

[19] 中华人民共和国国家知识产权局

[51] Int. Cl<sup>7</sup>



# [12] 发明专利申请公开说明书

H04N 7/12

H04N 7/24

H04N 7/32

H04N 1/417

[21] 申请号 03814629.0

[43] 公开日 2005 年 8 月 31 日

[11] 公开号 CN 1663258A

[22] 申请日 2003.6.27 [21] 申请号 03814629.0

[30] 优先权

[32] 2002.6.28 [33] US [31] 10/187,395

[86] 国际申请 PCT/US2003/020397 2003.6.27

[87] 国际公布 WO2004/004310 英 2004.1.8

[85] 进入国家阶段日期 2004.12.22

[71] 申请人 杜比实验室认证公司

地址 美国加利福尼亚州

[72] 发明人 G·A·迪莫斯

[74] 专利代理机构 北京纪凯知识产权代理有限公司

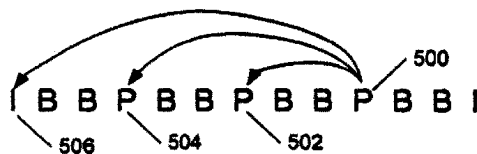
代理人 赵蓉民

权利要求书 3 页 说明书 35 页 附图 15 页

[54] 发明名称 改进的视频压缩帧插值

[57] 摘要

一种用于在视频图像压缩系统中改进一个或多个预测帧的图像质量的方法、系统和计算机程序，其中每个帧包括多个像素。某些类型的帧的宏块图像区域在某些情况下可以参考一个或多个可参考帧来编码，而在其他情况下可以参考两个或多个可参考帧来编码。这种编码可能包括插值，例如不相等加权的插值。图像区域的 DC 值或 AC 像素值也可以在加权或未加权的条件下进行插值。具有可变数量的双向预测帧的这类帧的编码模式可以动态地确定。可以以不同于显示顺序的传送顺序把帧从编码器传送到解码器。在运动向量补偿预测的过程中，可以把锐化和/或柔化滤波器应用于某些帧的图像区域。



I S S N 1 0 0 8 - 4 2 7 4

1. 一种用于视频图像压缩的方法，所述方法包括：  
提供一个包括图像区域的帧序列，所述帧包括预测帧和可参考帧；  
和  
通过参考所述序列中的两个或多个前面的可参考帧对至少一个预测帧的一个图像区域进行编码。
2. 根据权利要求1所述的方法，其中所述的至少一个预测帧包括一个可参考帧。
3. 根据权利要求1所述的方法，其中所述可参考帧包括预测帧和内部帧。
4. 根据权利要求1所述的方法，其中所述预测帧包括可参考预测帧和双向预测帧。
5. 根据权利要求4所述的方法，其中所述至少一个预测帧包括一个双向预测帧。
6. 根据权利要求5所述的方法，进一步包括：  
通过参考所述序列中的两个或多个后续的可参考帧对一个双向预测帧的一个图像区域进行编码。
7. 根据权利要求1所述的方法，进一步包括：  
识别所述两个或多个前面的可参考帧中的至少一个。
8. 根据权利要求1所述的方法，其中所述编码包括使用来自所述两个或多个前面的可参考帧的选定图像区域的不相等加权来编码。
9. 根据权利要求8所述的方法，进一步包括：  
识别所述的不相等加权。
10. 根据权利要求8所述的方法，其中所述的不相等加权包括大于

1 或小于 0 的权重。

11. 根据权利要求 8 所述的方法, 其中所述帧包括排列在图像区域中的帧, 和

其中所述编码包括使用对应于所述两个或多个可参考帧的不相等像素值来编码。

12. 根据权利要求 1 所述的方法, 其中所述的帧序列包括可参考帧和双向预测帧, 每个所述帧包括排列在宏块中的像素值。

13. 根据权利要求 12 所述的方法, 进一步包括:

按显示顺序基于来自一个或多个预测帧的运动向量使用直接模式预测来确定一个双向预测帧中的至少一个宏块。

14. 根据权利要求 13 所述的方法, 其中所述编码包括使用来自所述两个或多个前面的可参考帧的选定图像区域的一个不相等加权来编码。

15. 根据权利要求 14 所述的方法, 其中用小于 0 或大于 1 的帧比例分数对至少一个这样的运动向量进行换算。

16. 一种用于视频图像压缩的方法, 所述方法包括:

提供一个包括图像区域的帧序列, 所述帧包括双向预测 (B) 帧和可参考帧, 所述可参考帧包括预测 (P) 帧和内部 (I) 帧; 和

通过参考所述序列中的一个前面的可参考帧对至少一个 P 帧的一个图像区域进行编码, 其中所述的前面的可参考帧被至少一个居间的可参考帧从所述 P 帧中隔开。

17. 根据权利要求 16 所述的方法, 进一步包括:

识别所述的前面的可参考帧; 和  
用所述识别通知解码器。

18. 一种用于在具有可参考帧和双向预测帧序列的视频图像压缩系统中动态地确定具有可变数量的双向预测帧的这类帧的编码模式的

方法，所述方法包括：

选择一个初始序列，其开始于一个可参考帧，具有至少一个紧接着的后续双向预测帧，并以一个可参考帧结束；

增加一个可参考帧到所述初始序列的结尾以创建一个测试序列；

对照一个选定的评估标准对所述测试序列进行评估；

对于评估所述测试序列的每个满足步骤，在所述的所增加的可参考帧之前插入一个双向帧并重复所述的评估步骤；和

如果所述评估测试序列的步骤是不满足的，那么接受所述的前面的测试序列作为当前的编码模式。

19. 根据权利要求 18 所述的方法，进一步包括：

为一个选定的测试序列的帧的至少一个图像区域确定一个插值混合比例。

## 改进的视频压缩帧插值

### 相关申请的交叉参考

本申请是 2001 年 7 月 11 日提交的美国申请序列号 09/904203 的部分延续申请，并要求 2002 年 6 月 28 日提交的美国 C.I.P.申请序列号 10/187395 的优先权。

### 技术领域

本发明涉及视频压缩，更特别地是涉及在 MPEG 类的编码和解码系统中改进的视频压缩帧插值。

### 背景技术

#### MPEG视频压缩

MPEG-2 和 MPEG-4 是国际视频压缩标准，它们定义的各自的视频语法，提供了高效的方法来以更紧凑的编码数据表示图像序列。编码比特的语言就是所谓的“语法”。例如，几个标记可以表示整个样本块（例如对于 MPEG-2 是 64 个样本）。两种 MPEG 标准都描述了解码（重构）过程，即把编码的比特从紧凑的表示变换成图像序列的原始格式的一个近似。例如，在编码比特流中的一个标志可以表示在利用离散余弦变换（DCT）算法对随后的比特解码之前，是否使用预测算法先处理这些比特。包括解码过程的算法由这些 MPEG 标准所定义的语义来规定。这个语法可以被用于开发普通的视频特征，例如空间冗余、时间冗余、匀速运动、空间屏蔽等等。MPEG 解码器必须能够解析和解码输入的数据流，但是只要数据流符合相应的 MPEG 语法，就可以使用各种各样的可能的数据结构和压缩技术（虽然在技术上这背离了标准，因为语义不一致）。在某种替换语法内携带所需的语义也是可能的。

这些 MPEG 标准使用各种压缩方法，包括帧内方法和帧间方法。在大多数视频场景中，背景保持相对稳定而动作在前景中发生。背景

可能移动，但是通常许多场景是冗余的。这些 MPEG 标准通过创建称为“内”帧或“I 帧”的参考帧开始压缩。在不参考其他帧的情况下压缩 I 帧，因此 I 帧包含了视频信息的一个完整的帧。I 帧提供用于随机访问的数据比特流的入口点，但只能被适度地压缩。典型地，表示 I 帧的数据在比特流中每隔 12 到 15 帧放置（虽然在某些情况下在 I 帧之间使用更宽的间隔也是有用的）。据此，由于落入参考 I 帧之间的帧中只有一小部分不同于包围的（bracketing）I 帧，所以只有图像差别被捕获、压缩和存储。两种类型的帧被用于这种差别——预测帧（P 帧）和双向预测的（或插值的）帧（B 帧）。

P 帧一般根据过去的帧（I 帧或者前面 P 帧）来编码，而且通常被用作后续 P 帧的参考。P 帧受到相当大的压缩量。B 帧提供最高的压缩量，但是为了被编码既需要过去的参考帧也需要未来的参考帧。P 帧和 I 帧是“可参考帧”，因为它们可以被 P 帧或 B 帧参考。

宏块（macroblock）是图像像素区域。对于 MPEG-2，宏块是 4 个 8x8 DCT 块的 16x16 像素组，对于 P 帧加上一个运动向量，对于 B 帧加上一个或两个运动向量。可以使用帧内或帧间（预测的）编码对 P 帧内的宏块单独编码。可以使用帧内编码、前向预测编码、后向预测编码或同时使用前向和后向（即双向插值）预测编码对 B 帧内的宏块单独编码。在 MPEG-4 视频编码中所用的结构与此稍有不同，但是是类似的。

在编码之后，MPEG 数据比特流包括 I 帧、P 帧和 B 帧的序列。序列可以由 I 帧、P 帧和 B 帧的几乎任何模式组成（关于它们的位置有几个次要的语义限制）。但是，在业界实践中常见的是具有固定的帧模式（例如，IBBPBBPBBPBBPBB）。

### 运动向量预测

在 MPEG-2 和 MPEG-4 中（以及类似的标准，例如 H.263），B 型（双向预测的）帧的使用被证明是有益于压缩效率的。这种帧的每个宏块的运动向量可以由以下三种方法中的任何一种来预测：

模式 1：根据前面的 I 帧或 P 帧进行的前向预测（即非双向预测的帧）。

模式 2: 根据后续的 I 帧或 P 帧进行的后向预测。

模式 3: 根据后续和前面的 I 帧或 P 帧进行的双向预测。

模式 1 和用于 P 帧的前向预测方法相同。模式 2 除了向后根据后续帧来工作外, 是相同的概念。模式 3 是插值模式, 其结合了来自前面的和后续的帧的信息。

除了这三种模式, MPEG-4 还支持用于 B 帧的另一种插值运动向量预测模式: 直接模式预测, 其使用来自后续 P 帧的运动向量, 加上一个增量 ( $\delta$ ) 值 (如果来自共处的 P 宏块的运动向量被分成  $8 \times 8$  的模式——对于  $16 \times 16$  宏块产生 4 个运动向量——那么该增量被应用于 B 帧中的所有 4 个独立的运动向量)。后续 P 帧的运动向量指向前面 P 帧或 I 帧。一个比例被用于对后续 P 帧的运动向量进行加权。该比例是当前的 B 帧相对于后续的 P 帧和前面的 P (或 I) 帧的相对时间位置。

图 1 是根据现有技术的帧的时间线和 MPEG-4 直接模式运动向量。MPEG-4 直接模式 (模式 4) 的概念是指在每个居间的 B 帧中的宏块的运动可能接近被用于编码在随后 P 帧中的相同位置的运动。增量被用于对从后续 P 帧的相应运动向量 (MV) 103 导出的比例运动向量进行小的校正。图 1 显示的是给予每个中间 B 帧 104a、104b 的运动向量 101 和 102 的比例加权, 其作为前面 P 帧或 I 帧 105 和下一个 P 帧 106 之间的“时间距离”的函数。分配给相应的中间 B 帧 104a、104b 的运动向量 101 和 102 等于分配的权重值 (分别是  $1/3$  和  $2/3$ ) 乘以下一个 P 帧的运动向量 103, 加上增量值。

关于 MPEG-2, B 帧的所有预测模式在编码中被测试, 并且被比较以找到每个宏块的最佳预测。如果没有好的预测, 那么宏块被单独编码成“I” (“内部”) 宏块。编码模式被选为前向 (模式 1)、后向 (模式 2) 和双向 (模式 3) 编码中的最佳模式, 或者选为内部编码。关于 MPEG-4, 不允许选择内部编码。相反, 直接模式成为第四种选择。同样, 基于某种最佳匹配准则选择最优的编码模式。在基于 MPEG-2 和 MPEG-4 的软件编码器中, 使用 DC 匹配 (绝对差值的和, 或“SAD”) 来确定最佳匹配。

在编码的数据比特流中连续 B 帧的数量是由 MPEG 中的参数值“M”确定的。M 减 1 是每个 P 帧和下一个 P 帧 (或 I 帧) 之间的 B

帧的数量。因此，如果  $M=3$ ，则每个 P 帧（或 I 帧）之间就有 2 个 B 帧，如图 1 所示。在  $M$  值的限定方面（因而也限定了连续 B 帧的数量），主要的限制在于 P 帧（或 I 帧）之间的运动变化量变大。更多的 B 帧意味着 P 帧（或 I 帧）之间的时间更长。因此，运动向量的效率和编码范围限制构成了中间 B 帧数量的极限。

还很重要的是要注意到 P 帧携带“变化能量”和移动的图像流一起前进，因为每个解码 P 帧被用作预测下一个后续 P 帧的起始点。但是，B 帧在使用之后被丢弃。因此，任何用于创建 B 帧的比特仅被用于那个帧，而且与 P 帧不同，它们不提供帮助后续帧解码的校正。

## 发明内容

本发明涉及一种用于在视频图像压缩系统中提高一个或多个预测帧图像质量的方法、系统和计算机程序，其中每个帧包括多个像素。

一个方面，本发明包括确定双向预测帧的每个像素的值作为一个将一双向预测帧序列包围起来的非双向预测帧中的对应像素值的加权比例。在一个实施例中，该加权比例是进行包围的非双向预测帧之间的距离的函数。在另一个实施例中，加权比例是进行包围的非双向预测帧之间的距离和进行包围的非双向预测帧的等平均值的混合函数。

本发明的另一个方面，像素值的插值表示表示在线性空间中，或者在其他不同于原始非线性表示的优化的非线性空间中。

本发明的其他方面包括系统、计算机程序和方法，其包含：

- 具有包括图像区域的可参考帧序列的视频图像压缩系统，其中至少一个预测帧的至少一个图像区域参考两个或多个可参考帧被编码。
- 具有包括图像区域的可参考帧序列的视频图像压缩系统，其中至少一个预测帧的至少一个图像区域以显示顺序参考一个或多个可参考帧被编码，其中至少一个这样的可参考帧不是在显示顺序中最靠近至少一个预测帧的前面的可参考帧。
- 具有包括宏块的可参考帧序列的视频图像压缩系统，其中在至少一个预测帧内的至少一个宏块通过来自两个或多个可参考帧的插值而被编码。



