



(12) 发明专利

(10) 授权公告号 CN 101821734 B

(45) 授权公告日 2013. 09. 25

(21) 申请号 200880110557. 3

(51) Int. Cl.

(22) 申请日 2008. 08. 22

G06F 17/30(2006. 01)

(30) 优先权数据

(56) 对比文件

60/957, 446 2007. 08. 22 US

US 2006271947 A1, 2006. 11. 30, 全文.

12/174, 366 2008. 07. 16 US

US 2005198006 A1, 2005. 09. 08, 全文.

(85) PCT申请进入国家阶段日

CN 1758331 A, 2006. 04. 12, 全文.

2010. 04. 07

CN 1842151 A, 2006. 10. 04, 全文.

(86) PCT申请的申请数据

审查员 马春黎

PCT/US2008/074105 2008. 08. 22

(87) PCT申请的公布数据

W02009/026564 EN 2009. 02. 26

(73) 专利权人 谷歌公司

地址 美国加利福尼亚州

(72) 发明人 M·科维尔 J·亚格尼克

J·法斯特 S·巴拉甲

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华

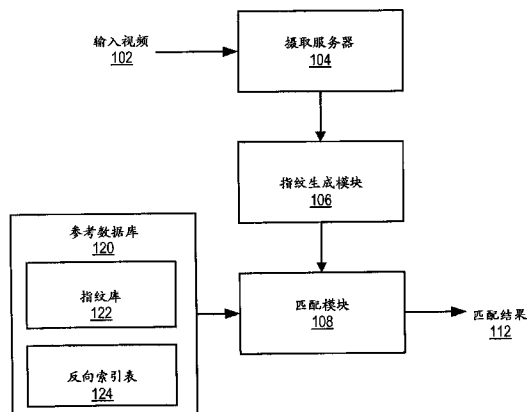
权利要求书4页 说明书13页 附图12页

(54) 发明名称

时基媒体之间的匹配的检测和分类

(57) 摘要

一种检测视频内容的部分之间的匹配的系统和方法。匹配模块接收输入视频和表示参考数据库中的参考视频的参考指纹的集合。匹配模块比较参考指纹和输入指纹,以生成来自参考视频集合的候选分段的列表。每个候选分段包括潜在地与输入视频相匹配的参考视频的时间局部化部分。对每个候选分段应用分类器以将分段分类为匹配分段或非匹配分段。然后基于分类为匹配的分段,输出标识来自参考视频集合的参考视频的匹配部分的结果。



1. 一种用于检测重复的视频内容的方法,所述方法包括:
接收表示输入视频的输入指纹;
获得表示参考视频集中的参考视频的参考指纹;
标识所述输入指纹与所述参考指纹之间的部分匹配;
基于标识的所述部分匹配生成来自所述参考视频集合的候选分段的列表,所述候选分段的列表包括所述参考视频集中的参考视频的时间局部化分段;
使用分类器将来自所述候选分段的列表的候选分段分类为匹配候选分段或非匹配候选分段;以及
基于对所述候选分段的分类来产生结果,所述结果标识来自所述参考视频集合的参考视频的匹配部分。
2. 根据权利要求 1 所述的方法,其中生成来自所述参考视频集合的所述候选分段的列表包括:
基于标识的所述部分匹配,确定初始候选参考视频的集合。
3. 根据权利要求 2 所述的方法,进一步包括:
获得所述输入指纹的输入子指纹的集合,其中每个输入子指纹包括输入子指纹关键词的序列;
获得所述参考视频的参考子指纹的集合,其中每个参考子指纹包括参考子指纹关键词的序列;以及
基于所述参考子指纹关键词与所述输入子指纹关键词之间的匹配,确定所述初始候选参考视频的集合。
4. 根据权利要求 2 所述的方法,进一步包括:
基于所述参考视频与所述输入视频之间的时间局部化匹配,确定每个参考视频的品质因数;以及
基于所述品质因数,确定所述初始候选参考视频的集合。
5. 根据权利要求 2 所述的方法,进一步包括:
按照从区别最大到区别最小的排序顺序,对表示所述参考视频的参考子指纹进行排序,区别最大的参考子指纹具有与所述输入指纹的每个输入子指纹最少的匹配关键词,并且区别最小的参考子指纹具有与所述输入指纹的每个输入子指纹最多的匹配关键词;以及
至少部分地基于所述参考子指纹的所述排序顺序,确定所述候选分段。
6. 根据权利要求 2 所述的方法,进一步包括:
按照从区别最大到区别最小的排序顺序,对表示参考视频的参考子指纹的参考关键词进行排序,区别最大的参考子指纹具有与所述输入指纹的每个输入子指纹最少的匹配关键词,并且区别最小的参考子指纹具有与所述输入指纹的每个输入子指纹最多的匹配关键词;以及
至少部分地基于所述参考关键词的所述排序顺序,确定所述候选分段。
7. 根据权利要求 2 所述的方法,进一步包括:
确定表示所述输入视频的输入视频子指纹与表示所述参考视频的参考视频子指纹之间的时间上连续的匹配;以及
基于所述时间上连续的匹配,确定所述候选分段的列表。

8. 根据权利要求 1 所述的方法,其中分类所述候选分段包括:
基于已知的匹配模型,确定指示所述参考视频与所述候选分段之间的匹配质量的质量测量;以及
基于所述质量测量,将所述候选分段分类为匹配候选分段或非匹配候选分段。
9. 根据权利要求 1 所述的方法,进一步包括:
接收表示来自所述输入指纹的部分值的输入关键词;以及
查询反向索引表以获得一个或多个子指纹标识符,所述一个或多个子指纹标识符标识具有与所述输入关键词相匹配的关键词的、来自所述参考视频集合的参考视频的时间局部化分段。
10. 根据权利要求 9 所述的方法,其中所述输入关键词包括来自输入子指纹的相邻值。
11. 根据权利要求 9 所述的方法,其中所述输入关键词包括来自输入子指纹的非相邻值。
12. 根据权利要求 9 所述的方法,其中如果超过预定百分比的表示所述参考视频的参考子指纹包含第一参考关键词,则所述第一参考关键词在所述反向索引表中被标识为第一级黑名单关键词,并且其中指示所述第一级黑名单关键词的标识符代码与所述第一参考关键词相关联地被存储在所述反向索引表中。
13. 根据权利要求 12 所述的方法,其中如果所述第一参考关键词不满足所述第一级黑名单关键词的标准并且超过另一预定义数目的表示所述参考视频的所述参考子指纹包含所述第一参考关键词,则所述第一参考关键词在所述反向索引表中被标识为第二级黑名单关键词,并且其中指示所述第二级关键词的标识符代码与所述第一参考关键词相关联地被存储在所述反向索引表中。
14. 根据权利要求 1 所述的方法,其中生成所述候选分段的列表包括:
确定来自所述参考视频的集合的、被标注为具有优质内容的优质参考视频;以及
基于所述优质参考视频的确定,生成所述候选分段的列表。
15. 一种用于检测重复视频内容的设备,包括:
用于接收表示输入视频的输入指纹的装置;
用于获得表示参考视频集合中的参考视频的参考指纹的装置;
用于标识所述输入指纹与所述参考指纹之间的部分匹配的装置;
用于基于标识的所述部分匹配生成来自所述参考视频集合的候选分段的列表的装置,所述候选分段的列表包括所述参考视频集合中的参考视频的时间局部化分段;
用于使用分类器将来自所述候选分段的列表的候选分段分类为匹配候选分段或非匹配候选分段的装置;以及
用于基于对所述候选分段的分类来产生结果的装置,所述结果指示来自所述参考视频集合的参考视频的匹配部分。
16. 根据权利要求 15 所述的设备,其中用于生成来自所述参考视频集合的所述候选分段的列表的装置包括:
用于基于标识的所述部分匹配确定初始候选参考视频的集合的装置。
17. 根据权利要求 16 所述的设备,进一步包括:
用于获得所述输入指纹的输入子指纹的集合的装置,其中每个输入子指纹包括输入子

指纹关键词的序列；

用于获得所述参考视频的参考子指纹的集合的装置，其中每个参考子指纹包括参考子指纹关键词的序列；以及

用于基于所述参考子指纹关键词与所述输入子指纹关键词之间的匹配确定所述初始候选参考视频的集合的装置。

18. 根据权利要求 16 所述的设备，进一步包括：

用于基于所述参考视频与所述输入视频之间的时间局部化匹配确定每个参考视频的品质因数的装置；以及

用于基于所述品质因数确定所述初始候选参考视频的集合的装置。

19. 根据权利要求 16 所述的设备，进一步包括：

用于按照从区别最大到区别最小的排序顺序对表示所述参考视频的参考子指纹进行排序的装置，区别最大的参考子指纹具有与所述输入指纹的每个输入子指纹最少的匹配关键词，并且区别最小的参考子指纹具有与所述输入指纹的每个输入子指纹最多的匹配关键词；以及

用于至少部分地基于所述参考子指纹的所述排序顺序确定所述候选分段的装置。

20. 根据权利要求 16 所述的设备，进一步包括：

用于按照从区别最大到区别最小的排序顺序对表示参考视频的参考子指纹的参考关键词进行排序的装置，区别最大的参考子指纹具有与所述输入指纹的每个输入子指纹最少的匹配关键词，并且区别最小的参考子指纹具有与所述输入指纹的每个输入子指纹最多的匹配关键词；以及

用于至少部分地基于所述参考关键词的所述排序顺序确定所述候选分段的装置。

21. 根据权利要求 16 所述的设备，进一步包括：

用于确定表示所述输入视频的输入视频子指纹与表示参考视频的参考视频子指纹之间的时间上连续的匹配的装置；以及

用于基于所述时间上连续的匹配确定所述候选分段的列表的装置。

22. 根据权利要求 15 所述的设备，其中用于分类所述候选分段的装置包括：

用于基于已知的匹配模型确定指示所述参考视频与所述候选分段之间的匹配质量的质量测量的装置；以及

用于基于所述质量测量将所述候选分段分类为匹配候选分段或非匹配候选分段的装置。

23. 根据权利要求 15 所述的设备，进一步包括：

用于接收表示来自所述输入指纹的部分值的输入关键词的装置；以及

用于查询反向索引表以获得一个或多个子指纹标识符的装置，所述一个或多个子指纹标识符标识具有与所述输入关键词相匹配的关键词的、来自所述参考视频集合的参考视频的时间局部化分段。

24. 根据权利要求 23 所述的设备，其中所述输入关键词包括来自输入子指纹的相邻值。

25. 根据权利要求 23 所述的设备，其中所述输入关键词包括来自输入子指纹的非相邻值。

26. 根据权利要求 23 所述的设备,其中如果超过预定义百分比的表示所述参考视频的所述参考子指纹包含第一参考关键词,则所述第一参考关键词在所述反向索引表中被标识为第一级黑名单关键词,并且其中指示所述第一级黑名单关键词的标识符代码与所述第一参考关键词相关联地被存储在所述反向索引表中。

27. 根据权利要求 26 所述的设备,其中如果所述第一参考关键词不满足所述第一级黑名单关键词的标准并且超过另一预定义数目的表示所述参考视频的所述参考子指纹包含所述第一参考关键词,则所述第一参考关键词在所述反向索引表中被标识为第二级黑名单关键词,并且其中指示所述第二级关键词的标识符代码与所述第一参考关键词相关联地被存储在所述反向索引表中。

28. 根据权利要求 15 所述的设备,其中用于生成所述候选分段的列表的装置包括:
用于确定来自所述参考视频集合的、被标注为具有优质内容的优质参考视频的装置;
以及

