

(12) 发明专利申请

(10) 申请公布号 CN 102422288 A

(43) 申请公布日 2012. 04. 18

(21) 申请号 201080021306. 5

(74) 专利代理机构 北京律盟知识产权代理有限
责任公司 11287

(22) 申请日 2010. 05. 20

代理人 孟锐

(30) 优先权数据

12/473, 321 2009. 05. 28 US

(51) Int. Cl.

G06F 17/30(2006. 01)

(85) PCT申请进入国家阶段日

2011. 11. 14

(86) PCT申请的申请数据

PCT/US2010/035523 2010. 05. 20

(87) PCT申请的公布数据

W02010/138367 EN 2010. 12. 02

(71) 申请人 贺利实公司

地址 美国佛罗里达州

(72) 发明人 罗伯特·麦克唐纳

克里斯托弗·T·丹凯尔

约翰·海明豪斯 阿里奇·彼得森

塔里克·巴基尔

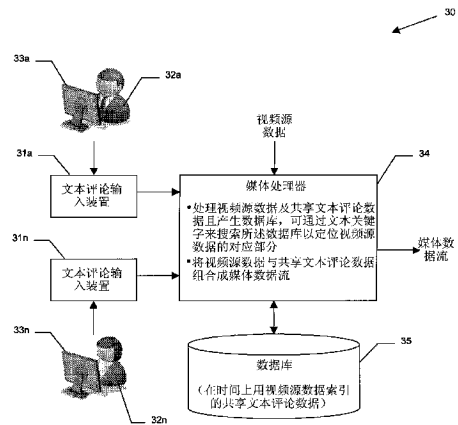
权利要求书 1 页 说明书 6 页 附图 9 页

(54) 发明名称

产生与视频源数据同步的音频触发器标记的多媒体系统及相关方法

(57) 摘要

一种多媒体系统 (130) 可包括至少一个音频评论输入装置 (141), 所述至少一个音频评论输入装置 (141) 经配置以准许至少一个评论员 (132) 基于观看来自视频源的视频数据而产生音频评论数据。此外, 媒体处理器 (134) 可与所述至少一个音频评论输入装置 (141) 协作且经配置以: 处理视频源数据及音频评论数据, 且针对所述音频评论数据中的预定音频触发器而从其产生与所述视频源数据同步的音频触发器标记; 及将所述视频源数据、所述音频评论数据及所述音频触发器标记组合成媒体数据流。



1. 一种多媒体系统,其包含:

至少一个音频评论输入装置,其经配置以准许至少一个评论员基于观看来自视频源的视频数据而产生音频评论数据;及

媒体处理器,其与所述至少一个音频评论输入装置协作且经配置以

处理视频源数据及音频评论数据,且针对所述音频评论数据中的预定音频触发器而从其产生与所述视频源数据同步的音频触发器标记,及

将所述视频源数据、所述音频评论数据及所述音频触发器标记组合成媒体数据流。

2. 根据权利要求1所述的多媒体系统,其中所述至少一个音频评论输入装置包含多个音频评论输入装置;且其中所述媒体处理器进一步经配置以基于预定音频触发器在设定时间内的多次出现而产生所述音频触发器标记。

3. 根据权利要求2所述的多媒体系统,其中所述媒体处理器进一步经配置以基于相应预定音频触发器在所述设定时间内从不同音频评论输入装置的多次出现而产生所述音频触发器标记。

4. 根据权利要求1所述的多媒体系统,其中所述媒体服务器进一步经配置以存储所述媒体数据流的与所述音频触发器标记的出现相关联的部分。

5. 根据权利要求1所述的多媒体系统,其中所述媒体处理器进一步经配置以基于所述预定音频触发器在所述音频评论数据中的出现而产生通知。

6. 根据权利要求1所述的多媒体系统,其进一步包含多个文本评论输入装置,所述多个文本评论输入装置经配置以允许多个评论员基于观看所述视频数据而产生共享文本评论数据;且其中所述媒体处理器进一步经配置以处理所述视频源数据及共享文本评论数据且从其产生数据库,所述数据库包含在时间上用所述视频源数据索引的共享文本评论数据,使得可通过文本关键字来搜索所述数据库以定位所述视频源数据的对应部分。

7. 一种多媒体数据处理方法,其包含:

使用至少一个音频评论输入装置产生音频评论数据,所述至少一个音频评论输入装置经配置以准许至少一个评论员在观看来自视频源的视频数据后即刻进行评论;

使用媒体处理器处理视频源数据及音频评论数据,且针对所述音频评论数据中的预定音频触发器而从其产生与所述视频源数据同步的音频触发器标记;及

使用所述媒体处理器将所述视频源数据、所述音频评论数据及所述音频触发器标记组合成媒体数据流。

8. 根据权利要求7所述的方法,其中所述至少一个音频评论输入装置包含多个音频评论输入装置;且所述方法进一步包含使用所述媒体处理器基于预定音频触发器在设定时间内的多次出现而产生所述音频触发器标记。

9. 根据权利要求8所述的方法,其中产生包含基于相应预定音频触发器在所述设定时间内从不同音频评论输入装置的多次出现而产生所述音频触发器标记。

10. 根据权利要求7所述的方法,其进一步包含使用媒体服务器来存储所述媒体数据流的与所述音频触发器标记的出现相关联的部分。

产生与视频源数据同步的音频触发器标记的多媒体系统及 相关方法

技术领域

[0001] 本发明涉及媒体系统领域,且更特定来说涉及用于处理视频、音频及其它相关联数据的多媒体系统及方法。

背景技术

[0002] 举例来说,从模拟到数字媒体系统的转变已允许先前不类似的媒体类型的组合,例如聊天文本与视频。将聊天文本与视频组合的一个示范性系统论述于颁发给德卫斯(DeWeese)等人的第 2005/0262542 号美国专利公开案。此参考文献揭示一种电视聊天系统,其允许电视观众在看电视的同时在聊天群组中参与与其它电视观众的实时通信。所述电视聊天系统的用户可参与与当前正在看相同电视节目或频道的其它用户的实时通信。

[0003] 另外,数字媒体格式的使用已增强产生并存储大量多媒体数据的能力。此外,随着增加的量的多媒体数据到来,存在处理所述数据的更大挑战。已开发用于增强视频处理的各种方法。一种此类方法论述于颁发给法西亚诺(Fasciano)的第 6,336,093 号美国专利。可分析与视频节目相关联的音频(例如,音频追踪或实况或记录的评论)以辨识或检测一个或一个以上预定声音模式,例如词语或声音效果。可使用所辨识或检测的声音模式通过在编辑期间控制视频捕获及/或递送来增强视频处理,或促进在编辑期间对剪辑或接合点的选择。

