



(12)发明专利申请

(10)申请公布号 CN 109101522 A

(43)申请公布日 2018.12.28

(21)申请号 201810609379.4

(22)申请日 2018.06.13

(30)优先权数据

62/522,423 2017.06.20 US

62/542,008 2017.08.07 US

15/728,261 2017.10.09 US

(71)申请人 三星电子株式会社

地址 韩国京畿道水原市灵通区三星路129号

(72)发明人 奇亮奭 李勇锡 杰森·马蒂诺

(74)专利代理机构 广州华进联合专利商标代理有限公司 44224

代理人 刘培培 黄隶凡

(51)Int.Cl.

G06F 17/30(2006.01)

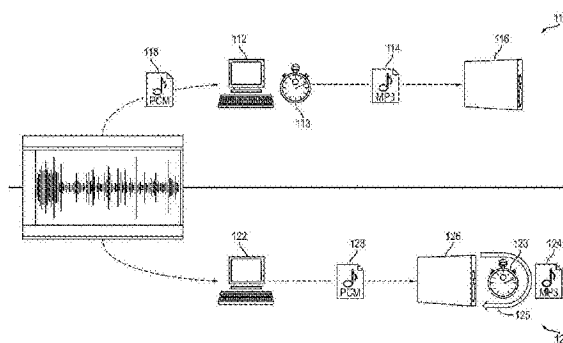
权利要求书3页 说明书12页 附图2页

(54)发明名称

数据压缩系统、有损数据压缩的方法和数据压缩的方法

(57)摘要

一种用于有损数据压缩的方法,所述方法包括:在存储装置处接收原始数据;接收要对旗标进行压缩的请求;存取板载数据压缩算法库,其包含分别与有损数据压缩方案对应的各种数据压缩算法;基于多个参数来选择所述数据压缩算法中的一者;在线运行所选择的所述数据压缩算法,使得当接收到所述原始数据时所述原始数据被所述存储装置压缩,且接着所述原始数据被作为压缩数据存储在该所述存储装置上;或者离线运行所选择的所述数据压缩算法,使得所述原始数据存储在该所述存储装置处,此后由所述存储装置根据所选择的所述数据压缩算法来压缩所述原始数据,且所述原始数据被作为压缩数据重新保存在所述存储装置处。也提供数据压缩系统和数据压缩的方法。



1. 一种用于有损数据压缩的方法,其特征在于,所述方法包括:
在存储装置处接收原始数据;
接收请求压缩旗标;
存取板载数据压缩算法库,所述板载数据压缩算法库包含分别与有损数据压缩方案对应的各种数据压缩算法;
基于多个参数来选择所述数据压缩算法中的一者;
以以下方式中的一者运行所选择的所述数据压缩算法:
在线运行,使得当接收到所述原始数据时所述原始数据被所述存储装置压缩,且接着所述原始数据被作为经压缩的数据存储在所述存储装置上;或者
离线运行,使得所述原始数据存储在所述存储装置处,此后由所述存储装置根据所选择的所述数据压缩算法来压缩所述原始数据,且所述原始数据被作为经压缩的数据重新保存在所述存储装置处。
2. 根据权利要求1所述的方法,其特征在于,所述参数包含在与所述原始数据对应的密钥值对的增强密钥中,或者包含在与所述存储装置相关联的服务品质偏好数据文件中。
3. 根据权利要求1所述的方法,其特征在于,还包括提醒请求压缩所述原始数据的用户提供所述参数。
4. 根据权利要求1所述的方法,其特征在于,还包括:
读取包含在所述原始数据中的与对象签名或对象扩展对应的元数据;
存取所述存储装置上的对象签名词典或所述存储装置上的对象扩展词典,以分别基于所述对象签名或所述对象扩展确定所述原始数据的数据类型;以及
基于所确定的所述数据类型来选择所述数据压缩算法。
5. 一种数据压缩方法,其特征在于,所述方法包括:
在存储装置处接收原始数据;
使用所述存储装置根据有损数据压缩方案对所述原始数据进行压缩,以产生经压缩的数据;以及
将所述经压缩的数据存储在所述存储装置上。
6. 根据权利要求5所述的方法,其特征在于,还包括使用所述存储装置确定用于对所述原始数据进行压缩的数据压缩方案。
7. 根据权利要求6所述的方法,其特征在于,还包括:
建立包含与压缩方案及参数有关的信息的服务品质偏好数据文件;
由所述存储装置识别与所述原始数据对应的对象文件;以及
根据所述信息及所述对象文件来确定用于对所述原始数据进行压缩的所述数据压缩方案。
8. 根据权利要求6所述的方法,其特征在于,还包括:
读取与所述原始数据对应的密钥值对的密钥的元数据;
根据所述元数据从所述存储装置上的数据压缩算法库选择数据压缩算法;以及
根据所选择的所述数据压缩算法来执行用于对所述原始数据进行压缩的所述数据压缩方案。
9. 根据权利要求5所述的方法,其特征在于,所述原始数据包括对象签名或对象扩展,

所述方法还包括：

存取所述存储装置上的对象签名词典或所述存储装置上的对象扩展词典，以分别基于所述对象签名或所述对象扩展确定所述原始数据的数据类型；以及

基于所确定的所述数据类型来执行数据压缩方案。

10. 根据权利要求5所述的方法，其特征在于，还包括根据用户请求将所述原始数据从主机计算机发送到所述存储装置。

11. 根据权利要求10所述的方法，其特征在于，还包括：

提供用于提醒用户提供信息的菜单；以及

根据由所述用户提供的所述信息来确定用于对所述原始数据进行压缩的数据压缩算法。

12. 根据权利要求5所述的方法，其特征在于，还包括将所述原始数据存储于所述存储装置上；

估算与所述存储装置对应的可用时间或资源；以及

在确定有足够的时间及资源可用时，对所述原始数据进行压缩。

13. 一种数据压缩系统，其特征在于，包括：

主机；以及

存储装置，被配置成从所述主机接收原始数据，且包括：

存储器；以及

处理器，被配置成执行指令，所述指令在由所述处理器执行时使所述处理器：

使所述存储装置根据有损数据压缩方案对所述原始数据进行压缩以产生经压缩的数据；以及

使所述存储装置将所述经压缩的数据存储于所述存储装置上。

14. 根据权利要求13所述的系统，其特征在于，所述指令在由所述处理器执行时还使所述处理器使所述存储装置确定用于对所述原始数据进行压缩的数据压缩方案。

15. 根据权利要求14所述的系统，其特征在于，所述指令在由所述处理器执行时还使所述处理器使所述存储装置：

存取包含与压缩方案及参数有关的信息的服务品质偏好数据文件；

识别与所述原始数据对应的对象文件；以及

根据所述信息及所述对象文件来确定用于对所述原始数据进行压缩的所述数据压缩方案。

16. 根据权利要求14所述的系统，其特征在于，所述指令在由所述处理器执行时还使所述处理器使所述存储装置：

读取与所述原始数据对应的密钥值对的密钥的元数据；

根据所述元数据从数据压缩算法库选择数据压缩算法；以及

根据所选择的所述数据压缩算法来执行用于对所述原始数据进行压缩的所述数据压缩方案。

17. 根据权利要求14所述的系统，其特征在于，所述原始数据包括对象签名或对象扩展，且

其中所述指令在由所述处理器执行时还使所述处理器使所述存储装置：

存取所述存储装置上的对象签名词典或所述存储装置上的对象扩展词典,以分别基于所述对象签名或所述对象扩展确定所述原始数据的数据类型;以及
基于所确定的所述数据类型来执行所述数据压缩方案。

