



(21) 申请号 202010183660.3

(22) 申请日 2020.03.16

(65) 同一申请的已公布的文献号

申请公布号 CN 113407375 A

(43) 申请公布日 2021.09.17

(73) 专利权人 浙江宇视科技有限公司

地址 310051 浙江省杭州市滨江区西兴街
道江陵路88号10幢南座1-11层、2幢A
区1-3楼、2幢B区2楼

(72) 发明人 刘灿铭 吴向兵 薛露萍

(74) 专利代理机构 北京品源专利代理有限公司

11332

专利代理师 孟金喆

(51) Int. Cl.

G06F 11/14 (2006.01)

(56) 对比文件

CN 104866610 A, 2015.08.26

CN 106844607 A, 2017.06.13

KR 101575246 B1, 2015.12.21

KR 20140083556 A, 2014.07.04

KR 20180043922 A, 2018.05.02

审查员 田亚平

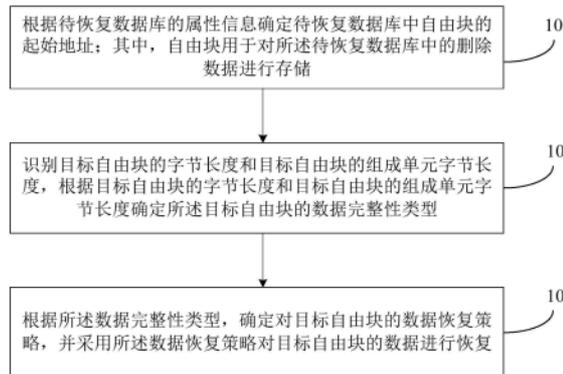
权利要求书2页 说明书11页 附图3页

(54) 发明名称

数据库删除数据的恢复方法、装置、设备和存储介质

(57) 摘要

本发明实施例公开了一种数据库删除数据的恢复方法、装置、设备和存储介质。该数据库删除数据的恢复方法包括：根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址；其中，自由块用于对待恢复数据库中的删除数据进行存储；识别目标自由块的字节长度和目标自由块的组成单元字节长度，根据目标自由块的字节长度和目标自由块的组成单元字节长度确定目标自由块的数据完整性类型；根据数据完整性类型，确定对目标自由块的数据恢复策略，并采用数据恢复策略对目标自由块的数据进行恢复。对自由块数据完整性的判断步骤简单，可复用性强，通过针对性的恢复策略实现对删除数据最大程度上的恢复，应用性强。



1. 一种数据库删除数据的恢复方法,其特征在于,包括:

根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;其中,所述目标自由块的组成单元包括:头字段、逻辑字段和数据字段,逻辑字段中由两种类型的单元组成,一种是HeaderSize,其表示自身以及逻辑字段中剩余单元所占字节数;另一种是Type单元,在逻辑字段中共有N个Type单元,并且在HeaderSize单元后顺序存储,TypeN表示数据库表中对应数据字段所占存储空间大小;所述目标自由块的数据完整性类型用于对Type单元的内容完整性进行表示;

根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

2. 根据权利要求1所述的方法,其特征在于,所述目标自由块的组成单元包括:头字段、逻辑字段和数据字段;

其中,所述头字段的字节长度为设定区间值;所述逻辑字段的各个单元用于对应描述所述数据字段的各个单元的字节长度。

3. 根据权利要求2所述的方法,其特征在于,识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型,包括:

若所述头字段的字节长度、所述逻辑字段的字节长度以及数据字段的字节长度之和等于目标自由块的字节长度,则确定所述目标自由块的数据完整性类型为第一类型;

若所述头字段的字节长度最小值、所述逻辑字段的字节长度以及数据字段的字节长度之和小于等于目标自由块的字节长度,则确定所述目标自由块的数据完整性类型为第二类型;

否则,确定所述目标自由块的数据完整性类型为第三类型。

4. 根据权利要求3所述的方法,其特征在于,根据所述数据完整性类型,确定对目标自由块的数据恢复策略,包括:

若所述目标自由块的数据完整性类型为第一类型,确定对目标自由块的数据恢复策略为顺序解码恢复;

若所述目标自由块的数据完整性类型为第二类型,确定对目标自由块的数据恢复策略为逆向解码恢复;

若所述目标自由块的数据完整性类型为第三类型,确定对目标自由块的数据恢复策略为模糊解码恢复。

5. 根据权利要求1所述的方法,其特征在于,根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址,包括:

根据数据库的存储格式要求获取待恢复数据库中的属性信息;其中,属性信息至少包括数据表信息;

根据所述数据表信息从所述待恢复数据库中确定目标叶子页信息;

根据所述目标叶子页信息确定自由块的起始地址。

6. 根据权利要求5所述的方法,其特征在于,根据所述数据表信息从所述待恢复数据库中确定目标叶子页信息,包括:

根据所述数据表信息确定待恢复数据库中所有数据表的根页编号;

根据预先存储的根页编号与叶子页之间的对应关系确定与所述根页编号关联的所有叶子页编号;

根据所述叶子页编号确定候选叶子页,并将所述候选叶子页中有存储记录数据的叶子页,作为目标叶子页。

7. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

获取待恢复数据库中的目标未使用空间;其中,目标未使用空间为非空;

采用模糊解码恢复策略对目标未使用空间的数据进行恢复。

8. 一种数据库删除数据的恢复装置,其特征在于,包括:

自由块确定模块,用于根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

数据完整性类型确定模块,用于识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;其中,所述目标自由块的组成单元包括:头字段、逻辑字段和数据字段,逻辑字段中由两种类型的单元组成,一种是HeaderSize,其表示自身以及逻辑字段中剩余单元所占字节数;另一种是Type单元,在逻辑字段中共有N个Type单元,并且在HeaderSize单元后顺序存储,TypeN表示数据库表中对应数据字段所占存储空间大小;所述目标自由块的数据完整性类型用于对Type单元的内容完整性进行表示;

数据恢复策略确定模块,用于根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

9. 一种设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-7中任一所述的数据库删除数据的恢复方法。

10. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1-7中任一所述的数据库删除数据的恢复方法。

数据库删除数据的恢复方法、装置、设备和存储介质

技术领域

[0001] 本发明实施例涉及数据恢复技术领域,尤其涉及一种数据库删除数据的恢复方法、装置、设备和存储介质。

背景技术

[0002] SQLite是一款轻型的数据库,它的设计目标是嵌入式的,目前已经在很多嵌入式产品中使用,占用资源非常低,在嵌入式设备中只需要几百KB的内存.SQLite以资源占用少、操作简洁、零管理成本、性能良好、代码开源等一系列特性广泛的应用于嵌入式设备、桌面应用以及Websites等领域。然而,一旦基于SQLite数据库的系统发生了数据丢失的情况,往往会给用户带来诸多不便。

[0003] 现有的进行SQLite数据库恢复的方法有预先设置若干种SQLite页特征规则,通过以SQLite页特征规则作为匹配依据对搜索区域内的数据进行匹配,获得当前移动终端上已删除的SQLite文件页碎片;最后对当前移动终端上已删除的SQLite文件页碎片内的数据进行解码。

[0004] 然而以页作为特征规则,使得数据恢复的规则筛选范围过大,恢复数据的精细化程度不高,导致数据恢复仍然不够全面,完整性也有待提高,且运行效率低下。

发明内容

[0005] 本发明实施例提供一种数据库删除数据的恢复方法、装置、设备和存储介质,通过对数据库中存放删除数据的自由块中数据完整性的判断,实现对数据库中已删除数据的准确恢复。

[0006] 第一方面,本发明实施例提供了一种数据库删除数据的恢复方法,包括:

[0007] 根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

[0008] 识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;

[0009] 根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

[0010] 第二方面,本发明实施例还提供了一种数据库删除数据的恢复装置,包括:

[0011] 自由块确定模块,用于根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

[0012] 数据完整性类型确定模块,用于识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;

[0013] 数据恢复策略确定模块,用于根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

- [0014] 第三方面,本发明实施例还提供了一种设备,包括:
- [0015] 一个或多个处理器;
- [0016] 存储装置,用于存储一个或多个程序,
- [0017] 当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如本发明任一实施例所述的数据库删除数据的恢复方法。
- [0018] 第四方面,本发明实施例还提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如本发明任一实施例所述的数据库删除数据的恢复方法。
- [0019] 本发明实施例通过对待恢复数据库的属性进行分析,确定存储了删除数据的自由块位置;并根据对自由块组成单元字节长度以及内容的识别,快速确定自由块的数据完整性;根据自由块的数据完整性采用对应的恢复实现对删除数据的恢复。对自由块数据完整性的判断步骤简单,可复用性广,通过针对性的恢复策略实现对删除数据最大程度上的恢复,应用性强。

附图说明

- [0020] 图1是本发明实施例一中的数据库删除数据的恢复方法的流程图;
- [0021] 图2(a)是本发明实施例一中的数据库完整性类型为第一类型的自由块的组成单元数据结构图;
- [0022] 图2(b)是本发明实施例一中的数据库完整性类型为第二类型的自由块的组成单元数据结构图;
- [0023] 图3是本发明实施例二中的数据库删除数据的恢复方法的流程图;
- [0024] 图4是本发明实施例三中的数据库删除数据的恢复装置的结构示意图;
- [0025] 图5是本发明实施例四中的设备的结构示意图。

具体实施方式

[0026] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释本发明,而非对本发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与本发明相关的部分而非全部结构。

[0027] 实施例一

[0028] 图1是本发明实施例一中的数据库删除数据的恢复方法的流程图,本实施例可适用于对SQLite数据库中的数据删除后需要在短时间内实现数据恢复的情况。该方法可以由数据库删除数据的恢复装置来执行,该装置可以采用软件和/或硬件的方式实现,并可配置在设备中,例如设备可以是后台服务器等具有通信和计算能力的设备。如图1所示,该方法具体包括:

[0029] 步骤101、根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储。

[0030] 其中,待恢复数据库是指在对数据库进行实际应用操作时随机性删除了一些数据的数据库。在本发明实施例中,待恢复数据库是指SQLite数据库,由于其应用于嵌入式开发领域,并且其具有轻型以及处理速度快等特点,因此对SQLite数据库中误删除数据可以进

行快速恢复。自由块是指是被删除的数据库中的数据单元集合,里面包含了一个或若干个被删除的数据单元及数据单元间的碎片,自由块会进入自由块链表等待回收。

[0031] 具体的,由于数据库是按照固有的编码格式对数据库内容进行编码存储,因此根据数据库存储的属性信息可以定位自由块的位置。并且由于自由块在数据库中以自由块链表的形式存在,当确定自由块的起始地址后,即可确定数据库中所有自由块的地址。

