



(12) 发明专利

(10) 授权公告号 CN 101742344 B

(45) 授权公告日 2011. 10. 19

(21) 申请号 200910204667. 2

(22) 申请日 2009. 10. 10

(30) 优先权数据

12/267, 719 2008. 11. 10 US

(73) 专利权人 香港城市大学

地址 中国香港九龙达之路 83 号

(72) 发明人 曾伟明

(74) 专利代理机构 中科专利商标代理有限责任

公司 11021

代理人 王波波

(51) Int. Cl.

H04N 13/00(2006. 01)

(56) 对比文件

US 2002/0176499 A1, 2002. 11. 28, 全文 .

US 2005/0041736 A1, 2005. 02. 24, 全文 .

US 2007/0109401 A1, 2007. 05. 17, 全文 .

CN 1981522 A, 2007. 06. 13, 全文 .

审查员 刘昶

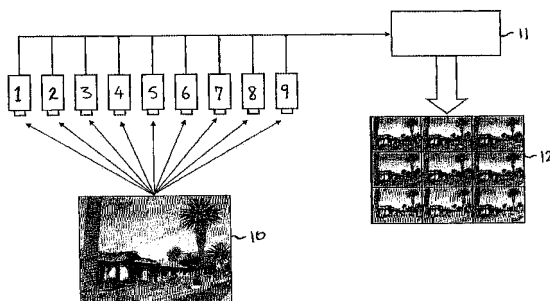
权利要求书 1 页 说明书 5 页 附图 3 页

(54) 发明名称

用于将多个视频信号编码成单个视频信号的方法

(57) 摘要

一种用于将多个视频信号编码成单个视频信号的方法, 该方法包括从相应的多个视频信号中提取多个背景图像; 将所述多个视频信号编码成所述单个视频信号, 所述单个视频信号包含用于重建所述多个视频信号的信息; 以及用所述背景图像来代替所述单个视频信号的帧。



1. 一种用于将多个视频信号编码成单个视频信号的方法,该方法包括:  
从相应的 N 个视频信号中提取 N 个背景图像;  
以 N 区块格式,将所述 N 个视频信号编码成所述单个视频信号,所述单个视频信号包含用于重建所述 N 个视频信号的信息;以及  
用所述背景图像来代替所述单个视频信号的帧。
2. 根据权利要求 1 所述的方法,其中,所述 N 个视频信号表示一个场景的不同视图或部分。
3. 根据权利要求 1 所述的方法,其中,从所述相应的 N 个视频信号中提取所述 N 个背景图像包括将所述 N 个视频信号分成所述 N 个背景图像和相应的 N 个前景图像序列。
4. 根据权利要求 3 所述的方法,其中,所述背景图像包括所述 N 个视频信号的低频成分。
5. 根据权利要求 3 所述的方法,其中,所述前景图像包括所述 N 个视频信号的高频成分。
6. 根据权利要求 3 所述的方法,其中,将所述 N 个视频信号编码成所述单个视频信号包括为所述 N 个视频信号的每个相应的帧压缩所述 N 个视频信号的相应的前景图像并将被压缩的帧图像平铺成单帧图像。
7. 根据权利要求 1 所述的方法,其中,将所述 N 个视频信号编码成所述单个视频信号包括为所述 N 个视频信号的每个相应的帧压缩所述 N 个视频信号的相应的帧图像并将被压缩的帧图像平铺成单帧图像。
8. 根据权利要求 1 所述的方法,其中,用所述背景图像代替所述单个视频信号的帧包括将所述背景图像与被编码的视频信号交织。
9. 根据权利要求 1 所述的方法,其中,所述背景图像是未压缩的。
10. 根据权利要求 1 所述的方法,该方法还包括在提取所述背景图像之前,将所述 N 个视频信号中的每一个分解成一系列的短的不相重叠的视频子序列,并对相应的短的不相重叠的视频子序列的组执行权利要求 1 的步骤。
11. 根据权利要求 10 所述的方法,该方法还包括:在分解之后但在提取之前,确定所述相应的短的不相重叠的视频子序列的组是否包含所述背景图像,如果所述视频子序列包含所述背景图像,则对所述相应的短的不相重叠的视频子序列的组执行权利要求 1 的步骤。
12. 根据权利要求 11 所述的方法,其中,如果所述视频子序列不包含所述背景图像,则略过提取和代替步骤,仅对所述相应的短的不相重叠的视频子序列的组执行编码步骤。
13. 根据权利要求 10 所述的方法,该方法包括:对一系列的相应的短的不相重叠的视频子序列的组分别执行权利要求 1 的步骤,以获得单个系列的短的不相重叠的视频子序列。
14. 根据权利要求 13 所述的方法,该方法还包括:结合所述单个系列的短的不相重叠的视频子序列以获得所述单个视频信号。
15. 根据权利要求 14 所述的方法,该方法还包括:在结合步骤之前,在所述单个系列的短的不相重叠的视频子序列中的每个视频子序列的开端和末尾插入标记。