用于基于所述优质参考视频的确定生成所述候选分段的列表的装置。

29. 一种用于检测重复视频内容的系统,包括:

摄取服务器,用于接收输入视频;

指纹生成模块,用于生成表示所述输入视频的输入指纹;

参考数据库,存储表示参考视频的集合的参考指纹;

匹配模块,其耦合至所述指纹生成模块和所述参考数据库,所述匹配模块用于:标识所述输入指纹与所述参考指纹之间的部分匹配;基于标识的所述部分匹配生成来自所述参考视频的集合的候选分段的列表;将所述候选分段分类为匹配分段或非匹配分段;以及基于对所述候选分段的分类来产生结果,所述结果标识来自所述参考视频的集合的参考视频的匹配部分。

30. 根据权利要求 29 所述的系统,其中所述参考数据库包括:

指纹库单元,用于存储所述参考指纹,其中每个参考指纹包括表示所述参考视频的时间局部化分段的参考子指纹的序列;以及

反向索引表单元,其存储所述参考视频的标识符与参考关键词之间的映射,其中所述参考关键词包括来自所述参考子指纹的值的局部化集合。

31. 根据权利要求 30 所述的系统,其中如果超过预定义百分比的表示所述参考视频的所述参考子指纹包含第一参考关键词,则所述第一参考关键词在所述反向索引表单元中被标识为第一级黑名单关键词,并且其中指示所述第一级黑名单关键词的标识符代码与所述第一参考关键词相关联地被存储在所述反向索引表单元中。

32. 根据权利要求 31 所述的系统,其中如果所述第一参考关键词不满足所述第一级黑名单关键词的标准并且超过另一预定义阈值的表示所述参考视频的所述参考子指纹包含所述第一参考关键词,则所述第一参考关键词在所述反向索引表单元中被标识为第二级黑名单关键词,并且其中指示所述第二级关键词的标识符代码与所述第一参考关键词相关联地被存储在所述反向索引表单元中。

时基媒体之间的匹配的检测和分类

技术领域

[0001] 本发明总体上涉及视频处理,更具体地涉及检测匹配的视频内容。

背景技术

[0002] 电子视频库可以包含成千上万的视频文件,这使得对这些库的管理成为极具挑战性的任务。视频托管网站需要用于标识未授权视频的机制。虽然一些文件可以通过文件名或用户提供的其他信息来标识,但是这些识别信息可能是错误的或不足以正确地标识视频。使用人工来手动地标识视频内容的备选方法昂贵且耗时。

[0003] 视频共享站点面临的另一个问题在于该站点可能包含相同视频内容的多个拷贝。这浪费了存储空间,并成为主机的一大开销。第三个问题在于:由于文件的数目巨大,很难按照对用户而言方便的方式来组织视频库。例如,搜索结果可能具有多个相同或非常相似的视频的拷贝,使得用户难以对结果进行导航。

[0004] 鉴于上述问题,需要一种自动地比较和匹配重叠视频内容的技术。

发明内容

[0005] 一种系统和方法,检测重复的视频内容。匹配模块接收表示输入视频的输入视频指纹(fingerprint)。匹配模块生成来自参考视频集合的候选分段的列表。每个候选分段包括参考视频集合中的参考视频的时间局部化部分。对每个候选分段应用分类器,以便将分段分类为匹配分段或非匹配分段。然后基于被分类为匹配的分段来产生结果,该结果标识来自参考视频集合的参考视频的匹配部分。

[0006] 在一个实施方式中,通过获得表示参考视频集合中的参考视频的参考指纹以及标识输入指纹与参考指纹之间的部分匹配来确定候选分段。然后基于标识的部分匹配来确定初始候选参考视频的集合。分析初始候选参考视频,以确定输入视频的分段与参考视频的分段之间的时间上连续的匹配。然后基于该时间上连续的匹配来选择候选分段。

[0007] 说明书中描述的特征和优势并非意在穷尽,特别地,根据附图、说明书和权利要求,多种附加的特征和优势对于本领域技术人员而言将是显然的。而且,应当注意,说明书中所使用的语言在原则上是为了易读和指示的目的而不是为了界定或限制发明主题而选择的。

附图说明

[0008] 图 1 是用于检测输入视频与参考视频集合之间的匹配的系统的一个实施方式;

[0009] 图 2 是用于将参考视频与输入视频进行匹配的参考数据库的一个实施方式;

[0010] 图 3 是用于检测匹配视频内容的过程的一个实施方式;

[0011] 图 4 是用于生成与输入视频潜在匹配的候选视频分段的列表的过程的一个实施方式;

[0012] 图 5 是示出了针对输入视频生成的 LSH 关键词的无重复(duplicate-free)列表

的表的一个实施方式；

[0013] 图 6 示出了用于根据与输入视频的部分匹配来确定参考视频的品质因数 (figure of merit) 的技术的一个实施方式；

[0014] 图 7 示出了用于对输入视频指纹进行排列的技术的一个实施方式；

[0015] 图 8 示出了用于对视频子指纹中的 LSH 关键词进行排列的技术的一个实施方式；

[0016] 图 9 示出了用于输入视频子指纹有序集合的 LSH 查找技术的一个实施方式；

[0017] 图 10 示出了用于确定匹配的输入视频子指纹与参考视频子指纹之间的映射的一个实施方式；

[0018] 图 11 是用于将参考视频分段分类为匹配分段或非匹配分段的过程的一个实施方式；以及

[0019] 图 12 是用于跨时间将输入视频的时间局部化块与参考视频进行匹配的技术的一个实施方式。

[0020] 附图仅为了示意的目的绘出了本发明的各个实施方式。本领域技术人员根据以下的描述将很容易认识到，在不脱离此处所描述的本发明的原理的情况下，可以采用此处示出的结构和方法的备选实施方式。

具体实施方式

[0021] 描述了一种系统和方法，用于确定输入媒体文件（例如视频或音频或两者）是否与媒体文件的参考集合（例如，视频和 / 或音频片段的数据库）中的参考媒体文件匹配或部分匹配。匹配检测过程可以检测媒体文件的部分（例如，20 秒的片段）之间的匹配，即使媒体文件不具有相同的开始和结束点，或者在匹配部分之前或之后出现的内容不同。另外，该过程足够鲁棒以承受由于低质量转码而出现的标准降低，并且对于一定量的时标修改（例如，较快或较慢地回放视频）是鲁棒的。该过程能够正确地分类“真阳”（数据库中具有一个或多个匹配媒体文件的情况）或“真阴”（数据库中没有相应的匹配的情况）两者。通常，该过程可以在紧迫的时间约束下（用以处理上传业务率）和 / 或使用有限量的存储器来检测匹配。

[0022] 图 1 示出了用于检测时基媒体之间的匹配的系统的一个实施方式。需要注意的是，虽然具体示例是在匹配视频内容的上下文中提供的，但是所描述的系统和方法可以用于其他类型的媒体内容匹配例如音频、图像等等。摄取 (ingest) 服务器 104 从视频源接收输入视频 102。例如，视频源可以是与摄取服务器 104 通过网络进行通信的客户端计算机。备选地，视频源可以是可通信地耦合至摄取服务器 104 的数据库或其他存储设备。例如，视频源可以是视频存储介质，例如 DVD、CD-ROM、数字视频录像机 (DVR)、硬盘驱动器、闪存或其他存储器。摄取服务器 104 还可以可通信地耦合至视频捕获系统（例如摄像机）以接收即时视频内容。

[0023] 指纹生成模块 106 从摄取服务器 104 接收输入视频 102，并生成表示输入视频 102 的“指纹”。指纹是表示例如压缩格式的视频文件中的某些或全部视频文件的空间、时间和 / 或结构特征的位向量。指纹基于视频的可视内容来标识视频，使得由压缩、解压、噪声、帧率、开始和结束时间、源分辨率等所导致的细微变化不会显著影响指纹。

[0024] 在一个实施方式中，指纹生成模块 106 将接收到的视频分为多个重叠的分段，并

且针对每个分段生成子指纹。分段的优选长度是 0.5 秒到 5.0 秒,但是也可以使用其他长度的分段。由此,每个子指纹表示媒体的时间局部化分段(例如,4 秒的视频或 1.5 秒的音频)。分段的开始时间通常以固定频率相间隔(例如,对于视频每隔 0.25 秒,或对于音频每隔 0.10 秒)。例如,针对从 0 秒到 4 秒的分段计算第一子指纹,针对从 0.25 秒到 4.25 秒的分段计算第二子指纹,针对从 0.50 秒到 4.50 秒的分段计算第三子指纹,以此类推。每个子指纹由子指纹标识符代码来参考,该子指纹标识符代码标识该子指纹所表示的视频的特定分段。例如,子指纹标识符代码可以包括(视频、偏移)对。标识符的“视频”部分唯一地标识参考视频数据库中的视频(例如,使用 32 位标识符)。标识符的“偏移”部分通过例如参考分段开始时间的偏移索引来标识视频的特定分段(或相应的子指纹)。例如,如果分段的开始间隔 0.25 秒,则在 0 秒开始的分段可以具有偏移索引 0,在 0.25 秒开始的分段具有偏移索引 1,在 0.5 秒开始的分段可以具有偏移索引 2,以此类推。分段也可以直接通过其开始时间来标识。子指纹的完整有序序列提供了视频的完全指纹。

[0025] 在一个实施方式中,每个子指纹包括值的向量,每个值取自有限大小的字母表(例如,256 大小的字母表,每个向量维编码为一个字节,但是没有顺序)中的无序值。例如,每个值可以基于使用熵编码的值的概率分布进行编码。通过将汉明(Hamming)距离度量替换为适合于从其中取得子指纹的空间的距离度量,来进行对度量空间的扩展。对非均匀采样的(生成了指纹的)序列的概述对于本领域技术人员而言将是易见的。为了清楚,本说明书将假设均匀采样,并将假设使用汉明距离的无序字母表以便比较。

[0026] 匹配模块 108 将输入视频 102 的指纹与表示参考视频集合的某参考指纹集合进行比较。参考指纹可以用于所有可用的参考视频或其子集。在某些情况下,指纹从指纹源中提供,而不提供作为指纹基础的初始参考视频。匹配模块 108 输出匹配结果 112,该匹配结果 112 标识与输入视频 102 的至少部分相匹配的一个或多个参考视频(或参考视频的部分)。以下参考图 3 对确定匹配的方法进行更加详细的描述。

[0027] 参考数据库 120 包括一个或多个指纹库 122 以及一个或多个索引表 124。指纹库 122 存储参考视频的指纹。参考指纹集合中的每个子指纹通过子指纹标识符来标识,该子指纹标识符标识特定的视频以及在对应于该子指纹所表示的视频分段的视频中的偏移。偏移可以是分段开始时间值的时间偏移,或者指示分段序列中位置的索引号。

[0028] 在一个实施方式中,某些参考视频的指纹可以使用指纹库 122 中的附加元数据来进行标记,例如参考视频包含“优质内容”的指示。被标记为“优质内容”的视频是那些值得增加保护级别并在匹配过程期间给予特殊考虑的视频,这将在以下进行描述。优质内容指定可以根据多个不同的因素来确定。在一个实施方式中,内容所有者确定将哪些内容指定为优质内容。例如,媒体公司可以从其拥有的视频中选择一定数目的“顶级”视频作为其最关注的视频。在另一实施方式中,优质内容的指定可以基于先前的匹配历史来确定。例如,先前已被确定为具有与随后上传的输入视频相匹配内容的参考视频可以被自动地指定为优质内容。在另一个实施方式中,优质内容的指定基于参考视频在参考库中的时间长度。例如,参考在其处于数据库中的第一个月可以被视为优质的,继而可选地移除其优质内容的指定。另外,优质内容的指定可以是多级别的。例如,可以对应于分析的不同级别,指定优质内容的不同级别。

