

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04N 7/26 (2006.01)



[12] 发明专利说明书

专利号 ZL 03141275.0

[45] 授权公告日 2009 年 1 月 21 日

[11] 授权公告号 CN 100455018C

[22] 申请日 2003.6.3 [21] 申请号 03141275.0

[30] 优先权

[32] 2002.6.3 [33] US [31] 60/385,965

[73] 专利权人 微软公司

地址 美国华盛顿州

[72] 发明人 A·图雷皮斯 S·李 F·吴

[56] 参考文献

US - 6005980A 1999.12.21

EP - 0782343A2 1997.7.2

US - 5132792A 1992.7.21

EP - 0944245A1 1999.9.22

Fast Motion Vector Estimation Using Multiresolution - Spatio - Temporal Correlations. Junavit Chalidabhongse ET AL. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7 No. 3. 1997

审查员 吴黄飞

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 陈斌

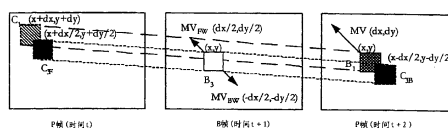
权利要求书 3 页 说明书 16 页 附图 11 页

[54] 发明名称

双向预测 (B) 图像的时间空间预测和多图像参照移动补偿的移动向量预测

[57] 摘要

提供了在一个视频序列中使用双向预测 (B) 图像的几个改进方案。在某些改进方案中通过使用空间预测技巧来增强直接模式编码和/或移动向量预测。在其他改进方案中移动向量预测包括时间距离和子块信息,例如,为了得到更准确的预测。这样的改进和其他这里展示的改进显著地提高了任何适用的视频编码系统/逻辑的性能。



1. 一种用于在视频编码器中编码视频数据的方法，所述方法包括：

对 B 图像中的至少一个直接模式大块作出空间/时间移动向量预测判决，其中对所述至少一个直接模式大块的空间/时间移动向量预测判决为所述至少一个直接模式大块指示出空间移动向量预测；

对所述 B 图像中的所述至少一个直接模式大块进行编码，其中所述编码对于所述至少一个直接模式大块中的一给定的直接模式大块包括：

从用于周围部分的参考图像中为所述给定的直接模式大块选择参考图像；

对于所述给定的直接模式大块执行空间移动向量预测；

用信号表示对所述至少一个直接模式大块的空间/时间移动向量预测判决信息，其中，所述用信号表示空间/时间移动向量预测判决信息向视频解码器通知对所述 B 图像中的所述至少一个直接模式大块的空间/时间移动向量预测判决。

2. 如权利要求 1 所述的方法，其特征在于，所述至少一个直接模式大块包括多个直接模式大块。

3. 如权利要求 2 所述的方法，其特征在于，所述多个直接模式大块在所述 B 图像的一个片段中。

4. 如权利要求 2 所述的方法，其特征在于，所述多个直接模式大块是多个 16×16 的大块。

5. 如权利要求 4 所述的方法，其特征在于，每个 16×16 的大块包括 4 个 8×8 的子块。

6. 如权利要求 1 所述的方法，其特征在于，所述为所述给定的直接模式大块选择参考图像的动作包括从用于所述周围部分的至少两个参考图像之中为所述给定的直接模式大块选择最小参考图像。

7. 如权利要求 1 所述的方法, 其特征在于, 所述空间移动向量预测包括中值移动向量预测。

8. 如权利要求 1 所述的方法, 其特征在于, 所述周围部分包括周围大块。

9. 一种用于在视频解码器中解码视频数据的方法, 该方法包括:

接收对 B 图像中的至少一个直接模式大块的用信号表示的空间/时间移动向量预测判决信息;

从所述用信号表示的空间/时间移动向量预测判决信息中确定对所述 B 图像中的所述至少一个直接模式大块的空间/时间移动向量预测判决, 其中对所述至少一个直接模式大块的空间/时间移动向量预测判决为所述至少一个直接模式大块指示出空间移动向量预测;

对所述至少一个直接模式大块进行解码, 其中所述解码对于所述至少一个直接模式大块中的一给定的直接模式大块包括:

从用于周围部分的参考图像中为所述给定的直接模式大块选择参考图像;

对于所述给定的直接模式大块执行空间移动向量预测。

10. 如权利要求 9 所述的方法, 其特征在于, 所述至少一个直接模式大块包括多个直接模式大块。

11. 如权利要求 10 所述的方法, 其特征在于, 所述多个直接模式大块在所述 B 图像的一个片段中。

12. 如权利要求 10 所述的方法, 其特征在于, 所述多个直接模式大块是多个 16×16 的大块。

13. 如权利要求 12 所述的方法, 其特征在于, 每个 16×16 的大块包括 4 个 8×8 的子块。

14. 如权利要求 9 所述的方法, 其特征在于, 所述为所述给定的直接模式大

块选择参考图像的动作包括从用于所述周围部分的至少两个参考图像之中为所述给定的直接模式大块选择最小参考图像。

15. 如权利要求 9 所述的方法，其特征在于，所述空间移动向量预测包括中值移动向量预测。

16. 如权利要求 9 所述的方法，其特征在于，所述周围部分包括周围大块。

双向预测(B)图像的时间空间预测和多图像参照移动补偿的移动向量预测

相关专利申请

这个美国专利许可非临时申请要求一起待批准的于 2002 年 6 月 3 日提交的, 标题为“双向预测(B)帧的时间空间预测和多帧参照移动补偿的移动向量预测”的美国专利许可临时申请 60/385,965 的优先利益, 并在这里通过引用包括其整个揭示的内容。

技术领域

本发明涉及视频编码, 并更特别地涉及提供改进的和不同类型的视频数据相关的编码和/或预测技巧的方法和装置。

背景技术

提高在视频编码中编码效率的动机导致了联合视频小组(JVT)(一个标准组织)的更加精细和复杂的对一个给定的大块说明移动信息的模型及模式的采用。这些模型和模式趋向于更好地利用可能存在于一个视频序列中间的时间冗余。参阅, 例如, ITU-T, 视频编码专家组(VCEG), “JVT 编码—(ITU-T H.261 & ISO/IEC JTC1 标准)—工作草案编号 2(WD-2)”, ITU-T JVT-B118, 2002 年 3 月; 和/或 Heiko Schwarz 和 Tomas Wiegand, “树结构的大块划分”, Doc. VCEG-N17, 2001 年 12 月。

对可以支持最新的模型和模式并还可能引入新模型和模式来利用改进的编码技巧的进一步改进的方法和装置有持续的需要。

发明内容

上述的和其他的需要, 例如通过一种在一系列的视频帧中编码视频数据的方法来处理。这个方法包括标识至少一个视频帧的至少一部分为一个双向预测(B)图像, 并选择性地至少使用空间预测来编码至少一个和 B 图像关联的移动参数来编码 B 图像。在某些示范性的应用中, B 图像可以包括一个块、一个大块、一个子块、一个片段或视频帧中其他类似的部分。例如, 当使用一个大块部分时, 这个方法产生一个直接大块。

在某些进一步的示范性的应用中，该方法进一步包括基于至少一个为视频帧的至少另一个部分的参照图像使用线性或非线性移动向量预测 B 图像。作为例子，在特定应用中，该方法使用中值移动向量来产生至少一个移动向量。

在其他示范性的应用中，除了空间预测，该方法也处理至少一个其他视频帧的至少一个其他部分来进一步选择性地使用时间预测编码至少一个和 B 图像相关的基于时间的移动参数来编码 B 图像。在一些实例中，时间预测包括双向时间预测，例如基于一个预测（P）帧的至少一个部分。

在某些其他应用中，该方法还选择性地确定适用的一个至少部分基于一个在预测视频帧和包括 B 图像的帧之间的时间距离的基于时间的移动参数的缩放比例。在某些应用中，时间距离信息被编码，例如，在一个头或其他类似的和已编码的 B 图像相关的数据布置中。

附图说明

本发明以举例方式并不仅限于附图中的图形来说明。同样的编号被使用在所有的图形中以参考类似的组件和/或特性。

图 1 为方框图，展示一个适合于使用本发明的某些应用的示范性的计算环境。

图 2 为方框图，展示一个适合于使用本发明的某些应用的示范性的有代表性的设备。

图 3 为示意图，展示根据本发明某些示范性的应用，和一个图片的各部分相关的空间预测。

图 4 为示意图，展示根据本发明某些示范性的应用，在 B 图像编码中的直接预测。

图 5 为示意图，展示根据本发明某些示范性的应用，当一个场景改变发生时或甚至当比较（collocated）块被内编码时会发生什么。

图 6 为示意图，展示根据本发明某些示范性的应用，所述移动假设为 0 时，在现有的编码解码器中处理比较的内块。

图 7 为示意图，展示根据本发明某些示范性的应用，当比较块的参照图像在后面的 P 图像中不为 0 时，直接模式如何被处理。

图 8 为示意图，展示根据本发明某些示范性的应用，一个示范性的方案其中 MV_{FW} 和 MV_{BW} 源于空间预测。

图 9 为示意图，展示根据本发明某些示范性的应用，空间预测如何解决场景改变及类似的问题。

图 10 为示意图，展示根据本发明某些示范性的应用，对在 B 图像编码中直接模式的联合时空预测。

图 11 为示意图，展示根据本发明某些示范性的应用，结合预测器(predictor)大块的参照图像信息对当前块的移动向量预测。

图 12 为示意图，展示根据本发明某些示范性的应用，如何使用多个候选者来进行直接模式预测，尤其是如果双向预测用在 B 图像中时。

图 13 为示意图，展示根据本发明某些示范性的应用，在使用未来和过去参照图像中如何限制 B 图像。

图 14 为示意图，展示根据本发明某些示范性的应用，为了时间直接预测，投射比较移动向量到一个当前参照。

图 15a-c 为示意图，展示根据本发明某些示范性的应用，对一个 MV 在不同配置中的移动向量预测器。

图 16a-c 为示意图，展示根据本发明某些示范性的应用，对一个有 8×8 分区的 MV 在不同配置中的移动向量预测器。

图 17a-c 为示意图，展示根据本发明某些示范性的应用，对一个有 8×8 分区附加预测器的 MV 在不同配置中的移动向量预测器。

具体实施方式

在一个视频序列中使用双向预测(B)图像的几个改进在下面并且表示在图中。在一些改进中，使用了空间预测技术以后，直接编码模式，或者说运动矢量预测得到了加强。在另一些改进中，运动矢量预测增加了时空和子信息，从而增强了预测能力，比如说更精确。因而，这些改进极大地增强了任何适用的视频编码系统/逻辑。