## 用于将多个视频信号编码成单个视频信号的方法

### 技术领域

[0001] 本发明涉及用于将多个视频信号编码成单个视频信号的方法,以及更具体地用于将其中的每个视频信号表示同一场景的不同视图的多个视频信号编码成单个视频信号的方法。

### 背景技术

[0002] 三维(3D)自动立体显示设备允许在不需戴偏振或彩色眼镜的情况下来观察多视图 3D 图像。自动立体显示器的示例可以在 PCT 专利公布 WO 2007/059054 以及美国专利公布 2002/0011969、2002/0036825 和 2007/0182738 中找到,它们的全部内容作为参考结合于此。

[0003] 为了将多视图 3D 视频信号集成为现有的视频链,提出了 N 区块(N-tile)格式,并用于上述自动立体显示系统中。在 N 区块格式中,视频的每个帧包括 N 个相同分辨率的图片,每个图片来自每个信道。为了与现有的 2D 视频链兼容,N 个图片中的每一个通过向下采样(downsample)被压缩,以使得所有的 N 个图片可以适合等价的 2D 帧。在图 1 中示出了 N 区块格式。多个(在示出的 9 个的情况下)视频信号通过 9 个视频摄像机 1-9 生成,每个视频摄像机对准同一场景的不同视图。9 个视频信号或信道中的每一个通过 N 区块处理器 11 进行向下采样,并通过均匀地将视频信号分布在 3×3 区块结构中来集成为单个图片帧 12。在示出的具有 9 个信道的实施方式中,每个信号被沿着水平和垂直方向向下采样 3 次,从而使它的面积减小其初始值的 1/9。用这种方法,视频序列的多个信道将被编码为普通的二维(2D)视频序列,使得多视图 3D 图像以与普通的二维视频信号相同的方式来被处理并适合现有的视频链。在接收端,聚集的信号将被解码回多个视频序列,并通过称为“插值”(interpolation)的过程来显示在自动立体显示器上。然而,每个视图的分辨率会显著地降低并且不符合当前的 HDTV 趋势。

[0004] 因为在每个信道的向下采样,图片质量在分辨率方面严重地降低。这通常在图片显示在自动立体显示器上时导致了图片的模糊不清,并且假象在静止或缓慢地变化的场景中变得更明显,在静止或缓慢地变化的场景中观察者有足够的时间来研究细节。虽然插值技术可以被应用来获得更好的视觉质量,但计算负担会严重增加并且在向下采样过程中丢失的细节信息没有被恢复。

### 发明内容

[0005] 本发明的目标是提供一种用于将多个视频信号编码成单个视频信号的方法,以及特别地多个视频信号中的每一个表示同一场景的不同视图或部分。这允许聚集的信号与现有的视频链和设备兼容。

[0006] 简单的说,本发明将高清晰度元素恢复到用 N 区块格式的透视图。视频信号的多个信道被编码成单个视频序列,可以用与普通 2D 视频信号相同的方式来被分配、压缩和记录,其中,每个视频信号对应于场景的唯一视图或部分。本发明不同于 N 区块格式,不同之

