



(12) 发明专利申请

(10) 申请公布号 CN 101855897 A

(43) 申请公布日 2010. 10. 06

(21) 申请号 200880115993. X

(22) 申请日 2008. 11. 10

(30) 优先权数据

07120629. 6 2007. 11. 14 EP

(85) PCT申请进入国家阶段日

2010. 05. 13

(86) PCT申请的申请数据

PCT/IB2008/054691 2008. 11. 10

(87) PCT申请的公布数据

W02009/063383 EN 2009. 05. 22

(71) 申请人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

(72) 发明人 B·佐特考 P·方塞卡 L·王

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

代理人 李舒 谭祐祥

(51) Int. Cl.

H04N 5/14(2006. 01)

G06F 17/30(2006. 01)

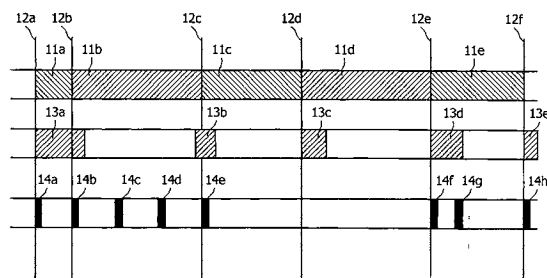
权利要求书 2 页 说明书 8 页 附图 4 页

(54) 发明名称

确定视听信号中的语义单元的起点的方法

(57) 摘要

一种确定片段(11)的起点(12)的方法,该片段对应于视听信号的语义单元,所述包括:处理该信号的音频成分以检测满足低音频功率标准的部分(14);以及处理该视听信号以识别对应于镜头的部分的边界。视听信号的视频成分被处理以评估用于识别由特定类型的至少一个镜头形成的视频部分的标准,该特定类型的至少一个镜头包括其中可能表示主持人的图像。如果满足低音频功率标准的部分(14)的至少一个终点位于所识别的视频部分(13)的边界之间的特定间隔处,则与满足低音频功率标准的部分(14)一致且位于所识别的视频部分的边界之间的点被选择为片段(11)的起点(12)。当确定没有满足低音频功率标准的部分与所识别的视频部分(13)一致时,该视频部分的边界被选择为片段(11)的起点(12)。



1. 一种确定片段 (11) 的起点 (12) 的方法, 该片段 (11) 对应于视听信号的语义单元, 包括:

处理信号的音频成分以检测满足低音频功率标准的部分 (14), 以及

处理视听信号以识别对应于镜头的部分的边界,

其中视听信号的视频成分被处理以评估用于识别由至少一个镜头形成的视频部分的标准, 该至少一个镜头满足用于识别包含图像的特定类型的镜头的标准, 主持人可能被表示在该图像中, 该视频部分仅仅包括该特定类型的镜头,

其中, 如果满足低音频功率标准的部分 (14) 的至少一个终点位于所识别的视频部分 (13) 的边界之间的特定间隔处, 则与满足低音频功率标准的部分 (14) 一致并位于所识别的视频部分 (13) 的边界之间的点被选择为片段 (11) 的起点 (12), 并且其中,

当确定没有满足低音频功率标准的部分与所识别的视频部分 (13c) 一致时, 选择该视频部分的边界作为片段 (11d) 的起点 (12d)。

2. 根据权利要求 1 的方法, 其中处理视听信号的视频成分包括: 评估用于识别所述特定类型的镜头的标准, 该评估包括确定镜头的至少一个图像是否满足与至少一个其它图像的相似度。

3. 根据权利要求 2 的方法, 其中评估用于识别所述特定类型的镜头的标准包括: 确定镜头的至少一个图像是否满足与该镜头中所包含的至少一个其它图像的相似度。

4. 根据权利要求 2 或 3 的方法, 其中评估用于识别所述特定类型的镜头的标准包括: 确定镜头的至少一个图像是否满足与至少一个其它镜头的至少一个其它图像的相似度。

5. 根据权利要求 4 的方法, 包括分析包含相似图像的镜头在视听信号上分布的均匀性。

6. 根据权利要求 1-5 中任一项的方法, 其中处理视听信号的视频成分包括: 评估用于识别所述特定类型的镜头的标准, 该评估包括分析该镜头中所包含的至少一个图像的内容以检测该镜头所包含的至少一个图像中表示的任意的人的面部。

7. 根据权利要求 1-6 中任一项的方法, 其中处理视听信号的视频成分以评估用于识别视频部分的标准包括下述的至少一个:

a) 确定镜头是否是连续的镜头序列的第一个, 该镜头的每一个被确定以满足用于识别所述特定类型的镜头的标准, 该特定类型的镜头包括其中可能表示主持人的图像, 且该序列具有比特定最小长度更大的长度, 以及

b) 确定镜头是否满足用于识别特定类型的镜头的标准, 该特定类型的镜头包括其中可能表示主持人的图像; 并且此外是否满足具有比特定最小长度更大的长度的标准。

8. 根据前述权利要求中任一项的方法, 包括: 当确定了满足低音频功率标准的多个部分 (14a, b, f, g) 的每一部分的至少一个终点位于所识别的视频部分 (13a, d) 的边界之间的特定间隔处时, 选择与所述多个部分 (14a, b, f, g) 中第一个出现的部分一致的点作为片段 (11a, e) 的起点 (12a, e)。

9. 根据权利要求 8 的方法, 进一步包括: 当至少确定第一与第二部分 (14a, b) 之间的间隔 (Δt_{ij}) 的长度超过特定阈值时, 选择与满足低音频功率标准的所述多个部分 (14a, b) 中的并在第一部分 (14a) 之后的第二部分一致的点作为另一个片段 (11b) 的起点。

