



(12)发明专利

(10)授权公告号 CN 106489241 B

(45)授权公告日 2018.04.17

(21)申请号 201580037251.X

(22)申请日 2015.06.26

(65)同一申请的已公布的文献号  
申请公布号 CN 106489241 A

(43)申请公布日 2017.03.08

(30)优先权数据  
1411531.5 2014.06.27 GB

(85)PCT国际申请进入国家阶段日  
2017.01.09

(86)PCT国际申请的申请数据  
PCT/EP2015/025042 2015.06.26

(87)PCT国际申请的公布数据  
W02015/197202 EN 2015.12.30

(73)专利权人 古鲁洛吉克微系统公司  
地址 芬兰图尔库

(72)发明人 托马斯·卡开宁 奥西·卡雷沃

(74)专利代理机构 北京英赛嘉华知识产权代理  
有限责任公司 11204  
代理人 王达佐 王艳春

(51)Int.Cl.  
H03M 7/30(2006.01)

(56)对比文件  
CN 102246137 A,2011.11.16,  
US 2014075150 A1,2014.03.13,  
US 7643505 B1,2010.01.05,  
CN 1259828 A,2000.07.12,

审查员 李云杰

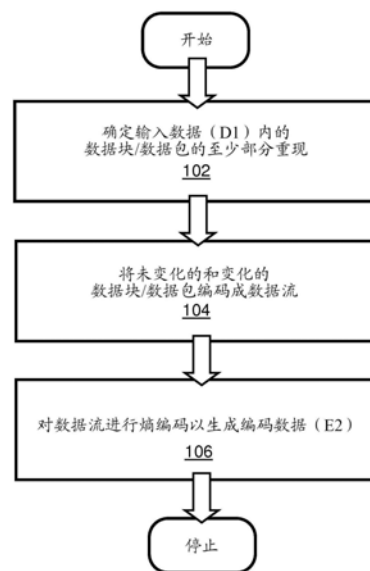
权利要求书5页 说明书26页 附图7页

(54)发明名称

编码器、解码器和方法

(57)摘要

用于对输入数据(D1)编码以生成对应编码数据(E2)的编码器,包括数据处理硬件,数据处理硬件可操作为:(a)确定输入数据(D1)内的数据块或数据包的至少部分重现,其中数据块或数据包包括多个字节;(b)采用至少一个参考符号来关联相互类似的数据块或数据包的重现,和/或指示输入数据(D1)内是否存在相互类似的数据块或数据包的重现;(c)采用多个变化符号,例如多个掩码位,来指示输入数据(D1)内的数据块或数据包的部分重现的变化数据元素和未变化数据元素以及变化数据元素的数据值的变化;以及(d)将至少一个参考符号和多个变化符号编码到编码数据(E2)中。本公开提供了使用编码器对输入数据(D1)编码以生成对应编码数据(E2)的方法。另外,本公开提供了对应的解码器,以及对编码数据(E2)解码以生成对应解码数据(D3)的对应方法。



1. 一种编码器,用于对输入数据(D1)编码以生成对应编码数据(E2),其特征在于,所述编码器包括数据处理硬件,所述数据处理硬件能够操作为:

(a) 确定所述输入数据(D1)内的数据块或数据包的至少部分重现,其中,所述数据块或数据包均包括多个字节;

(b) 采用至少一个参考符号来关联所述输入数据(D1)内的数据块或数据包的所述部分重现和对应参考数据块或数据包,和/或指示所述输入数据(D1)内是否存在数据块或数据包的所述部分重现,其中所述至少一个参考符号中的给定参考符号被用作相对指示器,所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置,或者所述至少一个参考符号中的给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置;

(c) 采用多个变化符号来指示所述输入数据(D1)内的数据块或数据包的所述部分重现中的变化数据元素和未变化数据元素,其中所述多个变化符号通过掩码位的方式实现,或者通过所述数据元素的数据值的变化实现;以及

(d) 将以下编码到所述编码数据(E2)中:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

2. 如权利要求1所述的编码器,其特征在于,所述数据处理硬件能够操作为通过使用指示未变化的一个或多个数据值来表示给定数据块或数据包的一个或多个未变化数据元素,其中,所述一个或多个数据值与所述输入数据(D1)中出现的数据值不同。

3. 如权利要求1所述的编码器,其特征在于,所述数据处理硬件能够操作为将以下编码到多个数据流中以提供所述编码数据(E2):

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

4. 如权利要求1所述的编码器,其特征在于,所述数据处理硬件能够操作为对所述输入数据(D1)编码,所述输入数据(D1)为一维数据或多维数据。

5. 如权利要求1所述的编码器,其特征在于,所述数据处理硬件能够操作为对所述输入数据(D1)编码,所述输入数据(D1)为音频数据、图像数据和视频数据中的至少一种。

6. 如权利要求1所述的编码器,其特征在于,所述编码器包括附加编码单元,所述附加编码单元用于将以下中的至少一部分编码到所述编码数据(E2)中:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值,

其中,所述附加编码单元能够操作为采用以下中的至少一种:熵调整编码、delta编码、ODelta编码、区间编码、行程长度编码(RLE)、拆分行程长度编码(SRLE)、插值编码。

7. 如权利要求1所述的编码器,其特征在于,所述数据处理硬件能够操作为确定在处理所述输入数据(D1)时用于所述数据块或数据包的最佳尺寸,并且在所述编码数据(E2)内提供指示用于所述数据块或数据包的所述最佳尺寸的信息。

8. 对输入数据(D1)编码以生成对应编码数据(E2)的方法,其中,所述方法包括处理所述输入数据(D1),其特征在于,所述方法包括:

(a) 确定所述输入数据(D1)内的数据块或数据包的至少部分重现,其中,所述数据块或数据包均包括多个字节;

(b) 采用至少一个参考符号来关联所述输入数据(D1)内的数据块或数据包的所述部分重现和对应参考数据块或数据包,和/或指示所述输入数据(D1)内是否存在数据块或数据包的所述部分重现,其中所述至少一个参考符号中的给定参考符号被用作相对指示器,所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置,或者所述至少一个参考符号中的给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置;

(c) 采用多个变化符号来指示所述输入数据(D1)内的数据块或数据包的所述部分重现中的变化数据元素和未变化数据元素,其中所述多个变化符号通过掩码位的方式实现或者通过所述数据元素的数据值的变化实现;以及

(d) 将以下编码到所述编码数据(E2)中:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

9. 如权利要求8所述的方法,其特征在于,所述方法包括:

通过使用指示未变化的一个或多个数据值来表示给定数据块或数据包的一个或多个未变化数据元素,其中,所述一个或多个数据值与所述输入数据(D1)中出现的数据值不同。

10. 如权利要求8所述的方法,其特征在于,所述方法包括将以下编码到多个数据流中以提供所述编码数据(E2):

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

11. 如权利要求8所述的方法,其特征在于,所述方法包括对所述输入数据(D1)编码,所述输入数据(D1)为一维数据或多维数据。

12. 如权利要求8所述的方法,其特征在于,所述方法包括对所述输入数据(D1)编码,所述输入数据(D1)为音频数据、图像数据和视频数据中的至少一种。

13. 如权利要求8所述的方法,其特征在于,所述方法包括:

(e) 使用附加编码单元,用于将以下中的至少一部分编码到所述编码数据(E2)中:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者

指示所述变化数据元素的数据值的变化数据值;以及

(f) 将所述附加编码单元设置为采用以下中的至少一个:熵调整编码、delta编码、ODelta编码、区间编码、行程长度编码(RLE)、拆分行程长度编码(SRLE)、插值编码。

14. 如权利要求8所述的方法,其特征在于,所述方法包括:

(g) 确定在处理所述输入数据(D1)时用于所述数据块或数据包的最佳尺寸;以及

(h) 在编码数据(E2)内提供指示用于所述数据块或数据包的所述最佳尺寸的信息。

15. 解码器,用于对编码数据(E2)解码以生成对应解码数据(D3),其特征在于所述解码器包括数据处理硬件,所述数据处理硬件能够操作为:

(a) 对所述编码数据(E2)解码,以识别:

(i) 至少一个参考符号,

(ii) 多个变化符号,所述多个变化符号指示所述解码数据(D3)内的数据块或数据包的部分重现的变化数据元素和未变化数据元素,所述多个变化符号通过掩码位或所述数据元素的数据值的变化的方式实现,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;

(b) 使用所述至少一个参考符号,以确定是否存在数据块或数据包的所述部分重现,并且生成用于所述解码数据(D3)内的数据块或数据包的所述部分重现的数据,其中所述至少一个参考符号中的给定参考符号被用作相对指示器,所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置,或者所述至少一个参考符号中的给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置,以及其中所述数据块或数据包均包括多个字节;

(c) 使用所述多个变化符号,以生成用于所述解码数据(D3)内的数据块或数据包的所述部分重现的所述变化数据元素的数据;以及

(d) 集合所生成的用于数据块或数据包的所述部分重现的数据以及所生成的用于数据块或数据包的部分重现的所述变化数据元素的数据,以生成对应解码数据(D3)。

16. 如权利要求15所述的解码器,其特征在于,所述数据处理硬件能够操作为通过指示未变化的一个或多个数据值对给定数据块或数据包的一个或多个未变化数据元素进行解码,其中,所述一个或多个数据值与在所述解码数据(D3)中出现的数据值不同。

17. 如权利要求15所述的解码器,其特征在于,所述数据处理硬件能够操作为对来自所述编码数据(E2)内提供的多个数据流的以下内容进行解码:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

18. 如权利要求15所述的解码器,其特征在于,所述解码数据(D3)被提供为一维数据或多维数据。

19. 如权利要求15所述的解码器,其特征在于,所述解码数据(D3)被提供为音频数据、图像数据和视频数据中的至少一种。

20. 如权利要求15所述的解码器,其特征在于,所述解码器包括附加解码单元,所述附

加解码单元用于对来自所述编码数据 (E2) 的以下中的至少一部分解码:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号, 以及

(iii) 当所述多个变化符号通过掩码位实现时, 指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值,

其中所述附加解码单元能够操作为采用以下中的至少一种: 熵调整解码、delta 解码、0Delta 解码、区间解码、行程长度解码、拆分行程长度解码、插值解码。

21. 如权利要求15所述的解码器, 其特征在于, 所述数据处理硬件能够操作为在所述编码数据 (E2) 内接收指示所述数据块或数据包的尺寸的信息。

22. 对编码数据 (E2) 解码以生成对应解码数据 (D3) 的方法, 其中, 所述方法包括处理所述编码数据 (E2), 其特征在于, 所述方法包括:

(a) 对所述编码数据 (E2) 解码, 以识别:

(i) 至少一个参考符号,

(ii) 多个变化符号, 所述多个变化符号指示所述解码数据 (D3) 内的数据块或数据包的部分重现的变化数据元素和未变化数据元素, 所述多个变化符号通过掩码位或所述数据元素的数据值的变化的方式实现, 以及

(iii) 当所述多个变化符号通过掩码位实现时, 指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;

(b) 使用所述至少一个参考符号, 以确定是否存在数据块或数据包的所述部分重现, 并且生成用于所述解码数据 (D3) 内的数据块或数据包的所述部分重现的数据, 其中所述至少一个参考符号中的给定参考符号被用作相对指示器, 所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置, 或者所述至少一个参考符号中的给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置, 以及其中所述数据块或数据包均包括多个字节;

(c) 使用所述多个变化符号, 以生成用于所述解码数据 (D3) 内的数据块或数据包的所述部分重现的所述变化数据元素的数据; 以及

(d) 集合所生成的用于数据块或数据包的所述部分重现的数据以及所生成的用于数据块或数据包的部分重现的所述变化数据元素的数据, 以生成对应解码数据 (D3)。

23. 如权利要求22所述的方法, 其特征在于, 所述方法包括:

通过指示未变化的一个或多个数据值对给定数据块或数据包的一个或多个未变化数据元素解码, 其中, 所述一个或多个数据值与在所述解码数据 (D3) 中出现的数据值不同。

24. 如权利要求22所述的方法, 其特征在于, 所述方法包括: 对来自所述编码数据 (E2) 内提供的多个数据流的以下内容进行解码:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号, 以及

(iii) 当所述多个变化符号通过掩码位实现时, 指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

25. 如权利要求22所述的方法, 其特征在于, 所述解码数据 (D3) 被提供为一维数据或多维数据。

26. 如权利要求22所述的方法,其特征在于,所述解码数据(D3)被提供为音频数据、图像数据和视频数据中的至少一种。

27. 如权利要求22所述的方法,其特征在于,所述方法包括:

(e) 使用附加解码单元,以对来自所述编码数据(E2)的以下中的至少一部分进行解码:

(i) 所述至少一个参考符号,

(ii) 所述多个变化符号,以及

(iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;以及

(f) 将所述附加解码单元设置为采用以下中的至少一种:熵调整解码、delta解码、ODelta解码、区间解码、行程长度解码、拆分行程长度解码、插值解码。

28. 如权利要求22所述的方法,其特征在于,所述方法包括:在所述编码数据(E2)内接收指示所述数据块或数据包的尺寸的信息。

29. 一种计算机程序产品,包括其上存储有计算机可读指令的非暂时性计算机可读存储介质,所述计算机可读指令能够通过包括处理硬件的计算机化设备执行,以执行如权利要求8或权利要求22所述的方法。

30. 一种编解码器,包括:

至少一个如权利要求1所述的用于对输入数据(D1)编码以生成对应编码数据(E2)的编码器;以及

至少一个如权利要求15所述的用于对所述编码数据(E2)解码以生成对应解码数据(D3)的解码器。

## 编码器、解码器和方法

### 技术领域

[0001] 本公开涉及用于对输入数据 (D1) 编码以生成对应编码数据 (E2) 的编码器;本公开还涉及对输入数据 (D1) 编码以生成对应编码数据 (E2) 的方法。另外,本公开涉及用于对编码数据 (E2) 解码以生成对应解码数据 (D3) 的解码器;本公开还涉及对编码数据 (E2) 解码以生成对应解码数据 (D3) 的方法。另外,本公开涉及包括其上存储有计算机可读指令的非暂时性计算机可读存储介质的计算机程序产品,计算机可读指令可通过电脑化设备执行,电脑化设备包括处理硬件以执行上述方法。另外,本公开涉及编解码器,编解码器包括至少一个上述编码器以及至少一个上述解码器。