处在于保留了在有限时间周期的静止或缓慢变化的内容的全分辨率。包含更严重的运动成分的剩余的内容用与 N 区块格式相同的分辨率来表示。理由是静止或缓慢变化的场景允许观察者有很多时间来仔细检查每个微小的细节,并且应该以较高的清晰度表示,而在场景的快速移动区域中的假象在人的感知中一般会更容易被宽恕。实验结果展现了用本发明的格式表示的多视图图像与用 N 区块格式表示的多视图图像相比的视觉质量的显著改善。

[0007] 本发明的另外的方面根据以下描述将变得更明显。

#### 附图说明

[0008] 现在将仅通过示例的方式并参考附图描述本发明的典型形式,其中:

[0009] 图 1 是现有技术的 N 区块编码格式的示意图;

[0010] 图 2 是多个视频信号的示意图,每个视频信号表示同一场景的不同视图或部分,每个视频信号在图中用多个连续的帧表示;

[0011] 图 3 是单个视频信号中的多个连续的视频帧的示意图,使用本发明的第一实施方式来编码单个视频信号;

[0012] 图 4 是单个视频信号中的多个连续的视频帧的第二示意图,使用本发明的第一实施方式来编码单个视频信号;

[0013] 图 5 是单个视频信号中的多个连续的视频帧的示意图,使用本发明的第二优选实施方式来编码单个视频信号;以及

[0014] 图 6 是单个视频信号中的多个连续的视频帧的示意图,使用本发明的第三实施方式来编码单个视频信号。

#### 具体实施方式

[0015] 图 1 示意性地示出了多个视频信号 21、22、23、24...29,其中的每一个可以使用图 1 中表示的 9 个摄像机 1-9 拍摄。当然,现在有 9 个这种视频信号,但为了说明性的目的,在图 2 中仅显示开始的 4 个信号 21、22、23、24 和最后的信号 29。为了说明本发明,每个视频信号 21-24...29 用多个连续的视频帧表示。为了清楚,仅有第一视频序列 21 的开始的四个帧 211、212、213、214 和第 9 个帧 219 以参考数字被提供。

[0016] 在本发明的方法中,多个视频信号 21-29 中的每一个被分解成 x 视频帧的多个相应的短的不相重叠的视频子序列。在每个子序列中的帧 x 的数量对于本发明来说不是关键的,并且可以仅包括几秒的视频。来自多视图视频信号的每一个的相应的子序列被称为子序列组 (GoS)。为了说明性的目的,每个 GoS 100、200、300 用虚线 101、201、301 分隔。GoS 包括多个相应的子序列,每个子序列包括同一场景的不同视频信号 21-29。在图 2 中示出的每个 GoS 100、200、300 中的每个子序列中有 10 个帧,但这对于本发明来说不是关键的。使用两个集成方法中的一个将在每个 GoS 100、200、300 中的多个相应的子序列集成为单个视频子序列,其中基于在每个 GoS 中的场景的特性来动态选择集成方法。然后,产生的表示每个 GoS 子序列的多个连续的单个视频序列被再结合成包含原始的多视图视频信号的信息的完整的单个视频序列。

[0017] 将在每个 GoS 中的多个子序列集成为单个子序列的方法基于每个 GoS 表示的场景的类型。如果由 GoS 表示的场景具有显著的背景成分,则使用第一方法集成 GoS 的多个

子序列,以下被称为类型-1GoS。如果在场景中没有显著的背景成分,则使用第二方法集成GoS的多个子序列,以下被称为类型-2GoS。场景的背景包括静止的或缓慢变化的内容。每个子序列的背景是它的低频成分,所述背景通过对多个子序列中的每一个应用边缘保留低通滤波器(EPF)来获得。可以通过从原始的输入图像中去除背景来获得前景、或高频成分、或每个子序列。EPF的阈值用于触发GoS为类型-1或类型-2。