以上描述了一些方法和装置作为示范，这里需要注意的是当前发明的这些技术并不局限于描述到的例子和相关的图，它们适用于其他类似的现存和未来视频编码系统。

在介绍这些示范性的方法和装置之前，在接下来的部分先介绍一下适合的示范性的操作环境，例如，以计算机或是其他类型的装置或程序的形式。

示范性的操作环境

根据附图，其中类似的引用编号指向类似的元素，本发明以应用在一个适合的计算环境中来展示。虽然不是必须的，本发明将在计算机可执行指令，如由个人计算机执行的程序模块的通用的语境中说明。

通常，程序模块包括执行特定的任务或应用特定的抽象数据类型的子程序、程序、对象、组件、数据结构等等。熟悉相关技术的人应理解本发明可以应用在其他计算机系统配置中，包括手持设备、多处理器系统、基于微处理器的或可编程的消费者电子设备、网络 PC、小型机、大型机、移动通讯设备，以及类似的设备。

本发明还可以应用在任务由通过一个通信网络连接的远程处理设备来执行的分布式计算环境。在一个分布式计算环境中，程序模块可以同时位于本地和远程存储设备。

图 1 展示了一个下面所述系统、装置和方法在其上应用的适合的计算机环境 120 的例子。示范性计算环境 120 仅为适合的计算机环境的一个例子并不意味着任何对在此说明的改进的方法及系统的使用或功能范围的限制。不应解释计算环境 120 为对在计算环境 120 中展示的组件的任何一种或组合有任何的关联性或必需性。

这里的改进的方法和系统可以运行在大量其他的通用或专用计算系统环境或配置中。适合的众所周知的计算机系统、环境，和/或配置的例子包括，但不限于，个人计算机、服务器计算机、瘦客户、胖客户、手持或膝上设备、多处理器系统、基于多处理器的系统、机顶盒、可编程消费者电子设备、网络 PC、小型机、大型机、包含上述以及类似系统或设备的任意组合的分布式计算机环境。

如图 1 所示，计算环境 120 包括形式为一个计算机 130 的一个通用计算设备。计算机 130 的组件可以包括一个或多个处理器或处理单元 132、一个系统存储器 134，及一个连接包括系统存储器 134 到处理器 132 的不同的系统组件的总线 136。

总线 136 代表几种总线结构中的任何一种或多种，包括一个存储器总线或存储器控制器、一个外围总线、一个加速图形接口，及一个处理器或使用多种总线结构中的任意结构的本地总线。作为例子，而非限制，这样的结构包括工业标准结构（ISA）总线，微通道结构（MCA）总线、增强的工业标准结

构（EISA）总线、视频电子标准协会（VESA）本地总线，及也称为包厢总线的周边元件扩展接口（PCI）总线。

计算机 130 通常包括不同的计算机可读媒体。这样的媒体可以为能由计算机 130 访问的任何可用的媒体，并包括易失的和非易失的媒体、可移动的和不可移动的媒体。

在图 1 中，系统存储器 134 包括易失的存储器形式的计算机可读媒体，如随机访问存储器（RAM）140，和/或非易失的存储器，如只读存储器（ROM）138。一个包含帮助在计算机 130 中的元件之间传输信息如在启动过程中的基本子程序的基本输入/输出系统（BIOS）142，被存储在 ROM 138 中。RAM 140 通常包含可以由处理器 132 立即访问到和/或处理的数据和/或程序模块。

计算机 130 可以进一步包括其他可移动的/不可移动的、易失的/非易失的计算机存储器媒体。例如，图 1 展示了一个用来读写一个不可移动的、非易失的磁媒体（未标出并通常称为一个“硬盘驱动器”）的硬盘驱动器 144，一个用来读写一个可移动的、非易失的磁盘 148（例如，一个“软盘”）的磁盘驱动器 146，及一个用来读写一个可移动、非易失的光盘 152 如一个 CD-ROM/R/RW，DVD-ROM/R/RW/+R/RAM 或其他的光媒体的光盘驱动器 150。硬盘驱动器 144、磁盘驱动器 146 和光盘驱动器 150 每个都通过一个或多个接口 154 连接到总线 136。

驱动器和相关的计算机可读媒体为计算机 130 提供计算机可读指令、数据结构、程序模块、及其他数据的非易失存储。虽然这里所述的示范性的环境使用一个硬盘、一个可移动磁盘 148 及一个可移动光盘 152，熟悉相关技术的人应理解其他类型的可以存储可由一个计算机访问的数据的计算机可读媒体，如盒式磁带、闪存卡、数字视频盘、随机访问存储器（RAM）、只读存储器（ROM）、及类似的媒体也可以用在示范性的操作环境中。

多个程序模块可以存储在硬盘、磁盘 148、光盘 152、ROM 138 或 RAM 140，包括，例如，一个操作系统 158，一个或多个应用程序 160，其他程序模块 162，及程序数据 164。

这里所述的改进的方法和系统可以应用在操作系统 158、一个或多个应用程序 160，其他程序模块 162，和/或程序数据 164 中。

一个用户可以通过输入设备，如键盘 166 和定点设备 168（如一个“鼠标”）提供命令和信息到计算机 130 中。其他输入设备（未标出）可以包括一个麦

克风、操纵杆、游戏垫、圆盘式卫星电视天线、串行口、扫描仪、相机等等。这些和其他输入设备通过一个连接到总线 136 的用户输入接口 170 连接到处理单元 132，但也可以通过其他接口和总线结构，如并行口、游戏口，或一个通用串行总线（USB）来连接。

一个显示器 172 或其他显示设备也通过一个接口，如视频适配器 174 与系统总线连接。除显示器之外，个人计算机通常包括通过输出周边接口 175 连接的其他周边输出设备（未标出），如音箱和打印机。

计算机 130 可以使用到一个或多个远程计算机，如远程计算机 182 的逻辑连接在一个网络环境中运行。远程计算机 182 可以包括多个或所有上述相对于计算机 130 的组件，

图 1 中所示逻辑连接是一个局域网（LAN）177 和一个通用广域网（WAN）179。这样的网络环境在办公室访问或企业访问的计算机网络、企业内部互联网和因特网中是很常见的。

当用在一个局域网联网环境中，计算机 130 通过一个网络接口或适配器 196 与局域网 177 相连。当用在一个广域网联网环境中，计算机通常包括一个调制解调器 178 或其他在广域网 179 上建立通讯的方法。调制解调器 178，可以为内置的或外置的，通过用户输入接口 170 或其他适合的机制连接到系统总线 136。

如图 1 所示，为一个 WAN 通过因特网的具体的应用。在这里，计算机 130 使用调制解调器 178 通过因特网 180 和至少一个远程计算机 182 建立通讯。

在一个网络环境中，所述相对于计算机 130 的程序模块，或其部分，可以存储在远程存储器存储设备中。因此，例如，如图 1 所示，远程应用程序 189 可以驻留在远程计算机 182 的一个存储器设备中。应理解所示网络连接仅为示范性的，也可以使用其他在计算机之间建立通讯连接的方法。

下面请看图 2，一个方框图，展示另一个也能够从这里揭示的方法和装置中受益的示范性的设备 200。设备 200 代表任何一种或多种操作上根据这里所述及等价的的方法和装置的所有或部分配置为处理视频和/或任何相关类型的数据的设备或装置。因此设备 200 可以采取如图 1 种的计算设备的形式，或其他形式，例如，一个无线设备、一个可移动通讯设备、一个个人数字助理、一个视频播放器、一个电视、一个 DVD 播放器、一个 CD 播放器、一个卡拉 OK 机、一个机顶盒、一个视频游戏机等等。在这个例子中，设备 200 包括配

置为处理视频数据的逻辑电路 202，一个配置为提供视频数据给逻辑电路 202 的视频数据源 204，及至少能够显示视频数据的至少一部分给用户看的一个显示模块 206。逻辑电路 202 代表硬件、固件、软件和/或它们的任何组合。在某些应用中，例如，逻辑电路 202 包括一个压缩器/解压缩器（codec）或类似的部分。视频数据源 204 代表任何可以提供、传输、输出，和/或至少暂时地存储适合由逻辑电路 202 处理的视频数据的机制。视频再现源展示为在设备 200 之内和/或在设备 200 之外。显示模块 206 代表用户可以直接或间接地观看这里所述的视频数据的视觉结果的任何机制。另外，在某些应用中，设备 200 还可以包括一些用来再现或处理和视频数据关联的音频数据的形式或能力。因此一个音频再现模块 208 在图中展示出。

有了图 1 和图 2 及其他类似它们的例子，下面的部分集中于某些可以至少部分使用这样的环境和这样的设备来应用的示范性的方法和装置。

编码双向预测（B）图像和移动向量预测

这个部分说明一些可以应用来在一个视频编码类似的系统或编码双向预测（B）图像和移动向量预测的示范性的改进。示范性的方法和装置可以用于预测移动向量和 B 图像直接模式设计的增强。这样的方法和装置特别适合于多图像引用编码解码器，例如，JVT，并可以得到相当大的编码增益，特别是对全景序列或场景改变。