[0032] 步骤102、识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型。

[0033] 其中,目标自由块是指当前恢复的自由块,例如在对待恢复数据库进行开始恢复时,目标自由块为自由块链表中的第一个自由块,当第一个自由块数据恢复完成后,目标自由块移动到下一个自由块,并以此类推。自由块的字节长度是指该自由块的整体所占大小,根据数据库的属性该自由块的第3个字节和第4个字节可以确定该自由块的字节长度。

[0034] 所述目标自由块的组成单元包括:头字段、逻辑字段和数据字段;

[0035] 其中,所述头字段的字节长度为设定区间值;所述逻辑字段的各个单元用于对应描述所述数据字段的各个单元的字节长度。

[0036] 如图2所示为自由块的组成单元数据结构图。其中,头字段中包括两个单元,可以表示为CellSize和RowId,CellSize表示单元内存所占的字节长度,根据数据库的属性可以判断CellSize所占的存储空间大小最大为2字节;RowId表示单元数据所占的行数,根据SQLite数据库实际应用场景,记录行数不会超过3个字节所表示的最大整数值,所以判断RowId最大占用字节不会超过3个字节。

[0037] 逻辑字段中由两种类型的单元组成,一种是HeaderSize,其表示自身以及逻辑字段中剩余单元所占字节数;另一种是Type单元,在逻辑字段中共有N个Type单元,并且在HeaderSize单元后顺序存储。TypeN表示数据库表中对应数据字段所占存储空间大小。根据数据库属性可知单元内容数据大小小于页大小,表中字段大小理应小于页的大小,因此,单个Type单元最大占用2字节。并且根据数据库属性可知HeaderSize单元所占空间大小为1个字节。

[0038] 数据字段中存放了被删除的有效数据内容,在一个自由块中与N个Type单元对应的有N个Data单元,Type单元中存放了对应Data单元所占空间大小。并且在自由块存储中,CellSize、RowId、HeaderSize、Type单元以及Data单元顺序存储。便于根据各自所占字节数确定单元类型。

[0039] 综上所述,在自由块中,除了Type单元和Data单元,其他三种类型的单元所占存储空间大小为3至6个字节。因此,根据自由块中依次确定的单元类型可以得到自由块组成单元字节长度,再结合识别得到的目标自由块的字节长度,可以确定自由块内数据的完整性。若自由块中数据不完整时,则说明Type单元和Data单元缺失,则无法确定该单元的字节长度,此时与自由块的字节长度不匹配。

[0040] 在本发明实施例中,可选的,识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型,包括:

[0041] 若所述头字段的字节长度、所述逻辑字段的字节长度以及数据字段的字节长度之

和等于目标自由块的字节长度,则确定所述目标自由块的数据完整性类型为第一类型;

[0042] 若所述头字段的字节长度最小值、所述逻辑字段的字节长度以及数据字段的字节长度之和小于等于目标自由块的字节长度,则确定所述目标自由块的数据完整性类型为第二类型;

[0043] 否则,确定所述目标自由块的数据完整性类型为第三类型。

[0044] 具体的,根据目标自由块的第3个字节和第4个字节可以确定自由块的字节长度,用Size(FreeBlock)表示。对于自由块中的HeaderSize、RowId、CellSize三个单元所占字节长度用m表示,并且根据上述说明, $m \in \{4, 5, 6\}$ 。示例性的,头字段的字节长度、逻辑字段的字节长度以及数据字段的字节长度之和等于目标自由块的字节长度可以用下面的等式进行表示:

$$[0045] \quad m + \sum_{i=1}^N \text{Length}(\text{Type}_i) + \sum_{i=1}^N \text{Size}(\text{Type}_i) = \text{Size}(\text{FreeBlock}), \quad m \in \{4, 5, 6\}$$

[0046] 其中,Length(Type_i)为Type_i占用的字节个数,Size(Type_i)为与Type_i对应的数据内容Data_i的大小。若该等式成立,表示N个Type单元完整,其中Type单元根据m+1字节后依次确定;若等式不成立,则说明Type单元缺失。当该等式成立时,确定所述目标自由块的数据完整性类型为第一类型,对应的数据完整性类型为第一类型的自由块的组成单元数据结构图如图2(a)所示,灰色部分表示m所占的字节数,对于该种类型的目标自由块Type₁单元完整。

[0047] 在上述等式不成立时,说明Type单元缺失。而在Type单元缺失的情况中存在一种情况:Type₁在删除数据时候,可能发生变化。即自由块内容中头字段占用3个字节,而Type₁可能已经被覆盖。在这种情况下,自由块字节长度和各组成单元的字节长度满足如下不等式:

$$[0048] \quad 4 + \sum_{i=2}^N \text{Length}(\text{Type}_i) + \sum_{i=2}^N \text{Size}(\text{Type}_i) \leq \text{Size}(\text{FreeBlock})$$

[0049] 示例性的,在上述不等式成立的基础上,说明Type₁被覆盖。目标自由块中缺少Type₁的内容,则在公式计算时从第2个Type单元开始确定,从第4个字节后开始依次确定Type单元的内容,若满足此不等式,表示从第二个Type单元开始确定到第N个Type单元的内容是正确的,此时确定所述目标自由块的数据完整性类型为第二类型,对应的数据完整性类型为第二类型的自由块的组成单元数据结构图如图2(b)所示,灰色部分所占的字节数为4,对于该种类型的目标自由块Type₁单元被覆盖。

