

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
H04N 7/15 (2006.01)



# [12] 发明专利申请公开说明书

[21] 申请号 200510082090.4

[43] 公开日 2006年2月1日

[11] 公开号 CN 1728815A

[22] 申请日 2001.11.9

[21] 申请号 200510082090.4

分案原申请号 01820446.5

[30] 优先权

[32] 2000.12.12 [33] US [31] 09/735716

[71] 申请人 英特尔公司

地址 美国加利福尼亚州

[72] 发明人 M·杨 B·-L·耶奥

L·-C·泰

[74] 专利代理机构 中国专利代理(香港)有限公司  
代理人 陈景峻

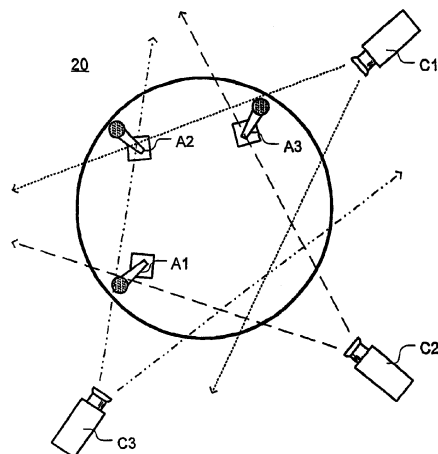
权利要求书5页 说明书10页 附图5页

[54] 发明名称

自动多摄像机视频合成

[57] 摘要

公开了一种方法和装置，用于从多个视频输入源(C1 - C3)中自动选择一个视频输出(50)。一种方法是，每个视频输入源都与一个或多个音频传感器(A1 - A3)相连接。最好，设置一个音频传感器以接收来自相关视频源的视场的有效覆盖接收方向的音频信号。自动选择器(40)在很短(例如0.5秒)的检测间隔内为每个音频传感器计算音频计数。在每个检测间隔，不同的视频源都有可能被选择为视频输出。自动选择器根据一个检测间隔的音频计数和视频源选择的最近时间历史而选择一个视频源。例如，若一个新的视频源刚刚被选择，则一个不同的源在几秒之内就不能被选择。时间历史的使用还可以增加使源的选择变换方式看起来自然的可能性。



1. 一种装置，包括一个计算机可读介质，该介质包含的计算机指令在执行时可使处理器或多个通信处理器执行从多个提供的视频源中自动选择一个视频源的方法，该方法包括：
- 5 将至少一个音频源与每个视频源相连接；
- 在一个检测间隔内估计从每个音频源接收的音频信号，从而为每个音频源形成音频计数，该计数表示在该检测间隔内该音频源的音频有效性级别；
- 10 保持视频源选择的至少一部分时间历史以及一组对话规则，对于多次连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性；
- 为该检测间隔对与每个视频源相关联的音频计数执行加权比较，其中该加权是基于视频源选择的部分时间历史和所述一组对话规则；
- 15 和
- 根据该加权比较选择一个当前视频源，
- 其中对于一个扩展的时间期间，最后选择的视频源被保持选择时，进一步用部分权重对与该视频源相关的音频计算进行加权，以便减小保持被选择的该视频源的可能性。
- 20 2. 如权利要求 1 的装置，其中将至少一个音频源与每个视频源的连接包括将该连接表示为一个矩阵，该矩阵指示音频源和视频源的每个连接的权重，其中加权比较的执行包括，通过将音频计数以矩阵中的对应权重进行加权，为每个视频源产生一个结合计数。
3. 如权利要求 1 的装置，其中加权比较的执行包括，根据一个加权函数将与最后选择的当前视频源相关联的音频计数进行加权，该加权函数随着最后选择的当前视频源保持被选择的时间长度而改变。
- 25 4. 如权利要求 1 的装置，其中音频计数的形成包括，将检测间隔分为多个时域子窗，对于每个子窗；
- 确定该子窗期间每个音频信号的变化； 和
- 30 对该子窗内具有最大变化的音频源增加音频计数。
5. 如权利要求 1 的装置，该方法还包括，当其中一个音频源和对应的视频源位于当前被选择的视频源位置的远程端时，通过一个通信

网络接收音频信号，并当其视频源被选择为当前视频源时通知远程端传送视频。

6. 如权利要求 1 的装置，还包括在选择当前视频源之后对每个音频源的音频计数复位。

5 7. 一种自动视频切换系统，包括：

多个音频输入，每一个都与至少一个对应的视频源相关联；

一个音频计数器，能够为每个音频输入计算音频计数，给定音频输入的该音频计数对应于在一个检测间隔该输入所接收的音频信号；

状态存储器，用于存储视频源选择的至少一部分时间历史；和

10 切换逻辑电路，用于根据检测间隔内与每个视频源相关联的音频计数的加权比较，为给定的该检测间隔选择一个当前视频源，该比较加权是基于视频源选择的部分时间历史，其中该切换逻辑电路根据包含对应于视频流中出现在视频切换点之后的一个点的音频输入的检测间隔，在一个给定的临时视频切换点选择一个当前视频源。

15 8. 如权利要求 7 的系统，其中所述切换逻辑电路包括一组对话规则，对于多次连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性。

9. 一种自动视频切换系统，包括：

20 多个音频输入，每一个都与至少一个对应的视频源相关联，并且至少一个音频输入与至少两个对应的视频源相关联；

一个音频计数器，能够为每个音频输入计算音频计数，给定音频输入的该音频计数对应于在一个检测间隔该输入所接收的音频信号；

状态存储器，用于存储视频源选择的至少一部分时间历史；和

25 切换逻辑电路，用于根据检测间隔内与每个视频源相关联的音频计数的加权比较，为给定的该检测间隔选择一个当前视频源，该比较加权是基于视频源选择的部分时间历史。

10. 如权利要求 9 的系统，其中所述切换逻辑电路包括一组对话规则，对于多次连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性。

30 11. 如权利要求 10 的系统，其中该系统能够在后期合成模式下进行操作，其中该切换逻辑电路根据包含对应于视频流中出现在视频切换点之后的一个点的音频输入的检测间隔，在一个给定的临时视频切

换点选择一个当前视频源。

12. 如权利要求 9 的系统，其中所述音频计数在每个检测间隔的开始复位。

5 13. 一种从多个提供的视频源中自动选择一个视频源的方法，该方法包括：

将至少一个音频源与每个视频源相关联；

在一个检测间隔内估计从每个音频源接收的音频信号，从而为每个音频源形成音频计数，该计数表示在该检测间隔内该音频源的音频有效性级别；

10 保持视频源选择的至少一部分时间历史；

为该检测间隔对与每个视频源相关联的音频计数执行加权比较，其中该加权是基于视频源选择的部分时间历史；

根据该加权比较选择一个当前视频源；和

15 将当前视频源的选择对应到视频流中该检测间隔结束之前的一个视频切换点。

14. 如权利要求 13 的方法，还包括保持一组对话规则，对于多次连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性，其中还根据所述一组对话规则进行加权。

