我们急需引入一些新的安全技术用来加强、完善对生物特征信息识别系统中生物特征数据的保护，才能本质上提高系统的安全性，使系统能更好的步入市场，服务于用户。

以虹膜识别系统为例，本文档提出对系统数据库中存放的用户虹膜信息进行加密，以信息补集的形式进行保存，确保在数据库被非法入侵时，个人虹膜的明文信息不会泄漏，同时也保证了识别过程的准确性与高效性。

## 1.2 相关工作

近些年来虹膜识别已经成为全球各高校研究室里的热点课题，大量的虹膜识别算法被先后提出，虽各有特点，但大多都集中于四个部分：虹膜图像获取、虹膜分段、虹膜结构分析和虹膜特征表示与匹配。其中大多数算法都是将虹膜特征表示为二进制串，期间有很多研究人员使用过滤器来产生虹膜图像的二进制表示。例如，Daugman就在文献[3]中提出了一种高效的虹膜识别系统，其使用Gabor滤波器来过滤虹膜结构，通过一系列的小波变换，最终以一个二进制串的形式表示一个虹膜的信息。在文献[4]中，Lim等人采用一个二维Haar小波转换来提取虹膜结构，最终同样是以一个二进制串的形式表示一个虹膜的信息。除了过滤器以外，还有一些其他的技术被用于虹膜特征提取和二进制编码。例如，在文献[14]中，Tisse等人使用二维Hilbert转换来提取虹膜特征，然后构建一个解析图像并产生二进制编码。在匹配和识别阶段，他们也使用了海明距离。在文献[15]中，局部直方图均衡化和自适应系数阐值被用于提取虹膜特征和创建二进制编码，再通过计算不同虹膜的特征之间的空间对应关系来实现匹配和识别。

从现今已有的虹膜识别工作来看，大多都是将虹膜信息转化为二进制的形式进行存储。而虹膜，作为人体的一大重要器官，其有终身不变的特点，如果虹膜信息被非法获取、利用，其后果不堪设想。大量的学者们也认识到了这个问题，同时有许多安全虹膜识别体制被提出。Davida等人在文献[5，6]中提出了首个使用密钥生成生物特征信息加密的系统。Hao、Anderson等人在文献[7]中提出了一个基于模糊承诺机制的安全虹膜识别体制[8]，并获得了不错的识别性能。

但是目前，大部分已有的安全虹膜识别体制的安全性并没有形式化的证明。并且几种在虹膜识别领域常用的安全技术已被证明是不安全的。所以，在虹膜识别体制中引入新型的安全加密技术是非常重要的。

## 1.3 特色描述

1. 高准确率的虹膜识别算法改进

近年来，虹膜识别已经成为了一个热点研究课题，前前后后已经有许多国内外的学者就这一课题在各大期刊上发表了相关文章。而就对虹膜图像编码这一问题也已经有了许多高效、准确的解决方案，所以这里我们经过实验对比，选择了Guillaume Sutra，Bernadette Dorizzi等人在文献[9]中提出的一种可行策略OSIRIS-V4.1，该策略将虹膜数据库CASIA-IrisV3-Interval[16]中每张虹膜图像转换为两长灰度图，即虹膜灰度图像和标准化的掩码灰度图像，再分别将其转换为二进制串。最后通过计算原始虹膜图像二进制串与待识别虹膜图像二进制的海明距离，以一定的阈值确定匹配结果。在我们的系统中，对该方案进行了必要的修改，加入掩码和移码策略，结合基于统计规律的修正海明距离估算公式可以使得虹膜识别的准确性进一步提升。

1. 基于信息负表示的安全的虹膜信息存储

生物特征识别是指在保护生物特征的前提下支持有效的身份识别功能。但是由于每次提取的生物特征信息会因环境等因素的影响而存在着微小的差异，所以这阶段的难点在于如何对虹膜数据进行加密并同时能够支持有效匹配。信息负表示是一种新的数据加密技术，由Esponda等人根据生物免疫系统中的负选择机制启发得到，取“正数据”（原始数据）的补集，再引入任意匹配符对原始数据进行压缩，可实现高效存储和操作。基于信息负的计算方法有很多，其中*p*-hidden算法有如下优点：

1. 生成难解的负数据库，使得攻击者难以由负数据库求逆恢复出原数据。
2. 能够细粒地控制所生成的数据库特征。众所周知，在工程应用中，细粒度的控制有着重要的作用。虽然，目前已有一些负数据库生成算法被提出，但是大部分己有的算法都难细粒度控制所生成的负数据库的特征（如负数据库中不同类型的记录的分布），从而难以细粒度地控制负数据库的难解性。
3. 基于*p*-hidden算法生成负数据库的海明距离估算法

根据文献[10]中的描述，对于给定任意串*x*和确定串*s*采用*p*-hidden算法生成的负数据*NDB*。可由*NDB*中位置集合与串*x*中对应位置的差异性，依据统计规律由以下式子估算串*x*与*s*之间的海明距离的期望值：



基于负数据库的海明距离估算法为负数据的模糊匹配提供了快速而高效的解决方法，它架起了负数据库和明文之间比对的桥梁。其中各参数的含义以及对公式在实际应用中的改进将在第二章中详细说明。

## 1.4 前景分析

1. 理论前景上

生物识别如今得到广泛关注，在各项生物特征信息中，虹膜最具有特殊性，每个人的虹膜都独一无二，伴随一生，并且活体虹膜难以伪造，适合用来进行身份识别认证。此时需要解决的就是数据库中的虹膜数据的安全问题，本系统在此角度上，完美解决的了上述问题，为个人隐私和系统安全做出了保障，具有广阔的发展前景

1. 科技前景上

如今嵌入式和移动设备硬件技术越来越发达，指纹，人脸，声音识别都已经全面普及，在如此迅速的发展状态下，更为精密的虹膜扫面器也终将会登上“历史的舞台”，本产品也能为虹膜识别提供后方保障，服务用户。

1. 市场前景上

据统计数据显示，2016年全球企业级虹膜识别终端出货量为1070万台，预计到2025年出货量将达到6160万台，复合年均增速为19.1%，对应市场规模将从2016年的6.77亿美元增长至2025年的41亿美元，复合年均增速19.7%。其中，2020年市场规模约为13.5亿美元，这仅仅是商业上的现状，更别说在高度安全级别的军事控制上，虹膜识别更是越来越受重视，无处不在。

