第 34 卷 第 5 期 计 算 机 工 程 2008 年 3 月
Vol.34 No.5 Computer Engineering March 2008
·人工智能及识别技术· 文章编号：1000—3428(2008)05—0182—03 文献标识码：A 中图分类号：TP391

基于比特流的虹膜特征数据的隐藏算法

叶学义
(杭州电子科技大学通信工程学院模式识别与信息安全实验室，杭州 310018)

摘 要：对生物特征数据的攻击是生物特征识别自身安全的主要威胁。为了提高虹膜特征数据的安全性，根据现有主要的虹膜识别方法中特征模板的数据特性和基于汉明距的比对方法，提出一种基于比特流的将虹膜特征模板数据嵌入人脸图像的数据隐藏算法。实验结果表明，该算法具有较强的隐蔽性，隐藏算法本身误码率为零，计算效率高，不会影响虹膜识别技术本身的性能，能够有效保护特征模板数据，增强虹膜识别系统自身的安全性。

关键词：虹膜识别；比特流；数据隐藏

Bit-stream Based Iris Features Data Hidden Method

YE Xue-yi
(Lab of Pattern Recognition & Information Security, School of Communications Engineering, Hangzhou Dianzi University, Hangzhou 310018)

【Abstract】 Attacks pointing to biometric data are the basic menace of biometrics self-security. For improving the security of iris features data, this paper proposes a bit stream based data hidden approach by embedding iris features data into a face image which is applicable to the dominant method of present iris recognition technologies. While having the high computation efficiency and the zero-decoding-errorrate, the method embedding a key biometric feature into another biometric feature for transmission in a biometrics security net can deceive attackers better, and also is not repressive to the excellent performance of the iris recognition. Experiments show that the data hidden approach can effectively protect iris features data and enhance the iris recognition system self-security.

【Key words】 iris recognition; bit stream; data hidden

1 概述

生物特征识别技术因为生物特征(指纹、虹膜、脸像等)自身固有的特性使得它超越和替代传统的身份识别手段成为可能，并且该技术已经在一些国家的某些应用领域被推广和使用[1]。但是正因为这些固有的特性，生物特征数据和拥有者直接相关，具备独有性及不可更改性，使得生物特征识别技术的研究不得不关注生物特征识别自身的安全性。如果一个注册用户的生物特征数据被非法窃取，那么可能引起的问题和解决的难度要远大于一个传统身份识别技术的使用者丢失了他的IC卡或者密码。因此，生物特征识别技术的有效使用是建立在这样一个基础上，即进入生物特征识别网络系统的生物特征数据只能来自于合法的拥有者[2]。

生物特征识别网络系统可能被攻击的情况，如图 1 所示。

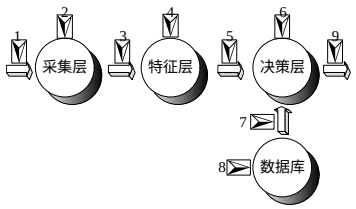


图 1 生物特征识别网络系统的结构示意及可能被攻击环节

在图 1 中，箭头 1 是指用伪造的生物特征来欺骗识别系统；箭头 9 是指篡改决策结果或者接管控制设备；而其余的无论是攻击系统的传输通道(箭头 3、5、7)，还是攻击系统的处理单元和存储单元(箭头 2、4、6、8)，其目的都是窃取和篡改合法的生物特征数据。所以，生物特征识别自身的安全性在很大程度上依赖于生物特征数据的安全性。

现有对生物特征数据安全性构成的威胁有很多方式，从数据本身来说主要可以分为以下两类：直接更改已注册身份的特征模板(替换)；根据获得的特征模板重构生物特征图像[3]，然后重新输入系统，获得已注册权限(伪造)。因为在生物特征识别技术的实际推广和应用中的生物特征隐私权问题，所以对于生物特征原始数据的保护尤为重视。一般在实际方案中原始数据往往只在中央数据库中存储或者根本不保留，也就是说在应用系统中存储和传输的通常是生物特征的模板数据。由此可见，生物特征模板数据的安全对于生物特征识别系统自身的安全性影响至关重要。

