(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107609149 B
(45)授权公告日 2020.06.19

(21)申请号 201710861497.X

(22)申请日 2017.09.21

(65)同一申请的已公布的文献号
申请公布号 CN 107609149 A

(43)申请公布日 2018.01.19

(73)专利权人 北京奇艺世纪科技有限公司
地址 100080 北京市海淀区北一街2号爱奇艺创新大厦10、11层

(72)发明人 李冠楠

(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319
代理人 莎日娜

(51)Int.Cl.
G06F 16/70(2019.01)
G06F 16/783(2019.01)

(56)对比文件
CN 103942337 A,2014.07.23,
CN 101101590 A,2008.01.09,
CN 101281534 A,2008.10.08,
CN 102024033 A,2011.04.20,
CN 103488764 A,2014.01.01,
CN 101577137 A,2009.11.11,
CN 101021855 A,2007.08.22,
CN 105828179 A,2016.08.03,
CN 102799605 A,2012.11.28,
CN 104093090 A,2014.10.08,
CN 104731938 A,2015.06.24,
CN 107066477 A,2017.08.18,
CN 103092958 A,2013.05.08,
CN 101079044 A,2007.11.28,
US 2003215110 A1,2003.11.20,

审查员 孟驭旋

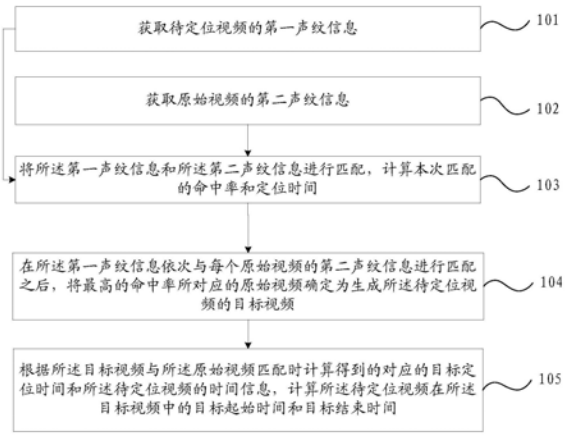
权利要求书4页 说明书11页 附图4页

(54)发明名称

一种视频定位方法和装置

(57)摘要

本发明提供了一种视频定位方法和装置，该方法包括：获取待定位视频的第一声纹信息；获取原始视频的第二声纹信息；将第一声纹信息和第二声纹信息进行匹配，计算本次匹配的命中率和定位时间；在第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成待定位视频的目标视频；根据目标视频与原始视频匹配时计算得到的目标定位时间和待定位视频的时间信息，计算待定位视频在目标视频中的目标起始时间和目标结束时间。本发明能够提升视频定位的准确度。



CN 107609149 B

1.一种视频定位方法,其特征在于,包括:

获取待定位视频的第一声纹信息;

获取原始视频的第二声纹信息;

将所述第一声纹信息和所述第二声纹信息进行匹配,计算本次匹配的命中率和定位时间;

在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后,将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频;

根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息,计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间;

其中,本次匹配的定位时间是按照以下方式确定的:

若t_1与t_2的差小于或等于第四预设阈值,则将$(t_1+t_2)/2$确定为本次匹配的定位时间;

若H_t^1与H_t^2的差大于第三预设阈值,且t_1与t_2的差大于第四预设阈值,则将t_1确定为本次匹配的定位时间;

其中,H_t^1表示最高特征命中率,H_t^2表示低于H_t^1的特征命中率,t1表示H_t^1的匹配时间,t2表示H_t^2的匹配时间,所述匹配时间为相互匹配的两个声纹特征在各自所属视频上所处的两个时间点之间的时间偏移。

2.根据权利要求1所述的方法,其特征在于,所述获取待定位视频的第一声纹信息,包括:

获取待定位视频的音频信息;

对所述音频信息作分类处理,获取所述音频信息中属于目标类型的目标音频信息;

提取所述目标音频信息的第一声纹信息。

3.根据权利要求1所述的方法,其特征在于,所述将所述第一声纹信息和所述第二声纹信息进行匹配,计算本次匹配的命中率和定位时间,包括:

将所述第一声纹信息和所述第二声纹信息进行匹配,得到匹配结果;

按照预设条件判断所述匹配结果是否有效;

若所述匹配结果有效,则计算本次匹配的命中率和定位时间;

若所述匹配结果无效,则将本次匹配的命中率记为零。

4.根据权利要求3所述的方法,其特征在于,所述第一声纹信息包括多个第一声纹特征,所述第二声纹信息包括多个第二声纹特征,所述将所述第一声纹信息和所述第二声纹信息进行匹配,得到匹配结果,包括:

将所述多个第一声纹特征中的每个第一声纹特征分别与所述多个第二声纹特征中的每个第二声纹特征进行匹配,得到各第一声纹特征的匹配结果;

其中,所述匹配结果包括:匹配分数以及匹配时间。

5.根据权利要求4所述的方法,其特征在于,所述按照预设条件判断所述匹配结果是否有效,包括:

判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}是否大于或等于第一预设阈值;

若M_{max}大于或等于第一预设阈值，则统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间t_i的第一声纹特征的个数；

对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1, t_2, t_3 \ldots t_n$，n为本次匹配的匹配时间总数；

在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数；

判断H_t^1个匹配分数中的最大值Max_{t1}是否大于或等于第二预设阈值；

若Max_{t1}大于或等于第二预设阈值，则将H_t^1与H_t^2的差与第三预设阈值进行比较，以及将t_1与t_2的差与第四预设阈值进行比较；

若t_1与t_2的差小于或等于第四预设阈值，或者t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值时，确定本次匹配的匹配结果有效；

否则，确定本次匹配的匹配结果无效。

6.根据权利要求5所述的方法，其特征在于，所述若所述匹配结果有效，则计算本次匹配的命中率，包括：

若t_1与t_2的差小于或等于第四预设阈值，则将H_t^1确定为本次匹配的命中率；

若H_t^1与H_t^2的差大于第三预设阈值，且t_1与t_2的差大于第四预设阈值，则将H_t^1确定为本次匹配的命中率。

7.根据权利要求2所述的方法，其特征在于，所述根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间，包括：