20 15. 一种从多个提供的视频源中自动选择一个视频源的方法，该方法包括：

将至少一个音频源与每个视频源相关联，并且至少一个音频源与至少两个视频源相关联；

25 在一个检测间隔内估计从每个音频源接收的音频信号，从而为每个音频源形成音频计数，该计数表示在该检测间隔内该音频源的音频有效性级别；

保持视频源选择的至少一部分时间历史；

为该检测间隔对与每个视频源相关联的音频计数执行加权比较，其中该加权是基于视频源选择的部分时间历史；

根据该加权比较选择一个当前视频源。

30 16. 如权利要求 15 的方法，还包括在选择当前视频源之后对每个音频源的音频计数复位。

17. 如权利要求 15 的方法，还包括保持一组对话规则，对于多次

连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性，其中还根据所述一组对话规则进行加权。

18. 如权利要求 16 的方法，包括将当前视频源的选择对应到视频流中该检测间隔结束之前的一个视频切换点。

5 19. 一种包括计算机可读介质的装置，所述可读介质包含计算机指令，当执行该指令时使得一个处理器或多个通信处理器实现一种从多个提供的视频源中自动选择一个视频源的方法，该方法包括：

将至少一个音频源与每个视频源相关联；

10 在一个检测间隔内估计从每个音频源接收的音频信号，从而为每个音频源形成音频计数，该计数表示在该检测间隔内该音频源的音频有效性级别；

保持视频源选择的至少一部分时间历史；

为该检测间隔对与每个视频源相关联的音频计数执行加权比较，其中该加权是基于视频源选择的部分时间历史；

15 根据该加权比较选择一个当前视频源；和

将当前视频源的选择对应到视频流中该检测间隔结束之前的一个视频切换点。

20. 如权利要求 19 的装置，其中的方法还包括保持一组对话规则，对于多次连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性，其中还根据所述一组对话规则进行加权。

21. 一种包括计算机可读介质的装置，所述可读介质包含计算机指令，当执行该指令时使得一个处理器或多个通信处理器实现一种从多个提供的视频源中自动选择一个视频源的方法，该方法包括：

25 将至少一个音频源与每个视频源相关联，其中至少一个音频源与至少两个视频源相关联；

在一个检测间隔内估计从每个音频源接收的音频信号，从而为每个音频源形成音频计数，该计数表示在该检测间隔内该音频源的音频有效性级别；

30 保持视频源选择的至少一部分时间历史；

为该检测间隔对与每个视频源相关联的音频计数执行加权比较，其中该加权是基于视频源选择的部分时间历史；

根据该加权比较选择一个当前视频源。

22. 如权利要求 21 的装置，其中的方法还包括在选择当前视频源之后对每个音频源的音频计数复位。

5 23. 如权利要求 21 的装置，其中的方法还包括保持一组对话规则，对于多次连续检测间隔，当一个视频源保持被选择时，所述对话规则增加视频源之间切换的可能性，其中还根据所述一组对话规则进行加权。

24. 如权利要求 22 的装置，包括将当前视频源的选择对应到视频流中该检测间隔结束之前的一个视频切换点。

## 自动多摄像机视频合成

## 5 技术领域

本发明涉及多摄像机视频系统，特别涉及一种自动多摄像机视频合成系统及其操作方法。

## 背景技术

10 在通常的视频传输和纪录领域中，经常从多个视点或位置同时捕捉视频。一个普通的例子就是体育广播：例如棒球比赛可能使用五个或更多个摄像机来从多个视角捕捉动作。一个或多个技术人员在这些摄像机之间进行切换，以提供包括比赛中此刻发生事件的最佳视角的电视信号。另一个例子是电影。然而电影编辑要在记录事件很长时间后，利用多个摄像机拍摄的大多数场景以特定的合成顺序而进行。

15 虽然可能比体育竞赛或电影少一些刺激，但多摄像机视频数据还有很多其它应用。例如，摄像机角度的选择可以对几乎任何录像或广播事件提供更丰富的纪录，而不论该事件是会议、演讲、视频会议或电子教室，下面说明一些例子。

20 两个研究者根据讲话者的动作模式提出了一种应用于视频会议的自动摄像机切换策略。参见1984年11月26-29日IEEE全球电话讨论会议记录第313-318页中F.Canavesio&G.Castagneri所作的“根据视频会议中的动作模式进行自动摄像机切换的策略”。该论文中所述的系统对六个视频会议参加者中的每一个都配有一个麦克风和一个摄像机。提供另外两个摄像机用于示出所有参加者的分屏全景的输入。一  
25 微处理器周期性的执行“有效讲话人识别处理”，用于检测所有的参加者中的哪个人正在讲话，并创建一个包括六个“讲话/不讲话”值的二进制有效图案。

30 在系统中输入多个时基阈值。该微处理器执行声音切换运算，确定七个摄像机视图（六个人的加一个全景）中的每一个对应用于哪个二进制有效图案。实际上，该算法根据谁正在讲话、哪个摄像机当前被选择、以及当前被选择的摄像机视图是否已经被保持最小的时间量来决定在新的估算间隔中使用哪个摄像机视图。如果检测到同时有

一个以上的讲话人或没有一个人讲话，则系统将在预定量的时间后切换到会议的全景图。通常，当检测到一个人讲话时，如果他们持续的讲话或中间只是稍有停顿，系统都将持续的选择最靠近该讲话人的视图。

## 5 发明内容

本发明包括用于选择多个并行视频输入中的一个作为视频输出的系统和方法。通常，进行切换决定时并不需要关于摄像机定位的特别知识或视频流自身的信息。相反，所述的实施例可以利用音频输入和

10 视频输入选择历史来选择一个视频输出。这样可极大的缓减系统校准和处理功率的需要，同时还可使视频输入选择的变化看起来自然。

这里研究的是在很多视频场景（特别是对话式场景）中，多个可用的摄像机角度中最好的是与有效讲话人（或其他声音源）的位置最直接相关的角度。所述实施例可很好的利用该研究，根据音频有效的位置来选择一个视频源。所述实施例使用一个相关矩阵，该矩阵将多个

15 音频输入中的每一个与至少一个视频输入相关联。当一个特定音频输入表现出音频有效性增加时，与该音频输入相关的视频输入将更可能被选为当前视频输出。

## 附图说明

通过阅读参照下面附图的说明可以更好的理解本发明，其中：

- 20 图 1 示出本发明一个实施例的典型的摄像机/麦克风结构；
- 图 2 示出根据本发明一个实施例的自动多摄像机视频合成系统的谈话图；
- 图 3 包括根据本发明一个实施例的视频自动选择器的方块图；
- 图 4 包括用于本发明一个实施例的音频配乐加权级别函数的图
- 25 表；
- 图 5 包括根据本发明一个实施例的视频源自动选择流程图；
- 图 6 示出用于本发明一个实施例的摄像机/麦克风结构；
- 图 7 示出使用根据本发明实施例的一个视频音频选择器的网络视频会议结构。