实现数据安全的方法主要可以分为 3 类：加密，信息隐藏[4]和数字水印。加密是指将关键信息利用密钥变成杂乱无章的信息或者对于非授权方没有意义的信息；信息隐藏是指将关键信息隐藏在宿主信息中，用于隐秘传输；数字水印是一种在加密和信息隐藏的基础上发展起来的数据嵌入技术，通常将商标或者公司的信息嵌入信息母体[5]。目前关于生物特征数据安全研究方面的文献并不多见，其中具有代表性的例如：文献[6]提出的一种基于局部块平均数字水印的方法，可以检测脸像和指纹是否被篡改；文献[7]提出的一种改变离散小波变换系数来实现数据隐藏的方法，应用于经过WSQ (Wavelet Scalar Quantization)压缩的指纹图像；文献[8]提出了一种根据密钥在时域将一幅水印图像嵌入一幅指纹图像，如

作者简介：叶学义(1973－)，男，博士，主研方向：图像处理，模式识别，智能信息处理，信息安全和计算机视觉

收稿日期：2007-03-30 **E-mail：**xueyiye@hdu.edu.cn

果该指纹图像在任何局部被更改，验证时将被发觉，称之为碎水印方法；文献[9]提出在嵌入水印时用指纹图像的梯度方向信息替代所有的非特征信息，随后又提出了一种改进方法，以保留图像的奇异点，使得嵌入水印的图像在进行指纹分类时不受影响。但迄今为止，还没有关于虹膜特征数据的隐藏和保护方面研究的文献资料，本文通过对目前生物特征数据的安全和标准，以及数据隐藏技术的研究等方面[10-11]的分析，提出了一种基于比特流的虹膜特征数据的隐藏方法，将虹膜特征模板的数据嵌入一幅人脸图像中，达到数据隐藏的目的，提高虹膜特征数据的安全性。实验表明，该方法取得了良好的效果。

2 基于数字水印的生物特征数据隐藏方法

基于幅度调制的数字水印方法是一种最常用的数字水印方法[9,11]，利用数字水印的方法来实现生物特征数据隐藏，一般将要隐藏的数据作为数字水印，而将被隐藏的数据作为宿主（通常是一幅图像）。隐藏的过程称为编码，取出隐藏数据的过程称为解码。简要描述如下[4]：

$$P_{WM}(i,j)=P(i,j)+(2s-1)\cdot P_{AV}(i,j)\cdot q\cdot (1+\frac{P_{SD}(i,j)}{A})\cdot(1+\frac{P_{GM}(i,j)}{B})\cdot\beta(i,j) \tag{1}$$

其中，$P(i,j)$ 是宿主图像中 (i,j) 位置的像素值；$P_{WM}(i,j)$ 是该位置编码之后的值；s 是要隐藏的二进制数据；$P_{AV}(i,j)$ 是像素 (i,j) 邻域像素的平均值；$P_{SD}(i,j)$ 是该邻域像素值的方差；$P_{GM}(i,j)$ 是该位置的梯度幅值；q 是水印强度，也就是水印对宿主数据的改变程度；A 和 B 分别表示 $P_{SD}(i,j)$ 和 $P_{GM}(i,j)$ 的权重；$\beta(i,j)$ 是调制指示符，调制细节和详细效果见文献[11]。

在解码时，根据编码时的密钥，从宿主图像中发现水印的位置开始，对每个经过水印处理的像素 $P_{WM}(i,j)$ 按式(2)计算：

$$\hat{P}(i,j)=\frac{1}{8}(\sum_{k=-2}^{2}P_{WM}(i+k,j)+\sum_{k=-2}^{2}P_{WM}(i,j+k)-2P_{WM}(i,j)) \tag{2}$$

式(3)计算的是在以 (i,j) 为中心的 5×5 十字邻域内的线性估计值，然后计算该估计值和水印像素的差值，如式(3)：

$$\delta=P_{WM}(i,j)-\hat{P}(i,j) \tag{3}$$

因为这种数字水印的嵌入方法属于冗余嵌入，其目的是为了增加解码的正确率，所以在计算 δ 时，先根据式(2)和式(3)计算嵌入了相同水印位的所有宿主图像的不同像素的多个 δ 值，再求平均 $\bar{\delta}$，以该值来作为判定解码时该水印位的值 $\hat{s}$，如式(4)所示：

$$\hat{s}=\begin{cases}1 & \text{if} \quad \bar{\delta}>\dfrac{\bar{\delta}_{R0}+\bar{\delta}_{R1}}{2}\\ 0 & \text{else}\end{cases} \tag{4}$$