双向预测（B）图像是多数视频编码标准和系统一个重要部分，因为它们趋向于提高这样的系统的编码效率，例如当和仅使用预测（P）图像比较时。这个在编码效率上的改进主要通过考虑可以有效地改进移动补偿的预测并因此允许显著地降低的残余信息的编码的双向移动补偿来应用。进一步来说，对在这样的图像中的一个大块/块的直接预测模式的引入 可以进一步显著地（例如，超过 10-20%）提高效率，因为没有移动信息被编码。这样可以，例如，通过允许同时预测正向和反向移动信息从使用在对应的一个随后的参照图像的大块的移动向量中直接得到来应用。

作为例子，图 4 展示在时间 $t+1$ 在 B 图像中的直接预测，在时间 t 和 $t+2$ 基于 P 帧编码，及适用的移动向量（MV）。在这里，假设图像中的一个对象正以一个恒定的速度移动。着使得可能预测在 B 图像中的一个当前位置而不用传输任何的移动向量。直接模式的移动向量 $(\overline{MV}_{fw}, \overline{MV}_{bw})$ 对在第一个随后的 P

参照图像中的比较 MB 的移动向量 $\overrightarrow{MV}$ 基本上这样计算：

$$\overrightarrow{MV}_{fw} = \frac{TR_B \cdot \overrightarrow{MV}}{TR_D} \text{ 和 } \overrightarrow{MV}_{bw} = \frac{(TR_B - TR_D) \cdot \overrightarrow{MV}}{TR_D}$$

这里 TR_B 为当前 B 图像和比较 MB 的正向 MV 指向的参照图像之间的时间距离，及 TR_D 为未来参照图像和比较 MB 的正向 MV 指向的参照图像之间的时间距离。

然而在一些情况下现有的直接模式不能提供足够的解决方案，因此不能有效地利用这种模式的属性。特别地，这种模式现有的设计通常强迫直接大块的移动参数，当在随后的 P 图像中的比较大块被内编码时，为 0。例如，如图 6 所示，在现有的编码解码器中处理比较内（collated intra）时其中移动假设为 0。这本质上意味着，对这种情况，B 图像大块将被编码为在第一个随后的及过去的 P 参照中的两个比较大块的平均。这立即产生了下面的问题；如果一个大块为内编码的，那么一个人如何知道它和它的参照图像的比较大块有多大关系。在一些情况下，它们可能几乎没有任何实际的关系。因此，可能降低直接模式的编码效率。一中极端的情况可以在图 5 所示的场景改变情况中看到。图 5 展示了当场景改变发生在视频序列中时将发生什么及/或当比较块为内编码时（intra）将发生什么。在这里，在这个例子中，明显没有关系存在于场景改变的两个参照图像中。在这样的情况下，双向预测几乎不能提供好处。因此，直接模式可能完全被浪费了。然而，现有的直接模式的应用限制它总要进行一个大块的双向预测。

图 7 为示意图，展示根据本发明某些示范性的应用，当比较块的参照图像在后面的 P 图像中不为 0 时，直接模式如何被处理。

当使用多图像参照移动补偿时，直接模式大块存在一个另外的问题。直到最近，例如，JVT 标准才提供了时间距离信息（ TR_B 和 TR_D ），因此允许正确地缩放参数。最近，这在编码解码器的新修订（参阅，例如，联合视频小组（JVT）ISO/IEC MPEG 和 ITU-T VCEG，“联合委员会草案（CD）联合视频规范（ITU-T Rec. H.265 | ISO/IEC 14496-10 AVC）”，ITU-T JVT-C167，2002 年 5 月，包括在这里引用）中有改变。在新修订中，随后的 P 图像的移动向量参数为直接模式预测被等量地缩放，而不需要参照图像信息。这会导致直接模式显著的性能下降，因为恒定移动的假设不再被支持。

然而，即使有时间距离参数，使用如先前定义的直接模式仍然不能确定为最适合的解决方案。特别是对更靠近第一个正向参照图像的 B 图像，与那个图像的关联可能会比随后的参照图像更强。一个可能包含这些情况的极端的例子可以是一个场景 A 改变到场景 B，然后移回到场景 A 的序列（例如，向在一个新闻公告牌上发生的那样，等等）。上述所有都可能显著地降低 B 图像编码的性能，因为直接模式在编码过程中不能被有效地利用。

考虑到这些和其他问题，不象前面其中只使用时间预测的直接模式的定义，根据本发明的某些方面，一种其中同时考虑到了时间预测和/或空间预测的新的直接大块类型被引入。例如，所用的预测的类型可以取决于第一个随后的 P 参照图像的参照图像信息的类型。

根据本发明的某些其他方面，一个人也可以进一步当使用多图像参照时，通过考虑时间距离，如果这个可用的话，相当地改进 P 和 B 图像的移动向量预测。

这些增强方式应用在下述某些示范性的方法和装置中。方法和装置可以获得显著的比特率减少而获得类似或更好的质量。

直接模式增强：

在多数已有的视频编码系统中，直接模式被设计为一个双向预测方案其中移动参数始终根据随后的 P 图像中的移动参数在时间上被预测。在这一部分，提供一种增强的直接模式技巧，其中在这样的预测中，也可以/作为其他选择考虑到空间信息。

一种或多种下面的示范性技巧可以根据需要应用，例如，取决于系统的复杂度和/或规范。

一种技巧时应用直接模式的移动向量参数的空间预测而不考虑时间预测。例如，空间预测可以使用现有的用来做移动向量编码（如，中值预测）的移动向量预测技巧来应用。如果使用了多图像参照，那么相邻块的参照图像也可以被考虑进来（即使没有这样的限制及通用的参照，如 0，可以始终被使用）。

移动参数和参照图像可以按如下方式预测并引用图 3 来展示和假设为可用的并为一个图像的一部分的部分 A-E（例如，大块，片段等）关联的空间预测。在这里，E 总的来说通过 A、B、C 作为其中值（A、B、C）来预测。如果 C 实际上是在图像之外，那么用 D 来代替。如果 B、C 和 D 都在图像之外，那

么只有 A 被用到，这里如果 A 不存在，那么使用 (0, 0) 来代替。熟悉相关技术的人应承认空间预测还可以应用在子块级别。

总的来说，空间预测可以看作是所有在一个图像或在同一图像中的一组大块/块中计算的可用的移动信息的一个线性或非线性函数。

有不同的可用方法可以安排用来预测直接模式的参照图像。例如，一种方法可以用来在预测中选择一个最小的引用图像。在另一种方法中，可以选择一个中值参照图像。在某些方法中，可以在一个最小参照图像和中值参照图像中选择，例如，如果最小为 0 时。在其他应用中，高优先级也可以给予垂直的或水平的预测器（A 和 B）由于它们和 E 可能有更强的相关性。

如果一个预测不存在（例如，所有周围的大块都以同样仅为 FW 或 BW 的方向来预测或为内编码的（intra）），那么仅使用存在的一个（单方向预测）或这样可以从可用的那一个来预测。例如如果正向预测是可用的那么：

$$\overrightarrow{MV}_{bw} = \frac{(TR_B - TR_D) \cdot \overrightarrow{MV}_{fw}}{TR_B}$$

如果随后的 P 引用如在现有的编码解码器中那样是非内编码的（intra），那么对大块进行时间预测。请看图 8，在其中 MV_{FW} 和 MV_{BW} 从空间预测（周围大块的中值 MV）得到。如果没有一个可用（即，没有预测器）那么使用单向预测。如果随后的 P 引用是内编码的（intra），那么可以如上所述使用空间预测来代替。假设没有限制存在，如果其中一种预测不可用那么直接模式成为一个单向预测模式。

这可以当场景改变时，例如，如图 9 所示，和/或即使当渐变存在于一个视频序列中时相当地有利于视频编码。如图 9 所示，空间预测可以用来解决场景改变的问题。

如果时间距离信息在一个编码解码器中不可用，那么当比较 P 参照块有一个非 0 参照图像时，时间预测将不象在块的直接模式中那么有效。在这样的情况下，空间预测也可以如上述来使用。作为其他选择，一个人可以估计缩放参数，如果周围的大块的一个也使用和比较 P 参照块相同的参照图像。进一步来说，可以对有一个非 0 参照的 0 移动（或接近于 0 的移动）进行特殊的处理。在这里，不管时间距离，正向和反向移动向量可以始终取 0。然而，最好的解决方案可以是始终检查周围大块的参照图像信息并基于其决定在这

样的情况下如何使用直接模式。

更特别地，例如，给定一个非 0 参照，下面的子情况可以被考虑：

情况 A：如果移动向量比较 P 块为 0，使用时间预测。

情况 B：如果所有周围的大块使用和比较 P 参照不同的参照图像，那么空间预测看起来为一个更好的选择而不使用时间预测。

情况 C：如果在 B 图像中的移动流看起来和 P 参照图像中的那一个很不一样，那么使用空间预测。

情况 D：直接模式大块的空间或时间预测可以在图像头中用信号表示。可以预先分析图像来决定使用哪一种。

情况 E：基于空间信息来修正空间预测参数（反之亦然）。因此，例如，如果都看起来有同样或近似相同的相位信息，那么空间信息可能是非常好的直接模式预测候选者。修正也可以对相位进行，因此修正预测的子像素准确度。

图 10 展示一个在 B 图像编码中对直接模式的联合时间预测。在这里，在这个例子中，直接模式可以为一个取决于可用信息的 1 方向到 4 方向模式。代替使用直接模式大块的双向预测，可以应用这种模式的一个多假设扩展并使用多个预测。