获取所述目标音频信息在所述待定位视频中对应的起始时间和结束时间；

计算所述起始时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标起始时间；

计算所述结束时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标结束时间。

8.一种视频定位装置，其特征在于，包括：

第一获取模块，用于获取待定位视频的第一声纹信息；

第二获取模块，用于获取原始视频的第二声纹信息；

匹配模块，用于将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间；

确定模块，用于在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频；

计算模块，用于根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间；

其中，本次匹配的定位时间是按照以下方式确定的：

若t_1与t_2的差小于或等于第四预设阈值，则将(t_1+t_2)/2确定为本次匹配的定位时间；

若H_t^1与H_t^2的差大于第三预设阈值，且t_1与t_2的差大于第四预设阈值，则将t_1确定为本次匹配的定位时间；

其中，H_t^1表示最高特征命中率，H_t^2表示低于H_t^1的特征命中率，t1表示H_t^1的匹配时间，t2表示H_t^2的匹配时间，所述匹配时间为相互匹配的两个声纹特征在各自所属视频上所处的两个时间点之间的时间偏移。

9.根据权利要求8所述的装置，其特征在于，所述第一获取模块包括：

第一获取子模块，用于获取待定位视频的音频信息；

分类子模块，用于对所述音频信息作分类处理，获取所述音频信息中属于目标类型的目标音频信息；

提取子模块，用于提取所述目标音频信息的第一声纹信息。

10.根据权利要求8所述的装置，其特征在于，所述匹配模块，包括：

匹配子模块，用于将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果；

判断子模块，用于按照预设条件判断所述匹配结果是否有效；

第一计算子模块，用于若所述匹配结果有效，则计算本次匹配的命中率和定位时间；

第二计算子模块，用于若所述匹配结果无效，则将本次匹配的命中率记为零。

11.根据权利要求10所述的装置，其特征在于，所述第一声纹信息包括多个第一声纹特征，所述第二声纹信息包括多个第二声纹特征，所述匹配子模块包括：

匹配单元，用于将所述多个第一声纹特征中的每个第一声纹特征分别与所述多个第二声纹特征中的每个第二声纹特征进行匹配，得到各第一声纹特征的匹配结果；

其中，所述匹配结果包括：匹配分数以及匹配时间。

12.根据权利要求11所述的装置，其特征在于，所述判断子模块包括：

第一判断单元，用于判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}是否大于或等于第一预设阈值；

统计单元，用于若M_{max}大于或等于第一预设阈值，则统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间t_i的第一声纹特征的个数；

排序单元，用于对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1$，$t_2$，$t_3$...$t_n$，n为本次匹配的匹配时间总数；

获取单元，用于在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数；

第二判断单元，用于判断H_t^1个匹配分数中的最大值Max_{t1}是否大于或等于第二预设阈值；

比较单元，用于若Max_{t1}大于或等于第二预设阈值，则将H_t^1与H_t^2的差与第三预设阈值进行比较，以及将t_1与t_2的差与第四预设阈值进行比较；

第一确定单元，用于若t_1与t_2的差小于或等于第四预设阈值，或者t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值时，确定本次匹配的匹配结果有效；

第二确定单元，用于若M_{max}小于第一预设阈值、或若Max_{t1}小于第二预设阈值、或若H_t^1与H_t^2的差小于第三预设阈值且t_1与t_2的差大于第四预设阈值，则确定本次匹配的匹配结果无效。

13.根据权利要求12所述的装置，其特征在于，所述第一计算子模块包括：

第三确定单元，用于若t_1与t_2的差小于或等于第四预设阈值，则将H_t^1确定为本次匹配的命中率；

第四确定单元，用于若H_t^1与H_t^2的差大于第三预设阈值，且t_1与t_2的差大于第四预设阈值，则将H_t^1确定为本次匹配的命中率。

14.根据权利要求9所述的装置，其特征在于，所述计算模块包括：

第二获取子模块，用于获取所述目标音频信息在所述待定位视频中对应的起始时间和结束时间；

第三计算子模块，用于计算所述起始时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标起始时间；

第四计算子模块，用于计算所述结束时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标结束时间。

一种视频定位方法和装置

技术领域

[0001] 本发明涉及视频处理技术领域，特别是涉及一种视频定位方法和装置。

背景技术

[0002] 目前，网络上存在着大量的短视频，这些短视频均来源于原始视频。在已发布的这些短视频中，当发生短视频部分损坏、视频信息不完整(例如原始视频发布时间不明、视频内容不完整)等情况时，则需要在原始素材库中查找到该短视频所属的原始视频，以及该短视频在该原始视频中所对应的视频片段，从而实现对短视频的二次加工或重新制作。

[0003] 而在现有技术中，主要是利用视频信息(例如感知哈希特征)来寻找短视频的视频来源，以及短视频在原始视频中的所在位置。但是，在短视频的加工过程中往往会对视频分辨率和编码格式等视频信息进行更改，从而导致短视频的画面内容与原始视频素材库中的原始视频存在差异。所以，基于视频信息的视频定位较为困难，难以确定短视频的素材来源，以及其在原始视频中的起始时间和结束时间。

发明内容

[0004] 本发明提供了一种视频定位方法和装置，以解决现有技术中的视频定位方案所存在的视频定位不准确的问题。

[0005] 为了解决上述问题，根据本发明的一个方面，本发明公开了一种视频定位方法，包括：

[0006] 获取待定位视频的第一声纹信息；

[0007] 获取原始视频的第二声纹信息；

[0008] 将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间；

[0009] 在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频；

[0010] 根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间。

[0011] 可选地，所述获取待定位视频的第一声纹信息，包括：

[0012] 获取待定位视频的音频信息；

[0013] 对所述音频信息作分类处理，获取所述音频信息中属于目标类型的目标音频信息；

[0014] 提取所述目标音频信息的第一声纹信息。

[0015] 可选地，所述将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间，包括：

[0016] 将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果；

[0017] 按照预设条件判断所述匹配结果是否有效；

[0018] 若所述匹配结果有效，则计算本次匹配的命中率和定位时间；

[0019] 若所述匹配结果无效，则将本次匹配的命中率记为零。

[0020] 可选地，所述第一声纹信息包括多个第一声纹特征，所述第二声纹信息包括多个第二声纹特征，所述将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果，包括：