10. 根据前述权利要求中任一项的方法, 包括: 对多个所识别的视频部分 (13) 的每一

个,连续地确定满足低音频功率标准的部分(14)的至少一个终点是否位于所识别的视频部分(13)的边界之间的特定间隔处。

11. 根据前述权利要求中任一项的方法,其中通过评估第一窗口上的平均音频功率相对于第二窗口上的平均音频功率来检测满足低音频功率标准的部分(14),该第二窗口大于该第一窗口。

12. 一种用于将视听信号分割为对应于语义单元的片段(11)的系统,该系统被配置为:

处理该信号的音频成分以检测满足低音频功率标准的部分(14),以及
处理该视听信号以识别对应于镜头的部分的边界,

其中视听信号的视频成分被处理以评估用于识别由至少一个镜头形成的视频部分(13)的标准,该至少一个镜头满足用于识别包含图像的特定类型的镜头的标准,主持人可能被表示在所述图像中,该视频部分仅仅包括该特定类型的镜头,并且其中该系统被布置成,

当确定满足低音频功率标准的部分(14)的至少一个终点位于所识别的视频部分(13)的边界之间的特定间隔处时,

选择与满足低音频功率标准的部分(14)一致并位于视频部分(13)的边界之间的点作为片段(11)的起点(12),并且其中

该系统被布置成当确定没有满足低音频功率标准的部分(14)与所识别的视频部分(13)一致时,选择该视频部分(13)的边界作为片段(11)的起点(12)。

13. 根据权利要求12的系统,被配置为执行根据权利要求1-11中任一项的方法。

14. 一种视听信号,被划分为对应于语义单元并具有由信号的配置指示的起点(12)的片段(11),该视听信号包括

音频成分,包含满足低音频功率标准的部分(14),以及

视频成分,包括视频部分,至少一个该视频部分满足用于识别由至少一个特定类型的镜头形成的视频部分的标准,该至少一个特定类型的镜头包括其中可能表示主持人的图像,并且至少一个该视频部分仅仅包含该特定类型的镜头,

其中满足低音频功率标准并具有位于满足所述标准的视频部分(13)的边界之间的特定间隔处的至少一个终点的至少一个部分(14)与片段(11)的起点(12)一致,并且其中

片段(11d)的至少一个起点(12d)与视频部分(13c)的边界一致,所述视频部分(13c)满足所述标准并与满足低音频功率标准的部分(14)不一致。

15. 根据权利要求14的视听信号,该视听信号可借助根据权利要求1-11中任一项的方法获得。

16. 一种计算机程序,其包括一组指令,当这组指令并入机器可读取介质中时能够使具有信息处理能力的系统执行根据权利要求1-11中任一项的方法。

确定视听信号中的语义单元的起点的方法

技术领域

- [0001] 本发明涉及一种确定片段的起点的方法,该片段对应于视听信号的语义单元。
- [0002] 本发明还涉及一种用于将视听信号分割为对应于语义单元的片段的系统。
- [0003] 本发明还涉及被划分成对应于语义单元并具有可识别起点的片段的视听信号。
- [0004] 本发明还涉及一种计算机程序。

背景技术

[0005] 2003 年 11 月 2-5 日在西安关于 Machine Learning and Cybernetics 的第二届国际会议论文集卷 5 的第 3008-3011 页 Wang, C. 等人的“Automatic story segmentation of news video based on audiovisual features and text information”涉及一种基于视听特征和文本信息的新闻故事自动分割方案,基本思想是:首先检测针对新闻视频的镜头边界,随后通过使用文本检测算法识别主题字幕帧以得到分割提示。在下一个步骤中,通过使用短时能量和短时平均零交叉率参数来检测静音片段。如果静音周期包含在连续的主题字幕开端之间,并且静音周期与镜头边界组的联合不是空的,则通过静音周期的半途的位置处的帧被选择为故事边界。如果连续的静音周期与主题字幕开端交替,并且静音周期与镜头边界组的联合是空的,则它示出新闻故事在一个主持人镜头内,并且在该故事周围没有镜头边界。连续的主题字幕开端对之间的最长静音周期被选择为故事边界。

[0006] 所述已知方法的问题在于,它依赖于静音周期的存在来确定故事边界。而且,为了使该方法起作用,必需按照顺序检测字幕。许多表示新闻项目的视听信号包括没有静音周期或字幕的新闻项目。

发明内容

[0007] 本发明的目的是提供一种用于在视听信号中相对精确地并在相对较大范围的新闻项目类型上检测具有与新闻项目的特性相似的特性的语义单元的起点的方法、系统、视听信号和计算机程序。

[0008] 该目的是通过根据本发明的确定片段的起点的方法实现的,该片段对应于视听信号的语义单元,该方法包括:

[0009] 处理信号的音频成分以检测满足低音频功率标准的部分,以及

[0010] 处理视听信号以识别对应于镜头的部分的边界,

[0011] 其中视听信号的视频成分被处理以评估用于识别由至少一个镜头形成的视频部分的标准,该至少一个镜头满足用于识别包括图像的特定类型的镜头的标准,主持人可能被表示在该图像中,该视频部分仅仅包括该特定类型的镜头,

[0012] 其中,如果满足低音频功率标准的部分的至少一个终点位于所识别的视频部分的边界之间的特定间隔处,则与满足低音频功率标准的部分一致并位于所识别的视频部分的边界之间的点被选择为片段的起点,并且其中,