[0050] 若上述两种情况均不符合,说明自由块起始地址偏移m个字节后确定的Type单元的内容是错误的,Type单元缺失了至少两个,表明该自由块中已删除数据已经被覆盖或成块删除,此时无法准确定位缺失Type单元的位置,则确定所述目标自由块的数据完整性类型为第三类型。

[0051] 步骤103、根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

[0052] 在自由块中,Type_N对应描述了Data_N的类型和大小,恢复删除数据需要依据

TypeN,可以根据整个数据单元大小以及Type的实际组成,依次取出DataN。因此恢复的准确性更依赖于Type的准确性。通过上述对自由块数据完整性的判断,可以对自由块中Type单元进行准确快速的识别,进而依据Type单元对自由块数据的完整性进行准确判断。

[0053] 在本发明实施例中,对于不同完整性类型的自由块中的数据采用不同的数据恢复策略,进行有针对性的恢复有利于提高数据恢复的准确性,并且在确定自由块的数据完整性时所采用的依据均为数据库的固有属性信息,因此判断过程简单快捷,提高了对自由块数据完整性的判断效率。

[0054] 在本发明实施例中,可选的,根据所述数据完整性类型,确定对目标自由块的数据恢复策略,包括:

[0055] 若所述目标自由块的数据完整性类型为第一类型,确定对目标自由块的数据恢复策略为顺序解码恢复;

[0056] 若所述目标自由块的数据完整性类型为第二类型,确定对目标自由块的数据恢复策略为逆向解码恢复;

[0057] 若所述目标自由块的数据完整性类型为第三类型,确定对目标自由块的数据恢复策略为模糊解码恢复。

[0058] 具体的,当目标自由块的数据完整性类型为第一类型时,说明自由块中各Type单元的内容是完整的,则可以依据Type单元中的内容对应取出Data单元中的数据进行恢复。示例性的,确定使等式成立的m值,确定m+1字节后的Type1单元,在Type1单元中存储了Data1单元的所占字节数,则根据TypeN单元后对应字节数即可确定Data1单元中的删除内容。以此类推,通过解码的方式读取Data单元中的删除数据,可逐个恢复Data单元中的数据,并且可保证对目标自由块中删除内容的精确恢复。

[0059] 当目标自由块的数据完整性类型为第二类型时,说明自由块中除了Type1单元以外各Type单元的内容是完整的,则采用逆向解码恢复的策略进行恢复。示例性的,先读取DataN中的内容,再依次读取Data2中的内容,最后根据确定的TypeN与Data2的位置可确定两者之间的数据即为Data1的内容。通过逆向解码恢复的策略可以实现对自由块中删除内容的精确恢复。

[0060] 当目标自由块的数据完整性类型为第三类型时,说明自由块中的内容缺失程度严重,这种情况下是无法进行精确恢复的,此时需要采用整体模糊解码恢复的方法,对自由块中的整体内容进行解码恢复,然后采用语义分析进行提炼,从中获取有用信息,尽可能提高对自由块中删除内容的恢复准确度。

[0061] 在一个可选的实施例中,可选的,所述方法还包括:

[0062] 获取待恢复数据库中的目标未使用空间;其中,目标未使用空间为非空;

[0063] 采用模糊解码恢复策略对目标未使用空间的数据进行恢复。

[0064] 对于数据库中删除数据除了存储在自由块中,还可能会存储在未使用空间中。对于未使用空间,除了数据库正在使用的内存空间以及自由块链表所占空间外为未使用空间。确定未使用空间中内容为非空的为目标未使用空间,目标未使用空间中存储待恢复数据库中的删除内容。

[0065] 对于目标非使用空间使用模糊解码恢复的策略进行恢复,对目标非使用空间中的整体内容进行解码恢复,然后采用语义分析进行提炼,从中获取有用信息,尽可能提高对目

标非使用空间中删除内容的恢复准确度。

[0066] 通过以上方式实现对数据库中删除数据的恢复,优势在于可以根据逻辑字段的判断,快速确定自由块是否完整,从而可快速根据对应恢复策略实现删除数据的恢复。判断方式简单快捷,可复用性广,应用性强。尤其对于用户误删除数据后,此时数据存储在自由块中且被覆盖的可能性较低,采用本发明实施例中的恢复方法可以实现对删除数据的快速准确恢复,提高数据恢复效率。

[0067] 本发明实施例通过对待恢复数据库的属性进行分析,确定存储了删除数据的自由块位置;并根据对自由块组成单元字节长度以及内容的识别,快速确定自由块的数据完整性;根据自由块的数据完整性采用对应的恢复实现对删除数据的恢复。对自由块数据完整性的判断步骤简单,可复用性广,通过针对性的恢复策略实现对删除数据最大程度上的恢复,应用性强。

[0068] 实施例二

[0069] 图3是本发明实施例二中的数据库删除数据的恢复方法的流程图,本实施例二在实施例一的基础上进行进一步地优化。如图3所示,所述方法包括:

[0070] 步骤301、根据数据库的存储格式要求获取待恢复数据库中的属性信息;其中,属性信息至少包括数据表信息。

[0071] 其中,数据库是按照一定的存储格式对文件属性以及内容进行存储。示例性的,SQLite数据库把每个数据库都存储到一个单独的文件中。SQLite数据库由固定大小的页(page)组成。页的大小可以在512~32768之间(包含这两个值,必须是2的指数),默认大小为1024个字节(1KB)。页大小可以在数据库刚刚创建时设置,一旦创建了数据库对象之后,这个值就不能再改变了。数据库中所有的页从1开始顺序编号。在具体的实现中,页号用4字节来表示,并限制最大页号不得超过 2^{31} 。第1个页被称为page1,第二个页被称为page2,以此类推。

