(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101253726 B
(45)授权公告日 2013.02.06

(21)申请号 200680031989.6

(22)申请日 2006.08.21

(30)优先权数据
11/218,261 2005.09.01 US

(85)PCT申请进入国家阶段日
2008.02.29

(86)PCT申请的申请数据
PCT/JP2006/316780 2006.08.21

(87)PCT申请的公布数据
WO2007/029529 EN 2007.03.15

(73)专利权人 三菱电机株式会社
地址 日本东京

(72)发明人 埃明·马丁尼安 安东尼·韦特罗
谢尔盖耶·M·叶哈宁
乔纳森·S·叶迪达

(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 李辉

(51)Int.Cl.
*H04L 9/32*(2006.01)
*H04L 9/08*(2006.01)

(56)对比文件
CN 1179658 A,1998.04.22, 说明书第3页第3行至第4页第28行.
CN 1427983 A,2003.07.02, 全文.
US 20040139332 A1,2004.07.15, 全文.
CN 1281608 A,2001.01.24, 全文.

审查员 李博

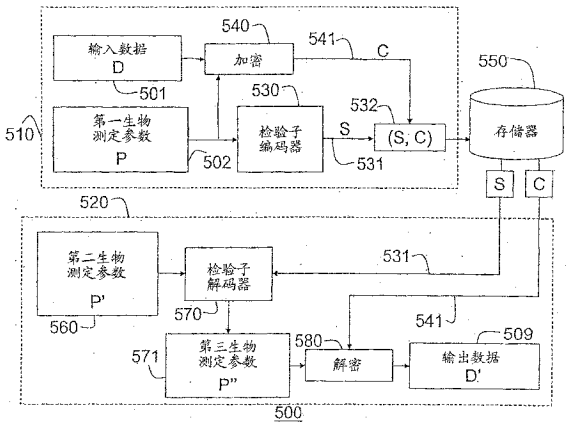
权利要求书 2 页 说明书 14 页 附图 12 页

(54)发明名称

在计算机可读介质中存储数据的计算机实施的方法

(57)摘要

从用户获取第一生物测定参数。根据所述生物测定参数对输入数据进行加密,以产生密文。利用检验子编码器对所述生物测定参数进行编码,以产生检验子码。将所述密文和所述检验子码彼此关联,并存储在计算机可读介质中,以使得只有同一用户能随后对所述密文进行解密。




CN 101253726 B

1. 一种在计算机可读介质中存储数据的计算机实施的方法，该方法包括以下步骤：

从第一用户获取第一生物测定数据；

根据所述第一生物测定数据生成第一生物测定参数；

利用所述第一生物测定参数对所述第一生物测定数据进行加密，以产生密文；

利用检验子编码器对所述第一生物测定参数进行编码，以产生检验子码；

将所述密文和所述检验子码相关联；以及

将所述密文和所述检验子码存储在计算机可读介质中。

2. 根据权利要求 1 所述的方法，该方法还包括以下步骤：

从第二用户获取第二生物测定数据；

根据所述第二生物测定数据生成第二生物测定参数；

利用检验子解码器和所述第二生物测定参数对所述第一生物测定参数进行解码，以产生第三生物测定参数；以及

利用所述第三生物测定参数对所述密文进行解密，以产生第三生物测定数据。

3. 根据权利要求 2 所述的方法，该方法还包括以下步骤：

比较所述第一生物测定数据和所述第三生物测定数据；以及

仅在所述第一和第二生物测定数据相同时才允许访问功能，否则拒绝访问。

4. 一种在计算机可读介质中存储数据的计算机实施的方法，该方法包括以下步骤：

从第一用户获取第一生物测定参数；

根据所述第一生物测定参数对输入数据进行加密，以产生密文；

利用检验子编码器对所述第一生物测定参数进行编码，以产生检验子码；

将所述密文和所述检验子码相关联；以及

将所述密文和所述检验子码存储在计算机可读介质中。

5. 根据权利要求 4 所述的方法，该方法还包括以下步骤：

从第二用户获取第二生物测定参数；

利用检验子解码器和所述第二生物测定参数对所述第一生物测定参数进行解码，以产生第三生物测定参数；以及

利用所述第三生物测定参数对所述密文进行解密，以产生输出数据。

6. 根据权利要求 5 所述的方法，其中，仅在所述第一用户与所述第二用户相同时，所述输入数据和所述输出数据才完全相同。

7. 根据权利要求 4 所述的方法，其中，所述输入数据是生物测定数据。

8. 根据权利要求 7 所述的方法，该方法还包括以下步骤：

从所述输入数据提取第一特征向量；

从第二用户获取生物测定数据，并从该生物测定数据提取第二特征向量；

比较所述第一特征向量和所述第二特征向量；以及

仅当所述第一和第二特征向量相同时才允许访问功能，否则拒绝访问。

9. 根据权利要求 8 所述的方法，其中，所述第一和第二生物测定参数分别与所述第一和第二特征向量不同。

10. 根据权利要求 7 所述的方法，该方法还包括以下步骤：

从所述输入数据提取完全特征向量；

根据所述完全特征向量构造误差直方图；

利用所述直方图将所述完全特征向量简化为检验子特征向量；

测量所述检验子特征向量的不同系数之间的相关性；以及

应用密度演化来设计用于所述检验子特征向量的检验子码。

11. 根据权利要求 10 所述的方法，其中，所述完全特征向量包括仅存储整数的完全硬特征向量以及存储整数和实数的完全软特征向量。

12. 根据权利要求 4 所述的方法，其中，所述检验子解码器使用置信传播网络。

13. 根据权利要求 12 所述的方法，其中，所述置信传播网络包括检验节点和变量节点，以及每对变量节点之间的相关性节点。

14. 一种在计算机可读介质中存储数据的计算机实施的方法，该方法包括以下步骤：

从用户获取生物测定数据；

根据所述生物测定数据生成加密密钥；

根据所述加密密钥对数据进行加密，以产生密文；

将所述加密密钥编码为编码后的密钥；以及

将所述编码后的密钥与所述密文关联地存储在计算机可读介质中。

15. 根据权利要求 14 所述的方法，该方法还包括以下步骤：

随后从所述用户重新获取生物测定数据；

根据所述重新获取的生物测定数据生成解码密钥；

利用所述解码密钥对所述编码后的密钥进行解码；以及

仅当所述解码密钥与所述加密密钥匹配时，才使用所述解码密钥对所述密文进行解密。

# 在计算机可读介质中存储数据的计算机实施的方法

## 技术领域

[0001] 本发明总体上涉及密码学领域，更具体地，涉及存储用于用户认证和数据加密的生物测定 (biometric) 参数。

## 背景技术

[0002] 传统的基于密码的安全系统

[0003] 传统的基于密码的安全系统通常包括两个阶段。具体地，在登记阶段，用户选择存储在诸如服务器的认证装置上的密码。在认证阶段为获得对资源或数据的访问，用户输入他们的密码，针对密码的存储版本检验所输入的密码。如果密码被存储为明文 (plain txt)，则获得该系统的访问的攻击者能够获取每个密码。因此，即便是个别的成功攻击都会危及整个系统的安全。

[0004] 如图 1 所示，传统的基于密码的安全系统 100 在登记阶段 10 将加密的 110 密码 101 存储 115 在密码数据库 120 中。如这里所定义的，该数据库可以存储在任何存储器或者其他计算机可读介质、磁带、闪存、RAM、ROM、磁盘等中。

[0005] 具体地，如果 X 是要存储 115 的密码 101，系统 100 实际上存储 f(X)，其中 f(.) 是某个加密或哈希 (hash) 函数 110。在认证阶段 20，用户输入候选密码 Y 102，系统确定 130f(Y)，并且仅当 f(Y) 与存储的密码 f(X) 匹配时才允许访问 150 该系统，否则拒绝 160 访问。

[0006] 作为优点，没有加密函数，加密的密码（通常非常难以逆转）对于攻击者来说是无用的。

[0007] 传统的基于生物测定的安全系统

[0008] 传统的生物测定安全系统具有与存储未加密的密码的基于密码的系统相同的弱点。具体地，如果数据库存储了未加密的生物测定参数，则这些参数容易遭到攻击和滥用。

[0009] 例如，在利用面部识别系统或者语音识别的安全系统中，攻击者能够搜索类似于攻击者的生物测定参数。在找到合适的生物测定参数后，攻击者可以修改这些参数以匹配该攻击者的外貌或者声音，从而获得未授权的访问。类似地，在利用指纹或虹膜 (iris) 识别的安全系统中，攻击者可以构造可以模仿匹配的指纹或虹膜以获得未授权的访问的装置，例如该装置可以是假指或者假眼。

[0010] 由于生物测定参数的固有的随时间的可变性，所以并不总是能够对这些参数进行加密。具体地，在登记阶段输入生物测定参数 X。利用加密或哈希函数 f(X) 对这些参数 X 进行加密并存储。在认证阶段，从同一用户获取的生物测定参数可能不同。例如，在利用面部识别的安全系统中，与认证期间相比，在登记期间用户的面部相对于相机可以有不同的方向。肤色、发型和面部特征都会改变。因此，在认证期间，加密的生物测定参数将不与任何存储的参数匹配，从而导致拒绝访问。

[0011] 纠错码

[0012] 字母表 Q 上的 (N,K) 纠错码 (ECC)C 包括长度为 N 的 $Q^K$ 个向量。可用 N 行 K 列的

生成矩阵 G 或用 N-K 行 N 列的奇偶校验矩阵 H 来描述线性 (N,K)ECC。名称“生成矩阵”是基于以下的事实：根据 w = vG，能够从任何长度 K 的输入行向量 v 通过将向量 v 右乘矩阵 G 来生成被表示为向量 w 的码字。类似地，为了检验向量 w 是否为码字，可以检验 $Hw^T = 0$ 是否成立，其中，列向量 $w^T$ 是行向量 w 的转置。

[0013] 在纠错码的标准应用中，将输入向量 v 编码成向量 w，并且对其进行存储或者进行发送。如果接收到向量 w 的受损版本，则解码器利用该码中的冗余信息 (redundancy) 进行纠错。直观来说，该码的纠错能力取决于该码中的冗余信息的量。

[0014] Slepian-Wolf 码、Wyner-Ziv 码及检验子码 (Syndrome Code)

[0015] 在某种意义上，Slepian-Wolf(SW) 码与纠错码相反。纠错码添加冗余信息并扩展数据，而 SW 码除去冗余信息并压缩数据。具体地，向量 x 和 y 表示相关数据的向量。如果编码器想要将向量 x 传送给已经具有向量 y 的解码器，则考虑到解码器已经具有向量 y 这一事实，编码器可以对数据进行压缩。

[0016] 举个极端的例子，如果向量 x 和 y 仅有一个位不同，则编码器可以通过简单描述向量 x 以及该不同的位置，来实现压缩。当然，对于更加实际的相关模型需要更复杂的码。

[0017] Slepian 和 Wolf 在“Noiseless coding of correlated information sources,”IEEE Transactions on Information Theory, vol.19, pp.471-480, July 1973, 以及 Wyner 和 Ziv 在“The rate-distortion function for source coding withside information at the decoder,”IEEE Transactions on Information Theory, vol.22, pp.1-10, January 1976 中描述了 SW 编码及相关的 Wyner-Ziv(WZ) 编码的基本理论。更加近期地，Pradhan 和 Ramachandran 在“DistributedSource Coding Using Syndromes(DISCUS):Design and Construction,”IEEETransactions on Information Theory, vol.49, pp.626-643, March 2003 中描述了这种码的具体实现。

[0018] 本质上，检验子码利用 N-K 行 N 列的奇偶校验矩阵 H 进行工作。为了将长度为 N 的二进制向量 x 压缩成长度为 K 的检验子向量，求出 S = Hx。解码通常取决于所使用的特定检验子码的详细内容。例如，如果检验子码基于网格结构 (trellis based)，则可以使用诸如熟知的 Viterbi 算法的各种基于动态编程的搜索算法，来找出与检验子码 S 和辅助信息 (side information) 序列相对应的最可能的源序列 x，如 Pradhan 等所述。

[0019] 另选地，如果使用低密度奇偶校验检验子码，则可以应用置信传播解码 (belief propagation decoding)，如 Coleman 等在“On some newapproaches to practical Slepian-Wolf compression inspired by channelcoding”, Proceedings of the Data Compression Conference, March 2004, pp.282-291 中所述。