[0004] 颁发给麦考恩(McKoen)等人的第 2008/0281592 号美国专利公开案揭示一种用于使用语音辨识技术产生的元数据评注视频内容的方法及设备。所述方法以在显示装置上再现视频内容来开始。从用户接收语音的段,使得所述语音段评注视频内容的当前正被再现的一部分。将所述语音段转换成文本段且所述文本段与所述视频内容的经再现内容相关联。以选择性地可检索的方式存储所述文本段,使得其与所述视频内容的所述经再现部分相关联。

[0005] 尽管此些系统提供优点,但仍可期望进一步的改进以便以对用户有所帮助的方式管理及存储多媒体数据。

发明内容

[0006] 鉴于前述背景,因此本发明的目的是提供一种用于多媒体数据处理的系统及相关方法。

[0007] 通过可包括至少一个音频评论输入装置的多媒体系统来提供此及其它目的、特征及优点,所述至少一个音频评论输入装置经配置以准许至少一个评论员基于观看来自视频源的视频数据而产生音频评论数据。此外,媒体处理器可与所述至少一个音频评论输入装置协作且经配置以处理视频源数据及音频评论数据并,且针对所述音频评论数据中的预定音频触发器而从其产生与所述视频源数据同步的音频触发器标记。所述媒体处理器可进一步经配置以将所述视频源数据、所述音频评论数据及所述音频触发器标记组合成媒体数据

流。因此,所述系统提供共享文本评论数据的可容易搜索的档案,所述共享文本评论数据有利地在时间上与所述视频源数据相关。

[0008] 更特定来说,所述至少一个音频评论输入装置可包含多个音频评论输入装置,且所述媒体处理器可进一步经配置以(举例来说)基于预定音频触发器在设定时间内(例如)从不同音频评论输入装置的多次出现而产生音频触发器标记。媒体服务器可进一步经配置以存储媒体数据流的与所述音频触发器标记的出现相关联的部分。

[0009] 通过举例的方式,所述媒体处理器可通过多路复用来组合视频数据馈送、音频数据馈送及音频触发器标记,以产生媒体数据流。此外,举例来说,所述媒体数据流可包含运动图片专家组(MPEG)传输流。另外,所述媒体处理器可进一步经配置以基于所述音频评论数据中的预定音频触发器的出现而产生通知。

[0010] 所述多媒体系统可进一步包括多个文本评论输入装置,所述多个文本评论输入装置经配置以准许多个评论员基于观看视频数据而产生共享文本评论数据。此外,所述媒体处理器可进一步经配置以处理所述视频源数据及共享文本评论数据且从其产生数据库,所述数据库包含在时间上用所述视频源数据索引的共享文本评论数据。此外,所述数据库可为可通过文本关键字搜索以定位所述视频源数据的对应部分。举例来说,所述媒体处理器可包含媒体服务器,所述媒体服务器可包括处理器及与其协作的存储器。

[0011] 相关多媒体数据处理方法可包括使用至少一个音频评论输入装置产生音频评论数据,所述至少一个音频评论输入装置经配置以准许至少一个评论员在观看来自视频源的视频数据后即刻进行评论。所述方法可进一步包括使用媒体处理器处理视频源数据及音频评论数据且针对所述音频评论数据中的预定音频触发器而从其产生与所述视频源数据同步的音频触发器标记。此外,可使用所述媒体处理器将所述视频源数据、所述音频评论数据及所述音频触发器标记组合成媒体数据流。

[0012] 相关物理计算机可读媒体可具有用于致使媒体处理器执行包括以下各项的步骤的计算机可执行指令:使用媒体处理器处理所述视频源数据及音频评论数据,且针对所述音频评论数据中的预定音频触发器而从其产生与所述视频源数据同步的音频触发器标记。进一步步骤可包括将所述视频源数据、所述音频评论数据及所述音频触发器标记组合成媒体数据流。

附图说明

[0013] 图1是根据本发明的示范性多媒体系统的示意性框图。

[0014] 图2是图1的系统的替代实施例的示意性框图。

[0015] 图3是更详细地图解说明图2的媒体服务器的示范性实施例的示意性框图。

[0016] 图4及图5是图解说明与图1及图2的系统相关联的方法方面的流程图。

[0017] 图6是根据本发明的另一示范性多媒体系统的示意性框图。

[0018] 图7是图6的系统的替代实施例的示意性框图。

[0019] 图8及图9是图解说明与图6及图7的系统相关联的方法方面的流程图。

具体实施方式

[0020] 现在下文将参照其中展示本发明的优选实施例的附图来更全面地描述本发明。然

而,本发明可体现为许多不同形式,且不应视为仅限于本文中所论述的实施例。而是,提供这些实施例以使本发明将全面且完整,且将本发明的范围完全传达给所属领域的技术人员。通篇中,相同编号指代相同元件,且使用撇号来指示替代实施例中的类似元件。

[0021] 如所属领域的技术人员应了解,本发明的部分可体现为一种方法、数据处理系统或计算机程序产品。因此,本发明的这些部分可采取以下形式:完全的硬件实施例;物理计算机可读媒体上的完全的软件实施例;或组合软件与硬件方面的实施例。此外,本发明的部分可为计算机可用存储媒体上的计算机程序产品,所述计算机可用存储媒体在所述媒体上具有计算机可读程序代码。可利用任何合适的计算机可读媒体,包括但不限于静态及动态存储装置、硬磁盘、光学存储装置及磁性存储装置。

[0022] 下文参照根据本发明实施例的方法、系统及计算机程序产品的流程图图解来描述本发明。应理解,所述图解中的框及所述图解中的框的组合可由计算机程序指令来实施。可将这些计算机程序指令提供到通用计算机、专用计算机或其它可编程数据处理设备的处理器,以产生机器,使得所述指令(其经由计算机或其它可编程数据处理设备的处理器执行)实施框中所指定的功能。

[0023] 也可将这些计算机程序指令存储于计算机可读存储器中,所述计算机可读存储器可引导计算机或其它可编程数据处理设备以特定方式发挥作用,使得存储于所述计算机可读存储器中的指令产生一制品,所述制品包括实施流程图框中所指定的功能的指令。也可将所述计算机程序指令加载到计算机或其它可编程数据处理设备上以致使在所述计算机或其它可编程设备上执行一系列操作步骤,以产生计算机实施的过程,使得所述指令(其在计算机或其它可编程设备上执行)提供用于实施流程图框中所指定的功能的步骤。