[0018] 在类型-1GoS的集成中,每一多个子序列的背景图像是使用EPF以全分辨率第一次提取的。多个子序列的每个帧中的背景图像是相同的,从而如果在GoS上有N个子序列,则获得N个背景图像。在示出的实施方式中,在每个GoS 100、200、300中有9个子序列,所以将会有9个背景图像。接下来,在多个子序列中的每一个的每个帧例图中,通过从帧中减去相应的背景图像来获得前景图像。然后,使用现有技术的N区块格式将在每个帧例图中的N个前景图像集成为单个图像,产生N区块格式前景帧的单个序列。N个背景图像与N区块前景帧交织以形成输出子序列。为了维持输入GoS和输出GoS之间的帧的相同数量,N个前景帧被N个背景图像代替。要被代替的前景帧可以由任何逻辑上可重复的序列来确定,例如用背景图像来代替前景图像的开始的N个奇数或偶数帧。因此,例如,GoS 100包括9个子序列(即,N=9)并且子序列长度为100帧,开始的9个奇数帧(即,帧1、3、5、7、9、11、13、15、17、19)被9个背景图像代替。这在图3中示出,其中第一、第二和第九个背景图像被分别标注为301、302和309,并且N区块编码的前景帧被标注为310。

[0019] 对于不具有显著的背景成分的类型-2GoS,在每个帧例图中的N个图像被编码成现有技术中公知的单个N区块格式。输出是N区块图片的序列。

[0020] 在最终的被集成的视频信号中,每个GoS子序列的开端和末尾由在子序列的第一个和最后一个帧的标记401、402指示。在优选实施方式中,通过消隐在每个子序列的每个帧中的第一个和最后一个视频行而便于使用适当的标记方法。然后,标记被插入到被消隐的视频行中。为了指示在最终的被集成的视频中的类型-1GoS子序列,在子序列中的第一个和最后一个帧中的第一个视频行被设置为白色峰值。为了指示在最终的被集成的视频中的类型-2GoS子序列,在子序列中的第一个和最后一个帧中的最后一个视频行被设置为白色峰值。

[0021] 图4示意性地示出了包括来自多个初始视频信号的信息的最终的单个视频序列,所述多个初始视频信号表示同一场景的不同视图或部分。图4的信号由单独的视频帧表示,其中,标记401和402表示被编码的GoS的开端和末尾帧。全分辨率的背景帧301、302、303...309与在一些序列的开端的N区块编码的帧310交织。

[0022] 在接收端,每个GoS被解码以重建多个信道视频子序列。类型-2GoS以与公知的N区块序列相同的方式被解码。每个区块被插值到全分辨率,并数字化为在自动立体显示器上的N个视图中的一个。对于类型-1GoS,解码过程包括三个阶段。第一,从子序列中提取N个背景图片。第二,N区块前景序列被解码成多个信道前景视频子序列。在编码过程中由背景图片代替的前景图像将在被解码的GoS中丢失。丢失的图片中的每一个将被事先解码的以前的邻近的前景图片取代。最后,每个信道前景图片被覆盖到其相应的背景图像上以重建原始图片。

[0023] 所提出的方法的优点是多个信道视频序列的静止或缓慢变化的内容将以全分辨率呈现在自动立体显示器上。快速变化的内容以与用N区块格式编码的内容的分辨率相同

的分辨率来被呈现。

[0024] 本发明还通过在用于将多个视频信号编码成单个视频信号的方法中的步骤的示例来示出,每个视频信号包括同一场景的不同视角:

[0025] (1) 首先,多个视频信号或信道被分割成多个连续的并且不相重叠的较短的子序列组 (GoS),多个视频信号中的每一个包括同一场景的不同视角。

[0026] (2) 第二,每个 GoS 被分类为类型 -1 或类型 -2GoS。

[0027] 根据步骤 (3) 到 (7) 来处理类型 -1GoS。

[0028] (3) 使用例如边缘保留低通滤波器的任何合适的背景提取方法从每一个类型 -1GoS 内的多个视频信号子序列中的每一个中提取背景图像。每个背景图像表示对在各自的视频信号子序列中的所有帧共用的静止内容。

[0029] (4) 前景图像对于在多个视频信号子序列中的每一个的每个帧而被提取。前景图像可以通过技术领域中公知的任何合适的方法找到,例如,通过从每个原始帧图像中减去各自的背景图像。