- 具有包括图像区域的可参考帧和双向预测帧序列的视频图像压缩系统，其中至少一个双向预测帧的至少一个图像区域被编码以包括至少两个以上的运动向量，每个这样的运动向量参考在至少一个可参考帧中的对应的图像区域。
- 具有包括图像区域的可参考帧序列的视频图像压缩系统，其中至少一个预测帧的至少一个图像区域被编码以包括至少两个运动向量，每个这样的运动向量参考在可参考帧中的对应的图像区域，其中这样的至少一个预测帧的每个这样的图像区域通过来自两个或多个可参考帧的插值而被编码。
- 具有包括图像区域的可参考帧和双向预测帧序列的视频图像压缩系统，其中至少一个双向预测帧的至少一个图像区域被编码作为来自两个或多个可参考帧的选定图像区域的不相等加权。
- 具有包括图像区域的可参考帧和双向预测帧序列的视频图像压缩系统，其中至少一个双向预测帧的至少一个图像区域通过来自两个或多个可参考帧的插值而被编码，其中两个或多个可参考帧中的至少一个在显示顺序中被至少一个居间的可参考帧与双向预测帧隔开，且其中这样的至少一个图像区域被编码作为这样的至少两个或多个可参考帧的选定图像区域的不相等加权。
- 具有包括图像区域的可参考帧和双向预测帧序列的视频图像压缩系统，其中至少一个双向预测帧的至少一个图像区域通过来自两个或多个可参考帧的插值而被编码，其中两个或多个可参考帧中的至少一个在显示顺序中被至少一个居间的后续可参考帧与双向预测帧隔开。
- 具有包括图像区域的可参考帧和双向预测帧序列的视频图像压缩系统，其中至少一个双向预测帧的至少一个图像区域被编码作为来自两个或多个可参考帧的选定图像区域的不相等加权。
- 具有预测帧和双向预测帧序列的视频图像压缩系统，每个帧包括排列在宏块中的像素值，其中双向预测帧中的至少一个宏块

是使用基于来自两个或多个预测帧的运动向量的直接模式预测确定的。

- 具有可参考帧和双向预测帧序列的视频图像压缩系统，每个帧包括排列在宏块中的像素值，其中双向预测帧中的至少一个宏块是使用基于来自显示顺序中的一个或多个预测帧的运动向量的直接模式预测确定的，其中这样的—个或多个预测帧中的至少一个在显示顺序中是在双向预测帧前面。
- 具有可参考帧和双向预测帧序列的视频图像压缩系统，每个帧包括排列在宏块中的像素值，其中双向预测帧中的至少一个宏块是使用基于来自一个或多个预测帧的运动向量的直接模式预测确定的，其中这样的—个或多个预测帧的至少一个在显示顺序中是在双向预测帧之后而且被至少一个居间的可参考帧与双向预测帧隔开。
- 具有帧序列的视频图像压缩系统，每个帧包括具有 DC 值的多个图像区域，每个这样的图像区域包括的每个像素具有 AC 像素值，其中至少一个帧的至少一个图像区域的 DC 值和 AC 像素值中的至少一个被确定为来自至少一个其他帧的对应的各自的 DC 值和 AC 像素值的加权插值。
- 具有可参考帧序列的视频图像压缩系统，每个帧包括具有 DC 值的多个图像区域，每个这样的图像区域包括的每个像素具有 AC 像素值，其中至少一个预测帧的至少一个图像区域的 DC 值和 AC 像素值中的至少一个是从两个或多个可参考帧的对应的各自的 DC 值和 AC 像素值插值得到的。
- 改进视频图像压缩系统中的两个或多个双向预测中间帧的序列的图像质量，每个帧包括具有 DC 值的多个图像区域，每个这样的图像区域包括的每个像素具有 AC 像素值，包括下列中的至少一个：确定双向预测中间帧的每个图像区域的 AC 像素值为包围双向预测中间帧序列的可参考帧中的对应 AC 像素值的第一加权比例；和确定这种双向预测中间帧的每个图像区域的 DC 值为包围双向预测中间帧序列的可参考帧中的对应 DC 值的第二加权比例。具有帧序列的视频图像压缩系统，所述帧

包括具有初始表示的多个像素，其中至少一个帧的像素是从至少两个其他帧的对应像素插值得到的，其中该至少两个其他帧的这种对应的像素在变换到不同的表示时被插值，并导致插值的像素变换回初始的表示。

- 在具有可参考帧和双向预测帧序列的视频图像压缩系统中，动态地确定具有可变数量的双向预测帧的这种帧的编码模式，包括：选择一个初始序列，其始于可参考帧，具有至少一个紧跟的后续双向预测帧，并以可参考帧结束；把可参考帧加到该初始序列的结尾以创建一个测试序列；对照选定的评估标准评估该测试序列；对于评估该测试序列的每个满足步骤，在加入的可参考帧前面插入双向帧并且重复评估步骤；且如果评估该测试序列不满足，那么接受前面的测试序列作为当前的编码模式。
- 具有可参考帧序列的视频图像压缩系统，可参考帧被至少一个双向预测帧隔开，其中这种双向预测帧的数量在这种序列中是变化的，且其中至少一个这种双向预测帧的至少一个图像区域是使用对应于至少两个可参考帧的像素值的不相等加权确定的。
- 具有帧序列的视频图像压缩系统，所述帧由编码器编码且由解码器解码，其中至少一个帧的至少一个图像区域是基于两个或多个其他帧的加权插值，这种加权插值基于编码器和解码器可用的至少一组权值，其中对选定的这种至少一组权值之一的指定从编码器传送到解码器以选择一个或多个当前有效的权值。
- 具有帧序列的视频图像压缩系统，所述帧由编码器编码且由解码器解码，其中至少一个帧的至少一个图像区域是基于两个或多个其他帧的加权插值，这种加权插值基于至少一组权值，其中至少一组权值被下载到解码器，且此后对选定的这种至少一组权值之一的指定从编码器传送到解码器以选择一个或多个当前有效的权值。
- 具有可参考帧序列的视频图像压缩系统，可参考帧由编码器编码且由解码器解码，其中在可参考帧序列中的预测帧从编码器

传送到解码器，其传送顺序不同于这种预测帧在解码之后的显示顺序。

- 具有可参考帧序列的视频图像压缩系统，可参考帧包括排列在图像区域中的像素，其中至少一个预测帧的至少一个图像区域是参考两个或多个可参考帧被编码的，其中每个这样的图像区域是使用对应于这样的两个或多个可参考帧的像素值的不相等加权确定的。
- 具有预测帧、双向预测帧和内部帧序列的视频图像压缩系统，每个帧包括图像区域，其中从锐化和柔化滤波器组中选择的至少一个滤波器在预测帧或双向预测帧的运动向量补偿预测的过程中被用于这种预测帧或双向预测帧的至少一个图像区域。

附图和下面的描述阐明了本发明的一个或多个实施例的细节。根据说明书、附图和权利要求，本发明的其他特征、目标和优点是明显的。

## 附图说明

图 1 是根据现有技术的帧的时间线和 MPEG-4 直接模式运动向量。

图 2 是根据本发明的这个方面的帧的时间线和比例像素加权值。

图 3 是根据本发明的这个方面的帧的时间线和混合的、比例的和相等的像素加权值。

图 4 是一个流程图，显示了作为可以被计算机实现的方法的本发明的一个说明性实施例。

图 5 显示当前 P 帧参考前两个 P 帧和一个前面 I 帧的多个先前参考的例子的图。

图 6A 是典型的现有技术 MPEG-2 编码模式图，显示了进行包围的 I 帧和/或 P 帧之间的恒定数量的 B 帧。

图 6B 是理论上可能的现有技术 MPEG-4 视频编码模式图，显示了进行包围的 I 帧和/或 P 帧之间的变化数量的 B 帧，以及 I 帧之间的变化的距离。

图 7 是一个编码模式图。

图 8 是一个流程图，显示了插值方法的一个实施例，DC 插值不同

于 AC 插值。

图 9 是一个流程图，显示了使用替换表示的亮度像素的插值方法的一个实施例。

图 10 是一个流程图，显示了使用替换表示的色度像素的插值方法的一个实施例。

图 11 是显示两个 P 帧中的每个帧的独特运动向量区域尺寸的图。

图 12 是显示 P 帧和 B 帧序列的图，带有 B 帧的插值权重，其被确定为与一个 2 远离 (2-away) 后续 P 帧的距离的函数，该 P 帧参考一个 1 远离 (1-away) 后续 P 帧。

图 13 是显示 P 帧和 B 帧序列的图，带有 B 帧的插值权重，其被确定为与一个 1 远离后续 P 帧的距离的函数，该 P 帧参考一个 1 远离前面 P 帧。

图 14 是显示 P 帧和 B 帧序列的图，其中后续 P 帧具有参考前面 P 帧的多个运动向量。

图 15 是显示 P 帧和 B 帧序列的图，其中最近的后续 P 帧具有一个参考前面的一个 P 帧的运动向量，且下一个最近的后续 P 帧具有多个参考前面 P 帧的运动向量。

图 16 是显示 P 帧和 B 帧序列的图，其中最近的前面 P 帧具有一个参考前面的一个 P 帧的运动向量。

图 17 是显示 P 帧和 B 帧序列的图，其中最近的前面 P 帧具有两个参考前面 P 帧的运动向量。

图 18 是显示 P 帧和 B 帧序列的图，其中最近的前面 P 帧具有一个参考前面 P 帧的运动向量。

图 19 是显示三个 P 帧 P1、P2、P3 的帧序列，其中 P3 使用有两个运动向量的插值参考，P1 和 P2 各一个运动向量。

图 20 显示四个 P 帧 P1、P2、P3 和 P4 的帧序列，其中 P4 使用有三个运动向量的插值参考，P1、P2 和 P3 各一个运动向量。

图 21 是显示 P 帧和 B 帧序列的图，其中不同的 P 帧具有参考前面不同的 P 帧的一个或多个的运动向量，并显示分配给特定 B 帧的各个前向和后向参考的不同权重。

图 22 是显示 P 帧和 B 帧序列的图，其中 P 帧的比特流顺序不同于

显示顺序。

图 23 是显示具有分配的权重的 P 帧和 B 帧序列的图。

图 24 是在帧内的物体的位置时间图。

在不同附图中的相同参考符号表示相同的元素。

## 具体实施方式

### 概述

本发明的一个方面是基于这样的认识：常见的做法是使用 M 值为 3，其在每个 P(或 I)帧之间提供了两个 B 帧。但是 M=2 和 M=4 或更高都是有用的。特别重要的是要注意到 M 的值（B 帧的数量加 1）也具有和帧速率的自然关系。在每秒 24 帧（fps），即电影的播放率，帧之间 1/24 秒的时间距离可以导致帧到帧的实质图像变化。但是在 60 fps、72 fps 或更高的帧速率，相邻帧之间的时间距离相应地减少。结果是随着帧速率的增加，越大的 B 帧数量（即 M 值越大）在压缩效率中变得有用和有益。

本发明的另一个方面是基于这样的认识：MPEG-2 和 MPEG-4 视频压缩都利用过于简单的插值方法。例如，对于模式 3，帧的每个宏块的双向预测是后续和前面的帧宏块的等平均，其被两个对应的运动向量取代。这个等平均适合 M=2（即单个的中间 B 帧），因为该 B 帧到前面 P(或 I)帧和后续 P(或 I)帧的时间相等。但是，对于更大的 M 值，使用等加权只能对中央对称的 B 帧（即如果 M=4, 6, 8 等等的中间帧）进行最佳预测。类似地，在 MPEG-4 直接模式 4，即使运动向量是比例加权的，每个中间 B 帧的预测像素值是前面 P(或 I)帧和后续 P 帧的对应像素的等比例。

因此，对于 M>2，把合适的比例权重应用于每个 B 帧的预测像素值是一个进步。当前 B 帧的每个像素的比例权重对应于当前 B 帧相对于前面和后续 P(或 I)帧的相对位置。因此，如果 M=3，第一个 B 帧就使用前面帧的对应像素值（调整的运动向量）的 2/3，和后续帧的对应像素值（调整的运动向量）的 1/3。

图 2 是根据本发明的这个方面的帧的时间线和比例像素加权值。每个中间 B 帧 201a、201b 的每个宏块内的像素值被加权成前面的 P 或

I 帧 A 和下一个 P 或 I 帧 B 之间的“距离”的函数，离 P 或 I 帧越近，加权值越大。也就是，双向预测 B 帧的每个像素值是进行包围的非双向预测帧 A 和 B 的对应像素值的加权组合。在这个例子中，对于  $M=3$ ，第一个 B 帧 201a 的加权等于  $2/3A + 1/3B$ ，第二个 B 帧 201b 的加权等于  $1/3A + 2/3B$ 。还显示的是，在常规的 MPEG 系统下被分配的等平均加权；每个 B 帧 201a、201b 的 MPEG-1、2 和 MPEG-4 加权等于  $(A + B) / 2$ 。

### 扩展动态范围和对比范围的应用

如果  $M$  大于 2，中间 B 帧中的像素值的比例加权将在许多情况下将会改进双向（模式 3）和直接（MPEG-4 模式 4 的有效性）编码。示例包括常见的电影和视频编辑效果，例如渐弱和交叉渐隐。这些类型的视频效果是 MPEG-2 和 MPEG-4 的有问题的编码示例，这是由于使用简单 DC 匹配算法，且通常使用的  $M = 3$ （即两个中间 B 帧），导致 B 帧的相等比例。这些情况下的编码通过根据本发明使用比例 B 帧插值得以改进。

比例 B 帧插值对于扩展动态和对比范围的编码效率的提高也有直接的应用。在图像编码中的常见事件是亮度的变化，这在当物体逐渐移进（或移出）阴影（软阴影边界）时发生。如果对数的编码表示被用于亮度（例如，由对数亮度  $Y$  具体表示的亮度），那么采光亮度变化将是 DC 偏移量变化。如果采光亮度减半，那么像素值将减少相等的量。因此，为了编码这个变化，应该发现一个 AC 匹配，且编码的 DC 差值应该应用于该区域。这种被编码进 P 帧的 DC 差值也应该在每个居间的 B 帧中成比例地应用（参见共同待决的美国专利申请 09/905039 号，标题为“Method and System for Improving Compressed Image Chroma Information”，其被转让给本发明的受让人，因而在在此通过引用将其并入，用于关于对数编码表示的额外信息）。

除了亮度中的变化，比例 B 帧插值也有益于对比度的变化。例如，随着飞机移向观众，移出云或薄雾，其对比度逐渐增加。这个对比度增加将被表示成在对应 P 帧编码宏块中的 DCT 的 AC 系数中的增加的幅度。同样，在居间的 B 帧中的对比度变化将由比例插值最接近地近

似，因此提高了编码效率。

随着帧速率变大和随着 M 值的增加，使用比例 B 帧插值提高动态范围和对比度编码效率变得日渐重要。

### 把高M值应用于时间分层

利用本发明的实施例允许增加 M 的值，从而增加了在进行包围的 P 帧和/或 I 帧之间的 B 帧的数量，同时保持或增加了编码效率。这种用法有益于许多应用，包括时间分层。例如，在题为“Temporal and Resolution Layering for Advanced Television”的美国专利 5988863 号中（其被转让给本发明的受让人，在此通过引用并入），注意到 B 帧是用于分层的时间（帧）速率的合适机制。这种速率的灵活性与可用的连续 B 帧的数量有关。例如，单个 B 帧（M=2）可以支持 72 fps 的流内的 36 fps 解码的时间层或 60 fps 的流内的 30 fps 解码的时间层。三个 B 帧（M=4）可以支持 72 fps 流内的 36 fps 和 18 fps 解码的时间层，以及 60 fps 流内的 30 fps 和 15 fps 解码的时间层。在 120 fps 流内使用 M=10 可以支持 12 fps、24 fps 和 60 fps 解码的时间层。M=4 也可以用于 144 fps 流，以提供在 72 fps 和 36 fps 解码的时间层。

作为对每个第 N 帧的改进，120 fps 或 72 fps 的多个帧可以被解码和按比例地混合，如在共同待决的美国专利申请 09/545233 号中描述的，该专利标题为“Enhancements to Temporal and Resolution Layering”（其被转让给本发明的受让人，在此通过引用并入），以改进 24 fps 结果的运动模糊特征。

