

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日
2017 年 10 月 5 日 (05.10.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/166494 A1

- (51) 国际专利分类号:
H04N 21/234 (2011.01)
- (21) 国际申请号: PCT/CN2016/088980
- (22) 国际申请日: 2016 年 7 月 6 日 (06.07.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610189188.8 2016 年 3 月 29 日 (29.03.2016) CN
- (71) 申请人: 乐视控股(北京)有限公司 (LE HOLDINGS (BEIJING) CO., LTD.) [CN/CN]; 中国北京市朝阳区姚家园路 105 号 3 号楼 10 层 1102, Beijing 100025 (CN)。乐视致新电子科技(天津)有限公司 (LE SHI ZHI XIN ELECTRONIC TECHNOLOGY (TIAN JIN) LIMITED) [CN/CN]; 中国天津市滨海新区生态城动漫中路 126 号动漫大厦 B1 区二层 201-427, Tianjin 300467 (CN)。
- (72) 发明人: 蔡炜 (CAI, Wei); 中国天津市滨海新区生态城动漫中路 126 号动漫大厦 B1 区二层 201-427, Tianjin 300467 (CN)。
- (74) 代理人: 北京同达信恒知识产权代理有限公司 (TDIP & PARTNERS); 中国北京市海淀区知春路 7 号致真大厦 A 座 1304-05 室, Beijing 100191 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA,

[见续页]

(54) Title: METHOD AND DEVICE FOR DETECTING VIOLENT CONTENTS IN VIDEO, AND STORAGE MEDIUM

(54) 发明名称: 一种视频中暴力内容的检测方法、装置及存储介质

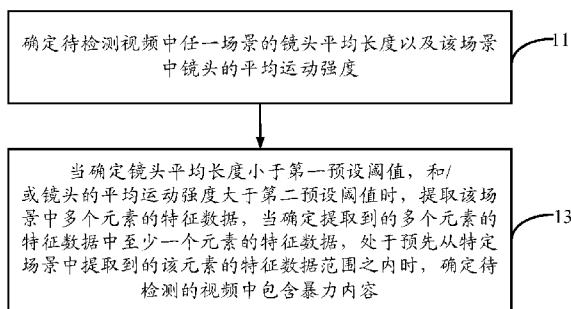


图 1

(57) Abstract: A method and device for detecting violent contents in a video and a storage medium, used to solve the issue in the prior art wherein false positive rate is high when detecting a video for violent contents, thereby improving the accuracy in detecting violent contents in videos. The method for detecting violent contents in a video comprises: determining an average length of a lens in any scene of a video to be detected and an average motion intensity of the lens in the scene; when it is determined that the average length of the lens is less than a first preset threshold and/or the average motion intensity of the lens is greater than a second preset threshold, extracting feature data of a plurality of elements in the scene; and when it is determined that the feature data of at least one element in the extracted feature data of the plurality of elements is within a feature data range of said element as extracted from a specific scene in advance, determining that the video to be detected contains violent contents.

(57) 摘要:

[见续页]

- 11 Determine an average length of a lens in any scene of a video to be detected and an average motion intensity of the lens in the scene
- 13 When it is determined that the average length of the lens is less than a first preset threshold and/or the average motion intensity of the lens is greater than a second preset threshold, extract feature data of a plurality of elements in the scene; and when it is determined that the feature data of at least one element in the extracted feature data of the plurality of elements is within a feature data range of said element as extracted from a specific scene in advance, determine that the video to be detected contains violent contents

WO 2017/166494 A1



RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG,

CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种视频中暴力内容的检测方法、装置及存储介质，用以解决现有技术在对视频中暴力内容进行检测时误判率高的问题，提高对视频中暴力内容检测的准确率。所述视频中暴力内容的检测方法，包括：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