[0072] 每个数据库可以由至少一个数据表组成,数据表信息可以通过数据库中系统表sqlite_master获取对应数据页的相关字段。系统表sqlite_master保存了数据库的概要信息。示例性的,可以根据表中第1个字段,类型Type确定数据表,通过第2个字段确定表名,通过第4字段确定表的RootPage(根页编号),通过第5个字段可以确定数据表Type字段的数量。可选的,属性信息还包括页大小、数据库大小以及数据库编码方式。示例性的,文件头偏移16字节后的2字节代表PageSize(页大小),偏移28字节后的4字节代表数据库大小,其中数据库大小以页为单位,偏移56字节后的4字节为数据库编码方式。其中,数据库编码方式用于模糊解码恢复策略中。

[0073] 由此可知,数据库的属性信息不局限与上述所描述出的内容,根据数据库的文件存储内容以及存储格式可以确定其他相关信息,在此不作赘述。

[0074] 在一个可行的实施例中,在根据数据库的存储格式要求获取待恢复数据库中的属性信息之前,还包括:确定数据库的类型。

[0075] 具体的,对数据库进行判断,确定该文件是否为SQLite数据库,可以根据文件头前端的16个字节来判断,以SQLite3为例,该内容应为“SQLite format3\000”。

[0076] 步骤302、根据所述数据表信息从所述待恢复数据库中确定目标叶子页信息。

[0077] 对于数据库中页的类型可以是:Btree页、空闲页或溢出页。一个数据库中可能没

有空闲页或溢出页,但必然有Btree页。Btree为一种树结构,又可以分为B-tree或B+tree,每一种树的结点又区分为内部页和叶子页。一个SQLite数据库由多个多重Btree构成。每个Btree占用至少一个完整的页,每个页是Btree的一个结点。每个表或索引的第1个页称为根页,所有表或索引的根页编号都存储在系统表sqlite_master中,并且其中系统表sqlite_master的根页为page1.SQLite数据库中第一个页永远都是Btree页。page1的前100个字节是一个对数据库进行描述的头文件。它包括数据库的版本、格式的版、页大小、编码等。

[0078] 自由块的起始地址与数据库中叶子页信息有关,因此需要确定数据库中目标叶子页信息,其中,目标叶子页为数据库中有存储记录数据的叶子页,目标叶子页信息包括叶子页的起始地址信息。

[0079] 在一个可行的实施例中,可选的,根据所述数据表信息从所述待恢复数据库中确定目标叶子页信息,包括:

[0080] 根据所述数据表信息确定待恢复数据库中所有数据表的根页编号;

[0081] 根据预先存储的根页编号与叶子页之间的对应关系确定与所述根页编号关联的所有叶子页编号;

[0082] 根据所述叶子页编号确定候选叶子页,并将所述候选叶子页中有存储记录数据的叶子页,作为目标叶子页。

[0083] 具体的,根据上述确定了数据库中数据表的RootPage,根据该编号可以确定数据表的所有叶子页,并判断该叶子页是否为有存储记录数据的B+tree叶子页,示例性的,判断方式可采用第一个字节是否为“0x0D”确定。根据上述判断方式确定含有存储记录数据的所有B+tree叶子页,作为目标叶子页。

[0084] 对于目标叶子页信息中的每个叶子页的起始地址可以根据数据表信息中的页大小进行确定。示例性的,根据如下公式确定叶子页的起始地址:

[0085] $PageAddress = (PageNum - 1) \times PageSize, PageNum \in \{4, 5, 6\}$

[0086] 其中,PageNum为页编号,根据数据库的属性可知,其取值区间在{4, 5, 6}中,PageSize为页大小。根据该公式可以确定所有目标叶子页的起始地址。

[0087] 步骤303、根据所述目标叶子页信息确定自由块的起始地址。

[0088] 由于自由块是以自由块链表的形式存在,并且首个自由块的地址可以根据叶子页偏移一个字节后的两个字节中的内容进行确定。因此根据确定出的所有目标叶子页信息中的叶子页地址偏移后的内容可以找到首个自由块的起始地址。

[0089] 步骤304、识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型。

[0090] 步骤305、根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

[0091] 本发明实施例通过对数据库存储格式的分析,获取其属性信息为准确定位自由块的地址提供依据。并且简单的确定方法可以提高自由块地址的确定效率,进而提高数据恢复的效率。

[0092] 实施例三

[0093] 图4是本发明实施例三中的数据库删除数据的恢复装置的结构示意图,本实施例

可适用于对SQLite数据库中的删除数据进行恢复的情况。如图4所示,该装置包括:

[0094] 自由块确定模块410,用于根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

[0095] 数据完整性类型确定模块420,用于识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;

[0096] 数据恢复策略确定模块430,用于根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

[0097] 本发明实施例通过对待恢复数据库的属性进行分析,确定存储了删除数据的自由块位置;并根据对自由块组成单元字节长度以及内容的识别,快速确定自由块的数据完整性;根据自由块的数据完整性采用对应的恢复实现对删除数据的恢复。对自由块数据完整性的判断步骤简单,可复用性广,通过针对性的恢复策略实现对删除数据最大程度上的恢复,应用性强。

[0098] 可选的,所述目标自由块的组成单元包括:头字段、逻辑字段和数据字段;