即使更高的帧速率可以使用在共同待决的美国专利申请 09/435277 号中所描述的方法合成，该专利申请题为“System and Method for Motion Compensation and Frame Rate Conversion”（其被转让给本发明的受让人，在此通过引用并入）。例如，通过运动补偿帧速率转换可以利用 72 fps 摄像底片（original）来创建 288 帧每秒的有效帧速率。使用 M=12，可以得到 48 fps 和 24 fps 的帧速率以及其他有用的速率，例如 144 fps、96 fps 和 32 fps（当然，底片为 72 fps）。使用这种方法的帧速率转换不需要是整倍数的。例如，120 fps 的有效速率可以根据 72 fps 的源创建，然后作为 60 fps 和 24 fps 速率的源（使用



M = 10)。

因此，时间分层有益于优化 B 帧插值的性能。上面所述的比例 B 帧插值使更多数量的连续 B 帧功能更高效，因此能够实现这些好处。

### 混合的B帧插值比例

在传统的系统中使用等平均加权作为 B 帧像素值的运动补偿模式预测器的一个原因是特定 B 帧前面或者后面的 P(或 I)帧可能是有噪声的，因此表示不理想的匹配。等混合在插值运动补偿块中会优化噪声的减少。存在使用量化的 DCT 函数来编码的差分残差 (difference residual)。当然，来自运动补偿比例的匹配越好，所需的差分残差比特越少，最后所得到的图像质量越高。

在有物体移近和移出阴影或者薄雾的情况下，M > 2 时的真正的比例提供更好的预测。但是，当光和对比度未发生变化时，等加权可被证明是更好的预测值，因为沿着运动向量向前移动宏块的误差将和来自后向位移块的误差平均，因此把每个误差减半。即使如此，更可能的是，较近的 P(或 I)帧的 B 帧宏块与较远的 P(或 I)帧相比与那个帧更相关。

因此，在某些情况下，例如区域的对比度或者亮度改变，理想做法是利用 B 帧宏块像素加权的真正比例(对于亮度和颜色)，如上所述。在其他情况下，可能更优的做法是利用等比例，如在 MPEG-2 和 MPEG-4 中。

本发明的另一方面利用这两个比例技术的混合(等平均和帧一距离比例)用于 B 帧像素插值。例如，对于 M = 3 的情况，1/3 和 2/3 比例的 3/4 可以与等平均的 1/4 混合，导致两个比例是 3/8 和 5/8。这个技术可以使用“混合因子” F 来概括。

加权 = F · (帧距离比例权重) + (1 - F) · 等平均权重

混合因子 F 的有用范围是从 1 到 0，1 表示完全比例插值，0 表示完全等平均(也可能使用相反的赋值)。

图 3 是根据本发明的这个方面的帧的时间线和混合的、比例的和相等的像素加权值。每个中间 B 帧 301a 和 301b 的每个宏块的像素值被加权为前面的 P 或者 I 帧 A 和下一个 P 或者 I 帧 B 之间的“时间距

离”的函数，以及 A 和 B 的等平均的函数。在这个例子中，对于  $M=3$  和混合因子  $F=3/4$ ，第一个 B 帧 301a 的混合加权等于  $5/8A+3/8B$ （也就是比例加权  $2/3A+1/3B$  的  $3/4$  加上等平均加权  $(A+B)/2$  的  $1/4$ ）。类似地，第二个 B 帧 301b 的加权等于  $3/8A+5/8B$ 。

混合因子 F 的值可以为全部的编码整体设定，或者为每个图像组 (GOP)、一个范围的 B 帧、每个 B 帧或者 B 帧内的每个区域设定（包括，例如细化到每个宏块，或者对于以  $8\times 8$  模式使用 P 向量的 MPEG-4 直接模式，甚至是单个的  $8\times 8$  运动块）。

为了节省比特并反映这样的事实即混合比例通常没有重要到要和每个宏块一起传送，混合的优化使用应该与要压缩的图像类型有关。例如，对于正在淡出、渐隐的图像或者整体采光或对比度正在逐渐改变的图像，混合因子 F 接近或等于 1（即选择比例插值）一般是最优的。对于没有这种采光或对比度变化的连续图像，那么较低的混合因子值例如  $2/3$ 、 $1/2$  或  $1/3$  可能形成最优的选择，从而保留比例插值的某些好处以及等平均插值的某些好处。所有在 0 到 1 范围内的混合因子值一般是有用的，因为对于任何给定的 B 帧，在这个范围内有一个特定值证明是优化的。

对于宽动态范围和宽对比度范围图像，混合因子可以依赖局部区域特征区域地确定。但是，通常宽范围的采光和对比度建议混合因子值支持完全比例插值，而不是等平均插值。

优化的混合因子通常是由经验确定的，虽然对于特定类型场景的经验可以被用于创建按场景类型的混合因子表格。例如，图像变化特征的确定可以被用于为帧或者区域选择混合比例。替换地，可以使用许多的候选混合因子（用于整个帧或者区域）编码 B 帧，评估每个混合因子以优化图像质量（例如，由最高的信噪比即 SNR 确定的）和最低的比特数。这些候选的评估然后可以被用于为混合比例选择最佳值。也可以使用图像变化特征和编码的质量/效率的组合。

靠近 B 帧序列中间的 B 帧或者来自低值 M 的 B 帧不会被比例插值过多影响，因为计算的比例已经接近等平均。但是，对于较高的 M 值，B 帧位置的极值可以被混合因子的选择显著影响。注意对于这些极限位置混合因子可以不同，这些极限位置比更中央的位置（其从偏

离平均获得很少好处或者没有获得好处)更多地利用该平均,因为它们已经具有邻近 P (或者 I) 帧的高比例。例如,如果  $M=5$ , 第一个和第四个 B 帧可能使用的混合因子  $F$  混合了更多的等平均,但第二个和第三个 B 帧可能使用严格的  $2/5$  和  $3/5$  的等平均比例。如果比例到平均混合因子 (proportion-to-average blend factor) 作为序列中 B 帧位置的函数变化,那么混合因子的变化值可以在压缩的比特流中传送或者作为辅助信息传给解码器。

如果要求静态的一般混合因子 (由于缺少传送值的方法),那么值  $2/3$  通常接近最优的,且可以在编码器和解码器中被选择作为 B 帧插值的静态值。例如,使用  $F=2/3$  的混合因子,对于  $M=3$  的连续帧,比例将是  $7/18$  ( $7/18=2/3*1/3+1/3*1/2$ ) 和  $11/18$  ( $11/18=2/3*2/3+1/3*1/2$ )。

### 线性插值

视频帧像素值一般存储在特定表示中,特定表示把原始的图像信息映射到数字值。这种映射可能导致线性的或者非线性的表示。例如,在压缩中使用的亮度值是非线性的。各种形式的非线性表示的使用包括对数、指数 (各种幂) 以及黑色校正的指数 (通常用于视频信号)。

在狭窄的动态范围上或者对于邻近区域的插值,非线性表示是可接受的,因为这些邻近的插值表示分段的线性插值。因此,亮度的小变化可以由线性插值合理地近似。但是,对于亮度的广泛变化,例如发生在宽动态范围和宽对比度范围图像中的,把非线性信号处理为线性的将是不准确的。即使对于正常的对比度范围图像,线性淡出和交叉渐隐可以被线性插值降低效果。有些淡出和交叉渐隐利用非线性的淡出和渐隐率,增加了进一步的复杂性。

对于使用比例混合,或者甚至是简单的比例或等平均插值的额外改进是在线性空间里表示的像素值上进行这种插值,或者在不同于原始非线性亮度表示的其他优化的非线性空间中进行这种插值。

这可以被完成,例如通过首先把两个非线性亮度信号 (来自前面和后续 P (或者 I) 帧) 转换成线性表示,或者不同的非线性表示。然后应用比例混合,在应用逆变换之后,在图像的原始非线性亮度表示中产生混合的结果。但是,比例函数已经在亮度信号的更优表示上执

行。

当颜色正在褪色或者变得更加饱和时，如与在薄雾和阴天中的变化关联的对比度变化中发生的，除了亮度之外，把这个线性或非线性变换有利地应用于颜色（色度）值也是有用的。

### 示例性实施例

图 4 是一个流程图，显示了本发明的一个示例性实施例，其作为可以被计算机实现的一个方法：

步骤 400：在视频图像压缩系统中，对于计算 B 帧的直接和插值模式，使用（1）帧距离比例，或者（2）等权重和帧距离比例的混合来确定插值值以应用于两个或多个双向预测中间帧的输入序列的每个像素，插值值来源于括起这种从源（例如视频图像流）输入的序列的至少两个非双向预测的帧。

步骤 401：相对于图像单元（例如图像组（GOP）、帧序列、场景、帧、帧内的区域、宏块、DCT 块或者类似有用的分组或像素的选择）优化插值值。插值值可以为整个编码期间静态地设定，或者为每个图像单元动态地设定。

步骤 402：进一步相对于场景类型或编码简单性优化插值值。例如，插值值可以被设定：静态地（例如 2/3 比例和 1/3 等平均）；比例地，为接近等平均的帧，但在邻近 P（或 I）帧附近要混合等平均；动态地，基于整体场景特征，例如淡出和交叉渐隐；动态地（和局部地），基于局部图像区域特征，例如局部的对比度和局部的动态范围；或者动态地（和局部地），基于编码性能（例如最高编码 SNR）和生成的最少编码比特。

步骤 403：传送合适的比例量给解码器，如果不是静态确定的。

步骤 404：可选地，为每个帧转换亮度信息为线性的或者替换的非线性表示，并把这个替换传送给解码器，如果不是静态确定的。

步骤 405：使用确定的插值值确定比例像素值。

步骤 406：如果有必要（由于步骤 404），再转换成原始表示。

### 扩展的P帧参考

如上所述，在现有技术 MPEG-1、2 和 4 压缩方法中，P 帧参考前

面的 P 或者 I 帧，且 B 帧参考最近的前面或后续的 P 和/或 I 帧。同样的技术被用于 H.261 和 H.263 运动补偿的 DCT 压缩标准，其包含了低比特率压缩技术。

正在开发的 H.263++ 和 H.26L 标准中，B 帧参考被扩展到指向没有直接包围当前帧的 P 或者 I 帧。也就是说，B 帧内的宏块可以指向前面 P 帧之前的一个 P 或者 I 帧，或者指向后续 P 帧之后的一个 P 或者 I 帧。由于每个宏块有一个或多个比特，跳过前面或后续 P 帧可以简单地表示。概念上，使用前面 P 帧用于 B 中的参考只要求存储空间。对于 H.263++ 或 H.26L 的低比特率编码使用，这是少量的额外存储。对于后续的 P 帧参考，必须相对于 B 帧编码修改该 P 帧编码顺序，以致于必须在居间的 B 帧之前解码未来的 P 帧（或者可能是 I 帧）。因此，编码顺序对于后续的 P 帧参考也是个问题。

P 帧和 B 帧类型之间的主要差别是：（1）B 帧可以被双向参考（最多每个宏块两个运动向量）；（2）在使用之后 B 帧被丢弃（这也意味着在解码过程中可以跳过它们以提供时间分层）；和（3）P 帧被用作“跳板”，一个到下一个，因为每个 P 帧必须被解码以用作每个后续 P 帧的参考。

作为本发明的另一个方面，根据一个或多个的前面的 P 或者 I 帧（不包括每个 P 帧只参考最近的前面 P 帧或者 I 帧的情况）解码 P 帧（与 B 帧相对）。因此，例如，每个宏块的两个或者多个运动向量可以被用于当前的 P 帧，所有运动向量在时间上指向后面（也就是指向一个或多个先前解码的帧）。这样的 P 帧仍然保持“跳板”特点。图 5 是一个显示多个先前参考的例子，由当前 P 帧 500 参考两个前面 P 帧 502、504 和前面 I 帧 506。

此外，在这样的 P 帧参考中可以应用宏块插值的概念，如上所述。因此，除了表示对一个以上的前面 P 或 I 帧的单个参考之外，也可以为每个这样的帧参考使用一个运动向量，以混合多个前面 P 或 I 帧的比例。例如，可以应用上面所述的使用具有两个帧参考的 B 帧插值模式的技术以允许 P 帧中的任何宏块利用两个运动向量来参考前面两个 P 帧或者一个前面的 P 帧和一个前面的 I 帧。这个技术在两个运动向量之间插值，但在时间上不是双向的（和 B 帧插值情况一样），因为两个

运动向量时间上都指向后面。存储开销减少到一个点，在这点，为这种并发参考在存储器中保存前面的多个 P 或 I 帧是相当实用的。

在应用这种 P 帧插值时，肯定需要选择前面两个或多个 P 帧（以及可选地，一个前面的 I 帧）的各种有用比例并通知解码器。特别是帧的等混合是有用的混合比例之一。例如，有前两个 P 帧作为参考，可以混合每个 P 帧的相等的  $1/2$ 。对于前三个 P 帧，可以使用  $1/3$  的相等混合。

两个 P 帧的另一个有用混合是  $2/3$  的最近的前面帧和  $1/3$  的最远的前面帧。对于前 3 个 P 帧，另一个有用混合是  $1/2$  的最近的前面帧， $1/3$  中间的前面帧和  $1/6$  的最远的前面帧。

在任何情况下，可以利用多个前面 P 帧（和可选地，一个 I 帧）的一组简单的有用混合，并将它们简单地从编码器通知到解码器。每当对优化图像单元的编码效率有用时，可以选择使用的特定混合比例。可以使用少量的比特选择许多的混合比例，只要适合想要的图像单元，这些比特可以被传送给解码器。

作为本发明的另一个方面，转变选择单个 P 帧参考从最近的前面 P（或 I）帧到更“远”的前面 P（或 I）帧也是有用的。在这个方式中，P 帧会利用每个宏块（或者可选地，在 MPEG-4 模式编码中的每个  $8 \times 8$  块）的单个运动向量，但是会利用一个或多个的比特表示该参考参考单个特定的前面帧。在这个模式中的 P 帧宏块不是插值的，但是会参考选定的前面帧，该帧从可能的两个、三个或者更多的前面 P（或 I）帧中选择用于参考。例如，2 比特编码可指定多至 4 个的前面帧中的一个作为单个选择帧。这个 2 比特编码可以在任何方便的图像单元被改变。

### 自适应的B帧数量

在 MPEG 编码中典型地使用 I、P、B 帧类型的固定模式。在 P 帧之间的 B 帧数量典型地是常数。例如，在 MPEG 编码中在 P（或 I）帧之间使用两个 B 帧是典型的。图 6A 是典型的现有技术 MPEG-2 编码模式图，显示了在进行包围的 I 帧 600 和/或 P 帧 602 之间的固定数量的 B 帧（即两个）。

MPEG-4 视频编码标准概念地允许在进行包围的 I 帧和/或 P 帧之间是变数量的 B 帧, 以及 I 帧之间变化的距离。图 6B 是理论上可能的现有技术 MPEG-4 视频编码模式图, 显示了在进行包围的 I 帧 600 和/或 P 帧 602 之间的变数量的 B 帧, 以及 I 帧 600 之间的变距离。

这个灵活的编码结构理论上可被用于通过将最有效的 B 帧和 P 帧编码类型匹配到运动图像帧来提高编码效率。虽然这个灵活性已经被特别地允许, 但是人们很少研究它, 而且在这种灵活的结构中实际确定 B 帧和 P 帧位置的机制也是未知的。

本发明的另一个方面在普通使用中把这里描述的概念应用于这个灵活的编码结构以及简单的固定编码模式。因此能够使用上面所述的方法来插值 B 帧, 而 P 帧可以参考一个以上的前面 P 帧或 I 帧并根据本说明书被插值。