[0029] 在一个实施方式中,针对参考视频的每个媒体轨道(例如音频或视频)生成独立

的指纹。另外,在一些实施方式中,可以针对每个媒体轨道生成多个指纹。例如,如果指纹生成过程是左右镜像敏感的,则该过程可以选择为每个视频轨道生成两个指纹(第二个针对镜像帧来运行指纹生成过程,其中,视频中的帧在指纹生成之前进行了左/右翻转)。针对每个轨道生成的指纹数目(以及所应用的转换的类型)可以取决于与参考视频相关联的元数据,以及参考视频是否被标记为“优质内容”。作为另一个示例,视频可以具有对应于不同声道的多个音频轨道,在这种情况下,每个音频轨道都可以具有指纹,或备选地只有选定的音频轨道(例如,立体声左和立体声右)可以具有指纹。

[0030] 在针对单个轨道生成多个指纹时,子指纹以相同的子指纹标识符代码与媒体类型(例如,音频或视频)标识符代码以及不同的子代码被添加到指纹库 122,以便在多个指纹之间进行明确区分。为了简化下文描述,假设针对每个轨道类型(例如,音频和视频)的指纹存储在独立的指纹库 122 中。而且,为了简化,所描述的示例情况假设每个参考视频仅具有一个视频轨道和一个音频轨道,并且每个轨道具有一个指纹。

[0031] 在一个实施方式中,基于指纹库 122 的内容创建反向索引表 124。如果库 122 周期性地更新(例如,当用户上传新的视频时),反向索引表 124 可以按调度间隔进行更新,或者在库内容变化的任何时候更新。在反向索引表 124 中,子指纹值(或称“子指纹关键词”)的子集提供了一种标识库表 122 中的参考子指纹的手段。每个子指纹关键词与包含该子指纹关键词的参考子指纹集合相关联地存储。构造该反向索引表 124 是为了提供近似最近邻居功能。

[0032] 图 2 中示出了参考指纹库 122 和使用位置敏感散列(LSH)的反向索引表 124 的一个示例。参考子指纹的集合对应于参考库 122 中的参考视频。每个子指纹与视频的一个分段相关联,并且使用表示法 $X@Y$ 来标识,其中 X 是视频的标识符,而 Y 是标识视频 X 的分段的偏移索引。子指纹集合 $(X@(Y_1) \dots X@(Y_n))$ 形成了视频 X 的指纹。例如,在图 2 中,参考视频 A 的指纹包括子指纹 $A@(0), A@(1), \dots$ 等等。每个子指纹对应于参考视频 A 的分段,该分段由偏移索引来标识。例如, $A@(0)$ 标识表示视频 A 在偏移索引 0 处的分段的子指纹, $A@(1)$ 标识表示视频 A 在偏移索引 1 处的分段的子指纹,以此类推。

[0033] 每个子指纹包括值的序列(例如,每个值可以是一个字节)。将值的序列分为多个 LSH 带,每个带对应于子指纹中值的一个子集。例如,LSH 带 0 包括子指纹的前四个值,LSH 带 1 包括子指纹的接下来的四个值,以此类推。在一个实施方式中,子指纹具有 25 个 LSH 带,每个包括 4 个字节值。子指纹的 LSH 带中值的集合对应于如上所述的子指纹的 LSH 带关键词,并且与包括该关键词的子指纹标识符相关联地存储在反向索引表 124 中。每个子指纹带关键词在此也称为“LSH 关键词”。

[0034] 例如,子指纹 $A@(0)$ 在 LSH 带 0 中具有值 $(65, 22, A5, B1)$ 。该值集合在反向索引 LSH 表 124 中由 LSH 关键词 $(65 \ 22 \ A5 \ B1+0)$ 表示。在该表示法中,+0 指示该关键词出现在 LSH 带 0 中。LSH 表 124 将每个 LSH 关键词映射至库 122 中包含该 LSH 关键词的每个子指纹。例如,该表与 LSH 关键词 $234344D2+1$ 相关联地存储子指纹 $A@(0)$ 和 $B@(0)$ 的标识符,因为每个子指纹包含 LSH 带 1 中的值 $(2343 \ 44 \ D2)$ 。需要注意的是,值 $(11 \ 34 \ 55 \ 56)$ 出现在子指纹 $B@(0)$ 的带 2 中和子指纹 $B@(1)$ 的带 1 中。然而,这些被认为是不同的 LSH 关键词,并且由此在反向索引表 124 中独立地索引,因为值序列在不同的 LSH 带中。

[0035] 在备选的实施方式中,LSH 带包括不相交(非相邻)的值子集而不是示出的连续

值。将值分组为 LSH 带取决于具体应用的约束。另一种备选方法使用从全部子指纹中导出的 LSH 关键词。例如,可以通过根据到随机但记忆的 (random-but-memorized) 分划面的多个投影、根据子指纹的短序列或者例如短支持窗口内的子指纹条目值的频率直方图来计算有符号的位代码,从而确定子指纹的 LSH 关键词。频率直方图方法可以使用固定点频率计数作为关键词,或者其可以使用该时段中最频繁的签名值的值作为关键词。

[0036] 在另一备选实施方式中,使用例如溢漏树 (spill tree) 或 M 树等不同于 LSH 的技术来提供最近邻居功能,其使用在子指纹向量级别测量的汉明距离作为其度量空间。在以下的描述中,反向索引称为 LSH 表,即使并不要求近似最近邻居功能由该特定数据结构来提供。

[0037] 在该实施方式中,针对共同 LSH 关键词选择性地修改反向索引表 124,以控制匹配过程中所使用的存储量和计算量。具体地,将表构造为标记有助于区别子指纹的可能性降低的 LSH 关键词。通过对参考指纹内的每个 LSH 关键词的频率和 / 或分布的各种测试来确定此状态。

[0038] 第一测试基于 LSH 关键词出现的频率。如果超过总数的预定义百分比的参考视频在其指纹中包含给定的 LSH 关键词,则该 LSH 关键词在 LSH 表 124 中标识为“第一级黑名单”关键词,并且将包含该关键词的子指纹标识符从表 124 中删掉。代替的是,指示第一级黑名单关键词的特殊标识符代码与该关键词相关联地存储在反向索引表 124 中。例如,在一个实施方式中,如果超过 5% 的参考视频包含关键词 (00 00 00 0A+1),则在反向索引表 124 中将该关键词标识为第一级黑名单关键词 (通过与该关键词相关联地存储特殊标识符代码,如“BL1”),并且不存储与该关键词相关联的子指纹标识符。

[0039] 如果不满足第一级黑名单的标准,但是包含给定 LSH 关键词的子指纹总数仍然在某个阈值之上,则在反向索引表 124 中将该 LSH 关键词以不同的方式标记为“第二级黑名单”关键词。例如,如果多于 10,000 个参考子指纹包含关键词 (00 00 00 0B+2),但是这 10,000 个参考子指纹都包含在参考视频的仅 1% 内,则指示“第二级黑名单”的特殊标识符与该关键词相关联地存储。在一个实施方式中,不存储包含第二级黑名单关键词的子指纹标识符的完全列表。代替的是,LSH 表 124 仅存储视频标识符的列表但是不存储偏移索引 (即,不标识具体分段)。另外,该表可以存储包含匹配关键词的每个视频中的子指纹数目的计数。例如,在表 124 中,关键词 (00 00 00 0B+2) 被标识为第二级黑名单关键词,并且特殊标识符代码 (如“BL2”) 与该关键词相关联地存储。代码 (G,6) 也与该关键词相关联地存储,指示视频“G”的指纹具有包含匹配关键词 (00 00 00 0B+2) 的 26 个不同的子指纹。包含匹配关键词的子指纹的个体子指纹标识符不进行单独存储。

[0040] 可以使用没有黑名单的附加 LSH 表来为杯标记为优质内容的参考视频提供附加索引。这些表仅包含在主 LSH 表中被列入某一级别黑名单的那些 LSH 关键词,并且仅包括被标记为优质内容的参考视频。另外,这些表还可以包括对参考视频分段的系数索引,否则其将由于黑名单而完全从反向索引中缺失。在一个实施方式中,稀疏索引确保:每个临界时间间隔 (例如,以不超过 20 秒的间隔) 至少存在一次所有参考视频都具有未列入黑名单的反向索引条目。将来自这些表的条目添加到以下描述的匹配过程中,如同它们来自主 LSH 表集合一样。

[0041] 指纹匹配

[0042] 当接收到输入视频 102 用于匹配时（例如，来自上传的内容或来自现有的视频数据库），按照与应用于参考视频相同的指纹生成过程对输入视频 102 进行指纹生成。然后匹配模块 108 确定输入视频 102 的哪些部分（如果有的话）与参考数据库 120 中的参考视频的部分相匹配。在一个实施方式中，匹配模块 108 根据图 3 中所示的三阶段过程来确定匹配。在阶段 1 中，匹配模块 108 为来自参考集合的输入视频 102 生成 (302) 候选匹配的候选列表。候选列表中的每个条目指示：(1) 输入视频中潜在地与候选相匹配的部分；(2) 候选参考匹配的视频标识符；以及 (3) 参考视频中潜在地与输入视频相匹配的部分。例如，候选列表可以包括如下表中指示的结果：

[0043]

| 参考视频 ID | 输入视频匹配的部分 | | 参考视频匹配的部分 | |
|------------|-----------|--------|-----------|--------|
| | 开始时间 | 结束时间 | 开始时间 | 结束时间 |
| A | 2.4 秒 | 12.4 秒 | 61.9 秒 | 71.9 秒 |
| F | 32.1 秒 | 37.6 秒 | 93.5 秒 | 99.0 秒 |
| T | 11.3 秒 | 31.3 秒 | 55.3 秒 | 75.3 秒 |

[0044] 在第二阶段，进一步对候选列表中的每个候选条目进行评估 (304)，以提供指示匹配是正确还是虚假的本地分类。该确定的基础是来自候选列表中指示的匹配部分内的证据 (evidence)。

[0045] 在第三阶段，将剩余的候选匹配进行组合和截断，以确定 (306) 跨时间以及可选地跨通道（例如，音频和视频）的匹配，从而提供最终的结果集合。以下对三阶段匹配过程的每个阶段进行更加详细的描述。

[0046] 在一个实施方式中，如果输入视频 102 较长（例如，长于 30 分钟），则可选地对输入视频进行预处理，以将输入视频分为可管理大小的“章节 (chapter)”（例如，32 分钟）。通常，章节的开始和结束时间这样来定义，使得章节之间具有一定的少量重叠（例如，每个结尾处的 2 分钟）。在后处理步骤中，将最终输出结果“缝合”回一起。为了示例目的，以下的描述假设输入视频具有可管理的持续时间（例如，小于 30-32 分钟）。

[0047] 阶段 1: 候选列表生成

[0048] 该阶段创建来自参考集合的参考分段的短列表以用于进一步考虑。该步骤有助于控制计算和存储使用，并且提供将匹配处理与整体数据库大小隔离开、而是具有仅与真正匹配条目的最大预期数目（例如，30-60 个条目）同样快增加的计算的方法。图 4 中示出了阶段 1 的示例过程。