### 背景技术

[0002] 为本公开提供技术背景的各种已知文献列在与本公开有关的附录中。

[0003] 通常,用来压缩数据的算法是基于无损压缩方法[3]或有损耗压缩方法[4]中的一种。在无损压缩中,各种文件(即数据(D1))以数据(D1)可稍后恢复为原样的方式进行压缩。

[0004] 通常,当对数据(D1)编码时,使用数据删除复制(data de-duplication)方法是公知的,该方法试图除去数据(D1)中的数据段(即当数据(D1)例如短暂地流动时反复出现而没有变化的那些数据段)的副本。已知的数据删除复制方法能够有效地寻找与期望数据块完全类似的这些数据块。

[0005] 通常,已知的数据删除复制方法通过使用各种不同的方法来寻找先前出现的数据段,诸如:

[0006] (i) 通过相对于参考数据段,检测给定数据段中的改变的数据元素的数;

[0007] (ii) 通过计算给定数据段和参考数据段的数据元素之间的绝对差的总和;

[0008] (iii) 通过利用冗余检验表;或者

[0009] (iv) 通过采用滑动块方法。

[0010] 整个文件也可被复制删除,在这种情况下,用来代替复制文件的符号产生出色的压缩比[5]。

[0011] 另外,可以后处理的方式执行数据删除复制,在这种情况下在数据(D1)已写入之后追溯地执行对关联数据的处理。可替代地,可实时执行数据删除复制,即正在数据(D1)进入给定系统时,在这种情况下给定辨识出的数据块没有完全写入,而是替代地参考之前的、与给定辨识出的数据块彼此类似的数据块。

[0012] 数据删除复制用于当代信息技术工业的各种分支中,诸如数据存储器中以及数据传送网络中。例如,数据删除复制用于云服务中、系统备份复制中、以及电子邮件服务中,其中连续地传送相互类似的文件,或仅仅稍微变化的基本上相互类似的文件。另外,在互联网通信网络中,发送对请求的响应,来回传送数据字节,并且这些字节主要容纳部分或全部的同互联网协议(IP)包数据;例如数据删除复制与广域网(WAN)优化有关。

[0013] 先前众所周知的是,与常规的数据压缩方法相比,已知的数据删除复制方法更具

有成本效率。然而,已知的数据删除复制方法具有许多缺点。首先,由于数据删除复制方法试图实现期望的数据压缩比,所以经常使用可观的数据存储和处理能力。通常,需要增加关联的搜索区域,即用于寻找类似性的大量存储器,来提高数据压缩比。另外,需要使用诸如滑动搜索法的CPU加强方法,来提高数据压缩比。滑动搜索法试图通过在搜索区域内移动至由用于实现滑动搜索法所采用的算法指向的方向,以原始方式来识别目标数据块或数据包。

[0014] 其次,已知的数据删除复制方法不能发现这样数据块或数据包,即内容稍有变化,但是相对于期望数据块仍包含许多未变化数据元素的数据块或数据包。

[0015] 第三,已知的数据删除复制方法可能导致数据断裂,尤其在实时执行与这些数据删除复制方法有关的处理时。

### 发明内容

[0016] 本公开旨在提供采用数据删除复制技术的改进的编码器,数据删除复制技术能够改进编码性能。

[0017] 另外,本公开旨在提供采用复制技术的改进的解码器,复制技术能够改进解码性能。

[0018] 在第一方面,本公开的实施方式提供了用于对输入数据(D1)编码以生成对应编码数据(E2)的编码器,其特征在于,编码器包括数据处理硬件,数据处理硬件能够操作为:

[0019] (a) 确定所述输入数据(D1)内的数据块或数据包的至少部分重现,其中,所述数据块或数据包均包括多个字节;

[0020] (b) 采用至少一个参考符号来关联所述输入数据(D1)内的相互类似的数据块或数据包的重现,和/或指示所述输入数据(D1)内是否存在相互类似的数据块或数据包的重现,其中给定参考符号被用作相对指示器,所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置,或者给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置;

[0021] (c) 采用多个变化符号来指示所述输入数据(D1)内的数据块或数据包的部分重现的变化数据元素和未变化数据元素,其中所述多个变化符号通过掩码位的方式实现,或者通过所述数据元素的数据值的变化实现;以及

[0022] (d) 将以下编码到所述编码数据(E2)中:

[0023] (i) 所述至少一个参考符号,

[0024] (ii) 所述多个变化符号,以及

[0025] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

[0026] 编码器的数据处理硬件可操作为确定输入数据(D1)内的数据块或数据包的至少部分重现。为此目的,可选地,编码器10的数据处理硬件可操作为确定用于数据块或数据包的最佳尺寸。另外,可选地,编码器的数据处理硬件可操作为在编码数据(E2)内提供指示用于数据块或数据包的最佳尺寸的信息。

[0027] 另外,编码器的数据处理硬件可操作为使用至少一个参考符号来关联相互类似的数据块或数据包的重现,和/或指示输入数据(D1)内是否存在相互类似的数据块或数据包



的重现。

[0028] 可选地,编码器的数据处理硬件可操作为将以下编码到多个数据流中以提供所述编码数据 (E2):

[0029] (i) 所述至少一个参考符号,

[0030] (ii) 所述多个变化符号,以及

[0031] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

[0032] 应理解,“变化符号”总是包含用于表示已经变化的值以及用于表示特定值或其 delta 变化将如何进一步发送所必需的所有信息。换言之,如果使用了 delta/0Delta 值,那么只有一个值增加到变化符号中,并且仅添加到该符号中。这样,这些单个值当然可插入包含那些 delta/0Delta 值的数据值的流中,但是如果指示变化/未变化的位/标记被增加到变化符号,那么将存在两个值而不是一个,这样当然将两个值插入两个数据流中。这些数据流之一将包含指示变化/未变化的位/标记,并且然后第二数据流将同样包含新值,或者与在相同位置中待复制的数据块/数据包的值有关的不同的值。

[0033] 关于“参考符号”,它表示数据块或数据包是否已被复制,并且还表示在部分复制或全部复制中使用了哪个数据块或数据包。换言之,在相对参考中或数据块自身索引或直接参考的指示器中的值“0”指示该数据块将不被复制。这样,使用直接或相对参考,其它值指示哪里可发现该数据块或数据包的重现。可通过不同程度的精度来执行直接和相对参考二者;换言之,地址/索引可对字节/字/块/包寻址/索引。当然,还可能通过直接参考来使用不是用于重现的可能位置的数据值,以指示关注的的数据块/数据包没有被复制。与使用数据块自身索引/指示器相比较,使用不可能直接参考的值产生一个特别的优点,即该值将总是同样的,即其例如可更好地熵压缩。在直接参考中的不可能值例如为负值或出现在数据块/数据包范围之外的值。在间接参考中,上述零值“0”总是指示好的(即,可能的)值,因为它每次都是同样的,并且相对地其总是指向相同的数据块,并且因而易于识别关注的的数据块/数据包不可具有重现的数据块/数据包。

[0034] 另外,编码器可操作为将 (c) 中的多个变化符号用作为多个掩码位,以指示输入数据 (D1) 内的数据块或数据包的部分重现的变化数据元素和未变化数据元素(值),并且还指示变化的值。可替代地,变化符号可包含与 delta 值内的变化值或未变化值以及值的变化有关的信息。

[0035] 可替代地,可选地,编码器的数据处理硬件可操作为通过使用指示未变化的一个或多个值来表示给定数据块或数据包的一个或多个未变化数据元素,其中,所述一个或多个值与所述输入数据 (D1) 中出现的那些值不同。

[0036] 另外,编码器的数据处理硬件可操作为将至少一个参考符号以及多个变化符号(例如,掩码位)编码到编码数据 (E2) 中。可选地,编码器的处理硬件可操作为将至少一个参考符号以及多个变化符号(例如,掩码位)编码到多个数据流中,以提供编码数据 (E2)。

[0037] 可选地,编码器的数据处理硬件可操作为对所述输入数据 (D1) 编码,所述输入数据 (D1) 为以下中的至少一种:一维数据、多维数据、音频数据、图像数据、视频数据、传感器数据、文本数据、二进制数据、医学数据。

[0038] 可选地,编码器包括附加编码单元,所述附加编码单元用于将以下中的至少一部

分编码到所述编码数据 (E2) 中:

[0039] (i) 所述至少一个参考符号,

[0040] (ii) 所述多个变化符号, 以及

[0041] (iii) 当所述多个变化符号通过掩码位实现时, 指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值,

[0042] 其中, 所述附加编码单元能够操作为采用以下中的至少一种: 熵调整编码、delta 编码、0Delta 编码、区间编码、行程长度编码 (RLE)、拆分行程长度编码 (SRLE)、插值编码。

[0043] 另外, 可选地, 编码器包括附加编码单元, 附加编码单元用于将至少一个参考符号和多个变化符号中的至少一部分编码到编码数据 (E2) 中。可选地, 附加编码单元可操作为使用以下中的至少一个: 熵调整编码、delta 编码、如于 2013 年 3 月 1 日提交的专利申请 GB1303661.1 中所描述的 0Delta 编码、具有如于 2014 年 7 月 21 日提交的专利申请 GB1412937.3 中描述的不同预测器的 0Delta 编码、1u 或 8u 区间编码、行程长度编码 (RLE)、如于 2013 年 3 月 1 日提交的专利申请 GB130360.3 中描述的拆分行程长度编码 (SRLE)、以及插值编码。

[0044] 本公开的实施方式的优势在于通过实现数据删除复制的方法, 编码器能够提供增强的编码性能。

[0045] 在第二方面, 本公开的实施方式提供了对输入数据 (D1) 进行编码以生成对应编码数据 (E2) 的方法, 其中方法包括处理输入数据 (D1), 其特征在于, 该方法包括:

[0046] (a) 确定所述输入数据 (D1) 内的数据块或数据包的至少部分重现, 其中, 所述数据块或数据包均包括多个字节;

[0047] (b) 采用至少一个参考符号来关联所述输入数据 (D1) 内的相互类似的数据块或数据包的重现, 和/或指示所述输入数据 (D1) 内是否存在相互类似的数据块或数据包的重现, 其中给定参考符号被用作相对指示器, 所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置, 或者给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置;

[0048] (c) 采用多个变化符号来指示所述输入数据 (D1) 内的数据块或数据包的部分重现的变化数据元素和未变化数据元素, 其中所述多个变化符号通过掩码位的方式实现, 或者通过所述数据元素的数据值的变化实现; 以及

[0049] (d) 将以下编码到所述编码数据 (E2) 中:

[0050] (i) 所述至少一个参考符号,

[0051] (ii) 所述多个变化符号, 以及

[0052] (iii) 当所述多个变化符号通过掩码位实现时, 指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

[0053] 可选地, 该方法包括通过使用指示未变化的一个或多个值来表示给定数据块或数据包的一个或多个未变化数据元素, 其中, 所述一个或多个值与所述输入数据 (D1) 中出现的那些值不同。

[0054] 可选地, 该方法包括将以下编码到多个数据流中以提供所述编码数据 (E2):

[0055] (i) 所述至少一个参考符号,

[0056] (ii) 所述多个变化符号, 以及

[0057] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

[0058] 可选地,所述方法包括对所述输入数据(D1)编码,所述输入数据(D1)为以下中的至少一种:一维数据、多维数据、音频数据、图像数据、视频数据、传感器数据、文本数据、二进制数据、医学数据。

[0059] 可选地,所述方法包括:

[0060] (e) 使用附加编码单元,用于将以下中的至少一部分编码到所述编码数据(E2)中:

[0061] (i) 所述至少一个参考符号,

[0062] (ii) 所述多个变化符号,以及

[0063] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;以及

[0064] (f) 将所述附加编码单元设置为采用以下中的至少一个:熵调整编码、delta编码、ODelta编码、区间编码、行程长度编码(RLE)、拆分行程长度编码(SRLE)、插值编码。

[0065] 更可选地,该方法包括:

[0066] (g) 确定在处理所述输入数据(D1)时用于所述数据块或数据包的最佳尺寸;以及

[0067] (h) 在编码数据(E2)内提供指示用于所述数据块或数据包的所述最佳尺寸的信息。

[0068] 在第三方面,本公开的实施方式提供了包括其上存储有计算机可读指令的非暂时性(即,非瞬时性)计算机可读存储介质的计算机程序产品,计算机可读指令可通过包括处理硬件的计算机化设备执行,以执行上述方法。

[0069] 在第四方面,本公开的实施方式提供了用于对编码数据(E2)解码以生成对应解码数据(D3)的解码器,其特征在于所述解码器包括数据处理硬件,所述数据处理硬件能够操作为:

[0070] (a) 对所述编码数据(E2)解码,以识别:

[0071] (i) 至少一个参考符号,

[0072] (ii) 多个变化符号,所述多个变化符号指示所述解码数据(D3)内的数据块或数据包的部分重现的变化数据元素和未变化数据元素,所述多个变化符号通过掩码位或所述数据元素的数据值的变化的方式实现,以及

[0073] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;

[0074] (b) 使用所述至少一个参考符号,以确定是否存在相互类似的数据块或数据包的重现,并且生成用于所述解码数据(D3)内的相互类似的数据块或数据包的重现的数据,其中给定参考符号被用作相对指示器,所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置,或者给定参考符号被用于指示所述参考数据块或数据包相对于所述给定数据块或数据包的位置,以及其中所述数据块或数据包均包括多个字节;

[0075] (c) 使用所述多个变化符号,以生成用于所述解码数据(D3)内的数据块或数据包的所述部分重现的所述变化数据元素的数据;以及

[0076] (d) 集合所生成的用于相互类似的数据块或数据包的所述重现的数据以及所生成的用于数据块或数据包的部分重现的所述变化数据元素的数据,以生成对应解码数据

(D3)。

[0077] 可选地,解码器的数据处理硬件可操作为在编码数据(E2)内接收指示数据块或数据包的尺寸的信息。

[0078] 解码器的数据处理硬件可操作为对编码数据(E2)解码,以识别至少一个参考符号以及多个变化符号。可选地,解码器的数据处理硬件可操作为对来自编码数据(E2)内提供的多个数据流的、以下内容进行解码:

[0079] (i) 所述至少一个参考符号,

[0080] (ii) 所述多个变化符号,以及

[0081] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

[0082] 可选地,在解码器中,解码数据(D3)被提供为以下中的至少一种:一维数据、多维数据、音频数据、图像数据、视频数据、传感器数据、文本数据、二进制数据、医学数据。

[0083] 解码器包括附加解码单元,所述附加解码单元用于对来自所述编码数据(E2)的以下中的至少一部分解码:

[0084] (i) 所述至少一个参考符号,

[0085] (ii) 所述多个变化符号,以及

[0086] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值,

[0087] 其中所述附加解码单元能够操作为采用以下中的至少一种:熵调整解码、delta解码、0Delta解码、区间解码、行程长度解码、拆分行程长度解码、插值解码。

[0088] 可选地,解码器的数据处理硬件能够操作为在所述编码数据(E2)内接收指示所述数据块或数据包的尺寸的信息

[0089] 可替代地,可选地,解码器的数据处理硬件可操作为对来自指示未变化的一个或多个值的给定数据块或数据包的一个或多个未变化数据元素进行编码,其中,该一个或多个值与在解码数据(D3)中出现的值不同。

[0090] 在第五方面,本公开的实施方式提供了对编码数据(E2)解码以生成对应的解码数据(D3)的简单且快速的方法,其中,所述方法包括处理所述编码数据(E2),其特征在于,所述方法包括:

[0091] (a) 对所述编码数据(E2)解码,以识别:

[0092] (i) 至少一个参考符号,

[0093] (ii) 多个变化符号,所述多个变化符号指示所述解码数据(D3)内的数据块或数据包的部分重现的变化数据元素和未变化数据元素,所述多个变化符号通过掩码位或所述数据元素的数据值的变化的方式实现,以及

[0094] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;

[0095] (b) 使用所述至少一个参考符号,以确定是否存在相互类似的数据块或数据包的重现,并且生成用于所述解码数据(D3)内的相互类似的数据块或数据包的重现的数据,其中给定参考符号被用作相对指示器,所述相对指示器指示给定数据块或数据包相对于其对应参考数据块或数据包的位置,或者给定参考符号被用于指示所述参考数据块或数据包相

对于所述给定数据块或数据包的位置,以及其中所述数据块或数据包均包括多个字节;

[0096] (c) 使用所述多个变化符号,以生成用于所述解码数据 (D3) 内的数据块或数据包的所述部分重现的所述变化数据元素的数据;以及

[0097] (d) 集合所生成的用于相互类似的数据块或数据包的所述重现的数据以及所生成的用于数据块或数据包的部分重现的所述变化数据元素的数据,以生成对应解码数据 (D3)。

[0098] 可选地,方法包括:对来自所述编码数据 (E2) 内提供的多个数据流的、以下内容进行解码:

[0099] (i) 所述至少一个参考符号,

[0100] (ii) 所述多个变化符号,以及

[0101] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值。

[0102] 可选地,解码数据 (D3) 被提供为以下中的至少一种:一维数据、多维数据、音频数据、图像数据、视频数据、传感器数据、文本数据、二进制数据。

[0103] 更可选地,方法包括:

[0104] (e) 使用附加解码单元,以对来自所述编码数据 (E2) 的以下中的至少一部分解码:

[0105] (i) 所述至少一个参考符号,

[0106] (ii) 所述多个变化符号,以及

[0107] (iii) 当所述多个变化符号通过掩码位实现时,指示所述变化数据元素的数据值或者指示所述变化数据元素的数据值的变化数据值;以及

[0108] (f) 将所述附加解码单元设置为采用以下中的至少一种:熵调整解码、delta解码、ODelta解码、区间解码、行程长度解码、拆分行程长度解码、插值解码。

[0109] 可选地,方法包括:在所述编码数据 (E2) 内接收指示所述数据块或数据包的尺寸的信息。

[0110] 在第六方面,本公开的实施方式提供了包括其上存储有计算机可读指令的非暂时性(即,非瞬时性)计算机可读存储介质的计算机程序产品,计算机可读指令可通过包括处理硬件的计算机化设备执行,以执行上述的方法。

[0111] 在第七方面,本公开的实施方式提供了包括上述编码器以及上述解码器的编解码器。

[0112] 本公开的附加方面、优势、特征和主题将通过结合所附权利要求解释的示例性实施方式的详细描述和附图而变得明显。

[0113] 将理解,在不偏离由所附权利要求限定的本公开范围的情况下,本公开的特征适合于以多种组合方式组合。

## 附图说明

[0114] 当结合附图阅读时,更好地理解以上内容 and 以下示例性实施方式的详细说明。为了示出本公开,在附图中示出了本公开的示例性结构。然而,本公开并不限于本文中公开的具体方法和装置。另外,本领域技术人员应理解附图不是按比例绘制的。只要可能,相同的元件由相同的标号来指示。

[0115] 现将仅通过示例的方式参照附图描述本公开的实施方式,在附图中:

[0116] 图1是根据本公开的实施方式的编码器和解码器的示意图,编码器用于对输入数据(D1)编码以生成对应的编码数据(E2),解码器用于对编码数据(E2)解码以生成对应的解码数据(D3),其中编码器和解码器共同形成编解码器;

[0117] 图2是根据本公开的实施方式的对输入数据(D1)编码以生成对应的编码数据(E2)的方法的步骤的图示;

[0118] 图3A和图3B共同为根据本公开的实施方式的编码过程的步骤的图示;

[0119] 图4是根据本公开的实施方式的对编码数据(E2)解码以生成对应的解码数据(D3)的方法的步骤的图示;以及

[0120] 图5A和图5B共同为根据本公开的实施方式的解码过程的步骤的图示。

[0121] 在附图中,使用带下划线的数字来表示下划线数字上的项或者与下划线数字相邻的项。未带下划线的数字涉及通过将未带下划线的数字与项连接的线所指示的项。当数字不带下划线且附有相关箭头时,不带下划线数字用于识别箭头所指的一般项。

### 具体实施方式

[0122] 以下详细描述示出了本公开的实施方式以及实现这些实施方式的方法。虽然已公开了执行本公开的最佳方式,但是本领域技术人员应理解用于执行或实践本公开的其他实施方式也是可能的。

[0123] 大体上,本公开的实施方式涉及处理数据的方法和用于执行这种数据处理的装置,其中执行数据删除复制和复制,这将在下文更详细地阐述。

[0124] 参照图1,本公开的实施方式涉及:

[0125] (i) 用于对输入数据(D1)编码以生成对应的编码数据(E2)的编码器10,以及对输入数据(D1)编码以生成编码数据(E2)的对应的方法;

[0126] (ii) 用于对编码数据(E2)解码以生成对应的解码数据(D3)的解码器20,以及对编码数据(E2)解码以生成解码数据(D3)的对应的方法;以及

[0127] (iii) 编解码器30,包括至少一个编码器和至少一个解码器的组合,即编码器10和解码器20的组合。

[0128] 可选地,如以无损耗操作方式,解码数据(D3)完全类似于输入数据(D1)。可替代地,可选地,如以有损耗操作方式,解码数据(D3)大致类似于输入数据(D1)。还可替代地,可选地,例如通过变换,解码数据(D3)不同于输入数据(D1),但与出现在输入数据(D1)中的信息保持大体上类似;例如,当还要求重新格式化解码数据(D3)时,解码数据(D3)被有用地形成为与输入数据(D1)不同,例如与不同类型的通信平台、软件层、不同类型的通信装置等兼容。虽然本公开的一些实施方式描述了关于出现于输入数据(D1)中的图像数据,即通过使用掩码位(mask bit)来表示,但是应理解本公开的实施方式能够对多种彼此不同类型的输入数据(D1)编码,例如传感器数据。为了概括所有类型的数据,例如在所附权利要求中使用术语“变化符号(change symbol)”,在输入数据(D1)包括图像数据的特殊示例情况下,数据符号实现为掩码位。

[0129] 编码器10包括用于处理输入数据(D1)的处理硬件。可选地,编码器10的处理硬件可操作为对输入数据(D1)编码,输入数据(D1)为以下中的至少一种但并不限于此:一维数

据、多维数据、音频数据、图像数据、视频数据、传感器数据、文本数据、二进制数据、以及医学数据。可选地，输入数据 (D1) 接收为数据流或文件。可选地，输入数据 (D1) 包括许多小的变化，例如结构化细节和噪音的混合。

[0130] 编码器10的处理硬件可操作为确定输入数据 (D1) 内的数据块或数据包的至少部分重现。为此目的，编码器10的处理硬件可选地可操作为确定用于数据块或数据包的最佳尺寸，如将在以下更详细地描述。另外，编码器10的处理硬件可选地可操作为在编码数据 (E2) 内，提供指示用于数据块或数据包的最佳尺寸的信息。

[0131] 为了确定数据块或数据包的至少部分重现，编码器10的处理硬件可操作为使用一个或多个适当的冗余检验方法来进行一个或多个冗余检验。可选地，使用冗余检验方法来选择数据块或数据包，所述数据块或数据包使未改变的数据元素的计数最大化，或使数据块或数据包的数据元素与对应的参考数据块或数据包的数据元素之间的绝对误差之和最小化。

[0132] 另外，可选地，使用冗余检验方法来计算用于给定数据块或数据包的一个或多个冗余检验值。在示例中，计算用于给定数据块或数据包的一个长的冗余检验值。在另一示例中，计算用于给定数据块或数据包的多个短的冗余检验值。

[0133] 如果计算出的用于给定数据块或数据包的冗余检验值与任何先前存储的信息没有匹配足够好，那么给定数据块或数据包被识别为非复制数据块或数据包。该非复制数据块或数据包如其在编码数据 (E2) 中那样写入或传递。另外，可选地，非复制数据块或数据包还可作为新参考数据块或数据包被插入至用于冗余检验目的的存储的信息序列中。

[0134] 有益地，冗余检验值可通过多种方式计算，但是主要考虑的是，冗余检验值应能够用足够的精度辨别非复制数据块或数据包。然而，应理解，如果给定任务可通过使用较少计算和数据存储资源密集和优化算法执行，那么通过计算不必要精确的冗余检验值来浪费编码器10的处理硬件的计算资源是没有益处的。

[0135] 另外，编码器10的处理硬件可操作为使用至少一个参考符号来关联相互类似的数据块或数据包的重现，和/或来指示输入数据 (D1) 内是否存在相互类似的数据块或数据包的重现。在这点上，编码器10的处理硬件可选地可操作为使用一个或多个指示器，用于指示数据块或数据包相对于它们的对应参考数据块或数据包的一个或多个至少部分重现。因此，所述至少一个参考符号被使用作为指向对应参考数据块或数据包的一个或多个指示器。

[0136] 可选地，一个或多个指示器为一个或多个相对指示器，该相对指示器指示输入数据 (D1) 内数据块或数据包的一个或多个至少部分重现与对应参考数据块或数据包的相关距离。换言之，一个或多个指示器可选地指示数据块或数据包的一个或多个至少部分重现相对于输入数据 (D1) 内的对应参考数据块或数据包的位置。

[0137] 可选地，一个或多个指示器包括具体位置的参考地址，这里可获得与它们的对应参考数据块或数据包相关的信息。可选地，一个或多个指示器包括在对编码数据 (E2) 解码以重建解码数据 (D3) 时需要使用的信息块的参考地址。换言之，一个或多个指示器指向它们的对应参考数据块或数据包的数据存储器位置。

[0138] 另外，可选地，一个或多个指示器可表示为以下所列中的至少一种：字节精确指示器（即直接寻址）、数据值精确指示器、字精确指示器、或块精确指示器（即，基于许多给定块

的寻址)。可选地,指示重现的地点(即“指示器”)可以相关的方式进行,诸如借助相关的每块符号值(即通过使用相关块精确移位)、相关字节精确移位、相关数据值精确移位、以及相关字精确移位等。

[0139] 另外,编码器10的处理硬件可操作为使用多个变化符号作为包括位/标记的多个掩码位以指示输入数据(D1)内数据块或数据包的部分重现的变化数据元素和未变化数据元素(即值)以及变化的值。可选地,与未变化数据元素有关的掩码位设置为值“0”,而与变化数据元素有关的掩码位设置为值“1”,或反之亦然。因而,多个掩码位描述了哪些数据元素是变化的,以及哪些数据元素是未变化的。可替代地,变化符号包含与delta值内的变化值或非变化值以及值的变化有关的信息。可选地,delta值被分配零“0”值以指示没有变化,以及被分配非零值以指示变化。

[0140] 可替代地,可选地,编码器10的处理硬件可操作为,通过使用指示没有变化的一个或多个值来表示给定数据块或数据包的一个或多个未变化数据元素,其中一个或多个值不同于在输入数据(D1)中出现的那些值。

[0141] 另外,编码器10的处理硬件可操作为将至少一个参考符号和多个变化符号编码到编码数据(E2)中。可选地,编码器10的处理硬件可操作为将至少一个参考符号和多个变化符号编码到多个数据流中以提供编码数据(E2)。

[0142] 如果使用无损耗操作方式,变化数据元素照样被写入或传送。如果使用有损耗操作方式,变化数据元素被量化。

[0143] 如果使用接近无损耗操作方式,变化数据元素的至少一部分被量化。为此目的,编码器10的处理硬件可选地可操作为根据输入数据(D1)的内容、类型和/或组成的分析,量化变化数据元素的仅一些部分。因而,编码器10能够适应性地改变输入数据(D1)和编码数据(E2)之间的数据压缩比。

[0144] 另外,变化数据元素被编码为例如以下中的至少一种:原始数据值、量化的原始数据值、delta值、量化的delta值、0Delta值、或量化的0Delta值。在本文中,“0Delta”指的是基于在二进制计数方法中概括的进行编码的微分形式,例如在专利文献GB1303661.1中所描述的,该专利通过引用并入本文。