其中，$\bar{\delta}_{R0}$ 和 $\bar{\delta}_{R1}$ 是解码时的判断阈值；根据上式从宿主图像中取出数字水印，存在一定误解码率。

从以上的编、解码过程可以看出这种方法的计算量比较大。对于一个 K bit 位的待隐藏数据，若宿主图像的大小为 S，冗余度 $\alpha=S/K$，通常 $\alpha>1$；那么该方法嵌入数据的计算量不小于：45 $K\alpha$ 次加法和 30 $K\alpha$ 次乘法；取出数据的计算量不小于：$\alpha(10S+11K)$ 次加法和 2 $\alpha(S+K)$ 次乘法；当随机数产生和参考码表的计算忽略不计时，总的计算量 $\alpha(10\ S+56\ K)$ 次加法和 2 $\alpha(S+16\ K)$ 次乘法。

3 基于比特流的虹膜特征数据的隐藏算法

根据上节的论述，这种基于数字水印的生物特征数据隐藏的方法存在 3 个需要改进的方面：(1)要求冗余度 $\alpha>1$；(2)误解码率不等于 0；(3)较大的计算量。如果采用这种方法来实现虹膜特征模板数据的隐藏，算法本身会在两方面带来明显不利的影响：(1)误解码率不等于 0，会增加识别的等错误率(EER)；(2)计算量较大，在执行 $1:N$ 搜索识别和 $m:N$ 列表搜索识别时，会直接影响搜索效率，尤其是当N较大时，影响更加明显。目前虹膜识别的一个出色的性能就是极高的搜索效率[12]。另外，如果通过增强水印强度来降低误码率，将会减弱数据隐藏的效果。

生物特征数据隐藏算法的研究不仅要能提高数据的安全性，而且应当不对原有的生物特征识别性能产生消极影响。由此，本文提出了一种基于比特流的虹膜特征模板数据的隐藏算法：该方法以人脸图像为宿主，具有更好的隐蔽性；算法本身的误解码率为 0，而且计算量很小；对冗余度没有要求，即冗余度 α 可以大于 1 也可以等于 1。算法的详细描述如下：

设虹膜特征模板为 $\boldsymbol{I}(k,l)$，是一个矩阵 $k\times l$ 表示的二进制序列，矩阵中每个元素为一个 bit 数；人脸图像为 $\boldsymbol{F}(m,n)$，表示一个 $m\times n$ 矩阵，矩阵中每个元素为一个字节表示的像素；在数据隐藏时以 $\boldsymbol{F}(m,n)$ 为宿主图像，而将 $\boldsymbol{I}(k,l)$ 作为嵌入的数字水印；要求 $(m\times n)\geq(k\times l)$（即满足 $\alpha\geq1$）。需要说明的是：(k,l) 表示 $\boldsymbol{I}(k,l)$ 中一个元素的行和列的坐标，$\boldsymbol{I}(k,l)$ 为二进制码流，$k\in[1,K],l\in[1,L]$，K 和 L 分别是矩阵 $\boldsymbol{I}(k,l)$ 的行数和列数的最大值；而 (m,n) 表示 $\boldsymbol{F}(m,n)$ 中一个元素的行和列的坐标，$\boldsymbol{F}(m,n)$ 是灰度图像（灰度级 0~255），$m\in[1,M],n\in[1,N]$，M 和 N 分别是矩阵 $\boldsymbol{F}(m,n)$ 的行数和列数的最大值。虹膜特征模板如图 2 所示。

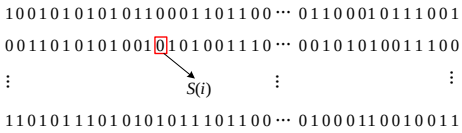


图 2 虹膜特征模板的二进制码流示意图

作为嵌入数据的虹膜特征模板 $\boldsymbol{I}(k,l)$ 的起始位可以从任意位开始，只要在密钥中标记起始位的位置，处理时将图 2 的码流首尾相接，然后依次隐藏所有的码值。如图 2 中所示的起始位选为方框套住的一个 bit，它的值记为 $S(i)$，$S(i)\in\{0,1\}$，i 表示在特征模板中的位置，且 $i=(k-1)\cdot L+l$；同理，作为宿主的人脸图像 $\boldsymbol{F}(m,n)$ 的初始像素也可以任意选取，也要在密钥中标记它的位置，记为 $P(j)$，如图 3 中所示的白色方框套住的那个像素，$j=(m-1)\cdot N+n$，对于灰度图像（灰度级 0~255）用一个字节来表示 $P(j)$，如图中箭头所指，框内分别是字节的每个 bit 位，下方分别是从 0~7 的位序。然后依次嵌入，处理时将 $\boldsymbol{F}(m,n)$ 的像素首尾续接，直至处理完所有的水印位。$\boldsymbol{F}(m,n)$ 可以是宿主图像本身，也可以是整个宿主图像的一部分。