[0049] 第一步是确定 (402) 存在于表示输入视频 102 的输入子指纹的整个序列中的 LSH 关键词的列表，以及被映射至反向索引表 124 中那些关键词的相应参考子指纹标识符。图 5 示出了该步骤 402 的一个示例实施方式。将每个输入视频子指纹 502 分为 25 个 LSH 带，每个包含 4 个字节值的关键词。在一个实施方式中，如图所示，移除相同 LSH 带内的重复关键词。需要注意的是，不同 LSH 带中的关键词即使包含相同的值也不认为是重复的。然后，从对应于输入子指纹集合中的每个唯一 LSH 关键词的 LSH 表 124 中获取参考子指纹标

识符。该步骤的输出是 LSH 关键词的无重复列表 504 以及被映射至 LSH 表 124 中的 LSH 关键词的每个参考子指纹标识符。此处,表示法 $Z@(X_1, X_2, X_3, \dots, X_n)$ 表示在视频 Z 中的时间偏移索引 $X_1, X_2, X_3, \dots, X_n$ 处找到关键词。如果任何关键词导致第一级黑名单或第二级黑名单,则将其在无重复列表 504 中进行标注。

[0050] 接下来,确定 (404) 初始候选视频以用于进一步考虑。在一个实施方式中,通过上述无重复列表 504 中标出子指纹的一般时间位置以及维护与该视频的每个部分相关联的 LSH 关键词数目的计数,来创建初始候选视频的列表。然后可以针对每个参考视频创建时间相关的匹配计数,其记录在每个参考视频的不同时间窗口期间参考视频子指纹关键词与输入视频子指纹关键词之间的匹配频率。例如,图 6 示出了参考视频 D 的直方图(当然,在实践中,时间相关的匹配计数仅存储在存储器中,而不需要显示或以其他方式另外呈现)。在示出的示例中,匹配计数功能是粗略量化的(例如,5 秒的分辨率)。由此,匹配计数功能维护发生在参考视频的每个 5 秒窗口内的、参考视频与输入视频之间的匹配关键词数目的计数。例如,在 0 秒到 5 秒之间的时间窗口中,存在来自参考视频 D 的子指纹关键词的 5 个实例与来自输入视频的关键词的无重复列表 504 中的关键词相匹配。在参考视频的 5-10 秒之间的时间窗口中存在 3 个匹配实例,在 10-15 秒之间的时间窗口中存在 1 个匹配实例,等等。

[0051] 然后基于匹配计数来获得每个参考视频的品质因数。然后该过程从按品质因数排列的列表的顶部选择多个视频,在未来的处理中将仅考虑所选的多个视频。在一个实施方式中,通过在输入视频的长度上对匹配计数求和来获得品质因数,其中开始和结束点被选择为使品质因数最大化。例如,在参考视频 D 中,对 0 秒到 45 秒之间的输入视频长度(45 秒)的匹配的总数求和,以获得品质因数 15。这可以实现为匹配分布与输入视频的长度的

卷积,随后是最大值选择: $\max_{\{t'\}} \sum_{t'=0}^L h(t-t')$, 其中 L 是输入视频的长度(例如,45 秒),而

$h(t)$ 是时间上的匹配计数。备选地,对计算进行修改以保证参考视频的每个量化窗口(例如,5 秒窗口)在其对品质因数做出贡献具有匹配的至少具有某个阈值,以及保证至少多个区段(section)(例如,3 个五分钟区段)具有非零贡献。在另一实施方式中,还针对每个参考视频的第二级黑名单对出现率进行计数。对那些具有第二级黑名单的参考,可以认为:被省略的偏移标识符的计数数目在未列入黑名单的已扫描 LSH 列表中出现的索引的范围内均匀分布。这将有效地降低允许这些条目提供非零贡献的阈值。在又一实施方式中,给予被标记为包括优质内容的参考视频以某个附加优先级。例如,将这些条目添加到初始候选视频列表的阈值(基于其被指定为优质)可以降低。

[0052] 创建初始候选视频列表的另一备选方法是:给予被观测为在之前检查的通道(例如,另一个音频轨道)上已经匹配的参考以附加优先级。例如,如果该过程已经完成了一轮针对输入媒体文件的音频通道匹配并且现在正在检查视频通道,则由于相同的参考将被添加到视频通道候选列表,该过程可以使视频声道检查较为容易。音频和视频通道经常是成对的(或至少选自备选的小组),所以如果假设音频通道与给定参考相匹配是合理的,则该过程可以配置为更为接近地检查视频通道同样匹配的可能性。用于该跨通道促进的逻辑依赖于匹配过程的约束。如果约束不是太严格,则可以简单地将在不同通道上匹配的所有参考添加到初始候选视频列表中。如果约束不允许该简单方法,则可以使用先前的匹配来降

低用于创建初始候选视频列表的阈值。这些阈值可以在先前匹配条目的完全持续时间上降低,或者仅在当前轨道的同步部分上降低。

[0053] 所有的后续处理将仅考虑作为初始候选参考视频列表的一部分的参考视频。接下来,将输入视频分为(406)时间局部化的块。在一个实施方式中,块的长度至多是必须检测的最短预期真正匹配的长度的一半。例如,为了检测 20 秒长度的匹配,块可以是 10 秒或更小。可以通过将输入视频任意切分为不重叠的 10 秒块或通过将输入视频切分为重叠的 10 秒块(例如,通过跨整个输入视频长度滑动 10 秒窗口)来形成这些块。备选地,可以通过使用视频或音频分析而确定的边界(例如,切割边界或高运动边界,当基于分析的边界扩展太宽时,利用均匀间隔的边界填充)来形成块。为了描述简单,所给出的示例使用任意切分的无重叠 10 秒块。

[0054] 仅考虑以上确定的初始候选视频列表这一限定可以这样来实现:利用初始候选视频列表对初始候选视频(例如,来自上述预选择过程的视频)的图(例如,散列图)进行预填,并且不允许向该图进一步添加其他参考视频。

[0055] 对输入视频的每个块独立地进行如下所述的处理。在一个实施方式中,对输入视频的当前处理块的子指纹进行排序,以确定将首先对哪些子指纹进行处理。在一个实施方式中,对子指纹进行排序,使得当前块中的子指纹列表从区别最大到区别最小进行排列(408)。在该实施方式中,对来自初始候选视频列表的、具有与每个输入视频子指纹相匹配的关键词的参考子指纹的总数进行计数。首先列出具有最少匹配的子指纹。例如,图 7 示出了来自当前处理的输入视频块的子指纹集合。针对每个输入子指纹中的每个 LSH 关键词示出匹配数目。例如,LSH 关键词 1C865002+0 与来自初始候选视频列表的 4 个参考子指纹具有匹配。针对每个输入子指纹对匹配数目进行求和。该和被用于将输入子指纹排列为从最少匹配(区别最大)到最多匹配(区别最小)。

[0056] 如果输入子指纹中的一个或多个 LSH 关键词将要列入第二级黑名单,则将子指纹的匹配候选的数目的和(跨 LSH 关键词列表中的视频求和,或备选地,跨候选视频求和)用于子指纹的该部分。如果子指纹中的一个或多个 LSH 关键词将要列入第一级黑名单中,则将某个较大的值用于输入子指纹所涉及的候选参考子指纹的数目(例如,100x 第二级黑名单阈值)。

[0057] 一旦针对当前块进行了输入子指纹集合的排列,则对子指纹进行处理,这开始于区别最大的子指纹(最少匹配)。处理每个子指纹从为空的候选图开始,除非将初始候选视频预填充进了图的顶层。在每个子指纹内,LSH 关键词也按照区别最大到最小的顺序排列(410),类似于上述对子指纹的排列。例如,如图 8 所示,为子指纹中的每个 LSH 关键词确定匹配参考子指纹候选的数目。然后按顺序从具有最少匹配(区别最大)的关键词开始处理关键词。使用 LSH 关键词的该排列,处理开始于第一 LSH 关键词,并且向候选图中添加具有匹配关键词的参考子指纹候选。例如,如图 9 所示,对有序的输入视频子指纹(并且在每个子指纹内按照 LSH 关键词的顺序)执行 LSH 查找。在执行每个查找时,存储得到的子指纹标识符。在一个实施方式中,根据以下限制来将候选(相应于参考子指纹)添加到候选图中:

[0058] (1) 候选来自于初始候选视频列表。该检查可以通过以上提到的图预填充步骤来隐式地完成。如果视频标识符在预填充图中,则其在初始候选列表中,并且过程可以继续进

行。如果该位置不在图中,则不记录候选匹配。

[0059] (2) 图中不同候选的数目不超过预定义的阈值。一旦其达到该阈值,随后的 LSH 关键词可以增加对现有候选的支持,但是不可以添加新的候选。再一次,优质内容候选可以例外。

[0060] (3) 对于将要允许的新候选,存在至少 $(t-1)$ 个剩余 LSH 块,其中 t 是 LSH 子指纹阈值(以下进行描述)。剩余 LSH 块的该计数包括将被列入第二级黑名单但是不被列入第一级黑名单的 LSH 块。

[0061] (4) 最后考虑将要列入第二级黑名单的 LSH 块,并为每个视频创建可见出现计数(seen-occurrence-count),但是跨越每个候选视频的列入黑名单的块来求和。

[0062] 得到的候选图具有来自初始候选视频列表的参考子指纹标识符以及匹配当前输入子指纹的偏移的有限列表。下一个步骤跨输入视频子指纹来结合候选,以确定输入子指纹支持哪个参考和偏移候选。输入视频的子指纹与参考子指纹候选之间的每个匹配针对输入视频与参考视频之间匹配的具体开始位置进行“投票”。通过从参考视频中的匹配位置减去对输入视频的偏移来确定开始位置。例如,如图 10 所示,偏移(2)处的输入视频子指纹与参考子指纹 $D@3$ 具有匹配。这为开始位置 $D@2$ 生成了“投票”。类似地,输入视频在偏移(3)处的子指纹与参考子指纹 $D@4$ 具有一个匹配。该匹配也为开始位置 $D@2$ 生成了“投票”。一般地,如果输入视频中偏移 X 处的子指纹与参考视频中的偏移 Y 具有匹配,这支持输入视频的起始与参考视频中的 $Y-X+1$ 位置相匹配这一推论。跨每个参考视频的子指纹对投票进行累加。实际上,结果是单位倾斜线的 Hough 变换,其给出了可能匹配的输入视频参考时间。

[0063] 对于输入视频中的每个子指纹类似地处理投票。在一个实施方式中,将具有针对开始位置的至少 t 个投票的参考偏移转移到保持针对当前输入视频块的候选支持的新候选图中。可以针对优质内容选择性地降低该阈值。例如,包括优质内容的候选即使比非优质内容的支持级别低,也可以转移到新候选图中。对于生成过多具有 t 个或更多投票支持的参考候选的输入子指纹来说,该阈值将较高。在每个输入子指纹上,向上调节该阈值,直到来自将通过的子指纹的候选的数目小于预定义的数目。例如,在一个实施方式中,允许 400,000 个候选对,其中每个候选对对应于针对输入视频的每 10 秒的唯一匹配参考或开始偏移。然而,所使用的候选的具体最大数目高度依赖于系统的计算和时间约束。通过将通过候选的支持添加到先前支持(来自当前输入视频块中先前考虑的输入子指纹),将通过候选转移至当前输入视频块的新候选图中。