[0145] 可选地,当原始值、delta值或0Delta值用于写入和传递关于参考数据块或数据包的变化数据元素时,未变化数据元素例如设置为值“0”以指示没有变化,或设置为不另外出现于数据中的值。当使用原始值但要求传递不另外出现于数据中的值时,后一选择为更好的解决方案。随后,变化数据元素被设置为其原始数据值或delta值,即在量化或没有量化的情况下与原始数据值和参考数据帧、数据块或数据包中对应数据值之差相等的数据值。如果变化数据元素设置为原始数据值,且该数据值正好等于“0”并且“0”还用来指示未变化数据元素,那么潜在地出现混乱,即数据模糊。因而,在本公开的实施方式中优选采用后一选择,即使用不另外出现于数据中的数据值来避免这种模糊。然后,用于指示未变化数据元素的数据值在编码数据(E2)中传递。当参考符号用于指示参考数据块或数据包,例如一些之前的数据块或数据包时,在GB141 1451.6中描述的方法是本公开的实施方式可使用的方法的好示例。

[0146] 如果使用有损耗操作方式,即使用了量化的形式,并且同时更新或增加参考符号,那么有益的是,当形成参考数据块时编码器10的处理硬件考虑将引起对应解码数据(D3)解



量化的那些变化。有益地,当用量化的原始值传递变化数据元素时,在确定待用于指示未变化数据元素的数据值时还考虑量化效果。当量化的delta值或量化的0Delta值用来传递变化数据元素时,那么当零用于未变化数据元素时,量化有益地不将变化的值变为零。

[0147] 另外,可选地,编码器10包括附加编码单元,附加编码单元用于将至少一个参考符号以及多个变化符号中的至少一部分编码到编码数据(E2)中。可选地,附加编码单元可操作为使用以下中的至少一个:熵调整编码、delta编码、0Delta编码、1u或8u区间编码、行程长度编码(RLE)、拆分行程长度编码(SRLE)、插值编码。

[0148] 此外,可选地,编码器10可操作为将编码数据(E2)通信至数据服务器和/或数据存储器(在图1中未示出),用于存储在数据库(在图1中未示出)中。数据服务器和/或数据存储器设置为能够访问解码器20,其有益地与编码器10兼容以用于随后对编码数据(E2)进行解码。

[0149] 在一些示例中,可选地,解码器20可操作为从数据服务器和/或数据存储器访问编码数据(E2)。

[0150] 在替代的示例中,可选地,编码器10可操作为通过通信网络或通过直接连接,将编码数据(E2)传递至解码器20。另外,应注意,设置有基于硬件的或基于软件的编码器的设备还可与设置有基于硬件的或基于软件的解码器的另一设备直接通信。

[0151] 在另一些替代示例中,可选地,解码器20被执行以从诸如硬盘和固态存储器(SSD)的非暂时性(即非瞬时性)计算机可读存储介质中获取编码数据(E2)。

[0152] 解码器20包括用于处理编码数据(E2)的处理硬件。

[0153] 可选地,解码器20的处理硬件可操作为在编码数据(E2)内接收指示数据块或数据包的尺寸的信息。

[0154] 解码器20的处理硬件可操作为对编码数据(E2)进行解码,以识别至少一个参考符号以及多个变化符号。可选地,解码器20的处理硬件可操作为对来自编码数据(E2)内提供的多个数据流的至少一个参考符号以及多个变化符号进行解码。

[0155] 可选地,解码器20的处理硬件可操作为对编码数据(E2)解码,以识别一个或多个指示器,所述指示器指示数据块或数据包相对于它们的对应参考数据包或数据块的一个或多个至少部分重现。

[0156] 解码器20的处理硬件然后可操作为使用至少一个参考符号,即一个或多个指示器,以生成用于编码数据(E2)内的数据块或数据包的部分重现的数据。一个或多个指示器可选地为一个或多个相对指示器,使得解码器20的处理硬件能够确定数据块或数据包相对于它们的对应参考数据包或数据块的部分重现的位置。

[0157] 另外,解码器20的处理硬件可操作为使用多个变化符号作为多个掩码位,以生成用于编码数据(E2)内的数据块或数据包的部分重现的变化数据元素的数据。可替代地,使用的变化符号包含与delta值内的变化值或无变化值以及值的变化有关的信息。

[0158] 可替代地,可选地,解码器20的处理硬件可操作为对来自指示无变化的一个或多个值的给定数据块或数据包的一个或多个未变化数据元素解码,其中该一个或多个值与解码数据(D3)中出现的那些值不同。

[0159] 随后,解码器20的处理硬件可操作为集合针对数据块或数据包的部分重现所生成的数据以及针对数据块或数据包的部分重现的变化数据元素所生成的数据,以生成解

码数据 (D3)。

[0160] 可选地,解码数据 (D3) 设置为以下中的至少一种:一维数据、多维数据、音频数据、图像数据、视频数据、传感器数据、文本数据、二进制数据、医学数据,但是不限于此。

[0161] 另外,可选地,解码器20包括附加解码单元,附加解码单元用于对来自编码数据 (E2) 的至少一个参考符号以及多个变化符号的至少一部分解码。可选地,附加解码单元可操作为使用以下中的至少一种:熵调整解码、delta解码、0Delta解码、1u或8u区间解码、行程长度解码、拆分行程长度解码、插值解码。

[0162] 此外,出现在编码器10和解码器20中的数据处理方法可选地包括关于专利申请 GB1411451.6中描述的子方法,上述专利通过引用并入本文,其中在表1中提供该专利申请的子方法。

[0163] 表1:编码器10和解码器20中使用的子方法

[0164]

使用范围	子方法细节
编码器 10	对输入数据 (D1) 编码以生成对应编码数据 (E2)

[0165]

	<p>的方法，其中，该方法包括将输入数据（D1）处理为数据块和/或数据包，其特征在于该方法包括：</p> <p>(i) 识别输入数据（D1）内的数据块和/或数据包的实质重现；</p> <p>(ii) 识别实质重现的数据块和/或数据包内何处元素是未变化的，和/或实质重现的数据块和/或数据包内何处元素是变化的；</p> <p>(iii) 通过使用至少一个对应符号或至少一个对应位对编码数据（E2）中的未变化数据元素进行编码，其中至少一个对应符号或至少一个对应位指示参考数据块和/或数据包中的未变化数据元素相对于对应元素缺少变化；以及</p> <p>(iv) 对编码数据（E2）中的变化数据元素编码。</p>
<p><b>解码器 20</b></p>	<p>对编码数据（E2）解码以生成对应的解码数据（D3）的方法，其中，该方法包括将编码数据（E2）处理为数据块和/或数据包，其特征在于，该方法包括：</p> <p>(i) 对编码数据（E2）解码以生成用于变化数据元素的数据，所述变化数据元素为在编码数据（E2）内的数据块和/或数据包的实质重现内变化的元素；</p> <p>(ii) 对编码数据（E2）解码以生成用于未变化数据元素的数据，所述未变化数据元素为在编码数据（E2）内的数据块和/或数据包的实质重现内未变化的元素，其中，未变化数据元素通过至少一个对应符号或至少一个对应位表示，该至少一个对应符号或至少一个对应位指示参考数据块和/或数据包中的未变化数据元素相对于对应元素缺少变化；以及</p> <p>(iii) 将在步骤（i）和步骤（ii）中对于变化数据元素和未变化数据元素生成的数据集合至数据块和/或数据包中，以生成解码数据（D3）。</p>

[0166] 图1仅为示例，不应过度地限制本文中权利要求的范围。应理解的是，用于编解码

器30的具体指示被提供为示例,并不应解释为将编解码器30限制为编码器和解码器的特定标号、类型或设置。本领域技术人员将领会到本公开的实施方式的许多变化、替代以及修改。

[0167] 可选地,编解码器30在单个设备内实现。可替代地,可选地,编解码器30在多个设备之间有效地实现。在示例中,编解码器30可选地实现为广播系统,其中,在第一空间位置处存在编码器10,在其他空间位置处存在多个解码器20。

[0168] 编解码器30可实现为以下中的至少一种:数据编解码器、音频编解码器、图像编解码器和/或视频编解码器。编解码器30能够在预处理阶段实时地压缩输入数据(D1)。

[0169] 另外,编解码器30可实现为提供实时数据传送网络编码方法,该编码方法特别在基于请求响应型通信的系统中可显著地节省数据传送所需的带宽,请求响应型通信诸如为用在网络浏览器和万维网(www)服务器中用于数据传送的超级文本传输协议(HTTP)[6]。除了HTTP协议之外,实时消息传输协议(RTMP)[7]经常用于音频和视频传递服务中以及回放视频服务中进行实时全双工数据传送。因而,编解码器30与基于HTTP的数据通信系统和基于RTMP的数据通信系统有关。

[0170] 另外,可选地,编码器10和解码器10可操作为实现用于HTTP和/或RTMP的区块传送编码。可选地,HTTP和/或RTMP使用请求和对请求的响应内的固定尺寸的数据块和/或数据包。

[0171] 编解码器30特别适合于用于传送尺寸被准确设置的数据块或数据包的通信协议。例如,因为在区块传送编码中传送的请求和对请求的响应使用了固定尺寸的数据块或数据包,所以在HTTP中使用区块传送编码是最佳使用区域。由于在HTTP上传送大量的数据,所以当使用编码器10对输入数据(D1)编码以生成对应编码数据(E2)时,将实现相当高的数据压缩比率。

[0172] 对于本公开的实施方式,通过使用HTTP获得的另一好处在于,几乎无一例外,HTTP请求/响应包含与正在传送什么以及何时传送有关的精确信息。这使得能够选择用于在对输入数据(D1)进行编码期间使用的最优冗余检验表。这潜在地提高了可实现的关联数据压缩比率。

[0173] 类似于HTTP,RTMP还限定了与具有特定限定尺寸的数据块或数据包的传送有关的规格。在互联网上,RTMP经常用来传送音频、图像或视频信息,特别是在实时通信中。在通信过程期间,用于待传送的数据块或数据包的尺寸根据采用的传送网络的响应时间和容量以预置间隔初始化。这潜在地提高了可实现的数据压缩比。

[0174] 如果使用了传送网络,并且给定的发送方还不清楚是否将采用根据本公开的编码方法,那么有可能实现转换用于数据接收器的信息的代理服务器,在这样情况下可在可能缓慢的接收器与快速的发射器之间实现期望的压缩比。在这个方案中,代理服务器通过将参考符号连同变化数据元素一起以及数据块或数据包中变化的和未变化数据元素的掩码位或其他指示发送至接收器,向接收器通知至少一个先前发送的数据块或数据包的未变化数据元素。

[0175] 在示例中,代理服务器为HTTP代理服务器或RTMP代理服务器,其相对于待优化的关联数据提供服务以及待服务的关联用户而位于最优地理位置处。可选地,采用根据本公开的实施方式的方法以使得诸如Windows Azure[8]的现成云端服务器方案更有效率。

[0176] 在最近几年(2010+)期间,数据、音频、图像以及视频传送的数量增加很快,以至于即使摩尔定律[9]也不能跟上这种发展。因而,编解码器30可能对更有效地传送数据非常有用。

[0177] 仅为了说明的目的,接下来提供在编码器10中执行的编码方法的数字示例。在这个示例中,输入数据(D1)包括可对应于单独的音频数据包、单独的图像块、HTTP包头等数据值。输入数据(D1)包括五个帧/块/包,其中每个帧/块/包包含54个数据值。应理解输入数据(D1)具有270个数据值,例如来自拆分成 $18 \times 3$ 个数据块的 $18 \times 15$ 图像。这导致分成五个数据块,每个数据块具有54个数据值。可替代地,为了压缩目的,输入数据(D1)还可认为是来自通过例如在美国专利文献US 8,675,731 B2中描述的块编码器生成的某一数据流的270个数据值,上述专利通过引用并入本文。

[0178] 输入数据(D1)包括270个数据值(熵等于1428.35比特 $\approx$ 179字节),并且表示如下:

[0179] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,120,250,64,0,  
 [0180] 118,6,206,99,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,  
 [0181] 31,254,192,146,80,16,0,254,185,9,0,0,0,12,41,157,180,29,0,16,243,42,  
 [0182] 42,172,8,0,69,0,0,40,121,13,64,0,118,6,206,80,62,241,193,52,172,16,  
 [0183] 17,60,34,201,192,220,189,31,183,3,31,254,201,204,80,16,1,2,175,204,  
 [0184] 0,0,0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,17,64,0,  
 [0185] 118,6,206,76,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,  
 [0186] 31,254,209,65,80,16,1,2,168,87,0,0,0,12,41,157,180,29,0,16,243,42,  
 [0187] 42,172,8,0,69,0,0,40,121,25,64,0,118,6,206,68,62,241,193,52,172,16,  
 [0188] 17,60,34,201,192,220,189,31,183,3,31,254,221,164,80,16,1,2,155,244,  
 [0189] 0,0,0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,67,64,0,  
 [0190] 118,6,206,26,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,  
 [0191] 31,254,233,149,80,16,1,2,144,3,0,0

[0192] 可选地,输入数据(D1)通过使用8u区间编码方法例如熵编码为253字节。通过参照这种8u区间编码,这意味着其是在根据本公开的编解码器中对8位数据值编码的区间编码方法。区间编码在[http://en.wikipedia.org/wiki/Range\\_coding](http://en.wikipedia.org/wiki/Range_coding)中进行了解释,上述网址中内容通过引用并入本文。

[0193] 现在,这个结果还包括在根据本公开的编解码器中发送当前22个替代的固定概率表中的概率表或相当好的固定表。如果已知较好的概率表并且需要发送和/或接近较少适用的概率表时,那么区间编码数据的共享有益地非常接近于熵,即在这个示例中为179字节。

[0194] 为了确定用于数据块的最佳尺寸(以下称为“最佳数据块尺寸”),编码器10的处理硬件操作为通过尺寸为7的初始数据块执行初始运行。有益地,初始数据块尺寸选择为相对小的数字,例如在3至8的范围内,以便能够精确检测最佳数据块尺寸。例如,数据块尺寸7导致如下所示的39个数据块:

[0195]  $270 = 38 \times 7 + 4$

[0196] 其中,38个数据块具有尺寸为7的数据值,以及最后一个数据块具有尺寸为4的数

据值。

[0197] 对于给定数据块,参考符号“1”用于到达给定数据块之前的一个数据块的对应参考数据块。同样地,如果参考数据块到达给定数据块之前的一个数据块和一个数据值,则使用参考符号“2”。因而,参考符号指示参考数据块相对于给定数据块的位置。