18. 根据权利要求13所述的系统,其特征在于,所述指令在由所述处理器执行时还使所述处理器使所述存储装置根据用户请求从主机计算机接收所述原始数据。

19. 根据权利要求18所述的系统,其特征在于,所述指令在由所述处理器执行时还使所述处理器:

使所述主机提供用于提醒正在请求对所述原始数据进行压缩的用户提供信息的菜单;
以及

使所述存储装置根据由所述用户提供的所述信息来确定用于对所述原始数据进行压缩的数据压缩算法。

20. 根据权利要求13所述的系统,其特征在于,所述指令在由所述处理器执行时还使所述处理器使所述存储装置:

将所述原始数据存储在该所述存储装置上;

估算与所述存储装置对应的可用时间或资源;以及

当有足够的时间及资源可用时,对所述原始数据进行压缩。

数据压缩系统、有损数据压缩的方法和数据压缩的方法

[0001] [相关申请的交叉参考]

[0002] 本申请主张在2017年8月7日提出申请的美国临时专利申请第62/542,008号的优先权及权利,且还主张在2017年6月20日提出申请的美国临时专利申请第62/522,423号的优先权及权利,并与同样主张所述美国临时专利申请第62/522,423号的优先权及权利且在2017年8月28日提出申请的美国非临时专利申请第15/688,779号相关,以上美国临时专利申请及美国非临时专利申请全文并入本申请供参考。

技术领域

[0003] 本公开的一些实施例大体来说涉及一种包括数据压缩方案的驱动装置。

背景技术

[0004] 数据压缩是用于提高对数据存储容量(例如硬存储装置)的有效使用的众所周知且有用的技术。然而,一些数据压缩方案或算法对于某些类型的数据而言比对于其他数据而言更行之有效。数据压缩可被称为“无损”数据压缩或“有损”数据压缩,在“无损”数据压缩中,可从经压缩版本获得准确的原始数据(original data),而在“有损”数据压缩情形中,则无法从经压缩版本恢复准确的输入数据。

[0005] 一般来说,数据压缩是在主机处发生的,主机接着将经压缩数据作为经压缩数据文件发送到驱动器(例如,硬盘驱动器(hard disk drive,HDD)或固态驱动器(solid state drive,SSD))。然而,驱动器上数据压缩(on-drive data compression)是驱动器从主机接收原始的、未经压缩的数据文件、对所述数据进行压缩且接着将经压缩的数据文件存储在驱动器上。

[0006] 之前的驱动器上数据压缩方法“笨拙地”对数据进行压缩,或者执行非智能的/过于简单化的数据压缩,其使用相同的数据压缩方法或算法对驱动器接收到的每一个原始数据文件进行压缩。之后,驱动器将经压缩的数据作为经压缩的数据文件存储在驱动器上。也就是说,非智能的驱动器上数据压缩是指驱动器取得从主机接收的每一数据,且使用相同的数据压缩算法对每一数据进行压缩,且接着将经压缩的数据文件存储在驱动器上。

[0007] 当使用非智能数据压缩时,驱动器会忽略数据的区别性特性,例如可指示以下的特性:数据是否已被主机压缩过;数据之前是否被加密过;驱动器的用户是否想要对数据进行压缩、以及想要压缩到何种程度、或者想要根据什么参数来进行压缩;以及各种其他因素。另外,当使用非智能数据压缩时,驱动器会忽略某些类型的数据与其他类型的数据相比更易于根据一种或多种特定的数据压缩方案来进行压缩的事实。

[0008] 之前实施的“笨拙的”或非智能的数据压缩版本(例如,通用数据压缩)是基于用户在力图将数据保存在驱动器上时不想采用会导致损失重要数据有损数据压缩方案这一假设来进行的。之前实施的非智能数据压缩版本因此总是涉及无损数据压缩,而非有损数据压缩。也就是说,由于有损数据压缩对特定类型的数据(例如,数据库、工作数据文件等)而言是有害的,且可能会潜在地引起数据毁坏,因此为维持数据的完整性,之前只实施无损数

据压缩方案。

[0009] 另外,之前实施的有损数据压缩方案总是由主机执行,且数据不被驱动器压缩。而是,驱动器简单地接收并存储从主机发送的经压缩的数据文件。主机通常对应于力图压缩及存储数据的用户的中央处理器(central processing unit,CPU)/计算机处理器。遗憾的是,与这种有损数据压缩相关联的计算操作通常占用大量的中央处理器循环,且还消耗对应的大量能量。

[0010] 公开以上信息仅是为了增强对本公开的背景的理解,且因此可能含有并非构成现有技术的信息。

发明内容

[0011] 本公开的实施例提供一种驱动器,所述驱动器无需主机的辅助便能够智能地执行有损数据压缩。

[0012] 根据一个实施例,提供一种用于有损数据压缩的方法,所述方法包括:在存储装置处接收原始数据;接收请求压缩旗标;存取板载数据压缩算法库,所述板载数据压缩算法库包含分别与有损数据压缩方案对应的各种数据压缩算法;基于多个参数来选择所述数据压缩算法中的一者;在线运行所述数据压缩算法,使得当接收到所述原始数据时所述原始数据被所述存储装置压缩,且接着所述原始数据被作为压缩数据存储在所述存储装置上;或者离线运行所述数据压缩算法,使得所述原始数据存储在所述存储装置处,此后由所述存储装置根据所选择的所述数据压缩算法来压缩所述原始数据,且所述原始数据被作为压缩数据重新保存在所述存储装置处。

