



(12)发明专利

(10)授权公告号 CN 103780263 B

(45)授权公告日 2017.04.12

(21)申请号 201310349363.1

(51)Int.Cl.

H03M 7/30(2006.01)

(22)申请日 2013.08.12

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 103780263 A

US 2004/0181648 A1, 2004.09.16,
US 2010/0121826 A1, 2010.05.13,
US 2010/0281004 A1, 2010.11.04,
US 5920723 A, 1999.07.06,
US 2011/0219357 A1, 2011.09.08,

(43)申请公布日 2014.05.07

审查员 孙韬敏

(30)优先权数据

10-2012-0117519 2012.10.22 KR

(73)专利权人 株式会社特迈数据

地址 韩国京畿道城南市

(72)发明人 安宰奭 朴相永

(74)专利代理机构 隆天知识产权代理有限公司

72003

代理人 金相允 向勇

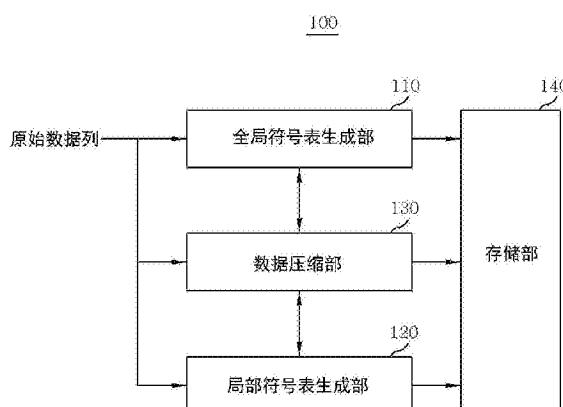
权利要求书2页 说明书6页 附图3页

(54)发明名称

数据压缩装置、数据压缩方法及记录介质

(57)摘要

本发明涉及数据压缩装置、数据压缩方法及记录介质，本发明是利用要压缩的整体数据中的一部分来生成对于重复的数据的全局符号表，生成不与全局符号表重复的局部符号表，来以块为单位压缩数据，从而能够增加压缩效率的技术。



1. 一种数据压缩装置,其特征在于,
包括:

全局符号表生成部,在输入的数据表的多个原始数据列中,搜索既定的基准块大小的原始数据列来选定为全局压缩对象,并生成对于上述全局压缩对象的全局符号表,

局部符号表生成部,在每当输入上述原始数据列时,搜索重复的原始数据列来选定为局部压缩对象,并生成对于上述局部压缩对象的局部符号表,以及

数据压缩部,根据上述全局符号表或上述局部符号表压缩上述原始数据列来生成全局压缩数据或局部压缩数据,将上述全局压缩数据或局部压缩数据生成为既定的大小的压缩块;

选定为上述局部压缩对象和上述全局压缩对象的原始数据列互不相同。

2. 根据权利要求1所述的数据压缩装置,其特征在于,上述数据压缩部在生成上述全局符号表之前,根据上述局部符号表压缩上述原始数据列来生成上述局部压缩数据。

3. 根据权利要求1所述的数据压缩装置,其特征在于,上述数据压缩部在每当输入上述原始数据列时,判断用于存储上述局部压缩数据或全局压缩数据的块的大小是否大于既定的大小,来生成上述压缩块。

4. 根据权利要求3所述的数据压缩装置,其特征在于,上述数据压缩部在用于存储上述局部压缩数据或全局压缩数据的块的大小大于既定的大小的情况下,将与当前为止输入的上述原始数据列相对应的上述局部压缩数据或全局压缩数据生成为上述压缩块。

5. 根据权利要求1所述的数据压缩装置,其特征在于,上述全局符号表生成部包括:

全局缓冲区,临时存储所输入的多个上述原始数据列;

全局数据模式搜索部,搜索临时存储的多个上述原始数据列中相同的原始数据列重复出现频率为基准来选定上述全局压缩对象;以及

全局符号分配部,向上述全局压缩对象分配全局符号。

6. 根据权利要求5所述的数据压缩装置,其特征在于,上述全局数据模式搜索部根据上述重复的原始数据列的长度和出现频率来选定上述全局压缩对象。

7. 根据权利要求1所述的数据压缩装置,其特征在于,上述局部符号表生成部包括:

局部缓冲区,临时存储所输入的多个上述原始数据列;