特别地, B 帧内的宏块可以利用与固定结构一样有效的适于灵活编码结构的比例混合。当 B 帧参考与最靠近的进行包围的 P 帧或 I 帧相比更远的 P 帧或 I 帧时, 也可以利用比例混合。

类似地, 在这个灵活编码结构中, P 帧可以参考一个以上的前面的 P 帧或 I 帧, 与固定模式结构一样有效。此外, 当这些 P 帧参考一个以上的前面的 P 帧 (可选地, 加上一个 I 帧) 时, 混合比例可以被应用于这种 P 帧中的宏块。

#### (A) 在灵活的编码模式中确定位置

以下方法允许编码器优化帧编码模式以及利用混合比例的效率。对于选定范围的帧, 可以尝试许多候选的编码模式以确定最优的或者接近最优的 (相对于指定标准) 模式。图 7 是一个可以被检验的编码模式图。以 P 或者 I 帧结束的初始序列 700 是任意选定的, 并被用作增加的附加 P 帧和/或 B 帧的基础, 然后评估附加 P 帧和/或 B 帧 (如下所述)。在一个实施例中, 一个 P 帧被加到初始序列 700 以创建用于评估的第一个测试序列 702。如果评估满足, 那么插入一个居间的 B 帧以创建第二个测试序列 704。对于每个满足的评估, 插入附加 B 帧以创建逐渐变长的测试序列 706~712, 直到评估标准变得不满足。到那一点, 前面的编码序列被接受。然后重复这个过程, 使用先前接受

的编码序列的结尾 P 帧作为增加新 P 帧的起始点, 然后插入新的 B 帧。

最优的或者接近最优的编码模式可以基于各种评估标准选择, 通常涉及各种编码特征的权衡, 例如编码的图像质量和所需的编码比特数量。常用的评估标准包括被使用的最少比特数量 (在固定的量化参数测试中), 或者最好的信噪比 (在固定的比特率测试中), 或者两者的组合。

最小化一个绝对差值和 (SAD) 也是常见的, SAD 形成 DC 匹配的测量。如在共同待决的题为“Motion Estimation for Video Compression System”的美国专利申请 09/904192 号 (其被转让给本发明的受让人, 因而通过引用并入) 中描述的, AC 匹配标准也是特定候选匹配的质量的有用测量 (本专利申请也描述其他有用的优化)。因此可以检验在所有宏块的最佳匹配上累积的 AC 和 DC 匹配标准, 以确定每个候选编码模式的整体匹配质量。当这个 AC/DC 匹配技术和每个帧模式类型的编码比特数量的估计一起使用时, 它可以增进或者取代信噪比 (SNR) 和使用最少比特的测试。典型的是, 编码 B 帧宏块所使用的量化参数值 (QP) 比用于 P 帧的更高, 其影响在各种候选编码模式内使用的比特数和质量 (通常作为信噪比测量)。

#### (B) 在灵活编码模式中的混合比例优化

可选地, 对于根据上述方法确定的每个候选模式, 可以测试混合比例相对一个或多个标准的适合性 (例如最优的或者接近最优的混合比例)。这可以被完成, 例如通过测试最好质量 (最低 SNR) 和/或效率 (使用的最少比特)。为 P 帧中的每个宏块使用一个或多个前面的参考也可以用同样的方式确定, 测试每个候选的参考模式和混合比例以确定一组或多组合适的参考。

一旦为下一步骤 (图 7 中的步骤 700) 选择了编码模式, 就可以为各种候选编码模式测试后续步骤 (步骤 702-712)。在这个方法中, 可以确定运动图像序列的更高效的编码。因此, 可以如上面的子节 (A) 描述的那样优化/提高效率; 可以在每个测试的编码步骤应用混合优化。

### DC对AC插值

在图像编码的许多情况中, 例如当使用图像帧的对数表示时, 帧



像素值的上述插值将最优地编码亮度中的变化。但是，在替换的视频“ $\gamma$ 曲线”、线性和其他表示中，把不同的插值混合因子应用于像素的DC值而不是AC值，这通常被证明是有用的。图8是一个流程图，显示了一种与AC插值明显不同的具有DC插值的插值方法的一个实施例。对于来自第一和第二输入帧的选定图像区域（通常是DCT块或者宏块）802、802'，减去每个这种区域的平均像素值804、804'，因此在选定的区域中把DC值（即整个选定区域的平均值）806、806'和AC值（即保留的有符号像素值）808、808'分开。然后各自的DC值806、806'可以乘以插值权重810、810'，其不同于乘以AC（有符号的）像素值808、808'使用的814、814'。然后，最新插值的DC值812和最新插值的AC值816可以被合并成818，产生选定区域的新预测820。

和本发明中的其他插值一样，可以将合适的权重通知每个图像单元的解码器。可以在许多编码值之间选择少量的比特，以及选择像素值的AC对DC方面的独立插值。

### 线性&非线性插值

插值是线性的加权平均。由于插值运算是线性的，且由于每个图像帧中的像素值通常是以非线性形式（例如视频伽马表示或者对数表示）表示的，插值过程的进一步优化成为可能的。例如，特定帧序列的像素的插值，以及从AC值分离的DC值的插值，有时用线性的像素表示将是最优的或者接近最优的。但是，对于其他的帧序列，如果像素是被表示成对数值或者其他的像素表示，那么这种插值将是最优的或者接近最优的。此外，用于插值U和V（色度）信号分量的最优表示或接近最优的表示可不同于Y（亮度）信号分量的最优表示或接近最优的表示。因此，本发明的一个有用方面是作为插值程序的一部分，把像素表示转换成替换的表示。

图9是一个流程图，显示了使用替换表示的亮度像素插值的方法的一个实施例。开始于初始表示（例如视频伽马或者对数）中的亮度（Y）像素的一个区域或者块（步骤900），像素数据被变换成不同于初始表示的替换表示（例如线性的、对数的、视频伽马）（步骤902）。然后如上所述，变换的像素区域或者块被插值（步骤906），并变换回

初始表示（步骤 906）。结果得到插值的像素亮度值（步骤 908）。

图 10 是一个流程图，显示了使用替换表示的色度像素插值的方法的一个实施例。开始于初始表示（例如视频伽马或者对数）中的色度（U, V）像素的一个区域或者块（步骤 1000），像素数据被变换成不同于初始表示的替换表示（例如线性的、对数的、视频伽马）（步骤 1002）。然后如上所述，变换的像素区域或者块被插值（步骤 1006），并变换回初始表示（步骤 1006）。结果得到插值的像素色度值（步骤 1008）。

各种表示之间的变换可以根据题为“Method and System for Improving Compressed Image Chroma Information”的美国专利申请 09/905039 号的讲解进行，其被转让给本发明的受让人，因此通过引用并入。注意，通常使用简单的查找表进行替换的表示变换及其逆变换。

作为本发明的这个方面的一个变化，用于 AC 插值的替换（线性或非线性）表示空间可不同于用于 DC 插值的替换表示空间。

和插值权重一样，选择哪种替换插值表示用于每个亮度（Y）和色度（U 和 V）像素表示，可以使用每个选定图像单元的少量比特通知解码器。

### 每个宏块的运动向量的数量

在 MPEG-2 中，P 帧中的每个 16x16 宏块允许一个运动向量。在 B 帧中，MPEG-2 允许每个 16x16 宏块最多两个运动向量，对应于双向的插值模式。在 MPEG-4 视频编码中，P 帧中的每个 16x16 宏块允许多至 4 个运动向量，对应于每个 8x8 DCT 块一个运动向量。在 MPEG-4 B 帧中，当使用插值模式时，每个 16x16 宏块允许最多两个运动向量。如果后续的对 P 帧宏块被设置为具有 4 个运动向量的 8x8 模式，在 MPEG-4 直接模式中的单个运动向量增量可以导致 4 个独立的“隐式”运动向量。这是在换算时间距离（该 B 帧在时间上更靠近该 P 帧的前面 P 帧或 I 帧参考）之后，通过将 16x16 B 帧宏块中所携带的一个运动向量增量增加到来自随后 P 帧宏块中的对应的 4 个独立运动向量中的每一个来实现的。

本发明的一个方面包括增加每个图像区域（例如宏块）的运动向量数量的选择。例如，有时证明每个 B 帧宏块具有两个以上的运动向

量是有益的。这些可以通过参考附加的 P 帧或者 I 帧和在加权和中具有 3 个或更多的插值项被应用。也可以应用附加的运动向量以允许 B 帧宏块的 8x8 DCT 块的独立向量。而且，可以使用 4 个独立增量以通过把独立增量应用于后续 P 帧的 4 个 8x8 区域运动向量中的每一个来扩展直接模式的概念。

此外，可以使用 B 帧插值技术修改 P 帧以在插值模式中参考多于一个的前面帧，使用上面描述的 B 帧的两个插值项技术。这个技术可以容易地扩展到两个以上的前面 P 帧或者 I 帧，所得到的插值在加权和中具有 3 个或更多的项。

和本发明的其他方面(例如像素表示和 DC 对 AC 插值方法)一样，可以使用每个图像单元的少量比特把特定的加权和传递给解码器。

在应用本发明的这个方面时，8x8 像素 DCT 块和运动向量域之间的对应不需要像 MPEG-2 和 MPEG-4 中那样严格。例如，使用除了 16x16、16x8 (只在 MPEG-4 中的隔行扫描中使用)、8x8 以外的替换区域尺寸用于运动向量可能是有用的。这种替换可能包括任何数量的有用区域尺寸，例如 4x8、8x12、8x16、6x12、2x8、4x8、24x8、32x32、24x24、24x16、8x24、32x8、32x4 等等。使用少量这种有用的尺寸，几个比特可以把运动向量区域尺寸和 DCT 块尺寸之间的对应通知解码器。在使用传统的 8x8 DCT 块的系统中，一组到运动向量域的简单对应简化运动补偿中的处理是有用的。在 DCT 块尺寸不是 8x8 的系统中，那么可以在指定的运动向量域中实现更大的灵活性，如在题为“Enhanced Temporal and Resolution Layering in Advanced Television”的共同待决的美国专利申请 09/545233 号中所描述的，其被转让给本发明的受让人，因此通过引用将其并入。注意，运动向量区域边界不需要对应于 DCT 区域边界。实际上，以运动向量区域界限落入 DCT 块内(且不在其界限上)的方式定义运动向量区域经常是有用的。

扩展运动向量域的灵活性的概念也适用于本发明的插值方面。只要指定了每个像素和指向一个或多个参考帧的一个或多个运动向量之间的对应，就可以使用本发明的所有通用性把上述插值方法应用于有用运动向量的全部灵活性。甚至当使用 P 帧时，对于每个前面帧参考，对应于每个运动向量的区域的尺寸可以不同，而且当使用 B 帧时，对

于每个前面和未来的帧参考，对应于每个运动向量的区域尺寸可以不同。如果在应用本发明的改进插值方法时运动向量的区域尺寸不同，那么插值反映重叠的共同区域。运动向量参考的重叠共同区域可以被用作当分别插值 DC 和 AC 像素值时，在其上确定 DC 项的区域。

图 11 显示了两个 P 帧 1104、1106 各自的独特运动向量区域尺寸 1100 和 1102。在根据这个发明计算插值时，确定运动向量区域尺寸的联合 1108。联合 1108 定义了所有的被认为是具有分配的运动向量的区域。

因此，例如在将 B 帧 1112 的 4x4 DCT 区域向后插值到前面 P 帧 1104 时，联合 1108 内的 4x4 区域 1110 会使用对应于前面 P 帧中的 8x16 区域 1114 的运动向量。如果前向预测，那么联合 1108 内的 1110 区域会使用对应于下一 P 帧中的 4x16 区域 1115 的运动向量。类似地，联合 1108 内的区域 116 的向后插值会使用对应于 8x16 区域 1114 的运动向量，而前向预测相同的区域会使用对应于 12x16 区域 1117 的运动向量。

在本发明的一个实施例中，利用两个步骤来实现普遍（即非统一尺寸）运动向量的插值。第一步是确定运动向量共同区域，如关于图 11 所描述的。这为每个前面帧或者后续帧参考建立了像素和运动向量之间的对应（即每个指定像素区域尺寸的运动向量的数量）。第二步是利用合适的插值方法和对于像素的每个区域有效的插值因子。编码器的任务是确保最优的或接近最优的运动向量区域和插值方法被指定，且所有像素使它们的向量和插值方法完全被指定。对于运动向量的固定模式（例如为整个帧指定每个 32x8 块一个运动向量），具有单个指定的插值方法（例如为整个帧指定的到被参考帧的每个距离的固定比例混合）的情况，这非常简单。如果对运动向量区域尺寸做区域性的改变，且区域尺寸的不同依赖于参考的是哪个前面帧或者后续帧（例如对于最近的前面帧是 8x8 块，对于下一个最近的前面帧是 32x8 块），这个方法也能变得相当复杂。此外，可以在帧内区域性地指定插值方法。

编码时，编码器的工作是确定比特的最优或接近最优的用法以在运动向量区域形状和尺寸之间选择，和选择最优或接近最优的插值方

法。还需要一个决定以指定被参考帧的数量和距离。通过详尽地测试大量候选运动向量区域尺寸、要参考的候选帧和每个这样的运动向量区域的插值方法来确定这些指定，直到找到最优或接近最优的编码。最优（相对于选定的标准）可以通过在编码块之后找到最小的 SNR 或者在编码块之后找到固定量化参数（QP）的最少比特数，或者通过应用其他合适的测量来确定。

### 直接模式扩展

在 MPEG-4 中的 B 帧宏块中使用的传统直接模式，在运动向量编码中是高效的，利用简单共同增量提供了 8x8 块模式的好处。直接模式基于时间距离在对应的宏块位置处对后续 P 帧的每个对应的运动向量进行加权，该 P 帧参考前面 P 帧。例如，如果  $M=3$ （即两个居间的 B 帧），用简单线性插值，第一个 B 帧使用  $-2/3$  倍的后续 P 帧运动向量以确定相对于这种 P 帧的像素偏移，和  $1/3$  倍的后续 P 帧运动向量以确定相对于前面 P 帧的像素偏移。类似地，第二个 B 帧使用  $-1/3$  倍的相同 P 帧运动向量以确定相对于这种 P 帧的像素偏移，和  $2/3$  倍的后续 P 帧运动向量以确定相对于前面 P 帧的像素偏移。在直接模式中，小的增量被加到每个对应的运动向量。作为这个发明的另一个方面，这个概念可以扩展至指向一个或多个  $n$  远离 P 帧的 B 帧参考，这些 P 帧又参考一个或多个前面或后续 P 帧或 I 帧，通过考虑帧距离来确定帧比例分数（frame scale fraction）。

图 12 是显示 P 帧和 B 帧序列的图，其中 B 帧的插值权重被确定为到 2 远离后续 P 帧的距离的函数，该 P 帧参考 1 远离后续 P 帧。在这个说明性的例子中， $M=3$ ，表示在进行包围的 P 帧 1204、1206 之间有两个连续 B 帧 1200、1202。在这个例子中，在下一个最近的后续 P 帧 1208 中的每个共存的宏块可指向该居间的（即最近的）P 帧 1204，且最初的两个 B 帧 1200、1202 可参考下一个最近的后续 P 帧 1208 而不是最近的后续 P 帧 1204，如在传统 MPEG 中的。因而，对于第一个 B 帧 1200，帧比例分数  $5/3$  乘以来自下一个最近的后续 P 帧 1208 的运动向量  $mv$  会被用作相对于 P 帧 1208 的像素偏移，且第二个 B 帧 1202 会使用  $4/3$  乘以那个相同的运动向量的偏移。