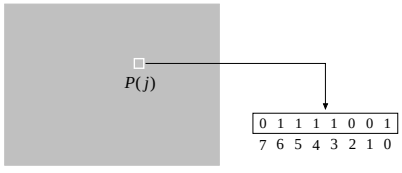


图 3 宿主图像 $F(m,n)$的示意图

如图 2 中右边的字节展开示意，从右到左表示从字节的低位到高位，分别用 $P^{(0)}(j),P^{(1)}(j),\cdots,P^{(7)}(j)$ ；$P^{(0)}(j)$ 作为标志位，从其余 7 位中可以任选一位作为参考位。例如选择 $P^{(1)}(j)$ 作为参考位，水印嵌入如式(5)：

$$P_{WM}^{(0)}(j)=\begin{cases}0, & \text{if } S(i)=P^{(1)}(j)\\ 1, & \text{if } S(i)\neq P^{(1)}(j)\end{cases} \tag{5}$$

$\boldsymbol{F}(m,n)$ 的像素 $P(j)$ 嵌入水印后记为 $P_{WM}(j)$，按照图 3 所示，即表示 $P_{WM}(j)$ 的一个字节中，除了 $P^{(0)}(j)$ 变成 $P_{WM}^{(0)}(j)$，其余位和 $P(j)$ 保持一致；然后依次取虹膜特征模板 $\boldsymbol{I}(k,l)$ 的下一个比特位作为水印，顺序嵌入 $\boldsymbol{F}(m,n)$ 的像素，直至每一个比特位都嵌入到一个像素中一次，数据隐藏处理结束。

取出水印时，也就是将虹膜特征模板 $\boldsymbol{I}(k,l)$ 从作为宿主图像 $\boldsymbol{F}(m,n)$ 中取出，根据隐藏时的密钥，得到 $\boldsymbol{F}(m,n)$ 初始嵌入位置的字节，记为 $P_{WM}(j)$，和虹膜特征模板 $\boldsymbol{I}(k,l)$ 的初始选取位置 $i,\ i\in[1,K\times L]$，解码处理如式(6)：

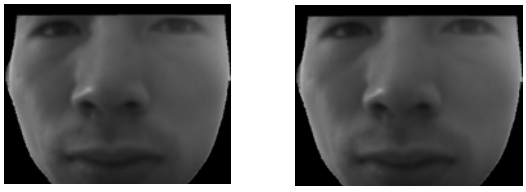
$$S(i)=\begin{cases}P_{WM}^{1}(j) & \text{if } P_{WM}^{0}(j)=0\\ \overline{P_{WM}^{1}(j)} & \text{if } P_{WM}^{0}(j)=1\end{cases} \tag{6}$$

其中，$\overline{P_{WM}^{1}(j)}$ 表示对 $P_{WM}^{1}(j)$ 取反，依照式(6)依次取出隐藏的 $S(i)$，放入确定的虹膜特征模板 $\boldsymbol{I}(k,l)$ 的位置，直至获得所有的模板数据位，完成解码。

上述的基于比特流的数据隐藏算法，直接将虹膜的特征模板隐藏到人脸图像中；且因为实际的计算过程无论是编码还是解码都仅仅是对位的比较和取反，操作次数由虹膜特征模板的码流长度决定；所以计算机来执行这样的操作效率很高。现有的虹膜识别算法和系统基本上是通过计算虹膜纹理编码的汉明距来完成比较和识别，所以对于利用上述数据隐藏算法处理的宿主图像，可以直接和数据库中存储的虹膜特征模板进行比对。只要根据密钥找到宿主图像初始位像素，比较确定的比特位，然后顺序操作即可完成，将解码和比较处理合并，进一步提高了计算效率。另外，从编码和解码的角度来看，算法本身不会造成误码。