[0198] 应理解的是,给定数据块的尺寸可以以许多不同方式而认为是最优的。然而,通常期望在复制之后将编码数据尺寸优化(例如最小化)。通常通过努力寻找尽可能大的且还可产生尽可能多的待复制的数据块的数据块尺寸来很好地完成这种优化。接下来的部分描述确定几乎最优的数据块尺寸的一种方式示例。根据本公开,还可使用其他方法来确定最优数据块尺寸。

[0199] 另外,应理解在待编码的数据是周期性的情况下,最优数据块尺寸通常是该周期的长度、该周期长度的倍数,或者其倍数将大致等于该周期长度的一些数。

[0200] 根据本公开,如果块/周期连续地同样重复多次,那么乘以待复制的所选元素和重复的计数也是可能的。当对数据编码时,这可能在许多情况下产生有效的结果,但是主要地,这可通过对复制符号进行熵编码或通过执行递归复制更容易且便利地进行。

[0201] 为了清晰起见,下列列出了第9数据块的示例并且在其对应的参考数据块下添加了下划线:

[0202] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,120,250,64,0,

[0203] 118,6,206,99,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,

[0204] 31,254,192,146,80,16,0,254,185,9,0,0,0,12,41,157,180,29,0,16,243,42,

[0205] 42,172,8,0,69,0,0,40,121,13,64,0,118,6,206,80,62,241,193,52,172,16,

[0206] 17,60,34,201,192,220,189,31,183,3,31,254,201,204,80,16,1,2,175,204,

[0207] 0,0,0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,17,64,0,

[0208] 118,6,206,76,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,

[0209] 31,254,209,65,80,16,1,2,168,87,0,0,0,12,41,157,180,29,0,16,243,42,

[0210] 42,172,8,0,69,0,0,40,121,25,64,0,118,6,206,68,62,241,193,52,172,16,

[0211] 17,60,34,201,192,220,189,31,183,3,31,254,221,164,80,16,1,2,155,244,

[0212] 0,0,0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,67,64,0,

[0213] 118,6,206,26,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,

[0214] 31,254,233,149,80,16,1,2,144,3,0,0

[0215] 应理解的是,实际的参考数据块通常从解码数据获得,但是,因为这个数据现在是无损编码,那么原始数据值还可用于参考数据块。当参考数据块到达示例性数据块之前的48个数据值时,参考符号“48”被分配至示例性数据块。

[0216] 在这个示例中,可选地通过使用下列方程使用数据值精确参考索引来计算参考数据块的第一索引:

[0217]  $\text{first\_index\_of\_reference\_data\_block} = \text{first\_index\_of\_current\_data\_block} +$

[0218]  $\text{block\_size} + \text{reference\_symbol} -$

[0219]  $\text{number\_of\_nearest\_reference.}$

[0220] (参考数据块的第一索引 = 当前数据块的第一索引 + 块尺寸 + 参考符号 - 最近的参考的数量)

[0221] 对于第9数据块的示例,方程得到诸如以下的值:

$$[0222] \quad 56+7+48-1=2$$

[0223] 其中,当前数据块的第一索引由下列方程计算:

$$[0224] \quad \text{first\_index\_of\_current\_data\_block} = (\text{number\_of\_current\_data\_block} - 1) * \\ [0225] \quad \text{block\_size},$$

[0226] (当前数据块的第一索引 = (当前数据块的数量 - 1) \* 数据块尺寸)

[0227] 其中,例如,当前数据块的数量为9,并且数据块尺寸为7。已添加下划线的第9数据块的参考符号,即“48”也来自下一章之后的参考符号。

[0228] 应理解,第一数据块不具有对应的参考数据块,因而不分配参考符号。然而,最后一个数据块被分配其对应的参考符号。还应理解,最后一个数据块是不完整的,因而参考数据块的仅四个第一数据元素的数据值用于对最后一个数据块进行编码。因而,38个参考符号被分配如下:

$$[0229] \quad 0, 0, 0, 0, 0, 0, 31, \underline{48}, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48,$$

$$[0230] \quad 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 34$$

[0231] 编码器10的处理硬件然后操作以计算最佳数据块尺寸,例如如下:

$$[0232] \quad \langle \text{最普通参考符号} \rangle + \langle \text{初始数据块尺寸} \rangle - \langle \text{用于最近的可能参考块的参考符号} \rangle$$

$$[0233] \quad = 48 + 7 - 1 = 54$$

[0234] 应理解,在这个具体情况中,用于该具体数据列的最优块尺寸计算为54,该值也是这个数据列的重复周期的长度。其他适合的数据块尺寸例如为:

$$[0235] \quad 162, 108, 27, 18, 9, 6$$

[0236] 可选地,在最佳块尺寸方程式中也可使用经常出现且具有比最普通参考符号更小的值的其它参考符号以代替最普通参考符号,以使得能够对较大数量的小数据块进行复制。应理解,最近的参考的数量有益地不用作最普通的参考数使用,因为其可能受到类似数据值的长连续集的影响,并且因而其可能是子最优的。

[0237] 以这种方式,编码方法包括在数据块自己的尺寸内寻找数据块的重现,以使得数据块不需要分成较小的数据段,即与已知的滑动搜索法类似。同样显然的是,这样不需要在具有全部长度和所有位置的所有先前存储的信息中反复地搜索复制的数据块,这会浪费宝贵的计算时间和资源。

[0238] 应理解,在计算用于数据块或包的最佳尺寸时,如果找到相当小但不为1的参考值以及许多使用的参考值,那么通常有益的是使用这些值来代替最常用的参考值。还应理解的是,有益地用相对小的数据块或数据包尺寸来执行初步搜索,以使得不丢失短的参考,但是无论如何用足够大的数据块或数据包尺寸进行初步搜索,以使得相同数据值的冗长链不导致第一参考值变为最普通的一个。在这种状况出现的情况中,有益地是选择一些经常出现的小的、但比1大的值,来计算用于数据块或数据包的最佳尺寸以代替最普遍使用的参考。

[0239] 现在,编码器10的处理硬件操作为通过最佳块尺寸,即为54的数据块尺寸执行下一运行。这产生如下的五个数据块:

$$[0240] \quad 0, 12, 41, 157, 180, 29, 0, 16, 243, 42, 42, 172, 8, 0, 69, 0, 0, 40, 120, 250, 64, 0,$$

$$[0241] \quad 118, 6, 206, 99, 62, 241, 193, 52, 172, 16, 17, 60, 34, 201, 192, 220, 189, 31, 183, 3,$$

- [0242] 31,254,192,146,80,16,0,254,185,9,0,0,
- [0243] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,13,64,0,118,
- [0244] 6,206,80,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,31,
- [0245] 254,201,204,80,16,1,2,175,204,0,0,
- [0246] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,17,64,0,118,
- [0247] 6,206,76,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,31,
- [0248] 254,209,65,80,16,1,2,168,87,0,0,
- [0249] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,25,64,0,118,
- [0250] 6,206,68,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,31,
- [0251] 254,221,164,80,16,1,2,155,244,0,0,
- [0252] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,121,67,64,0,118,
- [0253] 6,206,26,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,31,
- [0254] 254,233,149,80,16,1,2,144,3,0,0
- [0255] 在上文中,为了清晰起见,已对数据块的部分重现的变化数据元素添加了下划线。
- [0256] 另外,为了清晰起见,与最后四个数据块的216个数据元素有关的216个掩码位 ( $54 \times 4=216$ ) 表示如下:
- [0257] 0,1,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,
- [0258] 0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,1,1,1,1,0,0,
- [0259] 0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,
- [0260] 0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,1,1,0,0,
- [0261] 0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,
- [0262] 0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,1,1,0,0,
- [0263] 0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,
- [0264] 0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,1,1,0,0,
- [0265] 在上文中,与未变化数据元素有关的掩码位设置为值“0”,而与变化数据元素有关的掩码位设置为值“1”。
- [0266] 编码器10的处理硬件然后操作为将变化数据元素的数据值、与变化的和未变化数据元素有关的掩码位、以及与相互类似的数据块的部分重现有关的参考符号编码到多个数据流中,以提供编码数据(E2)。在本示例中,多个数据流包括如下表示的第一数据流、第二数据流、以及第三数据流。
- [0267] 第一数据流包括第一数据块的54个数据元素的数据值以及在后续数据块中的27个变化数据元素的数据值(熵=460.24位~ =58字节),并且表示如下:
- [0268] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,120,250,64,0,
- [0269] 118,6,206,99,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,
- [0270] 31,254,192,146,80,16,0,254,185,9,0,0,121,13,80,201,204,1,2,175,204,
- [0271] 17,76,209,65,168,87,25,68,221,164,155,244,67,26,233,149,144,3
- [0272] 第二数据流包括27字节( $216/8=27$ ),表示与后续数据块的216个数据元素有关的216个掩码位(熵=117.41位~ =15字节),并且表示如下:
- [0273] 0,0,12,2,0,48,15,0,0,130,0,0,12,3,0,128,32,0,0,195,0,0,32,8,0,192,



[0274] 48

[0275] 第三数据流包括指示用于后续数据块(熵=0位~0字节)的对应参考数据块的四个参考符号,并且表示如下:

[0276] 1,1,1,1

[0277] 这些参考符号指示第一数据块是用于第二数据块的对应参考数据块,第二数据块是用于第三数据块的对应参考数据块,诸如此类。在本文中,参考符号用作相对指示器,相对指示器指示数据块相对于它们的对应参考数据块或数据包的位置。

[0278] 由于参考符号根据它们何时顺序出现进行排序,所以显著地降低了通过第三数据流产生的熵。在替代的实现中,当参考符号以适应性次序顺序地出现时,参考符号可选地插入表中。在这种情况下,第三数据流可有效地被delta编码。

[0279] 在期望仅参考之前的数据块或数据包(即,恰好在给定数据块或数据包之前到达的数据块或数据包)的情况下,第三数据流不必写入或传送至编码数据(E2)中。

[0280] 也可以总是由块间数据值计算ODelta值的方式,通过使用一个或多个ODelta编码方法实现相同的示例。因而,不需要单独的掩码位(即,第二数据流),也不需要变化的值(即,第一数据流)。相反,通过使用ODelta值使用仅包含对应信息的第三数据流。在编码的ODelta值的第三数据流中,零“0”值对应于未变化的值,即零掩码位,以及另一编码的ODelta值对应于变化的值,即掩码中具有值“1”位。然而,这种对应性以这样的方式计算,即通常可通过读取当前编码的ODelta值直接知道ODelta值是出现变化,这确保了可对ODelta值解码。当使用Odelta编码时,可通过“0”用作最小值、“255”用作最大值并且“256”用作约束值的如下流代替以上最后一个示例中的第一数据流和第二数据流:

[0281] 0,12,41,157,180,29,0,16,243,42,42,172,8,0,69,0,0,40,120,250,64,0,

[0282] 118,6,206,99,62,241,193,52,172,16,17,60,34,201,192,220,189,31,183,3,

[0283] 31,254,192,146,80,16,0,254,185,9,0,0,

[0284] 0,1,19,0,0,0,0,0,237,0,0,0,0,0,0,

[0285] 0,0,0,0,0,0,0,0,0,0,0,0,0,0,9,58,0,0,1,4,246,195,0,0,

[0286] 0,4,0,0,0,0,0,252,0,0,0,0,0,0,0,

[0287] 0,0,0,0,0,0,0,0,0,0,0,0,0,8,117,0,0,0,0,249,139,0,0

[0288] 0,8,0,0,0,0,0,248,0,0,0,0,0,0,0,

[0289] 0,0,0,0,0,0,0,0,0,0,0,0,0,12,99,0,0,0,0,243,157,0,0

[0290] 0,42,0,0,0,0,0,212,0,0,0,0,0,0,0,

[0291] 0,0,0,0,0,0,0,0,0,0,0,0,0,12,241,0,0,0,0,245,15,0,0

[0292] 上述列出的数据流可通过例如使用区间编码、SMRLE编码或VLC编码有效地压缩。还应理解,可选地,可使用其他压缩方法对关注的编码ODelta值的熵进行编码。

[0293] 上述编码方法比已知的数据删除复制方法明显更有成本效率,因为它们潜在地能够:

[0294] (i) 除完全复制的数据块或数据包之外,发现部分复制的数据块或数据包;以及

[0295] (ii) 仅对编码数据(E2)中的变化数据元素进行编码。

[0296] 与输入数据(D1)相比,仅写入或传送编码数据(E2)中的变化数据元素减少了处理编码数据(E2)、编码数据(E2)内的大量编码字节和编码数据(E2)的熵[1,2]所需的时间。因

而,与已知数据删除复制方法相比,实现了明显高的数据压缩比。

[0297] 根据本公开的实施方式,当输入数据(D1)为多维数据时,数据值在数据块中限定的次序同样是重要性能。仅为了说明的目的,现在认为二维(2D)图像数据具有 $16 \times 16$ 数据值,该数据值可以两种方式分成16个数据块,即16个数据块中的每个具有 $1 \times 16$ 或 $16 \times 1$ 的数据值的第一种情况中,以及16个数据块中的每个具有 $4 \times 4$ 的数据值的第二种情况。应理解,第一种情况中不能适当地使用多维特征,即数据块之间的空间相关性,而在第二种情况中可以使用。

[0298] 另外,可选地,编码器10的附加编码单元操作为,通过采用一个或多个适当的熵编码方法进一步压缩三个数据流,即第一数据流、第二数据流以及第三数据流。附加地或可替代地,可选地,编码器10的处理硬件操作为使用上述编码过程进一步递归地压缩三个数据流。这在编码数据(E2)中得到了较高的数据压缩程度。

[0299] 编码数据(E2)然后从编码器10传递至解码器20。解码器20的附加解码单元然后操作为对编码数据(E2)进行解压缩。其后,解码器20的处理硬件操作为执行根据本公开的实施方式的解码过程。由于使用了无损操作方式,所以生成与输入数据(D1)完全类似的解码数据(D3)。因而,输入数据(D1)的原始数据元素和解码数据(D3)的数据元素之间的均方误差的和为零。

[0300] 现在将描述编码器10的有损耗操作方式,使用上述示例的相同的输入数据(D1)。在有损耗操作方式中,可以更加提高数据压缩比。

[0301] 可选地,在识别数据块或数据包的至少部分重现之前,对输入数据(D1)进行量化。在使用量化值4的示例情况中,如下列出了相对于为7的初始数据块尺寸所分配的38个参考符号:

[0302] 1,0,0,0,0,2,33,48,48,48,48,48,48,48,48,48,48,48,48,48,48,48,48,48,48,

[0303] 48,48,48,48,48,48,48,48,48,48,48,48,48,50

[0304] 在该示例情况中,最佳数据块尺寸为54。

[0305] 然后如下列所示的,编码器10的处理硬件操作为将变化数据元素的数据值、掩码位以及参考符号编码成三个数据流,以提供编码数据(E2)。

[0306] 第一数据流包括:第一数据块的54个数据元素的量化数据值,以及在后续数据块中的25个变化数据元素的量化数据值,并且表示如下:

[0307] 0,3,10,39,45,7,0,4,60,10,10,43,2,0,17,0,0,10,30,62,16,0,29,1,51,24,

[0308] 15,60,48,13,43,4,4,15,8,50,48,55,47,7,45,0,7,63,48,36,20,4,0,63,46,

[0309] 2,0,0,3,20,50,51,0,43,51,4,19,52,16,42,21,6,17,55,41,38,61,16,6,58,

[0310] 37,36,0

[0311] 第二数据流包括27字节,其指示与后续数据块的216个数据元素有关的216个掩码位,并且表示如下:

[0312] 0,0,8,2,0,48,14,0,0,130,0,0,12,3,0,128,32,0,0,195,0,0,32,8,0,192,48

[0313] 第三数据流包括指示用于后续数据块的对应参考数据块的四个参考符号,并且表示如下:

[0314] 1,1,1,1

[0315] 当需要时,解码器20的处理硬件操作为对编码数据(E2)(即,上述三个数据流)进

行解码,以生成解码数据(D3)。如此生成的解码数据(D3)表示如下:

[0316] 0,12,40,156,180,28,0,16,240,40,40,172,8,0,68,0,0,40,120,248,64,0,  
 [0317] 116,4,204,96,60,240,192,52,172,16,16,60,32,200,192,220,188,28,180,0,  
 [0318] 28,252,192,144,80,16,0,252,184,8,0,0,  
 [0319] 0,12,40,156,180,28,0,16,240,40,40,172,8,0,68,0,0,40,120,12,64,0,116,  
 [0320] 4,204,80,60,240,192,52,172,16,16,60,32,200,192,220,188,28,180,0,28,  
 [0321] 252,200,204,80,16,0,0,172,204,0,0,  
 [0322] 0,12,40,156,180,28,0,16,240,40,40,172,8,0,68,0,0,40,120,16,64,0,116,  
 [0323] 4,204,76,60,240,192,52,172,16,16,60,32,200,192,220,188,28,180,0,28,  
 [0324] 252,208,64,80,16,0,0,168,84,0,0,  
 [0325] 0,12,40,156,180,28,0,16,240,40,40,172,8,0,68,0,0,40,120,24,64,0,116,  
 [0326] 4,204,68,60,240,192,52,172,16,16,60,32,200,192,220,188,28,180,0,28,  
 [0327] 252,220,164,80,16,0,0,152,244,0,0,  
 [0328] 0,12,40,156,180,28,0,16,240,40,40,172,8,0,68,0,0,40,120,64,64,0,116,  
 [0329] 4,204,24,60,240,192,52,172,16,16,60,32,200,192,220,188,28,180,0,28,  
 [0330] 252,232,148,80,16,0,0,144,0,0,0

[0331] 在上文中,已单独示出了五个数据块,并且仅为了清晰起见,已对数据块的部分重现的变化数据元素添加了下划线。

[0332] 输入数据(D1)的原始数据元素和解码数据(D3)的数据元素之间的均方误差之和为535。这对应于输入数据(D1)和解码数据(D3)之间的一些数据损失。

[0333] 限定参考符号:

[0334] 可选地,参考符号“0”用于指示非复制的数据块或数据包,即该数据块或数据包不是任意之前的数据块或数据包的副本。

[0335] 可选地,当有“N”个不同的参考数据块或数据包时,使用参考符号“1”至“N”来指示部分复制的数据块或数据包,即该数据块或数据包是它们相应的参考数据块或数据包的部分副本。在本文中,“N”为大于1的任意正整数。

[0336] 在上文中描述的先前示例中,由于没有数据块是非复制的数据块或数据包,所以不使用参考符号“0”。相反,由于后续数据块中的每个是其对应参考数据块的部分副本,所以使用参考符号“1”。

[0337] 另外,当存在可用于给定数据块或数据包的多个参考数据块或数据包时,可从多个参考数据块或数据包中选择参考数据块或数据包。在选择期间,无需对所选的参考数据块或数据包和给定数据块或数据包之间的匹配妥协的情况下,有益地考虑参考数据块或数据包相对于给定数据块或数据包的距离,以找到具有下限值的可能的参考符号。

[0338] 可选地,当输入数据(D1)较大和/或在输入数据(D1)中存在大量冗余时,单独的参考符号用于完全复制的数据块或数据包,即数据块或数据包是它们各自的参考数据块或数据包的完全副本。

[0339] 在编码器10的一个实现中,当存在“N”个不同的参考数据块或数据包时,参考符号可选地限定如下:

[0340] (i) 参考符号“0”用于指示非复制的数据块或数据包;

- [0341] (ii) 参考符号“1”至“N”用于指示部分复制的数据块或数据包;以及
- [0342] (iii) 参考符号“1+N”至“N+N”用于指示完全复制的数据块或数据包。
- [0343] 在本文中为了说明的目的,现将考虑在输入数据(D1)中识别出四个参考数据块的示例。这四个参考数据块被分配如下的参考符号:
- [0344] 参考符号“1”分配至第一参考数据块;
- [0345] 参考符号“2”分配至第二参考数据块;
- [0346] 参考符号“3”分配至第三参考数据块;以及
- [0347] 参考符号“4”分配至第四参考数据块。
- [0348] 在这个示例中,值“N”为4(N=4)。
- [0349] 现在,如果输入数据(D1)中的第五数据块是第二参考数据块的部分副本,那么第五数据块通过使用参考符号“2”表示。如果输入数据(D1)中的第六数据块是第二参考数据块的完全副本,那么第六数据块通过使用参考符号“6”表示,即“2+N”。
- [0350] 在替代的实现中,可选地使用一位符号来指示给定数据块或数据包是否为完全副本或部分副本。除了用于给定数据块或数据包的参考符号之外,还使用一位符号。在本文中,参考符号指示用于给定数据块或数据包的对应参考数据块或数据包,而一位符号指示给定数据块或数据包是否为对应参考数据块或数据包的完全副本或部分副本。一位符号具有两个值,即“0”和“1”。例如,一位符号的“0”值可用于指示部分副本,而一位符号的“1”值可用于指示完全副本,或反之亦然。
- [0351] 在另一些其它的替代实现方中,可通过使用块索引、数据索引、运动向量、数据库参考、编码方法来限定部分复制的数据块或数据包。
- [0352] 可选地,当单独的参考符号不用于完全复制的数据块或数据包时,则与完全复制的数据块或数据包的所有未变化数据元素有关的掩码位设置为“0”值。
- [0353] 另一方面,当单独的参考符号用于指示完全复制的数据块或数据包时,完全复制的数据块或数据包不需要掩码位。同样地,当单独的参考符号用于指示非复制的数据块或数据包时,非复制的数据块或数据包不需要掩码位。
- [0354] 另外,可选地,通过使用如专利申请GB 1222240.2、US 13/715,405中所描述的数据库对掩码位进行编码,所述专利通过引用并入本文。当存在掩码位重新出现的高可能性时,与同样地对掩码位进行熵编码相比较,通过将参考传递至掩码位容易提高数据压缩比。
- [0355] 接下来,还将参照附图,即图2至图5B描述本公开的实施方式。
- [0356] 图2是根据本公开的实施方式的、对输入数据(D1)进行编码以生成对应编码数据(E2)的方法的步骤的图示。方法描绘为逻辑流程图中的步骤集合,逻辑流程图表示可以硬件、软件、或它们的组合实现的一系列步骤。
- [0357] 在步骤102处,确定输入数据(D1)内的数据块或数据包的至少部分重现。
- [0358] 在步骤104处,通过采用至少一个参考符号以及多个掩码位,对未变化的和变化的数据块或数据包进行编码。如之前描述的,采用至少一个参考符号来指示相互类似的数据块或数据包的至少部分重现,和/或来指示输入数据(D1)内是否存在相互类似的数据块或数据包的重现。另外,采用例如实现为多个掩码位的变化符号来指示数据块或数据包的部分重现的、变化的和未变化数据元素以及输入数据(D1)内的变化的值。可替代地,例如以如上文描述的方式使用delta值。

[0359] 根据步骤104,如之前所描述的,将至少一个参考符号以及多个变化符号编码成多个数据流。

[0360] 结合图3A和图3B,对步骤102和步骤104的编码过程进行了描述。

[0361] 接下来,在步骤106处,对多个数据流进行熵编码以生成编码数据(E2)。

[0362] 步骤102至106仅为说明性的,并且在不背离本文权利要求的范围的情况下,还可以提供其他的替换,其中增加有一个或多个步骤、去除一个或多个步骤、或以不同的顺序提供一个或多个步骤。

[0363] 图3A和图3B共同为根据本公开的实施方式的编码处理的步骤的图示。

[0364] 在步骤202处,如之前所描述,读取输入数据(D1)的给定数据帧/图/通道,以确定用于数据块或数据包的最佳尺寸。

[0365] 根据步骤202,将给定数据帧/图/通道按照用于数据块或数据包的最佳尺寸拆分成数据块或数据包。

[0366] 在步骤204处,读取数据块或数据包以进行对它们的对应参考数据块或数据包的搜索。

[0367] 接下来,在步骤206处,对于给定数据帧/图/通道中的给定数据块或数据包,确定给定数据块或数据包是否为参考数据块或数据包的副本。

[0368] 在步骤206处,如果确定出给定数据块或数据包不是副本,则执行步骤208。在步骤208处,为给定数据块或数据包写入“非复制的”参考符号。在一个示例中,将“非复制的”参考符号限定为参考符号“0”。另外,将给定数据块或数据包的所有数据值编码至编码数据(E2)中。

[0369] 在步骤206处,如果确定出给定数据块或数据包是参考数据块或数据包的副本,则执行步骤210。在步骤210处,确定给定数据块是否为参考数据块或数据包的完全副本或部分副本。当使用了有损耗操作方式时,按照用于有损耗操作方式所设置的质量水平来执行步骤210。

[0370] 在步骤210处,如果确定出给定数据块或数据包是部分副本,则执行步骤212。在步骤212处,为给定数据块或数据包写入指示参考数据块或数据包的参考符号。参考符号被用作指示器,该指示器指示参考数据块或数据包相对于给定数据块或数据包的位置。在一个示例中,当存在“N”个不同的参考数据块或数据包时,从参考符号“1”至“N”中选择参考符号。另外,将给定数据块或数据包的变化数据元素的数据值以及它们关联的掩码位编码到编码数据(E2)中。

[0371] 在步骤210处,如果确定出给定数据块或数据包是完全副本,则执行步骤214。在步骤214处,为给定数据块或数据包写入指示参考数据块或数据包的参考符号。在一个示例中,当存在“N”个不同的参考数据块或数据包时,从参考符号“1+N”至“N+N”中选择参考符号。在另一示例中,除了参考符号之外,写入指示完全副本的一位符号。

[0372] 步骤216接着步骤208、步骤212、以及步骤214。在步骤216处,确定在给定数据帧/图/通道中是否存在下一个数据块或数据包。如果确定存在下一个数据块或数据包,在步骤204处重新开始编码处理。否则,如果确定给定数据帧图/通道中不存在下一个数据块或数据包,则执行步骤218。

[0373] 在步骤218处,确定输入数据(D1)中是否存在下一个数据帧/图/通道。如果确定存

在下一个数据帧/图/通道,则在步骤202处重新开始编码处理。否则,如果确定在输入数据(D1)中不存在下一个数据帧/图/通道,则停止编码处理。

[0374] 另外,为编码处理而分配的后台存储器的量仅与当前数据块或数据包中的数据元素的量乘以可能参考的数量(即,数据块或包精确的参考索引)一样大,或者与当前数据块或数据包中的数据元素的量加上可能参考的数量减去一(即,数据值精确参考索引)一样大。最少地,其是具有一个数据块或数据包的存储器。另外,编码处理的结果(即编码数据(E2))可直接写入或者传送至原始存储器。这意味着不需要单独的传送存储器。因而,本方法能够起现场操作的作用,并且具有较高的成本效率。该方法例如可用于便宜的消费电子媒体产品,诸如智能电话、MP3播放器、平板电脑、电视、音频高保真(“hifi”)设备、电子书等。

[0375] 步骤202至218仅为说明性的,并且在不背离本文权利要求的范围的情况下,还可提供其他替换,其中增加有一个或多个步骤、去除一个或多个步骤、或以不同的顺序提供一个或多个步骤。例如,在不将单独的参考符号用于完全复制的数据块或数据包的替代实现中,去除了步骤210和步骤214。现在,在步骤206处,如果确定给定数据块或数据包是参考数据块或数据包的副本,则执行步骤212,而不是步骤210。

[0376] 本公开的实施方式提供了包括非短暂性(即,非瞬时性)计算机可读存储介质的计算机程序产品,在计算机存储介质上存储有计算机可读指令,计算机可读指令通过包括处理硬件的计算机化设备执行以执行如结合图2和图3A-3B所描述的方法。可选地,计算机可读指令可从软件应用商店,例如从“App商店”下载到计算机化设备。

[0377] 图4是根据本公开的实施方式的、对编码数据(E2)进行解码以生成对应解码数据(D3)的方法的步骤的图示。方法描绘为逻辑流程图中的许多步骤,逻辑流程图表示可以硬件、软件、或它们的组合实现的一序列步骤。

[0378] 在步骤302处,对编码数据(E2)进行熵解码以提供多个数据流。如之前所描述的,多个数据流包括以下中的一个或多个:

[0379] (i) 第一数据流,包括至少一个参考数据块和/或数据包的数据值,以及后续数据块的变化数据元素的数据值;