[0021] 将所述多个第一声纹特征中的每个第一声纹特征分别与所述多个第二声纹特征中的每个第二声纹特征进行匹配，得到各第一声纹特征的匹配结果；

[0022] 其中，所述匹配结果包括：匹配分数以及匹配时间，所述匹配时间为相互匹配的两个声纹特征在各自所属视频上所处的两个时间点之间的时间偏移。

[0023] 可选地，所述按照预设条件判断所述匹配结果是否有效，包括：

[0024] 判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}是否大于或等于第一预设阈值；

[0025] 若M_{max}大于或等于第一预设阈值，则统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间t_i的第一声纹特征的个数；

[0026] 对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1, t_2, t_3 \cdots t_n$，n为本次匹配的匹配时间总数；

[0027] 在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数；

[0028] 判断H_t^1个匹配分数中的最大值Max_{t1}是否大于或等于第二预设阈值；

[0029] 若Max_{t1}大于或等于第二预设阈值，则将H_t^1与H_t^2的差与第三预设阈值进行比较，以及将t_1与t_2的差与第四预设阈值进行比较；

[0030] 若t_1与t_2的差小于或等于第四预设阈值，或者t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值时，确定本次匹配的匹配结果有效；

[0031] 否则，确定本次匹配的匹配结果无效。

[0032] 可选地，所述若所述匹配结果有效，则计算本次匹配的命中率和定位时间，包括：

[0033] 若t_1与t_2的差小于或等于第四预设阈值，则将H_t^1确定为本次匹配的命中率，将$(t_1+t_2)/2$确定为本次匹配的定位时间；

[0034] 若H_t^1与H_t^2的差大于第三预设阈值，且t_1与t_2的差大于第四预设阈值，则将H_t^1确定为本次匹配的命中率，将t_1确定为本次匹配的定位时间。

[0035] 可选地，所述根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间，包括：

[0036] 获取所述目标音频信息在所述待定位视频中对应的起始时间和结束时间；

[0037] 计算所述起始时间和所述目标定位时间之和，得到所述待定位视频在所述目标视

频中的目标起始时间；

[0038] 计算所述结束时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标结束时间。

[0039] 根据本发明的另一方面，本发明还公开了一种视频定位装置，包括：

[0040] 第一获取模块，用于获取待定位视频的第一声纹信息；

[0041] 第二获取模块，用于获取原始视频的第二声纹信息；

[0042] 匹配模块，用于将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间；

[0043] 确定模块，用于在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频；

[0044] 计算模块，用于根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间。

[0045] 可选地，所述第一获取模块包括：

[0046] 第一获取子模块，用于获取待定位视频的音频信息；

[0047] 分类子模块，用于对所述音频信息作分类处理，获取所述音频信息中属于目标类型的目标音频信息；

[0048] 提取子模块，用于提取所述目标音频信息的第一声纹信息。

[0049] 可选地，所述匹配模块，包括：

[0050] 匹配子模块，用于将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果；

[0051] 判断子模块，用于按照预设条件判断所述匹配结果是否有效；

[0052] 第一计算子模块，用于若所述匹配结果有效，则计算本次匹配的命中率和定位时间；

[0053] 第二计算子模块，用于若所述匹配结果无效，则将本次匹配的命中率记为零。

[0054] 可选地，所述第一声纹信息包括多个第一声纹特征，所述第二声纹信息包括多个第二声纹特征，所述匹配子模块包括：

[0055] 匹配单元，用于将所述多个第一声纹特征中的每个第一声纹特征分别与所述多个第二声纹特征中的每个第二声纹特征进行匹配，得到各第一声纹特征的匹配结果；

[0056] 其中，所述匹配结果包括：匹配分数以及匹配时间，所述匹配时间为相互匹配的两个声纹特征在各自所属视频上所处的两个时间点之间的时间偏移。

[0057] 可选地，所述判断子模块包括：

[0058] 第一判断单元，用于判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}是否大于或等于第一预设阈值；

[0059] 统计单元，用于若M_{max}大于或等于第一预设阈值，则统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间t_i的第一声纹特征的个数；

[0060] 排序单元，用于对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1, t_2, t_3 \cdots t_n$，n为本次匹配的匹配时间总

数；

[0061] 获取单元，用于在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数；

[0062] 第二判断单元，用于判断H_t^1个匹配分数中的最大值Max_{t1}是否大于或等于第二预设阈值；

[0063] 比较单元，用于若Max_{t1}大于或等于第二预设阈值，则将H_t^1与H_t^2的差与第三预设阈值进行比较，以及将t_1与t_2的差与第四预设阈值进行比较；

[0064] 第一确定单元，用于若t_1与t_2的差小于或等于第四预设阈值，或者t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值时，确定本次匹配的匹配结果有效；

[0065] 第二确定单元，用于若M_{max}小于第一预设阈值、或若Max_{t1}小于第二预设阈值、或若H_t^1与H_t^2的差小于第三预设阈值且t_1与t_2的差大于第四预设阈值，则确定本次匹配的匹配结果无效。

[0066] 可选地，所述第一计算子模块包括：

[0067] 第三确定单元，用于若t_1与t_2的差小于或等于第四预设阈值，则将H_t^1确定为本次匹配的命中率，将$(t_1+t_2)/2$确定为本次匹配的定位时间；

[0068] 第四确定单元，用于若H_t^1与H_t^2的差大于第三预设阈值，且t_1与t_2的差大于第四预设阈值，则将H_t^1确定为本次匹配的命中率，将t_1确定为本次匹配的定位时间。

[0069] 可选地，所述计算模块包括：

[0070] 第二获取子模块，用于获取所述目标音频信息在所述待定位视频中对应的起始时间和结束时间；

[0071] 第三计算子模块，用于计算所述起始时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标起始时间；