一种视频中暴力内容的检测方法、装置及存储介质

本申请要求在 2016 年 03 月 29 日提交中国专利局、申请号为 201610189188.8、申请名称为“一种视频中暴力内容的检测方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明实施例涉及视频技术领域，尤其涉及一种视频中暴力内容的检测方法、装置及存储介质。

背景技术

暴力内容是一类特殊的激烈内容，在大多数的影视作品中都会出现暴力场面，而且暴力场面往往能够吸引观看者的注意，自动检测出影片中的暴力内容，可用于对影片内容的检索；还可以用于对影片的审查和后期处理。例如：通过检测出的暴力内容的多少来评定影片的级别，对于不适于儿童观看的部分可以进行过滤或覆盖。

发明人在实现本发明的过程中发现，目前，对视频中暴力内容的检测方法大多只利用了某一种信息特征对视频进行分析，难以取得满意的效果。具体来说：

方式一：通过找出视频中重复出现的相似可视内容少的镜头来确定视频的平均运动和持续时间，利用视频的平均运动和持续时间来对视频进行分类，这种方法很难区别暴力场面和有大量运动的体育节目；

方式二：分析视频中的音轨来定位视频中的暴力内容，由于视频中的声音常伴有大量噪声和许多相似的声音而产生较多的误判。

发明人在实现本发明的过程中，发现现有技术在对视频中暴力内容进行检测时，基于视频的平均运动和持续时间的检测方法，或者分析音轨的检测方法，均无法较为准确的检测出视频中的暴力内容，检测的误判率高。

发明内容

本发明实施例提供一种视频中暴力内容的检测方法、装置及存储介质，用以解决现有技术在对视频中暴力内容进行检测时误判率高的问题，提高

对视频中暴力内容检测的准确率。

第一方面，本发明实施例提供一种视频中暴力内容的检测方法，该方法包括：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

第二方面，本发明实施例提供一种视频中暴力内容的检测装置，该装置包括：第一处理单元，用于确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；第二处理单元，用于当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

第三方面，本发明实施例提供一种视频中暴力内容的检测装置，包括存储器以及一个或多个处理器；其中，所述存储器中存储有一个或多个程序；所述一个或多个处理器在执行所述存储器中存储的一个或多个程序时，执行下述操作：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

第四方面，本发明实施例提供一种存储介质，所述存储介质上存储有计算机可执行指令，所述计算机可执行指令响应于本发明实施例提供的视频中暴力内容的检测装置执行操作，所述操作包括：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的

该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

本发明实施例提供一种视频中暴力内容的检测方法、装置及存储介质，首先确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，与现有技术中基于视频运动和持续时间的检测方法，或者分析音轨的检测方法相比，提取场景中多个元素的特征数据，当确定场景中多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1为本发明实施例提供一种视频中暴力内容的检测方法的示意流程图；

图2为本发明实施例提供一种视频中暴力内容的检测方法的具体流程的示意流程图；

图3为本发明实施例提供一种视频中暴力内容的检测装置的结构示意图；

图4为本发明实施例中另一种视频中暴力内容的检测装置的结构示意图；

图5为本发明实施例中一种存储介质示意图。

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本

发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

本发明实施例提供一种视频中暴力内容的检测方法，如图 1 所示，该方法包括：

步骤 11，确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；

步骤 13，当确定镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容。

本发明实施例提供的方法中，首先确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，与现有技术中基于视频运动和持续时间的检测方法，或者分析音轨的检测方法相比，提取场景中多个元素的特征数据，当确定场景中多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

需要说明的是，由于大多数的暴力内容中都有人或物体快速、明显的运动，这样的运动往往是通过短时间的连续的视频镜头的切换来加以表现的，因此，把场景中的镜头平均长度作为衡量一个场景内是否包含暴力内容的一个标准，而镜头中的空间变化和镜头的持续时间决定了镜头中的运动强度，所以把镜头的平均运动强度作为衡量一个场景内是否包含暴力内容的另一个标准，基于这两个标准对视频中的每个场景进行预筛选，也即