局部数据模式搜索部,搜索临时存储的多个上述原始数据列中相同的原始数据列重复出现频率为基准来选定上述局部压缩对象;以及

局部符号分配部,向上述局部压缩对象分配局部符号。

8. 根据权利要求7所述的数据压缩装置,其特征在于,在生成了上述全局符号表的情况下,上述局部数据模式搜索部将选定为上述全局压缩对象的原始数据列从上述局部压缩对象中排除。

9. 根据权利要求7所述的数据压缩装置,其特征在于,在每当输入上述原始数据列时,上述局部数据模式搜索部更新上述局部压缩对象。

10. 根据权利要求1所述的数据压缩装置,其特征在于,上述基准块的大小设定为大于上述压缩块的大小。

11. 根据权利要求1所述的数据压缩装置,其特征在于,上述局部符号表存储于上述压缩块内。

12. 根据权利要求11所述的数据压缩装置,其特征在于,上述全局符号表存储于与存储上述全局压缩数据及上述局部压缩数据的空间相独立的空间。

13. 一种数据压缩方法,其特征在于,包括:

加载数据表的原始数据列的步骤;

针对既定的基准块的大小的上述原始数据列选定全局压缩对象的步骤;

判断是否已生成全局符号表的步骤,判断是否已生成对于上述全局压缩对象的全局符号表;

在生成了上述全局符号表的情况下,选定与上述全局压缩对象不同的局部压缩对象的步骤;

生成对于上述局部压缩对象的局部符号表的步骤;

根据上述全局符号表及上述局部符号表压缩上述原始数据列来生成全局压缩数据或局部压缩数据的步骤;以及

生成压缩块的步骤,存储既定的块的大小的上述原始数据列、上述全局压缩数据或上述局部压缩数据来生成压缩块。

14. 根据权利要求13所述的数据压缩方法,其特征在于,上述判断是否已生成全局符号表的步骤还包括如下步骤:若为生成上述全局符号表之前,则搜索上述原始数据列,将重复的原始数据列选定为上述局部压缩对象。

15. 根据权利要求14所述的数据压缩方法,其特征在于,还包括如下步骤:若为生成上述全局符号表之前,则根据上述局部符号表压缩上述原始数据列来生成局部压缩数据。

16. 根据权利要求13所述的数据压缩方法,其特征在于,上述生成压缩块的步骤包括如下步骤:每当输入上述原始数据列时,判断用于存储上述局部压缩数据或全局压缩数据的块的大小是否大于上述既定的大小。

17. 根据权利要求16所述的数据压缩方法,其特征在于,上述判断用于存储上述局部压缩数据或全局压缩数据的块的大小是否大于上述既定的大小的步骤还包括如下步骤:在用于存储上述局部压缩数据或全局压缩数据的块的大小大于既定的大小的情况下,将与当前为止输入的上述原始数据列相对应的上述局部压缩数据或全局压缩数据生成为上述压缩块。

18. 根据权利要求13所述的数据压缩方法,其特征在于,上述选定局部压缩对象的步骤包括每当输入上述原始数据列时更新上述局部压缩对象的步骤。

数据压缩装置、数据压缩方法及记录介质

技术领域

[0001] 本发明涉及数据压缩装置、数据压缩方法及用于该方法的计算机可读取的记录介质。更为详细地，本发明是利用要压缩的整体数据中的一部分来生成对于重复的数据的全局符号表，生成不与全局符号表重复的局部符号表，来以块为单位压缩数据，从而能够增加压缩效率的数据压缩方法装置、数据压缩方法及用于该方法的计算机可读取的记录介质的相关技术。

[0002] 本发明从作为知识经济部的产业融合源泉技术开发事业的一环来执行的研究中导出(10040937，用于日本大型计算机(日立、富士通)的集成解决方案的开发)。

背景技术

[0003] 数据压缩为用于将数据有效地存储于更小的存储空间的技术。为此，广泛使用在向原始数据适用特定压缩算法后，包含能够识别压缩算法的种类的信息来生成压缩数据的技术。通常，生成压缩数据的这种技术由甲骨文(Oracle)、结构化查询语言(SQL, structured query language)服务器、DB2等的数据库管理系统提供。在这种数据库管理系统中，向出现频率高的数据列分配例如数字或文字等符号来形成符号表，在压缩数据时利用符号表来压缩数据列。

发明内容

[0004] 作为以往的压缩方法中的一个，有在加载整体数据列后生成对于整体数据列的符号表来压缩数据的方法。在这种情况下，由于无法对在生成符号表之后新出现的重复的数据列进行压缩，因而存在降低压缩效率的问题。