[0099] 其中,所述头字段的字节长度为设定区间值;所述逻辑字段的各个单元用于对应描述所述数据字段的各个单元的字节长度。

[0100] 可选的,数据完整性类型确定模块420,包括:

[0101] 第一类型确定单元,用于若所述头字段的字节长度、所述逻辑字段的字节长度以及数据字段的字节长度之和等于目标自由块的字节长度,则确定所述目标自由块的数据完整性类型为第一类型;

[0102] 第二类型确定单元,用于若所述头字段的字节长度最小值、所述逻辑字段的字节长度以及数据字段的字节长度之和小于等于目标自由块的字节长度,则确定所述目标自由块的数据完整性类型为第二类型;

[0103] 第三类型确定单元,用于否则,确定所述目标自由块的数据完整性类型为第三类型。

[0104] 可选的,数据恢复策略确定模块430,具体用于:

[0105] 若所述目标自由块的数据完整性类型为第一类型,确定对目标自由块的数据恢复策略为顺序解码恢复;

[0106] 若所述目标自由块的数据完整性类型为第二类型,确定对目标自由块的数据恢复策略为逆向解码恢复;

[0107] 若所述目标自由块的数据完整性类型为第三类型,确定对目标自由块的数据恢复策略为模糊解码恢复。

[0108] 可选的,自由块确定模块410,包括:

[0109] 属性信息获取单元,用于根据数据库的存储格式要求获取待恢复数据库中的属性信息;其中,属性信息至少包括数据表信息;

[0110] 目标叶子页信息确定单元,用于根据所述数据表信息从所述待恢复数据库中确定目标叶子页信息;

[0111] 起始地址确定单元,用于根据所述目标叶子页信息确定自由块的起始地址。

[0112] 可选的,目标叶子页信息确定单元,具体用于:

- [0113] 根据所述数据表信息确定待恢复数据库中所有数据表的根页编号；
- [0114] 根据预先存储的根页编号与叶子页之间的对应关系确定与所述根页编号关联的所有叶子页编号；
- [0115] 根据所述叶子页编号确定候选叶子页,并将所述候选叶子页中有存储记录数据的叶子页,作为目标叶子页。
- [0116] 可选的,所述装置还包括目标未使用空间数据恢复模块,具体用于:
- [0117] 获取待恢复数据库中的目标未使用空间;其中,目标未使用空间为非空;
- [0118] 采用模糊解码恢复策略对目标未使用空间的数据进行恢复。
- [0119] 本发明实施例所提供的数据库删除数据的恢复装置可执行本发明任意实施例所提供的数据库删除数据的恢复方法,具备执行数据库删除数据的恢复方法相应的功能模块和有益效果。
- [0120] 实施例四
- [0121] 图5是本发明实施例四提供的一种设备的结构示意图。图5示出了适于用来实现本发明实施方式的示例性设备12的框图。图5显示的设备12仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。
- [0122] 如图5所示,设备12以通用计算设备的形式表现。设备12的组件可以包括但不限于:一个或者多个处理器或者处理单元16,系统存储装置28,连接不同系统组件(包括系统存储装置28和处理单元16)的总线18。
- [0123] 总线18表示几类总线结构中的一种或多种,包括存储装置总线或者存储装置控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构 (ISA) 总线,微通道体系结构 (MAC) 总线,增强型ISA总线、视频电子标准协会 (VESA) 局域总线以及外围组件互连 (PCI) 总线。
- [0124] 设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被设备12访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。
- [0125] 系统存储装置28可以包括易失性存储装置形式的计算机系统可读介质,例如随机存取存储装置 (RAM) 30和/或高速缓存存储装置32。设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统34可以用于读写不可移动的、非易失性磁介质(图5未显示,通常称为“硬盘驱动器”)。尽管图5中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如CD-ROM, DVD-ROM或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线18相连。存储装置28可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本发明各实施例的功能。
- [0126] 具有一组(至少一个)程序模块42的程序/实用工具40,可以存储在例如存储装置28中,这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本发明所描述的实施例中的功能和/或方法。
- [0127] 设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信,

还可与一个或者多个使得用户能与该设备12交互的设备通信,和/或与使得该设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且,设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图5所示,网络适配器20通过总线18与设备12的其它模块通信。应当明白,尽管图5中未示出,可以结合设备12使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0128] 处理单元16通过运行存储在系统存储装置28中的程序,从而执行各种功能应用以及数据处理,例如实现本发明实施例所提供的数据库删除数据的恢复方法,包括:

[0129] 根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

[0130] 识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;

[0131] 根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

[0132] 实施例五

[0133] 本发明实施例五还提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如本发明实施例所提供的数据库删除数据的恢复方法,包括:

[0134] 根据待恢复数据库的属性信息确定待恢复数据库中自由块的起始地址;其中,自由块用于对所述待恢复数据库中的删除数据进行存储;

[0135] 识别目标自由块的字节长度和目标自由块的组成单元字节长度,根据目标自由块的字节长度和目标自由块的组成单元字节长度确定所述目标自由块的数据完整性类型;

[0136] 根据所述数据完整性类型,确定对目标自由块的数据恢复策略,并采用所述数据恢复策略对目标自由块的数据进行恢复。

[0137] 本发明实施例的计算机存储介质,可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0138] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0139] 计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括——但不限