首先确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，确定该场景中可能包含暴力内容，将该场景加入候选场景，以进一步进行检测。其中，第一预设阈值和第二预设阈值可以根据经验值进行设定，例如：第一预设阈值的取值为 3，第二预设阈值的取值为视频画面面积的 1/6，当任一场景的镜头平均长度小于 3 秒，和/或场景中镜头的平均运动强度大于视频画面面积的 1/6 时，将该场景作为候选场景。

具体实施时，镜头中的空间变化和镜头的持续时间决定了镜头中的运动强度，为了有效的度量视频中的运动特征，首先抽取镜头中的运动序列。运动序列的抽取过程是：先将视频数据通过二维的小波分解生成一系列空间简化了的视频帧的灰度图像，再将这些图像中各个像素点的灰度在时间上的变化经过小波变换，过滤之后得到一组运动序列图像。采用这种小波分析的方法可以得到视频中运动对象的空间变化，最后生成的运动序列图像在运动对象的边界上有非零值，同时这种方法降低了计算的复杂程度。

接下来我们用下面的公式计算各个镜头的运动强度：

$$SS = \frac{1}{T} \sum_{i=b+1}^e \left\{ \sum_{m,n} |m_l^k(m,n)| \right\},$$

其中， $m_l^k(m,n)$ 是当前场景的运动序列图像在第 k 个镜头中的第 i 帧，m 和 n 是运动序列图像的水平 and 垂直分辨率，b 和 e 分别是第 k 个镜头的起始和结束帧号，T 是第 k 个镜头的长度 $T=e-b$ 。从上述公式中可以看出，持续时间越短、包含运动越多的镜头运动强度越大，计算各个镜头的运动强度之后，镜头的平均运动强度等于场景中所有镜头的运动强度之和与场景中的镜头总数之比。

具体实施时，场景中的镜头平均长度等于场景的总时间长度与该场景中的镜头数量之比。例如：假设一个场景的总时间长度为 300 秒，而该场景中包含 5 个镜头呈现的画面，则镜头平均长度为 60 秒。

具体实施时，根据场景中镜头平均长度和/或镜头的平均运动强度确定候选场景之后，为了提高检测准确率，进一步对候选场景进行检测，提取候选场景中多个元素的特征数据，检测候选场景中每个元素的特征数据是否处于预先从特定场景中提取到的该元素的特征数据范围之内，当确定提

取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容。其中，特定场景可以是一些已知的包含暴力内容的场景，例如：开枪场景、爆炸场景以及流血场景等。多个元素的特征数据，包括：该场景中每帧画面的图像特征数据以及该场景中的音频特征数据。

具体来说，预先从多个特定包含暴力内容的场景中提取多个元素的特征数据，得到每个元素的特征数据范围，当从候选场景中提取到的多个元素的特征数据中任一元素或多个元素的特征数据，处于该元素对应的特征数据范围内时，便可确定该候选场景中包含暴力内容，在通过镜头平均长度和镜头的平均运动强度检测的基础上，结合场景中多个元素的特征数据，当多个元素的特征数据包含每帧画面的图像特征数据以及该场景中的音频特征数据时，可以将可视特征与声音特征进行融合检测，提高了检测的准确率。

当然，本领域技术人员应当理解的是，从候选场景中提取到的多个元素的特征数据中，处于从特定场景中提取到的多个元素的特征数据范围内的元素数量越多，检测的准确率越高，当然，若从候选场景中提取到的多个元素的特征数据中，仅有一个元素的特征数据处于从特定场景中提取到的对应元素的特征数据范围之内，同样可以确定候选场景包含暴力内容。

作为较为具体的实施例，开枪场景和爆炸场景是最明显的包含暴力内容的场景，这些场景在影片中表现出一些独特的声音和图像特征，对于可视特征，也即图像特征，我们主要集中在对由开枪和爆炸引起的瞬时火焰的探测上。