[0005] 并且，作为其他方法，有加载整体数据列中的规定大小的数据列，则生成对于相关数据列的符号表并以一个块为单位来压缩数据的方法。在这种情况下，在一个块中出现的重复的数据列存在于多个块的情况下，重复的数据列整体存储于多个块内的符号表，因而存在降低压缩效率的问题。

[0006] 因此，本发明提供在向所要压缩的数据中插入新数据或出现在一个块的重复的数据列存在于多个块的情况下，也能够提高压缩效率的数据压缩装置、数据压缩方法及用于该方法的计算机可读取的记录介质。

[0007] 上述的本发明为数据压缩装置，其特征在于，包括：全局符号表生成部，在输入的数据表的原始数据列中，搜索既定的基准块大小的原始数据列来选定全局压缩对象，并生成对于上述全局压缩对象的全局符号表，局部符号表生成部，在每当输入上述原始数据列时，搜索重复的原始数据列来选定为局部压缩对象，并生成对于上述局部压缩对象的局部符号表，以及数据压缩部，根据上述全局符号表或上述局部符号表压缩上述原始数据列来生成全局压缩数据或局部压缩数据，将上述全局压缩数据或局部压缩数据生成为既定的大小的压缩块；选定为上述局部压缩对象和上述全局压缩对象的原始数据列互不相同。

[0008] 而且，本发明的特征在于，上述数据压缩部在生成上述全局符号表之前，根据上述

局部符号表压缩上述原始数据列来生成上述局部压缩数据。并且，本发明的特征在于，上述数据压缩部在每当输入上述原始数据列时，判断用于存储上述局部压缩数据或全局压缩数据的块的大小是否大于既定的大小，来生成上述压缩块。

[0009] 而且，本发明的特征在于，上述数据压缩部在存储上述局部压缩数据或全局压缩数据的块的大小大于既定的大小的情况下，将与当前为止输入的上述原始数据列相对应的上述局部压缩数据或全局压缩数据生成为上述压缩块。

[0010] 而且，本发明的特征在于，上述全局符号表生成部包括：全局缓冲区，临时存储上述原始数据列；全局数据模式搜索部，搜索临时存储的上述原始数据列的出现频率抽取重复的原始数据列来选定上述全局压缩对象；以及全局符号分配部，向上述全局压缩对象分配全局符号。

[0011] 而且，本发明的特征在于，上述全局数据模式搜索部根据上述重复的原始数据列的长度和出现频率来选定上述全局压缩对象。而且，本发明的特征在于，上述局部符号表生成部包括：局部缓冲区，临时存储上述原始数据列；局部数据模式搜索部，搜索临时存储的上述原始数据列的出现频率抽取重复的原始数据列来选定上述局部压缩对象；以及局部符号分配部，向上述局部压缩对象分配局部符号。

[0012] 而且，本发明的特征在于，在生成上述全局符号表的情况下，上述局部数据模式搜索部将选定为上述全局压缩对象的原始数据列从上述局部压缩对象中排除。并且，本发明的特征在于，在每当输入上述原始数据列时，上述局部数据模式搜索部更新上述局部压缩对象。而且，本发明的特征在于，上述基准块的大小设定为大于上述压缩块的大小。本发明的特征在于，上述局部符号表存储于上述压缩块内。本发明的特征在于，上述全局符号表存储于与存储上述全局压缩数据及上述局部压缩数据的空间相独立的空间。

[0013] 而且，本发明作为数据压缩方法，包括：加载数据表的原始数据列的步骤；针对既定的基准块的大小的上述原始数据列选定全局压缩对象的步骤；判断是否已生成全局符号表的步骤，判断是否已生成对于上述全局压缩对象的全局符号表；在生成了上述全局符号表的情况下，选定与上述全局压缩对象不同的局部压缩对象的步骤；生成对于上述局部压缩对象的局部符号表的步骤；根据上述全局符号表及上述局部符号表压缩上述原始数据列来生成全局压缩数据或局部压缩数据的步骤；以及生成压缩块的步骤，存储既定的块的大小的上述原始数据列、上述全局压缩数据或上述局部压缩数据来生成压缩块的步骤。

[0014] 而且，本发明的特征在于，在上述判断是否已生成全局符号表的步骤还包括如下步骤：若为生成上述全局符号表之前，则搜索上述原始数据列，将重复的原始数据列选定为上述局部压缩对象。在这里，本发明的特征在于，还包括如下步骤：若为生成上述全局符号表之前，则根据上述局部符号表压缩上述原始数据列来生成局部压缩数据。