如果被 B 帧参考的最近的后续 P 帧指向下一个最近的前面 P 帧，那么再次使用简单的帧距离以获得应用于运动向量的合适的帧比例分数。图 13 是显示 P 帧和 B 帧序列的图，其中 B 帧的插值权重被确定为到 1 远离后续 P 帧的距离的函数，该 P 帧参考 2 远离前面 P 帧。在这个说明性的例子中， $M=3$ ，B 帧 1300、1302 参考最近的后续 P 帧 1304，该 P 帧 1304 又参考 2 远离 P 帧 1306。因而，对于第一个 B 帧 1300，像素偏移分数是帧比例分数  $2/6$  乘以来自最近后续 P 帧 1304 的运动向量  $mv$ ，且第二个 B 帧 1302 会具有  $1/6$  乘以那个相同的运动向量的像素偏移，因为最近的后续 P 帧 1304 的运动向量指向 2 远离前面 P 帧 1306，该帧相距 6 个帧。

通常，在直接模式中 B 帧参考单个 P 帧的情况下，帧距离方法设定帧比例分数的分子等于该 B 帧到其参考的或者“目标”P 帧的帧距离，并设定分母等于从目标 P 帧到另一个被目标 P 帧参考的 P 帧的距离。对于从 B 帧到后续 P 帧进行的测量，帧比例分数的符号是负号，而对于从 B 帧到前面 P 帧进行的测量，帧比例分数的符号是正号。这种把帧距离或者帧比例分数应用于 P 帧运动向量的简单方法能够实现有效的直接模式编码。

此外，这个发明的另一个方面是允许直接模式应用于一个 P 帧的多个插值的运动向量参考。例如，如果从最近的和下一个最近的前面 P 帧插值一个 P 帧，那么根据本发明的这个方面的直接模式参考允许插值混合用于每个多参考直接模式 B 帧宏块。通常，一个 P 帧的两个或多个运动向量可以应用一个合适的帧比例分数。然后，两个或多个帧距离修正的运动向量可以和参考或指向那个 P 帧的每个 B 帧的对应插值权重一起使用（如下所述），以生成插值 B 帧宏块运动补偿。

图 14 是显示 P 帧和 B 帧序列的图，其中一个后续 P 帧具有多个参考前面 P 帧的运动向量。在这个例子中，B 帧 1400 参考后续 P 帧 P3。这个 P3 帧又具有两个运动向量  $mv1$  和  $mv2$ ，其参考对应的前面 P 帧 P2 和 P1。在这个例子中，可以使用两个加权项中的任一个或者这种加权项的组合在直接模式中对 B 帧 1400 的每个宏块进行插值。

B 帧 1400 的每个宏块可根据下列项构造成混合形式：

- 由  $mv1$  的帧比例分数  $1/3$  取代的帧 P2 的对应像素（这些像素

然后可能乘以某个比例权值  $i$ ) 加上由  $mv1$  的帧比例分数  $-2/3$  取代的帧  $P3$  的对应像素 (这些像素然后可能乘以某个比例权值  $j$ ); 以及

- 由  $mv2$  的帧比例分数  $2/3$  ( $4/6$ ) 取代的帧  $P1$  的对应像素 (这些像素然后可能乘以某个比例权值  $k$ ) 加上由  $mv2$  的帧比例分数  $-1/3$  ( $-2/6$ ) 取代的帧  $P3$  的对应像素 (这些像素然后可能乘以某个比例权值  $l$ )。

和所有的直接模式一样, 运动向量增量可以与  $mv1$  和  $mv2$  中的每一个一起使用。

根据本发明的这个方面, 如同多个前面  $P$  帧的情况一样, 使用相同的插值方法和运动向量帧比例分数应用,  $B$  帧中的直接模式预测宏块也能够参考多个后续的  $P$  帧。图 15 是显示  $P$  帧和  $B$  帧序列的图, 其中最近的后续  $P$  帧具有参考一个前面  $P$  帧的运动向量, 且下一个最近的后续  $P$  帧具有参考前面多个  $P$  帧的多个运动向量。在这个例子中,  $B$  帧 1500 参考两个后续  $P$  帧  $P2$  和  $P3$ 。帧  $P3$  具有两个运动向量  $mv1$  和  $mv2$ , 它们参考对应的前面  $P$  帧  $P2$  和  $P1$ 。帧  $P2$  有一个运动向量  $mv3$ , 其参考前面  $P$  帧  $P1$ 。在这个例子中, 使用 3 个加权项在直接模式中对  $B$  帧 1500 的每个宏块进行插值。假如这样的话, 运动向量帧比例分数可以大于 1 或小于 -1。

这种形式的直接模式  $B$  帧宏块插值的加权可以利用这里描述的插值的全部通用性。特别是, 为了获得各种图像单元的最佳性能 (例如质量对比特数), 可对每个权值或者权值的组合进行测试。用于这种改进直接模式的插值分数组可以每个图像单元用少量比特指定给解码器。

$B$  帧 1500 的每个宏块可根据下列项构造成混合形式:

- 由  $mv1$  的帧比例分数  $-5/3$  取代的帧  $P3$  的对应像素 (这些像素然后可能乘以某个比例权值  $i$ ) 加上由  $mv1$  的帧比例分数  $-2/3$  取代的帧  $P2$  的对应像素 (这些像素然后可能乘以某个比例权值  $j$ );
- 由  $mv2$  的帧比例分数  $-5/6$  取代的帧  $P3$  的对应像素 (这些像素然后可能乘以某个比例权值  $k$ ) 加上由  $mv2$  的帧比例分数  $1/6$

取代的帧 P1 的对应像素（这些像素然后可能乘以某个比例权值  $l$ ）；以及

- 由  $mv3$  的帧比例分数  $2/3$  取代的帧 P2 的对应像素（这些像素然后可能乘以某个比例权值  $m$ ）加上由  $mv3$  的帧比例分数  $1/3$  取代的帧 P1 的对应像素（这些像素然后可能乘以某个比例权值  $n$ ）。

和所有的直接模式一样，运动向量增量可以与  $mv1$ 、 $mv2$  和  $mv3$  中的每一个一起使用。

注意，当下一个最近的后续 P 帧参考了包围一候选 B 帧的最近的 P 帧时，经常出现特别有益的直接编码模式。

在 MPEG-4 中 B 帧的直接模式编码总是使用后续 P 帧的运动向量作为参考。根据本发明的另一个方面，B 帧参考前面 P 帧的共存宏块的运动向量也是可能的，这些运动向量有时将证明直接模式编码参考的有益选择。假如这样的话，当下一个最近的前面 P 帧被最近的前面 P 帧的运动向量参考时，运动向量帧比例分数将大于 1。图 16 显示了 P 帧和 B 帧序列，其中最近的前面 P 帧具有参考前面 P 帧的运动向量。在这个例子中，B 帧 1600 参考了 1 远离的前面 P 帧 P2。帧 P2 的运动向量  $mv$  参考下一个前面 P 帧 P1（相对于 B 帧 1600，2 远离）。图中显示了合适的帧比例分数。

如果根据多个向量和帧对最近的前面 P 帧进行插值，那么应用类似于结合图 14 所描述的那些方法以获得运动向量帧比例分数和插值权值。图 17 是显示 P 帧和 B 帧序列的图，其中最近的前面 P 帧具有两个参考前面 P 帧的运动向量。在这个例子中，B 帧 1700 参考前面 P 帧 P3。前面的帧 P3 的一个运动向量  $mv1$  参考下一个前面 P 帧 P2，而第二个运动向量  $mv2$  参考 2 远离的前面 P 帧 P1。图中显示了合适的帧比例分数。

B 帧 1700 的每个宏块可以根据下列项构造成混合形式：

- 由  $mv1$  的帧比例分数  $1/3$  取代的帧 P3 的对应像素（这些像素然后可能乘以某个比例权值  $i$ ）加上由  $mv1$  的帧比例分数  $4/3$  取代的帧 P2 的对应像素（这些像素然后可能乘以某个比例权值  $j$ ）；以及



- 由  $mv_2$  的帧比例分数  $1/6$  取代的帧 P3 的对应像素（这些像素然后可能乘以某个比例权值  $k$ ）加上由  $mv_2$  的帧比例分数  $7/6$  取代的帧 P1 的对应像素（这些像素然后可能乘以某个比例权值  $l$ ）。

当前面 P 帧（相对于 B 帧）的运动向量指向下一个最近的前面 P 帧时，没有必要只利用下一个最近的前面帧作为插值参考，如在图 16 中的。最近的前面 P 帧可以为运动补偿证明更好的选择。假如这样的话，最近的前面 P 帧的运动向量被缩短到从 B 帧到那个 P 帧的帧距离分数。图 18 是显示 P 帧和 B 帧序列的图，其中最近的前面 P 帧具有一个参考前面 P 帧的运动向量。在这个例子中，对于  $M=3$ ，第一个 B 帧 1800 会使用  $1/3$  和  $-2/3$  帧距离分数乘以最近的前面 P 帧 P2 的运动向量  $mv$ 。第二个 B 帧 1802 会使用  $2/3$  和  $-1/3$  帧距离分数（未示出）。将这种选择通知解码器以区分这个例子和图 16 中所示的例子。

和所有其他的编码模式一样，直接模式的使用最好包括根据其他可用的插值和单向量编码模式和参考帧对候选模式进行测试。对于直接模式测试，最近的后续 P 帧（可选地，和下一个最近的后续 P 帧或甚至更远的后续 P 帧，和/或一个或多个前面 P 帧）可被作为候选帧测试，并且少量的比特（典型地是一个或两个）可被用于指定由解码器使用的直接模式 P 参考帧距离。

### 扩展的插值值

在 MPEG-1、2、4 及 H.261 和 H.263 标准中指定了，B 帧使用前向参考帧和后向参考帧的像素值的等加权，如由运动向量所取代的。本发明的另一个方面包括各种有用的不相等加权的应用，这些不相等加权显著提高 B 帧编码效率，以及这种不相等加权到多于两个参考的扩展，这些参考包括两个或多个时间上向前或者向后的参考。本发明的这个方面还包括用于为 P 帧而被参考和插值的一个以上的帧的方法。此外，当两个或多个参考在时间上指向前，或者当两个或多个参考在时间上指向后时，那么有时使用负权重和超过 1.0 的权重是有用的。

例如，图 19 是显示三个 P 帧 P1、P2 和 P3 的帧序列，其中 P3 用两个运动向量使用插值参考，而 P1 和 P2 各一个。例如，如果连续的

变化正在 P1 和 P3 之间的帧范围发生，那么 P2-P1（即帧 P2 的像素值（被 P2 的运动向量所取代）减去帧 P1 的像素值（被 P1 的运动向量所取代））将等于 P3-P2。类似地，P3-P1 将是 P2-P1 和 P3-P2 大小的两倍。在这种情况下，可以通过公式从 P2 和 P1 差别地预测帧 P3 的像素值：

$$P3 = P1 + 2 \times (P2 - P1) = (2 \times P2) - P1$$

既然这样，P3 的插值权重对于 P2 是 2.0，对于 P1 是-1.0。

作为另一个例子，图 20 显示了四个 P 帧 P1、P2、P3 和 P4 的帧序列，其中 P4 使用具有三个运动向量的插值参考，而 P1、P2 和 P3 各一个。因此，由于 P4 是从 P3、P2 和 P1 预测的，所以应用三个运动向量和插值权重。在这种情况下，如果连续的变化正在这个帧范围上发生，那么 P2 - P1 将等于 P3 - P2 和 P4 - P3，而 P4-P1 将等于 3 x (P2 - P1)和 3 x (P3 - P2)。

因此，在这个例子中，基于 P2 和 P1 的 P4 预测为：

$$P4 = P1 + 3 \times (P2 - P1) = (3 \times P2) - (2 \times P1) \quad (\text{权重为 } 3.0 \text{ 和 } -2.0)$$

基于 P3 和 P1 的 P4 预测为：

$$P4 = P1 + 3/2 \times (P3 - P1) = (3/2 \times P3) - (1/2 \times P1) \quad (\text{权重为 } 1.5 \text{ 和 } -0.5)$$

基于 P3 和 P2 的 P4 预测为：

$$P4 = P2 + 2 \times (P3 - P2) = (2 \times P3) - P2 \quad (\text{权重为 } 2.0 \text{ 和 } -1.0)$$

然而可能的是，最靠近 P4（包括 P3 和 P2）的变化，比包括 P1 的预测是更可靠的 P4 的预测值。因此，通过给上面包括 P1 的两个项各 1/4 权重，并给仅包括 P3 和 P2 的项 1/2 权重，将导致：

$$\begin{aligned} & 1/2(2 P3 - P2) + 1/4(3/2 P3 - 1/2 P1) + 1/4(3 P2 - 2 P1) = \\ & 1 \ 3/8 P3 + 1/4 P2 - 5/8 P1 \quad (\text{权重为 } 1.375, 0.25 \text{ 和 } -0.625) \end{aligned}$$

据此，有时既使用大于 1.0 的权重又使用小于 0 的权重是有用的。在其它时候，如果从一个帧到下一个帧存在类似噪音的变化，具有 0.0 和 1.0 之间的适度的系数的正加权平均可产生 P4 的宏块（或者其他的像素区域）的最佳预测。例如，图 20 中 P1、P2 和 P3 各 1/3 的等加权在某些情况下可能形成 P4 的最佳预测。

注意最佳匹配的运动向量被用来确定 P1、P2、P3 等的区域，在这个例子中，计算时使用了 P1、P2、P3。在某些情况下，这个匹配可能最好是 AC 匹配，其允许变化的 DC 项通过 AC 系数预测。或者，如果

使用 DC 匹配（例如绝对差的和），那么经常能够预测 AC 系数的变化。在其他情况下，各种形式的运动向量匹配将形成具有各种权重混合的最佳预测。通常，特定情况的最佳预测是使用这里所描述的方法经验地确定的。

这些技术也适用于具有两个或多个在时间上指向后或指向前的运动向量的 B 帧。当在时间上指向前时，以上描述的用于 P 帧的系数模式被颠倒过来以准确地向后预测当前的 P 帧。使用本发明的这个方面，在向前和向后两个方向上均具有两个或多个运动向量是可能的，因此同时在两个方向预测。这些各种预测的合适的加权混合可以通过选择混合的加权来优化，混合加权最佳地预测了当前 B 帧的宏块（或其他像素区域）。

图 21 是显示 P 帧和 B 帧序列的图，其中不同 P 帧具有一个或多个参考前面不同 P 帧的运动向量，图中也显示了分配给被特定 B 帧参考的各自向前和向后参考的不同权重 a-c。在这个例子中，B 帧 2100 参考三个前面 P 帧和两个后续 P 帧。

在图 21 所示的例子中，帧 P5 必须被解码以使这个例子工作。有时以解码所需的顺序（“发送顺序”）在比特流中给帧排序是有用的，发送顺序没有必要是显示的顺序（“显示顺序”）。例如，在显示循环运动（例如物体的旋转）的帧序列中，特定的 P 帧可能更类似远处的 P 帧而不是最近的后续 P 帧。图 22 是显示 P 帧和 B 帧序列的图，其中 P 帧的比特流发送顺序不同于显示顺序。在这个例子中，帧 P3 更类似于帧 P5 而不是帧 P4。因此在 P4 之前发送和解码 P5，但在 P5 之前显示 P4 是有用的。优选地，当这种 P 帧可以被丢弃时（例如在比特流顺序中 n 个帧的终止时间或者在显示顺序中的帧 X 之后），每个 P 帧应该通知解码器。