在一种可能的实施方式中，本发明实施例提供的方法中，每帧画面的图像特征数据包括：每帧画面的颜色直方图；当多个元素的特征数据包括该场景中每帧画面的图像特征数据时，确定每帧画面的图像特征数据是否处于预先从特定场景中提取到的画面的图像特征数据范围之内，包括：针对该场景中的每帧画面，提取该帧画面的颜色直方图，当确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

具体实施时，与开枪相比，爆炸引起的火焰持续的时间长，而且在屏幕上覆盖的面积大，但由开枪和爆炸引起的火焰的共同特点是：都有以黄、橙或红色为主色调的颜色直方图，因此，我们预先定义了一个包含各种颜色范围的颜色模板，用候选场景的颜色直方图与预先定义的颜色模板进行比较，当候选场景的颜色直方图中黄色、橙色或红色的统计数量处在预先定义的颜色模板对应颜色的统计数量范围之内时，探测到场景中有火焰出现，候选场景中包含暴力内容。

在包含暴力内容的场景中，一些暴力行为（例如：开枪，刀刺，爆炸等）常常会导致流血事件的发生，在具体实施时，可以用颜色直方图判断场景中是否出现血色。但是，由于现实中有很多颜色与血色很接近，因此，不能仅通过场景的画面中血色像素的数量来判断流血事件的出现，需要结合相邻多帧画面中血色像素的数量做进一步判断，具体来说：

在一种可能的实施方式中，本发明实施例提供的方法中，当确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内之后，该方法还包括：确定该帧画面相邻多帧画面中预设数量个颜色的统计数量；确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内，包括：当确定该帧画面以及相邻多帧画面中预设数量个颜色中每个颜色的统计数量，随着多帧画面的时间顺序逐渐增多时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

具体实施时，在判断场景中是否有流血事件时，需要统计相邻的多帧画面中的血色像素的数量，在短时间内有明显的血色像素增加的情况，才被认为可能是发生了流血事件，也即在连续多帧画面中，血色像素的数量随着多帧画面的时间顺序逐渐增多时，确定场景中发生可能发生了流血事件。

在对视频中暴力内容进行检测时，仅凭可视特征的分析是很难确定场景中是否包含暴力内容的，还必须结合其它的特征分析。声音是视频中十分重要的部分，声音特征可以帮助观看者理解视频内容，特定的声音可以直接、快速的引起观看者的注意。本发明实施例中通过对音频数据的分析来辅助对暴力内容的检测。

在一种可能的实施方式中，本发明实施例提供的方法中，音频特征数据包括：音频数据的样本向量和协方差矩阵；当多个元素的特征数据包括该场景中的音频特征数据时，确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，包括：计算该场景中音频数据的样本向量和协方差矩阵，当确定该场景中音频数据的样本向量和协方差矩阵，与预先从特定场景中提取到的音频数据的样本向量和协方差矩阵的相似度大于第三预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

一般来说，包含暴力内容的场景常常伴随一些非语音的特殊声音（例如：爆炸声、尖叫声、枪声、玻璃的破碎声等）和特殊的背景音乐。通过高斯模型的方法，将视频中的伴随音频分为暴力声音和非暴力声音两种，作为进一步分析的依据，高斯模型提供了简单的计算复杂度，它的参数完全可以由各类样本向量的均值向量和协方差矩阵确定。

具体实施时，从大量视频中找出各种包含暴力内容的场景，将其中的音轨作为声音样本，样本向量由样本在时间上的采样得到，协方差矩阵提供了这种时间变化的紧凑表示，在检测候选场景是否包含暴力内容时，计算候选场景中音频数据的均值向量和协方差矩阵，就可以根据候选场景与声音样本之间均值向量以及协方差矩阵的相似度，确定候选场景中音频数据与声音样本的相似度，当候选场景与声音样本之间均值向量以及协方差矩阵的相似度大于第三预设阈值时，确定候选场景中包含暴力内容。其中，候选场景与声音样本之间均值向量以及协方差矩阵的相似度的计算方式可以采用现有技术，此处不再赘述，第三预设阈值可以根据经验值设定，例如：第三预设阈值的取值为 90。