[0072] 第四计算子模块，用于计算所述结束时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标结束时间。

[0073] 与现有技术相比，本发明包括以下优点：

[0074] 本发明通过利用待定位视频以及原始视频的声纹信息来确定该待定位视频源自哪个原始视频，以及其在原始视频中所在的准确位置，从而能够有效的恢复待定位视频在原始视频中的时间信息，匹配过程中无关视频信息，基于声纹信息的视频定位，提升了视频定位的准确度。

[0075] 此外，本发明使用声纹匹配得分、匹配时间及命中率判断匹配结果的有效性，使得定位精度可达到秒级。

附图说明

[0076] 图1是本发明的一种视频定位方法实施例的步骤流程图；

[0077] 图2是本发明的另一种视频定位方法实施例的流程图；

[0078] 图3是本发明的一种视频定位方法实施例的子流程图；

[0079] 图4是本发明的一种视频定位装置实施例的结构框图。

具体实施方式

[0080] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0081] 参照图1，示出了本发明的一种视频定位方法实施例的步骤流程图，具体可以包括如下步骤：

[0082] 步骤101，获取待定位视频的第一声纹信息；

[0083] 其中，为了确定某个待定位视频(例如短视频或者视频片段等)源自视频素材中的哪个原始视频，本发明实施例可以获取该待定位视频的声纹信息。

[0084] 其中，短视频是时间长度小于某个时间阈值(例如10分钟等等)的视频。

[0085] 步骤102，获取原始视频的第二声纹信息；

[0086] 其中，视频素材中包括很多产生例如短视频的原始视频，这里可以获取待检测的某个原始视频的声纹信息。

[0087] 步骤103，将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间；

[0088] 步骤104，在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频；

[0089] 其中，在将待定位视频的声纹信息与素材库中的每个原始视频的声纹信息都进行匹配之后，本发明实施例就可以将命中率最高的那次匹配所对应的原始视频确定为生成该待定位视频的目标视频。

[0090] 步骤105，根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间。

[0091] 其中，这里将目标视频与所述原始视频匹配时计算得到的定位时间，记为目标定位时间，所谓目标定位时间是一个时间值(例如5、6、7等数值)；而待定位视频的时间信息，则包括该待定位视频中每个视频帧所对应的时间点。

[0092] 最后，就可以根据在将待定位视频的声纹信息与该目标视频的声纹信息进行匹配后所得到的目标定位时间，以及该待定位视频的时间信息，计算该待定位视频在该目标视频中的具体位置，即，该待定位视频源自该目标视频中哪一段时间的视频片段。

[0093] 借助于本发明上述实施例的技术方案，本发明通过利用待定位视频以及原始视频的声纹信息来确定该待定位视频源自哪个原始视频，以及其在原始视频中所在的准确位置，从而能够有效的恢复待定位视频在原始视频中的时间信息，匹配过程中无关视频信息，基于声纹信息的视频定位，提升了视频定位的准确度。

[0094] 下面结合图2所示的本发明另一实施例的视频定位方法的流程图，以及图1来对本发明的上述技术方案进行详细阐述。

[0095] 其中，在一个实施例中，在执行步骤101时，可以对输入的待定位视频(下文以短视频为例进行说明)进行音视频分离处理，从而获得输入视频的视频画面和音频数据，然后再

对该音频数据进行声纹提取，从而获取该短视频的第一声纹信息。

[0096] 其中，音视频分离处理可以预先完成，也可以在需要声纹信息时完成，具体时机本发明并不限定。

[0097] 其中，在一个实施例中，在执行步骤101时，可以通过获取待定位视频的音频信息；对所述音频信息作分类处理，获取所述音频信息中属于目标类型的目标音频信息；提取所述目标音频信息(即，新闻音频数据)的第一声纹信息。

[0098] 其中，本例中的短视频为一个新闻短视频，如图2所示，由于新闻短视频中只有新闻主播的语音可以作为声纹匹配的准确依据，因此为了提升声纹匹配的准确度，这里需要对新闻短视频的音频信息作音频分类来获取属于语音类别的音频信息(即图2中的新闻报道语音)，并提取该音频信息的声纹信息。

[0099] 具体而言，短视频的片头片尾通常具有静音部分和音乐部分，因此，这里需要从短视频的音频信息中截取只包含语音类别的音频信息的声纹信息，即进行新闻报道语音的声纹信息提取。

[0100] 其中，可以采用计算RMS能量的方法来对音频数据进行静音检测，从而删除该音频数据中的片头及片尾的静音片段；其中，静音片段的能量至会小于预设能量阈值(例如-60)，从而实现静音片段的删除。

[0101] 然后，对删除静音片段后的音频数据进行逐帧的音频分类操作，寻找具有连续语音类别的片断作为输入的短视频的新闻内容，新闻内容持续时间的典型取值是30秒至3分钟中的任意一个值。

[0102] 在实际应用中可以采用时间窗(例如5s)来对删除静音片段后的音频数据进行5s音频片段的获取，并将其放入分类器中判断该5s音频片段属于语音类别还是音乐类别，那么如果属于语音类别则保存，而如果属于音乐类别则删除。然后，在对保存的这些属于语音类别的音频片段中，查看时间相邻的音频数据总共的持续时间是否超过15秒，如果是，则截取这些音频片段，。这样，连续截取到的这些音频片段的持续时间达到该时间阈值范围(30秒～3分钟)后，即得到了短视频的音频数据中属于新闻内容的音频数据。

[0103] 其中，在一个实施例中，在执行步骤102时，如图2所示，视频素材库中存储有很多原始视频素材，本发明实施例可以预先对视频素材库中的各视频文件进行音视频分离处理，从而获得各个原始视频素材的音频信息，接着再提取各原始视频素材的声纹信息，并对各原始视频素材的时间信息与以及声纹信息进行存储，用于后续查询；