总的来说，基于信息负表示的虹膜识别隐私保护系统具有优秀的商业、军事发展前景，值得进行深入的研究、完善、投资和开发。

# 第二章 作品设计与实现

## 2.1 系统方案

本节将对我们的作品“虹膜识别系统”做一个详细的架构设计描述。阐述了系统的整体框架，包括实际设备、软件逻辑的结构关系，说明了系统所采取的设计策略和软件在实际应用中的需求关系。

### 2.1.1 设计思路

在生物识别领域，虹膜识别已经被证明是安全的识别技术，以虹膜识别技术为基础的身份认证系统也已经得到一定的应用。我们的负虹膜识别系统的设计思路与其他虹膜识别系统大致相似，整个系统可以分为多个独立的模块，为提高系统内聚性，采用分布式系统设计，三层C/S结构，实现用户模块与管理模块、识别模块与匹配模块的分离。系统的组成及功能关系如图2-1所示。

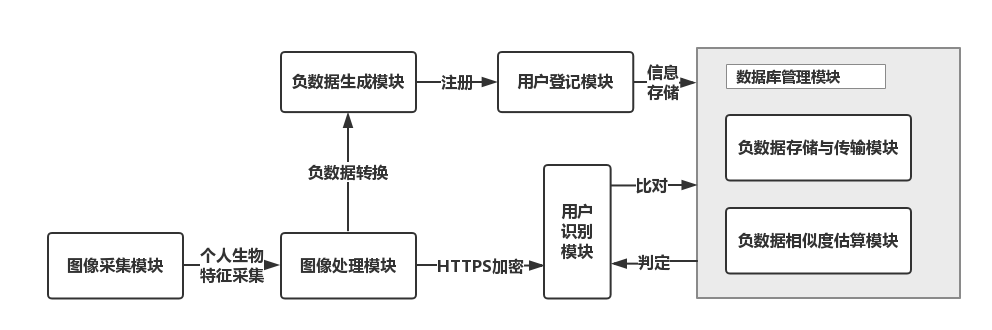


图2-1 系统组成

其中图像采集模块使用专用的虹膜采集设备，在图像处理模块对采集得到的图像进行切割、归一化以及编码工作，得到的二进制串可视作用户虹膜信息的明文。进行注册时需要使用用户虹膜信息明文，通过负数据生成模块得到对应负数据库，最终存储到管理端数据库。进行识别时则需要对用户虹膜信息明文通过HTTPS加密，并从用户识别模块发送到管理端进行负数据相似度估算比对，最后接收到从管理端返回的判定消息，完成识别。

### 2.1.2 系统架构

系统采用三层式C/S结构，实现中将业务逻辑层的关键模块“负数据生成模块”集成到扫描端，以消除虹膜信息传输过程中原始虹膜信息明文被拦截的风险。管理端完成虹膜负数据的接受存储以及负数据的相似度比对，同时管理端能够可视化的完成对注册用户的增删查改。扫描端与管理端的拓扑结构如图2-2。

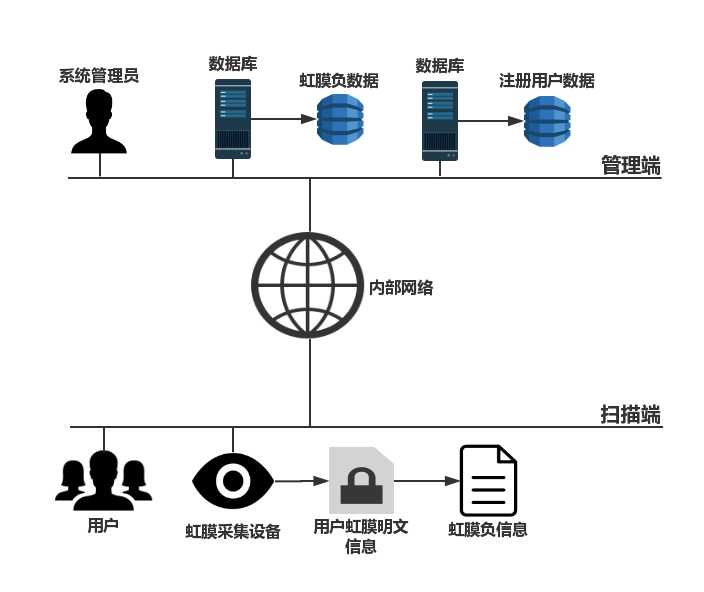


图2-2 系统拓扑结构

## 2.2 实现原理

这一节将通过详细介绍系统的三个主要功能来说明整个系统中一些重要的实现原理，其中三个主要功能分别是虹膜图像编码、负数据库生成和合法用户匹配。为了尽可能节省文本篇幅，也是为了尽力贴合大赛主题，以下对三个功能实现原理的叙述会有一定的侧重。下面就进入三个主要功能的详细介绍：

### 2.2.1 虹膜图像编码

近年来，虹膜识别已经成为了一个热点研究课题，前前后后已经有许多国内外的学者就这一课题在各大期刊上发表了相关文章。而就对虹膜图像编码这一问题也已经有了许多高效、准确的解决方案，所以这里我们经过实验对比，选择了Guillaume Sutra，Bernadette Dorizzi等人在文献[9]中提出的一种可行策略OSIRIS-V4.1，该策略将虹膜数据库CASIA-IrisV3-Interval[16]中每张虹膜图像转换为两张灰度图，即虹膜灰度图像和标准化的掩码灰度图像，再分别将其转换为二进制串。下面就该功能原理进行简要叙述：

1）人眼虹膜图像分割

虹膜数据库CASIA-IrisV3-Interval[16]中包括249个人的虹膜数据，其中可分为395个类（395）只眼睛，一共2639张图像。分割步骤的目的是找到虹膜的精确轮廓，即虹膜的内边界(瞳孔/虹膜)和外边界(虹膜/巩膜)，也就是将人眼虹膜图像分为两个区域：虹膜区域和非虹膜区域。通过分割步骤，我们可以得到一个二进制掩码，其中为1的像素点属于虹膜区域，为0的像素点不属于虹膜区域。此外，分割步骤的结果提供了两个轮廓(瞳孔轮廓和虹膜轮廓)，这将被虹膜区域归一化步骤使用。