在一种可能的实施方式中，本发明实施例提供的方法中，音频特征数据包括：音频数据的能量熵；当多个元素的特征数据包括该场景中的音频特征数据时，确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，包括：将该场景中的音频数据分为多段，计算每段音频数据的能量熵，当多段音频数据的能量熵中至少一段音频数据的能量熵小于第四预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

在对音频数据进行分析时，还需要对场景中的一些特殊声音进行分析，

许多包含暴力内容的场景，例如：击打、枪击、爆炸等，都伴有一些特殊的声音，而且这类场景往往在极短的时间内发生，突然爆发出一些声音。因此，在检测时将声音信号能量的突然变化作为检测场景中是否包含暴力内容的又一标准。为有效地度量这一特征，我们采用了“能量熵”规则。

具体来说，首先将候选场景的音频数据分割成若干片段，对每一片段计算其声音信号的能量，并除以音频数据的总能量进行归一化。每段音频数据的能量熵通过如下公式计算得到：

$$I = -\sum_{i=1}^J \sigma_i^2 \log_2 \sigma_i^2,$$

其中， I 为每段音频的能量熵， J 是将场景中的音频数据分为多段的总段数， σ^2 是第 i 段音频数据的归一化的能量值。

根据能量熵的计算过程可以看出，音频数据的能量熵的值可以反映声音信号的能量变化，能量基本恒定的音频数据具有较大的能量熵，而出现声音能量变化的音频数据的能量熵较小，且变化越大能量熵越小。如果场景的音频数据中存在能量熵小于第四预设阈值的音频数据，则确定场景中含有暴力内容。其中，第四预设阈值可以根据经验值进行设定，例如：第四预设阈值的取值为 6。

下面结合图 2 对本发明实施例提供的一种视频中暴力内容的检测方法的具体步骤进行详细说明，如图 2 所示，包括：

步骤 21，确定待检测视频中任一场景的镜头平均长度以及场景中镜头的平均运动强度；

步骤 22，判断镜头平均长度是否小于第一预设阈值，若是，则执行步骤 23，否则，执行步骤 29，其中，第一预设阈值根据经验值设定，例如：第一预设阈值取值为 3；

步骤 23，判断镜头的平均运动强度是否大于第二预设阈值，若是，执行步骤 24，和/或步骤 25，和/或步骤 26，和/或步骤 27，否则，执行步骤 29，其中，第二预设阈值根据经验值设定，例如：第二预设阈值取值为画面面积的 1/6，当然，本领域技术人员应当理解的是，在本发明其它实施例中，步骤 22 和步骤 23 也可以互换顺序；

步骤 24，确定场景中是否有火焰出现，具体来说：利用场景中每帧画面的颜色直方图与预先定义的颜色模板进行比较，判断场景的颜色直方图

中黄色、橙色或红色的统计数量是否处在预先定义的颜色模板对应颜色的统计数量范围之内，若是，执行步骤 28，否则，执行步骤 29；

步骤 25，确定场景中是否出现血色，且血色像素增多，具体来说：利用颜色直方图确定场景中是否出现血色，并统计连续多帧画面中血色像素的数量，判断血色像素的数量是否随多帧画面的时间顺序逐渐增多，若场景中出现血色，且逐渐增多，则执行步骤 28，否则，执行步骤 29；

步骤 26，确定场景中音频数据与声音样本的相似度是否大于第三预设阈值，具体来说，利用场景中音频数据与声音样本之间样本向量和协方差矩阵的相似度，确定场景中音频数据与声音样本的相似度是否大于第三预设阈值，若是，执行步骤 28，否则，执行步骤 29，其中，第三预设阈值根据经验值设定，例如：第三预设阈值取值为 90；