[0013] 所述参数可包含在与所述原始数据对应的密钥值对的增强密钥中,或者包含在与所述存储装置相关联的服务品质(QoS)偏好数据文件中。

[0014] 所述方法还可包括提醒请求压缩所述原始数据的用户提供所述参数。

[0015] 所述方法还可包括:读取包含在所述原始数据中的与对象签名或对象扩展对应的元数据;存取所述存储装置上的对象签名词典或所述存储装置上的对象扩展词典,以分别基于所述对象签名或所述对象扩展确定所述原始数据的数据类型;以及基于所确定的所述数据类型来选择所述数据压缩算法。

[0016] 根据另一个实施例,提供一种数据压缩方法,所述方法包括:在存储装置处接收原始数据;使用所述存储装置根据有损数据压缩方案对所述原始数据进行压缩,以产生压缩数据;以及将经压缩的所述数据存储在所述存储装置上。

[0017] 所述方法还可包括使用所述存储装置确定用于对所述原始数据进行压缩的数据压缩方案。

[0018] 所述方法还可包括:建立包含与压缩方案及参数有关的信息的服务品质(QoS)偏好数据文件;由所述存储装置识别与所述原始数据对应的对象文件;以及根据所述信息及所述对象文件来确定用于对所述原始数据进行压缩的所述数据压缩方案。

[0019] 所述方法还可包括:读取与所述原始数据对应的密钥值对的密钥的元数据;根据所述元数据从所述存储装置上的数据压缩算法库选择数据压缩算法;以及根据所选择的所述数据压缩算法来执行用于对所述原始数据进行压缩的所述数据压缩方案。

[0020] 所述原始数据可包括对象签名或对象扩展,所述方法还包括:存取所述存储装置

上的对象签名词典或所述存储装置上的对象扩展词典,以分别基于所述对象签名或所述对象扩展确定所述原始数据的数据类型;以及基于所确定的所述数据类型来执行所述数据压缩方案。

[0021] 所述方法还可包括根据用户请求将所述原始数据从主机计算机发送到所述存储装置。

[0022] 所述方法还可包括:提供用于提醒用户提供信息的菜单;以及根据由所述用户提供的所述信息来确定用于对所述原始数据进行压缩的数据压缩算法。

[0023] 所述方法还可包括将所述原始数据存储在其所述存储装置上;估算与所述存储装置对应的可用时间或资源;以及在确定有足够的时间及资源可用时,对所述原始数据进行压缩。

[0024] 根据另一个实施例,提供一种数据压缩系统,所述数据压缩系统包括:主机;以及存储装置,被配置成从所述主机接收原始数据,且包括存储器及处理器,所述处理器被配置成执行指令,所述指令在由所述处理器执行时使所述处理器:使所述存储装置根据有损数据压缩方案对所述原始数据进行压缩以产生经压缩的数据;以及使所述存储装置将经压缩的所述数据存储在其所述存储装置上。

[0025] 所述指令在由所述处理器执行时还可使所述处理器使所述存储装置确定用于对所述原始数据进行压缩的数据压缩方案。

[0026] 所述指令在由所述处理器执行时还可使所述处理器使所述存储装置:存取包含与压缩方案及参数有关的信息的服务品质(QoS)偏好数据文件;识别与所述原始数据对应的对象文件;以及根据所述信息及所述对象文件来确定用于对所述原始数据进行压缩的所述数据压缩方案。

[0027] 所述指令在由所述处理器执行时还可使所述处理器使所述存储装置:读取与所述原始数据对应的密钥值对的密钥的元数据;根据所述元数据从数据压缩算法库选择数据压缩算法;以及根据所选择的所述数据压缩算法来执行用于对所述原始数据进行压缩的所述数据压缩方案。

[0028] 所述原始数据可包括对象签名或对象扩展,且所述指令在由所述处理器执行时还可使所述处理器使所述存储装置:存取所述存储装置上的对象签名词典或所述存储装置上的对象扩展词典,以分别基于所述对象签名或所述对象扩展确定所述原始数据的数据类型;以及基于所确定的所述数据类型来执行所述数据压缩方案。

[0029] 所述指令在由所述处理器执行时还使所述处理器使所述存储装置根据用户请求从主机计算机接收所述原始数据。

[0030] 所述指令在由所述处理器执行时还可使所述处理器:使所述主机提供用于提醒正在请求对所述原始数据进行压缩的用户提供信息的菜单;以及使所述存储装置根据由所述用户提供的所述信息来确定用于对所述原始数据进行压缩的数据压缩算法。

[0031] 所述指令在由所述处理器执行时还可使所述处理器使所述存储装置:将所述原始数据存储在其所述存储装置上;估算与所述存储装置对应的可用时间或资源;以及当有足够的时间及资源可用时,对所述原始数据进行压缩。

[0032] 本文所公开的根据实施例的驱动器能够接收原始数据以及对所述原始数据执行有损数据压缩,从而允许用户改变与数据压缩对应的特定参数的控制级别。因此,通过卸载

原本可能会对用户系统造成相对实质性的负担的有损数据压缩活动,驱动器能够将主机中央处理器循环空出来。在一些实施例中,驱动器可在驱动器上实施额外的或增加的硬件资源,或者可实施操作系统 (operating system, OS) 级别及/或驱动器级别变化。

附图说明

[0033] 结合附图阅读以下说明可更详细地理解一些实施例,在附图中:

[0034] 图1示出有损数据压缩的传统方法的实例与根据本公开实施例的有损数据压缩的方法的实例之间的比较。

[0035] 图2示出根据本公开实施例的无需主机的辅助便能够在驱动器处执行智能有损数据压缩的有损压缩驱动器。

[0036] [附图标记的说明]

[0037] 1:密钥;

[0038] 2:值;

[0039] 3:数据压缩预测器;

[0040] 4:对象扩展词典;

[0041] 5:对象签名词典;

[0042] 6:熵估算算法库;

[0043] 7:压缩器;

[0044] 11:接收开关;

[0045] 12:出口开关;

[0046] 13:输出;

[0047] 14:服务品质偏好数据文件;

[0048] 15:板载装置性能简档/装置性能简档;

[0049] 16:压缩算法性能表;

[0050] 17:数据压缩算法库/板载编码库;

[0051] 110、120:有损数据压缩方案;

[0052] 112:主机;

[0053] 113:压缩/数据压缩/无损数据压缩;

[0054] 114:经压缩的数据文件;

[0055] 116:驱动器;

[0056] 118:原始数据文件/原始数据;

[0057] 122:主机/主机计算机/主机中央处理器;

[0058] 123:压缩/有损数据压缩/有损压缩/数据压缩;

[0059] 124:经压缩的数据文件/经压缩的数据/经压缩的有损数据文件;

[0060] 125:存储/保存/重新保存;