[0024] 首先参照图 1 到图 5,首先描述多媒体系统 30 及相关方法方面。特定来说,系统 30 以说明性方式包括多个文本评论输入装置 31a 到 31n,文本评论输入装置 31a 到 31n 经配置以准许多个评论员 32a 到 32n 基于观看来自视频源的视频数据而产生共享文本评论数据(在框 50 到 51 处)。通过举例的方式,文本评论输入装置 31a 到 31n 可为桌上型或膝上型计算机等,且评论员 32a 到 32n 可在相应显示器 33a 到 33n 上观看视频数据,但也可使用其它合适的配置,如所属领域的技术人员应了解。如本文中所使用,“视频数据”打算包括全运动视频以及运动图像,如所属领域的技术人员应了解。

[0025] 系统 30 进一步以说明性方式包括媒体处理器 34,媒体处理器 34 与文本评论输入装置 31a 到 31n 协作且经有利地配置以处理视频源数据及共享文本评论数据且从其产生数据库 35,数据库 35 包括在时间上用视频源数据索引的共享文本评论数据,使得可通过文本关键字来搜索所述数据库以定位视频源数据的对应部分(在框 52 处)。媒体处理器 34 可进一步经配置以(举例来说)将视频源数据及共享文本评论数据组合成媒体数据流,例如运动图片专家组(MPEG)(例如,MPEG2)传输流(在框 53 处),因此结束图 4 中所图解说明的方法(框 54)。

[0026] 在图 2 中所图解说明的实施例中,文本评论输入装置 31a' 及 31n' 经配置而以不同相应文本评论格式(此处为两个不同聊天文本格式)产生文本数据。更特定来说,文本评论输入装置 31a' 根据因特网中继聊天(IRC)格式产生聊天文本数据,而文本评论输入装置 31n' 根据 Adobe® Acrobat® Connect™(AC)格式产生聊天文本,如所属领域的技术人员应了解。然而,还应了解,也可使用这些示范性格式以外的其它合适的文本格式。

[0027] 因此,媒体处理器 34' 可进一步以说明性方式包括文本摄取模块 36', 文本摄取模块 36' 用于将不同文本评论格式调适成共用文本评论格式以供媒体处理器 34' 使用。更特定来说,文本摄取模块 36 可包括用于所述不同文本评论格式(IRC、AC 等)中的每一者的相应调适器 37a' 到 37n'。因此,文本摄取模块 36' 有利地可从各种不同系统抽取文本输入数据,例如聊天数据,且将各种格式转换或调适成适当的共用格式以供执行上述操作的媒体服务器 38' 使用。在图 3 中所示的实例中,所述媒体服务器以说明性方式包括处理器 39' 及与其协作以执行这些操作的存储器 40'。

[0028] 在一些实施例中,媒体服务器 38' 可进一步经配置以针对共享文本评论数据中的预定文本触发器而从共享文本评论数据产生文本触发器标记(在框 55' 到 56' 处)(图 5)。举例来说,基于共享文本评论数据中的一个或一个以上预界定文本触发器(例如,预界定关键字或短语)在设定时间内的出现,产生与视频源数据同步的文本触发器标记(例如,视频源数据在所述出现时间用视频数据的时间戳标记)。在一些实施例中,还可将所述文本触发器标记存储于数据库 35 中。如果需要,还可基于预界定文本触发器的出现产生通知(例如,电子邮件通知、弹出窗口等),并且警告预定文本触发器的出现的适当监督者或其它人员。

[0029] 举例来说,媒体处理器 34 可使用例如 MPEG2、MPEG4、H264、JPEG2000 等格式来执行媒体摄取。此外,可使用 MPEG 传输流或节目流、材料交换格式(MXF)、高级授权格式(AAF)、JPEG 2000 交互协议(JPIP) 等来执行例如归档、搜索及检索/导出等功能。如所属领域的技术人员应了解,也可使用其它合适的格式。可使用各种商业数据库系统来实施数据库 35, 也如所属领域的技术人员应了解。

[0030] 系统 30 因此可有利地用于其中一个或一个以上评论员将观看视频数据及评论的应用,且需要提供在时间上与视频数据相关的文本数据的可容易搜索的档案。此有利地允许用户快速地定位视频的可能的档案的有关部分,且避免搜索经过或观看不重要的视频及文本的长的部分或周期。所述系统可用于各种视频应用,例如电视剧或电影的观看、情报分析等。此外,系统 30 可有利地用以从数据库 35' 中所存储的文本产生总结报告。举例来说,在电视或电影观看的背景下,用户可在看电影的同时聊他们喜欢或不喜欢的东西。可由媒体处理器 34' 或其它计算装置通过存取数据库 35' 来产生针对视频的某些情景或部分、演员等使用了多少预定“喜欢”或“不喜欢”词语的总结报告。

[0031] 相关物理计算机可读媒体可具有用于致使媒体处理器 34 执行包括以下各项的步骤的计算机可执行指令:处理视频源数据及共享文本评论数据,且从其产生数据库 35, 数据库 35 包含在时间上用视频源数据索引的共享文本评论数据,其中可通过文本关键字来搜索所述数据库以定位视频源数据的对应部分。进一步步骤可包括将视频源数据及共享文本评论数据组合成媒体数据流。

[0032] 现在另外翻到图 6 到图 9,现在描述相关多媒体系统 130。借助于背景,尽管较易于产生并归档上述视频,但常常不存在用于在不将不期望的“聊天者”添加到多媒体文件的情况下添加来自视频分析员或评论员的音频评注或音频触发器的有效机制。举例来说,情报分析员连续看几个小时的视频数据流且评论其在视频流中看到了什么。许多评论可并非特定相关或是所关心的,但其他人可能需要重新观看当评论员或分析员识别出所关心项目时的那些时刻。然而,在许多小时的经归档音频/视频数据内找到这些特定所关心点可为

耗时且麻烦的。

[0033] 当前使用语音辨识系统,其可监视语音数据以找到特殊关键字。另一方面,举例来说,可使用一些媒体处理系统来将音频及标签短语多路复用成媒体流,例如 MPEG2 传输流。然而,系统 130 有利地允许监视来自视频分析员的语音以在特殊关键字或触发器发生时将其找到(即,实时地)、记录触发器标记且将所述触发器标记组合或多路复用成媒体容器,例如 MPEG2 传输流,又同时保持与视频及音频分离(即,不覆写在视频或数据馈送上)。