[0030] (5) 在每个帧例图中,用 N 区块格式将多个前景图像 (来自每个信道) 编码成单个子序列。

[0031] (6) 通过用背景图像来代替已知的前景帧,全分辨率的背景图像与 N 区块前景子序列交织。

[0032] (7) 在 GoS 序列的每个帧中的第一个和最后一个视频行被消隐,并且在第一个和最后一个帧中的第一个视频行被设置为白色峰值以识别类型 -1GoS。

[0033] 根据步骤 (8) 和 (9) 来处理类型 -2GoS。

[0034] (8) 使用 N 区块方法来将多个视频信号或信道编码成单个子序列。

[0035] (9) 在子序列中的每个帧中的第一个和最后一个视频行被消隐,并且在第一个和最后一个帧中的最后一个视频行被设置为白色峰值以识别类型 -2GoS。

[0036] (10) 被编码的类型 -1 和类型 -2GoS 子序列被链接以提供场景的单个视频序列,所述场景包含来自原始的多个视频信号的信息,所述视频信号包括同一场景的不同视角。

[0037] 在接收器或自动立体显示设备,根据步骤 (11) 和 (17) 将单个视频信号解码为多个视频信号以用于显示场景的 3D 图像。

[0038] (11) 通过识别在视频帧的第一行和最后一行中的白色峰值标记来将单个视频序列分成多个子序列,该子序列包含类型 -1GoS 和类型 -2GoS。

[0039] (12) 使用 N 区块方法将类型 -2GoS 解码成多个相应的子序列。通过插值将帧向上采样 (up-sample) 为原始大小。

[0040] 根据步骤 (13) 到 (17) 来解码类型 -1GoS。

[0041] (13) 全分辨率的背景图像从它们在序列中的帧位置被提取。

[0042] (14) 用从相邻的帧的插值代替丢失的帧,所述丢失的帧即被删掉以为背景图像让路的帧。

[0043] (15) 使用 N 区块方法将表示视频子序列的前景图像的帧的子序列解码成多个相应的子序列。每个帧通过插值被向上采样到其原始大小。

[0044] (16) 在多个相应的子序列中的每一个中的每个帧处,通过将各自的全分辨率的背景图像加入前景图像中来重建全帧图像。

[0045] (17) 来自每个类型 -1GoS 和类型 -2GoS 的各自的子序列被链接,以重建包括同一场景的不同视角的原始多个视频信号。

[0046] 应当理解,对于本领域技术人员来说显而易见的修改和变换不被理解为超出本发明的范围。例如,转向图 5 和图 6,示出了将背景图像交织到视频序列的 N 区块编码帧的不同方法。在图 5 中,背景图像 501、502、503...509 被置于 N 区块编码帧 510 的前面的序列中。图 6 表示另一个实施方式,其中,背景图像 601、602、603、604、605、606、607、608 也用 N 区块格式被编码,但以比前景图像更高的分辨率。用已知的 N 区块格式,表示同一场景的不同视图的多个视频信号中的每一个的相应的帧被平铺成输出视频序列的单个帧。根据本发明,多个背景图像可以被交织在单个输出视频内的不同位置。结果是输出信号的每个帧不需要包含来自多个原始输入信号中的每一个的数据。在图 5 的实施方式中,背景图像 601、602、603、604、605、606、607、608 通过占据之后与 N 区块前景图像交织的多个帧而以较高的分辨率用 N 区块格式被编码。