[0064] 一旦检查了当前块的所有子指纹,则将当前块、先前块和之后块的图加到一起,这给予了来自当前块的图的证据以最高权重,但是允许其他两个块图添加候选或增加对现有候选的支持。根据该组合映射,过程对该图使用模糊峰值采集(smear peak picking)(非最大抑制)以创建将在第二阶段考虑的候选列表(视频、偏移)。模糊峰值采集是单个候选内的开始时间的偏移上的简单卷积。一个示例实施方式包括通过三角窗进行卷积,该三角窗的三角宽度是用于对回放速度的可能改变的最大预期时间扩展的两倍。例如,在具有 10 秒大小的块的视频播放中,为了支持最多 10% 的加速或减慢,完全三角宽度将是 2 秒。卷积过程将增加存在针对邻近开始时间偏移的支持的峰的高度。非最大抑制是这样的过程,其强制被选作该阶段最终候选的峰属于该模糊时间信号的可区别的最大瓣(lobe)。此过程开

始于在信号中定位最高的极大值并记录该位置（以及值）。然后该过程将比某个时间间隔更接近该最大值的值归零（例如，当使用 10 秒块时，隔开 5 秒）。归零过程继续在时间上向外前进或后退，直到初始（模糊）函数持续单调下降。使用模糊峰值采集 / 非最大抑制有效地提供了更普遍的斜率和截距 Hough 变换功能，而不需要此类两维变换所需要的额外存储器。阶段 1 的输出是候选视频分段（每个对应于参考子指纹）的有限列表，将对其进行进一步考虑以确定输入视频与参考视频之间的匹配部分。对其设置的限制还高度依赖于系统，但是通常每 10 秒输入块低于 1000 个候选匹配是恰当的。

[0065] 阶段 2: 候选评估

[0066] 第二阶段处理考虑由第一阶段找到的候选视频分段，并确定某些、全部或没有候选视频分段是否为有效的匹配。因为过程允许参考集中有重复（全部或部分）的可能性，所以提供了较为一般的描述（而非每个探测块的单个是 / 否匹配）。第二阶段处理可以处理各种媒体，从与多种不同文件具有假匹配的非描述部分到时间敏感部分（即使是相同的参考材料，当其在时间上轻微偏移时，该时间敏感部分也匹配不佳）。

[0067] 为了处理该任务以及该材料范围，分类过程开始于针对每个候选视频分段来创建匹配质量测量。在一个实施方式中，动态时间规整 (DTW) 确定 (1102) 当前输入视频块与参考视频之间的跨时间最佳对准。DTW 的参数通过应用必须支持的匹配媒体内的时间失真量（例如，对于很多应用是大约 15% 时间加速或减慢）来确定。根据 DTW 的输出，在输入视频的个体子指纹与参考子指纹之间确定配对。对配对的集合进行评估，以生成 (1104) 输入视频与参考视频之间的匹配的向量描述。该匹配向量中的可能条目的示例可以是：

[0068] 1) 成对子指纹向量之间的累加汉明距离。

[0069] 2) 至少 80% 的成对向量相匹配的子指纹的百分比。

[0070] 3) 至少 60% 的成对向量相匹配的子指纹的百分比。

[0071] 4) 至少 40% 的成对向量相匹配的子指纹的百分比。

[0072] 5) 至少 20% 的成对向量相匹配的子指纹的百分比。

[0073] 6) 解码路径与最小方差 (LSE) 拟合直线路径之间的均方差 (MSE)。

[0074] 7) LSE 拟合直线解码路径的斜率。

[0075] 8) 与有效子指纹配对并且至少 10% 的成对向量匹配的有效子指纹的数目，其中使用先前的指纹生成过程期间采取的某一测量（例如，非空或非沉默）来确定有效 / 无效。

[0076] 9) 与无效子指纹配对的无效子指纹的数目。

[0077] 10) 与有效子指纹配对的无效子指纹的数目。

[0078] 11) 具有与其第一阶段候选列表中列出的相同视频以及近似相同偏移的相邻探测块的数目。

[0079] 12) 该配对在第一阶段证据收集期间接收到的投票的数目。

[0080] 13) 与先前检查的通道上的该参考的匹配的存在、置信度以及时间和偏移相似度。

[0081] 14) 将参考作为优质的指定。

[0082] 15) 生成该探测的用户的上传历史（例如，用户先前已经上传来自参考集合的内容）。

[0083] 16) 如果元数据关于探测和参考内容两者都可用，那些描述（例如锚文本）之间的

相似度。

[0084] 根据该向量描述,确定(1106)质量测量。质量测量的一个示例使用真正匹配的模型与虚假匹配的模型之间的似然率。适合的模型可以是全协方差高斯模型或对角线方差高斯混合模型,以及其他。备选地,通过将所有质量测量简单设置为0,可以跳过该质量测量步骤1106。

[0085] 可选地,一旦针对当前块的所有第一阶段候选计算了质量测量,可以对该集合使用总体(population)统计,以帮助在非描述分段与时间-偏移-敏感分段之间进行区分。这可以通过假设来自第一阶段候选列表的质量测量的扩展在这两者之间是不同的来进行。例如,可以是非描述内容具有在某种(阈值以上)程度上匹配同时又有区别的多个候选配对,时间敏感内容具有仅在某个轴上(例如解码路径的线性)但不在其他轴上良好匹配的单个候选对。可以通过利用总体正则化条目扩展匹配-配对描述,来提供对该类型的区别的某种程度的支持。这些可以包括:

[0086] 1) 配对(未正则化)的质量测量。

[0087] 2) 配对的排名,其中排名使用探测块的候选匹配的质量测量。

[0088] 3) 探测块的候选匹配的平均质量测量。

[0089] 4) 探测块的候选匹配的标准差。

[0090] 5) 探测块的候选匹配的逆标准差。

[0091] 6) 通过对探测块的候选的质量测量的均值和标准差对配对的质量测量进行的标准化。

[0092] 7) 通过对探测块的候选的质量测量的均值和标准差来进行正则化的配对相名。

[0093] 也可以通过将附加条目设置为0而省略该步骤。

[0094] 然后将该扩展的匹配-配对描述作为输入提供给真/假匹配配对的分类器。使用(1108)分类器将候选分类为有效或无效的匹配,以及提供匹配的置信度得分。分类器可以是任何多样的形式,不论基于神经网络状结构还是基于线性分类器。如果分类器指示接受该配对,则该配对与置信度得分一起被包括在向第三阶段提供的列表中。否则,其被忽略。该阶段的输出是有效匹配候选的列表。

[0095] 阶段3:候选组合和删减

[0096] 对于一些内容,存在通过第二阶段测试的大量候选,很多包括重叠内容。一个实施方式将候选列表缩减为更加简洁和更加可信,而非独立地列出这些候选中的每一个或忽略超出最终候选列表的某一长度限定的候选,。

[0097] 这如下实现:收集与不同输入视频块相匹配的参考视频部分,以及将这些分量组合为一个组合的匹配。例如,如图12所示,输入视频与参考视频之间的成分匹配(块1、块2、块3)组合成为一个组合的匹配。为了从删减阶段中去除消息,可以施加约束,例如:

[0098] 1) 至少2个不同的输入视频块为组合的匹配提供支持(其中输入视频块与参考视频的部分之间的任何信号配对仅可以支持一个组合匹配)。

[0099] 2) 组合匹配的平均匹配质量在某阈值之上。

[0100] 3) 跨支持组合匹配的块的输入视频与参考部分之间的差异小于某阈值。

[0101] 进行该匹配分组过程的一个方法是贪心算法,其中特定时间位置界限内的所有分量匹配被放入支持列表中。如果该全长列表显示跨分量块的输入视频与参考视频之间的偏

移差异太大,则将异常值(偏移中的)从该匹配中删掉,并且再次考虑该集合,直到集合长度为 2 或更小,或直到组合匹配通过其测试。如果组合匹配通过其测试,则将为组合匹配提供支持的所有成分匹配标志为匹配内容。另外,确定组合匹配的“阴影”中的其他匹配包括匹配内容,其中组合匹配的阴影是在输入视频与相同的参考视频之间具有相似的时间支持并且在两者之间具有相似的时间偏移的匹配部分。将出现阴影的一个示例是视频中具有长期不变的时间并且在匹配部分之间具有间隙的时候。如果组合匹配没有通过其测试,则删掉用作所提出的组合匹配的种子的成分匹配。然后在缩减的未要求的(unclaimed)列表上重复进行组合过程,直到该列表为空。

[0102] 最终阶段过程的一个实施方式被配置为避免匹配相似类型的通用低运动序列。该问题的一个示例是“正在说话的头部”视频。该视频轨道是清楚的,并且通常是清晰的,但是多种传统的指纹生成方法不能将黑色背景、黑色衣服与接近中心的浅色面部的视频内容区分开。例如,第一主题的新闻招待会的视频轨道,“George”可以与 George 举办的所有其他新闻招待会相类似的出现(多个指纹生成过程使用的细节级别),并且可能很容易与第二主体“Bill”举办的新闻招待会看起来非常像。由此,传统的指纹生成可能不期望地指示这些视频之间的匹配。视频通道的这些错误匹配将倾向于生成很多阴影匹配,并且将不会伴随音频通道之间的相应匹配。所以,为了避免报告这些错误匹配,在没有覆盖两个通道并且具有多于阴影-支持匹配的某个比率阈值的那些匹配上,过程可以降低该匹配中的置信度,或者,如果置信度可能太低,则将匹配全部移除。以这种方式,根据整个过程创建并返回了少量的具有多块支持的匹配。

[0103] 匹配结果

[0104] 最终输出提供已经确定为与输入视频(或输入视频的部分)相匹配的参考视频(或参考视频的标识部分)的列表。该确定可以用于多种目的。首先,如果匹配模块 108 确定上传的输入视频 102 是已经处于参考视频集合中的视频的副本,可以丢弃上传的输入视频 102 以便节省存储空间。其次,输入视频 102 可以用于探测例如版权保护的视频内容的参考视频。然后可以对这些视频进行标记或从参考视频集合中移除。有益地,所描述的系统和方法即使在紧迫的时间限制下和/或使用有限量的存储器,也能够有效地和精确地检测匹配。

[0105] 附加备选实施方式

[0106] 以上描述的一些部分按照信息操作的算法和符号表示给出了本发明的特征。这些算法描述和表示是本领域技术人员使用的方法,以便将其工作的本质最有效地传达给领域内的其他技术人员。这些操作,虽然功能性地或逻辑地进行了描述,但是理解为由计算机程序来实现。另外,已经多次证明,在不失一般性的情况下,将这些操作的排列作为模块或代码设备是适当的。

[0107] 然而,应当记住,所有这些和相似术语将要与适合的物理量相关联,并且仅是应用于这些量的方便标记。除非在本说明书中另有明显的特别说明,应当理解,在说明书全文中,使用术语诸如“处理”或“运算”或“计算”或“确定”或“显示”等等的讨论,是指在计算机系统存储器或寄存器或其他这样的信息存储、传输或显示设备内,计算机系统或类似的电子计算设备操作和转换表示为物理(电子)量的数据的数据的动作和过程。

[0108] 本发明的某些方面包括此处以算法形式描述的过程步骤和指令。应当注意,本发

明的过程步骤和指令可以在软件、固件或硬件中实现,并且在软件中实现时,可以进行下载以在实时网络操作系统使用的不同平台上存在并进行操作。

[0109] 本发明还涉及用于执行此处的操作的装置。该装置可以为所需要的目的而特别构造,或其可以包括由存储在计算机中的计算机程序可选地激活或重配置的通用计算机。这样的计算机程序可以存储在计算机可读存储介质中,例如但不限于任何类型的盘,包括软盘、光盘、CD-ROM、磁光盘、只读存储器 (ROM)、随机存取存储器 (RAM)、EPROM、EEPROM、磁卡或光卡、专用集成电路 (ASIC),或适于存储电子指令的任何类型的介质,并且每个都耦合至计算机系统总线。另外,本说明书中涉及的计算机可以包括单个处理器或可以是采用多处理器设计的架构以提高计算性能。