[0034] 更特定来说,所述多媒体系统以说明性方式包括一个或一个以上音频评论输入装置 141(例如,麦克风),音频评论输入装置 141 经配置以准许评论员 132 基于观看来自视频源的视频数据而产生音频评论数据(在框 150 到 151 处)。此外,媒体处理器 134 可与音频评论输入装置 141 协作且经配置以处理视频源数据及音频评论数据,且针对音频评论数据中的预定音频触发器而从其产生与视频源数据同步的音频触发器标记(在块 152 处)。媒体处理器 134 可进一步经配置以将视频源数据、音频评论数据及音频触发器标记组合(例如,多路复用)成媒体数据流(在框 153 处),因此结束图 8 中所图解说明的方法(框 154)。通过举例的方式,媒体处理器 134' 可通过多路复用来组合视频数据馈送、音频数据馈送及音频触发器标记以产生媒体数据流,例如(举例来说)将其多路复用成 MPEG2 传输流,但也可使用其它合适的格式。

[0035] 在图 7 中所图解说明的示范性实施例中,多个音频评论输入装置 141a' 到 141n' 由相应评论员 132a' 到 132n' 使用,且媒体处理器 134' 可进一步经配置以(举例来说)基于预定音频触发器在设定时间内从相同或从不同音频评论输入装置的多次出现而产生音频触发器标记(在框 155'、152' 处)。举例来说,此可有利地(例如)在第二分析员或评论员确认已发现特定项目或特定项目存在于视频馈送中时增加所要的事件的真正出现的置信率等。

[0036] 媒体处理器 134' 可进一步经配置以存储媒体数据流的与音频触发器标记的出现相关联的部分。根据一个示范性应用,音频触发器标记可用作视频记录系统的部分以仅记录并标记与特定触发器有关的视频数据馈送的那些部分。举例来说,所述系统可实施于数字视频记录器中,其中基于音频内容(例如,音频关键字或短语)而非题目、摘要等来记录电视节目。举例来说,用户可期望记录具有关于其喜爱的名人、当前事件等的评论的最近新闻剪辑。用户可将所关心人或事件的名字添加为预定音频触发器。媒体处理器 134' 有利地监视一个或一个以上电视频道,且一旦“听到”触发器,那么可任选地通过电视上的弹出窗口来通知用户等。举例来说,也可使用其它通知,例如电子邮件或 SMS 消息。系统 130' 还有利地以记录节目来开始且将音频触发器标记多路复用成视频数据。此后,用户可搜索记录或归档的多媒体节目以找到触发器且被提示视频馈送的当预定音频触发器出现时的确切位置。

[0037] 通过举例的方式,媒体处理器 134 可基于预定音频触发器的出现而开始记录且一直记录到节目的经排定结束时间为止。或者,媒体处理器 134 可记录一设定时间周期,例如几分钟、半个小时等。在其中数字视频记录器将最近观看的节目数据保持在数据缓冲器中的一些实施例中,媒体处理器 134 可有利地“回顾(reach back)”整个节目且从节目的开始为用户存储整个节目,如所属领域的技术人员应了解。

[0038] 另外,在一些实施例中,如上所述,媒体处理器 134' 可有利地经配置以基于预定

音频触发器在音频评论数据中的出现而产生通知（在框 157' 处）。同样，这些出现可包括一个或一个以上用户或监督者的显示器上的弹出窗口、电子邮件或 SMS 通知、自动化电话消息等，如所属领域的技术人员应了解。在视频 / 音频数据的未找到预定音频触发器的那些部分中，仍可将视频源数据及音频评论数据组合成媒体数据流而无音频触发器标记（在框 158' 处），如所属领域的技术人员应了解。对于上文所论述的系统 30' 也是如此，即，即使在不存在可用共享文本评论数据时仍可在媒体传输流中将视频源数据与音频数据（如果存在）组合。

[0039] 在此方面，在一些实施例中，可实施系统 30 及 130 的部分或将其组合在一起。举例来说，在系统 130' 中，包括多个文本评论输入装置 131a' 到 131n' 且其经配置以准许评论员 132a' 到 132n' 基于观看视频数据而产生共享文本评论数据，如上文所论述。也就是说，除基于预定音频触发器的出现而产生音频触发器标记以外，媒体处理器 134' 还可有利地产生在时间上用视频源数据索引的共享文本评论数据的上述数据库。此处同样，所述媒体处理器可实施为包括处理器 139' 及与其协作以执行上述功能的存储器 140' 的媒体服务器。

[0040] 上述系统及方法因此提供实时地自动添加有价值的信息以伴随视频数据而不添加不期望的聊天者的能力。具有事件标记的流可对在不需要操作者或用户看完整个经归档或经存储视频的情况下快速地识别重要事件有价值。此外，此方法有利地提供用以将有价值的音频评注组合或附加到实况或归档的视频的有效方式，此允许视频的用户在播放视频时看到触发器的弹出窗口或其它通知，以及搜索音频触发器点及被提示音频触发器点，而非看完整个视频。

[0041] 相关物理计算机可读媒体可具有用于致使媒体处理器 34 执行包括以下各项的步骤的计算机可执行指令：处理视频源数据及音频评论数据，且针对音频评论数据中的预定音频触发器而从其产生与视频源数据同步的音频触发器标记。如上文进一步论述，进一步步骤可包括将视频源数据、音频评论数据及音频触发器标记组合成媒体数据流。