[0015] 而且，本发明的特征在于，上述生成压缩块的步骤包括如下步骤：每当输入上述原始数据列时，判断用于存储上述局部压缩数据或全局压缩数据的块的大小是否大于上述既定的大小。

[0016] 而且，本发明的特征在于，在上述判断用于存储上述局部压缩数据或全局压缩数据的块的大小是否大于上述既定的大小的步骤中，还包括如下步骤：在用于存储上述局部压缩数据或全局压缩数据的块的大小大于既定的大小的情况下，将与当前为止输入的上述原始数据列相对应的上述局部压缩数据或全局压缩数据生成为上述压缩块。而且，本发明

的特征在于，上述选定局部压缩对象的步骤包括每当输入上述原始数据列时更新上述局部压缩对象的步骤。

[0017] 并且，本发明提供一种能够由计算机读取的记录介质，上述能够由计算机读取的记录介质记录有用于执行根据上述步骤中的至少某一个步骤所述的方法的计算机程序。

[0018] 对于数据压缩装置、数据压缩方法及用于该方法的计算机可读取的记录介质而言，本发明利用所要压缩的整体数据中的一部分来生成对于重复的数据的全局符号表，并且生成与全局符号表不重复的局部符号表，来以块为单位压缩数据，从而提供能够增加压缩效率的效果。

附图说明

[0019] 图1为表示本发明实施例的数据压缩装置的框图。

[0020] 图2为图1所示的全局符号表生成部的详细框图。

[0021] 图3为图1所示的局部符号表生成部的详细框图。

[0022] 图4为图1所示的数据压缩部的详细框图。

[0023] 图5为用于说明本发明实施例的数据压缩方法的流程图。

具体实施方式

[0024] 后述的对本发明的详细说明参照将能够实施本发明的特定实施例作为例示来图示的附图。将详细说明这些实施例以使技术人员能够充分实施本发明。本发明的多种实施例互不相同，但是应理解为没有必要相互排他。例如，记载于此的特定形状、结构及特性与一实施例相关地在不脱离本发明的精神及范围的同时能够以其他实施例来体现。并且，各公开的实施例内的个别结构要素的位置或配置应理解为在不脱离本发明的精神及范围的同时能够进行变更。因此，后述的详细说明并不具有限定性意义，若适当进行说明，本发明的范围仅仅由权利要求书中的主张相等同的所有范围和所附的权利要求书来限定。图中类似的附图标记指在多个方面相同或类似的功能。

[0025] 以下，参照附图来对本发明的优选实施例进行详细说明，以使本发明所属技术领域的普通技术人员能够容易实施本发明。

[0026] 图1为表示本发明实施例的数据压缩装置的框图。

[0027] 参照图1，本发明的实施例的数据压缩装置100包括全局符号表生成部110、局部符号表生成部120、数据压缩部130以及存储部140。全局符号表生成部110对所输入的数据表的原始数据列以既定的大小的一个基准块为单位搜索重复的原始数据列来选定全局压缩对象，并生成与全局压缩对象的原始数据列相对应的全局符号表。

[0028] 局部符号表生成部120以既定的大小的一个块为单位，对压缩原始数据列为止输入的原始数据列搜索重复的原始数据列来选定局部压缩对象，并生成与局部压缩对象的原始数据列相对应的局部符号表。在这里，优选地，本发明的实施例的所选定的局部压缩对象的原始数据列与全局压缩对象的原始数据列互不相同。

[0029] 而且，数据压缩部130根据全局符号表的完成与否来动态性地压缩所输入的原始数据列而生成压缩块，并存储于存储部140。具体地，数据压缩部130利用全局符号表以及局部符号表来生成压缩判断数据，根据压缩判断数据来生成压缩块，上述压缩块包含原始数

据列、局部压缩数据或全局压缩数据。在这里，压缩判断数据包含：选择信息，用于选择是否压缩原始数据列；大小信息，用于判断对于当前为止累积的原始数据列的压缩大小。

[0030] 图2为图1所示的全局符号表生成部的详细框图。

[0031] 参照图2，本发明的实施例的全局符号表生成部110包括全局缓冲区112、全局数据模式搜索部114以及全局符号分配部116。在这里，全局缓冲区112按顺序临时存储所输入的原始数据列。