2）虹膜区域归一化

归一化步骤基于分割步骤提供的两个轮廓，根据Daugman所提出的rubber-sheet方法[3]将虹膜区域转换成一个大小不变的带状矩形。

3）编码

编码步骤中使用Gabor滤波器对虹膜归一化图像进行滤波，提取虹膜纹理，得到一个虹膜模板。OSIRIS-V4.1中使用了三个复杂的Gabor滤波器对虹膜纹理模板进行处理：通过Gabor滤波器控制阈值来过滤虹膜纹理模板，得到6个二值化图像，但不是所有的图像的所以像素都会在合法用户匹配功能中被考虑到，因为在系统得到人眼虹膜图像前，系统并不能知道图像中的哪些像素是我们需要的，而哪些像素点是需要被系统考虑的则是由分割步骤提取的掩码决定的。结合分割步骤和归一化步骤，编码步骤的最终结果以一个长度合适的二进制串给出，即可视为用户虹膜信息的明文。

### 2.2.2 负数据库生成

用户的虹膜数据被编码为二进制串，这里使用*p*-hidden算法由虹膜编码得到的二进制串生成难解的负数据库。*p*-hidden算法中的正数据库仅包含一个二进制串，即是用户虹膜信息的明文，这个二进制串被称为隐藏串（实际上，为达到保护用户虹膜信息的目的，在得到负数据库后，该二进制串被直接删除）。为了简要起见，由*p*-hidden算法所生成的数据库被称为负数据库。下面结合*p*-hidden算法来说明系统生成负数据库的原理：

*p*-hidden算法由Liu、Luo和Yue在[11]中提出，其可用于生成相对于局部搜索策略难解的（即难以逆转的）*3-NDB*（即由三个参数细粒度控制3种不同类型的记录的分布的负数据库）。*p*-hidden算法的伪代码如图2-3所示。

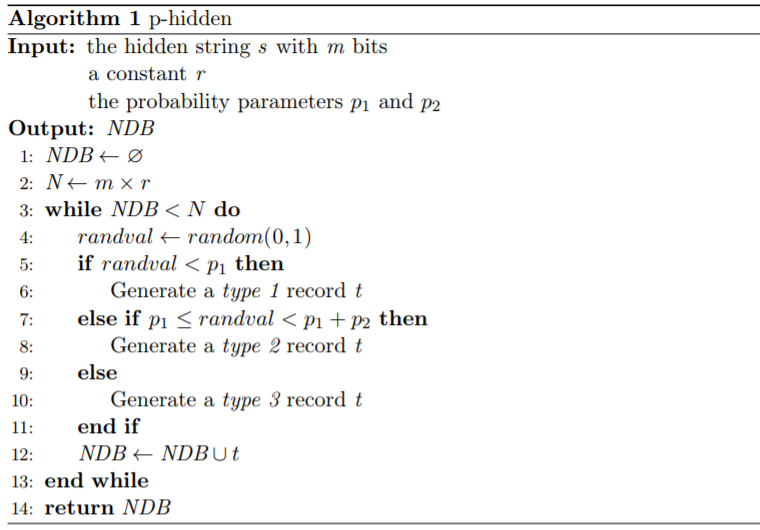


图2-3 *p*-hidden算法

在*p*-hidden算法中，*random*(0,1)用于生成区间[0,1)之间的随机数，隐藏串表示为*s*，即是由虹膜图像编码得到的二进制串，概率参数*p1、p2*，用于控制所生成的负数据库中不同类型的记录的分布，参数*r*用于控制所生成的负数据库中记录的数目，参数*m*保存二进制串*s*的长度。*p*-hidden算法通过迭代生成*N=m×r*条记录，并将这些记录加入对应的负数据库中。例如，*r*的值为3，每一张虹膜图像编码得到的二进制串将会被转成3条长度为*m*的二进制串加入对应的负数据库。而每一次迭代分别以概率*p1、p2*和*p3 = 1-p1-p2*生成一条类型1、类型2、类型3的记录*t*，每条记录都有且仅有3个确定位和*m-3*个非确定位。关于上述三种类型的记录分别定义如下：

类型1：在某一个合法的确定位置上，类型1的记录*t*与隐藏串*s*有不同值。在其他两个合法的确定位置上，类型1的记录*t*与隐藏串*s*有相同值。

类型2：在某两个合法的确定位置上，类型2的记录*t*与隐藏串*s*有不同值。在另外一个合法的确定位置上，类型2的记录*t*与隐藏串*s*有相同值。

类型3：在三个合法的确定位置上，类型3的记录*t*与隐藏串*s*均有不同值。

根据文献[11]，为了生成难解的负数据库（相对于局部搜索策略），参数*p1、p2*应该满足*4p1+2p2>3*且*p1+p2<1*。在参数*r*设置为常数的情况下，算法的计算复杂度为*O(m)*，所生成的负数据库的大小为*O(m)*。

### 2.2.3 合法用户匹配

从上述功能原理的叙述可以知道，用户的虹膜数据被编码为一个负数据库，其中包含多条二进制串。在文献[12]中，Liu给出了*p*-hidden负数据库上的海明距离的估算公式，文献[13]中，Zhao给出另一种更为简单的方法来估算*p*-hidden负数据库上的海明距离。结合上述两种方法，给定任意二进制串*x*和由二进制串*s*所生成的一个负数据库*NDB*，估算*x*和*s*的海明距离的公式为：



其中，*H(NDB,x)*表示*NDB*和*x*之间的海明距离，其可通过如下式子计算：



综上所述，由*p*-hidden算法生成的负数据库上的海明距离是能够被有效的估算的，所以在我们进行合法用户匹配的时候只需要将由用户识别时提供的虹膜图像编码得到的二进制串与系统中存储所有用户虹膜信息负数据库的数据库中的数据计算海明距离，以一定的阈值确定是否为同类虹膜，最终返回匹配结果。

但是*p*-hidden算法生成的一个负数据库包含了*m×r*条负数据记录，若使用文献[11]中所建议参数*r=6.5*，这将导致一个负数据库会存储大约一万条负数据，如果直接使用这些数据进行计算，系统的效率将会十分低下。考虑到用户量增大后，负数据库的数量过大可能导致匹配过程时间过长，为了减小系统的计算开支，提高系统效率和安全性，我们仔细研究了海明距离的计算公式，发现可以对负数据库进行有效的压缩存储而不影响海明距离的计算。在生成负数据库时，只要保留*p*-hidden算法生成的记录的每一个对应位上的*0*和*1*的个数，就能够利用海明距离计算公式进行有效的计算，进而实现有效识别。这样不仅大大提高了计算效率，占用存储空间也大大减小了。此外，压缩后的负数据库变得更加难解，提高了系统的安全性。