## 现有技术

[0020] 与本发明相关的现有技术归为三类。首先，大量现有技术描述了生物测定参数的与这些生物测定参数的安全存储不相关的详细特征提取、记录和使用。因为我们的发明与安全存储有关，并且很大程度上与如何获取生物测定参数的细节无关，所以忽略这类现有技术的细节。

[0021] 与本发明相关的第二类现有技术包括被设计用于生物测定参数的安全存储和认证的以下系统：美国专利 6,038,315，“Method and system fornormalizing biometric

variations to authenticate users from a public databaseand that ensures individual biometric data privacy”;Davida 等的“On enablingsecure applications through off-line biometric identification,”Proceedings ofthe IEEE Symposium on Security and Privacy, May 1998 ;Juels 等的“AFuzzy Vault Scheme,”Proceedings of the 2002 IEEE InternationalSymposium on Information Theory, June 2002 ;以及美国专利 6,363,485,“Multi-factor biometric authenticating device and method”。

[0022] 图 2 示出了美国专利 6,038,315 中描述的基本方法的某些细节。在登记阶段 210，以被表示为 E 201 的比特序列的形式获取生物测定参数。接下来，从二进制纠错码中选出随机码字 W 202，并利用异或 (XOR) 函数 220 将其与参数 E 进行相加组合以产生基准 R 221。可选地，还可对基准 R 进行编码 230。在任何情况下，基准 R 都存储在密码数据库 240 中。

[0023] 在认证阶段 215，提供生物测定参数 E’ 205 以用于认证。该方法通过实质上从 R 减去 E’ 以获得 Z ＝ R-E’ ＝ W+E-E’ 251，利用 E’ 求出 250 R 的 XOR。然后，利用纠错码对该结果进行解码 260，以得到 W’ 261。在步骤 270 中，如果 W’ 与 W 匹配，则允许 271 访问，否则拒绝 272 访问。

[0024] 该方法实质上是测量汉明 (Hamming) 距离，即，在登记的生物测定参数 E 201 和认证生物测定参数 E’ 205 之间不同的位的数量。如果这种差异小于某一预定阈值，则允许访问。因为该方法仅存储基准 R，而不存储实际的生物测定参数 E，因此该方法是安全的。

[0025] Davida 等和 Juels 等描述了图 2 中所示的方法的变型。具体地，两种变型都在登记阶段中利用纠错码对生物测定数据进行编码，然后进行操作，以保护所得到的码字。Davida 等通过仅发送校验位来隐藏该码字，而 Juels 等添加被称为“壳 (chaff)”的一定量的噪声。

[0026] 美国专利 6,363,485 描述了将生物测定数据与纠错码和某些秘密信息（诸如密码或者个人识别号 (PIN)）进行组合以生成密钥的方法。将诸如 Goppa 码或 BCH 码的纠错码被应用于各种 XOR 运算。

[0027] 除了图 2 所示的固定数据库访问控制系统外，第三类现有技术包括利用生物测定以保护数据，特别是用于包括了存储器的移动设备（诸如笔记本电脑、PDA、蜂窝电话及数字照相机）的数据保护。因为移动设备容易丢失或者被盗，因此需要保护存储在移动设备中的数据。

[0028] 图 4 例示了用于存储数据 D 401 的现有方法的问题。在编码过程 410 中，从用户获得生物测定参数 P 402，并将其用作加密 440 数据 D 的密钥，以产生密文 C 441。P 和 C 都保存在存储器 450 中。当用户希望对数据 401 进行解密 420 时，从用户获取生物测定参数 P’ 460，并将其与所存储的生物测定参数 P 402 相比较 465。如果 P’ 与 P 匹配 470，则系统允许访问并利用 P 对所存储的密文 C 441 进行解密 480，以产生数据 D 401，否则不对数据进行解密 471。

[0029] 仅在没有危及到存储介质的安全的情况下，这种现有技术的系统才是有效的。如果攻击者可以访问该介质，则攻击者可以获取 P 并对数据进行解码。

[0030] 现有技术的问题

[0031] 第一，基于位的现有技术方法提供了不确定的安全性。此外，生物测定参数常常是实数值或者整数值，而不是二进制值。现有技术一般都假设生物测定参数由均匀分布的随机位组成，并且假设难以根据所存储的生物测定数据正确地求出这些位。实际上，生物测定

参数常常有偏离，这会对安全性产生负面影响。而且，即使攻击者仅恢复所存储的生物测定参数的近似版本，攻击也可能造成严重损害。现有技术方法没有被设计为防止攻击者根据编码版本估计出实际生物测定参数。

[0032] 例如，美国专利 6,038,315 依赖于以下事实：基准值 R = W+E 通过添加随机码字 W 来有效地对生物测定参数 E 进行加密。然而，该方法只能获得较差的安全性。有很多方法可以从 R 恢复 E。例如，如果向量 E 仅有几个位等于 1，则 R 和 W 之间的汉明距离较小。因此，纠错解码器能够容易地从 R 恢复 W，并由此也恢复 E。另选地，如果码字的分布较差，例如，如果码的加权谱 (weight spectrum) 小，并且很多码字都群集在全零向量周围，则攻击者可以根据 R 获得 E 的良好近似。

[0033] 第二，除了不确定的安全性外，现有技术的方法还有增加了所存储的数据量的实际弊端。因为生物测定数据库常常存储许多单个用户的数据，该额外的存储会极大增加系统的成本和复杂性。

[0034] 第三，很多现有技术的方法要求具有高计算复杂性的纠错码或算法。例如，现有技术的 Reed-Solomon 和 Reed-Muller 解码算法通常具有在所编码的生物测定参数的长度上至少是四次并通常是更高次的计算复杂性。

[0035] 第四，对于现有技术中已知的移动安全系统，存在基本结构上的基本问题。诸如图 4 所示的移动安全系统仅在没有危及该移动安全系统本身的安全时才是有效的。返回到笔记本电脑上的移动安全系统的示例，其安全性仅在攻击者不能在物理上访问存储有 P 和 C 的介质时才是有效的。如果攻击者可以访问该介质，例如通过从笔记本电脑上移除硬盘，则攻击者立即获取了作为用于产生 C 并由此对 C 进行解密的加密密钥的 P。

[0036] 现有的移动安全系统的主要困难在于：与用户的生物测定参数相对应的加密密钥是存储在设备中的。因此，如果该设备被盗，则可以利用所存储的参数对数据进行解码。

[0037] 因此，需要一种存储生物测定参数及对应的加密信息的方法，使得即使攻击者可以访问加密的信息和生物测定参数的存储版本，也不可能对数据进行解码。

## 发明内容

[0038] 例如从人脸、声音、指纹和虹膜获得的生物测定参数常常可以用于用户认证和数据访问控制。因为这些生物测定参数是连续的而且对于同一用户一次读取与下次读取可能不同，所以不能象对密码一样在数据库中以哈希 (hashed) 或加密形式存储生物测定参数。例如，面部的外观或者声音的音调会随时间有轻微变化。如果在数据库中存储生物测定参数，则它们易于受到“泄漏一次、到处运行 (break once，run everywhere)”攻击。

[0039] 本发明的一个实施方式利用检验子码来保护生物测定数据，例如，基于 Wyner-Ziv 或 Slepian-Wolf 编码的检验子码。这些检验子码可以安全地存储在数据库中，而仍然可以容许原始生物测定数据的固有可变性。

[0040] 具体地，根据本发明的生物测定检验子具有以下属性：第一，这些检验子码有效地隐藏或者加密了与原始生物测定特性相关的信息，使得在危及检验子数据库的安全的情况下，所存储的检验子码对绕过 (circumvent) 系统的安全性毫无用处。第二，可以对每个存储的检验子码进行解码，以产生原始生物测定参数，并对用户行认证或者对利用生物测定数据进行了加密的数据进行解密。

[0041] 所述检验子码可用于用户认证和数据加密。

## 附图说明

[0042] 图 1 是现有技术的基于密码的安全系统的框图；

[0043] 图 2 是现有技术的基于生物测定参数的安全系统的框图；

[0044] 图 3 是根据本发明的一个实施方式的生物测定安全系统框图；

[0045] 图 4 是现有技术用于保护数据的安全系统的框图；

[0046] 图 5 是根据本发明的一个实施方式的数据安全系统的框图；

[0047] 图 6 是根据本发明的一个实施方式的生物测定安全系统的框图；

[0048] 图 7 是根据本发明的一个实施方式的构造检验子码的过程的框图；

[0049] 图 8 是根据本发明的一个实施方式的生成直方图的过程的框图；

[0050] 图 9 是根据本发明的一个实施方式的选择特征向量的过程的框图；

[0051] 图 10 是根据本发明的一个实施方式的测量系数间相关性的框图；

[0052] 图 11 是根据本发明的一个实施方式的具有相关节点的置信传播因子图；以及

[0053] 图 12 是根据本发明的一个实施方式的检验子码的设计框图。

## 具体实施方式

[0054] 我们的发明的实施方式包括以下部分：用于安全地存储生物测定参数的检验子编码器和哈希方法；用于安全地存储使用生物测定密钥进行了加密的数据的、基于检验子码的加密方法；以及用于诸如前两种方法的安全生物测定应用的优化检验子码的方法。我们以独立的部分来描述各种方法。

[0055] 用于保护生物测定参数的检验子和哈希方法

[0056] 图 3 示出了根据本发明的基于检验子和哈希的生物测定安全系统 300。根据本发明的方法利用检验子码来压缩所测量的生物测定参数，以产生压缩的检验子码。不同于传统的压缩，由检验子码所产生的检验子码与原始生物测定数据无关。因此，所存储的检验子码不能用于对原始生物测定数据的近似进行解码。将所得到的压缩检验子码和该检验子码的哈希存储在生物测定数据库中。

[0057] 为了对用户进行认证，再次测量生物测定参数。将该生物测定参数与所存储的检验子码进行组合，以对原始的生物测定参数进行解码。如果检验子解码失败，则拒绝用户访问。如果检验子解码成功，则使用该原始生物测定参数对该用户的认证进行检验。

[0058] 登记阶段

[0059] 在登记阶段 310，获取用户的生物测定数据。例如，该生物测定数据是从面部图像、语音记录、指纹图像或虹膜扫描获得的。在下文中，生物测定数据是指从用户感测的、测量的或者获取的原始生物测定信号。可以从该生物测定数据提取特征。将这些特征设置为 d 维特征向量。该特征向量形成登记生物测定参数 301。用于从各种形式的生物测定数据中提取特征的方法在现有技术中是已知的，如上所述。以下更详细地描述将该特征向量转换成生物测定参数和最佳的检验子码。

[0060] 利用检验子编码器 330 对生物测定参数 E 301 进行编码，以产生登记检验子码 S 331。接下来，对登记检验子码 S 应用 340 消息认证码或哈希函数，以产生登记哈希 H 341。

该哈希函数可以是由 Ron Rivest 在“TheMD5 Message Digest Algorithm,”RFC 1321, April 1992 中所描述的已知的 MD5 加密哈希函数。将登记检验子码哈希对 (S, H)331、341 存储在生物测定数据库 350 中。

[0061] 可以使用任何类型的检验子码,例如上述 SW 码或 WZ 码。本发明的优选实施方式使用根据所谓的“重复积累码”(“乘积积累码”) 获得的码以及我们称为“扩展汉明积累码”的码。

[0062] 我们通常称这些为串联积累 (SCA, serially concatenated accumulate) 码。对于通常意义上关于这些类型的码的更多信息,参见 J.Li 等的“ProductAccumulate Codes :A Class of Codes With Near-Capacity Performance andLow Decoding Complexity,”IEEE Transactions on Information Theory, vol.50, pp.31-46, January 2004 ;M.Isaka. 和 M.Fossorier 的“High Rate SeriallyConcatenated Coding with Extended Hamming Codes,”IEEECommunications Letters,2004 ; 以 及 D.Divsalar 和 S.Dolinar 的“Concatenation of Hamming Codes and Accumulator Codes with High OrderModulation for High Speed Decoding,”IPN Progress Report 42-156,JetPropulsion Laboratory, Feb.15,2004。

[0063] 由 Yedidia 等在 2004 年 8 月 27 日提交的美国专利申请序列号 10/928,448, “Compressing Signals Using Serially-Concatenated AccumulateCodes”(在此通过引用将其并入) 描述了基于本发明所使用的 SCA 码的本发明的优选检验子编码器的操作。