于无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0140] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0141] 注意,上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解,本发明不限于这里所述的特定实施例,对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此,虽然通过以上实施例对本发明进行了较为详细的说明,但是本发明不仅仅限于以上实施例,在不脱离本发明构思的情况下,还可以包括更多其他等效实施例,而本发明的范围由所附的权利要求范围决定。

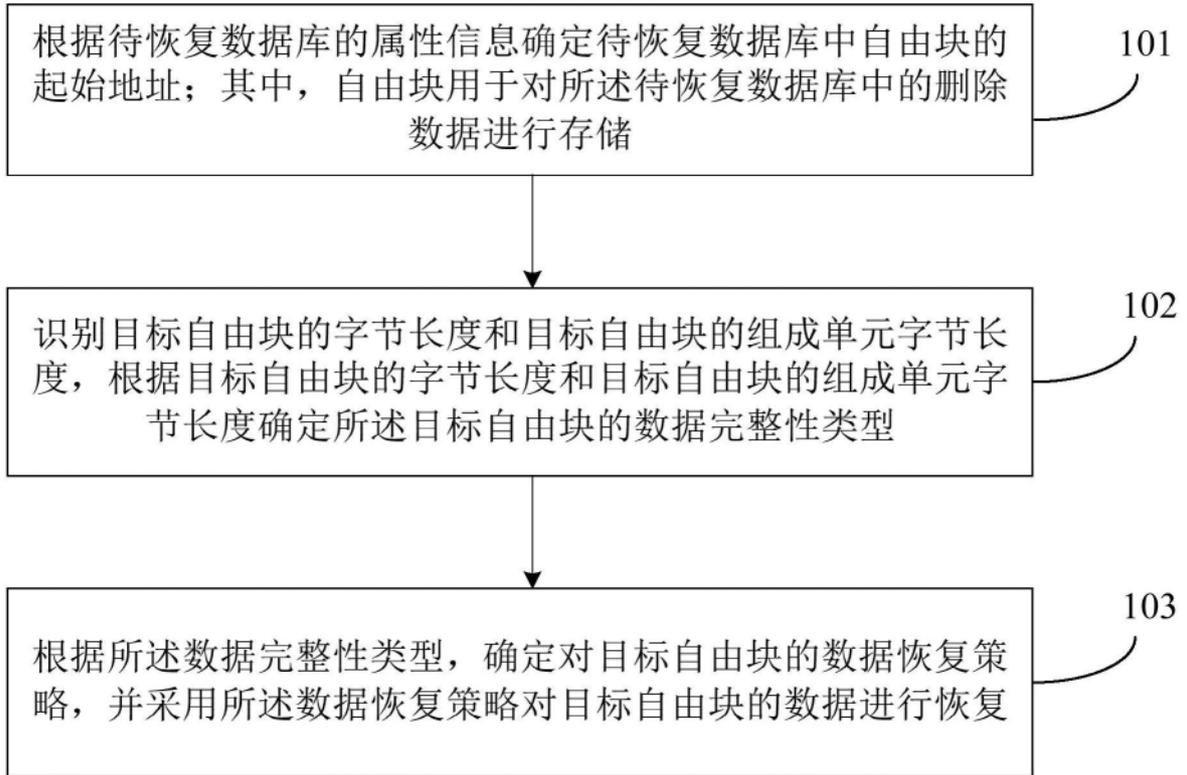


图1



图2(a)



图2(b)