[0110] 此处提出的算法和显示并不固有地涉及任何特定计算机或其他装置。各种通用系统也可以与根据此处教导的程序一起使用,或可以证明方便构造更专用的装置来执行所需要的方法步骤。多种这样的系统所需要的结构将在以上描述中出现。另外,本发明不参考任何特定程序语言进行描述。可以理解,可以使用多种程序语言来实现此处描述的本发明的教导,并且对具体语言的任何引用提供用于可实施性和本发明的最佳模式的公开。

[0111] 最后,应当注意,本发明书中使用的语言原则上是为了易读和教导的目的而选择的,而不是为了界定或限制发明主题而选择的。因此,本发明的公开是示意性的,而不是对本发明的范围的限制。

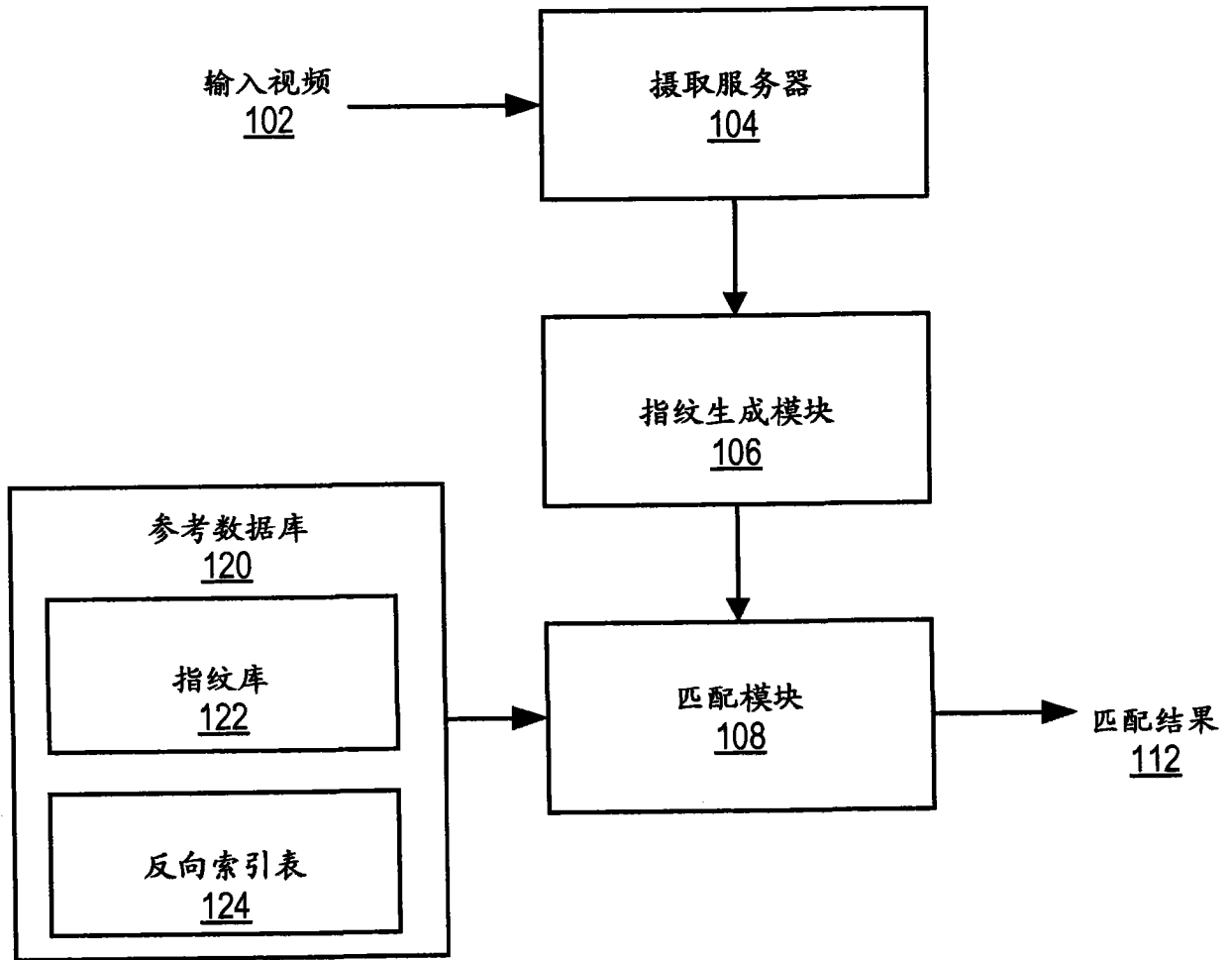


图 1

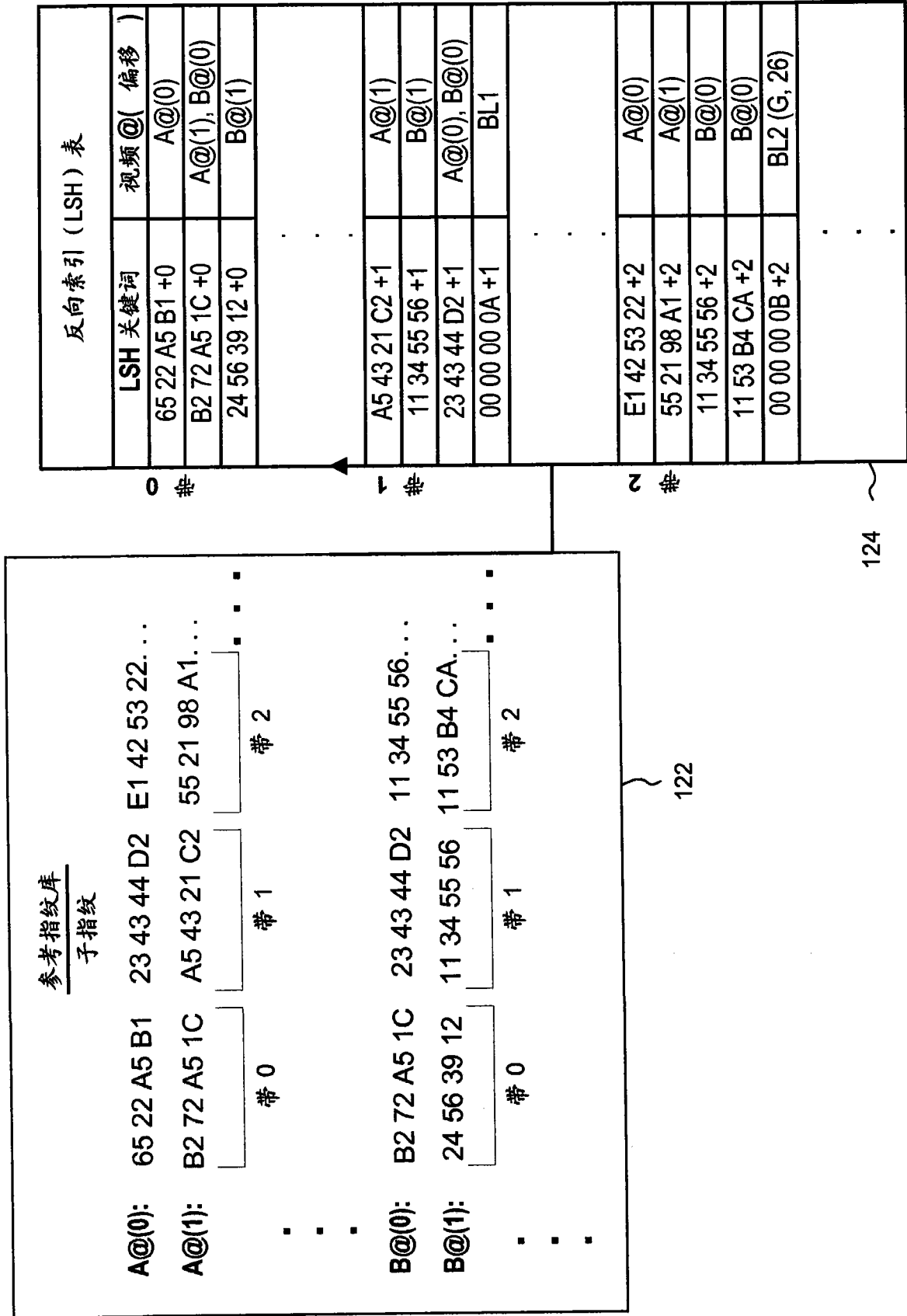