[0064] 本发明的用于生物测定参数 301 的检验子编码器 330 具有多个优点。检验子编码器 330 能够对整数值输入进行操作。相反,现有技术的编码器通常是对二进制值输入进行操作。该检验子编码器具有非常高的压缩比,以使生物测定数据库 350 的存储要求最小化。该检验子编码器是速率自适应的,而且能够以增量的方式 (incremental fashion) 进行操作。可以根据需要发送更多的位、而不浪费先前发送的检验子位中的信息。

[0065] 认证阶段

[0066] 在认证阶段 320 中,再次从用户获取生物测定数据。提取特征以获得认证生物测定参数 E’360。搜索数据库 350 以找到用于该用户的匹配的登记检验子码 S 331 和登记哈希 H 341。

[0067] 该搜索可以检查数据库 350 中的每个条目 (S-H 对),或者可以利用启发式有序搜索以加快查找匹配的条目的过程。具体地,如果我们将数据库中的第 i 检验子码哈希对表示为 $(S_i, H_i)$,则穷举搜索首先将检验子解码为 E’和 $S_1$,并将检验子解码器输出的哈希与 $H_1$ 比较。如果拒绝访问,利用 $(S_2, H_2)$ 尝试相同的过程,然后使用 $(S_3, H_3)$,等等,直到尝试了所有条目或者允许访问为止。

[0068] 如果诸如登记用户名称的辅助信息可用,则可以使用该辅助信息来加快搜索。例如,在登记阶段,将该登记用户名称的哈希与 S-H 对一起存储。然后,在认证阶段,用户提供认证用户名称,并且系统求出该认证用户名称的哈希,在数据库中搜索具有匹配的哈希登记用户名称的 S-H 对,并尝试利用所得到的 S-H 对来对 E’进行认证。

[0069] 具体地,将检验子解码器 370 应用于具有用作“辅助”信息的认证参数 E’360 的登记检验子 S。检验子解码器在本领域通常是已知的。通常,利用了置信传播或 turbo 码的解码器具有优异的容错性并且复杂性低。检验子解码器 370 的输出是经解码的登记参数

E”371。解码值 E”371 是用于产生检验子码 S 331 的原始生物测定参数 E 301 的估计值。将哈希函数 340 应用于 E”371 以产生认证哈希 H’381。

[0070] 对登记值 H 341 和认证值 H’381 进行比较 390。如果这些值不匹配，则拒绝访问 392。否则，值 E”381 基本上与原始生物测定参数 E 301 匹配。这种情况下，可以允许用户访问 391。

[0071] 此外，在经解码的参数 E”381 和认证生物测定参数 E’360 之间进行直接比较，以对用户进行认证。例如，如果 E’和 E”与面部识别系统中的生物测定参数相对应，则可以将用于比较面部之间的相似性的传统算法应用于这些参数 E’和 E”。

[0072] 基于检验子的数据加密

[0073] 图 5 示出用于对数据 501 进行编码 510 和解码 520 的方法 500。在编码过程 510 中，从第一用户获得第一生物测定参数 P 502。这些参数用于对输入数据 D 501 进行加密 540，以产生密文 C 541。然而，与现有技术相比，这些第一生物测定参数 P 从不存储在存储器中。相反，检验子编码器 530 对该第一生物测定参数 P 进行编码，以产生检验子码 S 531，将 S 和 C 彼此关联，并且将 (S，C) 对 532 存储在存储器 550 中。在本发明的一个实施方式中，输入数据是在登记过程期间从用户获取的原始生物测定数据。

[0074] 当人们希望对密文 541 进行解码 520 时，从第二用户获取第二生物测定参数 P’560。使用第二生物测定参数对所存储的检验子码 S 531 进行检验子解码 570 以产生第三生物测定参数 P”571。然后，使用第三生物测定参数 P”对密文 C 541 进行解密 580，以产生输出数据 D’509。显然，如果第二或第三生物测定参数与第一生物测定参数不匹配，则输出数据 D’509 与输入数据 D 501 不匹配。只有在第一用户和第二用户是同一人时，输出数据才会与输入数据匹配。

[0075] 该方法具有以下优点。如果攻击者获得了对检验子码和密文 (S，C) 的访问，不能对数据进行解密。这是因为不能从检验子码恢复加密密钥 ( 即，第一生物测定参数 P)。此外，由于检验子码的纠错特性，即使第二生物测定参数 P’与第一生物测定参数 P 稍有不同，适当设计的检验子解码器也能够成功地产生与用作加密密钥 P 502 的第一生物测定参数完全相同的第三生物测定参数 P”。

[0076] 检验子编码提供了安全存储生物测定参数的有效方法，并且可以应用于安全存储生物测定信息的其他方法。应当注意，可以从生物测定数据提取特征向量。因此，可以使用对应的特征向量来代替任何上述生物测定参数。

[0077] 以加密形式存储生物测定参数的附加优点是，这使得安全生物测定存储应用能够对与在生物测定识别应用中使用的特征向量不同的特征向量进行操作。例如，指纹识别系统通常利用基于从指纹图像提取的所谓的“细节特征 (minutiae)”的特征向量。类似地，虹膜识别系统常常使用通过使虹膜图像通过一组 Gabor 过滤器而提取的特征。

[0078] 在很多情况下，用于生物测定识别 ( 例如面部识别或指纹识别 ) 的理想特征向量可以不同于用于检验子编码 / 解码的理想特征向量。在很多情况下，这是由于这样的事实 :用于训练识别或鉴别系统的分类器 ( 例如，基于高斯混合模型 (GMM)、神经网络或隐式 Markov 模型的分类器 ) 的过程产生了与用于训练直方图 (histogram)( 该直方图与检验子编码器及解码器的置信传播解码器一起使用，如这里所述 ) 的过程所产生的特征向量不同的特征向量。

[0079] 图 6 示出了用于存储所输入的生物测定数据 601 的加密版本的方法 600。如上所述，从用于测量或感测用户的生物测定特征的原始信号获得这些生物测定数据。

[0080] 在访问控制系统的登记阶段 610 中，例如，从用户获取第一生物测定数据 B 601。然后，从第一生物测定数据 B 601 获得第一生物测定参数 P 602 的特征向量。利用第一生物测定参数 P 作为加密密钥对第一生物测定数据 B 进行加密，以产生密文 C 641。此外，对第一生物测定参数进行检验子编码 630 以产生检验子码 S 631。然后将相关联的 (S,C) 对 632 存储在生物测定数据库 650 中。

[0081] 在认证阶段 620，从用户获得认证第二生物测定数据 B'660。该第二数据用于生成第二生物测定参数 P' 661 的特征向量。然后，检验子解码器 670 对第一生物测定参数进行解码，以产生第三生物测定参数 P"671。然后，将第三生物测定参数用作密钥对密文 C 进行解密 680，以产生第三生物测定数据 B"681。此时，通过生物测定识别方法 690 对认证生物测定数据 B' 和解码的生物测定数据 B" 进行比较，以确定是否允许 691 或拒绝 692 访问特定功能。跟前面一样，只有当第一和第三生物测定数据完全相同，即，第一和第二用户是同一人时才允许访问。

[0082] 在另一变型例中，比较步骤可以利用从生物测定数据所提取的特征向量。这些特征向量不必与生物测定参数相同。此外，被比较的两个特征向量仅需要基本上相同，这是因为验证步骤可以使用完全不同的过程。因此，这些特征向量可以允许生物测定数据中的更宽范围的变化（这在时间上表征了特定用户）。

[0083] 图 6 中所示的过程具有一些优点。认证系统可以在步骤 690 中利用传统的识别系统。此外，可以独立于生物测定验证步骤 690 所使用的参数或特征向量，来选择由检验子编码器 / 解码器使用的生物测定参数 P 和 P'。另外，检验子码是安全存储生物测定参数的有效方法。然而，还可以通过其中安全生物测定存储方法可以使用独立于生物测定验证方法的特征向量的方式，将图 6 中的方法应用于安全存储生物测定参数的其他方法。

[0084] 设计用于安全生物测定参数的最佳检验子码

[0085] 通常，在使用检验子码来保护生物测定参数和生物测定特征时，存在安全性与准确性之间的折衷。具体地，任何检验子码的关键参数是该码中的位的数量。具有大量位的检验子码传递更多与生物测定数据有关的信息，并且使得更容易容许该生物测定数据中的噪声和变化。相反，较小的检验子码向攻击者提供更少的信息，但更容易出错。

[0086] 在一个极端，当检验子码的长度与基本生物测定数据的长度基本相同时，可以容许任何量的噪声，这是因为可以仅从检验子码完全恢复原始生物测定数据。当然，在这种情况下，获得该检验子码的攻击者也可能恢复生物测定数据，从而危及到系统的安全。

[0087] 在另一个极端，在攻击者不能从检验子码恢复生物测定数据的情况下，位的数量非常小的检验子码提供了极好的安全性。然而，在这种情况下，登记生物测定数据和认证生物测定数据之间所允许的变化就受到了限制。

[0088] 显然，基于检验子的编码器和解码器应当选择平衡了生物测定变化的安全性和容限的检验子码长度。然而，精心设计的检验子码可以提高容错性。

[0089] 如图 12 所示，利用下面的术语来描述检验子码的设计。该设计从生物测定数据 1201（例如，面部图像）开始。从生物测定数据提取完全特征向量 1202。将完全特征向量 1202 简化为检验子特征向量 1203，并且将检验子特征向量用于设计最佳检验子码 1204。

[0090] 图 7 示出了用于构造最佳检验子码的过程 700。获取训练生物测定数据 1201。这些生物测定数据用于生成 800 误差直方图 890。误差直方图用于选择 900 用于检验子码的特征向量。在该环境中，我们使用术语“完全特征向量”1202 来表示所有生物测定参数，并且术语“检验子特征向量”1203 是指完全特征向量的子集。可以将检验子特征向量转换到不同的特征空间。

[0091] 在选择了检验子特征向量 1203 后，我们测量 1000 检验子特征向量的不同系数之间的相关性。然后，通过利用检验子特征向量的误差统计和系数间相关性，我们应用密度演化 (density evolution)740 来搜索可以产生特定长度的最佳检验子码 1204 的次数分布 (degree distribution)。在选择了检验子特征向量和检验子码后，我们构造 1100 利用了系数间相关性的置信传播解码器。

[0092] 在更详细地描述图 7 的每个部分之前，我们还定义以下术语。我们使用术语“硬”特征向量来指代特征向量的量化版本，并使用术语“软”特征向量来指代与“硬”特征向量相比，该特征向量的未量化或者更精细量化版本。因为某些生物测定参数可以包括在大数值范围上的整数和实数，所以使用量化。加密、密钥生成以及其他认证过程对于小范围的整数工作得最好。

[0093] 我们区分“硬”特征向量和“软”特征向量的原因在于 :检验子码是由“硬”特征向量构成的。因此，“硬”特征向量通常被量化。相反，在认证阶段，检验子解码器将“软”特征向量与检验子码进行组合，以对“硬”特征向量进行解码。因此，不需要对“软”特征向量进行量化，或者可以对“软”特征向量不同地进行量化，以提高系统的容错性。

[0094] 通常，可以存在从生物测定数据提取完全特征向量的多种方式，以及从完全特征向量提取“硬”和“软”特征向量的多种方式。在这些情况下，我们对每种可能性都应用图 7 的过程，并选择在训练期间产生最佳的整体结果的特征向量。

[0095] 构造误差直方图

[0096] 图 8 示出了用于生成误差直方图 890 的过程 800。首先，我们获取 810 在不同场合采用的特定用户的训练生物测定数据。接下来，我们选择 820 一对生物测定参数 B 和 B’，并求出完全“软”特征向量 VS(B)830 和完全“硬”特征向量 VH(B’)840。然后，对完全特征向量中的每个位置或维度 i，我们根据位置 i 的 VS(B) 来估计 845 对应位置 i 的 VH(B’) 的值，并确定 850 该估计是否正确。如果该估计不正确，则我们在误差直方图 890 中的位置 i 处使 VH(B’) 和 VS(B) 的相应值的一个柄 (bin) 递增 870。在对每个位置 i 完成该处理后，我们检查 860 是否已经处理了所有的生物测定参数对 B 和 B’。如果没有，则我们返回到步骤 820 并选择另一对生物测定参数。如果已经处理了所有的对，则完成误差直方图并结束 880 该过程。