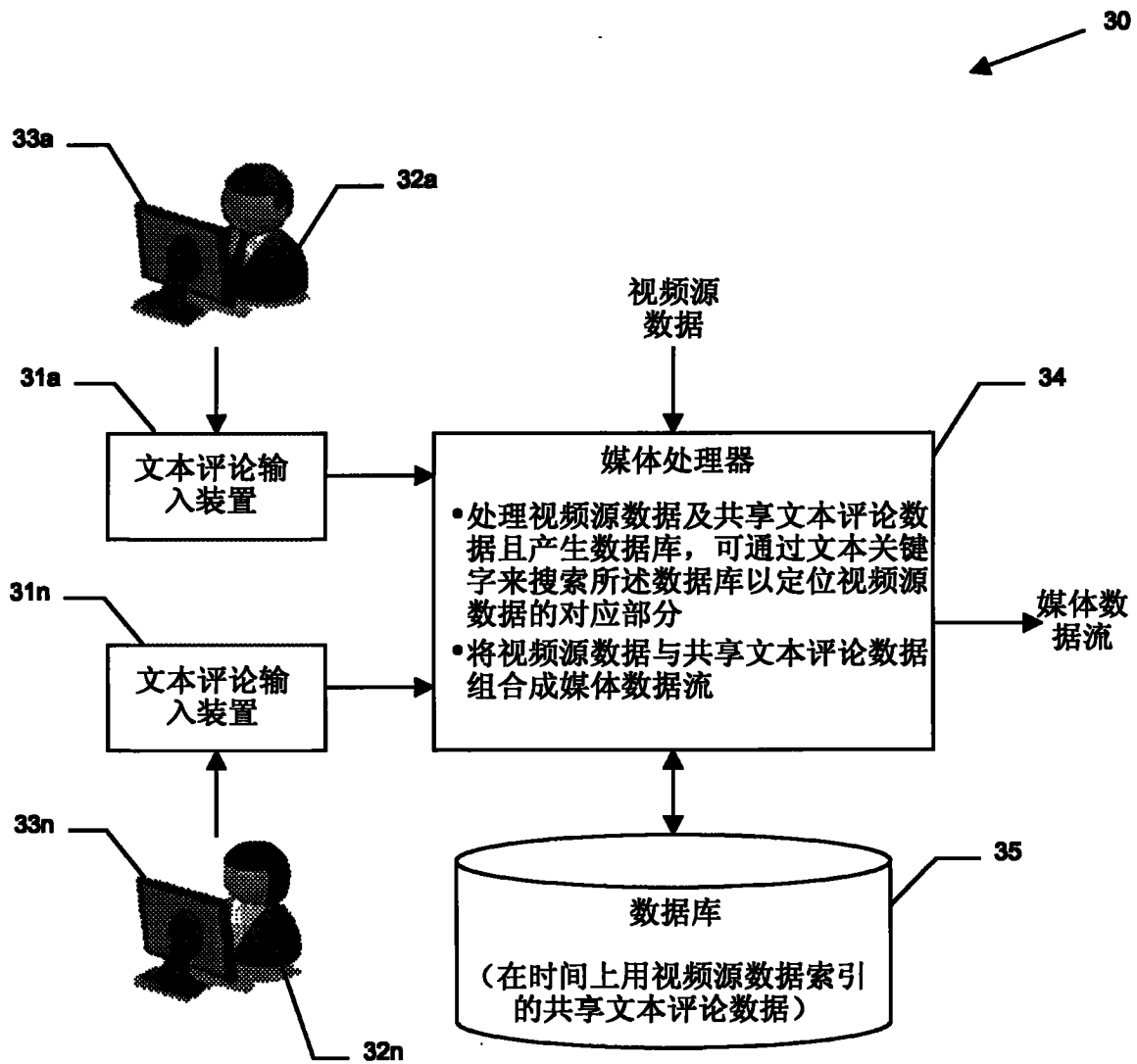


图 1

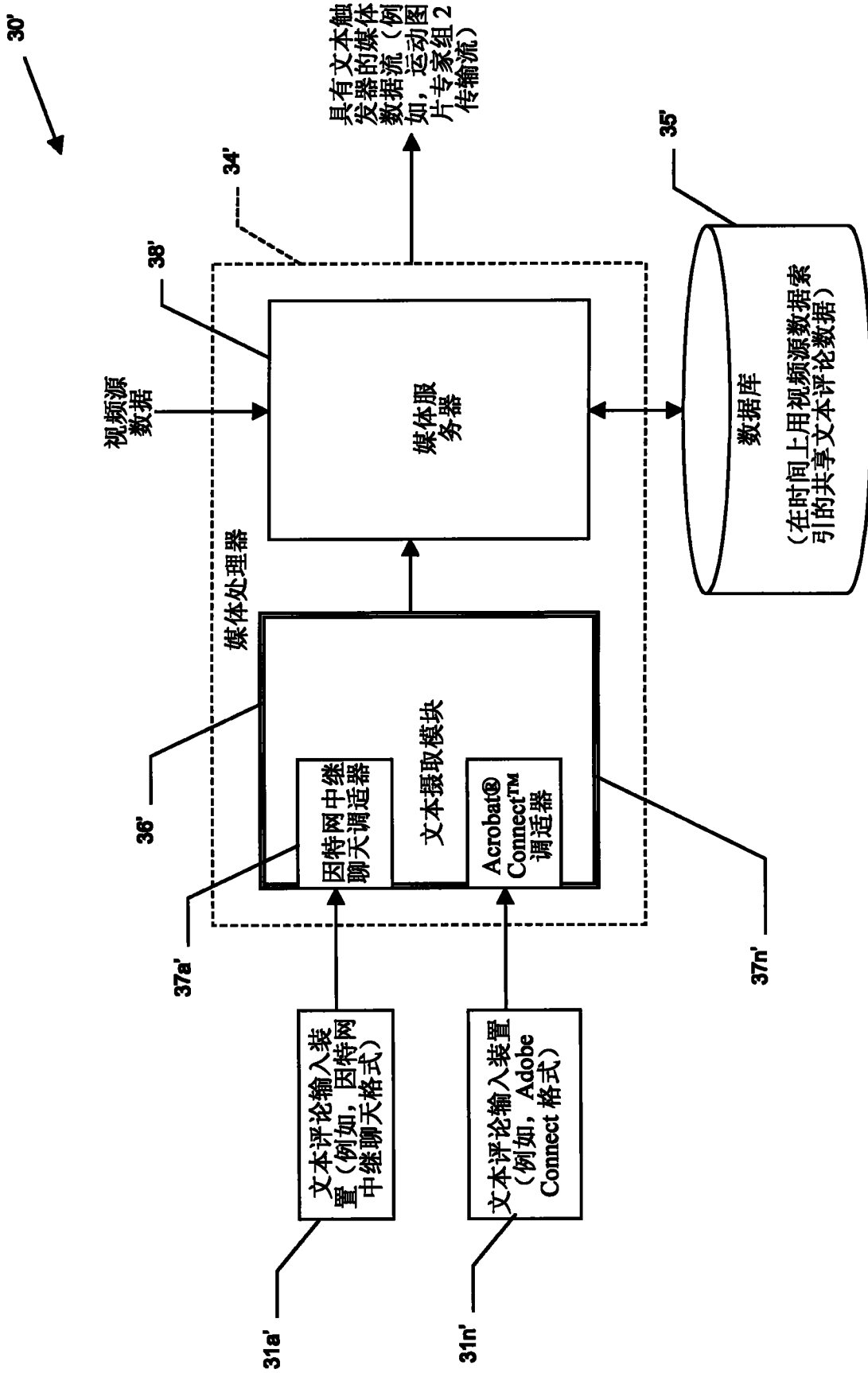


图 2