## 30 具体实施方式

在所述实施例部分的基于音频有效性而进行视频输出选择的同时，最好将其它因素也考虑到选择过程中。这些其他因素中最显著的



是视频源选择的时间历史。例如一个人可在合成视频中的讲话中占用较长的时间。虽然在该延长的间隔期间，通常有一个摄像机角度为“最佳”，但“最佳”的合成视频流有时也应当示出，例如更宽角度的视图、从另一个比最佳视点稍差的视点看到的讲话人或沉默的参加者的视图。通过使用时间历史，所述的实施例可以在使一个“最佳”音频信道占主导的同时，选择一个混合的摄像机角度。或者作为另一例子，可以使用时间历史来提高平滑度标准，例如通过提供它，每当选择一个新的摄像角度时，该角度将在某初始时间接收到更大的保持被选择的可能性。

所述实施例提出了在 Canavesio 和 Castagneri 的系统中没有出现的几个特征。例如，利用本系统，音频计数将不局限于二进制处理，摄像机/麦克风的联系变的灵活，允许多个麦克风与一个摄像机联系，反之亦然。这将不确定性的方面引入了视频切换的计时，并引入了背景技术中没有提到的多个可能的切换条件（根据所给出的讲话模式）。在至少一个实施例中，可在不同的远距离位置产生视频输入，各个位置可能接收到与其它位置不同的视频会议的视图。

更为重要的是，所述实施例可防止在一个很长的时间段中对一个摄像机的连续选择，即使该摄像机示出一个长篇大论的讲话人的特写镜头也不例外。当一个人独占整个会议时，通过插播其他参加者的简短镜头和或全景镜头，可以增加视频输出的视觉效果。

为了进一步说明这些原则，参照图 1 的会议室结构 20 来说明第一实施例。在结构 20 中，在会议桌周围放置三个麦克风 A1、A2、A3。会议桌周围同样安装了三个视频摄像机 C1、C2、C3。摄像机 C1 的视野包括麦克风 A1，摄像机 C2 的视野包括麦克风 A2，摄像机 C3 的视野包括麦克风 A3。

图 2 示出一种视频切换系统 30 的结构，它用于例如图 1 所示的结构 20 的麦克风/摄像机结构。一视频自动选择器 40 接收到各麦克风 A1、A2 和 A3 的音频输入。视频开关 50 接收到各摄像机 C1、C2、C3 的视频输入。视频自动选择器 40 产生一个音频输出和一个切换命令。视频开关 50 利用该切换命令将视频源 C1、C2、C3 中的一个输出的视频切换为视频输出。

图 3 包含视频自动选择器 40 的更详细的方块图。下面将顺序说明

图 3 种的各方块。

参数接口 42 向切换逻辑电路 44 提供操作参数。最佳的，这些参数中至少一部分可以根据预先设定的模式单独或成组的由用户调节。下面结合各参数应用来说明它们的含义。

5           一音频计数器 45 接收音频输入 A1、A2 和 A3。检测间隔长度输入参数（参数接口 42 输入）确定计算音频计数的时间长度。当新的检测间隔将要开始时，切换逻辑电路 44 向计数器 45 提供一复位信号。该音频计数器为各音频输入 A1、A2、A3 形成音频计数，表现为在检测间隔期间音频输入的音频有效性的级别。在该间隔结束时，向切换逻辑  
10           电路 44 发送该音频计数。

切换逻辑电路 44 进行实际判断，决定在给定的检测间隔期间，选择哪个视频源作为当前视频源。该决定部分地基于在该检测间隔期间的音频计数。该决定也基于存储在状态存储器 46 中的变量，下面利用参数接口 42 提供的参数进行说明。

15           状态存储器 46 存储至少一部分视频源选择的时间历史。在一个简单的情况中，该历史包括上次选择的视频源和相关的视频片段长度（表示上次选择的视频源保持被选择状态的时间长度）。该历史也可根据类型对视频片段分类。该历史中还有其他有用的项目，例如各没有被选择的视频源从上次被选择直到现在的时间长度和/或以前 n 个视频片  
20           段的历史。

可选择的，视频自动选择器 40 可包括一音频混频器 48。切换逻辑电路 44 可以根据当前的视频选择向混频器 48 提供一个音频控制。或者，可以固定混频器 48，例如使其平均的混合所有的音频输入以形成一个音频输出。

25           视频自动选择器 40 的一个操作方法如下。该系统考虑在视频信道/音频传感器相关参数中指定的视频信道和音频传感器。这种相关可以是手动的（由用户设定），固定的（例如，各视频摄像机连接一个定向的麦克风），或者根据物理结构信息（例如已知的麦克风位置和摄像机视野）自动设定。这种相关可为 1: 1，即一个音频输入对应一个  
30           视频输入；在其他实施例中，一个音频输入可以对应多个视频输入，反之亦然。在图 1、2 和 3 的第一实施例中，A1 对应 C1，A2 对应 C2，A3 对应 C3。这种对应可以表示为一个 N×M 矩阵，其中 N 为音频输入的

数目， $M$  为视频输入的数目。当对应关系为 1: 1 (即  $N=M$ ) 时，可以使用两个长度为  $N$  的阵列： $A[i]$  表示音频输入， $V[i]$  表示对应视频输入。

图 5 包括一个流程图 60，它包括一个视频源选择处理的循环步骤。  
5 该系统以多种方式实现视频源选择时间历史。首先是在判断方块 62 的每个源选择循环开始时发生。每当系统切换到一个不同的视频源时，该系统可以被控制为在至少最小视频片段长度（例如，在示意性例子中为 4 秒）内锁定该视频源。在该初始间隔期间，该系统只进行计时，直到达到该最小视频片段长度为止，系统不做任何处理，只是循环执行  
10 方块 62 和 64。

当片段长度最后达到了该最小片段长度时，从块 62 转到执行块 66。在块 66，音频检测间隔开始。该检测间隔可根据系统模式，以多种方式中的一种与视频切换点相关。至少有两种模式，即实时合成模式和后期（例如档案式）合成模式。在实时合成模式中，视频切换利用一个包括前面音频采样的音频检测间隔。例如，如果  $t_s$  为可能的切换点，且检测间隔的长度为  $T_E$ ，则用于切换点  $t_s$  的检测间隔从  $t_s - T_E$   
15 开始。后期合成模式由于不具有实时限制，因此可以对音频采样进行“预测”，看看下面谁将讲话，从而可能切换视频源并实现在新的讲话人开始讲话的同时播出他的画面。例如，在后期合成模式中，对视频切换点  $t_s$  的检测间隔可以在  $t_s - T_E/5$  处开始，并持续到  $t_s + 4T_E/5$  结束，从而对切换点之前的讲话人给出一些加权，并对切换点之后将要讲话的人给出加权重值。  
20