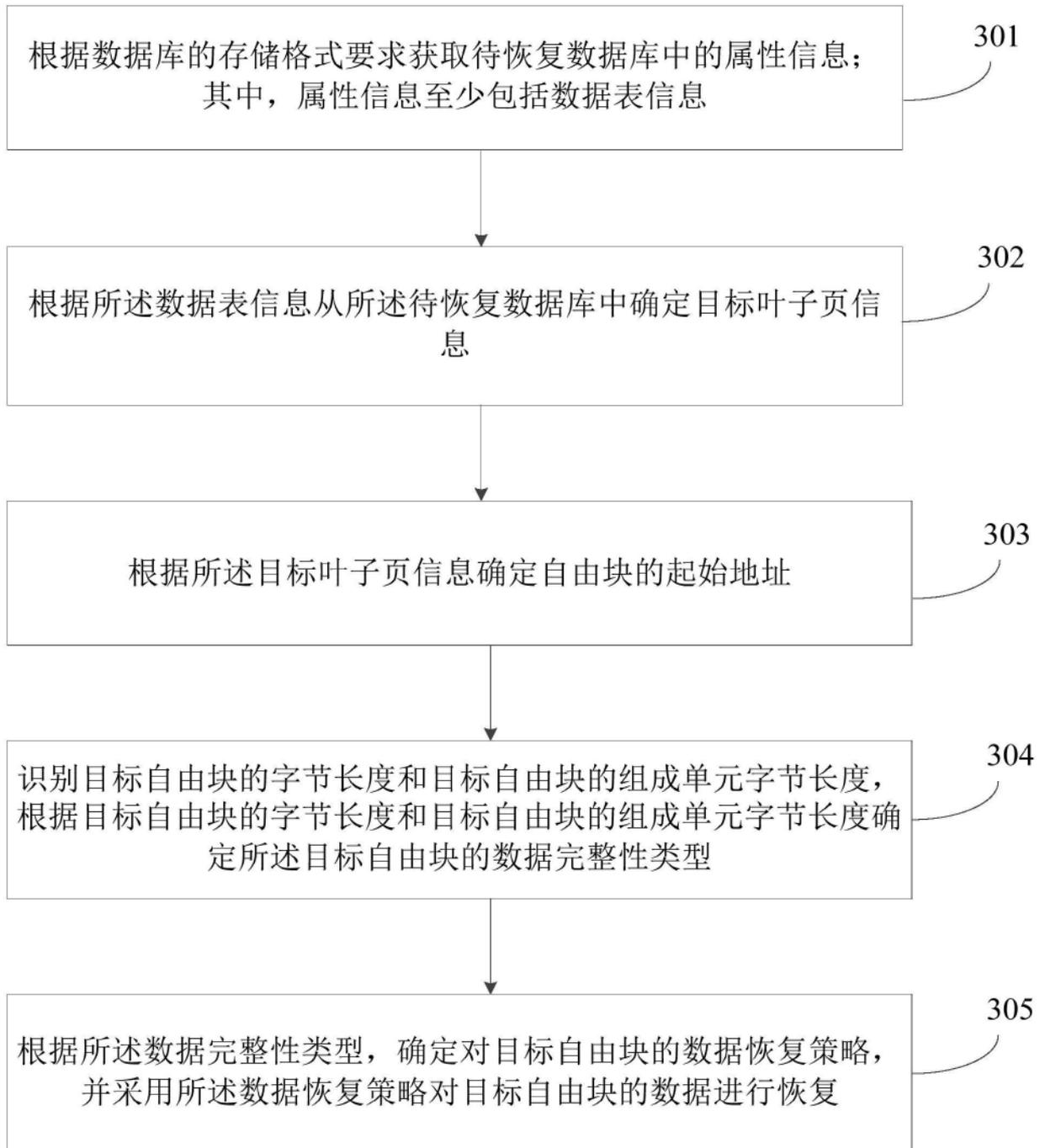


图3



图4

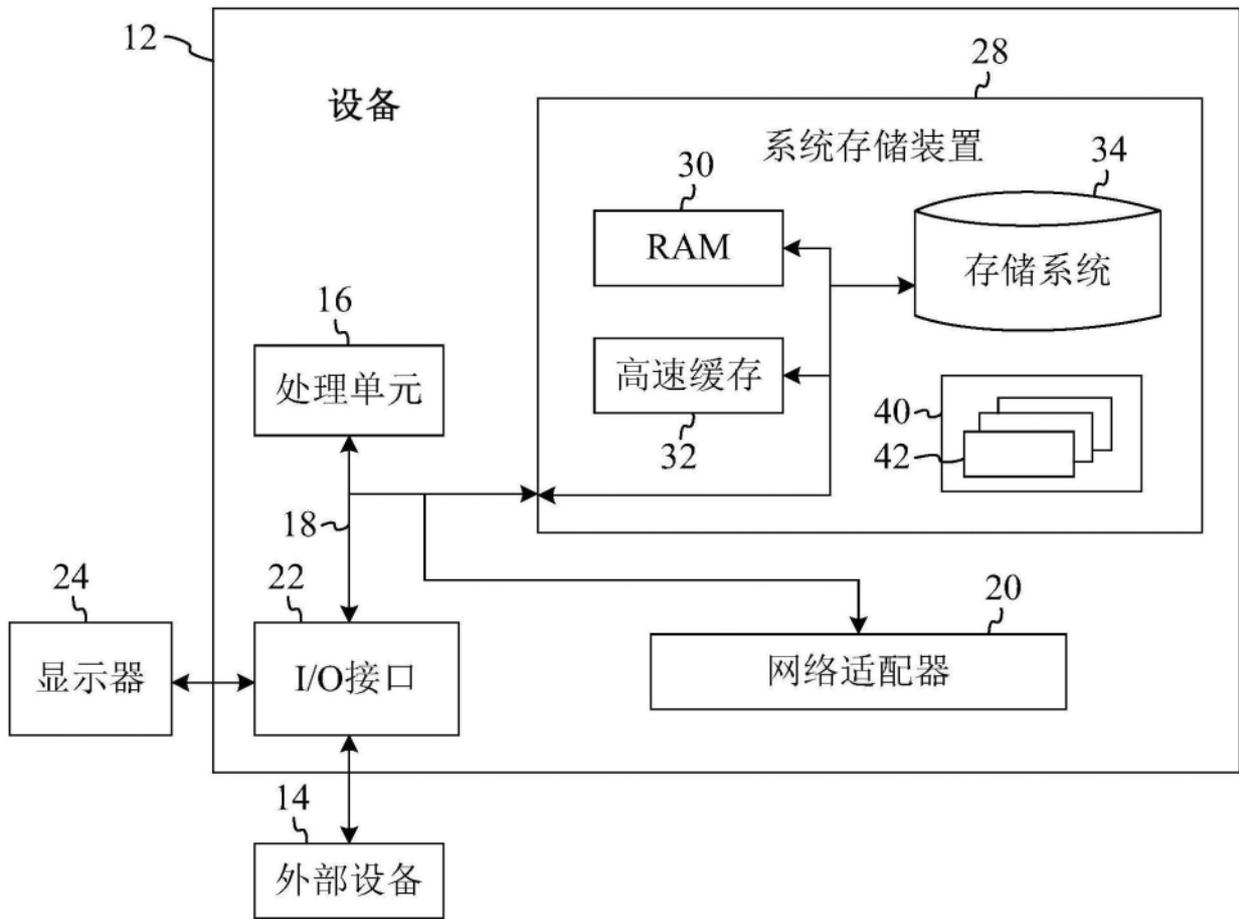


图5