[0097] 选择检验子特征向量

[0098] 图 9 示出了在图 8 的误差直方图的辅助下选择特征向量的过程 900。首先，从最可靠到最不可靠的位置 920 对误差直方图进行排序 910。具体地，如果 E(i) 是在根据 VS(B) 的位置 i 预测 VH(B’) 的位置 i 时的平均误差，则当 E(i) ＜ E(j) 时认为位置 i 比位置 j 更可靠。在对误差直方图进行排序后，我们将误差直方图中的次最可靠位置包括 930 在检验子特征向量中，构造 940 当前检验子特征向量的最佳检验子码，并测试 950 包括了最新位置是否增加了安全性或容错性。如果增加了安全性或容错性，则我们继续向检验子特征向量

添加另外的位置。否则，我们从特征向量中去除 960 最新添加的位置，并且我们终止 970 该过程。

[0099] 如果希望指定安全性等级并使容错性最优，则以下步骤可用于步骤 940 和 950：首先，在步骤 940 中，通过根据固定的次数分布生成具有 S 检验子的低密度奇偶校验 (LDPC) 码，来构造具有与特征向量中的当前的位置的数量相对应的长度 N 的新检验子码。在这种情况下，通过固定量 N-S 使安全性级别保持恒定，并在整个过程中使其保持恒定。然后，从数据库中选择生物测定数据的随机生物测定抽样，并通过应用 LDPC 码的奇偶校验矩阵将其映射到检验子码，并且利用应用于来自同一用户的另一随机生物测定抽样的置信传播，来对所得到的检验子码进行解码。多次重复该过程而产生对于该给定特征向量的检验子码的容错性的估计。另选地，如果在该设计过程中可容许更大的计算复杂性，则可以使用 Richardson 等在“Design of capacity-approaching irregular low-densityparity-check codes”，IEEE Transactions on Information Theory，vol.47，issue2，pp.619-637，February 2001 中讨论的密度演化处理（在此通过引用将其并入），来对该码优化次数分布，并更准确地估计误差概率。

[0100] 如果希望指定容错性的级别并获得最好的安全性，以下步骤可用于步骤 940 和 950：首先，在步骤 940 中，利用密度演化设计具有与特征向量中的当前的位置的数量相对应的长度 N 的新检验子码。具体地，利用密度演化来构造一系列不同等级 (rate) 的码，直到找到满足通过密度演化所评估的指定级别的容错性的最高等级的码。

[0101] 我们将通过该过程所选择的特征向量称为“检验子特征向量”，因为它是为该检验子码专门设计的特征向量。我们注意到，该特征向量可以具有与为诸如面部或对象识别的生物测定识别而构造的其他类型的特征向量不同的属性。

[0102] 测量系数间相关性

[0103] 在已经选择了检验子特征向量之后，下一步骤是测量系数间相关性。不能从根据图 7 生成的误差直方图中提取该信息，因为该误差直方图是针对完全特征向量 1202 生成的，而步骤 900 仅选择完全特征向量中的位置的子集来产生检验子特征向量 1203。

[0104] 图 10 示出了用于测量二进制检验子特征向量中的第一阶相关性的过程 1000。也可以将该过程应用于非二进制特征向量或者更高阶的相关性。首先，从生物测定训练数据集中选择元素，并从该元素提取检验子特征向量。然后，将计数器变量 i 初始化 1010 为零。接下来，我们测试 1020 位置 i 是 0 还是 1，并且，当为前者时前进到步骤 1030，当为后者时前进到步骤 1040。然后，我们测试 1030 位置 i-1（即，前一位置）是 0 还是 1，并使该直方图中的适当柄递增 1035。直观地，柄 p00 对后面跟着 0 的 0 的出现进行计数，柄 p01 对后面跟着 1 的 0 的出现进行计数，等等。接下来，我们使计数器 i 递增 1050，测试 1060 是否还有更多的位置保留在检验子特征向量中，并且我们对下一位置重复该过程。否则，如果我们已经处理了每个位置，则我们终止 1070 该过程。

[0105] 在对生物测定训练集中的每个元素执行图 10 中的过程之后，我们将柄 p00、p01、p10 和 p11 的值除以生物测定训练集的大小，以测量该检验子特征向量的第一阶相关性。

[0106] 利用密度演化来构造最佳检验子码

[0107] 在已经选择了检验子特征向量 1203 并已经测量了系数间相关性之后，我们随后利用密度演化来设计检验子码 1204。具体地，对于 LDPC 检验子码，我们设计检验子码的次

数分布。有两个理由需要这样的设计。首先，在步骤 900 中设计的初始检验子码可能使用了在选择检验子特征向量 1203 之前选择的固定次数分布。第二，初始检验子码没有利用系数间相关性的知识，因为只可以在选择了检验子特征向量之后测量该相关性。

[0108] 为了实际上构造最佳次数分布，我们应用密度演化技术来产生多个候选次数分布。

[0109] 然而，现有技术中已知的传统密度演化处理没有考虑系数间相关性。因此，虽然通过密度演化产生的候选次数分布对于没有系数间相关性的情况可能是合适的，但当存在系数间相关性时它们通常会有不同的表现。

[0110] 为了获得检验子码的最佳次数分布，我们比较通过对生物测定训练数据集的密度演化而获得的候选次数分布，并选择表现最好的次数分布。在另选实施方式中，我们修改传统的密度演化算法，以考虑系数间相关性。

[0111] 构造检验子码的置信传播解码器

[0112] 在设计检验子码中的最后步骤是构造相关的置信传播检验子解码器。用于没有系数间相关性的应用的置信传播解码器是已知的。然而，传统的解码器没有设计用来解决系数间相关性。具体地，传统的置信传播方法利用 Kschischang 等在“Factor graphs and the sum-productalgorithm，”IEEE Transactions on Information Theory，vol.47，issue 2，pp.498-519，February 2001（在此通过引用将其并入）中所述的和 - 乘公式，将“消息”从变量节点传递到检验节点并再次传回来。

[0113] 如图 11 所示，本发明的置信传播因子图的构造 1100 除了传统检验节点 1110 和变量节点 1120 外还包括相关性节点 1130。具体地，将相关性节点添加在每对连续的变量节点之间。对将消息从变量节点传递到相邻的检验节点的方法进行改进，以包括与其他消息相乘的、来自每个相邻的相关性因子节点的附加消息。

[0114] 具体地，利用 Kschischang 等的表示，如果 $\mu_{y\to f}(x)$ 是从检验节点 f 到变量节点 y 的对于状态 x 的输入消息，而 L(x) 是来自左边的相关性节点的输入消息，则从该变量节点到右边的相关性节点的输出消息为

[0115] $L(x)\cdot\prod\mu_{y\to f}(x)$，

[0116] 而到左边的相关性节点的输出消息为

[0117] $R(x)\cdot\prod\mu_{y\to f}(x)$，

[0118] 其中，R(x) 是来自右边的相关性节点的输入消息。

[0119] 我们还描述了一种根据本发明的实施方式的用于将消息传送到相关性节点及从相关性节点传送消息的方法。具体地，我们描述确定消息 L(x) 和 R(x) 的过程。如果 μ(0) 是到左边的相关性节点的输入消息，则该相关性节点的右侧的输出消息（其为到该相关性节点的右边的变量节点的输入消息）为

[0120] $L(0)=p00\cdot\mu(0)+p10\cdot\mu(1)$

[0121] 和

[0122] $L(1)=p10\cdot\mu(0)+p11\cdot\mu(1)$，

[0123] 其中，p00、p01、p10 和 p11 项为如图 10 所示测量的第一阶相关性值。

[0124] 类似地，该相关性节点的左侧的输出消息（其为到该相关性节点的左边的变量节点的输入消息）为

[0125] $R(0) = p00 \cdot \mu(0) + p01 \cdot \mu(1)$

[0126] 和

[0127] $R(1) = p01 \cdot \mu(0) + p11 \cdot \mu(1)$。

[0128] 用于虹膜生物测定参数的检验子码设计

[0129] 接下来，我们描述我们如何将上述过程应用于虹膜生物测定参数的具体示例。我们选择完全“硬”特征向量作为从如 J.Daugman 在“How irisrecognition works,”IEEE Transactions on Circuits and Systems for VideoTechnology, vol.14, issue 1, pp.21-30, January 2004（在此通过引用将其并入）中所述的一组 Gabor 过滤器提取的一系列位。

[0130] 当完全“硬”特征向量是二进制时，我们选择完全“软”特征向量为四进制的。具体地，我们选择位置 i 处的完全“软”特征向量的值为其位置应该在“硬”特征向量中的值的最佳推测值 (guess)，并且我们还附加表示可靠性级别的位。具体地，我们附加表示对于该位置的判决我们确信还是不确信的位。

[0131] 例如，“硬”特征向量的某些位置的特征可能是难以预测的，例如，因为这些特征被眼睑或睫毛覆盖，并且这些位置应该采用“不确信”可靠性值。

[0132] 接下来，我们利用生物测定训练数据来生成误差直方图，如以上针对图 8 所述，然后应用图 9 的特征向量设计方法。尽管完全特征向量的长度大约为 10,000，但是我们发现与很多位置相关联的特征是不可靠的。例如，与眼睛上部相对应的特征向量的分量通常被眼睑或者睫毛所覆盖。在通过图 9 的过程丢弃用途最小的位置之后，在检验子特征向量中我们还剩有大约 2,000 个最可靠的位置。

[0133] 如果我们这时停止图 7 中的步骤 900，则所得到的检验子码将不是容错性的，不能容许单个用户的虹膜生物测定参数中的自然变化。具体地，在某天所获取的用户的虹膜的检验子码与另一天所获取的同一虹膜的另一生物测定参数进行组合，大约有 12％的时间不能解码。这证明了对图 7 中的其余步骤的需要。

[0134] 在我们利用图 10 中的过程测量了第一阶相关性之后，我们检测到“硬”检验子特征向量中的位取与相邻位相同的值的可能性大约是取与该相邻位相反的值的可能性的两倍。然后，我们继续图 7 中的步骤 740，以利用密度演化来构造经优化的检验子码，来利用高相关性。最后，我们按照步骤 1100 来构造置信传播解码器，以考虑高的第一阶相关性。

[0135] 按照这些步骤产生了检验子码，这些检验子码与我们的初始码相比，具有高出一个数量级以上的可靠性，由此证明了按照图 7 中的整个过程的优点。

[0136] 本发明的效果

[0137] 本发明实现了基于生物测定参数的安全用户认证。因为存储了检验子码而不是原始生物测定数据，所以本发明是安全的。这防止了获得对数据库的访问的攻击者获知基本的生物测定数据。

[0138] 可以利用来自多个描述 (multiple descriptions) 的已知问题的传统工具（例如，参见 V.K.Goyal 的“Multiple description coding:compressionmeets the network,”IEEE Signal Processing Magazine, vol.18, pp.74-93, September 2001），来限制原始生物测定参数 E 的最优可能估计（攻击者仅能够使用检验子码 S 来作出该估计）。此外，当通过绝对误差、平方误差、加权误差测量还是任何任意的误差函数来测量该估计的

质量时，可以进行这些限制。相反，所有的现有技术方法都是基于二进制值的。因此，安全性取决于汉明距离。

[0139] 实质上，检验子码 S 的安全性是归因于它是原始生物测定参数 E 的压缩版本这一事实。此外，该压缩表示与 E 的“最低有效位”相对应。利用数据压缩理论的已知工具，可以证明：如果使用了具有高压缩的检验子码，则这些最低有效位最多只能够产生原始参数 E 的较差估计，例如，参见 Effros 的“Distortion-rate bounds for fixed-and variable-ratemultiresolution source codes,”IEEE Transactions on Information Theory, vol.45, pp.1887-1910, September 1999 以及 Steinberg 和 Merhav 的“Onsuccessive refinement for the Wyner-Ziv problem,”IEEE Transactions onInformation Theory, vel.50, pp.1636-1654, August 2004。