组合上面的讨论，直接模式大块可以基于可用的信息用从一个到四个可能的移动向量来预测。例如，这可以基于比较 P 参照图像大块和当前 B 图像的周围大块的模式来决定。在这样一种情况下，如果空间预测和时间预测差别很大，它们中的一种可以被选中作为唯一的预测。因为上述空间预测可能使用一个和时间的那一个不同的参照图像，同一个大块可能用超过 2 个参照图像来预测。

JVT 标准并不限制第一个未来参照为一个 P 图像。因此，在这样的一个标准中，一个图像可以如图 12 所示为一个 B 图像，甚至一个多重假设（MH）图像。这意味着更多的移动向量被赋给每个大块。这意味着一个人也可以使用这个属性通过更有效地利用附加的移动信息来提高直接模式的有效性。

在图 12 中，第一个随后的参照图像是一个 B 图像（图像 B_8 和 B_9 ）。这使得一个人可以使用更多的直接模式预测候选者，尤其是在 B 图像中使用双向预测时。

特别地，一个可以执行下面的操作：

a.) 如果在第一个未来参照中的比较参照块正在使用双向预测，对应的移动向量（正向或反向）被用来计算当前块的移动向量。因为对应于一个未来参照图像的参照的反向移动向量，在估计当前移动参数时需要特别注意。例如，请看图 12，其中第一个随后的参照图像是一个 B 图像（图像 B_8 和 B_9 ）。这使得一个人可以使用更多的直接模式预测候选者，尤其是在 B 图像中使用双向预测时。因此，如图所示，由于 B_8 、和 P_6 之间的时间距离， B_8 的反向移动向量 $\overrightarrow{MV}_{B8bw}$ 可以计算为 $2 \times \overrightarrow{MV}_{B7bw}$ 。类似地对 B_9 反向移动向量可以取 $\overrightarrow{MV}_{B7bw}$ ，如果这些指 B_7 。一个人也可以限制这些来引用第一个是的 P 图像，在这种情况下这些移动向量可以相应地被缩放。类似的结论可以对正向移动向量作出。多图像参照或内编码的（intra）大块可以类似于前面的讨论来处理。

b.) 如果使用对比较块的双向预测，那么，在这个例子种，一个人可以通过投影和反转参照的反向和正向移动向量对一个大块估计四个可能的对直接模式情况的预测。

c.) 选择性的投影和反转可以根据时间距离来使用。根据这种解决方案，一个人从对预测更可靠的参照图像中选择移动向量。例如，考虑图 12 中的示意图，一个人将注意到 B_8 比 P_6 更接近 P_2 。这意味着反向移动向量 B_7 不可能为一个很可靠的预测。在这种情况下，直接模式移动向量因而只可以从 B_7 的正向预测来计算。然而，对 B_9 ，两个移动向量看起来对预测都是足够的并因而都可以使用。这样的决定/信息也可以在图像头中来决定/支持。其他情况和规则也可以应用。例如，也可以考虑一个预测的附加的空间置信度和/或一个移动向量相位。注意，特别地，如果正向和反向移动向量没有关系，那么反向移动向量对使用可能非常不可靠。

B 图像的单图像参照：

存在一种特殊情况，仅对 B 图像使用一个图像参照（虽然，通常一个正向和一个反向参照是需要的）而不管有多少参照图像在 P 图像中使用。从在当前 JVT 编码解码器中编码向量的观察来看，例如，注意，如果一个人比较使用 B 图像的单图像参照和多图像参照情况，即使 P 图像对多图像情况的编码性能几乎总是优于使用单图像，而这对 B 图像来说不一定总是这样。

这个观察的一个原因是为每个大块使用的参照图像的开销。考虑 B 图像比

P 图像更加依赖于移动信息，参照图像信息开销 降低了以一个给定的比特率为残余信息传输的比特数，从而降低了效率。一个很简单并有效的解决方案可以是每个反向或正向移动补偿仅选择一个图像参照，因此不需要传输任何参照图像信息。

这通过引用图 13 和 14 来说明。如图 13 所示，B 图像可以通过仅使用一个未来和过去参照图像来限制。因此，对直接模式移动向量计算，移动向量的投影是必要的。一个比较 MV 的到当前参照用来做时间方向预测的投影在图 14 中展示（注意可能 $TD_{D,0} > TD_{D,1}$ ）。因此，在这个例子中，直接模式移动参数通过投影引用对两个参照图像的其他参照图像的移动向量，或使用如图 13 所示的空间预测来计算。注意这样的选项不仅使得可能减少 B 图像编码复杂度，还趋向于减少内存需求，因为如果 B 图像允许引用 B 图像时，更少的 B 图像（例如，最大为两）需要被存储。

在某些情况下，第一个未来参照图像的一个参照图像可以不再出现在参照缓冲中。这会立即对直接模式大块的估计产生一个问题并需要特殊处理这样的情况。如果使用单个图像参照，显然没有这样的问题。然而，如果需要多图像参照，那么可能的解决方案包括投影移动向量到任一个正向参照图像，及/或到离不可用的图像最近的参照图像。任一个解决方案都是可行的，这里空间预测可以再次成为一个作为其他选择的解决方案。

单图像和多图像参照移动补偿移动向量预测的改进

多图像参照移动补偿的移动向量预测会显著地影响 B 和 P 图像编码的性能。现有的标准，例如，JVT，并不总是考虑在预测中使用的大块的参照图像。这样的标准唯一的考虑是当仅有一个预测大块使用同一个参照时。在这样的情况下，仅有那个预测器被用在移动预测中。如果仅有一个或所有的预测器使用一个不同参照时，不考虑参照图像。

在这样的情况下，例如，并根据本发明某些进一步的方面，一个人可以根据它们对当前参照的时间距离缩放预测器。请看图 11，展示了考虑预测器大块（Pr）的参照图像信息以及正确调整（例如，缩放预测器）的性能的一个当前块（C）的移动向量预测。

如果预测器 A、B 和 C 使用分别有时间距离 TR_A 、 TR_B 和 TR_C 的参照图像，并且当前参照图像有一个等于 TR 的时间距离，那么中值预测器按下面公式计

算：

$$\overrightarrow{MV}_{pred} = TR \times Median\left(\frac{\overrightarrow{MV}_A}{TR_A}, \frac{\overrightarrow{MV}_B}{TR_B}, \frac{\overrightarrow{MV}_C}{TR_C}\right)$$

如果使用整数计算，可能更简单地把乘法放在中值中，因此提高准确率。除法也可以用移位来代替，但那样会降低性能，在其中处理带符号的移位（ $-1 \gg N = -1$ ）可能是需要的。因此在这样的情况下有对进行适合的缩放可用的时间距离信息是很重要的。这也可以提供在头中，如果否则不能预测的话。

如上所述的移动向量预测基本上是偏向于中值的（Median biased），即在一组预测器中的中值被选中用来预测。如果一个人对一个移动向量（MV）只使用一种类型的大块（例如 16×16 ），那么这些预测器可以按例如，如图 15 所示来定义。在这里，MV 预测器展示为对一个 MV。在图 15a 中，MB 不在第一行或最后一列。在图 15b 中，MB 在最后一列。在图 15c 中 MB 在第一行。

JVT 标准在这上面进一步通过也考虑只有三个预测器中的一个存在（即，大块是内编码的（intra）或在多图像预测的情况下使用一个不同的参照图像）的情况来改进。在这样的情况下，只有现有的或同样的参照预测器被用来预测并且所有其他的预测器没有被检查。

内编码并不总是意味着一个新的物体出现或者场景改变。可能相反地，例如，是移动估计和补偿不足以代表当前物体（例如，搜索区域、使用的移动估计算法，残余（residue）的量子化，等等）这样的情况，并且可以通过内编码来取得更好的结果。可用的移动预测器仍然可以足够提供一个好的移动向量预测器解决方案。

令人感兴趣的是考虑一个大块中的每一个被赋予不同的移动信息的子块。MPEG-4 和 H.263 标准，例如，最多可以有四个这样的子块（例如，大小为 8×8 ），这里 JVT 标准允许最多有十六个子块而也可以处理不同的块大小（例如， 4×4 ， 4×8 ， 8×4 ， 8×8 ， 16×8 ，及 16×16 ）。另外 JVT 也允许 8×8 内编码子块，因此使得情况更加复杂。

考虑 JVT 和 MPEG-4/H.263 的一般情况（ 8×8 和 16×16 ），一个 16×16 大块的预测器集合在有相应地类似于图 15a-c 的安排的图 16a-c 中展示。在这里展示的移动向量预测器是一个有 8×8 分区的 MV。技术所述的预测器可以在一些情况下产生合理的结果，看起来它们并未足够地包含所有可能的因此。

下面看图 17a-c, 这也是有类似于图 15a-c 的安排。在这里, 在图 17a-c, 有两个也可以在 预测相位 (C_1 和 A_2) 考虑的附加的预测器。如果也考虑 4×4 的块, 这增加 4 种可能的预测器。

代替使用三个预测器 A、B 和 C (或 A_1 、B, 和 C_2) 的一个中值, 一个人可以有一些附加的, 并显然更可靠的选项。因此, 例如, 一个人可以看到预测器 A_1 和 C_2 本质上互相非常接近并且很可能它们在预测相位上不是很有代表性。相反, 选择预测器 A_1 、 C_1 , 和 B 由于它们的分离看起来是一个更可靠的解决方案。一个其他选择也可以是选择 A_2 代替 A_1 但那也过于接近预测器 B。模拟表明第一种情况通常是一个更好的选择。对最后一列 A_2 可以用来代替 A_1 。对第一行 A_1 或 A_2 甚至它们的平均值都可以使用。对这个应用高至 1% 的增益在 JVT 中被注意到。