[0061] 126:驱动器/智能有损数据压缩驱动器/存储装置/智能压缩驱动器;

[0062] 128:未经压缩的数据/原始数据文件/原始数据/数据文件/原始的未经压缩的数据/数据/对象原始数据文件。

具体实施方式

[0063] 通过参照对实施例及附图的以下详细说明,可更容易地理解本发明概念的特征及其实现方法。在下文中,将参照附图更详细地阐述实施例。然而,本发明可被实施为各种不同形式,而不应被视为仅限于本文所示出的实施例。确切来说,提供这些实施例作为实例是为了使本公开将透彻及完整,并将向所属领域中的技术人员充分传达本发明的各个方面及特征。因此,可不再阐述对于所属领域中的一般技术人员完整地理解本发明的各个方面及特征而言并非必需的工艺、元件及技术。除非另外注明,否则在所有附图及书面说明通篇中相同的参考编号表示相同的元件,且因此,将不再对其予以重复说明。另外,为使说明清楚起见,可不再示出与实施例的说明无关的部件。在附图中,为清晰起见,可夸大各个元件、各个层及各个区的相对大小。

[0064] 在以下说明中,出于解释目的,阐述各种具体细节来提供对各种实施例的透彻理解。然而,显而易见的是,可不使用这些具体细节或者可使用一种或多种等效配置来实践各种实施例。在其他实例中,以方块图形式示出众所周知的结构及装置以避免不必要地混淆各种实施例。

[0065] 应理解,当称一元件、层、区或组件位于另一元件、层、区或组件“上(on)”、“连接到(connected to)”或“耦合到(coupled to)”另一元件、层、区或组件时,所述元件、层、区或组件可直接位于所述另一元件、层、区或组件上、直接连接到或直接耦合到所述另一元件、层、区或组件,抑或可存在一个或多个中间元件、层、区或组件。然而,“直接连接/直接耦合(directly connected/directly coupled)”则是指一个组件与另一个组件直接连接或直接耦合,而不具有中间组件。同时,例如“位于...之间(between)”、“直接位于...之间(immediately between)”或者“相邻于...(adjacent to)”及“紧邻于...(directly adjacent to)”等阐述各个组件之间的关系的其他表达可类似地加以解释。另外,还应理解,当称一元件或层“位于”两个元件或层“之间(between)”时,所述元件或层可为所述两个元件或层之间的唯一元件或层,抑或也可存在一个或多个中间元件或层。

[0066] 本文所用术语仅是出于阐述特定实施例的目的而并非旨在限制本发明。除非上下文清楚地另外指明,否则本文所用单数形式“一(a及an)”旨在也包括复数形式。还应理解,当在本说明书中使用用语“包括(comprises、comprising)”、“具有(have、having)”及“包含(includes、including)”时,是指明所陈述特征、整数、步骤、操作、元件及/或组件的存在,但不排除一个或多个其他特征、整数、步骤、操作、元件、组件及/或其群组的存在或添加。本文所用用语“及/或”包括相关列出项中的一个或多个项的任意及所有组合。

[0067] 本文所用用语“实质上(substantially)”、“大约(about)”及类似用语用作近似用语、而并非作为程度用语,并且旨在考虑到所属领域中的一般技术人员将知的测量值或计算值的固有偏差。本文所述“大约”或“近似地”包括所阐述值且意指处于所属领域中的一般技术人员所确定的特定值的偏差的可接受范围内,此考虑到所说的测量以及与特定品质的测量相关联的误差(即,测量系统的限制)。举例来说,“大约”可意指处于一个或多个标准偏差范围内,或处于所述值的 $\pm 30\%$ 、 20% 、 10% 、 5% 内。另外,在阐述本发明的实施例时使用“可(may)”是指“本发明的一个或多个实施例”。本文所用用语“使用(use)”、“正使用(using)”及“被使用(used)”可被视为分别与用语“利用(utilize)”、“正利用(utilizing)”

及“被利用(utilized)”同义。另外,用语“示例性(exemplary)”旨在指实例或例示。

[0068] 当某一实施例可被以不同方式实施时,特定工艺次序可与所阐述的次序不同地执行。举例来说,两个连续阐述的工艺可实质上同时执行或以与所阐述的次序相反的次序执行。

[0069] 在本文中参照剖视图阐述各种实施例,所述剖视图为实施例及/或中间结构的示意性例示。因此,预期会因例如制造技术及/或容差而导致相对于例示形状的变化。另外,本文所公开的具体结构说明或功能说明仅是例示性的,目的在于阐述根据本公开概念的实施例。因此,本文所公开的实施例不应被视为仅限于各个区的特定例示形状,而是应包含由例如制造引起的形状偏差。举例来说,被例示为矩形的植入区通常应具有圆形特征或曲线特征及/或在其边缘存在植入浓度的梯度而非从植入区到非植入区为二元变化。同样地,通过植入而形成的隐埋区可在所述隐埋区与在进行植入时所经过的表面之间的区中引起一些植入。因此,附图中所例示的区为示意性的且其形状并非旨在例示装置的区的实际形状且并非旨在进行限制。另外,如所属领域中的技术人员应认识到,可采用各种不同的方式来对所阐述的实施例进行修改,而此完全不背离本公开的精神及范围。

[0070] 根据本文所述本发明的实施例的电子装置或电装置及/或任何其他相关装置或组件可使用任何适合的硬件、固件(例如,应用专用集成电路(application-specific integrated circuit))、软件、或软件、固件、及硬件的组合来实施。举例来说,可将这些装置的各种组件形成在一个集成电路(integrated circuit,IC)芯片上或单独的集成电路芯片上。此外,可将这些装置的各种组件实施在柔性印刷电路膜(flexible printed circuit film)、带载封装(tape carrier package,TCP)、印刷电路板(printed circuit board,PCB)上、或形成在一个衬底上。此外,这些装置的各种组件可为在一个或多个计算装置中由一个或多个处理器运行、执行计算机程序指令并与用于执行本文所述各种功能性的其他系统组件进行交互的进程或线程(thread)。计算机程序指令存储在可在例如(举例来说)随机存取存储器(random access memory,RAM)等标准存储器装置的计算装置中实施的存储器中。计算机程序指令也可存储在例如(举例来说)紧凑型光盘只读存储器(compact disc read only memory,CD-ROM)、闪存驱动器(flash drive)或类似元件等其他非暂时性计算机可读媒体中。另外,所属领域中的技术人员应知,在不背离本发明示例性实施例的精神及范围的条件下,可将各种计算装置的功能性组合或整合成单一的计算装置,或者可使一特定计算装置的功能性跨越一个或多个其他计算装置分布。