[0013] 当确定没有满足低音频功率标准的部分与所识别的视频部分一致时,选择该视频

部分的边界作为片段的起点。

[0014] 镜头是在一个连续的运动期间真实或虚拟像机记录的邻接图像序列,该连续运动表示在一个场景中的时间和空间中连续的动作。低音频功率标准可以是相对于信号音频成分的其它部分的低音频功率标准、绝对标准或这两个标准的组合。尽管在此主要参照新闻广播来描述所述方法,但是由作为主持人 (compère) 的人引入的项目构成的其它类型的视听信号可以被类似地分割。

[0015] 当确定没有满足低音频功率标准的部分与满足用于识别该特定类型的镜头的标准镜头一致时,通过选择至少一个特定类型的适当的主持人镜头的边界作为片段的起点,这确保起点与满足用于识别合适的主持人镜头或主持人镜头的不中断序列的标准的部分相关联。因此,即使新闻项目不以静音开始或不包含静音,合适的主持人镜头的点仍将被识别为新闻项目的起点。因为与满足低音频功率标准的部分一致并位于所识别的视频部分的边界之间的点被选择为片段的起点,如果满足低音频功率标准的部分的至少一个终点位于所识别的视频部分的边界之间的间隔处,则起点被相对精确地确定。特别地,当新闻读者宣布桥接两个连续的新闻项目时,起点可被确切地确定。这是因为恰在新闻读者移到下一个新闻项目之前可能存在对应于低音频功率部分的暂停。上述效应是独立于视听信号中存在的主持人镜头的类型而实现的。定位合适的主持人镜头和满足低音频功率标准的部分是足够的。因此,该方法适于许多不同类型的新闻广播。

[0016] 在一个实施例中,处理视听信号的视频成分包括评估用于识别所述特定类型的镜头的标准,该评估包括确定镜头的至少一个图像是否满足与至少一个其它图像的相似度。

[0017] 一种效果是利用了主持人镜头的特性,该特性是主持人镜头在新闻广播的整个过程中是相对静止的,不必依赖于对任意特定类型的内容的检测。因此,该方法适合于在宽范围的新闻广播中使用,不管背景的类型、副标题或标志的存在或主持人镜头的其它特性,还包括该主持人如何显示(全身,在桌子或讲台之后,等等)。

[0018] 在一个变型中,评估用于识别特定类型的镜头的标准包括:确定镜头的至少一个图像是否满足与该镜头中所包含的至少一个其它图像的相似度。

[0019] 该变型利用下述事实:主持人镜头是相对静止的。该主持人一般是固定不动的,并且背景不会改变太多。

[0020] 在一个变型中,评估用于识别特定类型的镜头的标准包括:确定镜头的至少一个图像是否满足与至少一个其它镜头的至少一个其它图像的相似度。

[0021] 该变型利用下述事实:来自特定源的节目中的不同主持人镜头彼此在很大程度上相似。具体地,节目主持人一般是同一个人并且一般被表示在同一位置,背景相同。

[0022] 该方法的一个实施例包括分析包含相似图像的镜头在视听信号上分布的均匀性。

[0023] 广播中的项目趋于相似的长度,使得主持人镜头在节目中应当相对均匀地分布。彼此相似但不重新出现的邻接镜头将趋于成为相同的单个语义单元的部分而不是主持人镜头。

[0024] 在一个实施例中,处理视听信号的视频成分包括:评估用于识别特定类型的镜头的标准,该评估包括分析该镜头中所包含的至少一个图像的内容以检测该镜头所包含的至少一个图像中表示的任意的人的面部。

[0025] 该实施例在检测宽范围广播中的主持人镜头方面是较有效的。它对文化差异相对

不关心,因为在几乎所有广播文化中主持人的面部在主持人镜头中是显著的。

[0026] 在一个实施例中,处理视听信号的视频成分以评估用于识别视频部分的标准包括下述的至少一个:

[0027] a) 确定镜头是否是连续的镜头序列的第一个,该镜头的每一个被确定满足用于识别特定类型的镜头的标准,该特定类型的镜头包括其中可能表示主持人的图像,该序列具有比特定最小长度更大的长度,以及

[0028] b) 确定镜头是否满足用于识别特定类型的镜头的标准,该特定类型的镜头包括其中可能表示主持人的图像;并且此外是否满足具有比特定最小长度更大的长度的标准。

[0029] 该实施例在增加识别与节目主持人的一个引导对应的视听信号的一部分的全部的机会方面是有效的。特别地,在发生快速变回到节目主持人或两个节目主持人之间的地方,这些不被错误地识别为对新项目(例如新的新闻项目)的引导,而被识别为对一个特定新闻项目的引导的继续。

[0030] 该方法的一个实施例包括:当确定了满足低音频功率标准的多个部分的每一部分的至少一个终点位于所识别的视频部分的边界之间的特定间隔处时,选择与该多个部分中第一个出现的部分一致的点作为片段的起点。

[0031] 一种效果是:在主持人镜头或主持人镜头的背靠背序列内存在一个项目时,该项目的起点也被相对可靠地确定。

[0032] 一个变型进一步包括:当至少确定第一与第二部分之间的间隔的长度超过特定阈值时,选择与满足低音频功率标准的多个部分中的并在第一部分之后的第二个部分一致的点作为另一个片段的起点。

[0033] 因此,在主持人镜头或主持人镜头的不中断序列内存在一个项目并且在相同的主持人镜头或主持人镜头的不中断序列内开始下一个项目时,实现项目的分割而没有错过任何起点。

[0034] 该方法的一个实施例包括:对多个所识别的视频部分的每一个,连续地确定满足低音频功率标准的部分的至少一个终点是否位于所识别的视频部分的边界之间的特定间隔处。

[0035] 一种效果是视听信号被相对有效地分割,因为下一项目的起点一般是前一项目的终点。因此,连续地处理主持人镜头(在该方法中片段的至少一个起点被确定为与每个主持人镜头一致)是实现完全分割成视听信号的语义单元的高效方式。