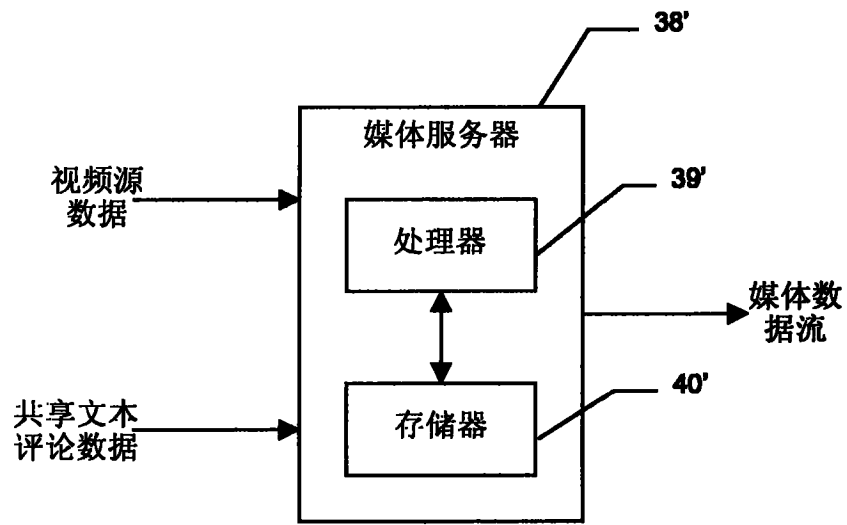


图 3

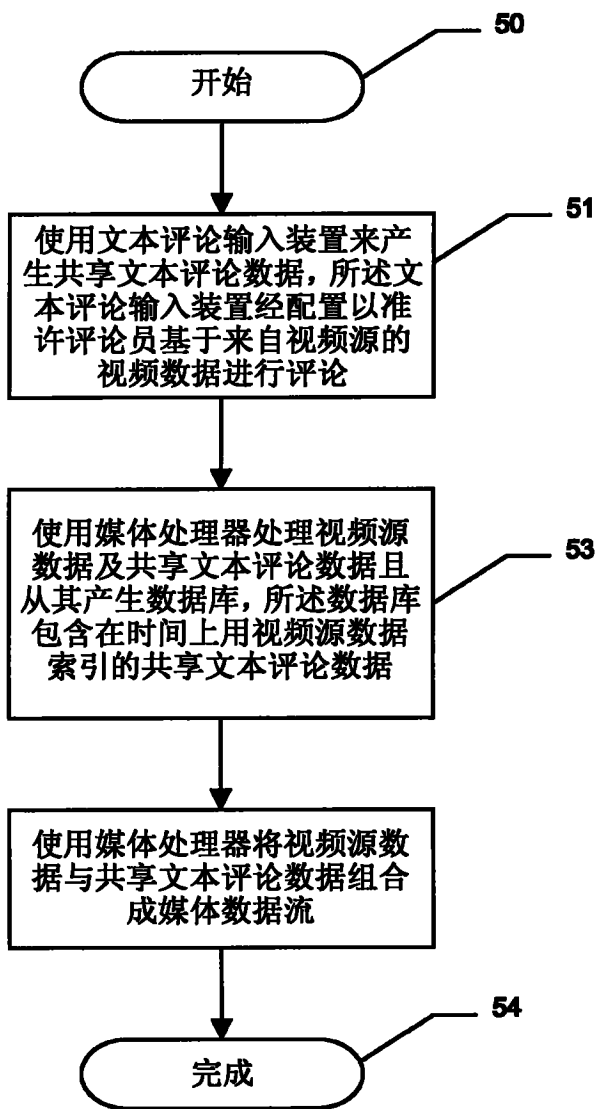


图 4

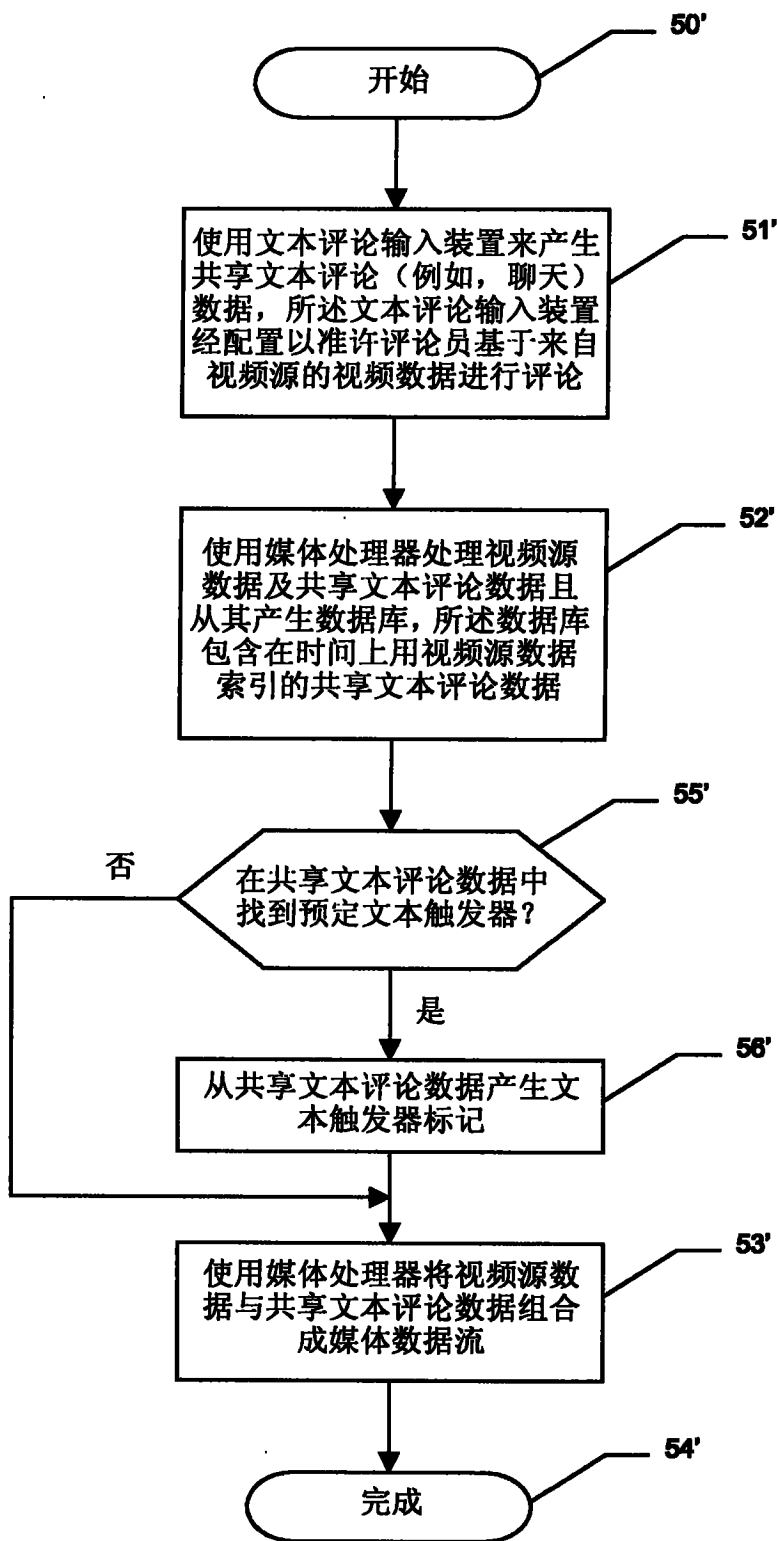