[0071] 除非另外定义,否则本文所用所有用语(包括技术及科学用语)的含义均与本发明所属领域中的普通技术人员所通常理解的含义相同。还应理解,用语(例如在常用词典中所定义的用语)应被解释为具有与其在相关技术的上下文及/或本说明书中的含义一致的含义,且除非在本文中明确定义,否则不应将其解释为具有理想化或过于正式的意义。

[0072] 如上所述,在传统上,在无需主机的辅助的条件下由硬存储装置执行的非智能数据压缩总是为无损压缩的形式。另外,在传统上,所有的有损数据压缩均是由主机/主机计算机/主机中央处理器执行的,而不是由驱动器执行的。然而,可存在其中不同类型的数据适合于较高程度压缩的场景,在这些场景中有损数据压缩是可接受的或者甚至是所期望的(例如,当编辑及/或存储音乐数据文件、视频数据文件或其他多媒体数据文件时)。此外,可能会存在以下情况:在驱动器处压缩原始数据文件、同时将原本被主机中央处理器消耗用

于执行压缩的资源空出来可能是有用的。

[0073] 本公开的实施例提供一种存储装置(例如固态驱动器(SSD)),所述存储装置能够执行有损数据压缩及存储而在执行数据压缩时无需对应主机的辅助,从而将主机的中央处理器空出来用于其他用户活动。

[0074] 图1示出有损数据压缩的传统方法的实例与根据本公开实施例的有损数据压缩的方法的实例之间的比较。

[0075] 参照图1,如前面所阐述,传统的驱动器116没有在无需求主机112的辅助的条件下执行有损数据压缩。而是,如图1所示,通过以下步骤来执行有损数据压缩方案110:在主机112对数据进行压缩113之前将原始数据文件118提供到主机112;以及将经压缩的数据文件114发送到驱动器116以进行存储,从而会耗尽主机112的系统资源。也就是说,在传统上,如果要对原始数据文件118(例如,脉冲编码调制(PCM)格式的原始声音/音频数据文件)进行压缩,则用户向主机112发出指令来将原始数据文件118压缩113成经压缩的数据文件114(例如,压缩成动态图像专家压缩标准音频层面3(MP3)数据文件),且主机112将接着执行数据压缩113来创建经压缩的数据文件114。之后,主机112会将经压缩的数据文件114发送到驱动器116以在驱动器116上进行存储。

[0076] 因此,尽管之前的技术使驱动器116能够对原始数据118执行无损数据压缩113(此将会保护数据保真度(data fidelity)),然而驱动器116不能够执行有损数据压缩。当数据保真度损失对于用户而言是可接受的且被允许的、或者是用户请求时(例如,当用户偏好高容量的有损数据压缩时,例如可能是对音乐及视频进行编辑的情形),有损数据压缩在传统上将在对应的主机机器上(例如,在主机112上)占用大量处理器循环。

[0077] 有鉴于上述,如果有损数据压缩的任务可被从位于用户的主要中央处理器中的主机卸载,则可有所助益。因此,根据本公开实施例的驱动器126(例如,硬存储装置或固态驱动器(SSD))在从主机/主机计算机/主机中央处理器122接收原始的未经压缩的数据/原始数据文件128之后执行一个或多个有损数据压缩方案120,并对原始数据文件128进行压缩123以产生经压缩的数据文件124,经压缩的数据文件124会接着被存储/保存125在驱动器126上。

[0078] 根据本实施例,用户可指令主机122(例如,通过指令中央处理器设定“请求压缩”旗标)压缩123原始数据128(例如,当用户已完成对将以经压缩的形式被保存在驱动器126上的视频或音频数据文件进行编辑时),从而使驱动器126能够使用有损数据压缩来自动地对原始数据文件128进行压缩123以在驱动器126上产生经压缩的数据文件124,而无需使用主机122的资源。智能有损数据压缩驱动器126能够通过对数据进行分析(例如,通过对数据文件的类型、可用数据压缩方法及算法、用户参数及服务品质(quality of service,QoS)要求等进行分析)以识别适合进行有损数据压缩123的情形来实现自动数据压缩。之后,驱动器126可根据所选择的数据压缩方案来执行有损数据压缩123,且可将经压缩的数据文件124保存125在驱动器126上。应注意,驱动器126可在接收到数据128后便对数据128进行压缩123,或者可在对数据进行压缩123及将经压缩的数据124存储125在驱动器126上之前将数据以未经压缩格式存储少量的时间。

[0079] 图2示出根据本公开实施例的无需主机的辅助便能够在驱动器处执行智能有损数据压缩的有损压缩驱动器。

[0080] 参照图2,如上所述,用户可将原始数据文件128(例如,PCM格式的原始声音数据文件)从主机122直接发送到驱动器126。之后,驱动器126能够将原始数据文件128识别为可被适合转换成有损的、经压缩的数据文件124的原始数据文件。之后,驱动器126可执行有损数据压缩123来在驱动器126内创建经压缩的数据文件124(例如,MP3数据文件),从而将原始数据文件128转换成经压缩的有损数据文件124,且之后将经压缩的有损数据文件124存储/保存125在驱动器126自身上。因此,本实施例是一种存储装置(例如,驱动器)126,其能够自动地或半自动地执行得到良好控制的有损数据压缩123、从而将原本将由主机122用来执行数据压缩113的中央处理器时钟循环及资源空出来以用于其他活动。

[0081] “智能压缩(smart compress)”驱动器126可包括用于数据管理的命令及控制协议以及相对增强的盘上(例如,固态驱动器上)资源。举例来说,不同于传统的硬盘,驱动器126可包括数据压缩算法库17,数据压缩算法库17包括各种有损数据压缩机制。板载编码库/数据压缩算法库17可含有与用于执行各种有损数据压缩方案及编码方案的各种数据压缩算法对应的代码。驱动器126还可包括被扩展成包含有损数据压缩参数的服务品质(QoS)偏好数据文件14。驱动器126可使用这些参数来确定如何对原始数据128进行压缩。

[0082] 来自用户的每一个有损数据压缩请求最终均与将进行压缩123的原始数据文件128相关联。数据压缩请求可含有明确地指明作为压缩123的结果所应采用的文件格式的直接命令、或者明确地指明用于引导驱动器126执行压缩123的参数的直接命令。作为另外一种选择,数据压缩请求可对应于密钥元数据中含有的嵌入式命令,且驱动器126可基于密钥元数据来确定如何执行压缩123。