[0380] (ii) 第二数据流,包括多个掩码位;和/或

[0381] (iii) 第三数据流,包括至少一个参考符号。

[0382] 在步骤304处,对来自多个数据流的未变化的以及变化的数据块进行解码。根据步骤304,对至少一个参考符号以及多个变化符号进行解码,以生成用于数据块或数据包的部分重现的数据,并且生成用于数据块或数据包的部分重现的变化数据元素的数据。

[0383] 结合图5A和图5B,描述了步骤304的解码过程。

[0384] 接下来,在步骤306处,将对于数据块或数据包的至少部分重现生成的数据以及对于数据块或数据包的部分重现的变化数据元素生成的数据集合,以生成解码数据(D3)。

[0385] 步骤302至306仅为说明性的,并且在不背离本文权利要求的范围的情况下,还可以提供其他的替代,其中增加一个或多个步骤、去除一个或多个步骤、或以不同的顺序提供一个或多个步骤。

[0386] 图5A和图5B共同为根据本公开的实施方式的解码处理的步骤的图示。

[0387] 在步骤402处,从编码数据(E2)中提供的第三数据流读取给定参考符号。

[0388] 在步骤404处,确定给定参考符号是否为“非复制的”参考符号。换言之,确定给定参考符号是否对应于非复制的数据块或数据包。

[0389] 在步骤404处,如果确定出给定参考符号是“非复制”的参考符号,那么执行步骤406。在步骤406处,对非复制的数据块或数据包的所有数据值进行解码,以生成用于非复制的数据块或数据包的数据。

[0390] 在步骤404处,如果确定出给定参考符号不是“非复制”的参考符号,那么执行步骤408。在步骤408处,确定给定参考符号是否为“完全复制”的参考符号。换言之,确定给定参考符号是否对应于完全复制的数据块或数据包。

[0391] 在步骤408处,如果确定给定参考符号不是“完全复制”的参考符号,那么执行步骤410。在步骤410处,对与部分复制的数据块或数据包的数据元素有关的掩码位进行检查,从而对部分复制的数据块或数据包的变化数据元素的数据值进行解码。另外,从给定参考符号指向的参考数据块或数据包中获取未变化元素的数据值。根据步骤410,将未变化元素的数据值与变化数据元素的数据值组合,以生成用于部分复制的数据块或数据包的数据。

[0392] 在步骤408处,如果确定给定参考符号是“完全复制”的参考符号,那么执行步骤412。在步骤412处,从给定参考符号指向的参考数据块中获取未变化元素的数据值。对未变化元素的数据值进行设置,以生成用于完全复制的数据块或数据包的数据。

[0393] 步骤414接着步骤406、步骤410、以及步骤412。在步骤414处,确定第三数据流中是否存在下一个参考符号。如果确定存在下一个参考符号,那么在步骤402处重新开始解码处理。

[0394] 否则,如果确定不存在下一个参考符号,则执行步骤416。在步骤416处,将在步骤406、步骤410、以及步骤412处生成的用于数据块或数据包的数据集合,以生成解码数据(D3)。

[0395] 应理解,不管所有数据是否完成,每当对这样一部分数据的解码完成时,单独更新通道、图或帧是可能的。

[0396] 步骤402至416仅为说明性的,并且在不背离本文权利要求的范围的情况下,还可以提供其他替换,其中增加一个或多个步骤、去除一个或多个步骤、或以不同的顺序提供一个或多个步骤。

[0397] 本公开的实施方式提供了包括非短暂性(即,非瞬时性)计算机可读存储介质的计算机程序产品,在计算机可读存储介质上存储有计算机可读指令,计算机可读指令可通过包括处理硬件的计算机化设备执行,以执行如结合图4和图5A-B所描述的方法。可选地,计算机可读指令可从软件应用商店,例如从“App商店”下载到计算机化设备。

[0398] 在不背离如所附权利要求限定的本发明的范围的情况下,可对上文描述的本发明的实施方式进行修改。用于描述并限定本发明的诸如“包括(including)”、“包括(comprising)”、“合并(incorporating)”、“包括(consisting of)”、“具有(have)”、以及“是(is)”的表达旨在以非排他性方式解释,即允许还表现未明确描述的项、成分或元素。提及的单数还解释为涉及复数。包括在所附权利要求中的括号内的数字旨在帮助理解权利要求,并且不以任何方式解释为限制通过这些权利要求所要求保护的主体。

[0399] 附录

[0400]

[1]	熵 (信息理论) -Wikipedia, 免费百科全书 (2013 年 9 月 27 日访问), URL:	<a href="http://en.wikipedia.org/wiki/Entropy_%28information_theory%29">http://en.wikipedia.org/wiki/Entropy_%28information_theory%29</a>
[2]	Shannon, Claude E. (1948) (2012 年 11 月 28 日访问), 通信的数学理论, URL:	<a href="http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf">http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf</a>
[3]	无损压缩-Wikipedia, 免费百科全书 (2013 年 9 月 27 日访问), URL:	<a href="http://en.wikipedia.org/wiki/Lossless_compression">http://en.wikipedia.org/wiki/Lossless_compression</a>
[4]	有损压缩-Wikipedia, 免费百科全书 (2013 年 9 月 27 日访问), URL:	<a href="http://en.wikipedia.org/wiki/Lossy_compression">http://en.wikipedia.org/wiki/Lossy_compression</a>
[5]	数据复制删除-Wikipedia, 免费百科全书(2013 年 9 月 27 日访问), URL:	<a href="http://en.wikipedia.org/wiki/Data_deduplication">http://en.wikipedia.org/wiki/Data_deduplication</a>
[6]	超文本传输协议-HTTP/1.1 (2013 年 9 月 27 日访问), URL:	<a href="http://www.w3.org/Protocols/rfc2616/rfc2616.html">http://www.w3.org/Protocols/rfc2616/rfc2616.html</a>
[7]	实时消息传送协议 (RTMP) 规范  Adobe 开发人员连接 (2013 年 9 月 27 日访问), URL:	<a href="http://www.adobe.com/devnet/rtmp.html">http://www.adobe.com/devnet/rtmp.html</a>
[8]	Windows Azure: 微软云平台 云主机 云服务 (2013 年 9 月 27 日访问), URL:	<a href="http://www.windowsazure.com/en-us/">http://www.windowsazure.com/en-us/</a>
[9]	摩尔定律-Wikipedia, 免费百科全书 (2013 年 9 月 27 日访问), URL:	<a href="http://en.wikipedia.org/wiki/Moore%27s_law">http://en.wikipedia.org/wiki/Moore%27s_law</a>
[10]	US 专利 US7643505B1 (Colloff): “实时压缩和解压缩的方法和系统”	