图 5

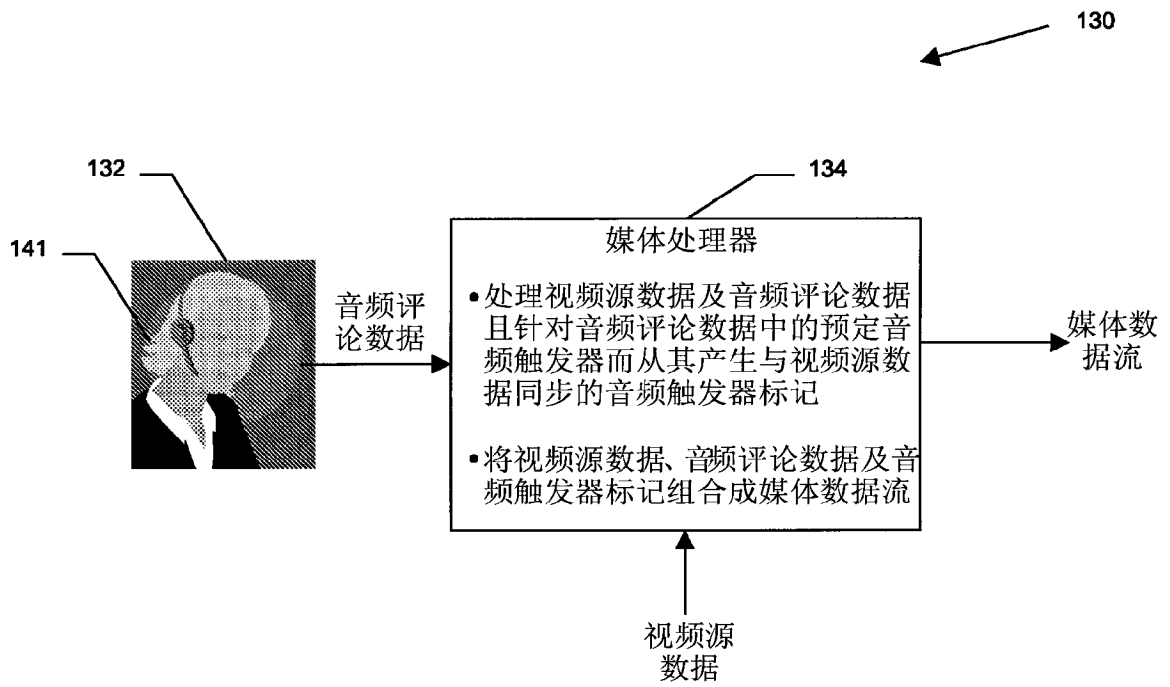


图 6

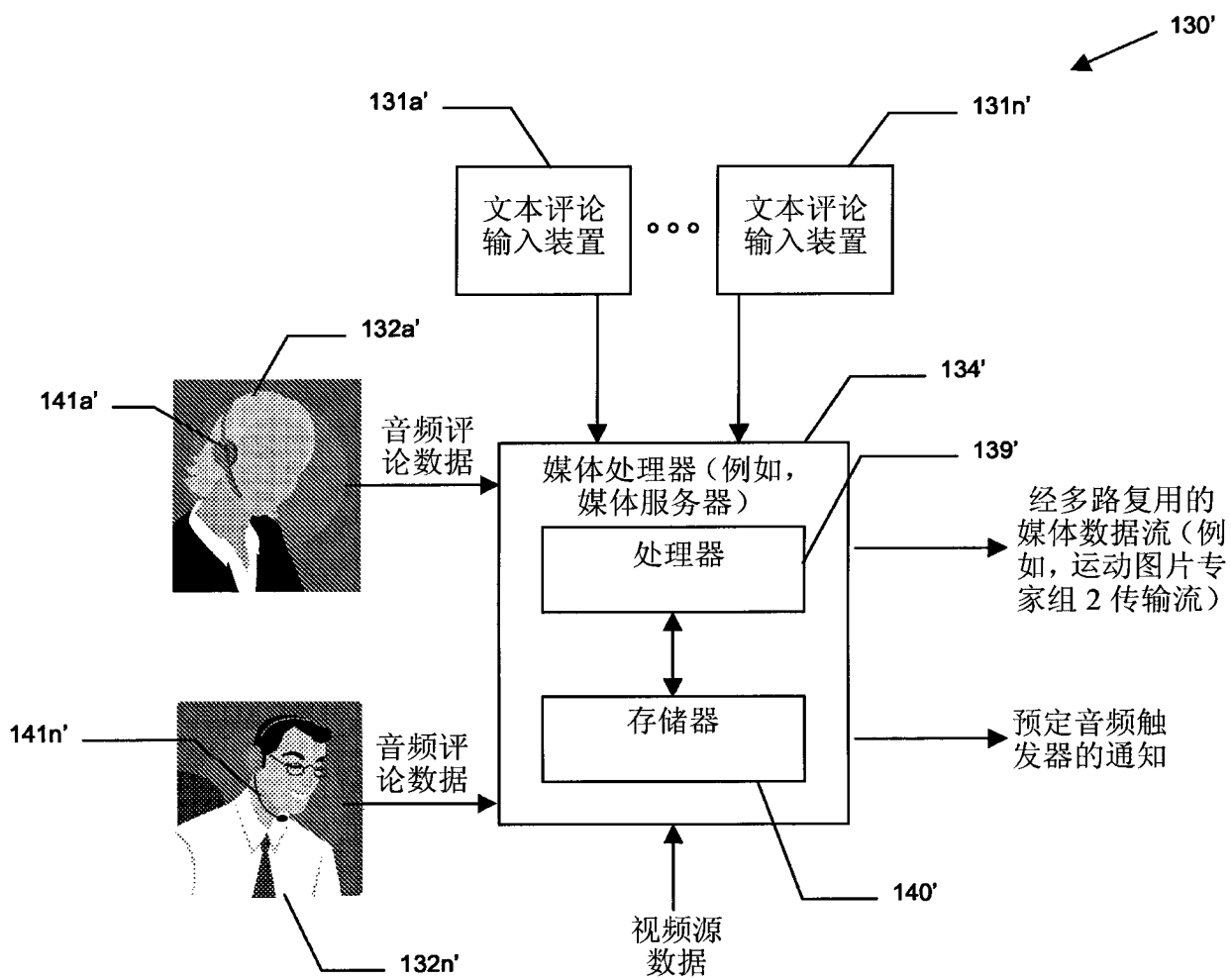