[0032] 而且，全局数据模式搜索部114搜索通过全局缓冲区112输入的原始数据列的出现频率来抽取重复的数据列。全局数据模式搜索部114根据重复的数据列的长度和出现频率来选定全局压缩对象。

[0033] 具体地，全局数据模式搜索部114按照重复的数据列的长度相对长、出现频率多的顺序来将n个数据列选定为全局压缩对象。在这里，优选地，全局数据模式搜索部114将所输入的整体原始数据列分为至少一个的基准块并按照各基准块抽取重复的数据列。

[0034] 例如，在整体原始数据列的大小为10G字节的情况下，全局数据模式搜索部114可搜索10M字节的原始数据列来选定全局压缩对象。为此，全局数据模式搜索部114从全局缓冲区112中删除在重复的数据列中出现频率低的数据列，来能够形成可供下一个原始数据列存储于全局缓冲区112的空间。

[0035] 全局符号分配部116向选定为全局压缩对象的数据列分配数字或文字等的全局符号来完成全局符号表并存储于存储部140。

[0036] 图3为图1所示的局部符号表生成部的详细框图。

[0037] 参照图3，本发明的实施例的局部符号表生成部120包括局部缓冲区122、局部数据模式搜索部124以及局部符号分配部126。局部缓冲区122按照顺序临时存储所输入的原始数据列。

[0038] 局部数据模式搜索部124抽取通过局部缓冲区122输入的原始数据列中重复的数据列来选定局部压缩对象。并且，在每当输入原始数据列时，局部数据模式搜索部124判断重复的数据列的长度和出现频率来更新局部压缩对象。

[0039] 在这里，优选地，本发明的实施例的局部数据模式搜索部124根据是否完成全局符号表来选定局部压缩对象。具体地，在完成全局符号表的情况下，局部数据模式搜索部124排除全局压缩对象的原始数据列来选定局部压缩对象。由此，能够防止存储于全局符号表和局部符号表的数据列互相重复。

[0040] 而且，局部符号分配部126向选定为局部压缩对象的数据列分配数字或文字等的局部符号来生成局部符号表。在这里，优选地，每当更新局部压缩对象时也一并更新局部符号表。

[0041] 图4为图1所示的数据压缩部的详细框图。

[0042] 参照图4，本发明的实施例的数据压缩部130包括判断部132、压缩判断数据生成部134以及压缩块生成部136。在这里，判断部132对关于全局符号表的信息进行判断。判断部132判断是否存在全局符号表，判断结果为不存在全局符号表的情况下，向全局符号表生成部110请求生成全局符号表。而且，在全局符号表存在的情况下，判断部132判断是否完成全局符号表。在完成全局符号表的情况下，判断部132向局部符号表生成部120以及压缩判断数据生成部134传递全局符号表的完成信息。

[0043] 压缩判断数据生成部134利用局部符号表以及全局符号表来生成压缩判断数据。具体地,压缩判断数据生成部134根据局部符号表或全局符号表来预期压缩原始数据列时的大小。预期结果,在数据的大小大于压缩之前的情况下,压缩判断数据生成部134生成维持原始数据列的选择信息,在数据的大小小于压缩之前的情况下,生成压缩原始数据列的选择信息。

[0044] 而且,压缩判断数据生成部134利用选择信息来对累积的原始数据列进行将要压缩的整体预期数据的大小与既定的压缩块的大小的比较来生成大小信息。在这里,优选地,压缩块的大小小于基准块。例如,在基准块为10M字节的大小的情况下,可将压缩块的大小设定为8K字节。

[0045] 压缩块生成部136根据压缩判断数据,来在将要压缩原始数据列的整体预期数据的大小为既定的压缩块的大小的情况下,将当前为止输入的原始数据列压缩为局部压缩数据或全局压缩数据,并生成压缩块来存储于存储器140。在这里,优选地,压缩块的大小小于基准块。例如,在基准块为10M字节的大小的情况下,可将压缩块的大小设定为8K字节。

[0046] 并且,在生成一个压缩块时,压缩块生成部136将最终更新的局部符号表存储于存储部140。在这里,优选地,局部符号表存储于相关压缩块内。另一方面,优选地,全局符号表存储于与在存储部140中存储压缩块的空间相独立的其他空间。