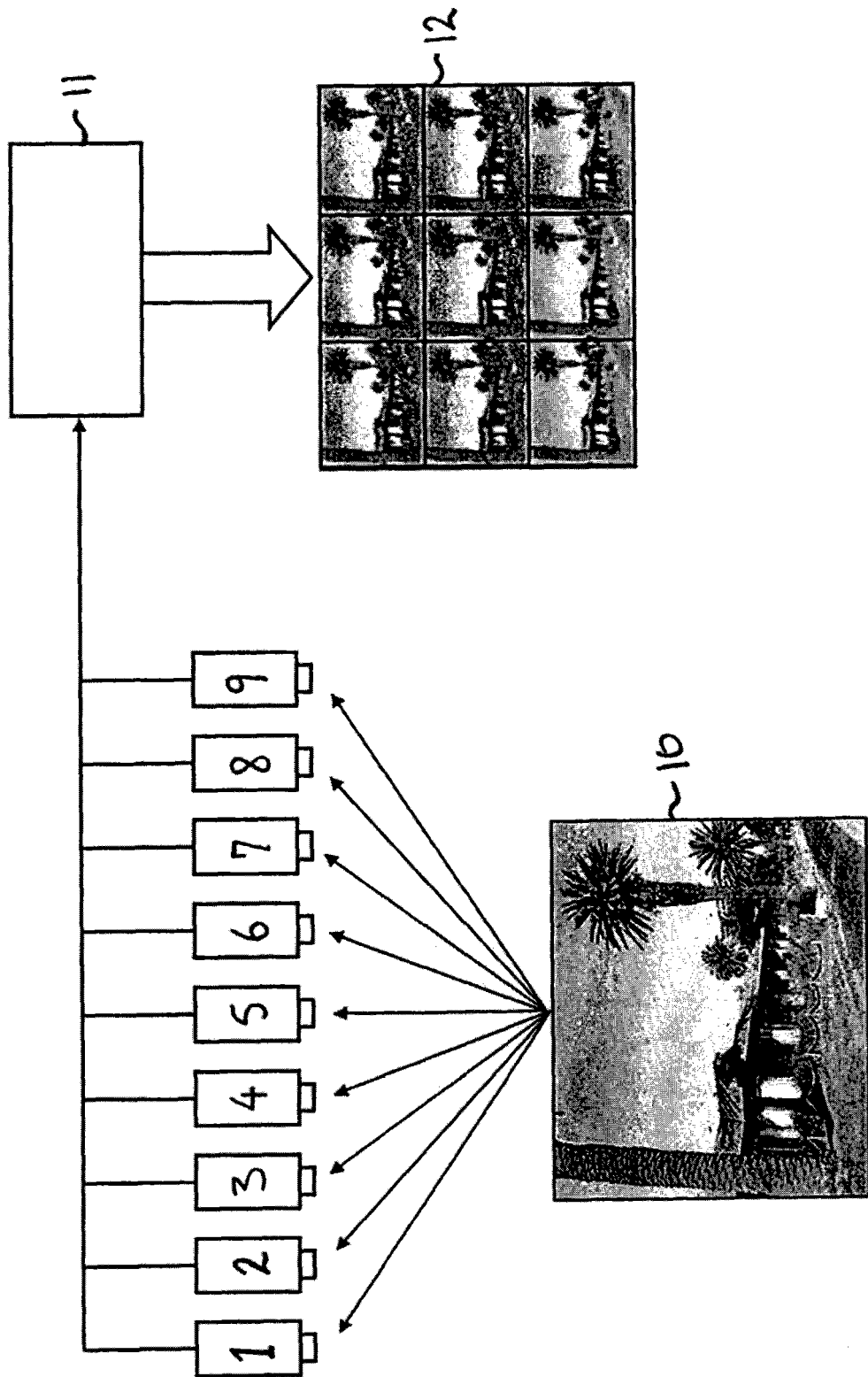


图 1



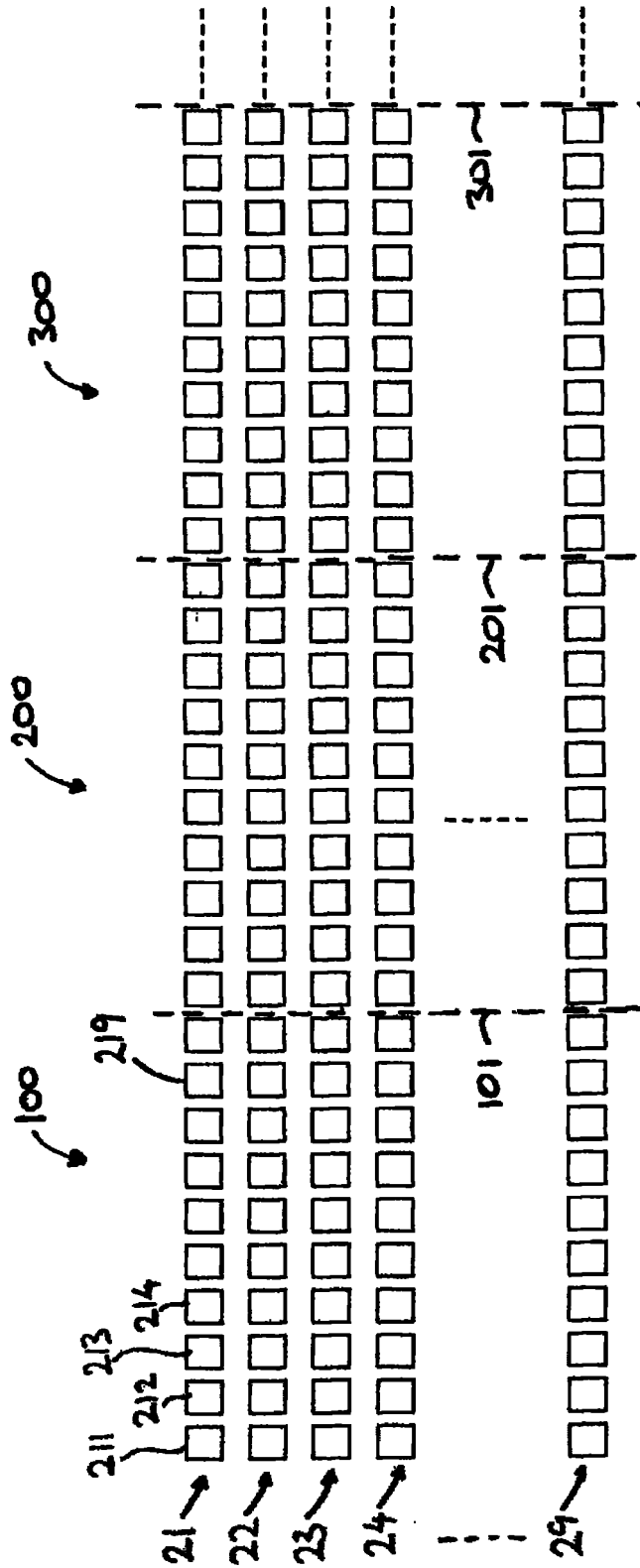


图 2