前面的情况对最后一列增加了一些检验。通过查看图 17b, 例如, 很明显这趋向于提供可用的最好的分区。因此, 一个可选的解决方案可以是选择 A_2 、 C_1 , 和 B (从左上位置)。然而, 这并不总是被推荐的, 因为这样的应用会反过来一些右边 (right) 的预测器的性能。

一个作为其他选择的解决方案可以是使用一个大块中的预测器的平均值。中值可以计算如下:

$$\overline{MV}_{pred} = Median (Ave (\overline{MV}_{C_1}, \overline{MV}_{C_2}), Ave (\overline{MV}_{A_1}, \overline{MV}_{A_2}), \overline{MV}_B)$$

对中值行/列计算, 中值可以计算为:

$$\begin{aligned} \overline{MV}_{pred} = Median (Median (\overline{MV}_{C_1}, \overline{MV}_{C_2}, \overline{MV}_D), \dots \\ Median(\overline{MV}_D, \overline{MV}_{A_1}, \overline{MV}_{C_1}), Median(\overline{MV}_B, \overline{MV}_{A_1}, \overline{MV}_{A_2})) \end{aligned}$$

另一个可能的解决方案是一个中值 5 解决方案。这可能是最复杂的解决方案由于计算 (例如可以使用快速排序或冒泡排序), 但可能产生最好的结果。如果考虑 4×4 的块, 例如, 那么也可以使用中值 9:

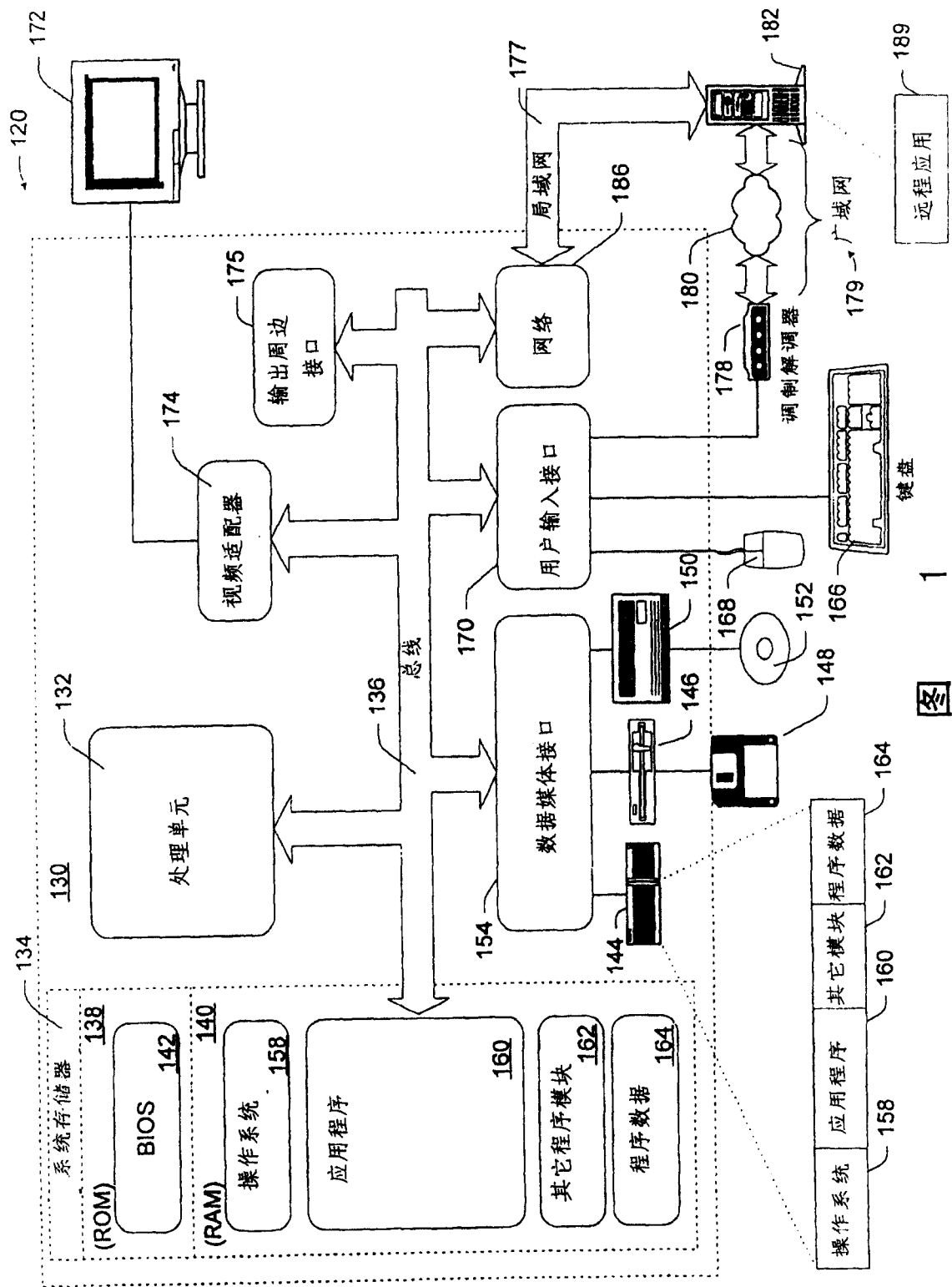
$$\overline{MV}_{pred} = Median (\overline{MV}_{C_1}, \overline{MV}_{C_2}, \overline{MV}_D, \overline{MV}_B, \overline{MV}_{A_1}, \overline{MV}_{A_2})$$

考虑 JVT 允许内编码的子块存在于一个间大块 (例如, 树状大块结构), 这也可以考虑在移动预测中。如果一个被用来做 MV 预测的子块 (例如, 来自于上面或左边的大块) 是内编码的, 那么相邻的子块可以用来代替。因此, 如果 A_1 是内编码的而 A_2 不是, 那么 A_1 可以被 A_2 替代用在预测中。进一步的可能性是用来自左上位置的 MV 预测器替换一个缺少的内编码大块。在图 17a 中, 例如, 如果 C_1 是缺少的那么可以使用 D 来代替。

在上面的部分，展示了 B 图像直接模式和移动向量预测的几个改进。展示了空间预测也可以用于直接模式大块；其中为了更准确的预测，移动向量预测应考虑时间距离和子块信息。这样的考虑会显著的改进任何适用的视频编码系统的性能。

结论

虽然上面的说明使用了特定于结构特性和/或方法论的动作的语言，应理解在所附的权利要求中定义的发明并不局限于上述特定的特性或动作。更确切地，特定的特性和动作以应用本发明的示范性的形式在这里揭示。