负数据库的压缩思路如下图所示：

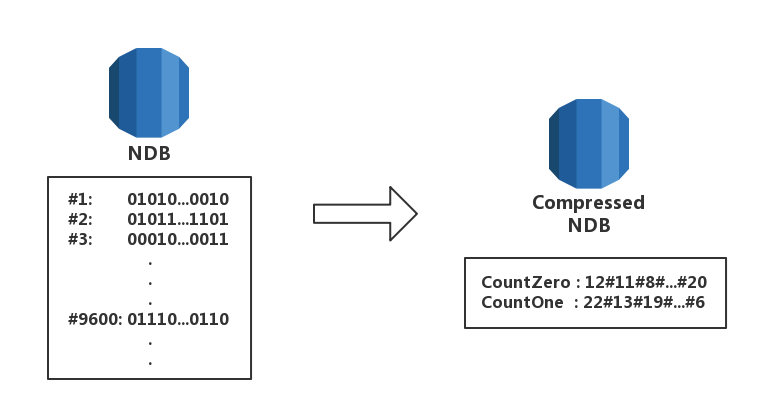


图2-4 负数据库压缩

其中NDB中数据包括了近万条虹膜信息负数据，每一条负数据是一条长度进1600位的二进制串，经过压缩得到的CompressedNDB中只有两条数据，分别记录了*0*和*1*在对应位上的总数。例如，图2-4中生成的压缩负数据库的第1位有12个*0*和22个*1*。

为使用压缩的负数据库进行海明距离计算，给出变形后的海明距离计算公式如下：



其中*xi*表示虹膜信息二进制串*x*中第*i*位的取值，*Zi*表示负数据库*NDB*中所有负数据在第*i*位上*0*的总数，*Oi*表示负数据库*NDB*中所有负数据在第*i*位上*1*的总数。

### 2.2.4 负数据库的安全性说明

在我们的系统中，用户的虹膜信息是以负数据库的形式存储在管理端的数据库中的。相比于传统的方式，由于对负数据库求逆恢复出正数据的信息是*NP*难解的，所以即使管理端的数据库被入侵，用户的虹膜信息也不会因为负数据库被非法获取而泄露。下面将对负数据库的难解性进行说明。

对于一个给定的二进制串*s*，串长为*m*。假定生成负数据库为*NDB={t1,…,tN}*，该负数据库可通过如下步骤转化为一个*SAT*实例*C=c1^…^cN*。

1. *C*中包含*m*个变量*{v1,…,vm}*，第*k*个变量对应上述负数据库*NDB*中每条负数据的第*k*个位置。
2. *NDB*中第*i（i=1,…N）*条记录*ti*可对应于*C*中的第*i*个子句*ci*：



其中*ti,j=\** 表示*NDB*中第*i*条数据的第*j*位与给定串*s*中的第*j*位相同。

对于给定串*s*，将其第*1*位到第*m*位上的值赋予变量*v1,…,vm*，由负数据库的生成算法可以知道，*s*与*NDB*中任意记录都不完全匹配。根据上式，将*sj*赋值给*vj*后，就能满足子句*ci*。因此，对于给定串*s*，能够生成*C*中所有子句，故对于任意给定串*s*都能生成对应于*C*的解。

综上可知，由于负数据库于SAT实例之间具有一一对应的关系，所以对给定负数据库求逆恢复出正数据的信息等价于求解SAT问题，而后者是*NP*难解的，故对给定负数据库求逆恢复出正数据的信息也是*NP*难解的。这就从根本上解决了数据库被非法获取带来的信息泄露问题，因为即使得到虹膜信息的完整负数据库，也没办法恢复出对应的虹膜信息编码。

## 2.3 系统流程

我们所开发的这套嵌入式系统需要搭载到满足硬件要求的嵌入式设备上，并事先完成必要的配置工作。实际使用时的流程可分为两类，一是登记时的信息录入部分，二是识别时的识别匹配部分。下面就这两类流程进行分别叙述：

### 2.2.1 登记

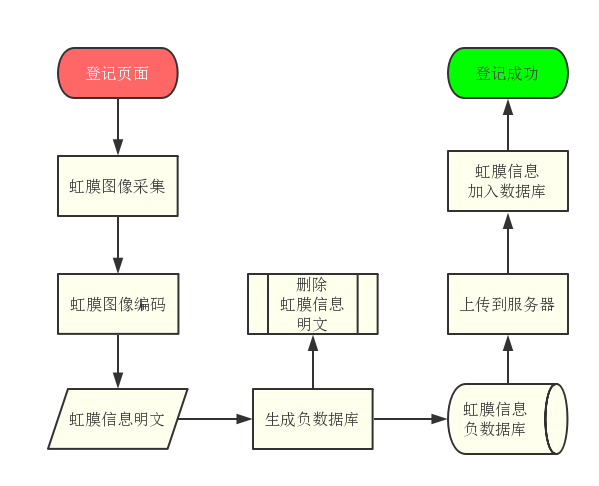


图2‑5 登记流程

登记的流程如图2-5所示，下面对其进行详细的叙述。

1. 虹膜图像采集

用户使用人眼虹膜采集设备（后文叙述中简称为采集设备）进行虹膜图像采集，在一定时间内（取决于采集设备的采集速度）采集一定量的人眼虹膜图像。如果用户在采集设备还未采集到足够数量的人眼虹膜图像前移动眼睛范围超出采集设备的工作范围导致系统采集已采集到的人眼虹膜图像数量不足，则系统不会进行后续的登记流程，而是留在虹膜图像采集界面；正常情况下，系统从采集设备得到足量的人眼虹膜图像，进入登记流程的下一步。

1. 虹膜图像编码

系统将得到的人眼虹膜图像进行分割、归一化和编码等处理，最后对每一张人眼虹膜图像得到一个二进制串，即是虹膜信息的明文。

1. 明文转负数据库

在上文中已经说到，虹膜信息的明文泄露是非常危险的（原因这里不再赘述），所以我们的系统将其转为其负数据库。由明文转负数据库的原理在2.1节中已经有了详细说明，这里就不再赘述。在得到了虹膜信息的负数据库后，系统立即删除虹膜的明文信息（所以系统不存在被攻击而泄露用户虹膜信息明文的问题），同时将虹膜信息的负数据库上传到管理端，登记成功，登记流程结束。