音频计数器 (scorer) 45 在检测间隔开始时复位各音频输入。考虑到数字实施，其中每个音频输入都是一个脉冲编码调制的采样流，采样被认为成组。例如，利用以 8kHz 采样的音频流，可以认为采样为具有 50 个连续采样的组（例如，6.25 毫秒的时间子窗）。  
25

对于每个时间子窗和音频输入，可以确定最大和最小采样值。计算这两个采样值之间的差值，表示出（近似的）该子窗的音频输入的最大峰-峰幅度。

30 在该子窗的末端，将对该音频输入计算出的差值与对该子窗的其他音频输入计算出的差值进行比较。计算所得的差值最大的输入“赢得”该子窗，且它的音频计数加一。或者，利用具有最大计算差值的

计数对所有的计算差值进行归一化，然后使每个计数加上该归一化后的差值。

在检测间隔中对每个子窗连续进行该处理。在检测间隔的末端，各音频计数表示对应的音频输入具有最大的计算差值的子窗的数目。

5 在本实施例中，当前的视频片段可归类为常规视频片段和临时视频片段这两种类型中的一个。常规的视频片段是被选择作为音频计数加权比较的结果（马上将要说明）。临时视频片段是当由于前面的常规片段已经达到一定长度，需要采取其他措施来增加进行切换的可能性时被选择的片段。

10 块 68 根据当前片段类型进行切换。当片段类型为临时类型时，计数越过加权函数而执行块 72。但如果片段连续经过该循环，则意味着该片段的选择是在头对头非加权比较的计数情况下进行的，因此块 72 将片段类型从临时片段设为常规片段。

15 在当前视频片段已经是常规片段时，对该片段的音频计数在块 70 加权。对应该音频计数的加权是该片段长度的一个函数。

图 4 示出一个可能的加权函数。在新片段的开始处，该源的加权与所有其他源的加权相同。但是，随着片段长度不断增加，当前源加权逐步降低，从而增加了选择其他源的可能性。

20 虽然可以使用平滑的加权函数，但图 4 的阶梯式函数可根据基于片段长度的查询操作来确定加权。计算一个整数索引  $n$ ，其中  $T_L$  为当前片段长度， $T_p$  为最佳视频片段长度， $C$  为调节图 4 中的  $x$  比例的常数（这样衰减函数处理很快）：

$$n = \min(\lfloor C * T_L / T_p \rfloor, 8)$$

25 该索引  $n$  用于从阵列  $W$  中选择一个加权，该阵列  $W$  从 0 到 8 索引，其中  $W = \{1.0, 1.0, 0.9, 0.85, 0.8, 0.5, 0.375, 0.2, 0.1\}$ 。

30 在包括上述  $W$  值的一个示例结构中， $T_p=9$  且  $C=5$ 。对应当前片段的加权缓慢减小，直到该片段达到最佳片段长度（九秒），从阵列元素 0 到达 4 时为止。当片段达到九秒的长度时，其相关的音频计数将接收到一个加权  $W[5] = 0.5$ ，这极大的增加了当在第一个九秒中没有选择其他源时，选择其他源的可能性。

在当前源已持续了延长的时间周期时，块 74 可进一步增加另一个源被选择的可能性。在块 74，当前片段长度与多个优选的片段长度（这里设为 3.0）进行比较。在此例中，一旦该片段长度超过了因数为 3 的优选长度，则分支进行到块 76。在块 76，当前源的计数被乘以另一个加权 0.5。同样，其“将成为类型”（WBT）被设为临时类型。将成为类型是一个新的片段在此循环期间被选择时将被指定的片段类型。因此 WBT=临时表示在当前片段达到一个延长的长度后选择一个新的源，这个增加的长度是为了加入其他插入的源。注意若分支不进行到块 76，则到块 78 的分支设定 WBT=常规，表示不需块 76 的相乘就可选择任何新的源。

块 80 在当前检测间隔比较加权的音频计数。例如，通过将加权计数存储在索引的阵列  $v[i]$  中，块 80 可确定最大计数的索引  $i$ 。然后块 82 将最大计数的索引与当前视频源的索引进行比较。若两个索引相同，最后被选择的源已被再次选择，则当前片段继续。这样，块 84 增加当前片段的长度。

当最大计数的索引与当前视频源的索引不匹配时，开始一个新的源的新片段。块 86 将当前索引设定为最大计数的索引，将片段长度复位为 0，并将片段类型设为等于将成为类型。块 84 和 86 都返回到块 62 开始下一个循环。

根据本发明的实施例，还可实施音频和视频源的多种其他结构。例如，图 6 示出了使用三个基本摄像机（C1, C2, C3）和四个麦克风的结构 90。摄像机 C1 与两个麦克风 A1-L 和 A1-R 相联，这两个麦克风都在摄像机 C1 的视野中。这种相联可以采用多种方式。

一种将多个音频输入与一个视频源相联接的方法是单独地考虑将每个音频源并将最高的源（计数）作为相应视频源的原始计数。这就是对相应于一个视频源的音频源的矢量取无穷模方。也可使用其它的模方，例如 1-模方（两个源的平均）或是 2-模方（两个源的均方根）。

将多个摄像机与多个音频输入连接的另一种方法是规定一个相关矩阵。这样不仅可实现多个音频传感器与一个视频源相连，还可实现多个视频源与一个音频源相连。例如，在图 6 中示出了第四摄像机 C4。摄像机 C4 示出了整个会议桌的一个宽视角视图。下面是一个可能的相关矩阵：

	视频源	音频传感器			
		A1-L	A1-R	A2	A3
	C1	1.0	1.0	0.0	0.0
	C2	0.0	0.0	1.0	0.1
5	C3	0.0	0.0	0.0	1.0
	C4	0.2	0.2	0.1	0.1

通过这个矩阵，摄像机 C4 的宽视角场景可接收每个音频输入的计数的一部分，从而有可能在一个人讲话时间过长时该系统偶尔的选择宽视角场景。同样示出的是音频传感器 A3 和摄像机 C2 的较小加权。摄像机 C2 示出了环绕音频传感器 A3 区域的近距离视图，从而在 A3 附近的讲话者讲话时间过长时偶尔选择该视图作为一个变换的视图。

虽然之前的例子示出了一个单会议房间多摄像机的结构，但根据本发明的视频自动选择并不局限于这种应用。例如，图 7 示出了通过一个通信网络 96（即，一个电路交换网络，一个分组交换网络，或两者的结合）进行操作的实施例 92。