[0083] 在驱动器126以有损方式对文件进行压缩123时,驱动器126可使用例如请求数据压缩算法或最终格式等信息来确定使用何种压缩方案(例如,从数据压缩算法库17选择哪一种数据压缩算法)。也就是说,驱动器126可使用各种信息来得知驱动器126应使用某一种有损数据压缩算法对原始数据文件128进行压缩、或者得知驱动器126最终应根据特定的有损存储格式(例如,MP3数据文件、联合图像专家小组(JPEG)数据文件、动态图像专家小组(MPEG)数据文件等)来保存经压缩的数据文件124。

[0084] 作为另外一种选择,驱动器126可使用参数及设定值(例如,被存储在服务品质偏好数据文件14中)来确定何时及如何对数据进行有损压缩。举例来说,服务品质文件可指示将某种扩展形式的所有文件压缩123成JPEG数据文件。服务品质文件还可含有压缩算法的压缩细节,例如使用户可对具体数据压缩方案或品质设定值进行选择或微调、设定子采样量或对各种其他参数进行控制。

[0085] 因此,驱动器126可使用信息来理解所接收到的任意给定原始数据文件所需要的数据压缩格式及相关联的参数。根据本公开的实施例,可作为至少三个不同的选项中的一者或多者来递送及确定数据压缩格式及参数以使驱动器126能够得知这些信息,此将在以下进一步阐述。

[0086] 作为驱动器126的增强特征的另一个实例,驱动器126可包括编码加速硬件(encoding acceleration),且可能联机地或离线地执行重复数据删除(dedupe)且还能够在线地或离线地进行无损数据压缩。当有损数据压缩123是以在线操作执行时,数据是在(或近乎/实质上在)驱动器126接收到原始数据文件128时被压缩123。当有损数据压缩123是以离线模式执行时,未经压缩的数据128可在驱动器126上存储不确定的时间量。接着,当

驱动器126确定有足够的时间/资源时,或者当驱动器126决定使用存储在驱动器126上的未经压缩的数据128来工作时,数据文件128可被压缩123并且可作为经压缩的数据文件124被重新保存125到驱动器126。驱动器126是决定进行在线操作还是进行离线操作可根据由驱动器126作出的与压缩所需的可用时间及/或系统资源的估算量对应的估算而定。

[0087] 这种估算可基于机器学习预测,且可基于对驱动器126的板载装置性能简档(onboard device performance profile)15的分析来确定。举例来说,装置性能简档15可包括与驱动器126的正在使用的物理能力的记录对应的信息(例如,空闲的存储器、处理器能力等),且可使用压缩算法性能表16来预测给定任务可能需要的空闲资源的量。基于装置性能简档15及压缩算法性能表16中含有的信息,驱动器126可基于对将要进行压缩的原始数据文件128的类型的判断来确定来自数据压缩算法库17的哪一种压缩算法将最有效地利用可用装置资源。

[0088] 压缩算法性能表16可含有关于多种压缩算法的信息(例如,每一种压缩算法的平均压缩比、每一种压缩算法的吞吐量、每一种压缩算法的计算及存储器资源要求等),且可能基于文件类型及可用吞吐量(可通过来自服务品质偏好数据文件14的服务品质需求来确定)从数据压缩算法库17选择特定数据压缩算法,且可能基于装置性能简档15选择装置能力。

[0089] 驱动器126的其他增强特征可包括熵估算算法库6、对象签名词典5及/或对象扩展词典4,它们能够辨识不同类型的原始媒体数据文件(例如,通过识别原始数据文件的扩展或签名)。

[0090] 熵估算算法库6可含有用于对任意经采样数据的熵进行测试从而确定数据文件128的压缩系数级别的代码。基于熵的计算可用于搜索值2的一些部分以预测文件可被压缩到何种程度。基于熵的计算基于熵数据或随机数据更难压缩这一原则(例如,已经过压缩的文件具有较高的熵特性,且未经压缩的文件更有可能具有更可压缩的有序数据)来进行。对象签名词典5可含有已知的对象/文件签名,所述已知的对象/文件签名可对应于相应的压缩技术。对象扩展词典4可含有一系列已知的对象/文件扩展以及对这些类型的对象对于各种压缩技术中的每一种而言可压缩程度如何的评估。对象扩展词典4将这一信息提供到数据压缩预测器3以便于压缩。因此,对象扩展词典4及对象签名词典5可向数据压缩预测器提供信息以便于压缩(部分地通过对已经过无损压缩的文件进行检测,且不再对它们进行重复压缩)。

[0091] 因此,当接收到原始数据文件128时,可使用驱动器126的各个组件来识别将要进行压缩的原始数据128的类型,所述类型可用于影响数据压缩预测器3及压缩器7的操作,此又可用于从数据压缩算法库17选择恰当的数据压缩算法。举例来说,数据压缩预测器3可使用对象扩展词典4、对象签名词典5及/或熵估算算法库6中含有的信息来确定或帮助确定值2的压缩系数。

[0092] 之后,压缩器7可基于所指示的参数来执行数据压缩。数据压缩预测器3及压缩器7可识别数据文件,辨识数据文件是何种类型,并基于例如以下因素来智能地应用最有效的有损数据压缩算法:直接用户需求、驱动器126的瞬态性能、根据服务品质偏好数据文件14而定的服务品质需求(其可用于确定应使用何种数据压缩方案来对数据进行压缩)、存储在压缩算法性能表16中的历史数据压缩算法性能等。数据压缩预测器3及压缩器7随后也可确

定存在更好的数据压缩算法可用(例如,如果更多处理器空间变得可用,有更多缓冲器空间变得可用等)。

[0093] 一旦数据压缩预测器3确定了如何对原始数据文件128进行压缩,则将信息提供到接收开关11,接收开关11接着在以下两者之间切换:将数据(例如,值2)输出到压缩器7;或者通过将数据发送到出口开关12进行输出13来对压缩器7进行旁通(如果决定不进行压缩,例如当接收到未被识别为需要进行无损压缩的文件类型时)。因此,数据压缩预测器3确定压缩应如何发生,且接收开关11将数据发送到压缩器7或将数据发送到出口开关12进行输出13。

[0094] 以下提出确定用于对驱动器126接收到的原始数据文件128进行压缩123的恰当的压缩方案及参数的三种选项。

[0095] 作为第一选项,数据压缩格式及参数可被确定为增强保存命令的显式部分(explicit part)。

[0096] 在传统的系统中,被保存到驱动器126的每一个原始数据文件128均可包括对驱动器126的最终保存命令(order),从而向驱动器126指示显式指令来使用某一种数据压缩类型及使用某些数据压缩参数保存某一种原始数据文件128。