### 2.2.2 识别

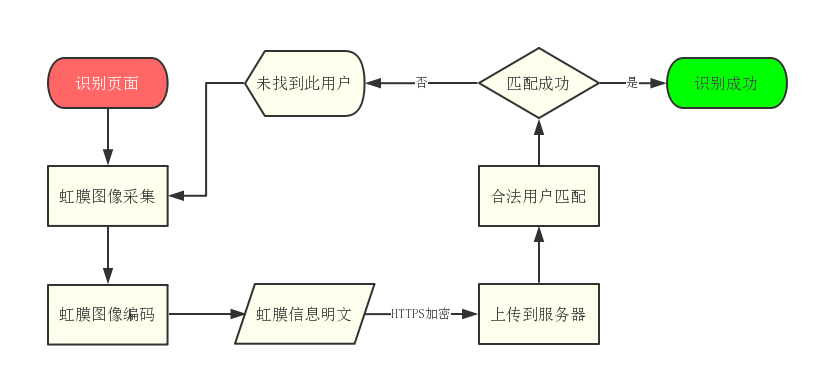


图2-6 识别流程

识别的流程如图2-6所示，很容易可以发现识别的流程与登记的流程非常相似，所以完全相同的部分在下文不再赘述。

1. 虹膜图像采集
2. 虹膜图像编码
3. 明文信息HTTPS加密

对虹膜图像编码后得到虹膜信息二进制串使用HTTPS加密，并传输到管理端。

1. 虹膜信息匹配

管理端通过解密扫描端传入的虹膜加密信息得到待匹配虹膜信息二进制串，并使用2.2.3节中的适用于负数据库的海明距离估算公式，与数据库中的负数据库进行海明距离比对，大于一定阈值视为异类虹膜，反之为同类虹膜。如果没有找到同类虹膜，则说明系统的数据库中并没有该用户的登记信息，提示未找到此用户；正常情况下，在数据库中找到了同类虹膜，则确定识别用户的身份，识别成功，识别流程结束。

## 2.4 功能

1. 虹膜图像编码

在扫描端进行合法用户身份认证和管理端进行合法用户信息添加都需要在第一步将用户的虹膜图像作为输入，使用该功能输出一段包含用户独有生物ID的二进制串。后续的所有功能都将以此为基础。

1. 负数据库生成

为保护用户的独有生物ID——虹膜信息二进制串，需要使用该功能对虹膜信息二进制串进行加密。加密的方式就是生成其对应的*NP*难解的负数据库以代替原有的二进制串，达到保护用户独有生物ID的目的。

1. 用户信息（基本信息和虹膜信息）录入

在管理端，可由管理员进行控制，使用该功能向系统数据库中添加合法用户的虹膜负数据库。

1. 使用基于负数据库的海明距离估算公式进行身份认证

该功能用于合法用户匹配阶段。管理端使用基于负数据库的海明距离估算公式将待认证用户虹膜信息与数据库中所有负数据库进行比对，当海明距离估算结果小于一定阈值（由实验得出），则可认为是合法用户；否则认为是非法用户。

1. 管理端可对连接设备信息进行管理

管理端可以连接到多个扫描端，在管理端可以对任意的扫描端进行管理，包含删除任意扫描端、添加扫描端。对于任意一个扫描端，可在管理端对其设定一个认证等级，对应该扫描端所能访问的范围（具体内容由应用场景决定）。

1. 管理端可对合法用户信息进行管理

在管理端，可由管理员进行控制，对数据库中现有的合法用户进行管理，包含删除任意用户和添加用户信息。对于任意一位用户，可在管理端对其设定一个认证等级，对应该用户所能访问的范围（具体内容由应用场景决定）。

## 2.5 指标

1. 负数据库生成算法计算复杂度

对于给定串*x*，设串长为*|x|*，生成负数据库大小为*|NDB|*。本系统采用*p*-hidden算法，所以在登记阶段，生成负数据库的计算复杂度为*O(r×|x|)*。

1. 合法用户匹配计算复杂度

对于给定串*x*，设串长为*|x|*，生成负数据库大小为*|NDB|*，假设当前管理端的虹膜数据库中已存储的负数据库数目为*|P|*。使用改进的基于负数据库的海明距离估算方法，对于每一次海明距离估算都可以在线性时间内完成，计算复杂度为*O(|x|)*，而对于每一次合法用户匹配，最坏的情况即是非法用户匹配，需要在管理端进行*|P|*次海明距离估算。所以管理端合法用户匹配过程的总的计算复杂度为*O(|x|×|P|)*。

1. 数据库的拒识率FFR和误识率FAR均小于0.3%

误识率（FAR）是指在标准指纹数据库上测试虹膜识别算法时，不同虹膜的匹配分数大于给定阈值，从而被认为是相同虹膜的比例，简单地说就是“把不应该匹配的虹膜当成匹配的虹膜”的比例。 拒识率（FRR）是指在标准指纹数据库上测试虹膜识别算法时，相同虹膜的匹配分数低于给定阈值，从而被认为是不同虹膜的比例，简单地说就是 “把应该相互匹配成功的虹膜当成不能匹配的虹膜”的比例。

1. 不可逆性、可撤销性和不可连接性

对于虹膜模板保护，不可逆性是指攻击者难以由认证数据恢复出原始虹膜信息；可撤销性是指一旦认证数据被泄露，用户或服务器可撤销该认证数据，再生成新的认证数据；不可连接性是指不同应用中的虹膜认证数据之间不可交叉匹配，以避免泄露用户的活动信息。

## 2.6 作品展示

通过本章前五节的介绍，我们已经比较全面的介绍了作品的实现原理、设计思路以及功能指标等相关内容。在这一节中，我们将会按照作品功能逻辑，以作品界面截图加上文字辅助描述的形式对我们的作品“负虹膜识别系统”，做一个作品展示。