图 7 示出了三个“会议点” 100, 200, 300。这些点可表示一个三维的视频会议，但也可适用于其他应用，例如保卫，赌博，电子教室等。每个会议点包含至少一个摄像机（即 C1），音频传感器（即 A1）和显示器（即 D1），它们都与编解码器（coder）（即 102）相连。每个编解码器 102, 202, 302 通过网络 96 与视频自动选择器 94 进行通信。

一种操作的方法是，编解码器 100 将来自 C1 的视频和来自 A1 的音频编码，将编码的视频和音频流传送至视频自动选择器 94。编解码器 202 和 302 对他们的视频和音频流也进行相应的操作。每个编解码器可根据已知的视频和/或音频编码/压缩/打包算法进行操作。视频自动选择器 94 然后可使用音频流（和选择时间历史）来选择对每个会议点将提供哪个视频流。这种功能操作与上述的单个房间的例子类似，除了每个会议点可在任一给定时间接收与其他会议点不同的视频源。在自动选择器 94 中，切换逻辑电路的一个单独部分可与每个端点相匹配，每个部分都保持对该端点的选择时间历史。切换的视频流到达它们的目的端点，其中各个编解码器对它们解码并将它们显示在附加的显示器上。音频流可以被自动选择器 94 所混合或切换，或者每个端点

都可以从其它的端点接收到音频流。

操作可以在几个方面被改善。当自动选择器不进行视频流的实际处理时，他们可能不需要视频流通过自动选择器。相反，自动选择器 94 可以通知该端点何时（以及到其他哪个端点）提供输出视频流，并且该视频流可以通过网络 96 直接在端点之间传送。这种改进的一个优点是减少未被选择的视频源所占用的带宽。自动选择器 94 还可提供所需的桥接和转换功能，并可作为该会议的通用连接点。

图 7 的通用示例可以有多种方式的变形。该自动选择器可以以物理方式安置在其中一个端点处。某些端点可以只接音频，或者只是接收。一个端点还可以具有多个音频和/或视频传感器，这样，该端点的编码器可以编码多个码流，或者可以使用一个另外的自动选择器（例如上述单房间的例子）为该端点选择一个输出码流。

上述的实施例特别适用于可编程数字处理系统的实施，例如用于自动选择任务的专用数字信号处理器，或是编程执行自动选择任务的通用计算机。例如，在通用计算机中，音频输入可被提供至一个或多个与计算机相连的声卡。该声卡对音频信号进行数字化处理，并将数字化后的音频信号提供至计算机的中央处理单元和附加存储器。该计算机用于运行一个软件处理以恢复该数字音频信号，计算音频计数，并根据在先的描述进行周期性地视频切换决定（可选择的，音频计数或音频计数的一部分可在声卡上进行）。该软件处理产生视频切换信号送至，例如，执行视频切换的单独处理或装置，视频卡，或是作为网络信号送至另一个计算机。

根据可编程计算机的例子，本发明的一个方面包括一个装置，该装置包括一个包含计算机指令的计算机可读介质，当执行该指令时，可使一个或多个处理器根据自动选择一个视频源的方法而进行操作。例如，在上述的计算机例子中，软件处理可被存储在磁性或光学媒体上，半导体存储器中，或是可通过网络访问的远程位置。

虽然音频计数和视频源选择中包含有特定的计算方法，这些都只是举例。本领域的普通技术人员在阅读了本说明后可以实现能达到与本公开实施例类似功能的许多种变形。例如除了对当前视频源进行下降加权计算外，未被选择的源对于它们保持未选择的状态也可进行上升加权计算。临时片段与常规片段的最小片段长度可以是不同的，和/

或使用不同的加权函数。也可以混合应用其他原理，例如每 T 秒某个源必须被选择至少一次。也可以使用不同的原理应用于更复杂的场景，例如对于包含一个主要讲话人的交替切换和包含多个主讲话人，每人短时讲话的交替切换，使用不同的加权函数或检测间隔。

5 音频计数在系统和系统之间可能非常不同。所公开的计数方法，在抑制喧哗，和例如物体落地或咳嗽这样的瞬时噪音时可正常执行，但更复杂的计数方法也可被使用，包括语音识别和讲话者识别。音频计数的输入也可被简化——除了接收模拟或脉冲编码调制音频流的自动选择器，其也可接收其他信号。该其他信号可以是来自一个端点的  
10 周期性声音有效检测（VAD）信号，以执行自身的声音有效性检测。该其他信号还可以是，例如子带编码的音频信号。在后一种情况下，音频计数器还可检测子带中的能量分配以进行计数，而不需要重新创建 PCM 码流以对音频计数。

本领域的普通技术人员应当认识到这里所述的概念可以被实施为  
15 特定的例子应用于许多其它的有利方面。特别的，本领域的普通技术人员在阅读了本说明后应当认识到所描述的例子仅仅是大量变形实施例中的一个。音频流可以被嵌入在视频流中。音频计数器，和实施例的其他元件不需要进行数字化处理。视频流不需要与传统的模拟摄像机相关联，例如该码流可以是视频摄像机数字输入信号，数字化压缩  
20 的视频信号，记录的视频信号，计算机产生的视频信号等。这种微小的改变被包含在本发明中，并落在权利要求的保护范围之内。

前述例子都是示例性的。虽然说明书中多处使用“一”，“一个”，“另一个”，或“一些”实施例，这并不意味着每个这种引用都是指相同的实施例，或者该特征只用于单个实施例。