如果从一个小的选择集中选择权重，那么少量比特就能够通知解码器要使用哪个权重。和这里所描述的所有其他权重一样，这可以每一个图像单元通知解码器一次，或者在解码过程中权重变化是有用的任何其他点上。

下载新的权重集也是可能的。在这种方式中，少量的权重集可以在给定时间上有效。这允许少量的比特通知解码器在解码过程中在任

何给定点将使用哪个有效的权重集。为了确定合适的权重集，可以在编码过程中测试大量的权重。如果发现小的子集提供高效率，那么可以通知解码器使用那个子集。因此能够只用几个比特把该子集的特定元素通知解码器。例如 10 个比特可以选择 1024 个子集元素中的一个。此外，当改变特定的小子集以维持效率时，可以把新的子集通知解码器。因此，编码器可以对比更新权重集所需的比特数，动态地优化在权重集元素中选择所需的比特数。此外，少量的短编码能被用于表示常见的有用权重，例如 1/2、1/3、1/4 等等。在这种方式中，少量的比特能被用于表示权重集，例如用于 P 帧中的 K-前向-向量预测（其中  $K=1, 2, 3 \dots$ ），或者 B 帧中的 K-前向-向量和 L-后向-向量预测（其中 K 和 L 从 0, 1, 2, 3  $\dots$  中选择），或者 P 帧中的 K-前向-向量和 L-后向-向量预测（其中 K 和 L 从 0, 1, 2, 3  $\dots$  中选择），作为当前 M 值的函数（即 B 帧相对于邻近 P（或 I）帧的相对位置）。

图 23 是显示具有分配权重的 P 帧和 B 帧序列的图。B 帧 2300 具有权重 a-e，其值是从 B 帧权重集 2302 的表格分配的。P 帧 2304 具有权重 m 和 n，其值是从 P 帧权重集 2306 的表格分配的。某些权重可以是静态的（即不变地下载到解码器），并由编码器通知。其他权重可以被动态地下载然后通知。

这个相同的技术可被用于动态地更新权重集以对比于 AC 插值选择 DC 插值。此外，可以表示编码值，其对比于变换值（在替换的线性或非线性表示中）的线性插值选择一般的（线性）插值（通常以非线性表示表示的像素值的插值）。类似地，这种编码值可以表示哪些这样的插值应用于 AC 值或 DC 值或者是否分开预测的 AC 和 DC 部分。

有效的子集也可被用于最少化在目前使用的权重系数集之间选择所必需的比特数。例如，如果 1024 个下载的权重集被保存在解码器中，那么可能在帧的一个特定部分期间需要大概 16 个是有效的。因此，通过选择（从 1024 个中选出）16 个权重集的哪些子集是有效的，只需要使用 4 个比特选择这 16 个中的哪个权重集是有效的。可以为这些最常见子集使用短编码来表示这些子集，因而允许少量的比特在常用的子集中选择。

## 柔化和锐化

和通过减去平均值从 AC 信号中简单地分开 DC 分量一样，在运动向量补偿预测的过程中，其他的滤波运算也是可能的。例如，各种高通、带通和低通滤波器可被应用于像素区域（例如宏块）以提取各种频率带。然后，当进行运动补偿时，可以修改这些频率带。例如，在有噪音的移动图像上滤出最高的频率以柔化（使之不太清晰或者稍微地模糊）该图像通常可能是有用的。更加柔化的图像像素，与用于量化的更陡的斜矩阵（tilt matrix）结合（更陡的斜矩阵忽略当前块中的更高频率的噪音），通常形成更高效的编码方法。已经有可能在每个图像单元的量化斜矩阵中表示变化。还可能下载亮度和色度的定制斜矩阵。注意不管斜矩阵是否改变，都可以提高运动补偿的效率。然而，通常最有效的是既改变斜矩阵又改变在运动补偿过程中应用的滤波器参数。

常见做法是使用降低的色度编码分辨率和色度特定的斜矩阵。但是，在这个例子中，色度编码分辨率是静态的（例如 4:2:0 水平和垂直地编码半分辨率或者 4:2:2 只是水平地编码半分辨率）。可以根据本发明的这个方面通过在运动补偿过程中把一个动态滤波器处理应用于色度和亮度（独立地或合作地）来提高编码效率，每个图像单元选择一个。

题为“Enhanced Temporal and Resolution Layering in Advanced Television”的美国专利申请 09/545233 号（前文引用过）描述了改进的负半周（截断的正弦函数）位移滤波器的使用。这些滤波器具有的优点是当进行运动向量位移的碎片像素（fractional-pixel）部分时，它们保持清晰度。在整像素（integer pixel）位移点和碎片像素位移点，使用降低或提高其清晰度的滤波器可更优地移动某些宏块（或其他有用的图像区域）。例如，对于“变焦”（其中，帧中的某些物体随着时间移出焦点和帧的其他部分进入焦点）来说，转变是清晰度和柔度的变化之一。因此，能够在图像的某些区域提高清晰度而在其他区域降低清晰度的运动补偿滤波器能够提高编码效率。特别地，如果图像的区域正移出焦点，那么降低清晰度可能是有益的，这将柔化图像（因而可能产生更好的匹配）和减少颗粒和/或噪音（因而可能提高编码效

率)。如果图像的区域正进入焦点，那么保持最好的清晰度或者甚至使用更大的负半周正弦滤波器值提高清晰度可能是有益的。

在编码过程中，色度滤波也能够从提高和降低清晰度中受益。例如，4:2:0 编码（水平和垂直地编码半分辨率色度）的许多编码效率优点可以通过为色度使用更柔的运动补偿滤波器而在 U 和/或 V 通道中保持全分辨率来实现。仅仅当在 U 和 V 通道中的色彩细节是高时，才有必要选择最清晰的位移滤波器；在有高色彩噪音或者颗粒的地方较柔的滤波器将更有益。

除了焦点的变化之外，从一个帧到下一个帧的方向和运动模糊变化量也是普遍的。当运动图像胶片帧的速率为 24 fps 时，即使一个简单的对话场景从一个帧到下一个帧在运动模糊方面也有显著的变化。例如，在一个帧中上嘴唇可能是模糊的，且在下一个帧中清晰化，这完全是由于在照相机中打开快门期间的嘴唇运动。对于这样的运动模糊，不仅在运动补偿过程中具有锐化和柔化（模糊化）滤波器而且具有锐化和柔化的方向方面将是有益的。例如，如果可以确定运动的方向，那么沿那个方向的柔化或锐化可被用于对应图像特写的移动或停止。用于运动补偿的运动向量本身可以相对于任何周围的帧（或对应区域）为特定帧（或帧内的区域）提供关于运动量、运动量的变化（即运动模糊）的一些有用信息。特别是，运动向量是 P 帧之间的最佳运动匹配，而运动模糊来自帧中的在打开快门时间期间的运动。

图 24 是一个帧内的物体位置对时间的图。照相机的快门只在部分帧时间期间打开。当快门打开时，物体的任何运动导致模糊。运动模糊量是由在快门打开期间的位置变化量表示的。因此，当快门打开时位置曲线 2400 的斜率是运动模糊的度量。

运动模糊量和运动方向也可以从清晰度量度、周围的运动向量（图像区域匹配的地方）、特征拖影检测和帧区域的人为辅助指定的组合中确定。可以基于确定的运动模糊量和运动方向选择滤波器。例如，对比所确定的运动模糊量和运动方向，各种滤波器的映射可以经验地确定。

当结合本发明的其他方面时，这种智能应用的滤波器可以显著地提高压缩编码效率。可以选择少量的这种滤波器，用少量的比特通知

解码器。同样，这能够是以每个图像单元一次或者在解码过程中的其他有用点上实现。与权重集一样，可以使用一组动态装入的滤波器，以及有效的子集机制以最少化在最有利的滤波器参数集合之间选择所需的比特数。

## 实现

本发明可以在硬件或软件，或者两者的组合（例如可编程的逻辑阵列）中实现。除非另外指定，作为本发明的一部分而包括的算法本质上不与任何特定的计算机或者其他装置相关。特别地，各种通用目的的设备可以与根据本文所述编写的程序一起使用，或者可能更方便的是构造更特殊的装置（例如集成电路）以执行特定功能。因此，本发明可以在一个或多个可编程计算机系统上执行的一个或多个计算机程序中实现，每个可编程计算机系统包括至少一个处理器、至少一个数据存储系统（包括永久的和易失的存储器和/或存储单元），至少一个输入设备或端口，和至少一个输出设备或端口。程序代码被应用于输入数据以执行这里描述的功能并生成输出信息。输出信息以公知的方式被应用于一个或多个输出设备。

每个这样的程序可以用任何希望的计算机语言（包括机器的、汇编的、或高级过程化的、逻辑的、面向对象的编程语言）实现以与计算机系统通信。在任何情况下，语言可以是编译语言或解释语言。

每个这样的计算机程序优选存储在或者下载到存储介质或设备上（例如固态存储器或介质，或者磁介质或光介质），这些存储介质或设备是通用或专门目的的可编程计算机可读的，当存储介质或设备被计算机系统读取以执行这里描述的程序时，这样的计算机程序用于配置和操作计算机。也可以考虑将本发明系统实现成计算机可读存储介质，其被配置为具有一计算机程序，如此配置的存储介质使计算机系统以特定的和预定的方式运行以执行本文所描述的功能。

已经描述了本发明的多个实施例。但是，应该理解的是，在不脱离本发明的精神和范围的情况下可以进行各种修改。例如，上文所述的某些步骤可能是顺序无关的，因此可以以一种与所述不同的顺序执行。因此，其他的实施例均在所附权利要求的范围内。

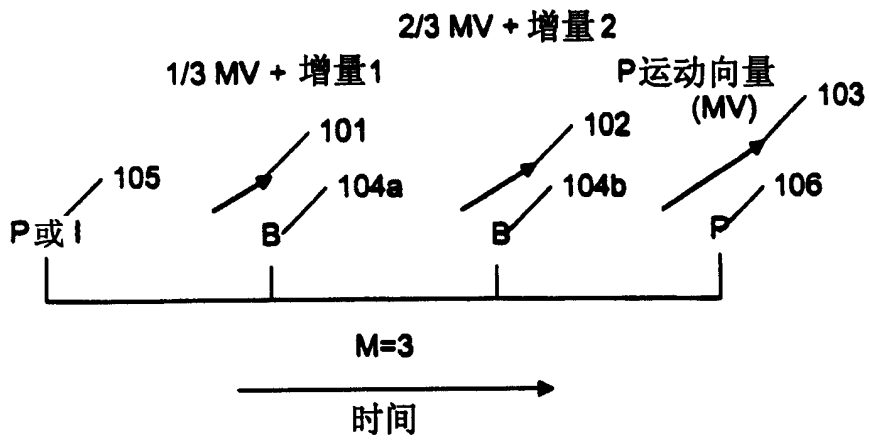


图1  
比例运动向量加权  
(现有技术)