### 2.6.1 管理端



图2-7 管理端登录界面

图2-7所展示的是管理端的登陆界面，需要使用管理员账户登录（管理员账户需要后台添加）。在该界面输入用户名和密码，点击“登录”按钮，随即提交用户输入并在管理员用户信息数据库中进行精准匹配。若匹配失败，则提示用户名或密码不符；否则进入管理端主界面。点击“退出”按钮则退出管理端。

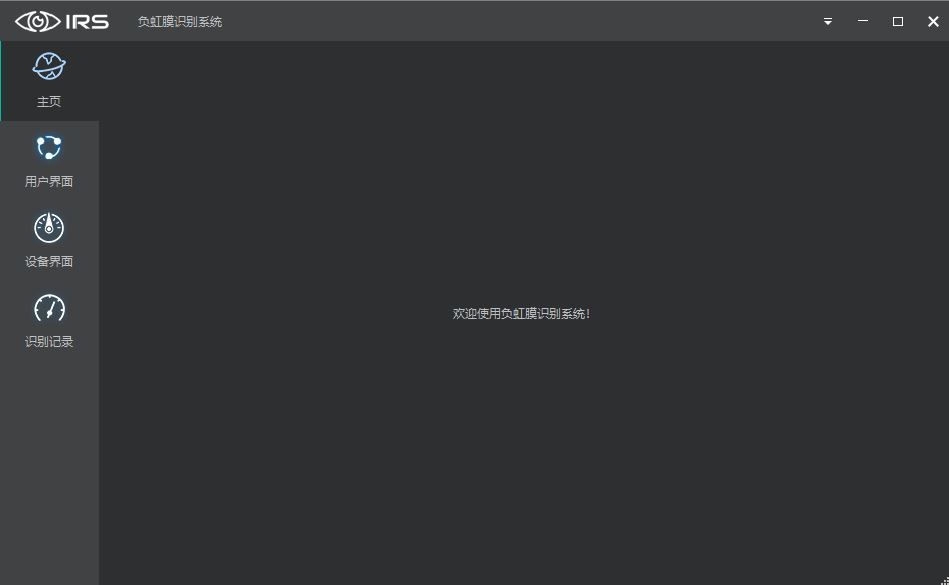


图2-8 管理端主界面

图2-8所展示的是管理端主界面的“主页”。由于现有的作品可以应用到多种实际场景，所以“主页”只显示了欢迎语。

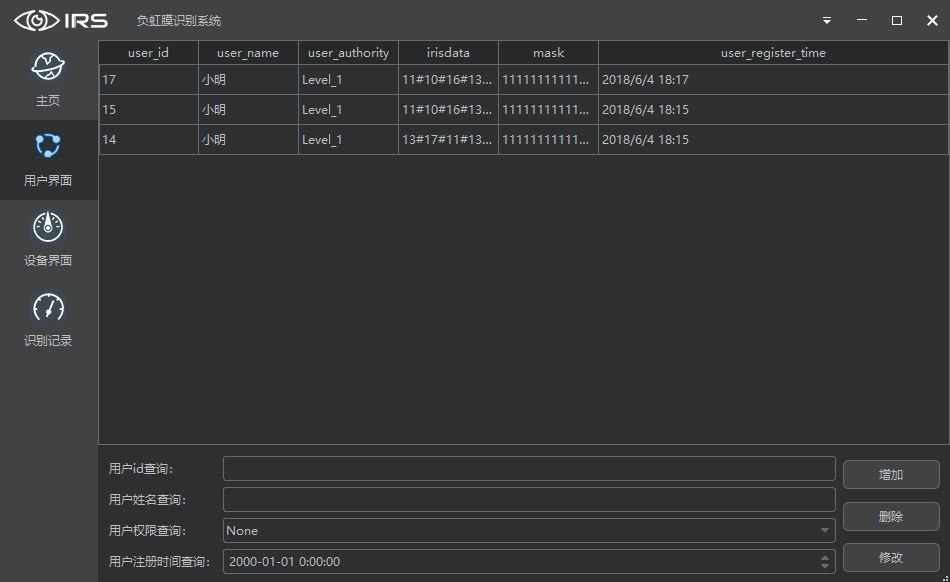


图2-9 管理端主界面

图2-9所展示的是管理端主界面的“用户界面”。该界面由一张用户信息表和下方的功能栏组成。其中，用户信息表所显示的是当前已加入系统的合法用户信息，包括用户名、认证等级和压缩虹膜负数据库等信息，表格内容还可视具体应用场景进行扩充；下方的功能栏则包括了多条件用户查询输入框，以及对合法用户增、删、改的功能按钮。

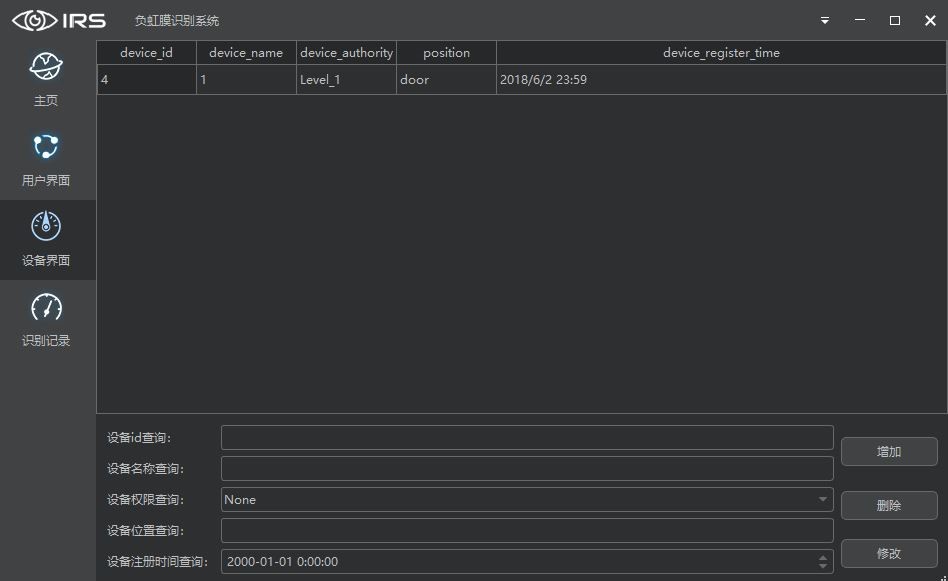


图2-10 管理端主界面

图2-10所展示的是管理端主界面的“设备界面”。该界面与“用户界面”相似，由一张扫描端设备信息表和下方的功能栏组成。其中，设备信息表所显示的是已连接管理端的扫描端信息，包括认证等级和扫描设备ID等信息；下方功能栏则包括了多条件设备查询输入框，以及对扫描端设备增、删、改的功能按钮。