25



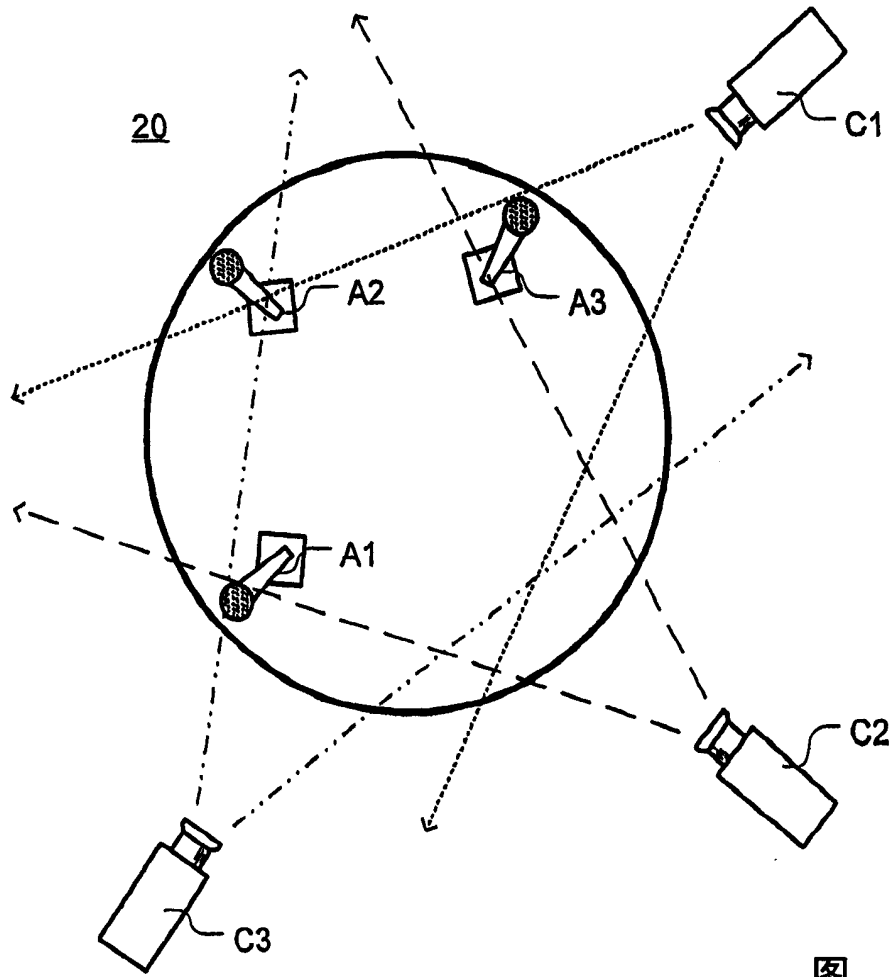


图 1

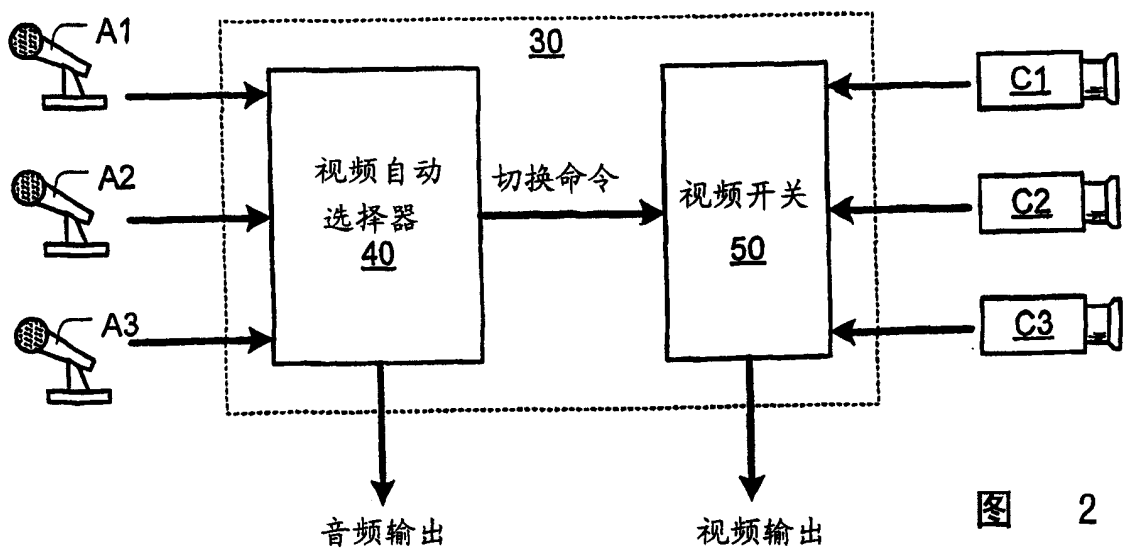


图 2