[0104] 在本步骤中，就可以查询某个原始视频素材的声纹信息来与步骤101中的声纹信息进行匹配。

[0105] 其中，在一个实施例中，在执行步骤103时，可以将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果；按照预设条件判断所述匹配结果是否有效；若所述匹配结果有效，则计算本次匹配的命中率和定位时间；若所述匹配结果无效，则将本次匹配的命中率记为零。

[0106] 其中，所述第一声纹信息包括多个第一声纹特征，所述第二声纹信息包括多个第二声纹特征。

[0107] 那么在执行将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果的步骤时，就可以使用现有技术中的声纹匹配工具来将所述多个第一声纹特征中的每个第一

声纹特征分别与所述多个第二声纹特征中的每个第二声纹特征进行匹配，得到各第一声纹特征的匹配结果。

[0108] 其中，所述匹配结果包括：匹配分数以及匹配时间，其中，所述匹配时间为相互匹配的两个声纹特征在各自所属视频上所处的两个时间点之间的时间偏移。可选地，该匹配结果也可以包括与第一声纹特征匹配的目标第二声纹特征。其中，在计算匹配分数和匹配时间时，需要借助于与第一声纹特征匹配的目标第二声纹特征。

[0109] 具体而言，例如短视频的声纹特征有50个，当前匹配的原始视频素材的声纹特征有100个，那么通过声纹匹配工具，可以对短视频的50个声纹特征中的每个声纹特征都在上述100个声纹特征中匹配到一个声纹特征，从而输出来自原始视频素材的与上述短视频的50个声纹特征一一匹配的50个声纹特征，以及相互匹配的这两个声纹特征之间的匹配度(即匹配分数)，以及匹配时间。

[0110] 其中，就匹配时间而言，例如短视频的声纹特征1与素材1的声纹特征2匹配成功，那么该声纹特征1在短视频所处的时间位置t1与该声纹特征2在原始视频素材1中所处的时间位置t2之间的时间差(即时间偏移)就是该声纹特征1的匹配时间。

[0111] 其中，在执行按照预设条件判断所述匹配结果是否有效的步骤时，可以通过图3所示的以下子步骤S1～S6来实现：

[0112] S1，判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}是否大于或等于第一预设阈值；

[0113] 例如，在短视频的50个声纹特征的50个匹配结果中，确定最高的匹配分数M_{max}，判断M_{max}是否大于或等于β·N；如果否，则确定本次匹配的匹配结果无效(即和该原始视频素材的声纹信息匹配无效，则执行S4，更换素材库中的下一个视频素材的声纹特征进行重新匹配)；如果是，则执行S2；

[0114] 其中，N为输入的短视频的声纹特征个数(这里为50)，而β的典型取值是0.05。

[0115] 另外，需要注意的是，这里的第一预设阈值可以根据N来确定，并不限定为β与N的乘积，也可以是β与N相加等其他运算得到的第一预设阈值。

[0116] S2，统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间ti的第一声纹特征的个数；对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1, t_2, t_3 \cdots t_n$，n为本次匹配的匹配时间总数；在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数；

[0117] S3，判断H_t^1个匹配分数中的最大值(这里用Max_{t1}表示)是否大于或等于第二预设阈值(例如α·M_{max}，α的取值是0.25)；

[0118] 若Max_{t1}小于α·M_{max}，则确定本次匹配的匹配结果无效，则执行S4；若Max_{t1}大于或等于α·M_{max}，则执行S5；

[0119] 另外，需要注意的是，这里的第二预设阈值可以根据M_{max}来确定，并不限定为α与M_{max}的乘积，也可以是α与M_{max}相加等其他运算得到的第二预设阈值。

[0120] S5，将H_t^1与H_t^2的差(即，$H_t^1 - H_t^2$)和第三预设阈值(例如，$\theta \cdot \mathrm{H}_t^1$，θ的优选取值

是0.3)进行比较,以及将t_1与t_2的差(即,t_1-t_2)和第四预设阈值(例如τ)进行比较;

[0121] 若t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差小于第三预设阈值,即,(t_1-t_2)大于τ,且$(H_t^1-H_t^2)$小于$\theta\cdot H_t^1$,则确定本次匹配的匹配结果无效,执行S4,更换素材库中的下一个视频素材的声纹特征进行重新匹配;

[0122] 若t_1与t_2的差小于或等于第四预设阈值,即,若(t_1-t_2)小于或等于τ,则S6,确定本次匹配的匹配结果有效;

[0123] 若t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值,即,$(H_t^1-H_t^2)$大于$\theta\cdot H_t^1$,且(t_1-t_2)大于τ,则S6,确定本次匹配的匹配结果有效。

[0124] 另外,需要注意的是,这里的第三预设阈值可以根据H_t^1来确定,并不限定为θ与H_t^1的乘积,也可以是θ与H_t^1相加等其他运算得到的第三预设阈值。

[0125] 其中,若上述匹配结果无效,则继续获取下一个原始视频的第二声纹信息,并将所述第一声纹信息与所述下一个原始视频的第二声纹信息进行匹配,循环上述步骤,在此不再赘述;

[0126] 其中,在一个实施例中,在执行若所述匹配结果有效,则计算本次匹配的命中率和定位时间时,可以通过以下方式来计算:

[0127] 若(t_1-t_2)小于或等于第四预设阈值(此时不论若$(H_t^1-H_t^2)$和$\theta\cdot H_t^1$的大小关系如何),则将H_t^1确定为本次匹配的命中率,将(t_1+t_2)/2(即,t_1和t_2的平均数)确定为本次匹配的定位时间;

[0128] 若$(H_t^1-H_t^2)$大于第三预设阈值,且(t_1-t_2)大于第四预设阈值,则将H_t^1确定为本次匹配的命中率,将t_1确定为本次匹配的定位时间。

[0129] 其中,在一个实施例中,在执行步骤105时,可以通过以下方式来实现:

[0130] 获取所述目标音频信息在所述待定位视频中对应的起始时间和结束时间;计算所述起始时间和所述目标定位时间之和,得到所述待定位视频在所述目标视频中的目标起始时间;计算所述结束时间和所述目标定位时间之和,得到所述待定位视频在所述目标视频中的目标结束时间。