[0036] 在该方法的一个实施例中,通过评估第一窗口上的平均音频功率相对于第二窗口上的平均音频功率来检测满足低音频功率标准的部分,其中第二窗口大于第一窗口。

[0037] 一种效果是“静音周期”相对于背景音频水平而被确定。因此,例如在主持人暂停同时背景主题正播放时,或在主持人镜头为位置上的主持人时,广播中的暂停被可靠地识别。

[0038] 根据另一个方面,根据本发明的用于将视听信号分割为对应于语义单元的片段的系统被配置为:处理该信号的音频成分以检测满足低音频功率标准的部分,以及

[0039] 处理该视听信号以识别对应于镜头的部分的边界,

[0040] 其中视听信号的视频成分被处理以评估用于识别由至少一个镜头形成的视频部分的标准,该至少一个镜头满足用于识别包含图像的特定类型的镜头的标准,主持人可能

被表示在图像中,该视频部分仅仅包括该特定类型的镜头,并且其中该系统被布置成,

[0041] 当确定满足低音频功率标准的部分的至少一个终点位于所识别的视频部分的边界之间的特定间隔处时,

[0042] 选择与满足低音频功率标准的部分一致的并位于视频部分的边界之间的点作为片段的起点,并且其中该系统被布置成当确定没有满足低音频功率标准的部分与所识别的视频部分一致时,选择视频部分镜头的边界作为片段的起点。

[0043] 在一个实施例中,该系统被配置为执行根据本发明的方法。

[0044] 根据另一个方面,根据本发明的视听信号被划分为对应于语义单元并具有由信号的配置指示的起点的片段,并且包括

[0045] 音频成分,包含满足低音频功率标准的部分,以及

[0046] 视频成分,包括视频部分,至少一个该视频部分满足用于识别由至少一个特定类型的镜头形成的视频部分的标准,该至少一个特定类型的镜头包括其中可能表示有主持人的图像,并且至少一个该视频部分仅仅包含该特定类型的镜头,

[0047] 其中满足低音频功率标准并具有位于满足用于识别特定类型镜头的标准的镜头的边界之间的特定间隔处的至少一个终点的至少一个部分与片段的起点一致,并且其中

[0048] 片段的至少一个起点与满足该标准并与满足低音频功率标准的部分不一致的视频部分的边界一致。

[0049] 在一个实施例中,视听信号可借助根据本发明的方法获得。

[0050] 根据本发明的另一方面,提供一种计算机程序,其包括一组指令,当这组指令并入机器可读取介质中时能够使具有信息处理能力的系统执行根据本发明的方法。

附图说明

[0051] 现在将参照附图更详细描述本发明,在附图中:

[0052] 图 1 是具有硬盘存储设备的综合接收解码机的简化框图;

[0053] 图 2 是示出视听信号的部分的示意图;

[0054] 图 3 是确定视听信号中新闻项目的起点的方法的流程图;以及

[0055] 图 4 是示出图 3 所示方法的细节的流程图。

具体实施方式

[0056] 综合接收解码机 (IRD) 1 包括网络接口 2、解调器 3 和解码器 4,用于接收数字电视广播、视频点播服务等等。网络接口 2 可连到基于数字、卫星、陆地或 IP 的广播或窄播网络。解码器的输出包括一个或多个节目流,其包括例如 MPEG-2 或 H. 264 或类似格式的(压缩的)数字视听信号。对应于节目或事件的信号可以存储在例如硬盘、光盘或固态存储设备等大容量存储设备 5 上。

[0057] 存储在大容量存储设备 5 上的视听数据可以被用户访问以便在电视系统(未示出)上回放。为此,IRD 1 设有用户接口 6,例如电视系统的屏幕上显示的遥控和图形菜单。IRD 1 由中央处理器 (CPU) 7 控制,中央处理器使用主存储器 8 执行计算机程序代码。为了回放和显示菜单,IRD 1 进一步设有视频编码器 9 和音频输出级 10,用于生成适合于电视系统的视频和音频信号。CPU 7 中的图形模块(未示出)生成由 IRD 1 和电视系统提供的图

形用户接口 (GUI) 的图形成分。

[0058] 尽管广播提供者将会把节目流分割为多个事件并且包含用于识别这些事件的辅助数据,但是这些事件一般将对应于完整的节目,例如完整的新闻节目,其在这里将被用作一个实例。

[0059] 越来越多的新闻节目在电视和因特网上广播。几乎每个频道具有它自己的每日新闻播报,并且许多专用的新闻频道也已经变得可获得。大量可获得的内容使用户几乎不可能观看所有这些内容。而且,大部分新闻项目(与单独的主题相关的新闻节目内的单独语义单元)通常从较早的新闻节目开始重复。如果用户最近已经观看了新闻节目,则他自然可能对再次观看相同的新闻项目不感兴趣。用户一般也不会对观看所有可获得的新闻项目感兴趣。

[0060] IRD 1 被编程以执行这样的例行程序:其使得它能够得到完整的新闻节目(例如,在节目流中识别的)并且检测该节目中新的新闻项目的起点,由此使得新闻节目能够分割为单独的语义单元,该语义单元比设有表示节目的视听数据的辅助数据中识别的语义单元小。

[0061] 图 2 是显示新闻广播的部分的示意性时间线。视听信号的片段 11a-e 对应于单独的新闻项目,并且被示出在表示地面实况(groundtruth)的上部时间线中。边界 12a-f 表示每下一个新闻项目的起点,其对应于前面的新闻项目的终点。