4　实验结果和分析

本文的数据隐藏算法，对于作为宿主的人脸图像并不是无损的。但是因为在数据嵌入时仅仅改变了图像中被嵌入水印的像素字节的最低位，对灰度值的影响是 1，而且改变的概率仅仅是 0.25。这样即使在最极端的情况下，例如某个像素的灰度值加 1，而所有相邻的像素都减 1，从视觉效果上来说，基本上没有什么变化。如图 4 所示，随机抽取了一幅人脸图像中一半的像素进行上述极端情况下的处理后，所得到的结果和原始图像进行比较，并不能通过观察判断宿主图像是不是进行了数据隐藏的处理。



(a)隐藏前　　(b)隐藏后

图 4　宿主图像在数据隐藏前后的比较

这种基于比特流的虹膜特征数据隐藏算法对增强虹膜识别的数据安全是具有现实意义的。因为即使经过隐藏的生物特征数据被截获，首先根据数据本身不能确定是否进行了数据隐藏；其次即使确定经过了隐藏处理，如果没有密钥，想得到正确的虹膜特征模板的数据是相当困难的。所以该算法能够在不影响虹膜识别的搜索效率的前提下，有效地保证了虹膜特征模板的安全，从而增强了身份识别系统自身的安全性。并且文献[13]表明作为宿主的人脸图像在经过该数据隐藏算法处理之后，对于目前的脸相识别方法，可以直接进行识别处理，无须进行解码恢复成原始的人脸图像。

另外，为了评估该算法对虹膜识别本身的性能产生的影响，本文进行了如下实验：实验中采用的数据库[14]包括 1 200 幅虹膜样本图像(共 30 人每人 40 个图像样本)，根据文献[13]的方法得到相应的虹膜特征模板，再利用本文提出的方法实现虹膜特征模板的数据隐藏，得到一个对应的同样大小的嵌入了虹膜特征模板的新的人脸图像库；然后分别对原虹膜特征模板库和得到的新库，利用汉明距进行了比对实验，根据比对结果检测处理前后虹膜识别的性能的变化。处理前后识别性能的DET（修正的ROC曲线）曲线如图 5 所示，实验中类内比较次数为：$30\cdot C_{40}^{2}=23\ 400$，类间的比较次数为：$C_{1\ 200}^{2}-30\cdot C_{40}^{2}=696\ 000$。结果表明这种数据隐藏算法不会影响基于汉明距的虹膜识别方法的性能。

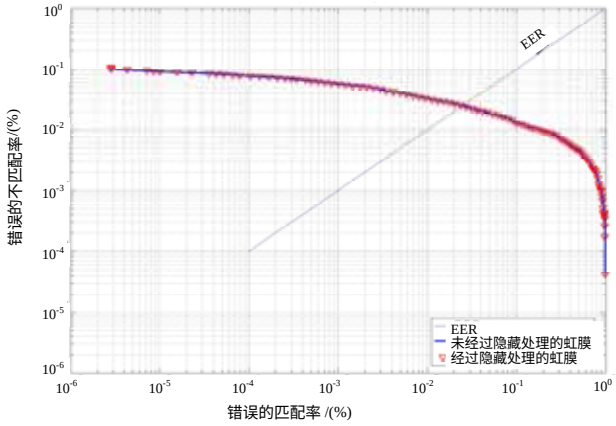


图 5　处理前后虹膜识别性能的 DET 曲线

5　结束语

本文对生物特征识别系统自身的安全性进行分析，并讨论了生物特征识别系统可能被攻击的途径，认为对生物特征数据的攻击是当前生物特征识别自身安全的主要威胁之一。基于此，针对现有主要的虹膜识别的特征模板和基于汉明距的比对方法的特性，提出了一种基于比特流的将虹膜特征模板数据嵌入人脸图像的数据隐藏算法。

参考文献

[1] Higgins P T. An Introduction of Biometrics[C]//Proceedings of Biometrics Consortium Conference. Arlington, VA, USA: [s. n.], 2005.

[2] Schneier B. The Uses and Abuses of Biometrics[J]. Comm. ACM, 1999, 42(8): 136.

[3] Edler A. Can Sample Images Be Regenerated form Biometric Templates[C]//Proceedings of Biometrics Consortium Conference. Arlington, VA, USA: [s. n.], 2003.

[4] Jain A K, Uludag U. Hiding Biometric Data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(11).

[5] Huartung F, Kutter M. Multimedia Watermarking Techniques[J]. Proc. of the IEEE, 1999, 87(7): 1079-1107.

（下转第 187 页）