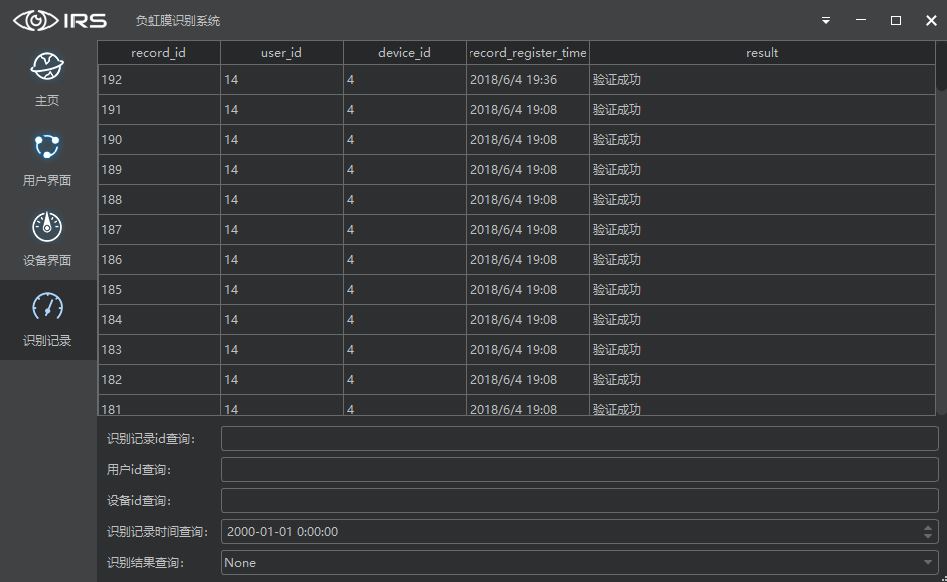


图2-11 扫描端主界面

图2-11所展示的是管理端主界面的“识别记录”。该界面由一张识别记录表和多个查询输入框组成。其中，识别记录表所显示的是所有已连接的扫描端进行识别的记录，每一条记录包含识别动作发生的扫描端设备ID和识别结果等信息；下方的多个输入框则支持多条件记录查询。

### 2.6.2 扫描端



图2-12 扫描端主界面

图2-12所展示的是扫描端主界面。扫描端需嵌入到人眼扫描设备当中，通过接收扫描设备所提供的人眼图像作为输入，将识别结果作为输出反馈给用户。为方便就现有的虹膜数据库CASIA-IrisV3-Interval[16]，现有作品的扫描端中将人眼图像输入环节从扫描设备分离，通过上传图片的形式模拟人眼扫描图像输入；模拟设备ID也需在实验时手动输入。将该作品应用到实际场景中时，上述两个功能都将由扫描设备自动完成。主界面的下方还包括一个“识别”按钮和“退出”按钮，这两个按钮在实际应用中可以考虑由扫描设备结合控制逻辑代码实现自动化控制。

# 第三章 作品测试与分析

## 3.1 测试方案

1. 测试在同一个人录入三次的情况下，系统识别的误识率（FAR）和拒识率（FRR）。
2. 检测负信息是否具有可逆性。

## 3.2 测试环境

|  |  |  |  |
| --- | --- | --- | --- |
| 操作系统 | 相关开发库 | 开发环境 | 数据库 |
| Windows10 | Opencv,qt5 | Qt creator | Mysql5.7 |

表3-1

## 3.3 测试数据

数据集：中科院虹膜数据集CASIA-IrisV3-Interval

数据规模：共计249人次，2639张扫描图像

### 3.3.1 同一个人多次录入测试

|  |  |  |
| --- | --- | --- |
| 同一个人虹膜录入次数 | 拒识率 | 误识率 |
| 1 | 2% | 0.1% |
| 2 | 1.2% | 0.05% |
| 3 | 0.4% | 0 |
| 4 | 0.4% | 0 |
| 5 | 0.4% | 0 |
| 6 | 0.4% | 0 |
| 7 | 0.3% | 0 |
| 8 | 0.2% | 0 |

表3-2

### 3.3.2 模拟数据库入侵测试

可获取的数据库内虹膜数据：



实际对应的测试虹膜原始数据：

0010100110000110011011010010001101111111001100111100110001011111110011101111011011011110011111111111000101011111111000011101111100110011101111111110111100011111111111100011111111111110111111110011001000001110000110001111111011101000111111110011000111111001000100011111110110010111110111100001010001001110001000111101111101000011111110010000101001000100001101101111001100110000111100110011010100011111000000000011101100000111001101000000001111000111000010111110011011111111001100001100111101100111100000111111000010000000

## 3.4 结果分析

从3.3.1节测试中可以看出，在同一个人录入数据小于三条时，拒识率较高，而当其少于三条时，拒识率的提高不是很明显，因此我们最终选定的同一个人录入三次是比较合理的选择。而误识率始终不会大于0.1%，说明系统能够保证足够的安全性，不会出现非认证人员可以进入的情况。

在3.3.2节测试中，数据库存储下来的信息只是统计信息，完全无法恢复出由0和1组成的虹膜原始数据，但是我们系统是可以通过这些信息进行模糊查询的，而即使有人拿到这些数据依然不会造成信息泄露，保证了系统足够的安全性和实用性。

# 第四章 创新性说明

在网络发展十分发达的今天，生物信息识别越来越受市场的青睐。但是，我们在享受便利的同时，也留下了信息泄露的安全隐患。生物信息作为我们但目前虹膜信息保护领域基本的思路都是进行加密防破解，但是这种方式无法从根本上解决这个问题。

因此，我们采用了一种新型的隐私保护技术——信息负表示技术。该技术使用负数据库存储原始数据补集的压缩形式，对其求逆恢复出原始数据是一个*NP*难问题，也就是说，即使数据库信息被窃取，也不会被还原出原始虹膜信息，这保证了系统的安全性，也是我们的重点。相比于传统的加解密算法，信息负表示具有一些独特的性质，它可在不需“解密”的情况下进行一些计算和数据库操作，例如并集、插入和查询等。它可以在保护虹膜数据隐私的同时，支持基于海明距离的相似度估算，实现模糊识别认证。 这些性质使得该技术能很好的用于虹膜模板保护。此外，基于信息负表示的虹膜识别机制具有良好的可扩展性，它可用于支持一些扩展的虹膜识别策略（如移位和掩码策略等），以提升识别的准确性。总体来说，本系统的主要创新点可罗列如下：