图 2

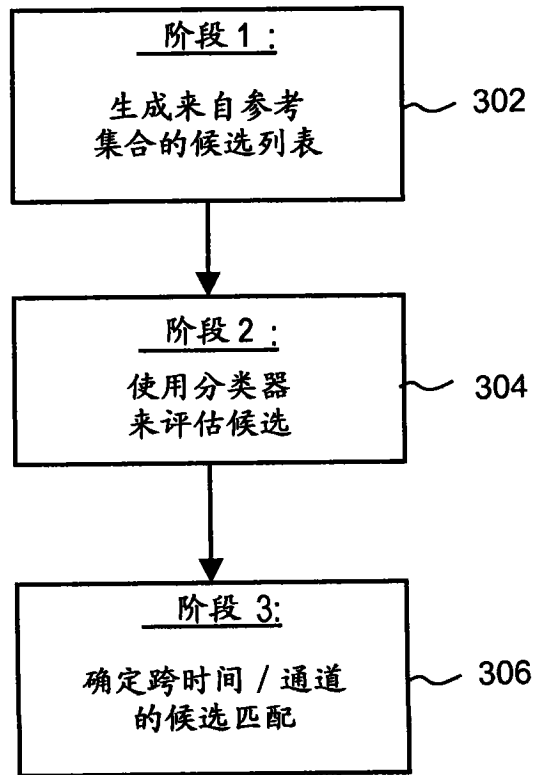


图 3

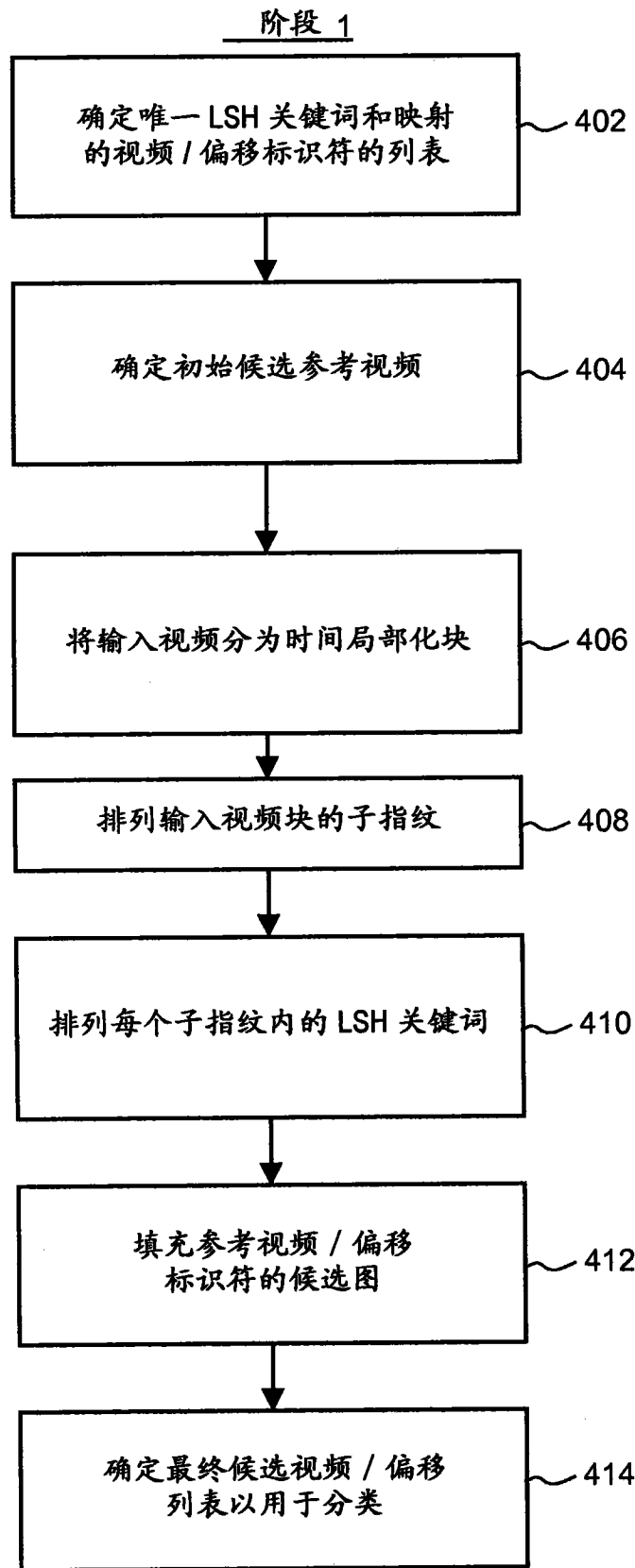


图 4

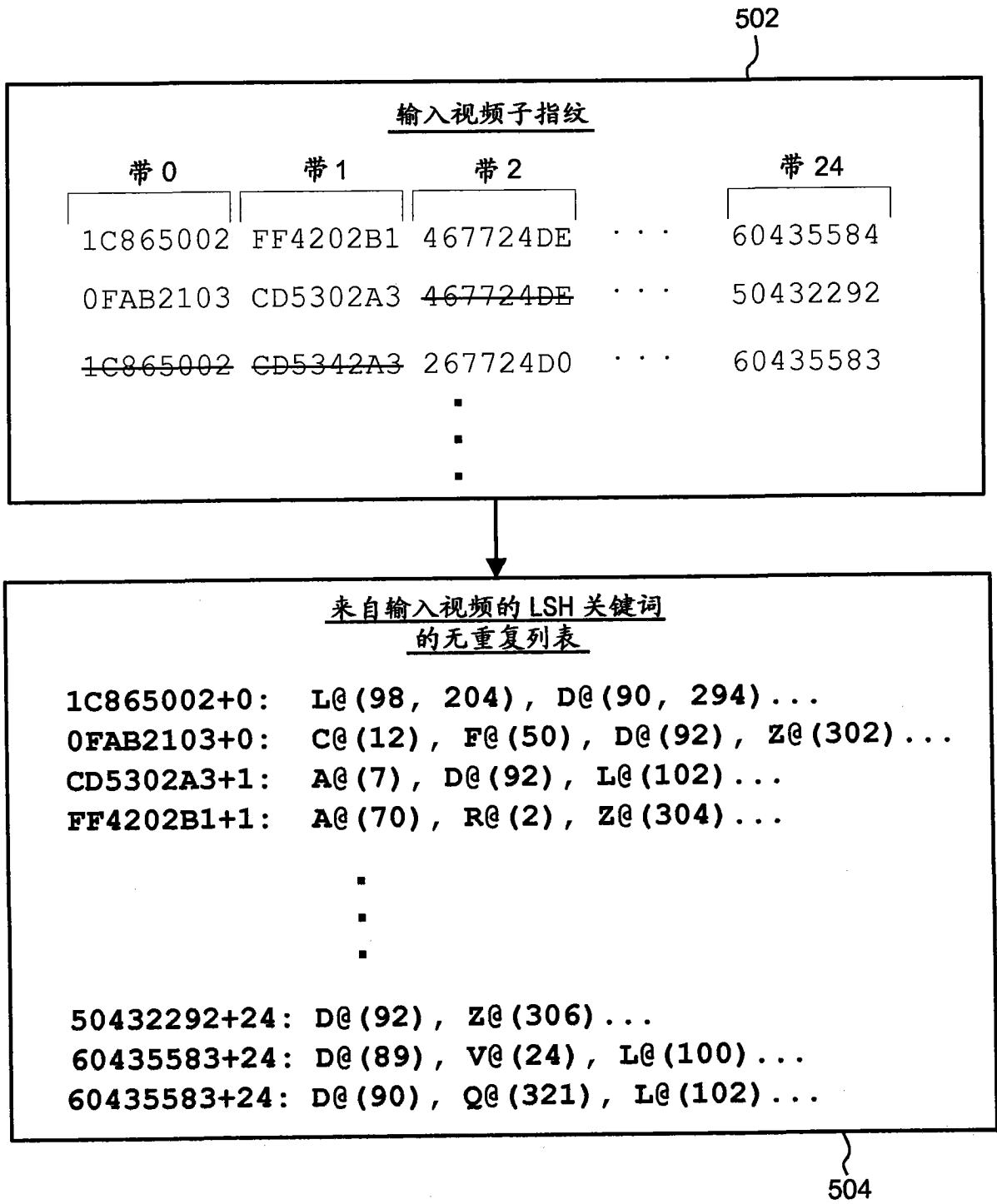


图 5

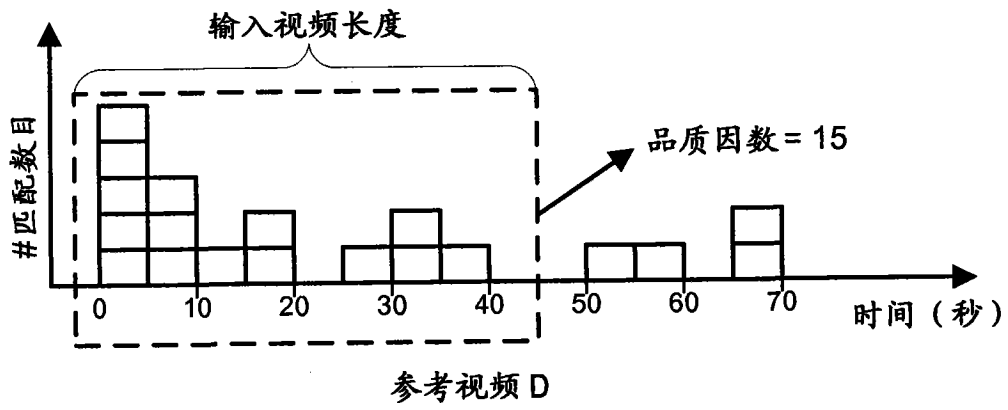


图 6

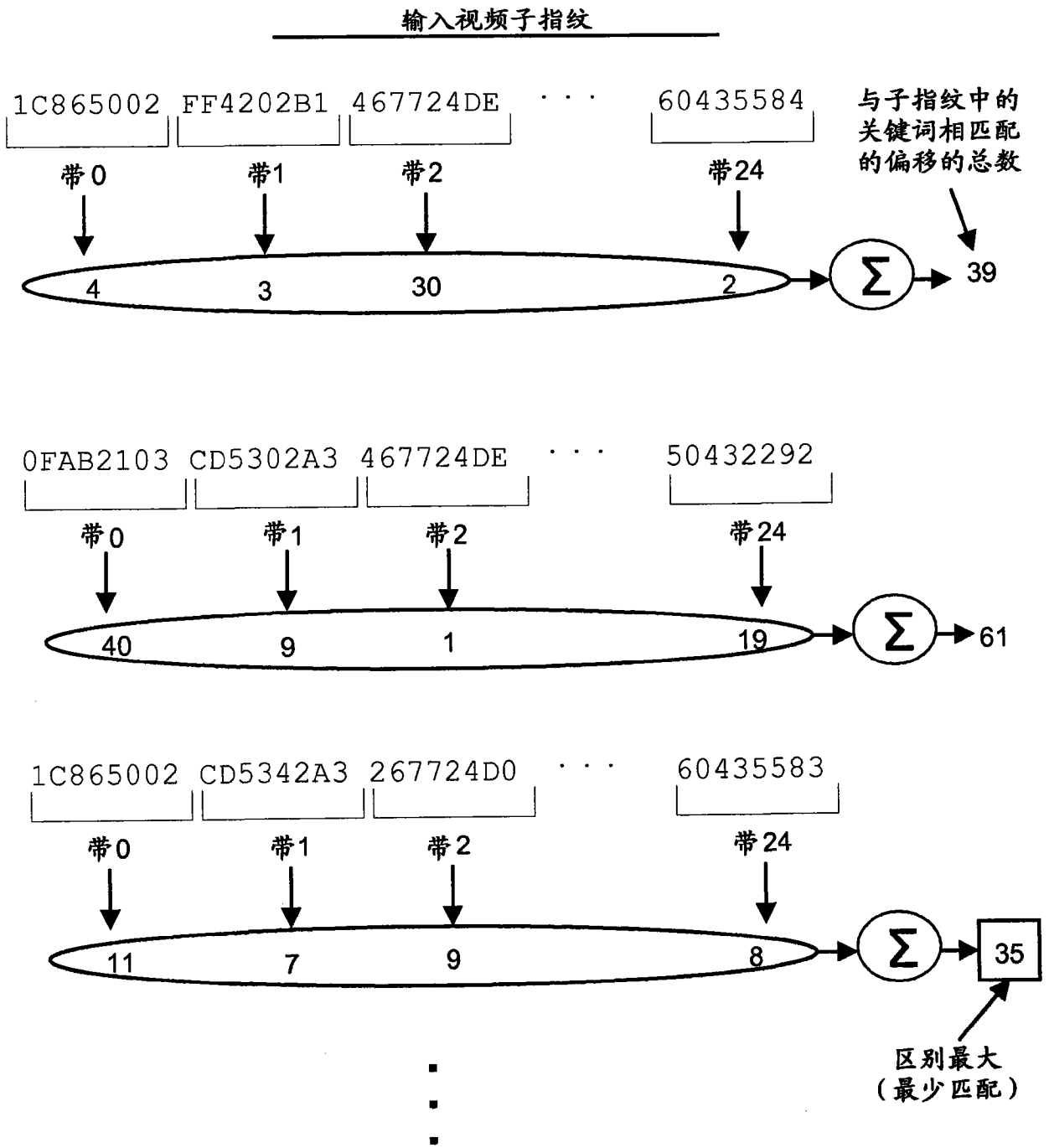


图 7

子指纹中的匹配关键词的计数

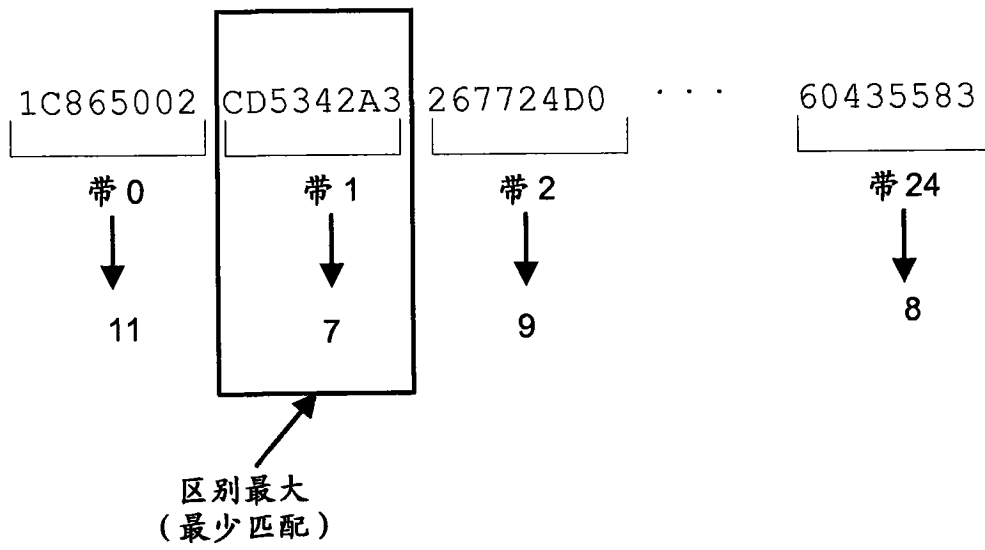
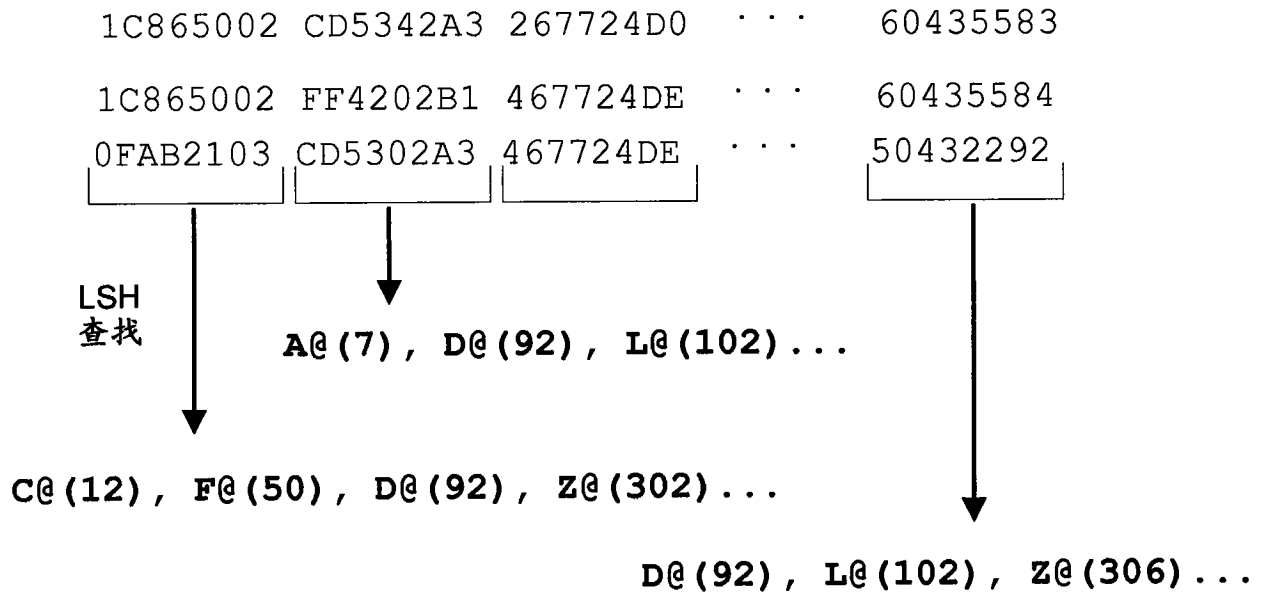