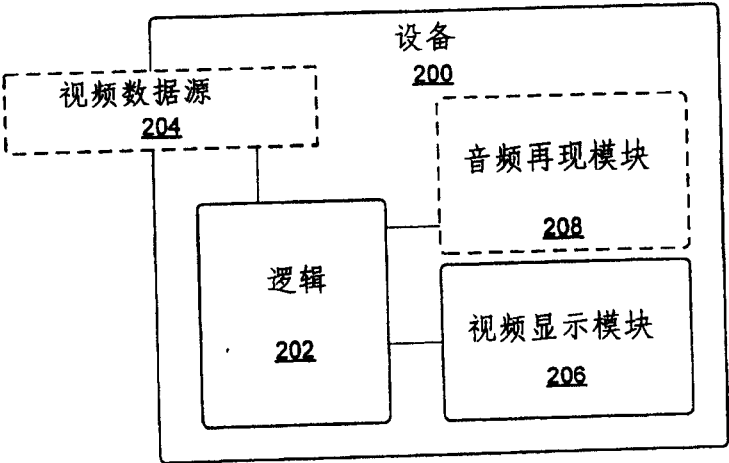


图 2

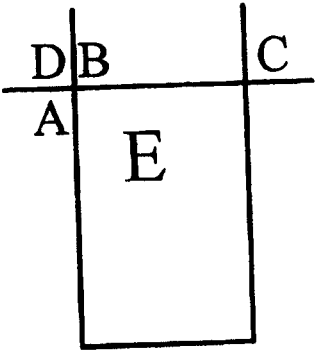


图 3

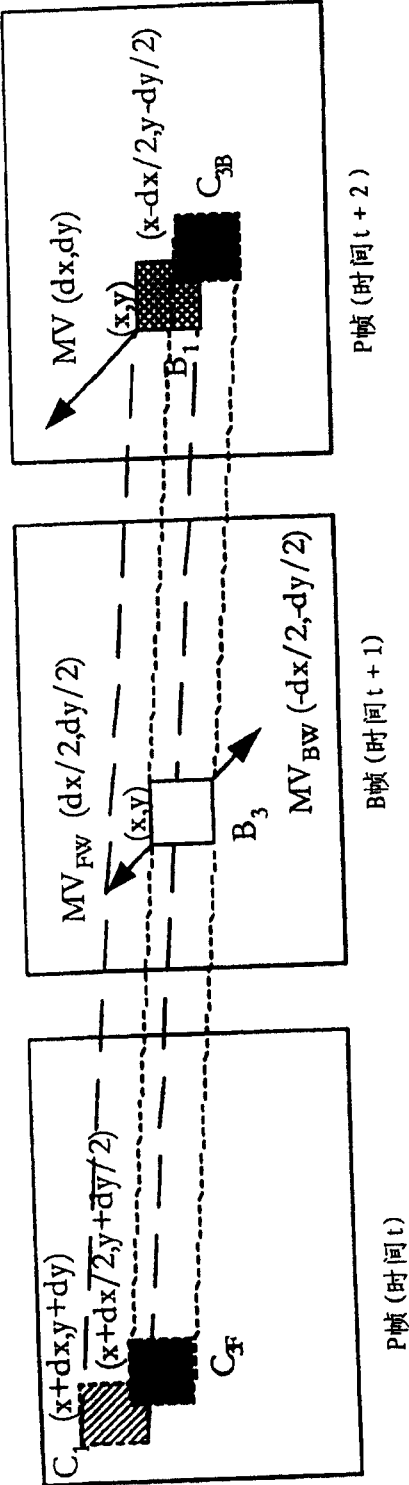


图 4

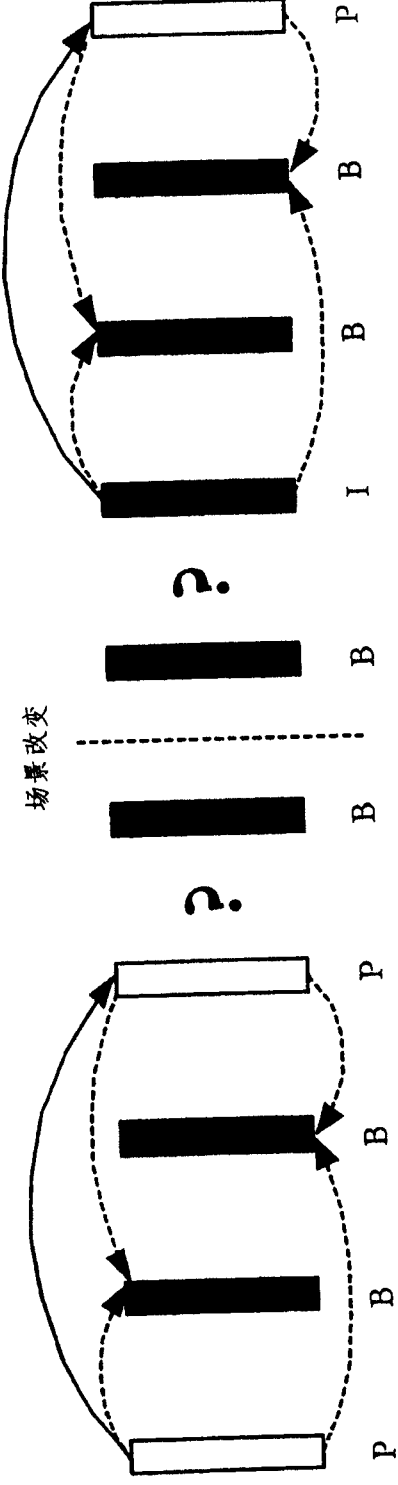


图 5

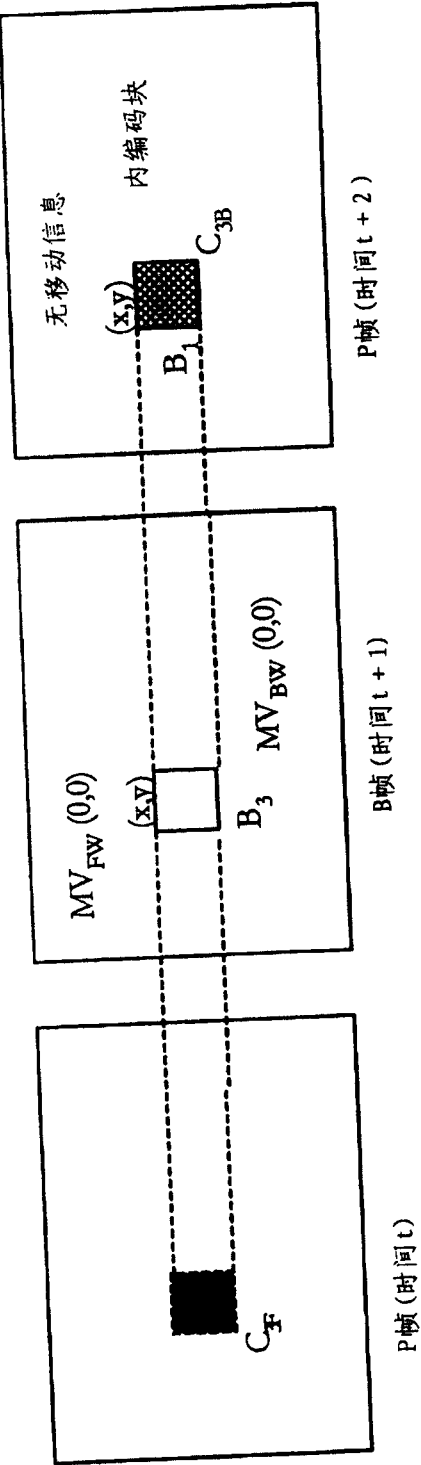


图 6