1. 利用信息负表示技术和*p*-hidden算法保护虹膜信息，提高了虹膜信息的安全性。
2. 通过负信息的不可逆性，系统即使在数据库信息泄露的情况下也不会泄露个人的虹膜信息。
3. 采用基于信息负表示技术的海明距离估算方法，使模糊匹配的效率大大提高（相对于hash链），从而使得系统的实用性有了保障。
4. 利用海明距离计算公式的特性，通过改进已有的基于负数据库的海明距离计算公式，对负数据库进行压缩。将原有负数据库中的*m×r*条负数据压缩成了两条，大大提高了系统的计算效率。
5. 压缩的负数据库使得系统的安全性得到进一步提升。
6. 系统应用了移位和掩码策略，以提升虹膜识别性能。

# 第五章 总结

通过前文的论述，我们已大致阐明了整个系统的构思、设计、开发与测试过程，这里再对其做一个简单的总结。整个系统最主要的工作流程如图2-4和2-5所示，分别是注册阶段和识别阶段。识别阶段的任务是通过嵌入式设备，如智能门锁，实现实时虹膜扫描，然后将扫描数据上传到局域网管理端，通过后台服务进行虹膜负数据库的查询与匹配，最后将识别结果返回给前端嵌入式设备，根据结果决定访问者是否具有进入权限。由于系统是一个分布式系统，分为扫描端和服务端，我们将两端的功能分别部署，这使得系统的适用范围被大大地扩展了。

本系统所使用的核心技术是信息负表示技术、负数据库信息匹配技术和虹膜识别技术，信息负表示技术是由生物免疫系统启发得到的一种新型的隐私保护和安全技术。由其生成的负数据库本身具有一些特性，使其可用于构建一些安全协议和体制。现如今，虽然信息负表示技术已被应用于一些场景中，但是，负数据库的特性还未得到充分的挖掘，我们实现的负虹膜识别系统只是其适用场景中的一小部分。

同时，本系统的应用范围十分广泛，凡是需要较高安全性的认证系统都可以改进成为负认证系统，例如存储公司重要财产的保险箱的身份识别和认证系统、国防以及其他涉密领域。另外负虹膜识别系统可以支持多生物特征信息识别，以提升识别精确度和安全性。

我们对于本系统的未来充满信心，并且为系统预留了很多可扩展项，例如，可将本系统扩展为多模式生物特征负信息识别系统，结合诸如指纹、面相等识别技术，大大增强系统的安全性以适用相关应用场景。本系统中主要使用改进的海明距离估算作为相似度计算方法，其实还可进一步挖掘负数据库的性质，以支持其他的相似度计算方法，如欧几里得距离等，以扩展负虹膜识别的适用范围。

最后，我们要由衷地感谢我们的指导老师，感谢他对我们开发过程无私的指导和帮助；还要感谢我们小组中的每一位成员为本系统的开发所付出的努力与牺牲。

# 参考文献

[1] F.Esponda, E.S.Ackley, S.Forrest, and P.Helman, "Online negative Database, " in *Proceedings of the Third International Conference on Artificial Immune Systems,* Catania, Sicily, Italy, 2004, pp. 175-188.

[2] F.Esponda, S.Forrest, and P.Helman,"Enhancing Privacy Negative Representations of Data," University of New Mexico TR-CS-2004-18,2004.

[3] J.G.Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vo1.15, no. 11, pp. 1148-1161, 1993.

[4] Lim S, Lee K, Byeon O, and T.Kim, "Efficient Iris Recognition Through Improvement of Feature Vector and Classifier, " *ETRI Journal*, vol. 23, no.2, pp.1-2, 2001.

[5] G. I. Davida, Y. Frankel, and B. J. Matt, "On Enabling Secure Applications Off-line Biometric Identification," in *Proceedings of the 1998 IEEE Symposium on Security and Privacy*, 1998, pp. 148-157.

[6] G. I. Davida, Y.Frankel, and B. J. Matt, "On the Relation of Error Correction and Cryptography to an Off-line Biometric Based Identification Scheme," in *Proceedings of the 1999 Workshop on Coding and Cryptography (WCC99)*, 1999, pp. 129-138.

[7] F. Hao, R. Anderson, and J. Daugman, "Combining Crypto with Biometrics Effectively," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1081-1088, 2006.

[8] A. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme," in *Proceedings of the 6th ACM Conference on Computer and Communications Security*, 1999, pp. 28-36.

[9] Sutra G, Dorizzi B, Garcia-Salicetti S, et al. *A Biometric Reference System for Iris, OSIRIS version 4.1(2013)*. Telecom Sud Paris, France, Tech. Rep, 2012.

[10] R. Liu, "Research on the Negative Database Generation Algorithms and Applications," PhD Dissertation, Department of Computer Science and Technology, University of Science and Technology of China, 2013 (In Chinese).

[11] R. Liu, W. Luo, and L.Yue, "The *p*-hidden Algorithm: Hiding Single Databases More Deeply," *International Journal of Immune Computation* , vol. 2, no. 1, pp. 43-55, March 2014.

[12] R. Liu, W. Luo, and L. Yue, "C1assifying and Clustering in Negative Databases," *Frontiers of Computer Science*, vol. 7, no. 6, pp. 864-874, December 2013.

[13] 赵冬冬, "信息负表示的若干应用方案研究", PhD Dissertation, Department of Computer Science and Technology, University of Science and Technology of China, May,2016.

[14] C.-1. Tisse, L. Martin, L. Tomes, and M. Robert, "Person Identification Technique Using Human Iris Recognition," in *Proceedings of the 2002 Vision Interface*, 2002, pp. 294-299.

[15] P. Thoonsaengngam, K. Horapong, S. Thaini-mit, and V. Areekul, "Efficient Iris Using Adaptive Quotient Thresholding," in *Proceedings of the 2006 International Conference on Biometrics*, 2006, pp. 472-478.

[16] *CASIA-IrisV3 [Online]*, 2005. Available: http://biometircs.idealtest.org/.