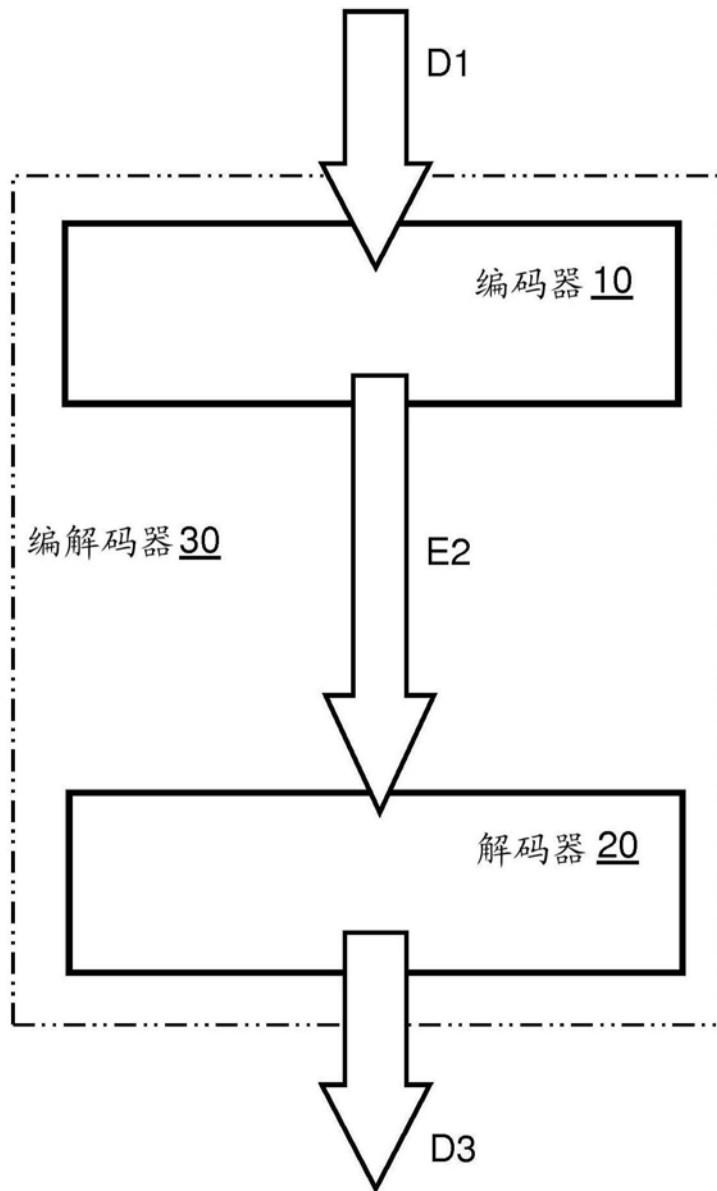


图1

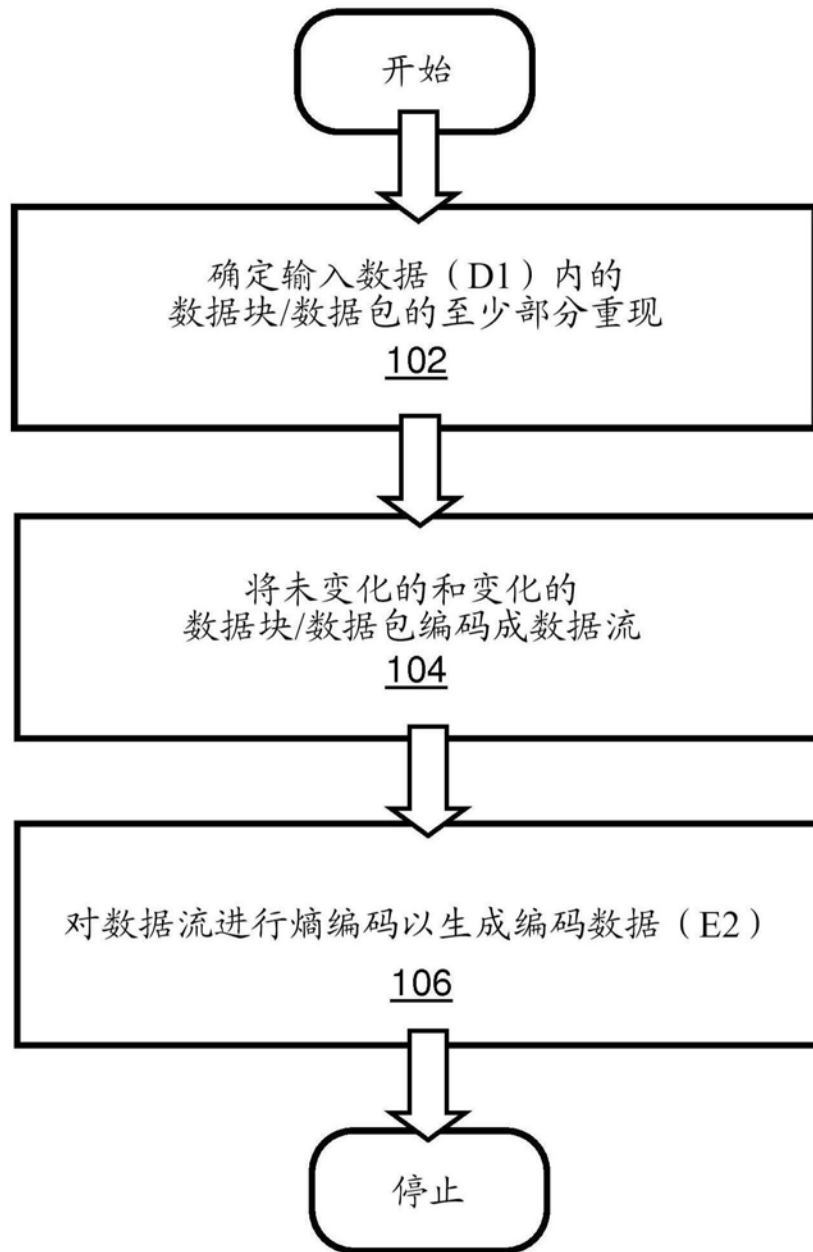


图2

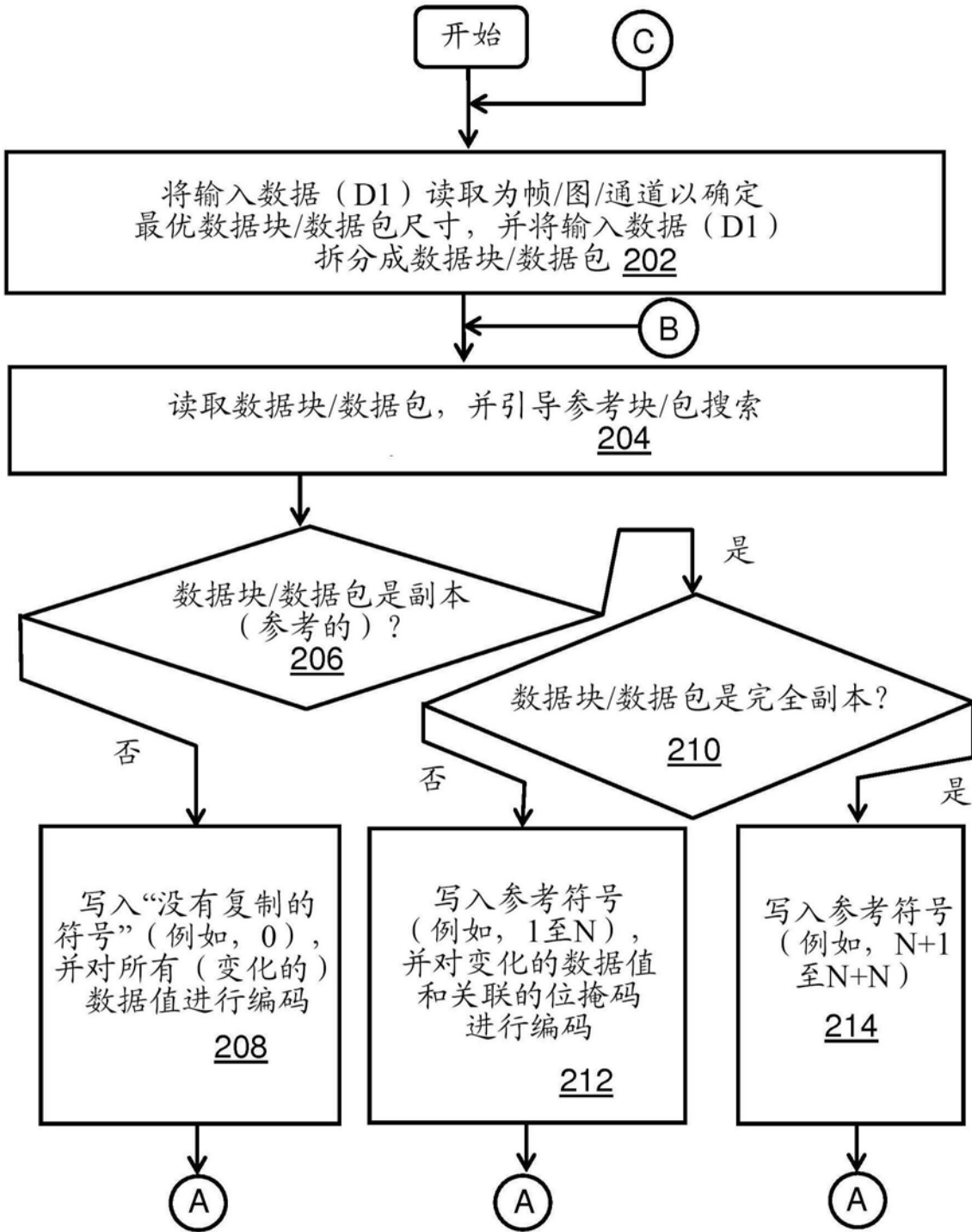


图3A

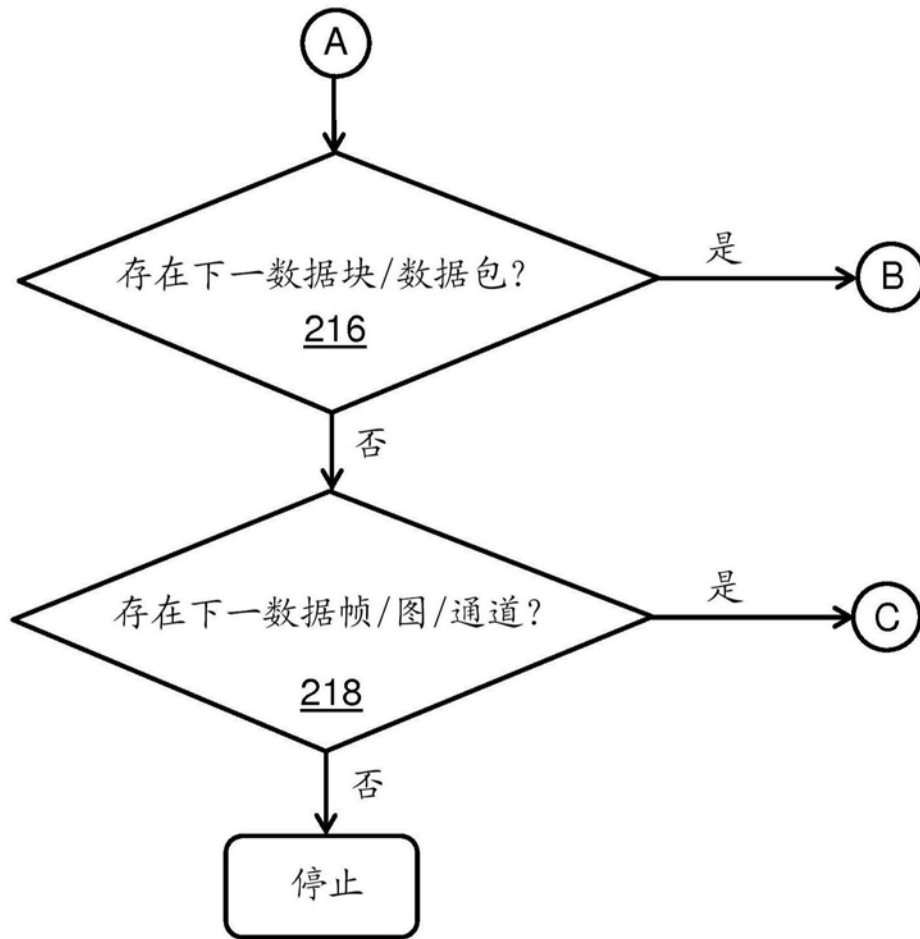


图3B

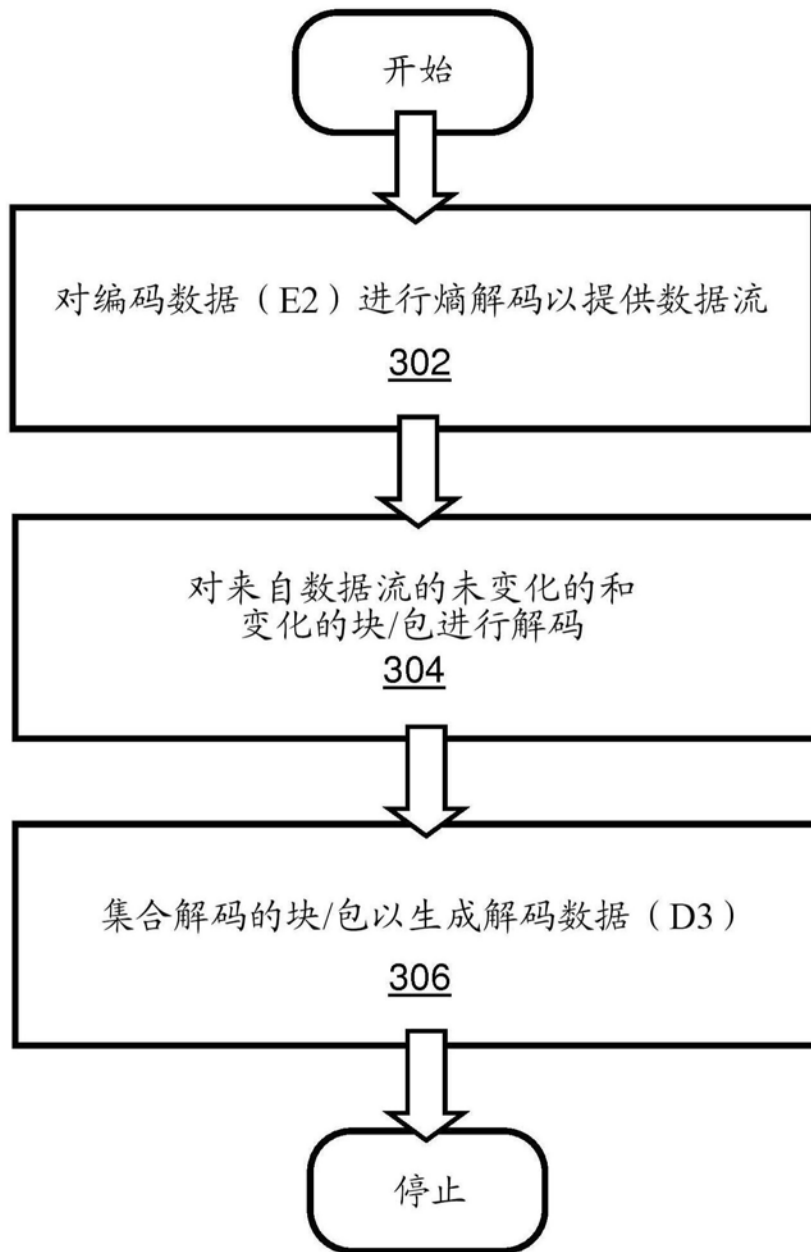


图4

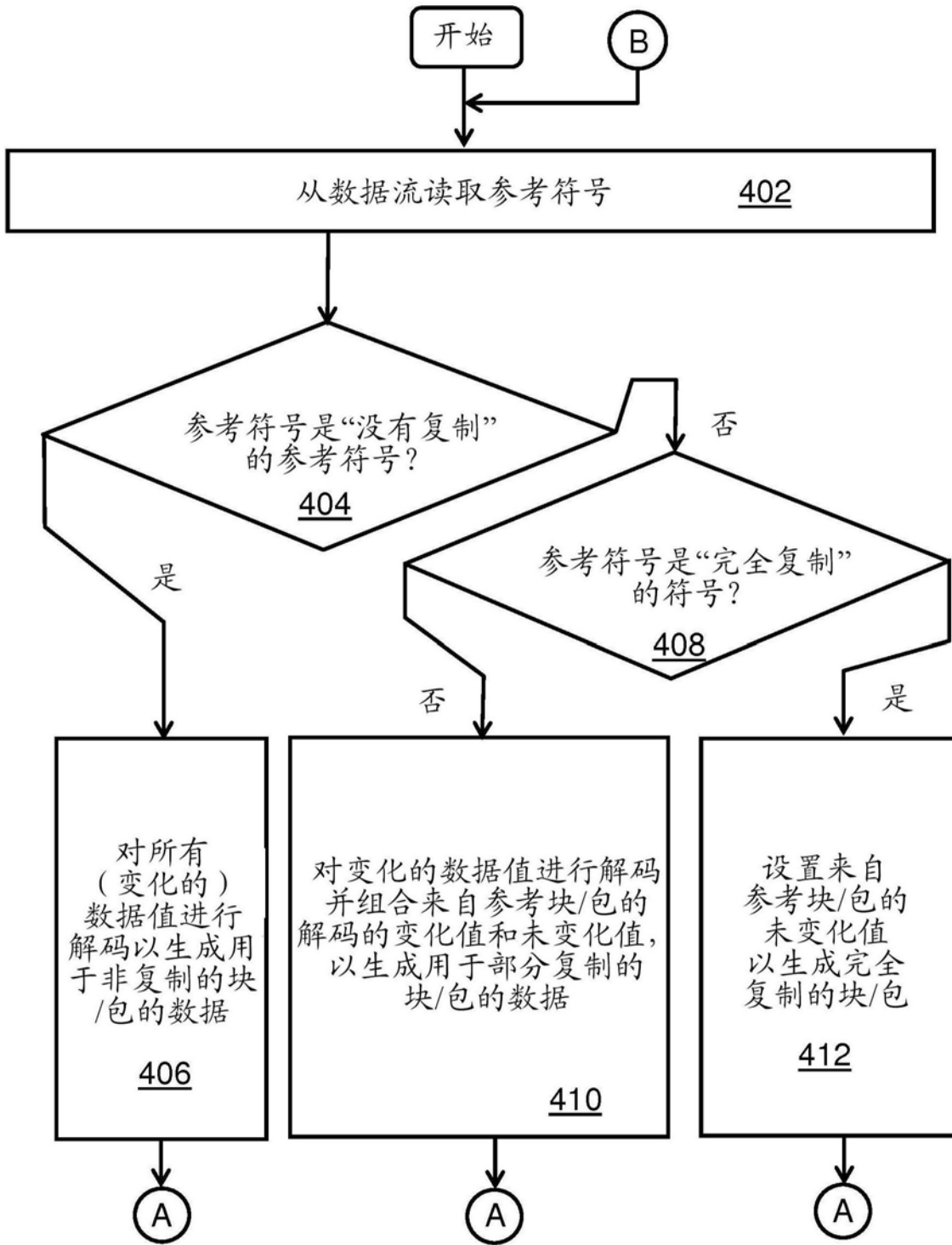


图5A

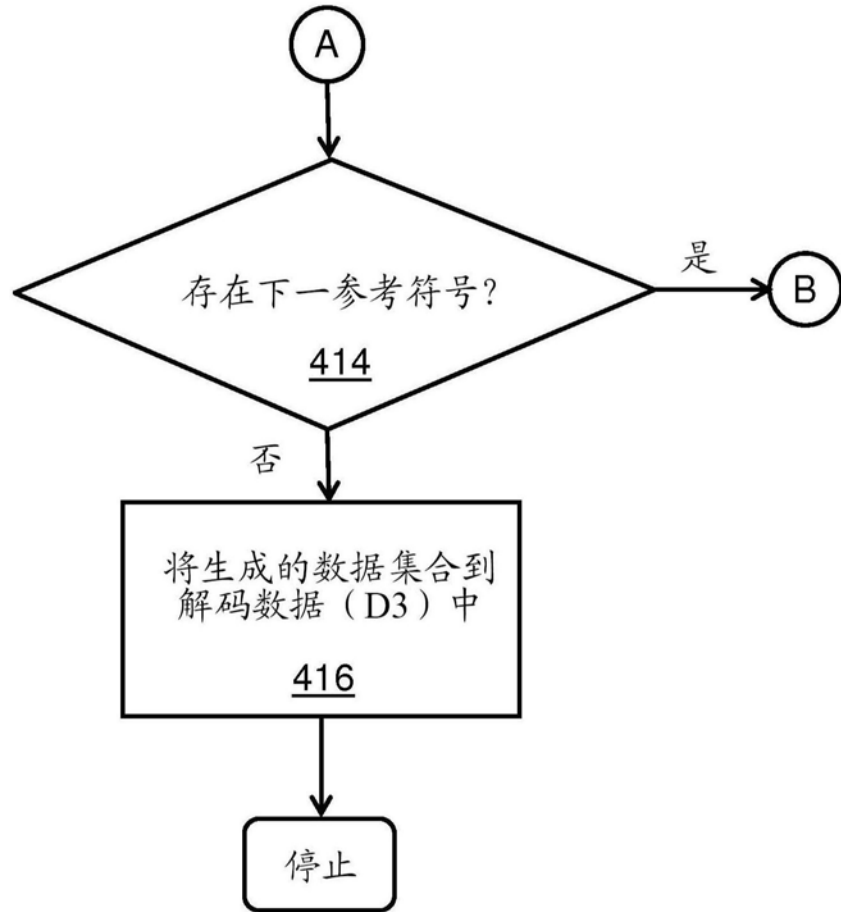


图5B