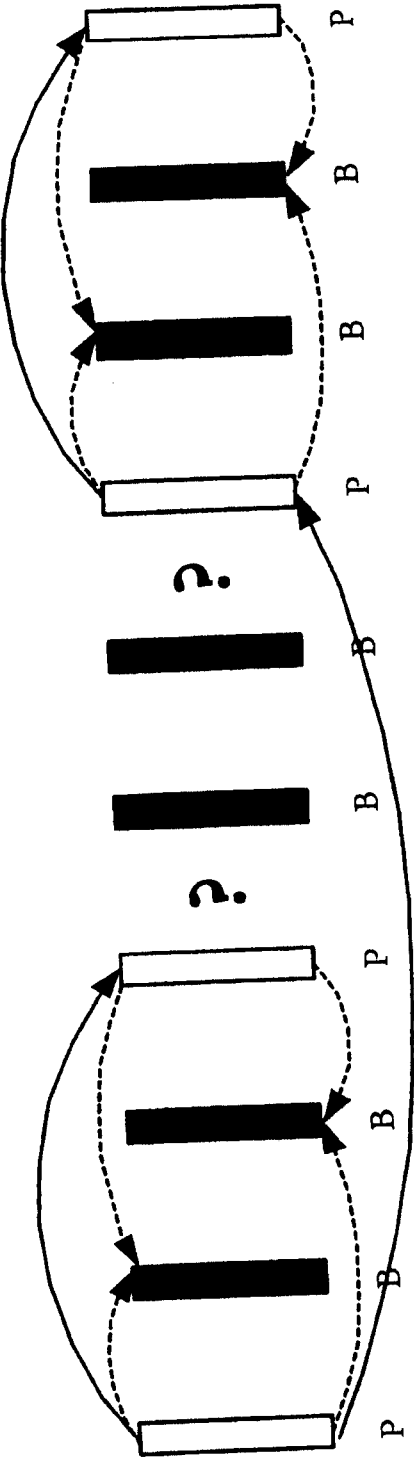


图 7

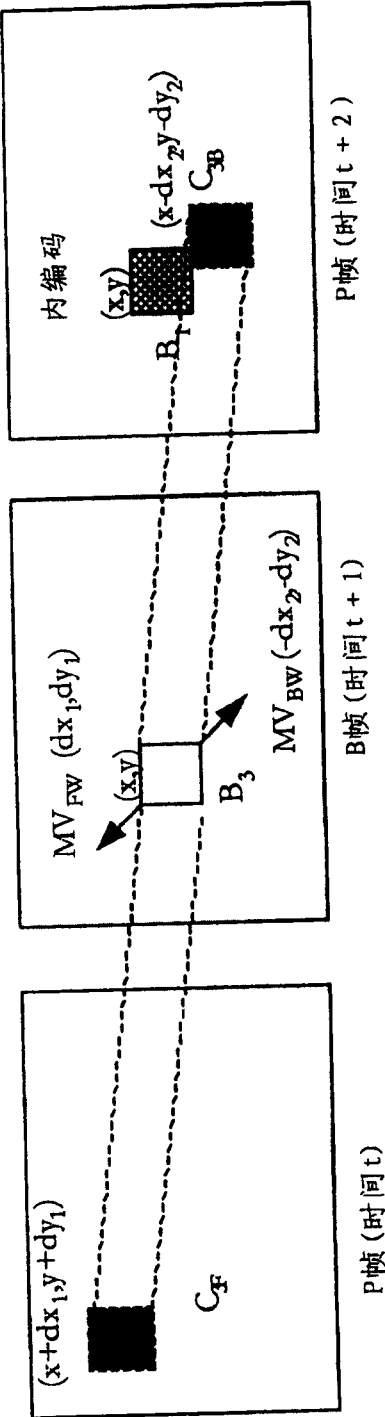


图 8

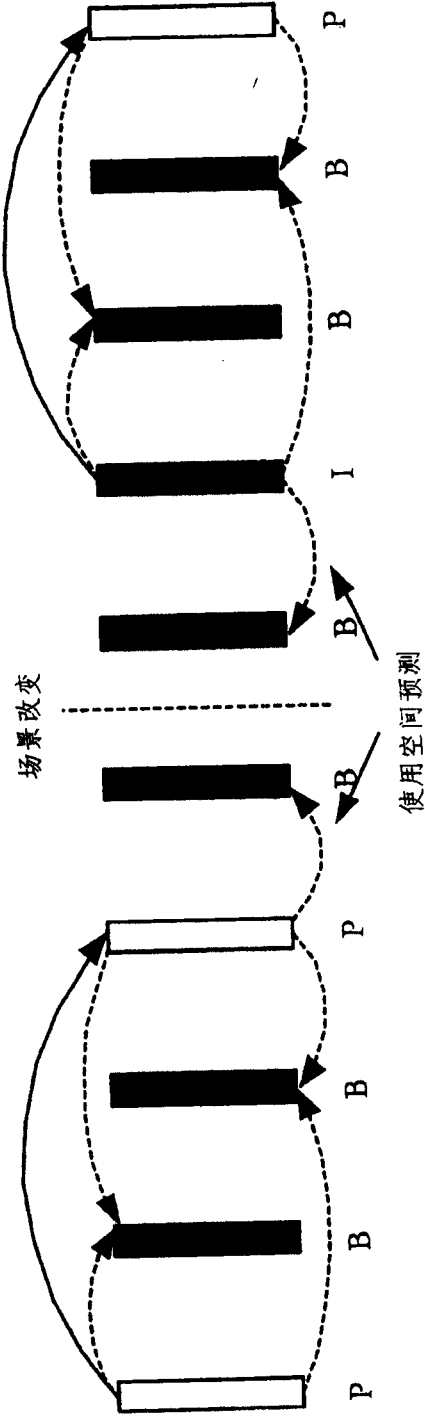


图 9

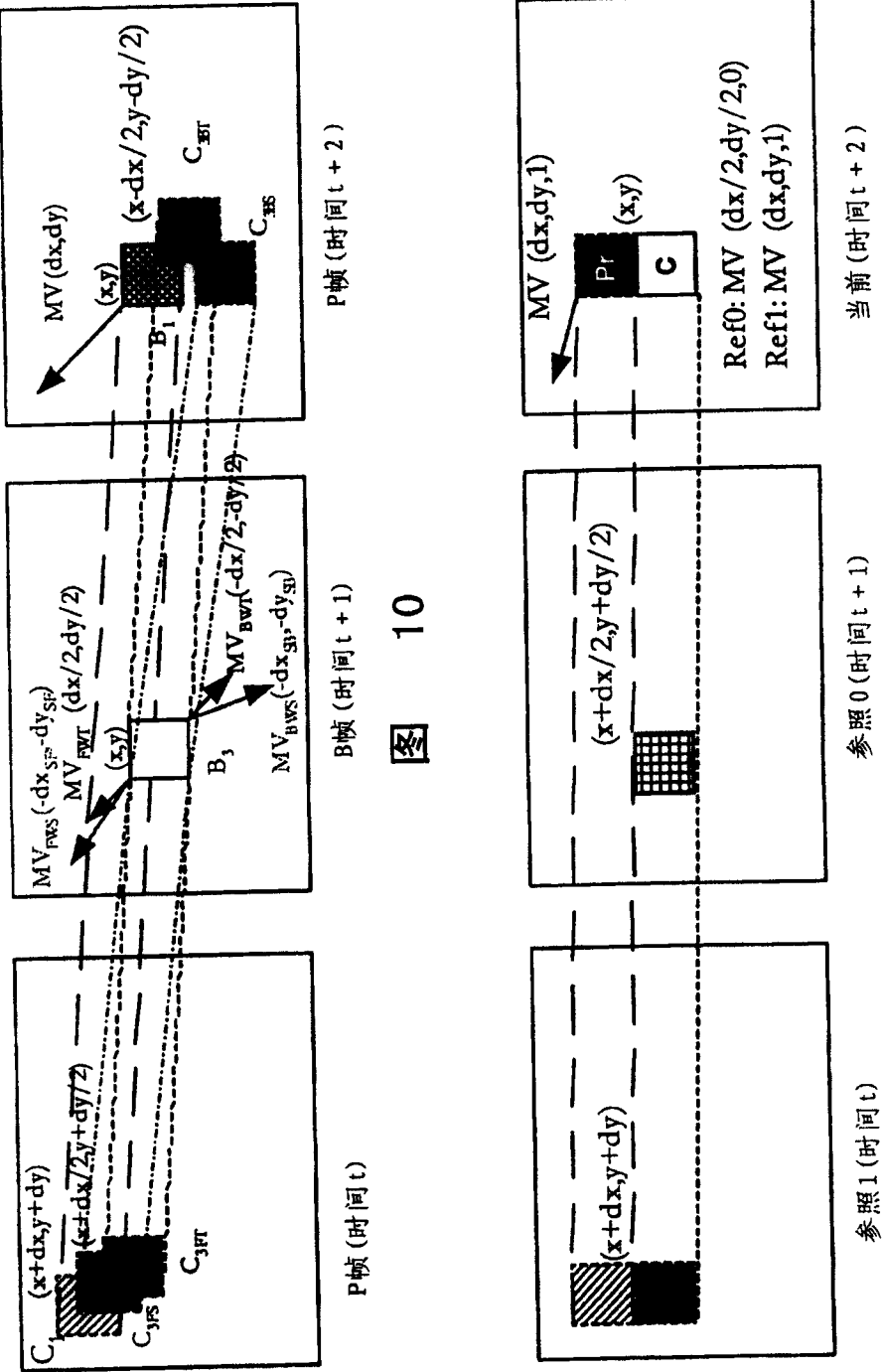


图 10

图 11

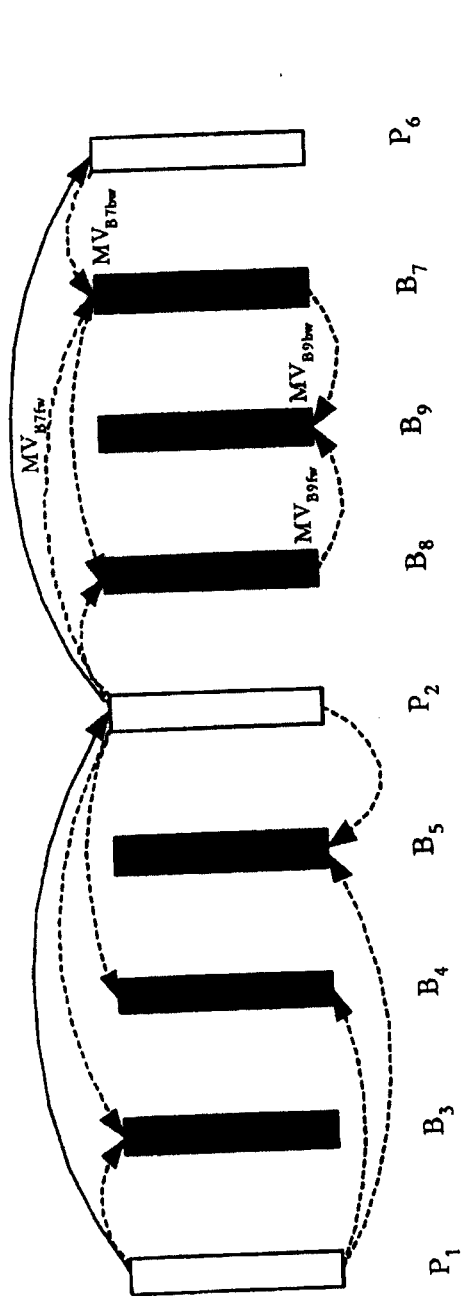


图 12