[0131] 综上,本发明实施例通过使用音频信息进行短视频新闻的定位查询,能够解决视频画面尺寸及画质不同导致的匹配困难,有效恢复短视频新闻的时间信息,并且,与视频画面匹配相比,具有存储数据量小、计算复杂度低的优点;

[0132] 另外,本发明实施例结合音频分类与声纹技术进行匹配查询,避免短视频片头/尾非新闻内容对定位结果的影响,提升定位准确度;

[0133] 此外,本发明实施例使用声纹匹配得分、匹配时间及匹配命中率判断匹配结果的有效性,使得定位精度可达到秒级。

[0134] 借助本发明实施例的上述视频定位方法,当短视频破损或者需要更新信息进行二次加工时,本发明能够有效准确的定位到该短视频的原始视频,以及该短视频在该原始视频中的起始和终止位置,能够实现新闻拆条样例的自动标注。

[0135] 另外，需要注意的是，虽然上述具体实例是以新闻视频的语音来作为目标类型的目标音频信息的，但是本发明的目标类型并不限于语音类型，也可以是音乐类型等其他需要进行定位的音频类型，方法类似，在此不再赘述。

[0136] 需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明实施例并不受所描述的动作顺序的限制，因为依据本发明实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本发明实施例所必须的。

[0137] 与上述本发明实施例所提供的方法相对应，参照图4，示出了本发明一种视频定位装置实施例的结构框图，具体可以包括如下模块：

[0138] 第一获取模块31，用于获取待定位视频的第一声纹信息；

[0139] 第二获取模块32，用于获取原始视频的第二声纹信息；

[0140] 匹配模块33，用于将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间；

[0141] 确定模块34，用于在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频；

[0142] 计算模块35，用于根据所述目标视频与所述原始视频匹配时计算得到的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间。

[0143] 可选地，所述第一获取模块31包括：

[0144] 第一获取子模块，用于获取待定位视频的音频信息；

[0145] 分类子模块，用于对所述音频信息作分类处理，获取所述音频信息中属于目标类型的目标音频信息；

[0146] 提取子模块，用于提取所述目标音频信息的第一声纹信息。

[0147] 可选地，所述匹配模块33包括：

[0148] 匹配子模块，用于将所述第一声纹信息和所述第二声纹信息进行匹配，得到匹配结果；

[0149] 判断子模块，用于按照预设条件判断所述匹配结果是否有效；

[0150] 第一计算子模块，用于若所述匹配结果有效，则计算本次匹配的命中率和定位时间；

[0151] 第二计算子模块，用于若所述匹配结果无效，则将本次匹配的命中率记为零。

[0152] 可选地，所述第一声纹信息包括多个第一声纹特征，所述第二声纹信息包括多个第二声纹特征，所述匹配子模块包括：

[0153] 匹配单元，用于将所述多个第一声纹特征中的每个第一声纹特征分别与所述多个第二声纹特征中的每个第二声纹特征进行匹配，得到各第一声纹特征的匹配结果；

[0154] 其中，所述匹配结果包括：匹配分数以及匹配时间，所述匹配时间为相互匹配的两个声纹特征在各自所属视频上所处的两个时间点之间的时间偏移。

[0155] 可选地，所述判断子模块包括：

[0156] 第一判断单元，用于判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}

是否大于或等于第一预设阈值；

[0157] 统计单元，用于若M_{max}大于或等于第一预设阈值，则统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间t_i的第一声纹特征的个数；

[0158] 排序单元，用于对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1, t_2, t_3 \cdots t_n$，n为本次匹配的匹配时间总数；

[0159] 获取单元，用于在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数；

[0160] 第二判断单元，用于判断H_t^1个匹配分数中的最大值Max_{t1}是否大于或等于第二预设阈值；

[0161] 比较单元，用于若Max_{t1}大于或等于第二预设阈值，则将H_t^1与H_t^2的差与第三预设阈值进行比较，以及将t_1与t_2的差与第四预设阈值进行比较；

[0162] 第一确定单元，用于若t_1与t_2的差小于或等于第四预设阈值，或者t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值时，确定本次匹配的匹配结果有效；

[0163] 第二确定单元，用于若M_{max}小于第一预设阈值、或若Max_{t1}小于第二预设阈值、或若H_t^1与H_t^2的差小于第三预设阈值且t_1与t_2的差大于第四预设阈值，则确定本次匹配的匹配结果无效。

[0164] 可选地，所述第一计算子模块包括：

[0165] 第三确定单元，用于若t_1与t_2的差小于或等于第四预设阈值，则将H_t^1确定为本次匹配的命中率，将$(t_1+t_2)/2$确定为本次匹配的定位时间；

[0166] 第四确定单元，用于若H_t^1与H_t^2的差大于第三预设阈值，且t_1与t_2的差大于第四预设阈值，则将H_t^1确定为本次匹配的命中率，将t_1确定为本次匹配的定位时间。

[0167] 可选地，所述计算模块35包括：

[0168] 第二获取子模块，用于获取所述目标音频信息在所述待定位视频中对应的起始时间和结束时间；

[0169] 第三计算子模块，用于计算所述起始时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标起始时间；

[0170] 第四计算子模块，用于计算所述结束时间和所述目标定位时间之和，得到所述待定位视频在所述目标视频中的目标结束时间。

[0171] 对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0172] 本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

[0173] 本领域内的技术人员应明白，本发明实施例的实施例可提供为方法、装置、或计算

机程序产品。因此，本发明实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0174] 本发明实施例是参照根据本发明实施例的方法、终端设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0175] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0176] 这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上，使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0177] 尽管已描述了本发明实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本发明实施例范围的所有变更和修改。

[0178] 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

[0179] 以上对本发明所提供的一种视频定位方法和一种视频定位装置，进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