[0047] 以下,如下对本发明的实施例的数据压缩方法进行说明。

[0048] 图5为用于说明本发明实施例的数据压缩方法的流程图。

[0049] 参照图5,加载数据表的原始数据列并依次临时存储于全局缓冲区112以及局部缓冲区122(步骤S1)。之后,判断部136判断是否存在全局符号表(步骤S2)。判断结果,在不存在全局符号表的情况下,判断部136向全局符号表生成部110请求生成全局符号表。这样,全局数据模式搜索部114抽取重复的原始数据列,直到输入原始数据列达到既定的基准块的大小为止,由此选定全局压缩对象。而且,全局符号分配部116向选定为全局压缩对象的原始数据列分配全局符号来生成全局符号表(步骤S3)。

[0050] 像这样,在生成全局符号表的过程中,判断部136继续判断是否完成全局符号表(步骤S4)。判断结果,在完成全局符号表的情况下,判断部136向局部符号表生成部120传递表示已完成全局符号表的生成的信息。这样,局部数据模式搜索部124则排除存在于全局符号表的原始数据列,搜索在每当输入原始数据列时重复的数据列来选定局部压缩对象。之后,局部符号分配部126向选定为局部压缩对象的原始数据列分配局部符号来生成局部符号表(步骤S5)。

[0051] 相反,在步骤S4中的判断结果,在未完成全局符号表的情况下,局部数据模式搜索部124搜索每当输入原始数据列时重复的数据列来选定局部压缩对象。而且,局部符号分配部126向选定为局部压缩对象的原始数据列分配局部符号来生成局部符号表(步骤S6)。

[0052] 之后,压缩判断数据生成部134利用全局符号表或局部符号表来生成压缩判断数据(步骤S7)。之后,压缩块生成部136根据压缩判断数据来判断将要压缩的原始数据列的整体预期数据的大小是否大于既定的压缩块的大小(步骤S8)。判断结果,在整体预期数据的大小小于既定的压缩块的大小的情况下,继续接收下一个顺序的原始数据列。

[0053] 相反,在整体预期数据的大小大于既定的压缩块的大小的情况下,将与当前为止输入的原始数据列相对应的局部压缩数据或全局压缩数据生成为一个压缩块。并且,将所

生成的压缩块存储于存储部140(步骤S9)。此时,当前为止更新的局部符号表存储于相关压缩块内。

[0054] 以上,根据本发明的具体结构要素等的特定事项和限定的实施例及附图来进行说明,但是这仅仅是为了有助于本发明的更为全面性的理解而提供的,本发明并不限定于上述实施例,本发明所属技术领域的普通技术人员能够通过这种记载进行多种修改及变形。