[0140] 第二，因为伪造至少是与在基本哈希函数中找到冲突一样困难，所以本发明是安全的。具体地，如果经解码的生物测定参数 E”的哈希 H’与原始哈希 H 匹配，则系统在认证阶段仅接受检验子对 (S, H)。对于加密哈希函数（例如 MD5），通常认为不可能找到与 E 不同但是具有与 E 的哈希匹配的哈希的元素 E”。因此，如果检验子解码成功地对具有正确的哈希的 E”进行了解码，则系统可以确信 E”实际上与 E 相同，并且利用原始生物测定参数进行所有认证判决。

[0141] 第三，本发明在产生检验子 S 时对原始生物测定参数 E 进行压缩。用于很多用户的生物测定数据库能够要求大量的存储，尤其是在生物测定数据问题需要大量数据（例如，面部图像或者语音信号）的情况下。因此，减少所要求的存储可以实现在成本和容错性方面的很大改善。相反，用于生物测定数据的安全存储的大部分现有技术方法由于加密或纠错的开销而实际上增加了所存储的数据的大小，因此与不安全的系统相比需要更多的存储。

[0142] 第四，因为本发明建立在检验子码的理论上，所以本发明可以应用复杂的码构造和解码算法。具体地，根据本发明的检验子编码使得便于使用对二进制和多级码构造进行解码的软解码（其使用已知的 Viterbi 算法、置信传播和 burbo 解码）。相反，因为多数现有技术方法基于二进制码、Reed-Solomon 码和代数解码，所以当生物测定数据取实值（与二进制值相反）时，不能有效地应用软解码。例如，一些方法特别要求在登记阶段利用随机码字计算生物测定数据的 XOR，以产生基准，并要求在认证阶段利用生物测定数据计算该基准的 XOR。

[0143] 第五，尽管与安全生物测定相关的大多数现有技术利用了纠错编码，但是本发明利用检验子编码。纠错编码的计算复杂性通常在输入大小上是超线性的。相反，通过利用各种类型的基于低密度奇偶校验的检验子编码，很容易构造检验子编码器，其中检验子编码的计算复杂性在输入大小上仅是线性的。

[0144] 第六，通过利用检验子编码框架，能够使用如同由 Yedidia 等描述的 SCA 码一样的强大的新的嵌入式检验子码。这些码使得检验子编码器能够在登记期间估计生物测定数据的固有可变性，并仅对足够的检验子位进行编码以使得能够成功地进行检验子解码。

[0145] 第七，如上所述的检验子码可以用于对数据进行加密。此外，描述了使得能够设计具有给定性能级别和容错性的最佳检验子码的方法。

[0146] 尽管通过优选实施方式的示例描述了本发明，但是可以理解的是，可以在本发明

的精神和范围内进行各种其他的修改和改进。因此，所附权利要求的目的是覆盖落入本发明的真正精神和范围内的所有这些修改和改进。

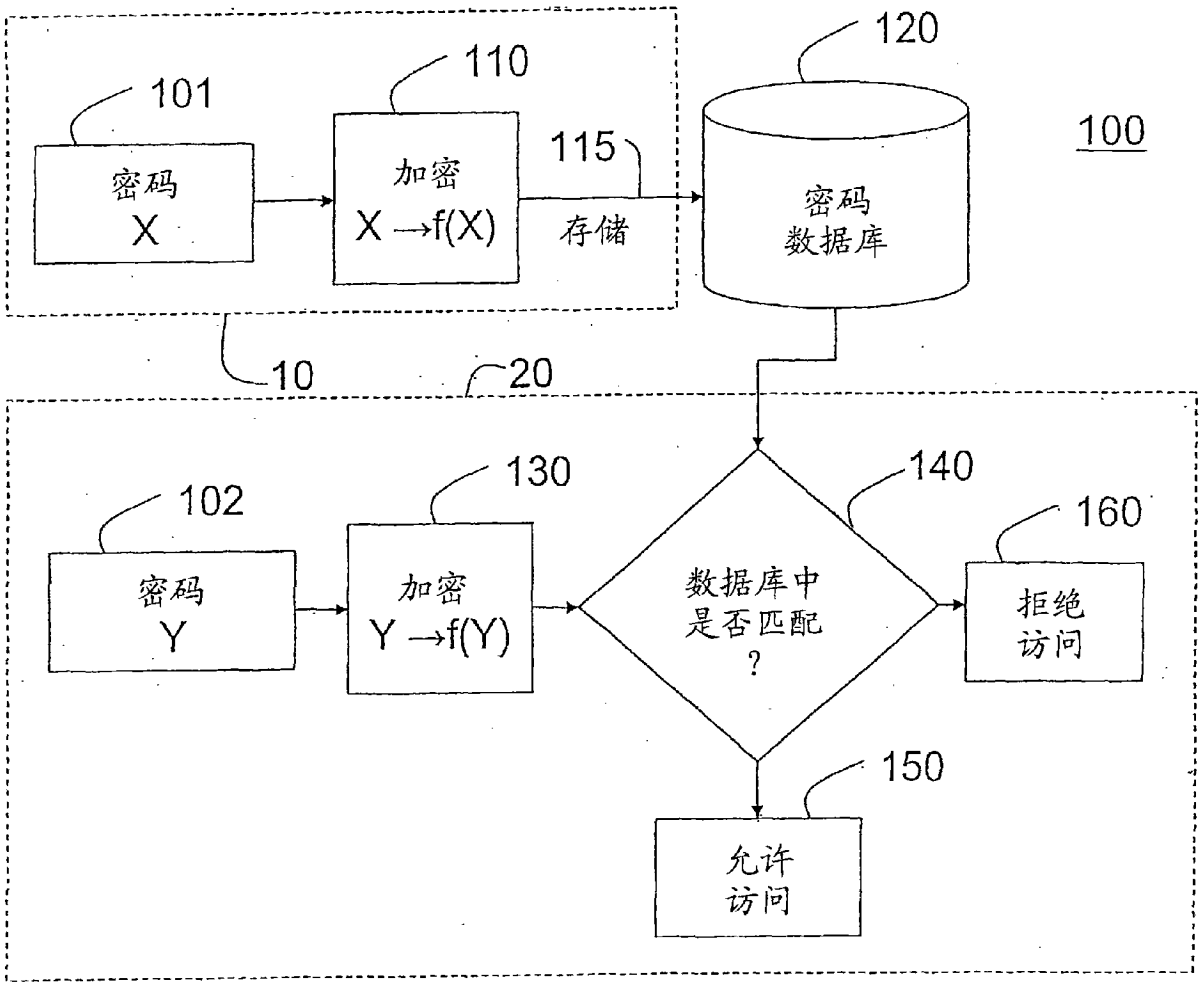

101
密码
X
110
加密
X →f(X)
115
存储
120
密码
数据库
100
10
20
102
密码
Y
130
加密
Y →f(Y)
140
数据库中
是否匹配
?
160
拒绝
访问
150
允许
访问

图 1

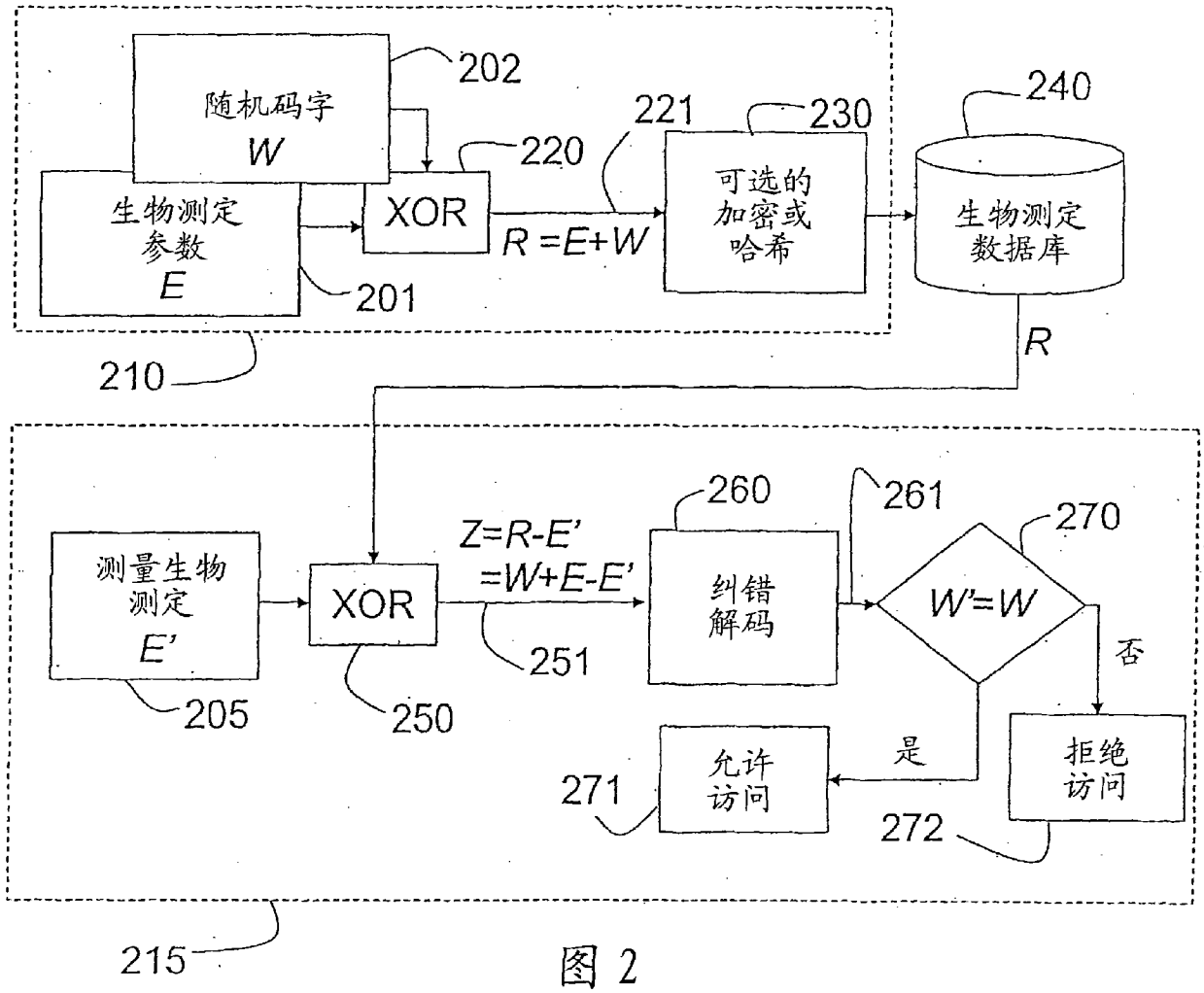

随机码字
W
202
生物测定
参数
E
201
XOR
220
R =E+W
221
230
可选的
加密或
哈希
240
生物测定
数据库
210
R
测量生物
测定
E'
205
XOR
250
Z=R-E'
=W+E-E'
251
260
纠错
解码
261
W'=W
270
否
是
允许
访问
271
拒绝
访问
272
215