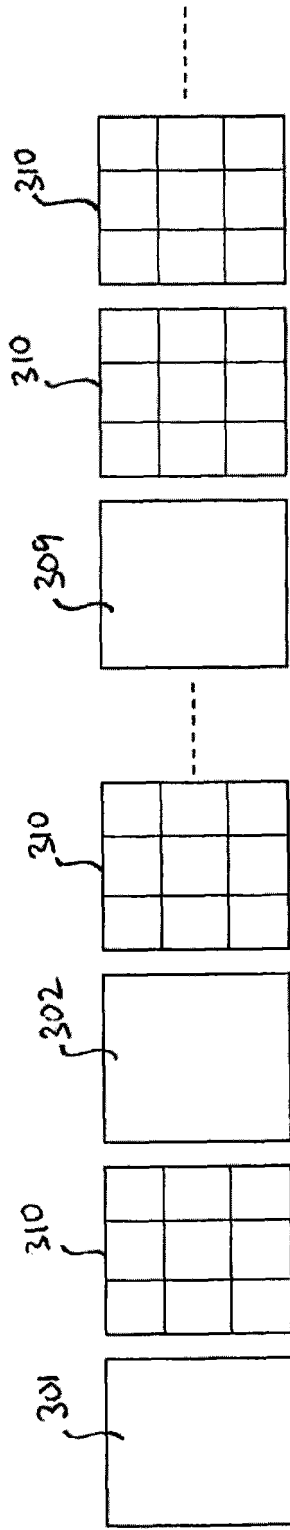


图 3

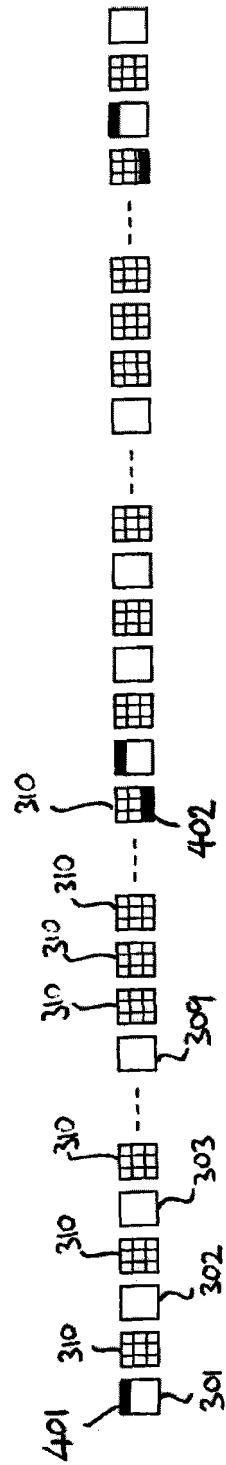


图 4