[0097] 相比之下,使用这一选项,当用户向主机122发出指令来命令驱动器126将原始数据文件128保存为经压缩的数据文件124时,数据压缩格式及参数可被作为保存命令的一部分得到明确确定。在本文中所提出的三个选项中,这一选项一般来说可与基于区块的存储装置一起使用(这是因为根据以下所述第二选项,基于区块的存储装置不具有密钥值元数据选项)。然而,这一第一选项的实施方式可能需要对传统的基于区块的数据文件系统的驱动器/操作系统层进行改变。

[0098] 对于驱动器/操作系统层的改变可包括对引起向驱动器126发出写入命令的所有用户活动进行扫描以确定存在对原始数据文件128进行有损压缩123的意图或机会(例如,用户有意将位图保存为JPEG,或者有意将位图从未经压缩驱动器移动到驱动器126)。对驱动器/操作系统层的另一种改变可包括中断数据压缩请求来确认用户的意图、以及提醒用户提供关于压缩123的参数信息(例如,用户可在主机122处与菜单进行交互以使得用户能够选择关于有损数据压缩123的参数)。另外,可使用额外的经修改命令来传送身份及参数信息以及使驱动器126能够识别哪些逻辑块地址(logical block address, LBA)与目标数据文件相关联。

[0099] 在获得与所建议的对驱动器/操作系统层的改变中的一者或多者相关联的信息之后,可将原始数据文件128发送到驱动器126,且操作系统及驱动器可将用户选择的信息从主机122转发到驱动器126。这一选项还可实施与驱动器126进行通信的某些增强协议。

[0100] 作为第二选项,可将数据压缩格式及相关联的参数明确地编码在密钥1及值2((KV)对)的密钥1中。驱动器126可对增强命令或命令集作出回应,或者可检测与原始数据128相关联的充实密钥1,其中可有损压缩数据的写入伴随有“请求压缩”旗标、数据压缩格式识别、及/或数据压缩格式参数。

[0101] 在根据这一选项的基于KV的系统(KV-based system)中,可将数据压缩格式及参数的相关信息存储在密钥1的用于指示应如何存储数据的元数据部分中。举例来说,密钥1可具有指示经压缩的数据文件124应使用某一数据压缩设定值及某一子采样量而被存储为

JPEG数据文件的元数据。举例来说,驱动器126的数据压缩预测器3及压缩器7可使用密钥1中含有的信息(例如,元数据)来判断是否应对数据文件128进行修改/压缩以及应如何执行数据压缩。

[0102] 另外,密钥1可含有指示数据文件128的类型的文件/对象扩展,所述文件/对象扩展可提供关于是否执行压缩/如何执行压缩的信息。在对象扩展词典4中可维持有一系列已知的对象扩展,且数据压缩预测器3可使用来自对象扩展词典4的信息来确定数据压缩123的类型。

[0103] 密钥1还可含有文件/对象签名,即一行代码,所述一行代码提供关于存储在KV对的值2中的数据文件128的类型的信息。在对象签名词典5中可维持有一系列已知的数据类型的签名,且来自对象签名词典5的信息可被提供到数据压缩预测器3以确定压缩123使用的类型。

[0104] 因此,驱动器126可接收密钥1来指示所接收到的数据是原始媒体数据文件,且密钥1可含有应如何进行数据压缩的参数。因此,对象特征词典5或对象扩展词典4可通过读取密钥1的信息/元数据来检测与值2对应的所接收到的原始数据文件128应被压缩,且可使用元数据来确定应使用来自数据压缩算法库17的何种数据压缩算法。

[0105] 驱动器126可读取密钥1的元数据部分以确定如何对数据进行压缩123。当数据文件128被作为KV对象发送到KV存储体时,KV存储体可读取密钥1,可辨识并读取元数据部分,且可确定驱动器126被指令使用所指示的数据压缩格式及参数以所选择类型的有损数据压缩123来保存125经压缩的数据文件124。通过使用元数据部分中含有的信息,驱动器126可使用所指定的数据压缩算法及参数来对相关联的原始数据文件128进行压缩123。

[0106] 因此,数据压缩格式及参数可为KV存储系统的显式部分。KV对的密钥1可含有包含隐藏信息的元数据部分,其中可对参数进行编码以使驱动器126能够识别可压缩的数据。这种元数据信息的用户级输入(user-level entry)可使用KV存储软件堆栈来完成。也就是说,用户可对想要进行有损压缩123的数据文件128标记标签,与所述标签对应的信息包括在KV对的密钥1中。

[0107] 这一选项为用户提供“细粒度”控制,容易实施,且可在作出非常小的改变的条件实施。根据这一选项,用户可理解当某一类型的原始数据文件被保存到驱动器126时,将对所述数据进行压缩。

[0108] 作为第三选项,数据压缩格式及参数可作为较大的服务品质(QoS)偏好数据文件14的一部分被以隐式方式进行。服务品质偏好数据文件14可产生(例如,通过用户)且使用驱动器126来建立以规定确切的或通用的有损压缩及存储选项。服务品质偏好数据文件14可含有数据压缩策略的特定及/或全局实施方式的设定值(例如,以指示将应用于给定类型的所有数据文件128的默认数据压缩类型及参数、以要求预留原始数据文件128的备份副本、以在给定状况下阻止或延迟有损数据压缩123等)。

[0109] 服务品质偏好数据文件14可含有由驱动器126用于识别对象原始数据文件128的信息(例如,密钥1中的信息、数据文件扩展、首标签名(header signature)等)。在驱动器126识别到与服务品质偏好数据文件14适配的媒体数据文件后,可通过智能驱动器126来自动地实施服务品质偏好数据文件14中所含有的选项。驱动器126可使用服务品质偏好数据文件14中的信息来选择并自动应用恰当的数据压缩算法及参数。在一些实施例中,全局服

务品质需求可对冲突的用户指令或关键字段指定 (key-field designation) 进行覆盖或延迟。

[0110] 因此,驱动器126可被配置成当用户使主机122向驱动器126发送特定类型的原始数据文件128时,驱动器126可根据服务品质偏好数据文件14以对应的方式来压缩123及保存125数据。这一选项在理论上使得在对通用规则进行初始配置之后用户的控制较少,此可应用于数以千计的数据文件128,从而不再需要来自用户的一些精细控制。然而,这一选项能够实现全局数据压缩控制,以使得服务品质偏好数据文件14能够实现数据压缩机制的全局配置,且可减少侵入性的用户交互。应注意,根据本实施例的驱动器126可包括额外的硬件资源,例如较快的控制器处理器、较大的写入缓冲器、较大的控制器处理器随机存取存储器分配、及/或数据压缩/编码加速器中的一者或多者。