图 2

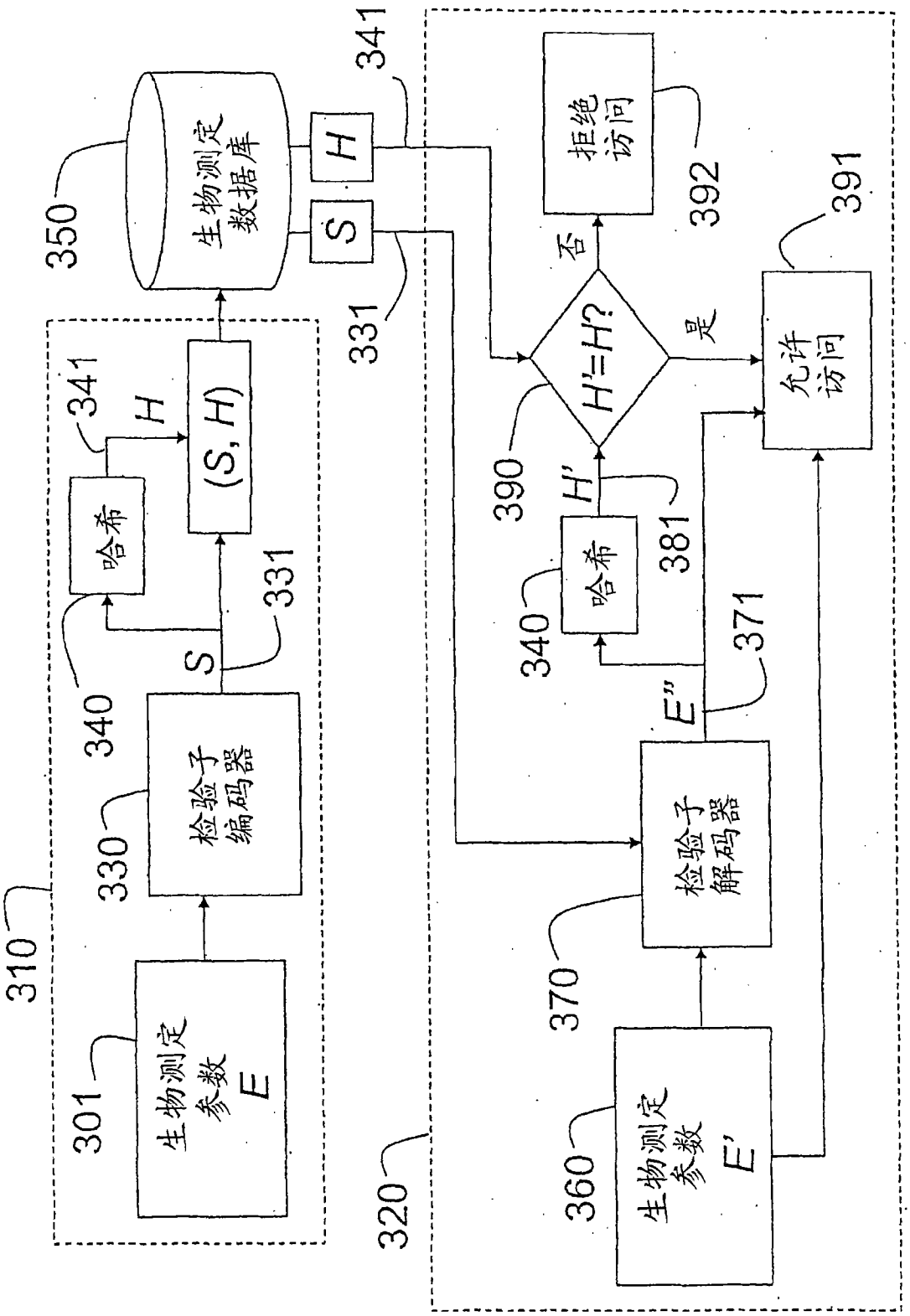

生物测定参数 E
检验子编码器
330
340
S
331
哈希
341
H
(S, H)
生物测定数据库
350
S
H
341
331
生物测定参数 E'
360
检验子解码器
370
E''
371
哈希
340
H'
381
390
H'=H?
否
是
拒绝访问
392
允许访问
391
310
320
300

图 3

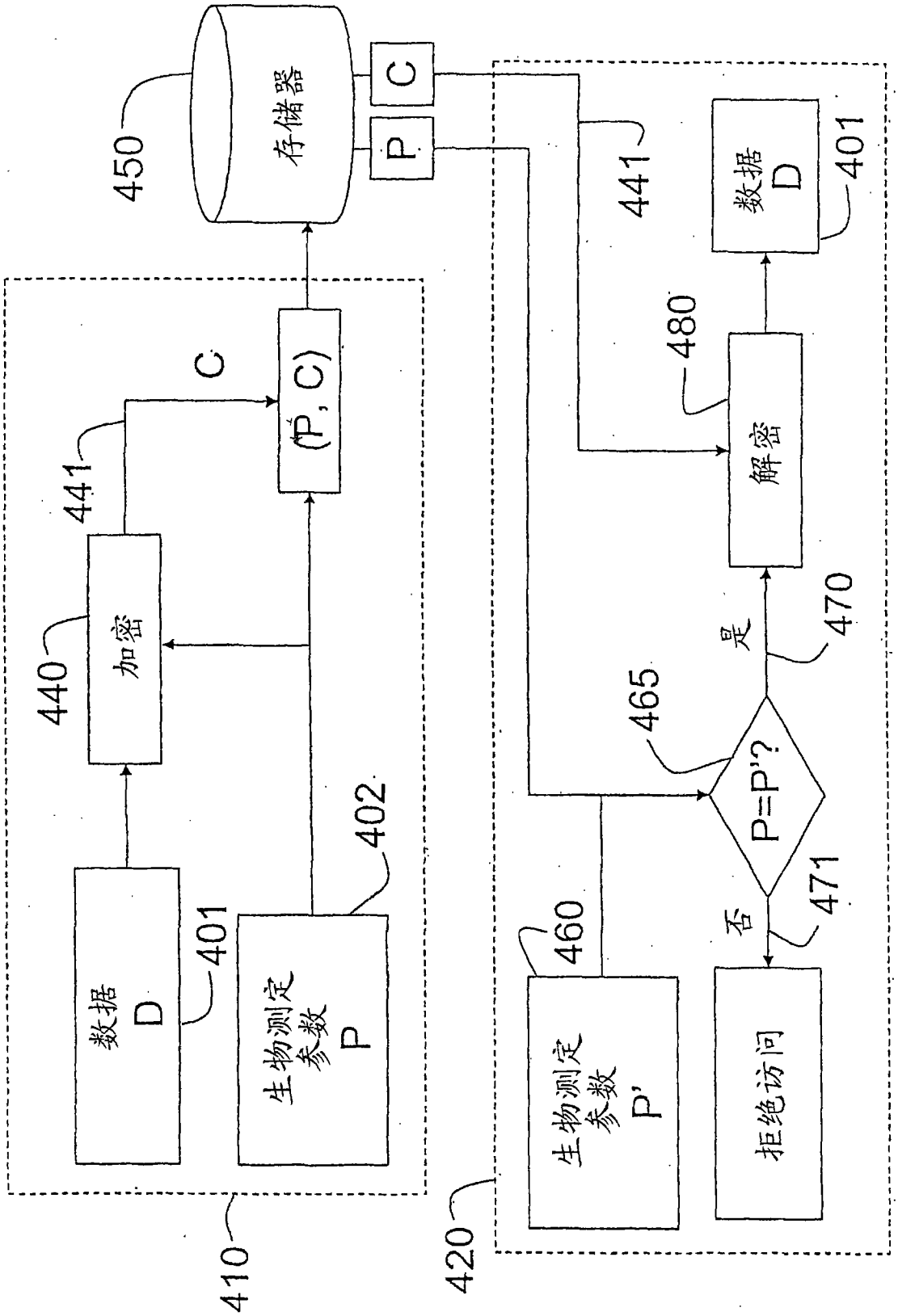

450
存储器
C
P
441
数据
D
401
480
解密
(P', C)
C
441
加密
440
数据
D
401
生物测定
参数
P
402
410
420
生物测定
参数
P'
460
465
P=P'?
是
470
否
471
拒绝访问

图 4

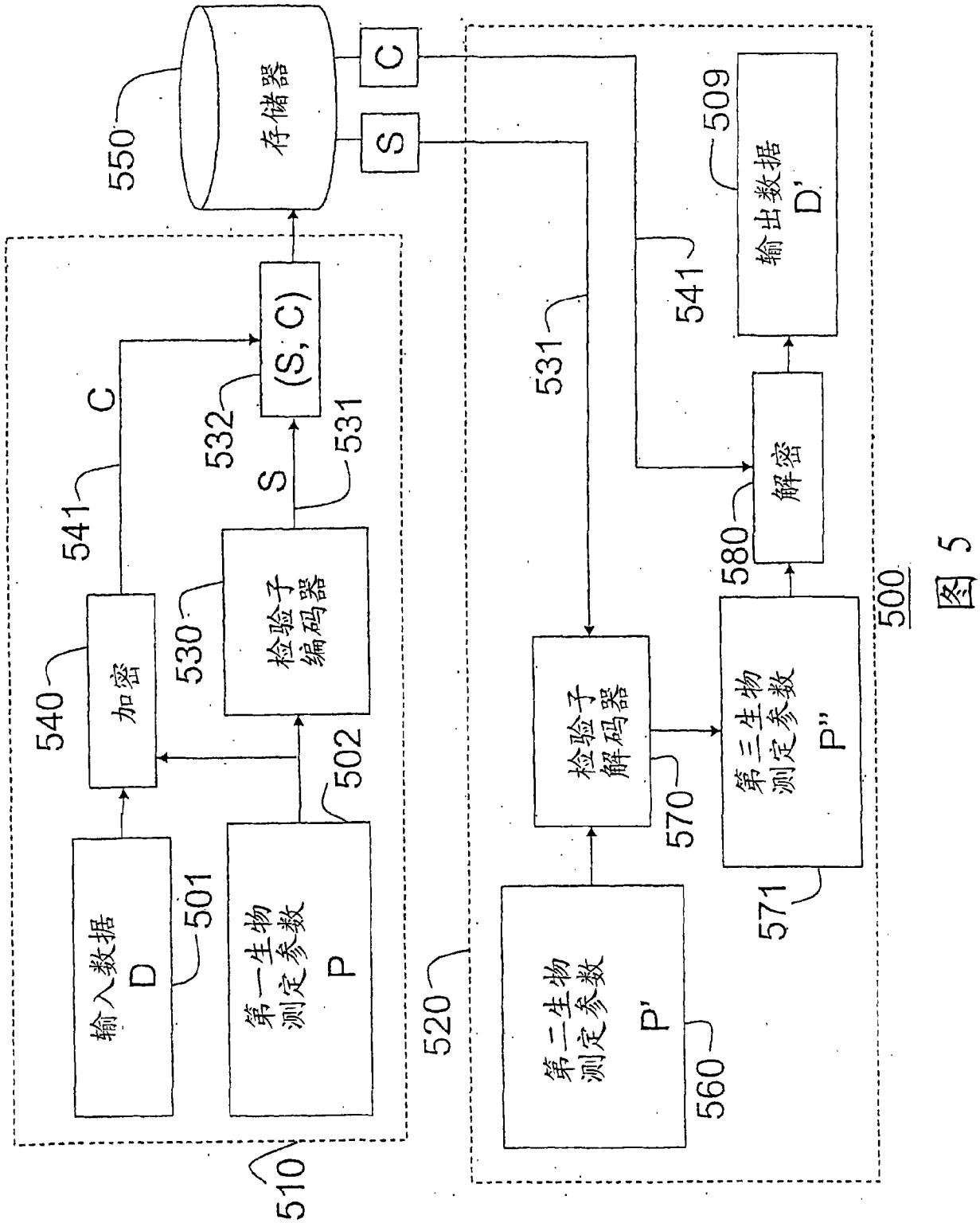

存储器
550
C
S
509
输出数据
D'
541
531
(S; C)
532
S
531
C
541
加密
540
检验子
编码器
530
502
输入数据
D
501
第一生物
测定参数
P
510
520
检验子
解码器
570
第二生物
测定参数
P'
560
第三生物
测定参数
P''
571
解密
580
500