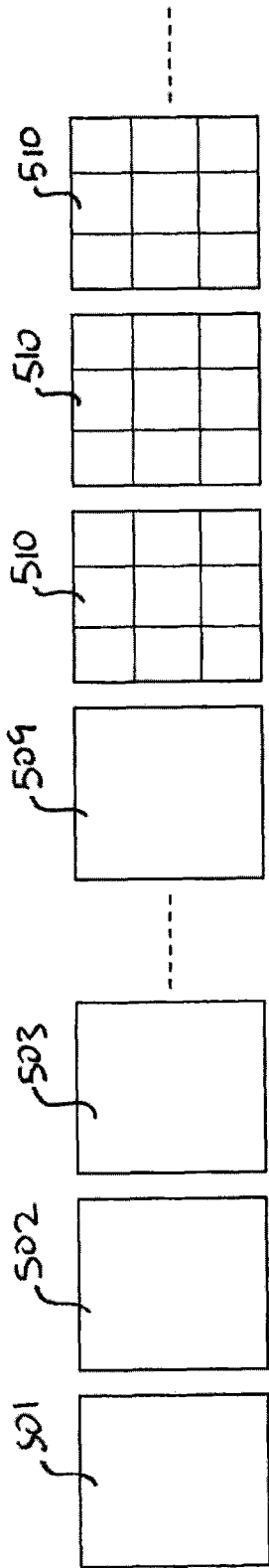


图 5

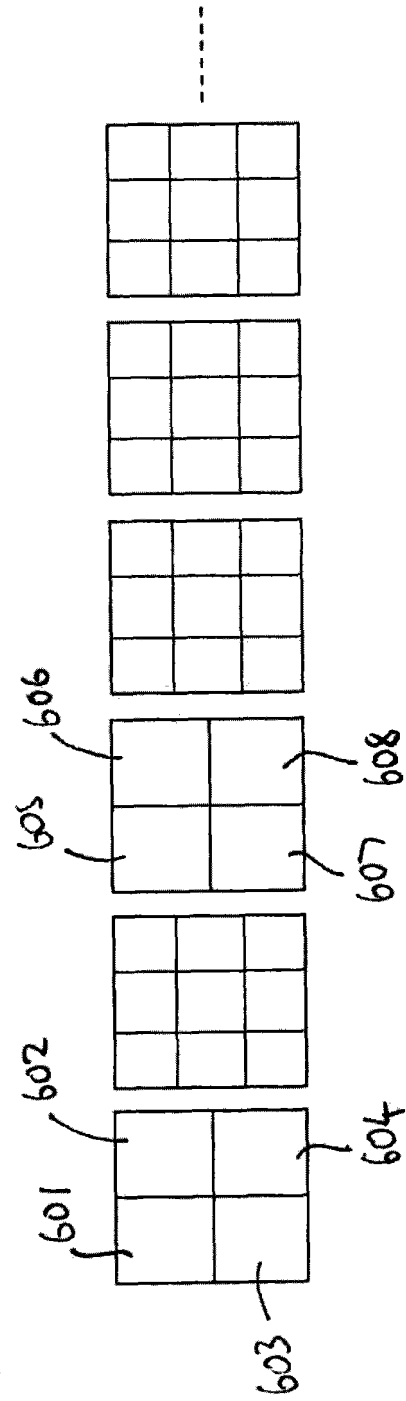


图 6