[0111] 因此,上述实施例提供一种能够进行智能有损数据压缩的自给自足的存储装置。

[0112] 以上是对示例性实施例的例示,而不应被视为对示例性实施例的限制。尽管已阐述了几个示例性实施例,但是所属领域中的技术人员将容易地理解,在不实质上背离示例性实施例的新颖教示及优点的条件下,在示例性实施例中可进行许多修改。因此,所有这些修改均旨在包含在由权利要求书所界定的示例性实施例的范围内。在权利要求书中,手段附加功能条款旨在涵盖执行所述功能的本文所述结构,且不仅涵盖结构等效形式且还涵盖等效结构。因此,应理解,以上是对示例性实施例的例示,而不应被视为仅限于所公开的具体实施例,且对所公开的示例性实施例的修改形式以及其他示例性实施例也旨在包含于所附权利要求书的范围内。本发明概念是由以上权利要求书界定,其中所述权利要求书的等效形式包含在本文中。

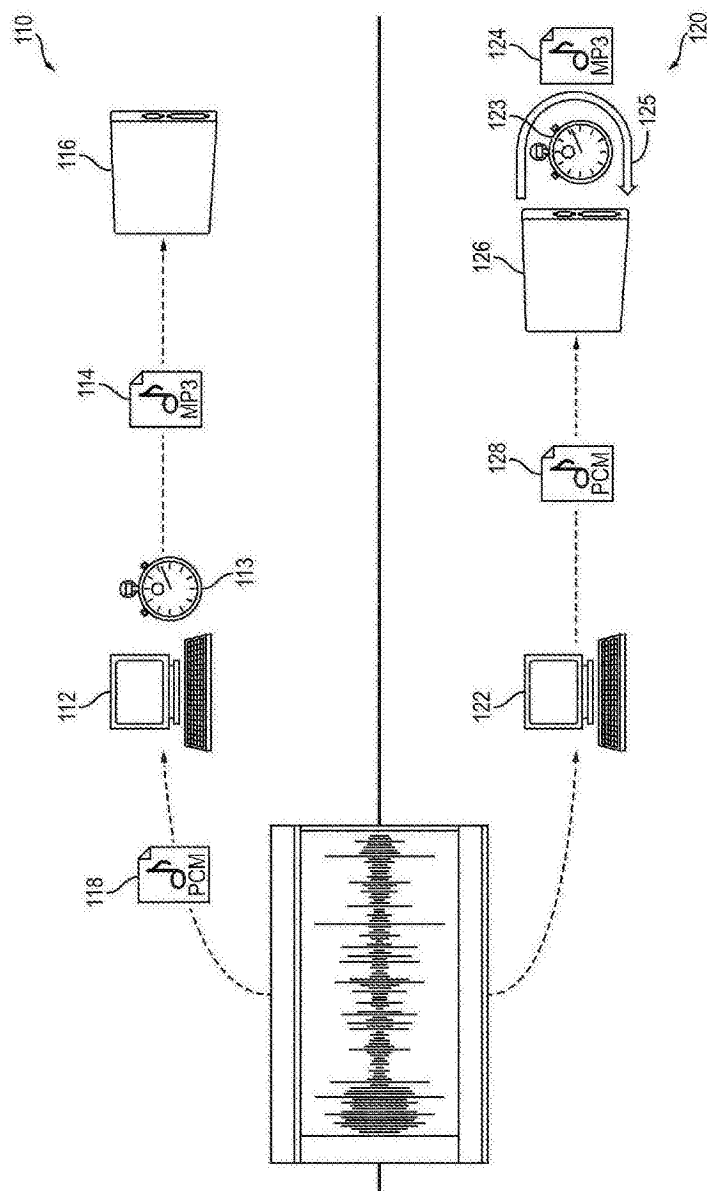


图1

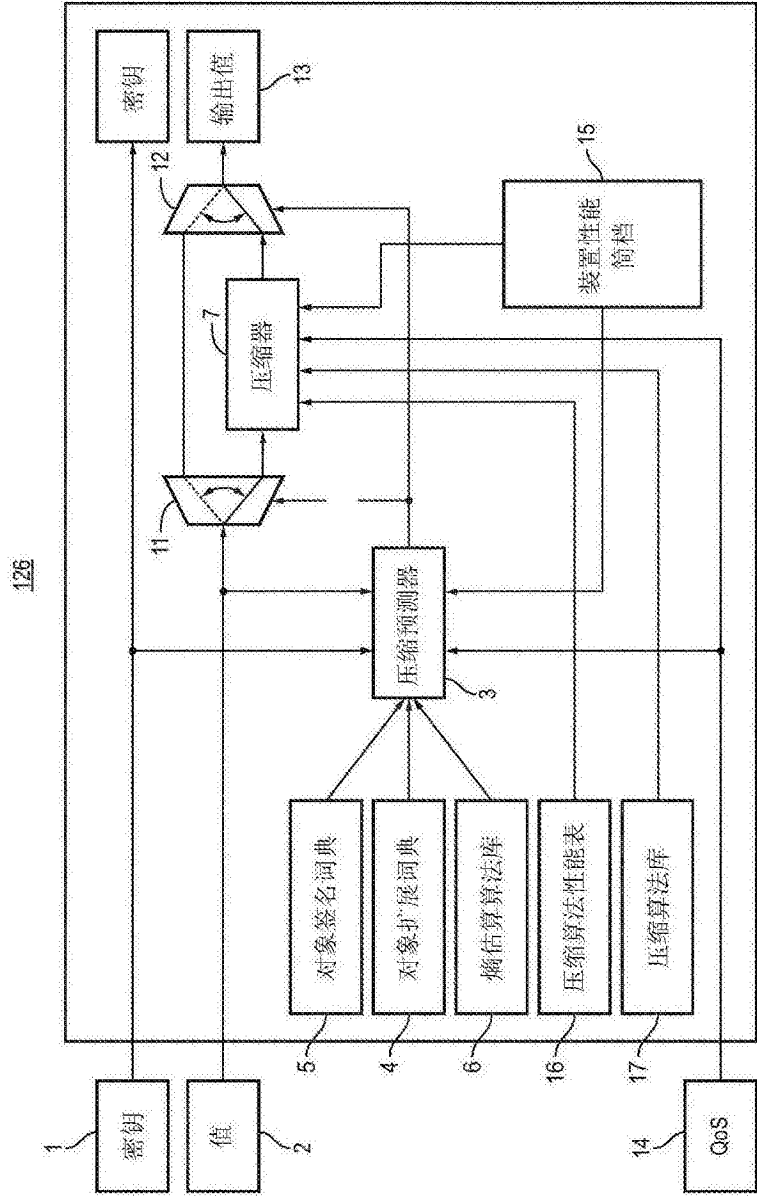


图2