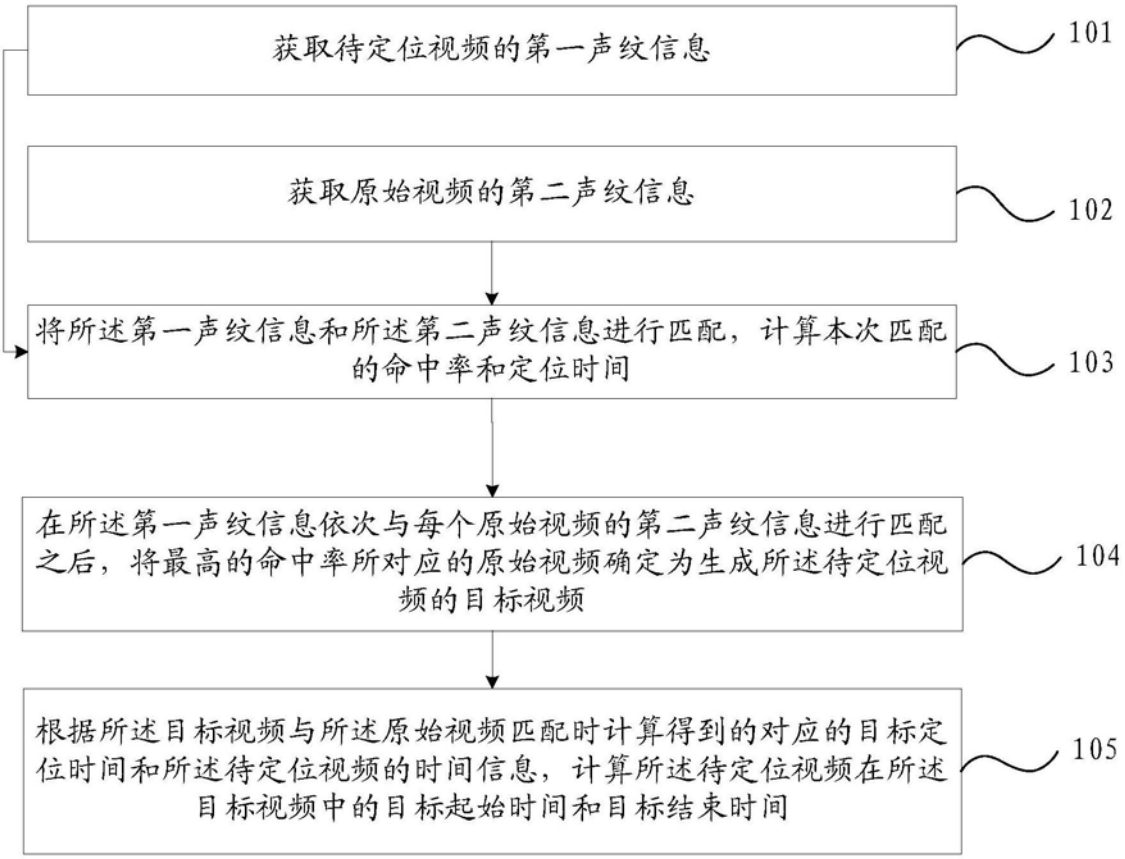

获取待定位视频的第一声纹信息
101
获取原始视频的第二声纹信息
102
将所述第一声纹信息和所述第二声纹信息进行匹配，计算本次匹配的命中率和定位时间
103
在所述第一声纹信息依次与每个原始视频的第二声纹信息进行匹配之后，将最高的命中率所对应的原始视频确定为生成所述待定位视频的目标视频
104
根据所述目标视频与所述原始视频匹配时计算得到的对应的目标定位时间和所述待定位视频的时间信息，计算所述待定位视频在所述目标视频中的目标起始时间和目标结束时间
105

图1

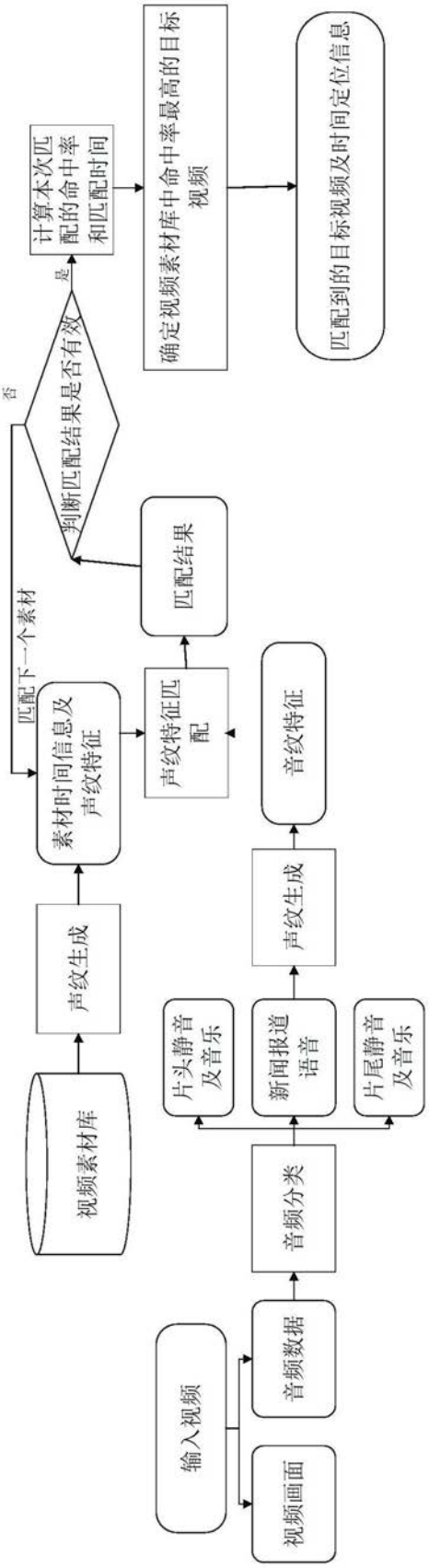

输入视频
视频画面
音频数据
音频分类
片头静音及音乐
新闻报道语音
片尾静音及音乐
声纹生成
音纹特征
视频素材库
声纹生成
素材时间信息及声纹特征
声纹特征匹配
匹配结果
判断匹配结果是否有效
否
匹配下一个素材
是
计算本次匹配的命中率和匹配时间
确定视频素材库中命中率最高的目标视频
匹配到的目标视频及时间定位信息