[0062] 视听信号的视频成分包括对应于图像或半像的视频帧(例如 MPEG-2 或 H. 264 视频帧)的序列。邻接帧的分组对应于镜头。在该上下文中,镜头是在一个连续的移动期间真实或虚拟像机记录的邻接图像序列,并且所述图像序列中每个图像序列表示场景中在时间和空间中的连续动作。在镜头之中,一些镜头表示一个或多个新闻读者,并且在图 2 中被表示为主持人镜头 13a-e。主持人镜头被检测并且被用于确定片段 11 的起点 12,如下面将会解释。

[0063] 视听信号的音频成分包括其中具有相对较低强度的音频信号的部分,在这里被称为静音周期 14a-h。这些静音周期也被 IRD 1 使用以确定对应于新闻项目的视听信号的片段 11 的起点 12。

[0064] 参照图 3 和 4,当被提示分割对应于新闻节目的视听信号时,IRD1 获得对应于该视听信号的数据(步骤 15)。随后,进行对静音周期 14 的定位(步骤 16)并且识别镜头边界(步骤 17)。当然,与存在的新闻项目相比存在更多的镜头,因为一个新闻项目一般由多个镜头组成。这些镜头被分类为主持人镜头和其它镜头(步骤 18)。

[0065] 在一个实施例中,定位静音周期的步骤 16 包括在短时间窗口上将音频信号强度与对应于绝对值的阈值(例如预定值)比较。在另一个实施例中,在第一移动窗口上的平均音频功率与在以与第一窗口相同速率处理的第二窗口上的平均音频功率的比被确定。第二窗口大于第一窗口,即它对应于视听信号的音频成分的更大部分。实际上,针对长周期的步行平均(例如对应于以正常呈现速度下的 20 秒)与针对短周期(例如 1 秒)的步行平均相比较。当长期与短期平均的比在比第二阈值长的间隔上大于阈值(例如 10)时,认为静音周期 14 已被检测到。第二阈值足够高以确保仅仅显著的暂停被归类为静音周期并且是低音频功率标准的一部分。在一个实施例中,仅仅确定了例如 1-5kHz 的特定频率范围内的音频功率。

[0066] 例如,识别镜头的步骤 17 可以包括识别视频信号的视频成分中的突然转变或者分析由视频编码标准定义的特定类型的视频帧出现的顺序。该步骤 17 也可以与后续步骤 18 组合,使得仅仅检测主持人镜头。在这样的组合的实施例中,相邻的主持人镜头可以合并为一个镜头。

[0067] 分类镜头的步骤 18 包括评估用于识别包含视频帧的镜头的标准,在该视频帧中可能出现一个或多个主持人。该标准可以是包括若干子标准的标准。在该步骤 18 中执行下列评估中的一个或多个。

[0068] 首先,IRD 1 可以确定在考虑中的镜头的至少一个图像是否满足与相同镜头中所包含的至少一个其它图像(更具体地,均匀分布在该镜头上的一组图像)的相似度。这用于识别相对静止的镜头。相对静止的镜头一般对应于主持人镜头,因为主持人或多个人在进行他们的广播的同时不会移动太多,而且所捕获的他们的图像的衬托背景也不会变化太多。

[0069] 其次,IRD 1 可以确定在考虑中的镜头的至少一个图像是否满足与在新闻节目中的多个其它镜头(例如所有后续镜头)中每一个的至少一个图像的相似度。如果该镜头相似于多个其它镜头的每一个并且这些相似的其它镜头被分布为使得它们的分布超过分布的均匀度的阈值,则该镜头(和这些其它镜头)被确定为对应于主持人镜头 13。

[0070] 可以例如通过分析所述镜头中包含的选定图像的色彩直方图的平均来确定镜头的相似性。可替代地,可以通过分析每个镜头的选定的一个或多个图像的特定空间频率分量的时间发展,并且随后比较这些发展以确定相似镜头,由此确定相似性。也可以使用镜头特征(如在镜头期间像素变化的量或镜头中存在的移动量)来确定镜头中所包含的图像彼此有多相似。其它相似度也是可能的,并且它们可以单独地或以组合方式应用以确定在考虑中的镜头与其它镜头有多相似,或镜头中所包含的图像彼此有多相似。

[0071] 分布的均匀度可以是相似镜头之间的时间间隔的标准偏差,或相对于该时间间隔的平均长度的标准偏差。其它度量也是可能的。

[0072] 第三,可替代地或此外对相似性的评估,考虑中的镜头中所包含的单独图像的内容可被分析以确定它是否是主持人镜头。特别地,可以执行前景/背景分割,以便典型地针对主持人镜头的特定类型的元素的存在来分析图像。例如,可以执行面部检测和识别算法。所检测到的面部可以与存储在大容量存储设备 5 中的已知主持人的数据库进行比较。在另一个实施例中,从新闻节目的多个镜头中提取面部。聚类算法用于识别在整个新闻节目中复现的那些面部。包含超过预定数量的一个或多个图像(其中表示了复现的面部)的那些镜头被确定以对应于主持人镜头 13。

[0073] 该步骤 18 的所有上述变型可以在帧上、半像上执行,而不是在图像上执行。

[0074] 据观察,用于识别主持人镜头的标准可被限制为仅仅是一个或多个特定类型的主持人镜头。特别地,该标准可以包括拒绝很短(例如短于 90 秒)的镜头。可以应用其它类型的过滤器。

[0075] 在主持人镜头 13 已被识别并且静音周期 14 被定位之后,启发式逻辑用于确定对应于新闻项目的片段 11 的起点 12。镜头且特别是主持人镜头 13 被连续地处理,因为一个片段 11 的起点 12 是前一个片段 11 的终点,使得对至少主持人镜头 13 的连续处理是最有效的。