图 5

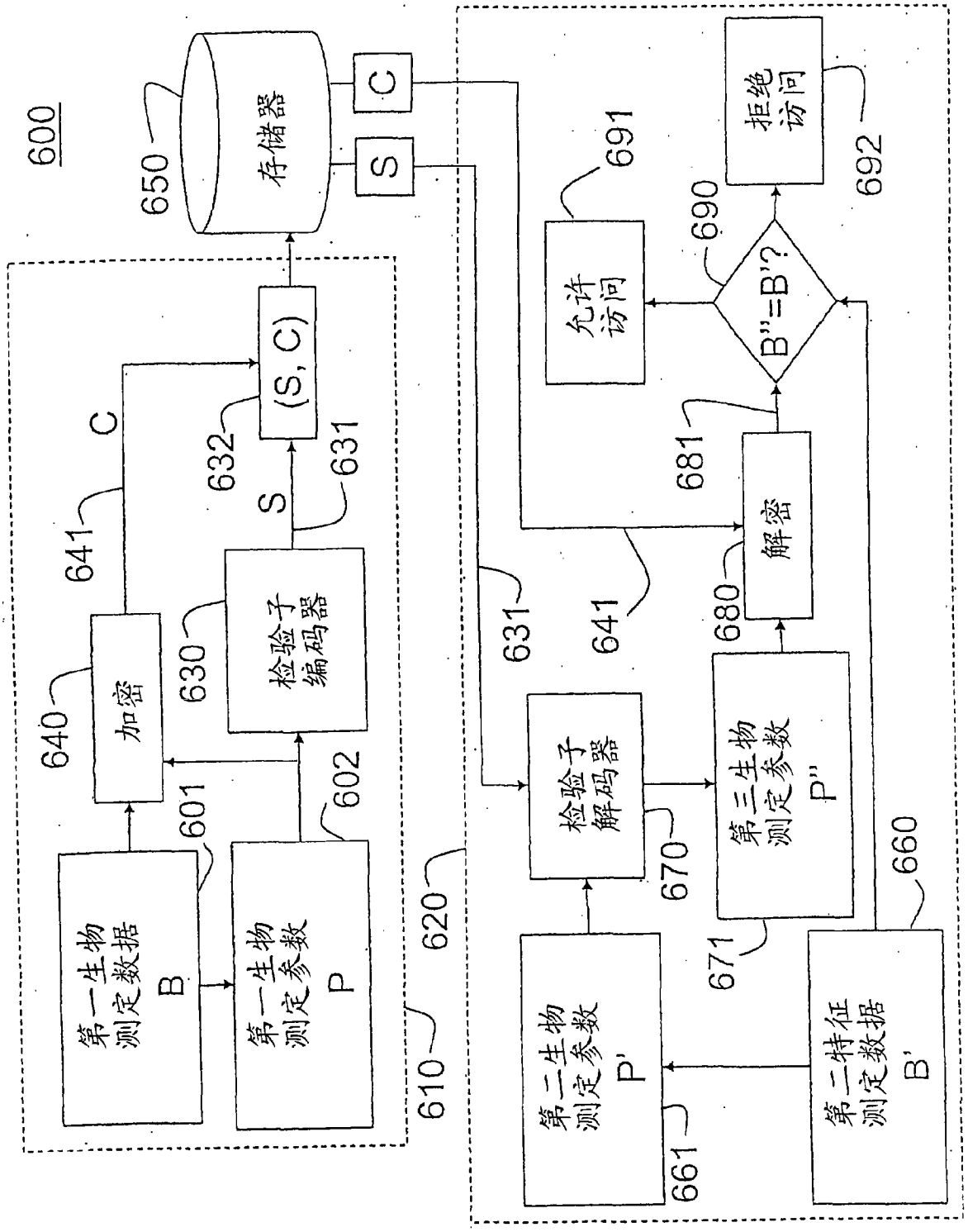

600
650
存储器
C
S
610
620
第一生物测定数据 B
第一生物测定参数 P
601
602
加密
640
641
C
检验子编码器
630
(S, C)
632
S
631
631
641
检验子解码器
670
第二生物测定参数 P'
661
第三生物测定参数 P''
671
680
解密
681
B''=B'?
690
允许访问
691
拒绝访问
692
660
第二特征测定数据 B'

图 6

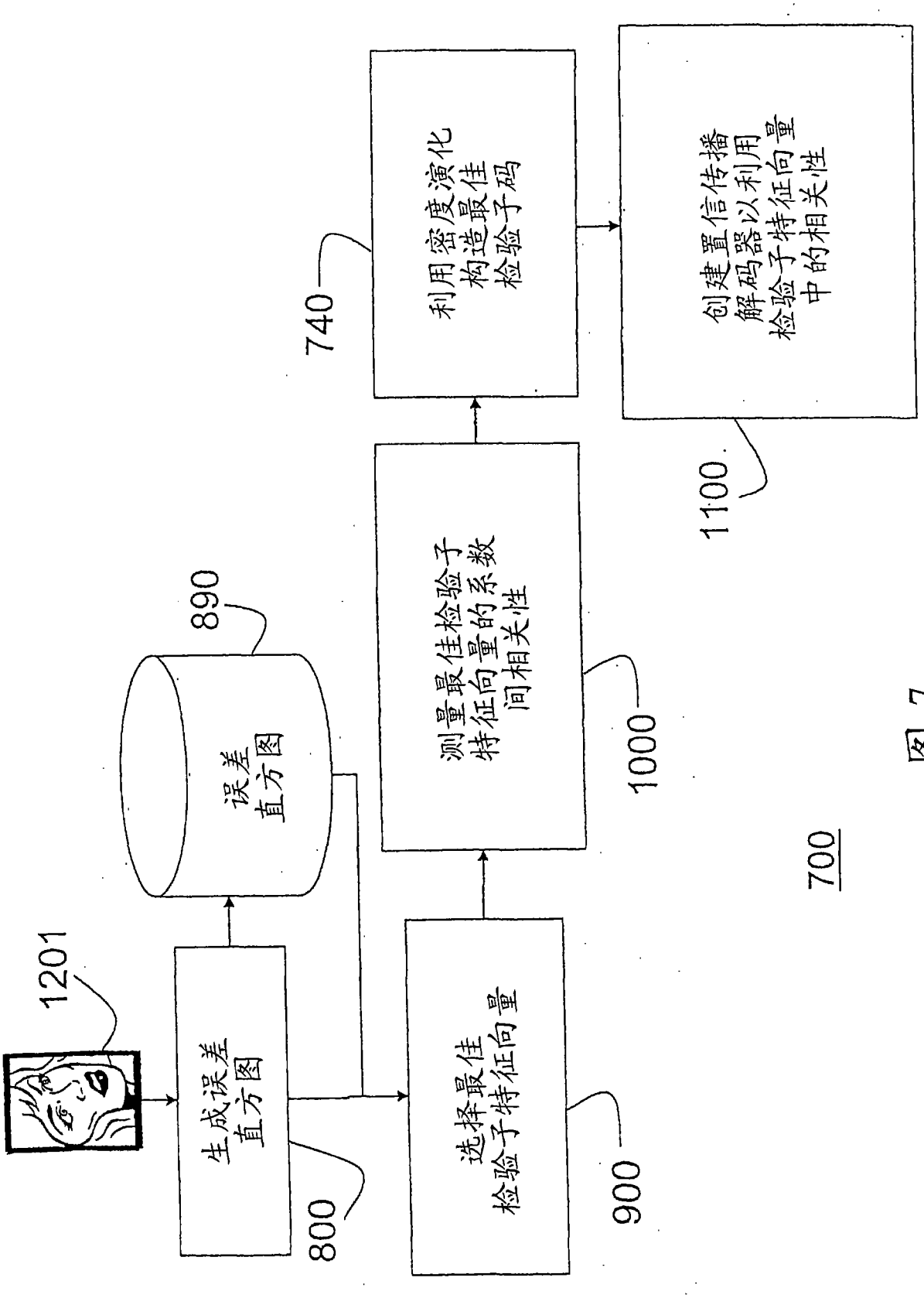

1201
生成误差直方图
800
误差直方图
890
选择最佳检验子特征向量
900
测量最佳检验子特征向量的系数间相关性
1000
利用密度演化构造最佳检验子码
740
创建置信传播解码器以利用检验子特征向量中的相关性
1100
700

图 7

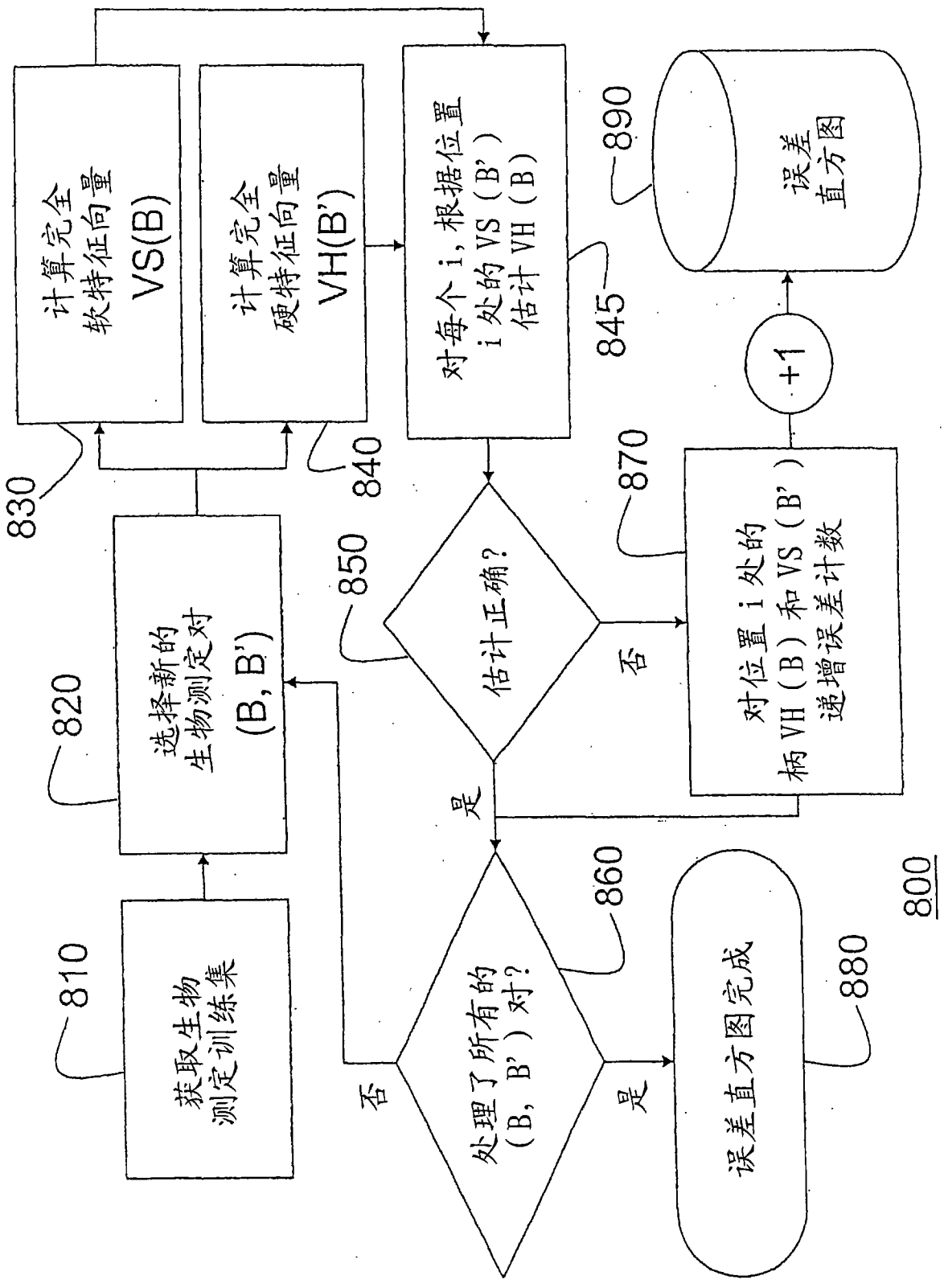

获取生物测定训练集
810
选择新的生物测定对 (B, B')
820
计算完全软特征向量 VS(B)
830
计算完全硬特征向量 VH(B')
840
对每个 i，根据位置 i 处的 VS（B'）估计 VH（B）
845
估计正确?
850
否
对位置 i 处的柄 VH（B）和 VS（B'）递增误差计数
870
+1
误差直方图
890
是
处理了所有的（B, B'）对?
860
否
是
误差直方图完成
880
800