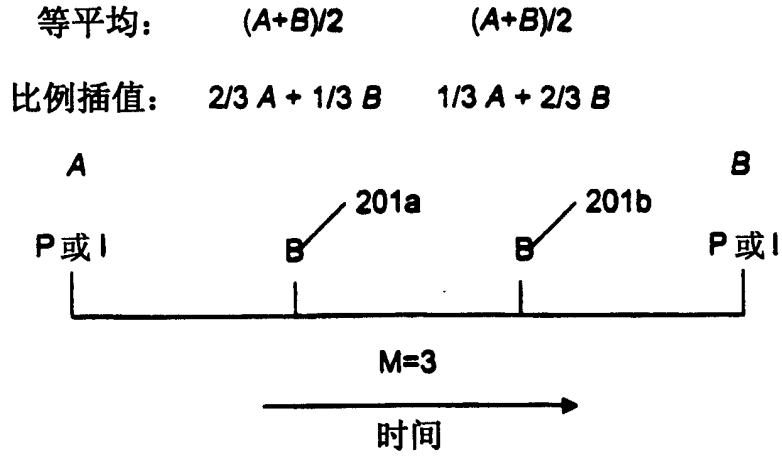


图2  
像素值比例加权

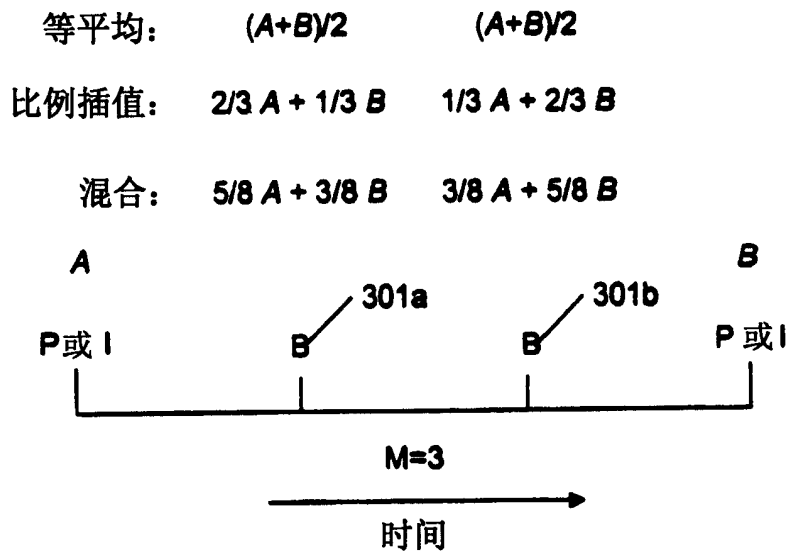


图3  
混合像素值加权

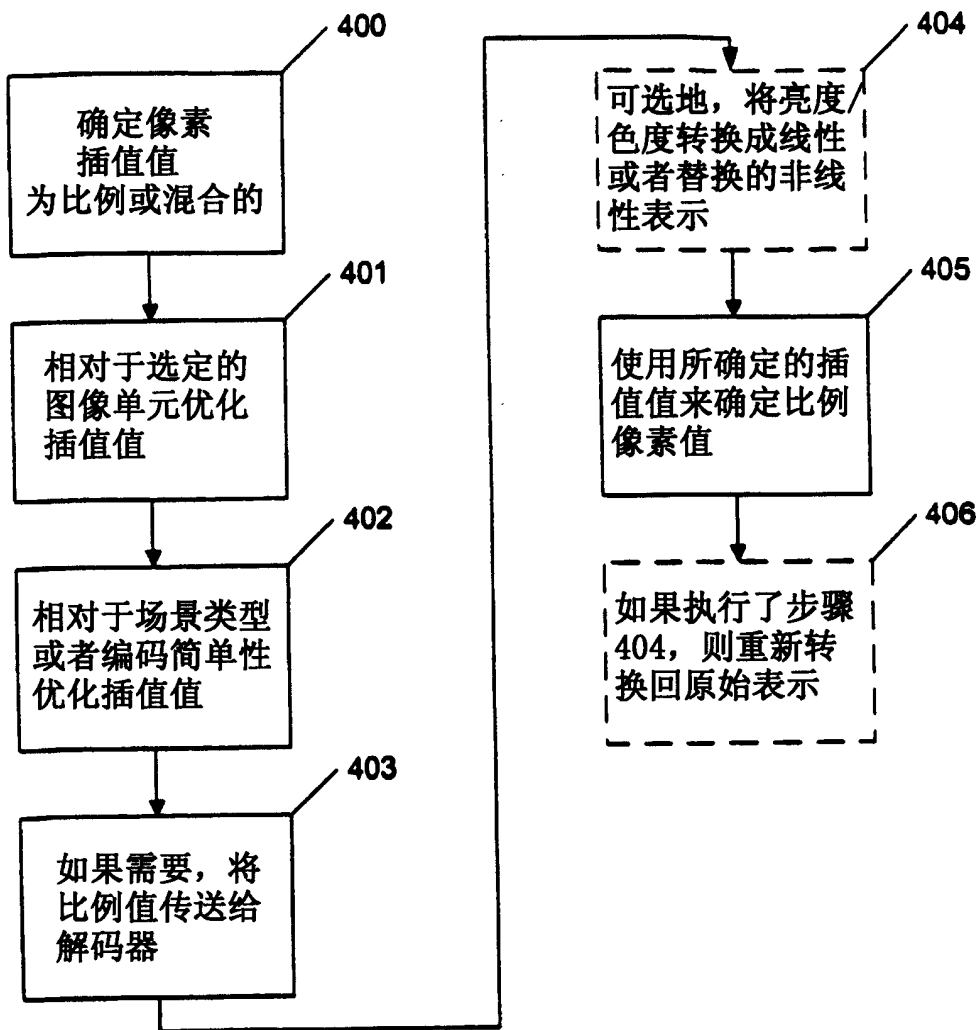


图4

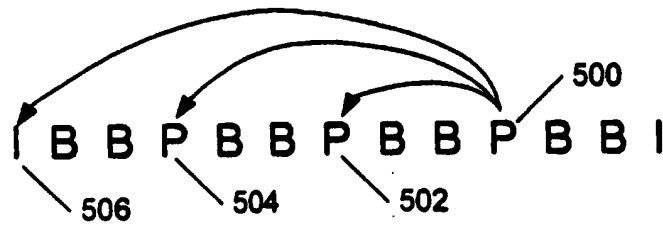


图5

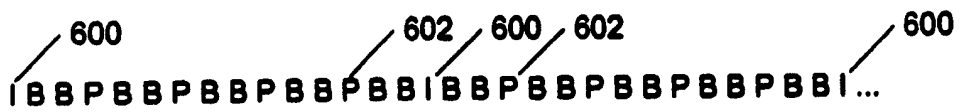


图6A  
现有技术

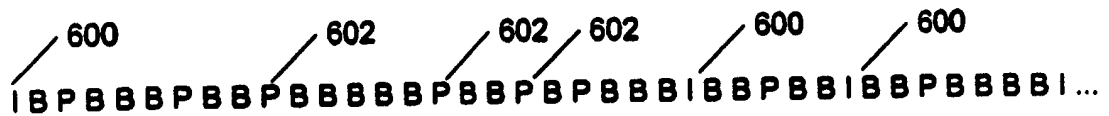


图6B  
现有技术

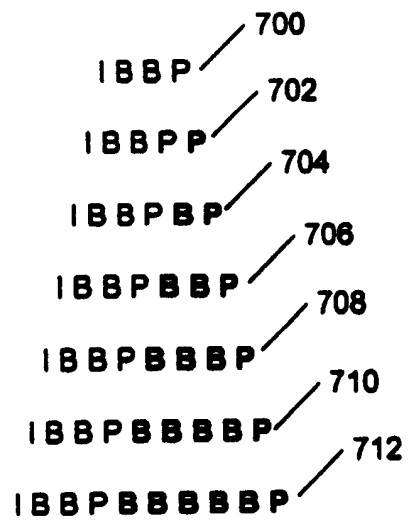


图7

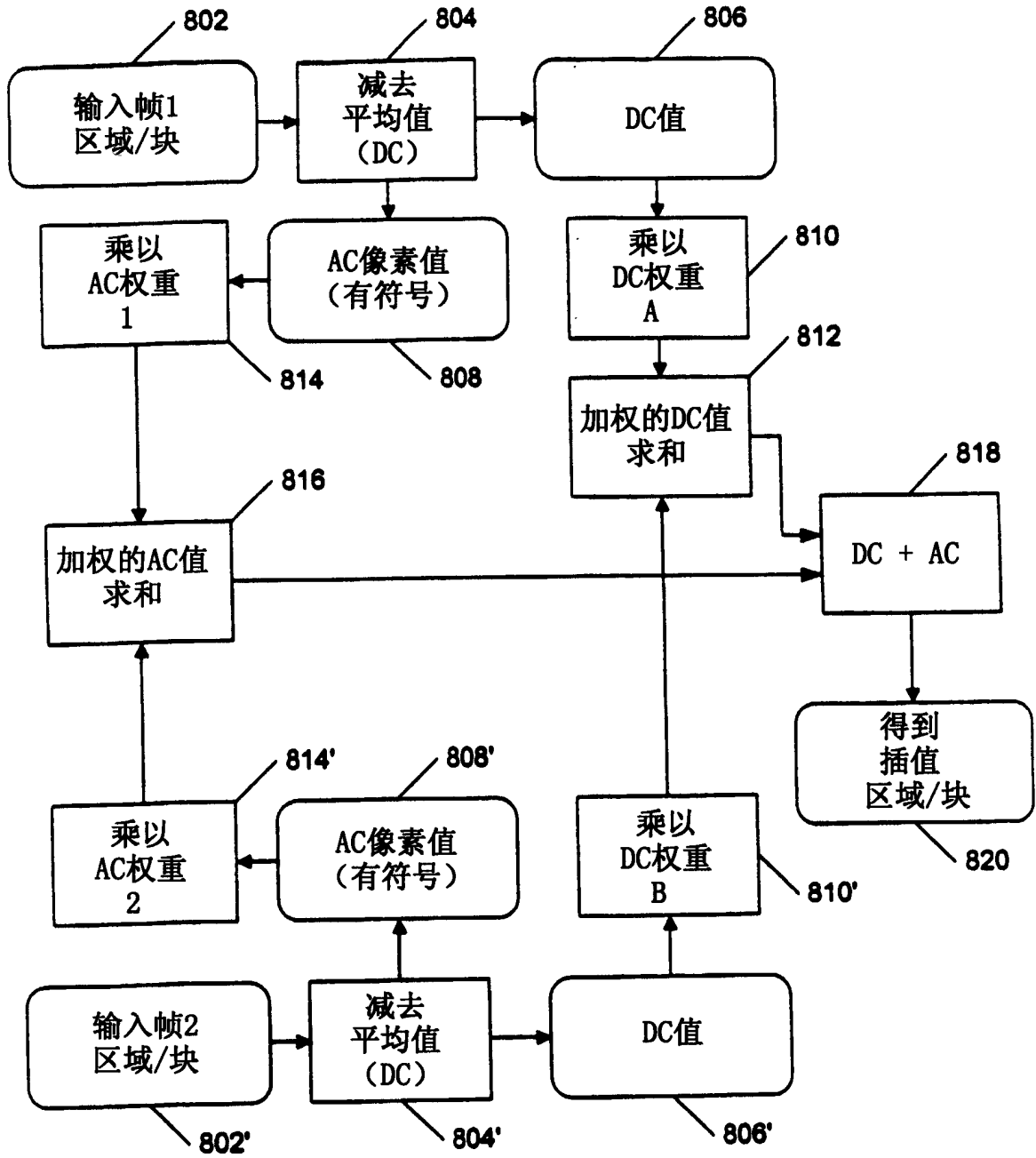


图8

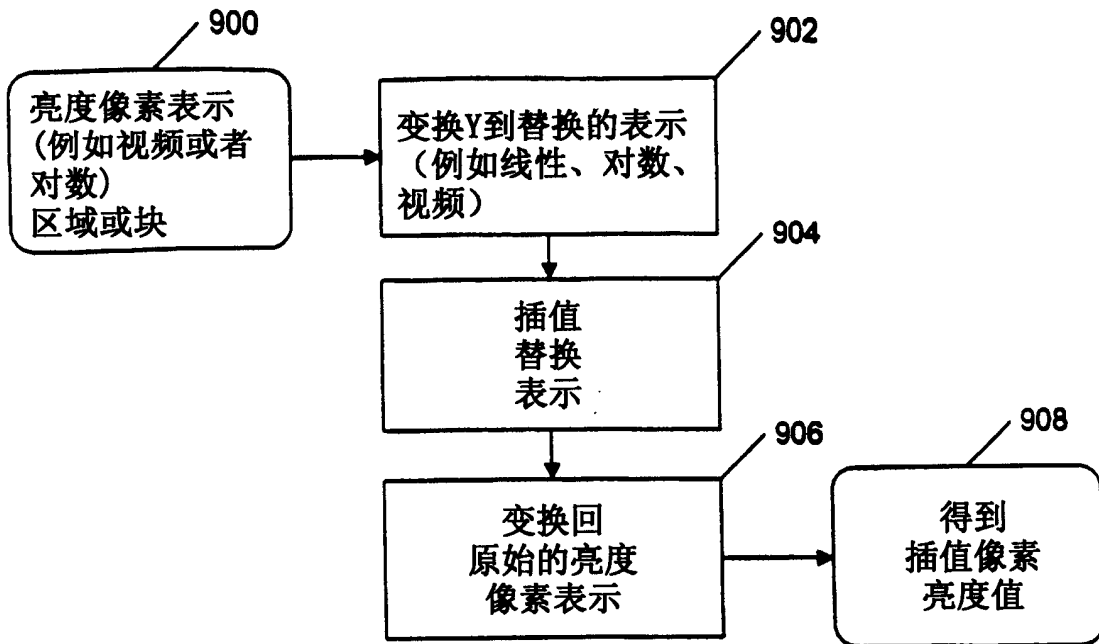


图9

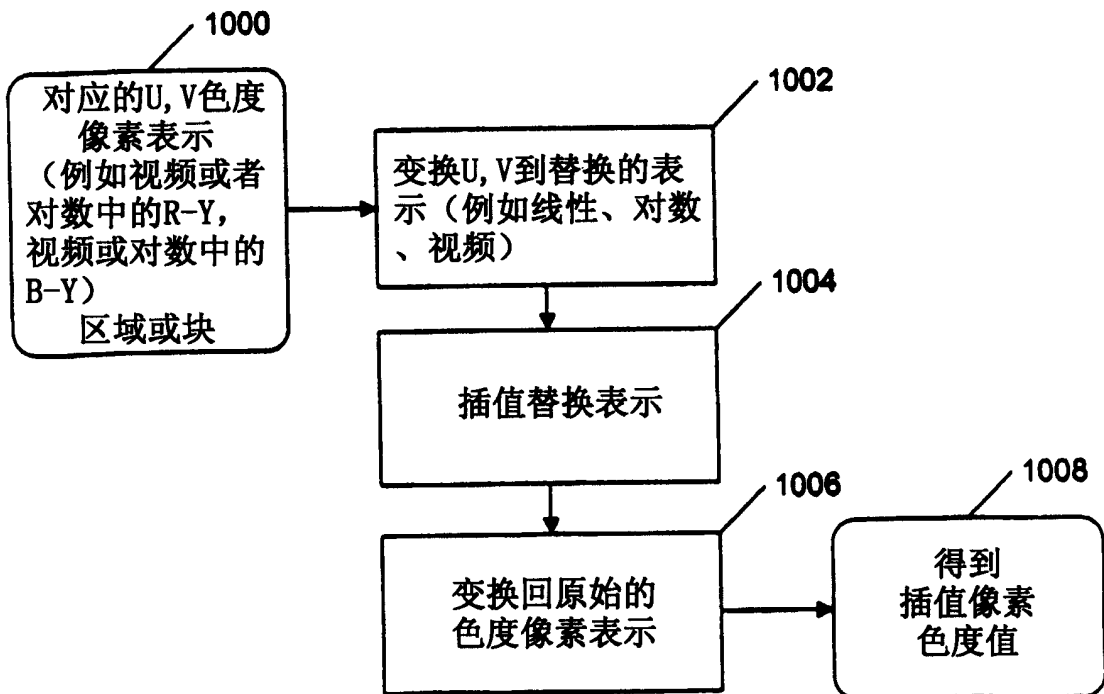


图10

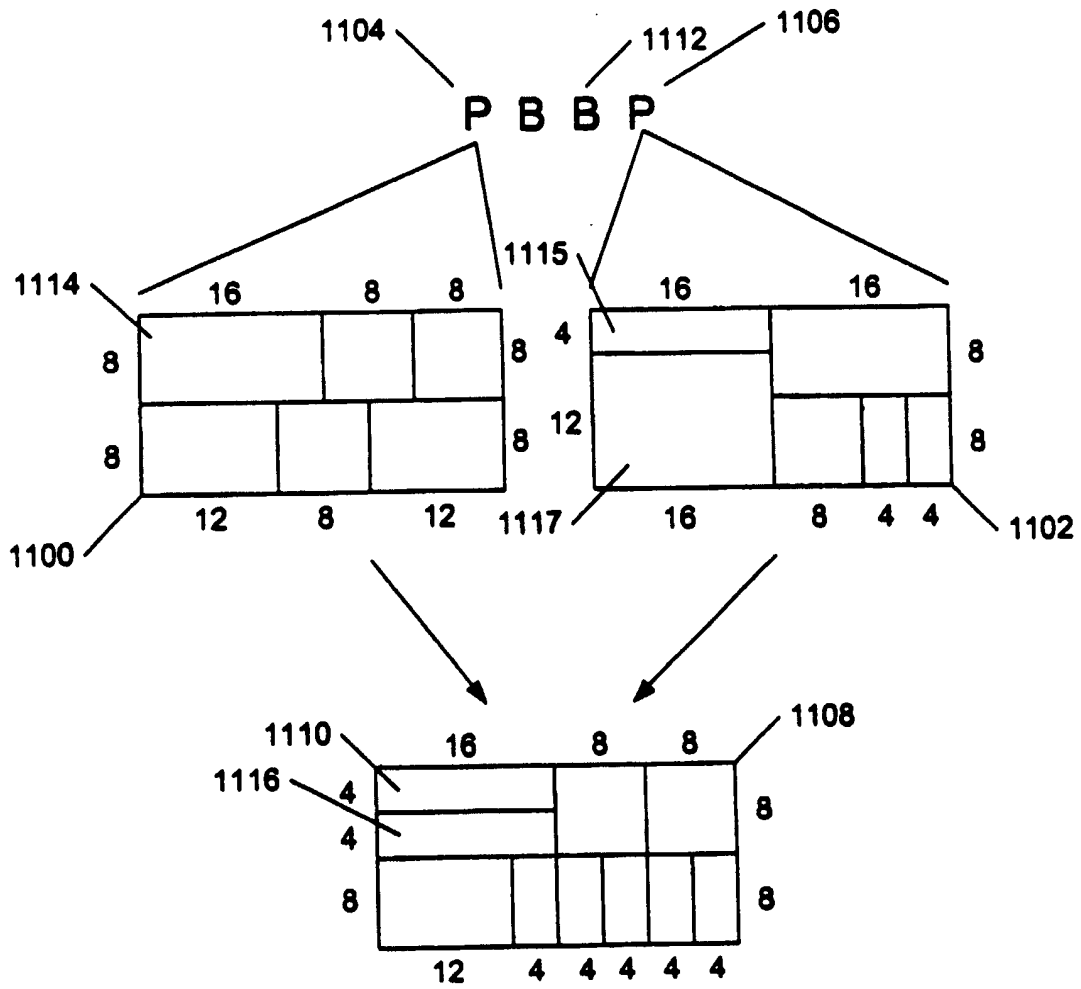


图11

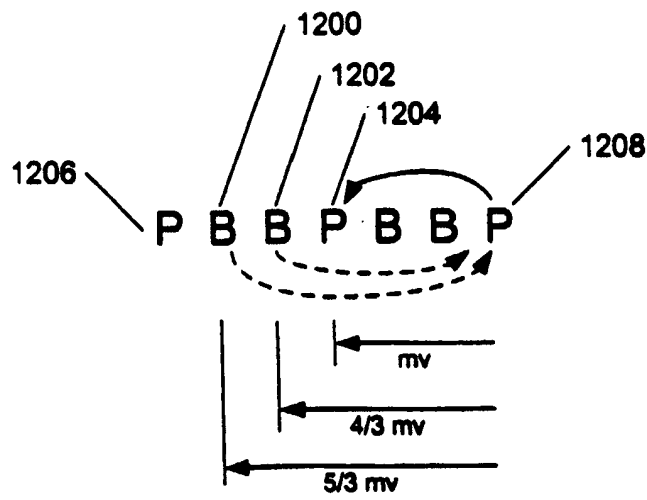