图2

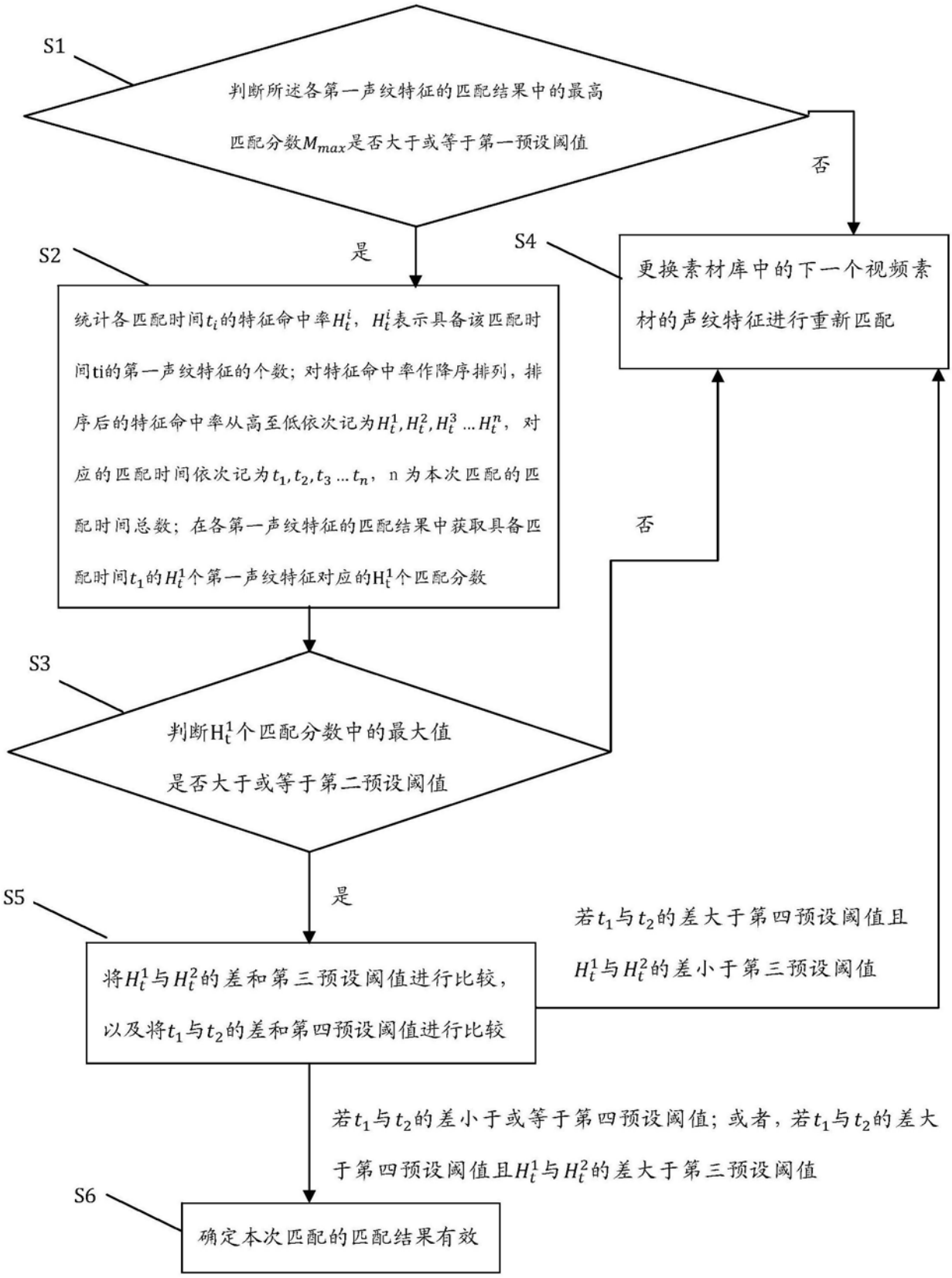

S1
判断所述各第一声纹特征的匹配结果中的最高匹配分数M_{max}是否大于或等于第一预设阈值
否
是
S2
统计各匹配时间t_i的特征命中率H_t^i，H_t^i表示具备该匹配时间ti的第一声纹特征的个数；对特征命中率作降序排列，排序后的特征命中率从高至低依次记为$H_t^1, H_t^2, H_t^3 \ldots H_t^n$，对应的匹配时间依次记为$t_1, t_2, t_3 \ldots t_n$，n为本次匹配的匹配时间总数；在各第一声纹特征的匹配结果中获取具备匹配时间t_1的H_t^1个第一声纹特征对应的H_t^1个匹配分数
S4
更换素材库中的下一个视频素材的声纹特征进行重新匹配
S3
判断H_t^1个匹配分数中的最大值是否大于或等于第二预设阈值
否
是
S5
将H_t^1与H_t^2的差和第三预设阈值进行比较，以及将t_1与t_2的差和第四预设阈值进行比较
若t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差小于第三预设阈值
若t_1与t_2的差小于或等于第四预设阈值；或者，若t_1与t_2的差大于第四预设阈值且H_t^1与H_t^2的差大于第三预设阈值
S6
确定本次匹配的匹配结果有效

图3

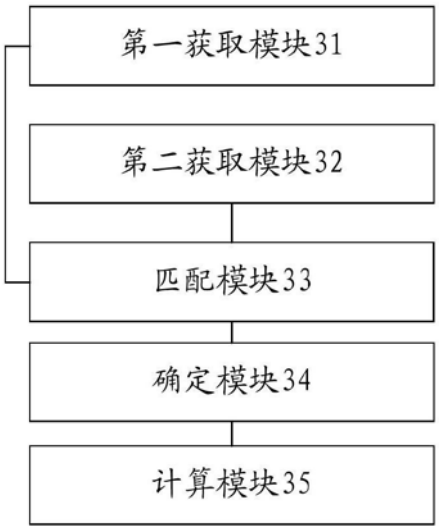


图4