[0076] 至少一个起点 12 与每个主持人镜头 13 相关联,不管在该主持人镜头 13 期间是否出现任何静音周期 14。事实上,如果确定了没有对应于静音周期 14 的音频成分的部分具有位于主持人镜头 13 的边界内的间隔上的至少一个终点,则这个主持人镜头 13 的起点被识别为片段 11 的起点 12(步骤 19)。因此,如果在主持人镜头 13 期间没有检测到静音,例如因为恰在主持人镜头 13 之前出现了静音周期,则该新闻项目在主持人镜头 13 的开端处被分割。例如,图 2 中的第三主持人镜头 13c 与静音周期 14 不重叠,并且因此它的起点被识别为第四片段 11d 的起点 12d。

[0077] 如果仅仅一个静音周期 14 具有位于主持人镜头 13 的边界内的间隔上的至少一个终点,则与静音周期 14 一致的点被选择(步骤 20)作为片段 11 的起点 12。该点可以是静音周期 14 的起点或对应于静音周期 14 的间隔上某处(例如半途通过)的点。扩展到下一个镜头的静音周期 14 在所示的实施例中未被考虑。事实上,主持人镜头 13 的边界之间的间隔(静音周期 14 的至少该终点必须位于该间隔上)一般地在与主持人镜头 13 的端部边界有些距离处结束,例如在 5 秒与 9 秒之间或在镜头长度的 75%处结束。然而,在所示的实施例中,该间隔对应于完整的主持人镜头 13。使用所示的启发式方法,与图 2 中的第二主持人镜头 13b 一致的第五静音周期 14e 被识别为第三片段 11c 的起点 12c。

[0078] 如果确定了多个静音周期 14 具有位于在考虑中的主持人镜头 13 的边界之间的间隔上的至少一个终点(图 4),则与这些静音周期中的第一个出现的静音周期一致的点被选择为片段的起点(步骤 21)。因此,在图 2 中,第一静音周期 14a 和第二静音周期 14b 都与第一主持人镜头 13a 一致。第一镜头周期 14a 被选择为第一片段 11a 的起点 12a。相似地,第六静音周期 14f 和第七静音周期 14g 具有至少一个位于第四主持人镜头 13d 的边界内的间隔上的终点。与第六静音周期 14f 一致的点被选择为第五片段 11e 的起点 12e。

[0079] 情况可能是这样的:新闻项目完全地包含在单个主持人镜头 13 的边界内。该主持人一般将会会在新闻项目之间暂停,或者在该点可能发生两个主持人之间的交接。在任一种情况中,将存在短静音。IRD 1 确定在考虑中的主持人镜头 13 的总长度 Δt_{shot} (步骤 22)。IRD 1 还确定在主持人镜头 13 期间出现的第一个静音周期与下一个静音周期之间的每个间隔 Δt_{ij} 的长度(步骤 23)。如果这些间隔 Δt_{ij} 的任意一个的长度超过特定阈值,则在第一间隔端部处超过阈值的静音周期是另一个片段 11 的起点 12。该阈值可以是主持人镜头 13 的总长度 Δt_{shot} 的一部分。在所示的实施例中,如果静音周期之间的间隔 Δt_{ij} 的任意一个的长度超过第一阈值 Th_1 并且主持人镜头 13 的总长度 Δt_{shot} 超过第二阈值 Th_2 ,则仅仅选择另一个起点(步骤 24)。这些步骤 23、24 可以通过从与第二起点一致的静音周期计算间隔长度来重复执行,从而找到在考虑中的主持人镜头 13 内的第三起点,等等。参照图 2,第一静音周期 14a 和第二静音周期 14b 均与第一主持人镜头 13a 一致。第二静音周期 14b 被选择为第二片段 11b 的起点 12b,因为第一主持人镜头 13a 足够长并且第一静音周期 14a 与第二静音周期 14b 之间的间隔也足够长。相反地,第六静音周期 14f 与第七静音周期 14g 之间的间隔太短并且 / 或者第四主持人镜头 13d 太短。

[0080] 从图 2 中显见,第三和第四静音周期 14c、d(其没有与主持人镜头 13 的边界之间的间隔上的点一致的至少一个终点)不被选择为对应于新闻项目的片段 11 的起点 12。

[0081] 通过对与新闻项目对应的片段 11 的起点 12 的位置的确定,视听信号可以被索引以允许快速访问特定新闻项目,例如通过存储表示起点 12 的数据,该数据与包括视听数据

的文件相关联。可替代地,该文件可以被分割为单独的文件以便分开处理。在任一情况下,IRD 1 能够向用户提供更多的个性化新闻内容,或至少允许用户在按该方式分割的新闻节目的内部导航。例如,IRD 1 能够为用户呈现一种略过用户不感兴趣的那些新闻项目的简单方式。可替代地,该设备可以为用户呈现对新闻节目中存在的所有项目的快速概览,并且允许用户选择他或她感兴趣的那些项目。

[0082] 应当注意,上面描述的实施例说明而非限制本发明,并且本领域技术人员将能够设计许多可替代实施例,而不脱离所附权利要求的范围。在权利要求中,括号之间的任何附图标记不应当被解释为限制该权利要求。动词“包括”及其变化形式的使用不排除权利要求中叙述的元件或步骤之外的元件或步骤的存在。元件之前的冠词“一”或“一个”不排除多个这样的元件的存在。本发明可以借助包含若干不同元件的硬件实现,并且借助适当编程的计算机实现。在列举了若干装置的设备权利要求中,这些装置中的若干个可以由同一项硬件体现。在相互不同的从属权利要求中记载某些措施这个纯粹事实并不表示这些措施的组合不能被有利地使用。