图 8

有序的输入视频子指纹



| LSH 查找结果 | |
|----------|---|
| A@(7): | 1 |
| D@(92): | 3 |
| L@(102): | 2 |
| C@(12): | 1 |
| F@(50): | 1 |
| Z@(302): | 1 |
| Z@(306): | 1 |
| . | . |
| . | . |
| . | . |

图 9

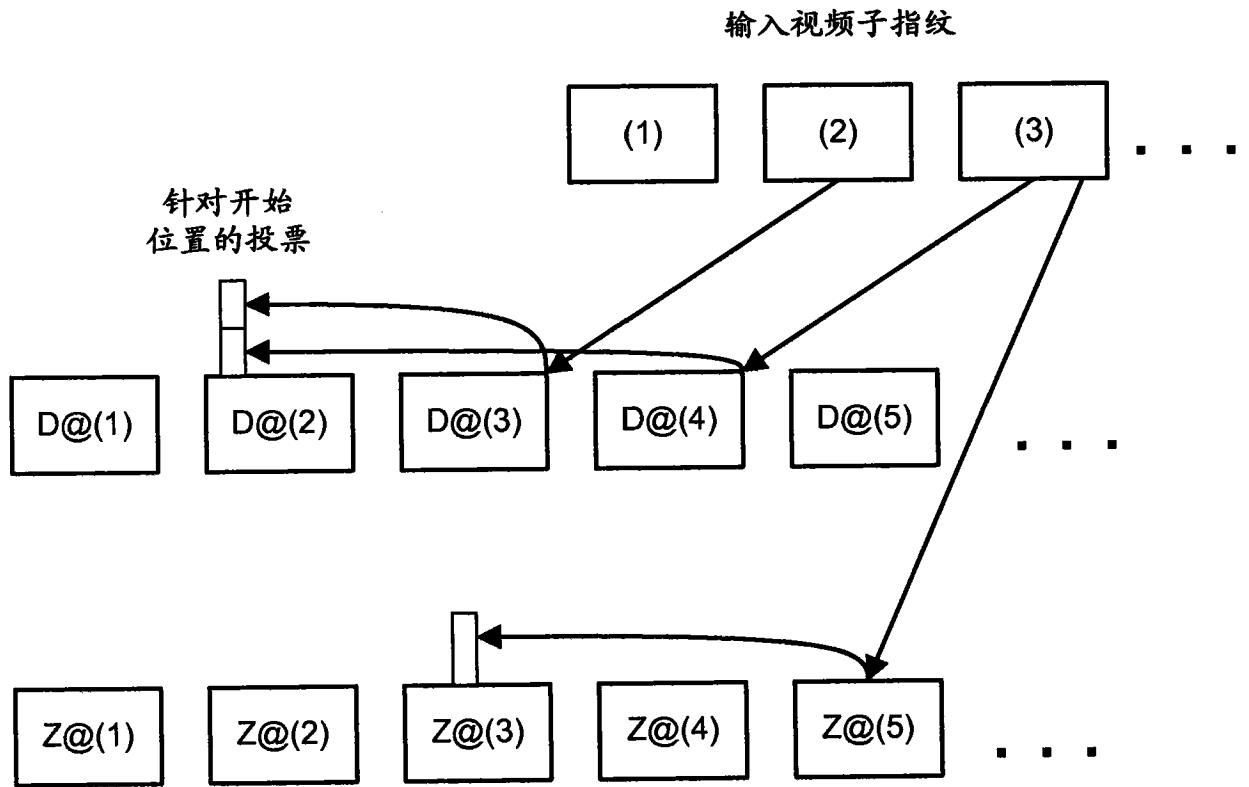


图 10

阶段 2

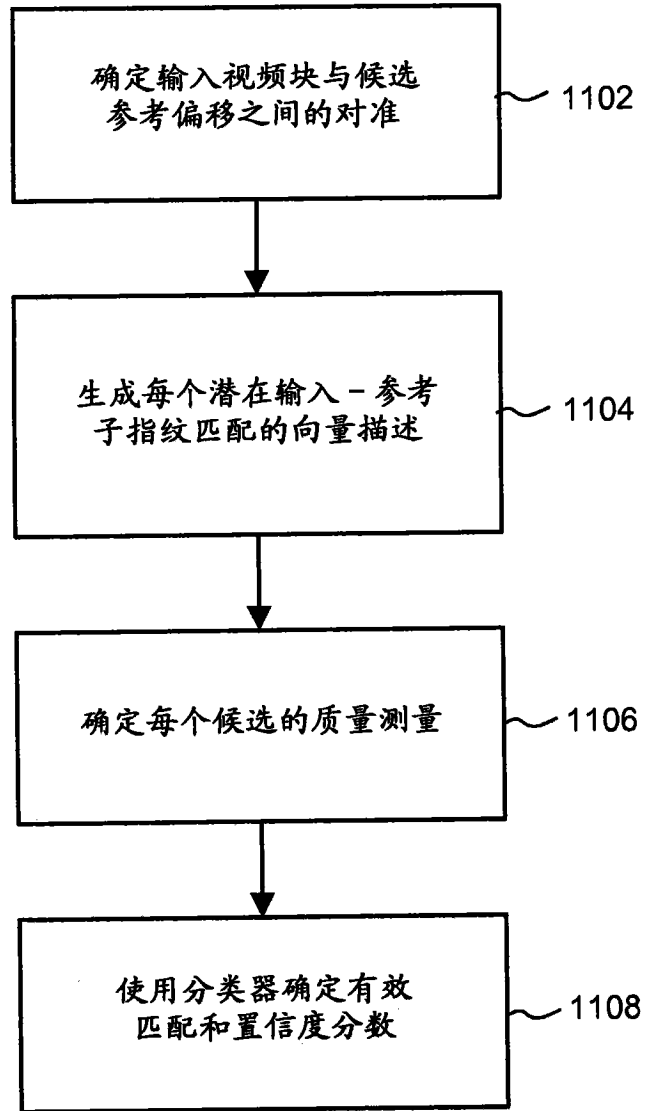


图 11

阶段 3

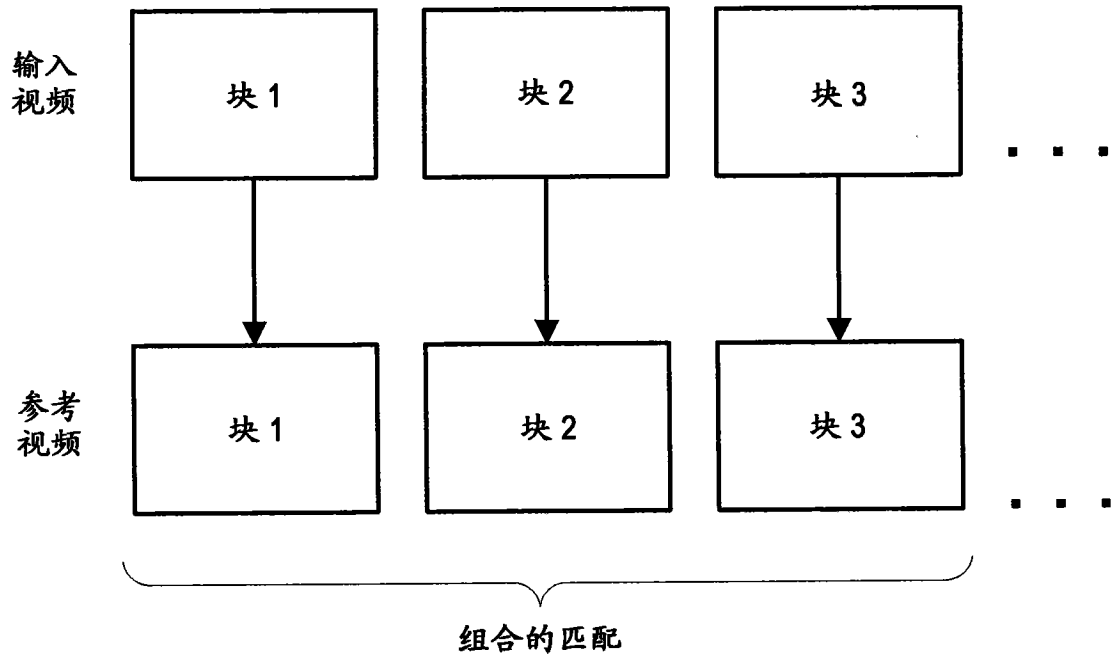


图 12