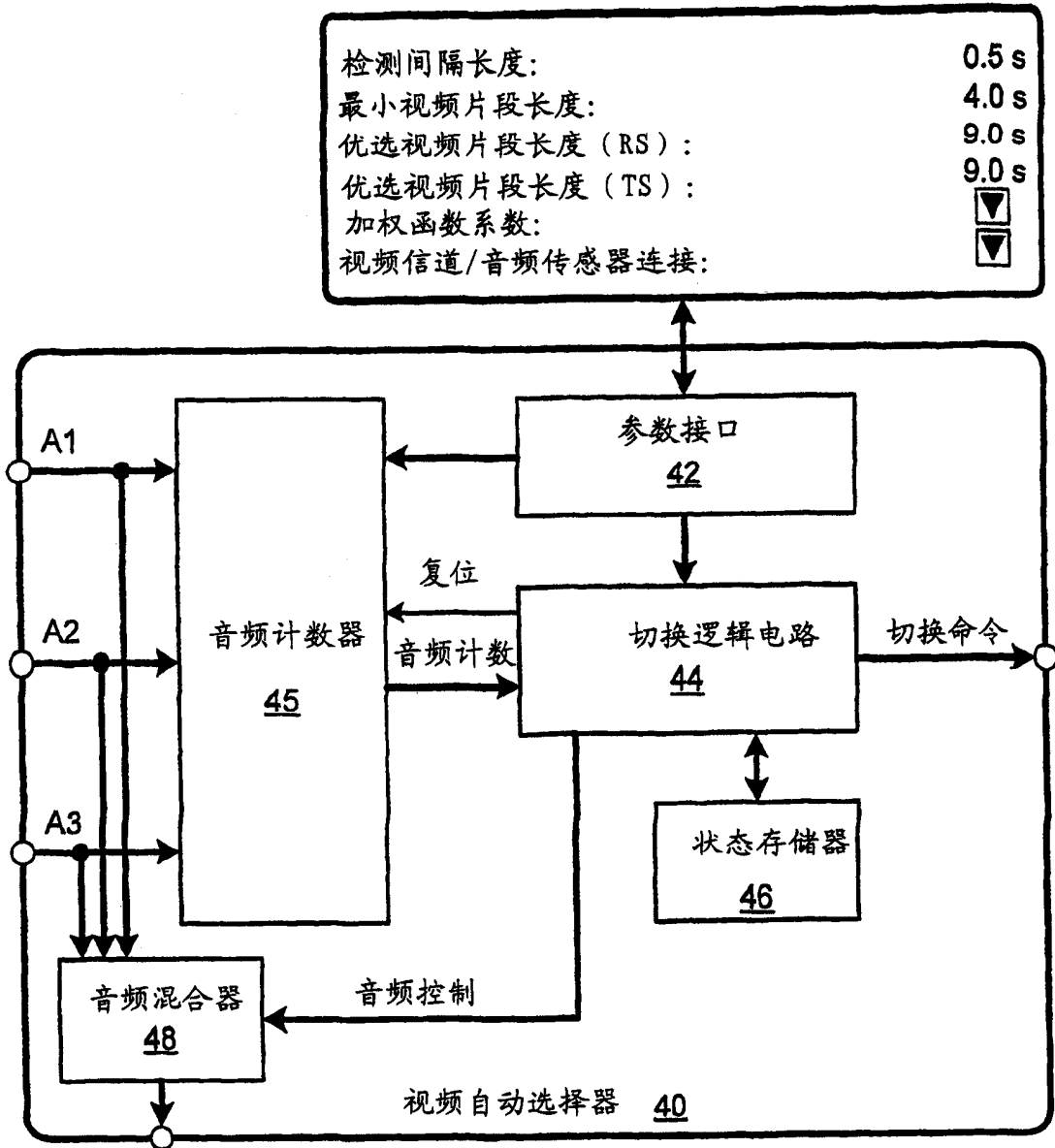


图 3

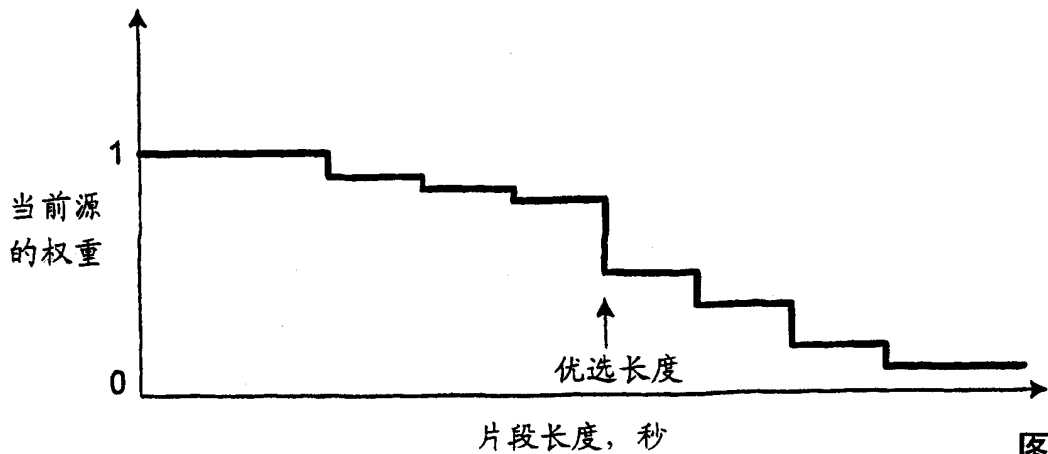


图 4

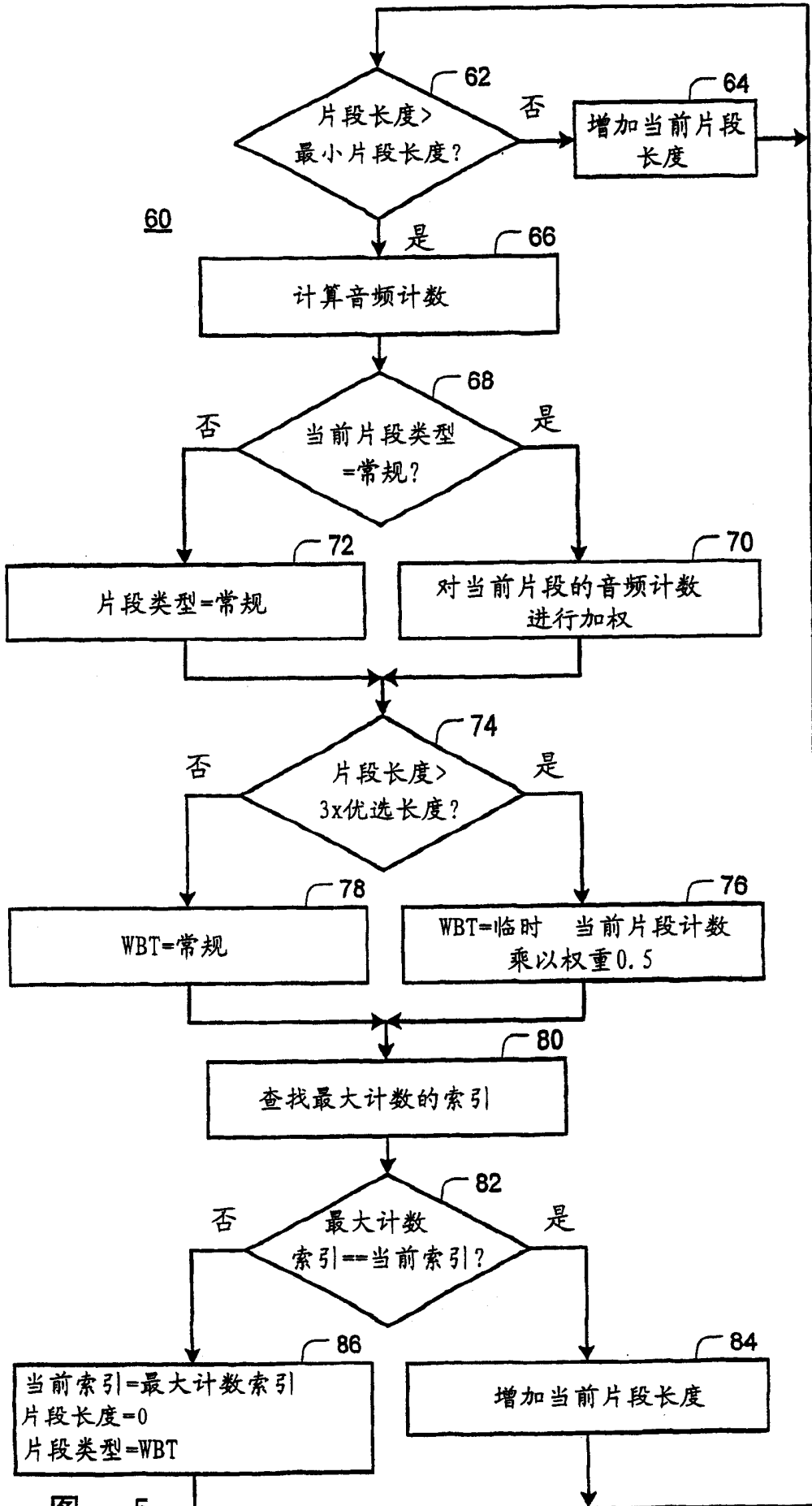