图 8

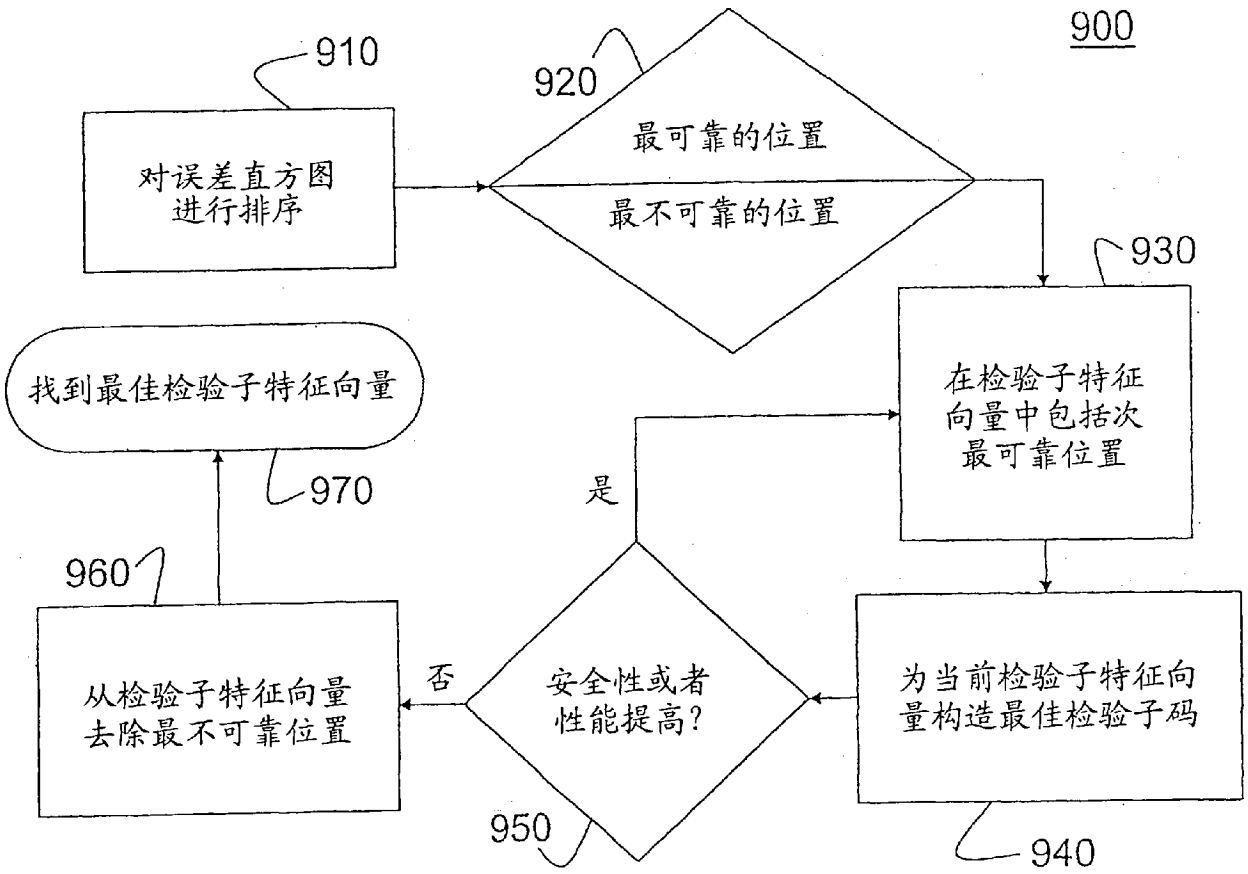

900
910
对误差直方图进行排序
920
最可靠的位置
最不可靠的位置
930
在检验子特征向量中包括次最可靠位置
为当前检验子特征向量构造最佳检验子码
940
安全性或者性能提高?
950
是
否
从检验子特征向量去除最不可靠位置
960
找到最佳检验子特征向量
970

图 9

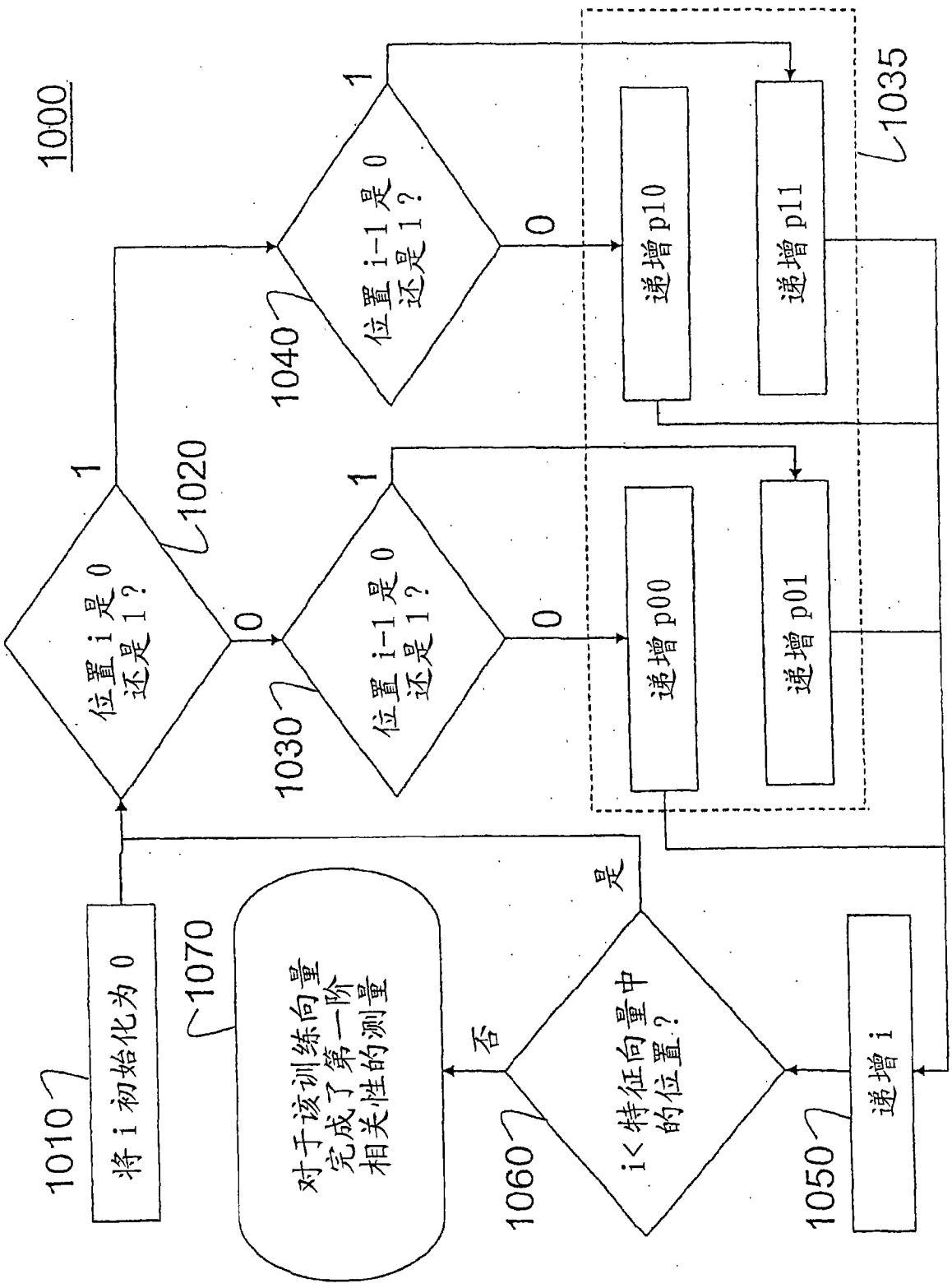
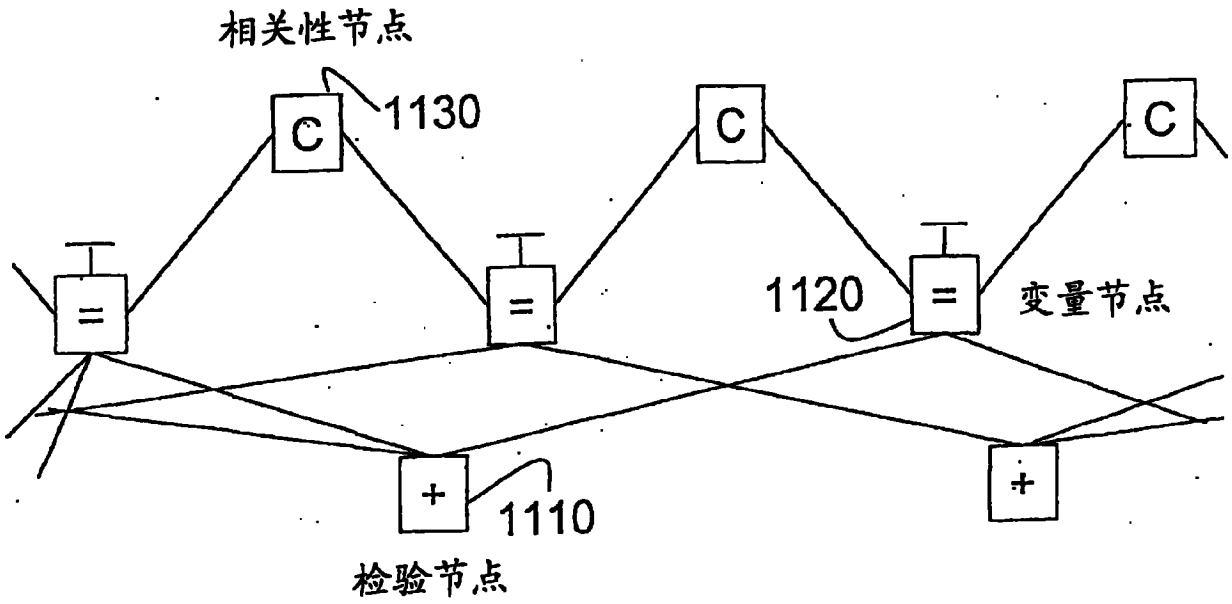

1000
将 i 初始化为 0
1010
对于该训练向量
完成了第一阶
相关性的测量
1070
i< 特征向量中
的位置?
1060
否
是
递增 i
1050
位置 i 是 0
还是 1?
1020
0
1
位置 i-1 是 0
还是 1?
1030
递增 p00
递增 p01
位置 i-1 是 0
还是 1?
1040
递增 p10
递增 p11
1035

图 10




1100

图 11

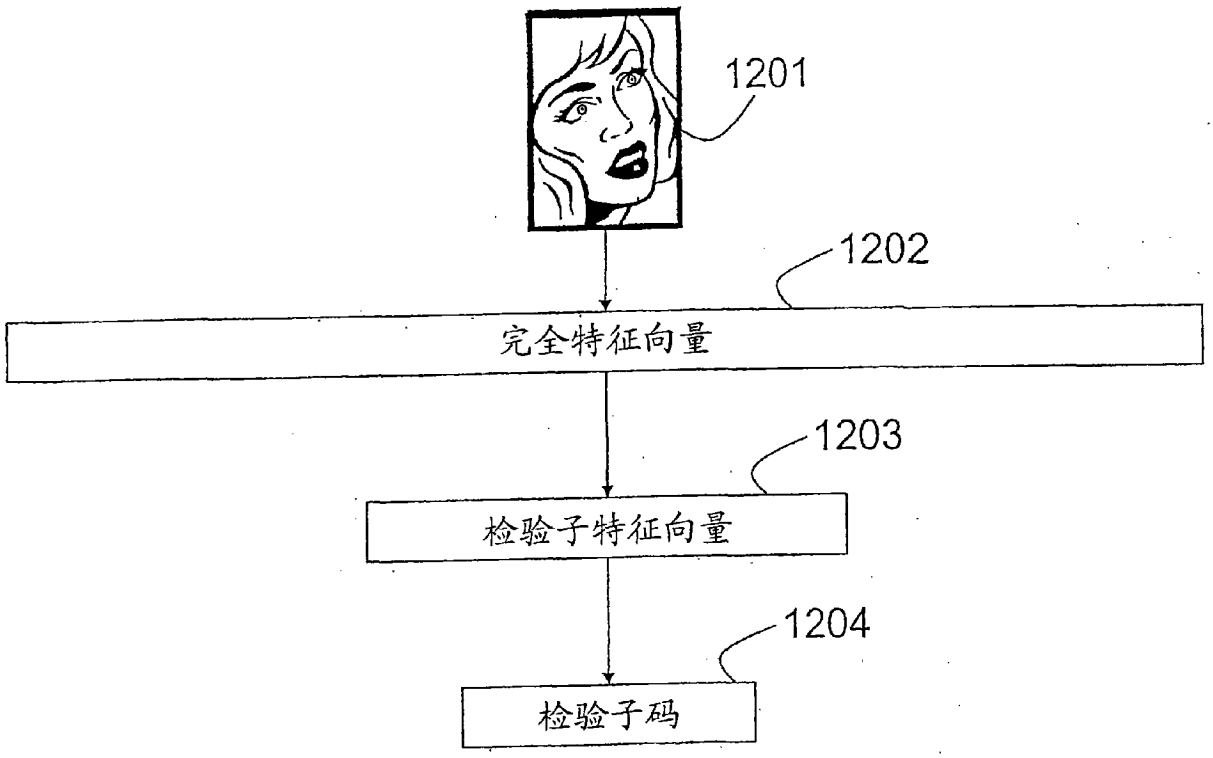



图 12