图12

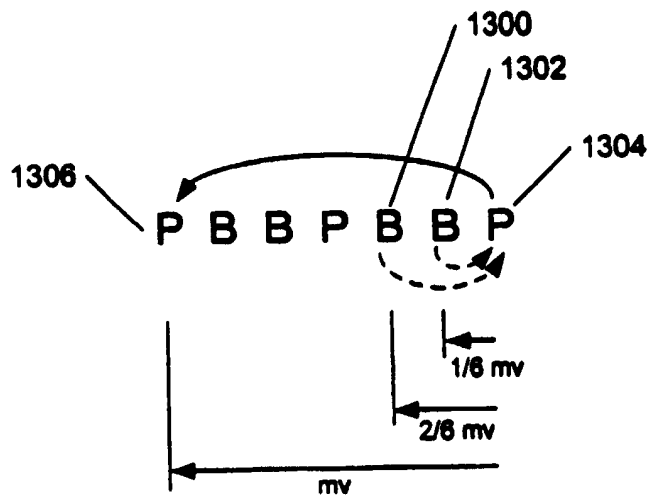


图13



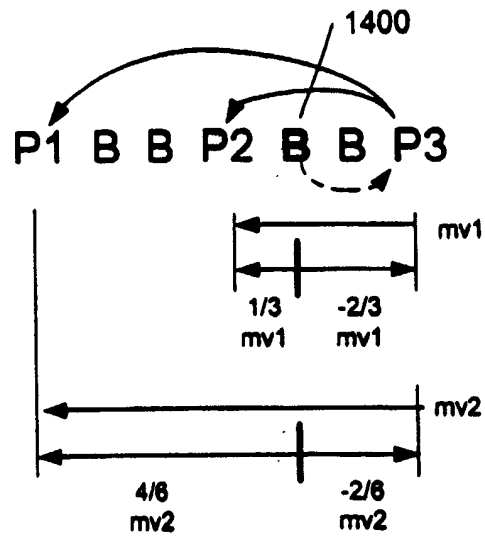


图14

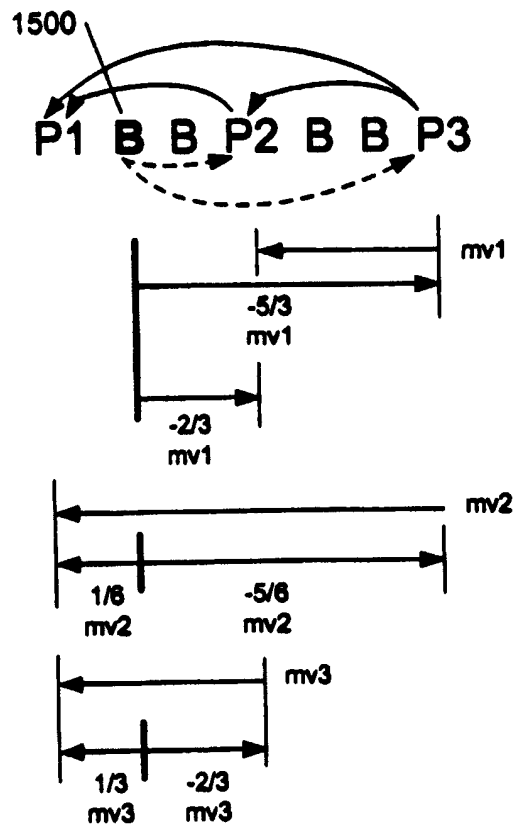


图15

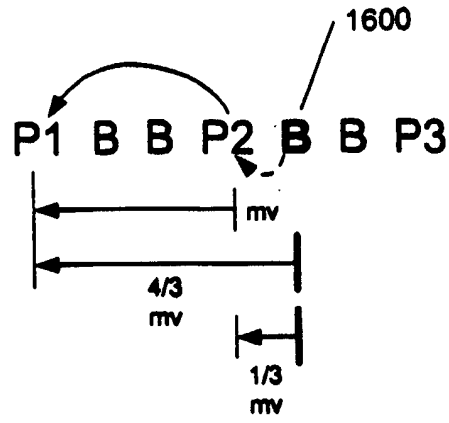


图16

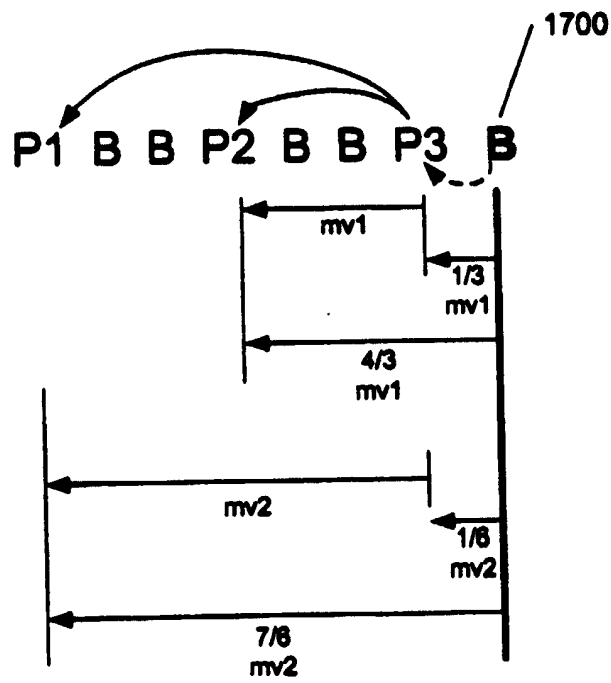


图17

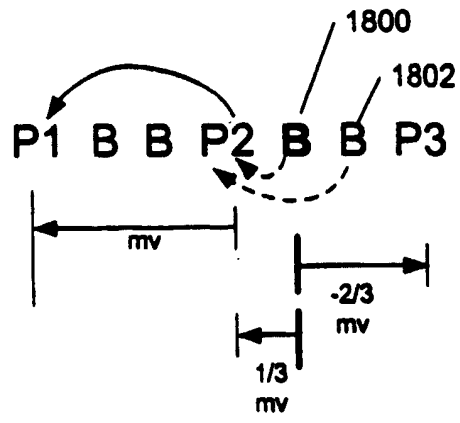


图18

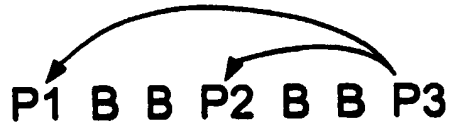


图19

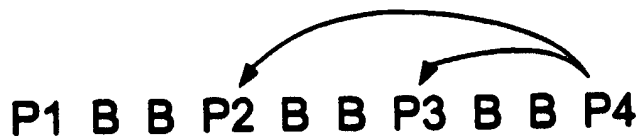


图20

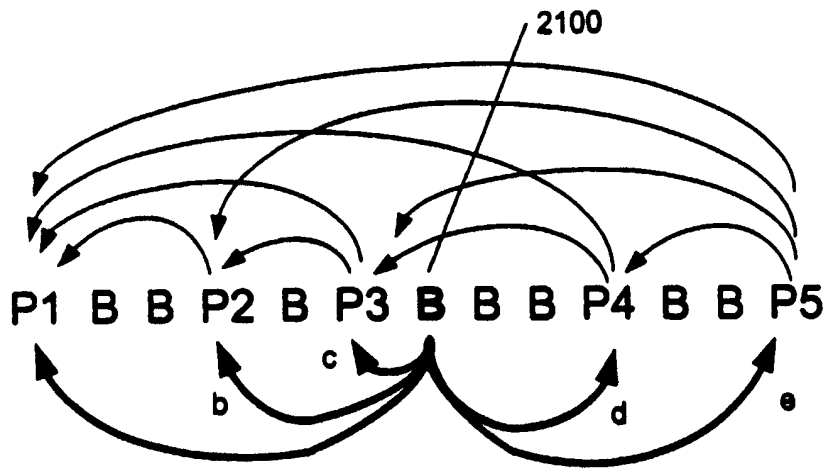


图21

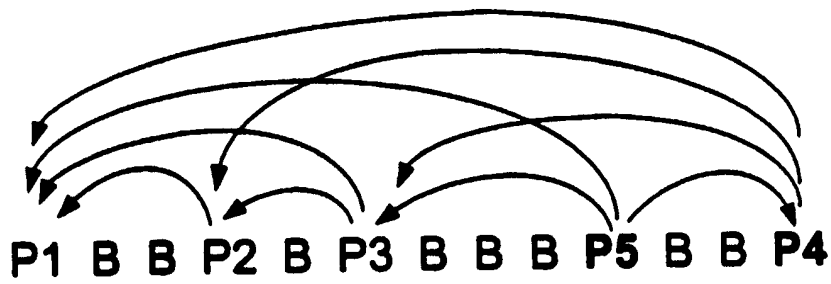


图22

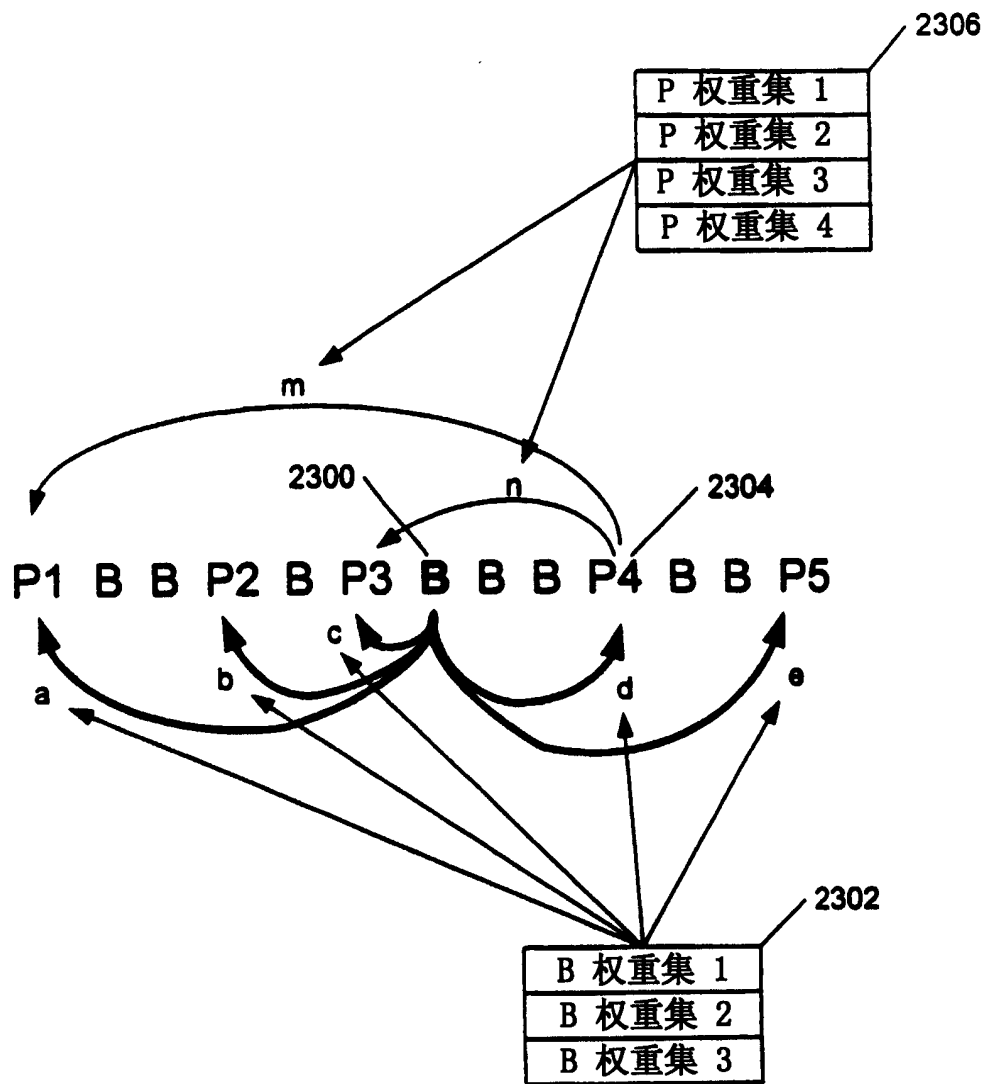


图23

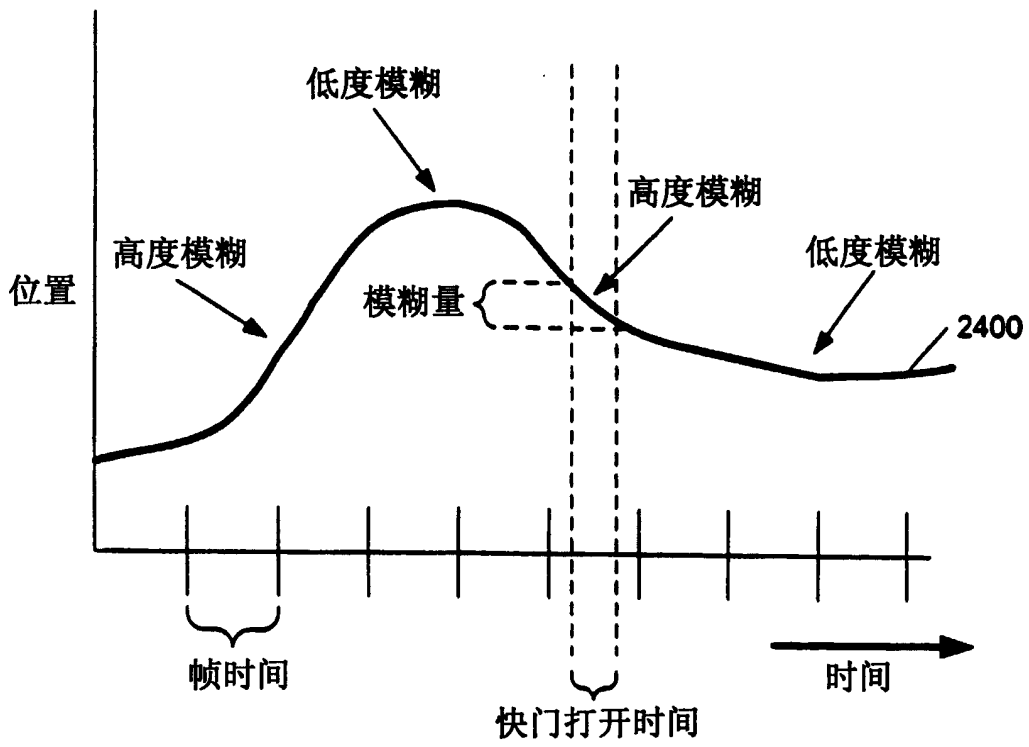


图24