[0055] 因此,本发明的思想不应局限于上述所说明的实施例而定,不仅是权利要求书,而且与该权利要求书等同或等价变形的所有技术方案均视为属于本发明的思想范畴。

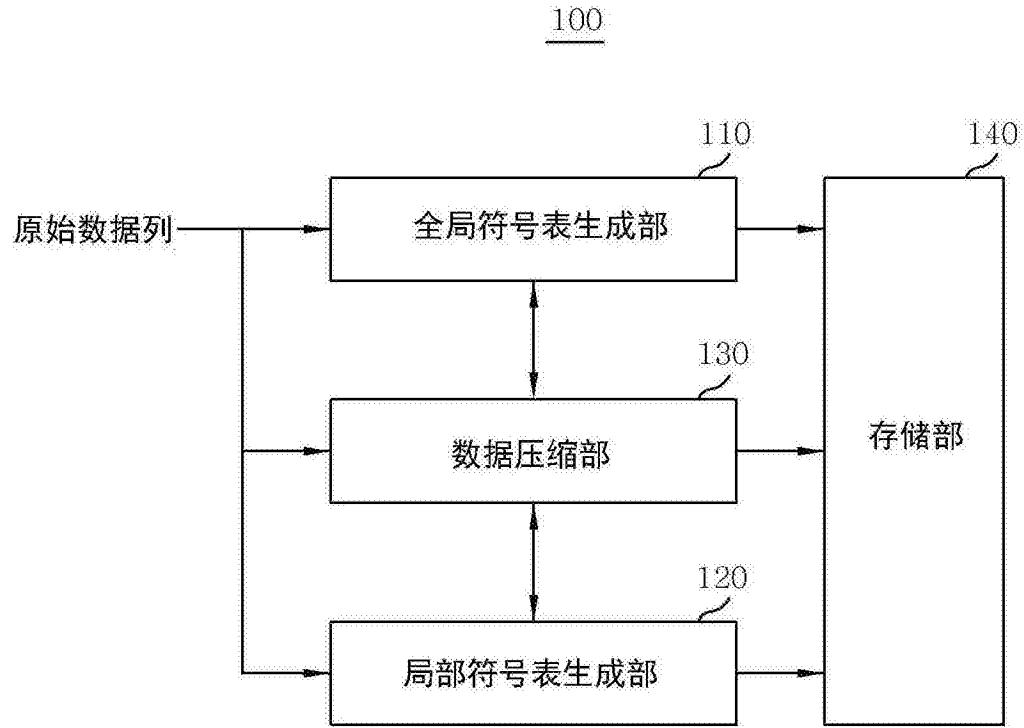


图1