图 5

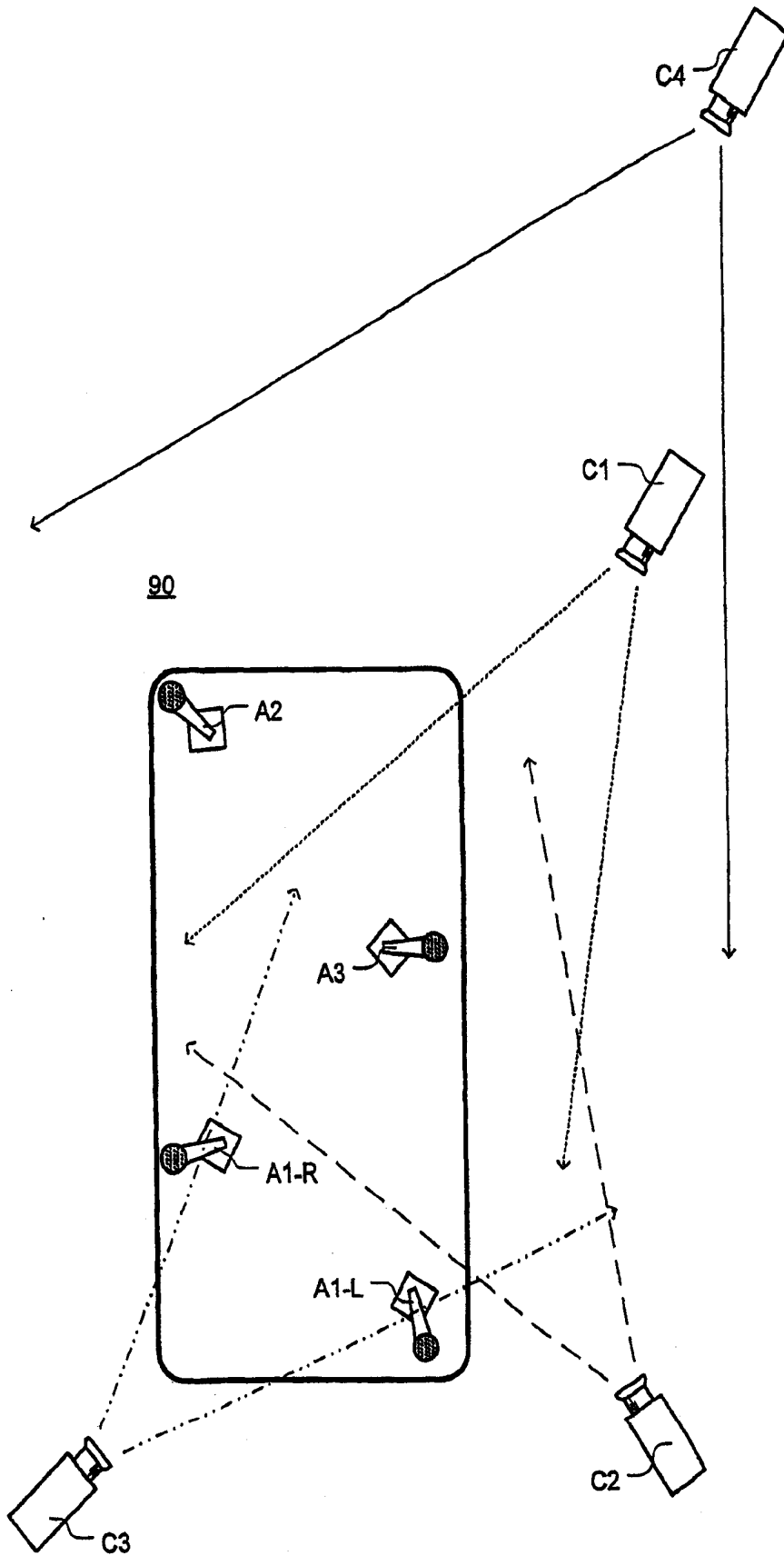


图 6

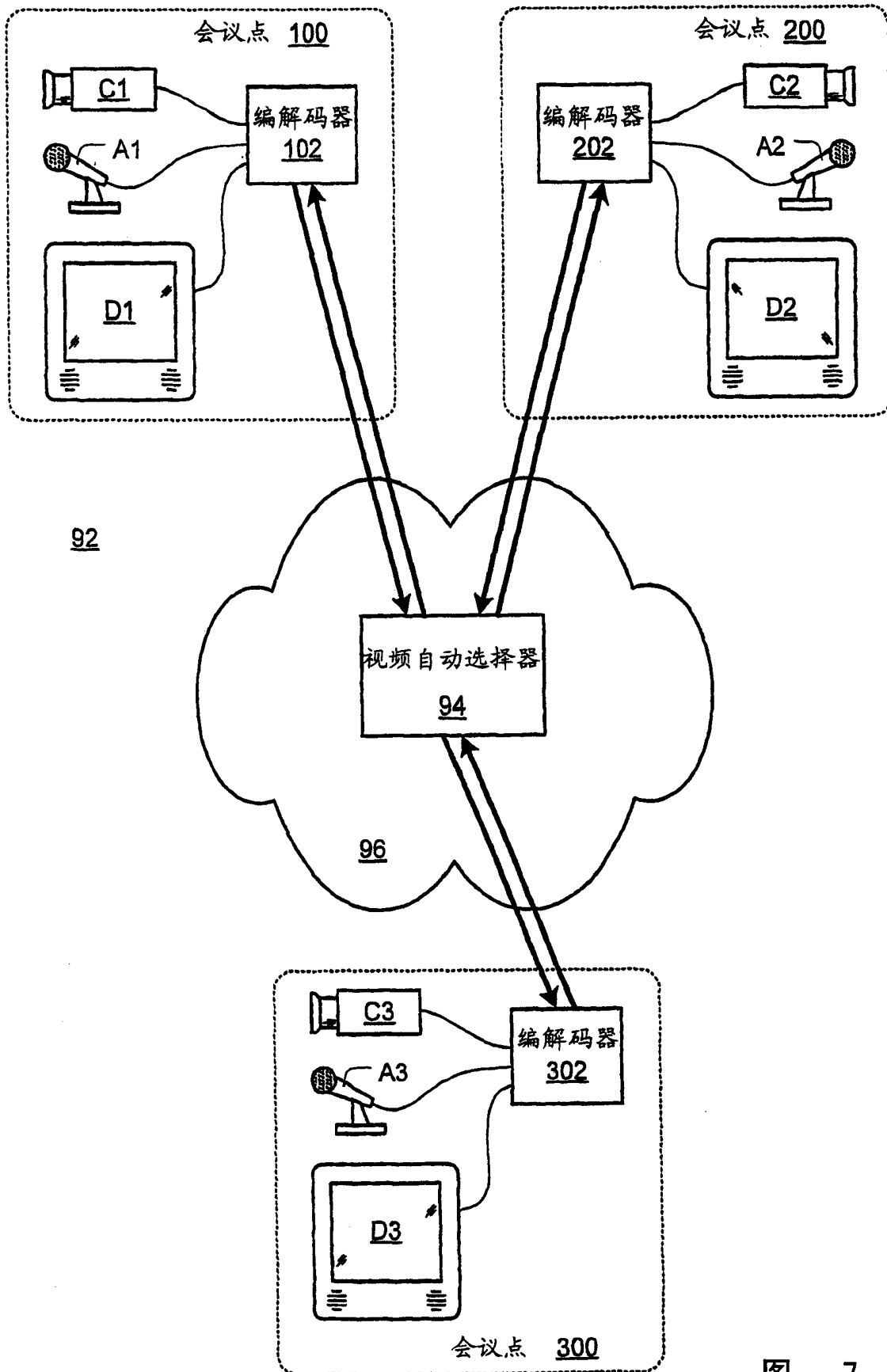


图 7