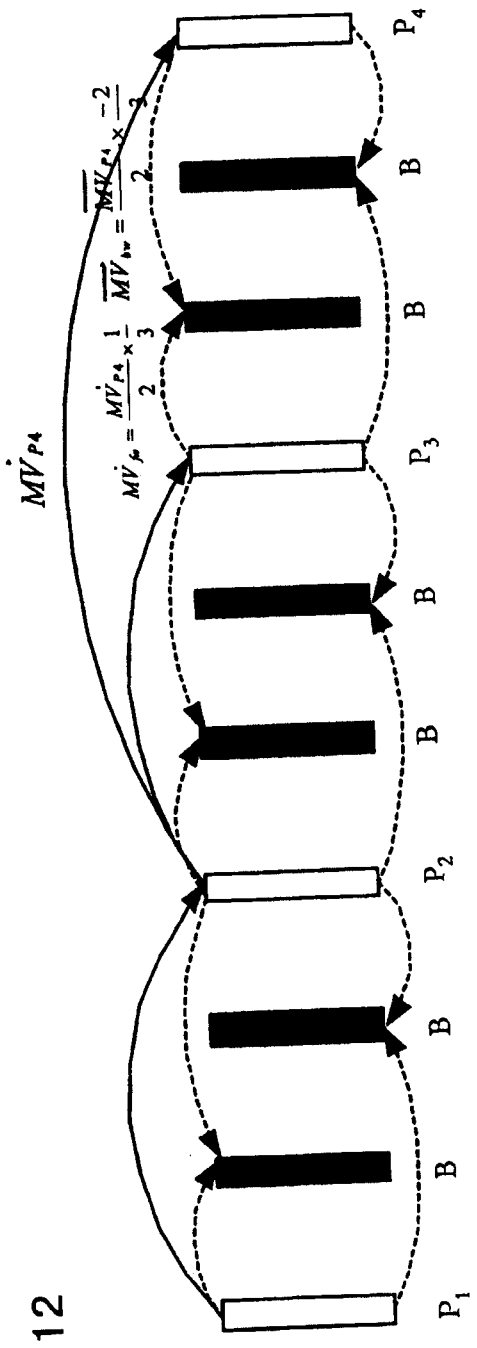


图 13

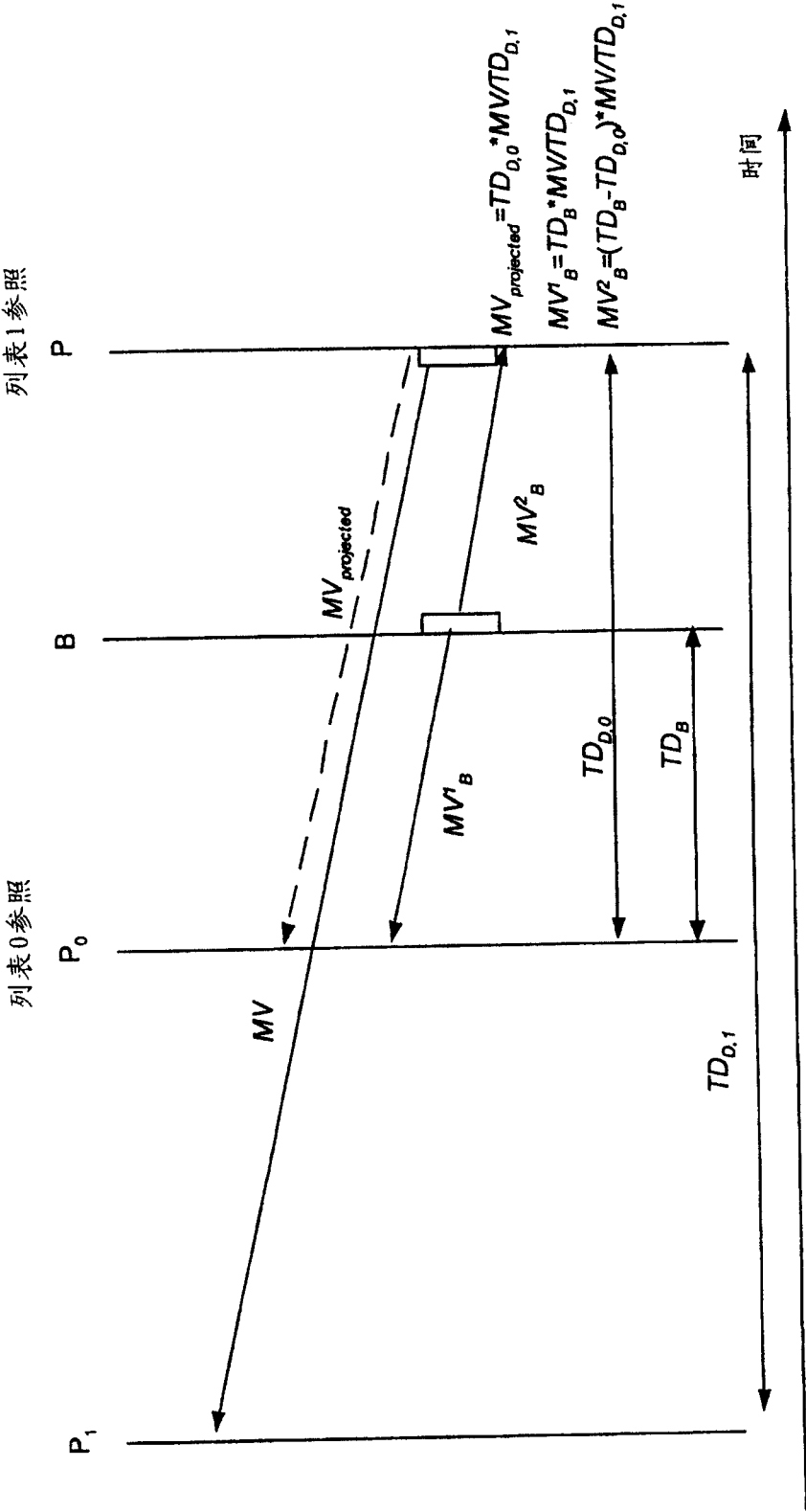


图 14

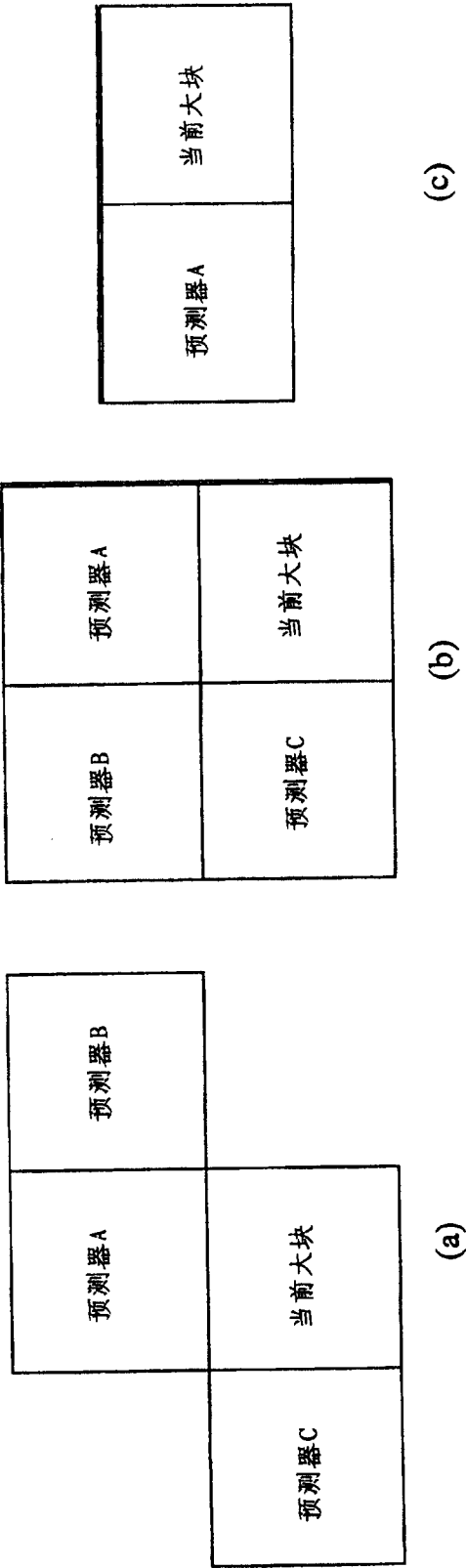


图 15

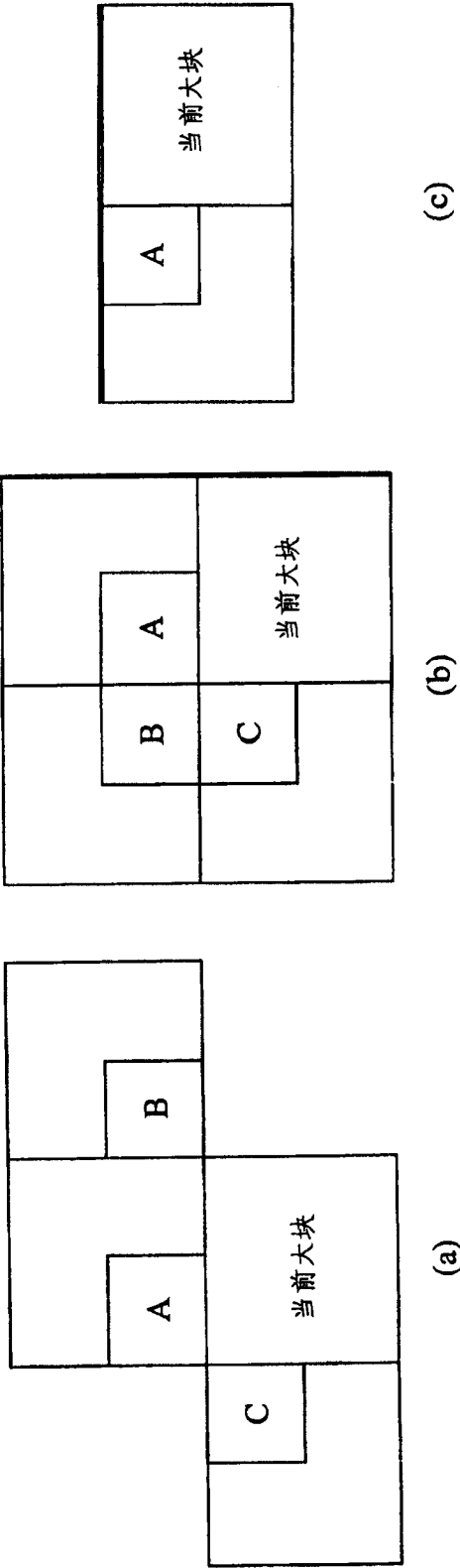


图 16

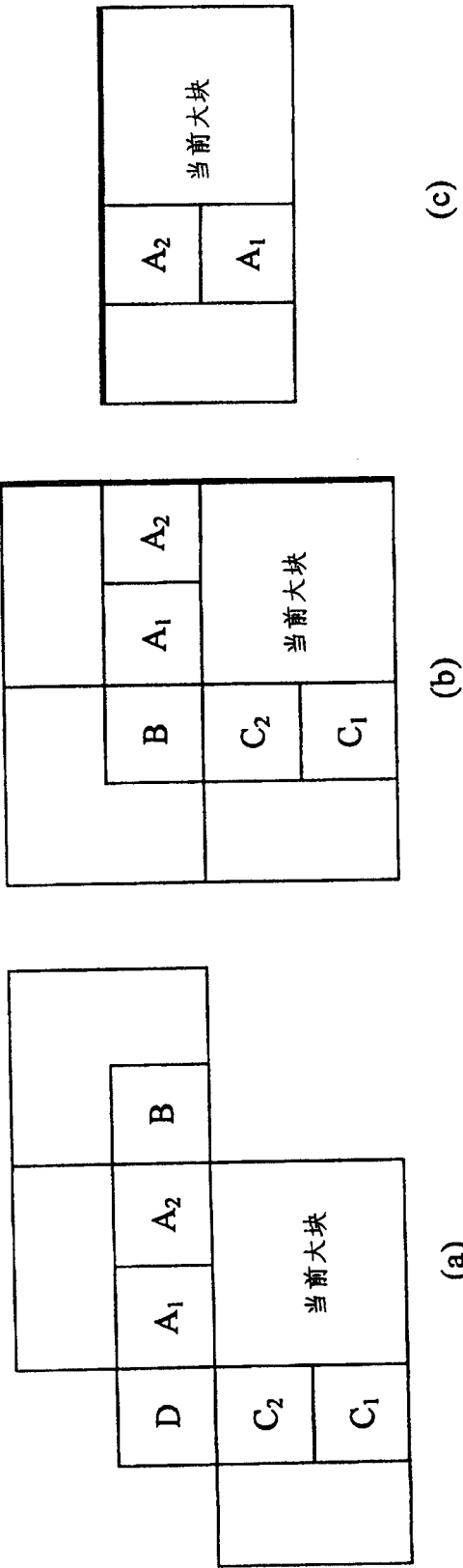


图 17