[0083] 尽管已经描述了一种使用 IRD 1 的实现方式,但是这里所概述的方法可以容易地在个人或手持计算机、数字电视机或类似设备上实现。

[0084] 如本领域技术人员将会明白的,“装置”的意思是包括在操作中执行或被设计以执行指定功能的任何硬件(比如单独或集成电路或电子元件)或软件(比如程序或程序的一部分),无论该功能单独地还是与其它功能结合,无论孤立方式还是与其它元件协作。“计算机程序”将被理解为意思是存储在计算机可读取介质(比如光盘)上的、可经由网络(比如因特网)下载的或以任何其它方式可销售的任何软件产品。

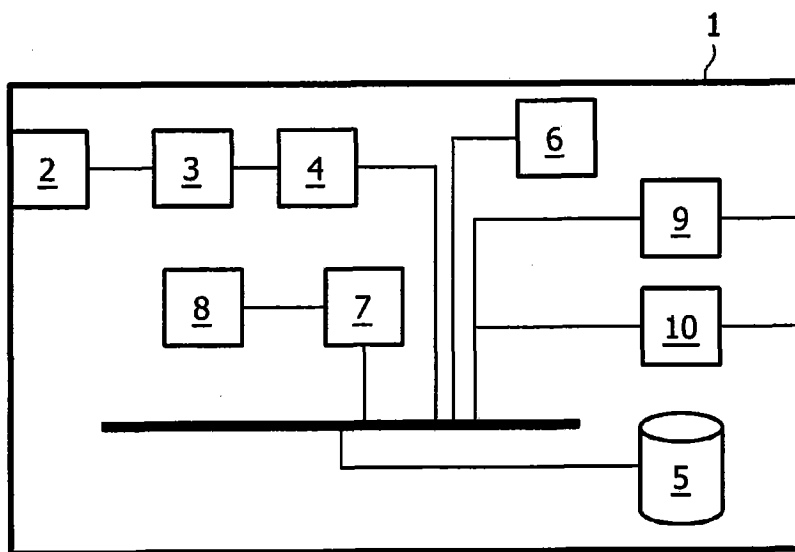


图 1

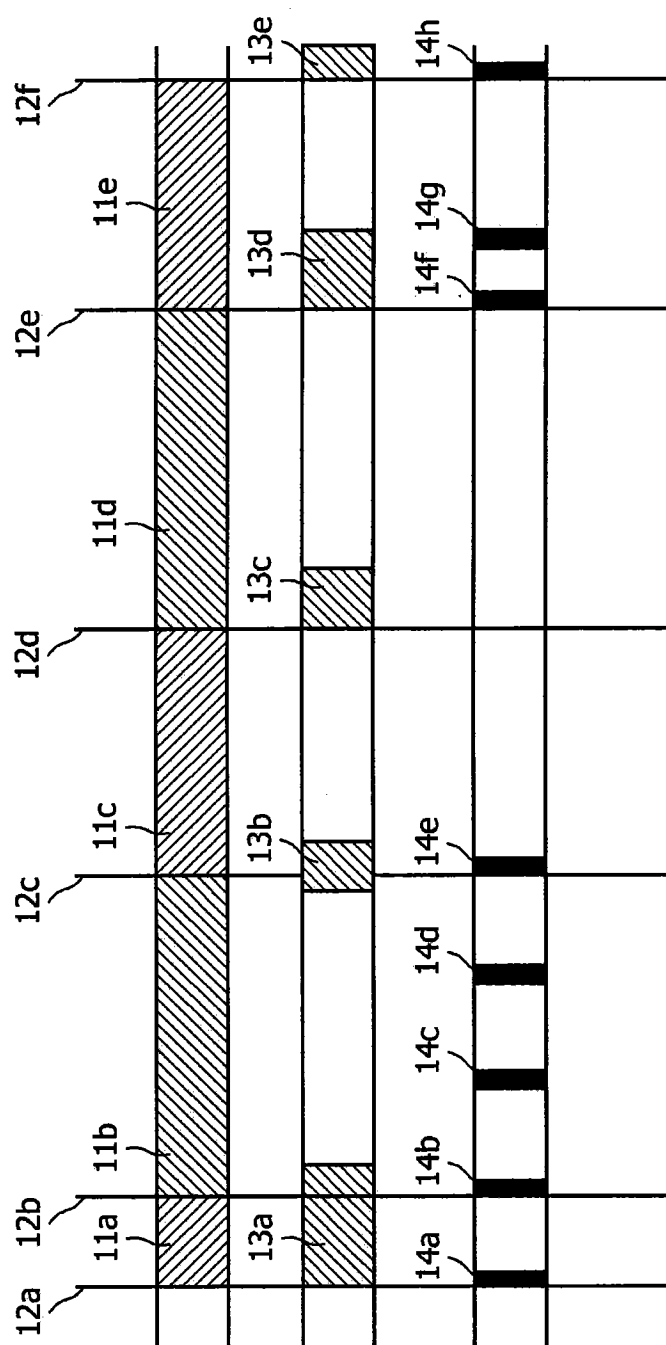


图 2

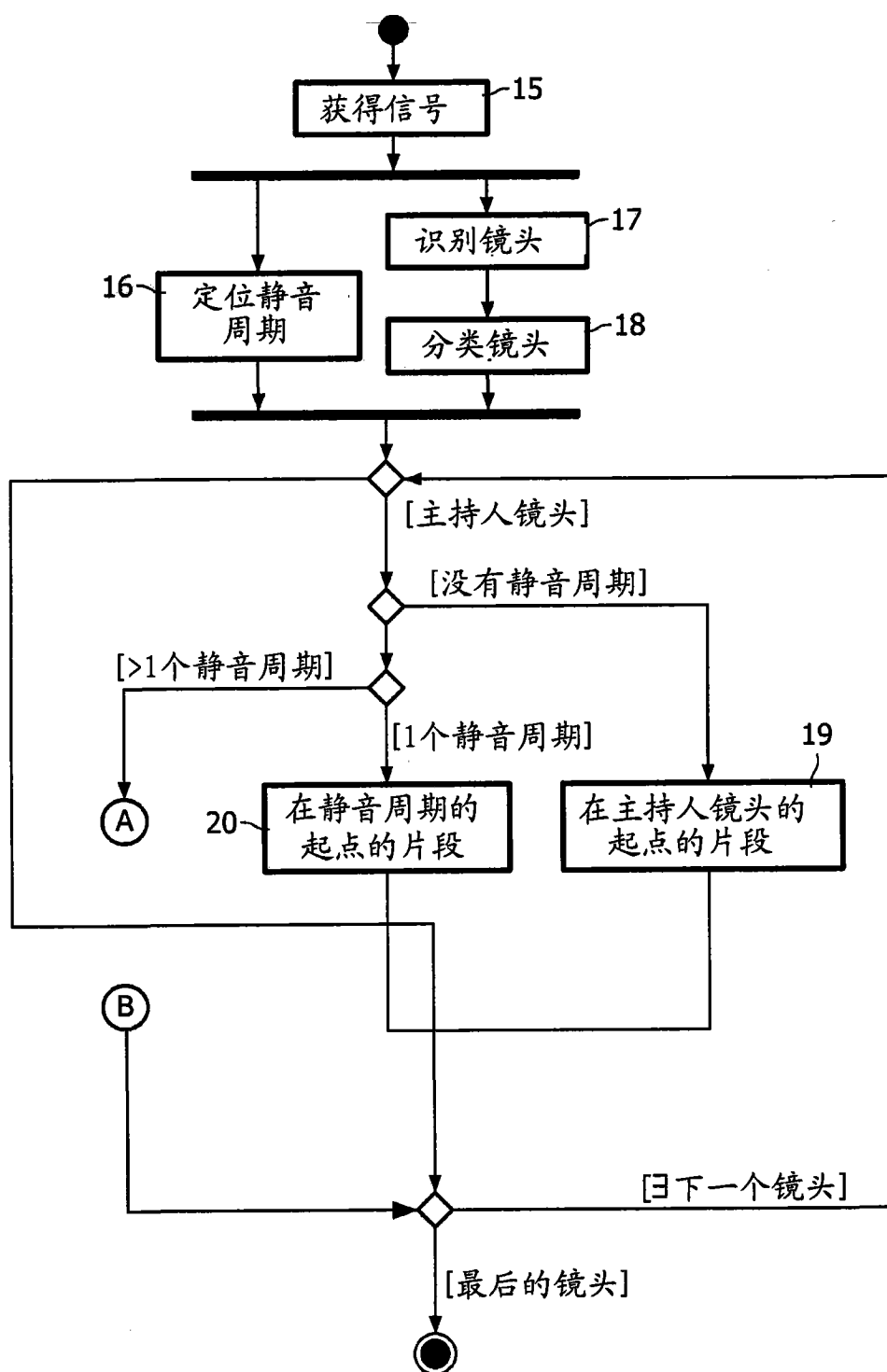


图 3

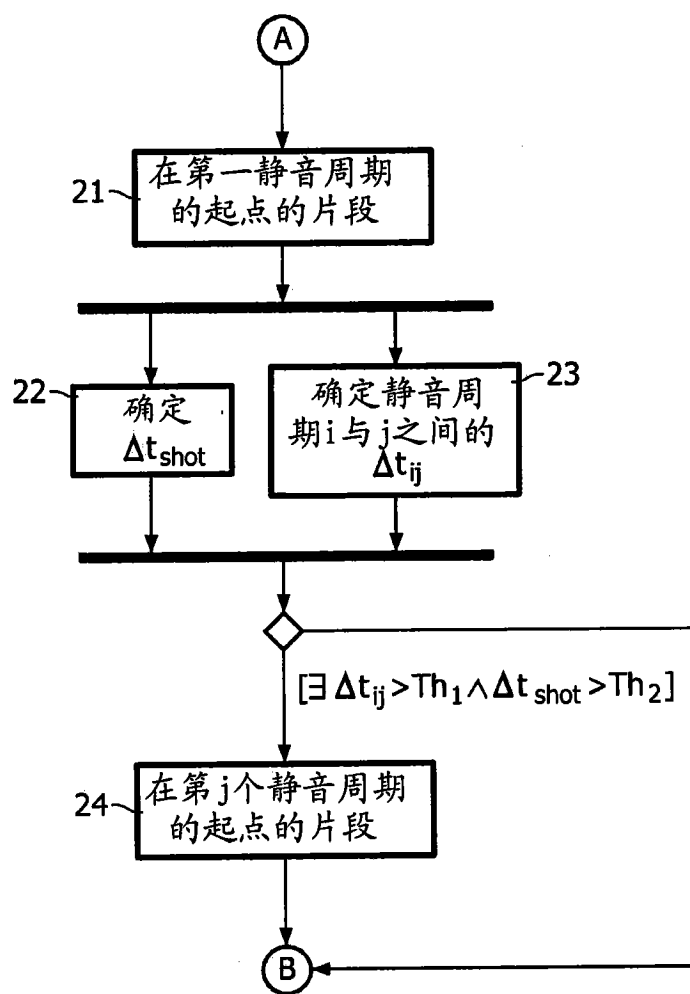


图 4