图 7

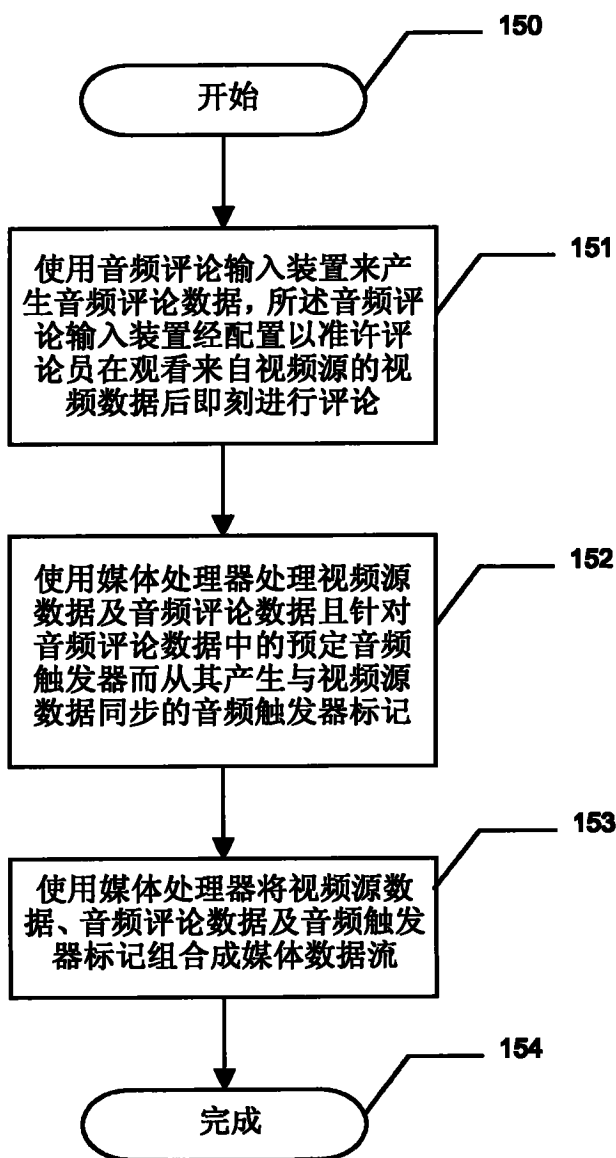


图 8

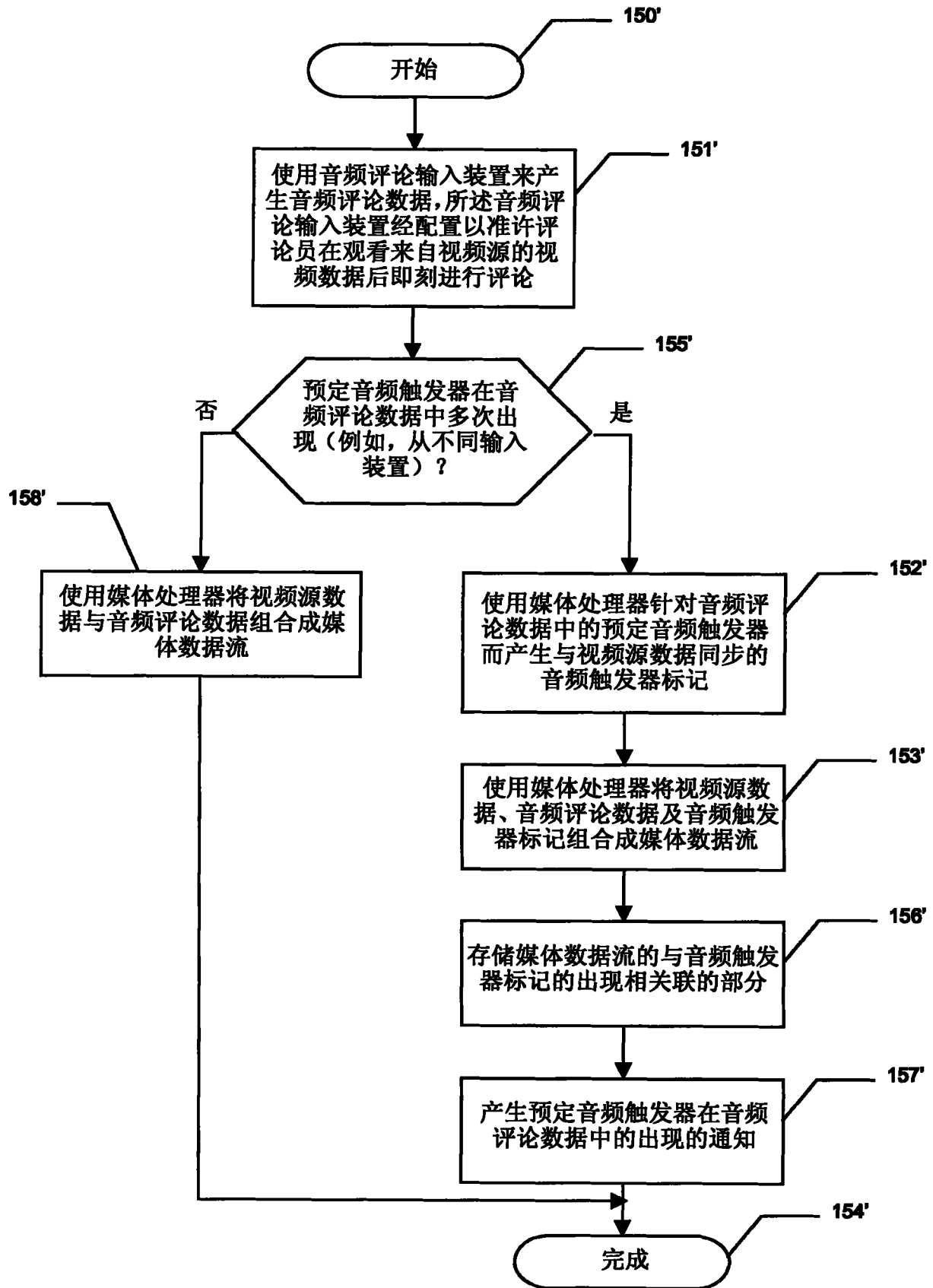


图 9