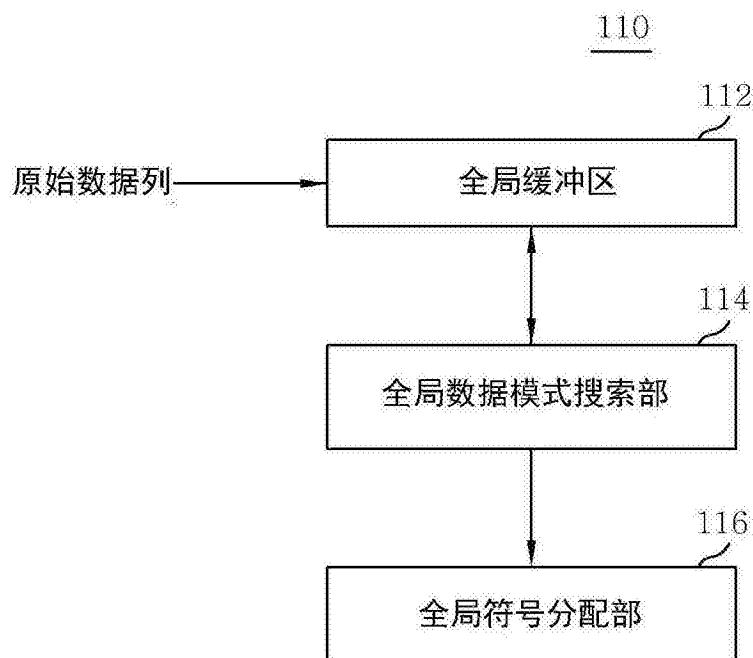


图2

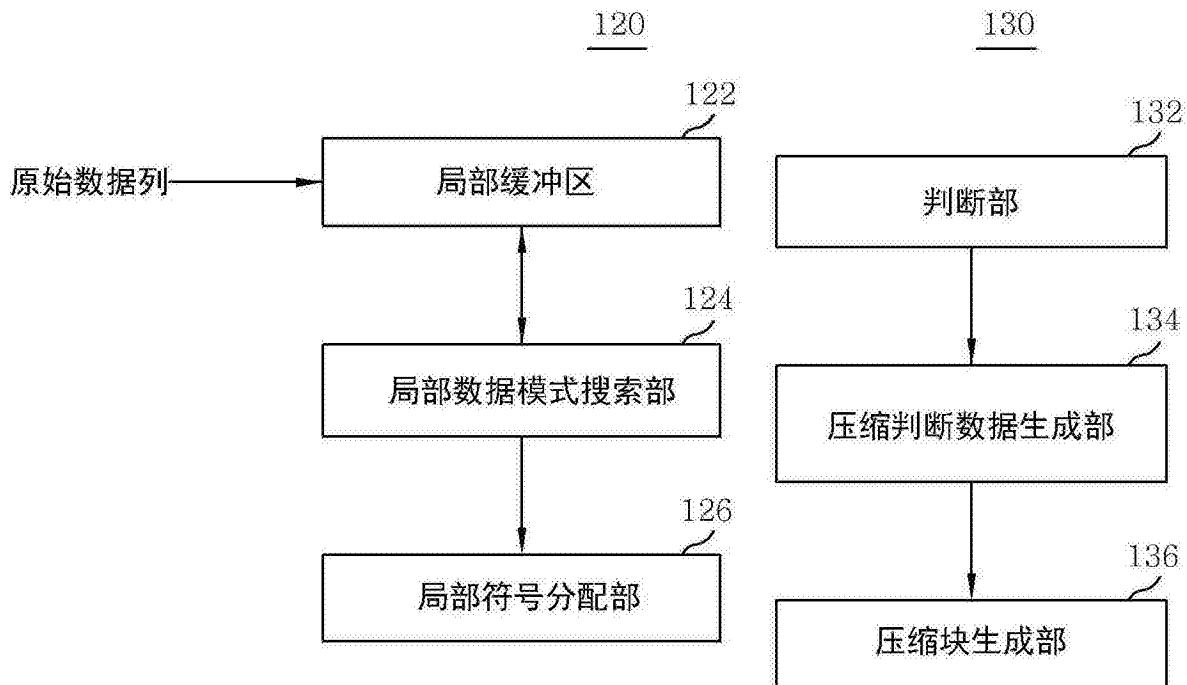


图3

图4

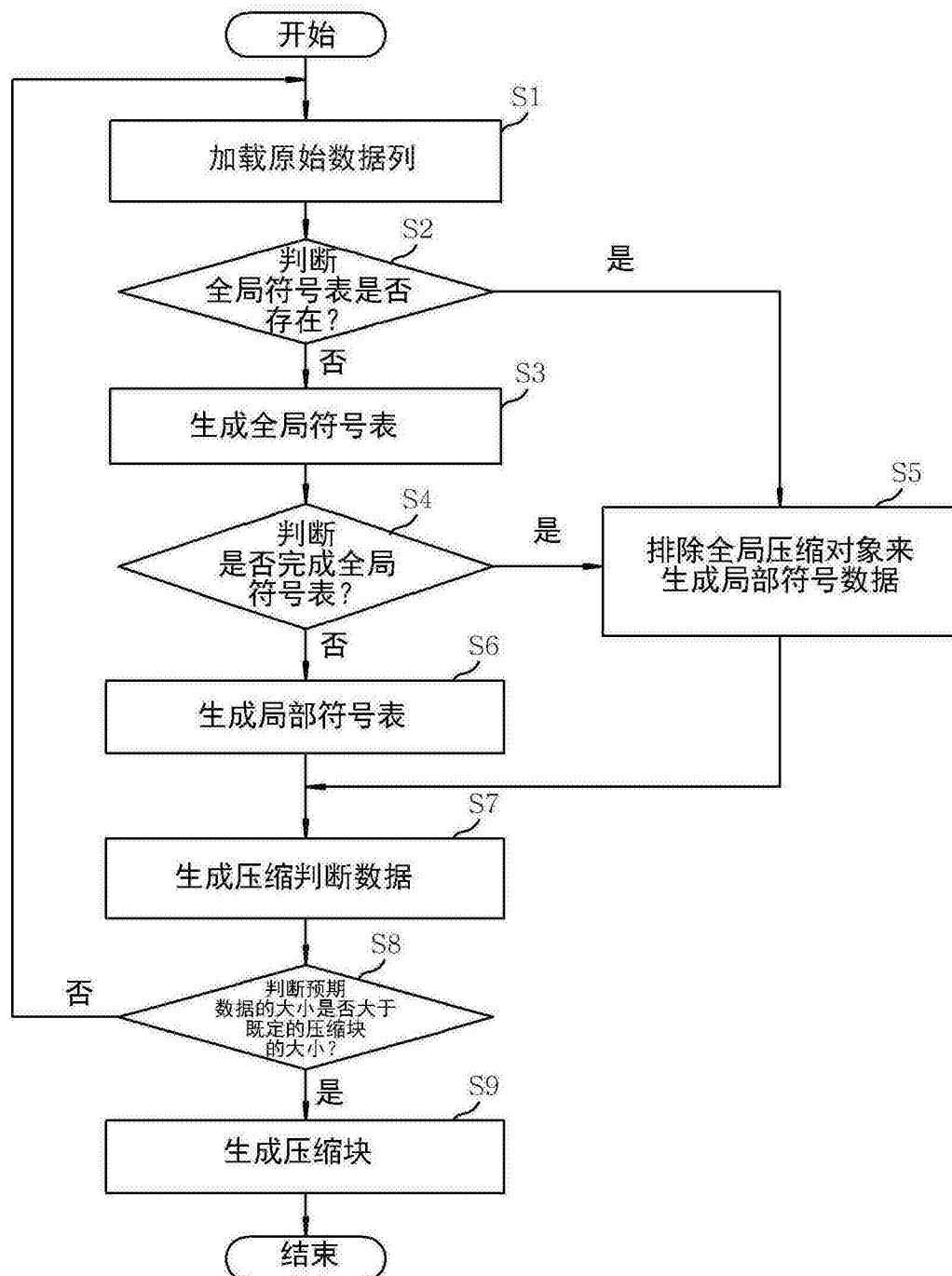


图5