步骤 27，判断场景的音频数据中是否存在能量熵小于第四预设阈值的片段，若是，执行步骤 28，否则，执行步骤 29，其中，第四预设阈值根据经验值设定，例如：第四预设阈值取值为 6；

步骤 28，当步骤 24、步骤 25、步骤 26 以及步骤 27 中至少一个的判定结果为是时，确定当前场景中包含暴力内容，也即待检测视频中包含暴力内容；

步骤 29，当步骤 22 的判定结果为否，或者步骤 23 的判定结果为否，或者步骤 24、步骤 25、步骤 26 以及步骤 27 的判定结果均为否时，确定当前场景中不包含暴力内容，也即待检测视频中不包含暴力内容。

本发明实施例，首先确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，具体来说，提取该场景中每帧图像的颜色直方图、音频数据的样本向量和协方差矩阵以及音频数据的能量熵，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

本发明实施例提供一种视频中暴力内容的检测装置，如图 3 所示，该装置包括：第一处理单元 31，用于确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；第二处理单元 33，用于当确定镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容。

本发明实施例提供的装置中，首先确定待检测视频中任一场景的镜头

平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，与现有技术中基于视频运动和持续时间的检测方法，或者分析音轨的检测方法相比，提取场景中多个元素的特征数据，当确定场景中多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

在一种可能的实施方式中，本发明实施例提供的装置中，多个元素的特征数据，包括：该场景中每帧画面的图像特征数据以及该场景中的音频特征数据。

在一种可能的实施方式中，本发明实施例提供的装置中，在一种可能的实施方式中，本发明实施例提供的装置中，每帧画面的图像特征数据包括：每帧画面的颜色直方图；当多个元素的特征数据包括该场景中每帧画面的图像特征数据时，第二处理单元 33 确定每帧画面的图像特征数据是否处于预先从特定场景中提取到的画面的图像特征数据范围之内，具体用于：针对该场景中的每帧画面，提取该帧画面的颜色直方图，当确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

在一种可能的实施方式中，本发明实施例提供的装置中，当第二处理单元 33 确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内之后，第二处理单元 33 还用于：确定该帧画面相邻多帧画面中预设数量个颜色的统计数量；第二处理单元 33 确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内，具体用于：当确定该帧画面以及相邻多帧画面中预设数量个颜色中每个颜色的统计数量，

随着多帧画面的时间顺序逐渐增多时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

在一种可能的实施方式中，本发明实施例提供的装置中，音频特征数据包括：音频数据的样本向量和协方差矩阵；当多个元素的特征数据包括该场景中的音频特征数据时，第二处理单元 33 确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，具体用于：计算该场景中音频数据的样本向量和协方差矩阵，当确定该场景中音频数据的样本向量和协方差矩阵，与预先从特定场景中提取到的音频数据的样本向量和协方差矩阵的相似度大于第三预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

在一种可能的实施方式中，本发明实施例提供的装置中，音频特征数据包括：音频数据的能量熵；当多个元素的特征数据包括该场景中的音频特征数据时，第二处理单元 33 确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，具体用于：将该场景中的音频数据分为多段，计算每段音频数据的能量熵，当多段音频数据的能量熵中至少一段音频数据的能量熵小于第四预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

在一种可能的实施方式中，本发明实施例提供的装置中，第二处理单元 33 通过如下公式计算每段音频数据的能量熵：

$$I = -\sum_{i=1}^J \sigma_i^2 \log_2 \sigma_i^2 ; \text{ 其中, } I \text{ 为每段音频的能量熵, } J \text{ 是将场景中的}$$

音频数据分为多段的总段数, σ_i^2 是第 i 段音频数据的归一化的能量值。

在一种可能的实施方式中，本发明实施例提供的装置中，镜头的平均运动强度等于场景中所有镜头的运动强度之和与场景中的镜头数量之比，其中，第一处理单元 31 通过如下公式计算场景中每个镜头的运动强度：

$$SS = \frac{1}{T} \sum_{i=b+1}^e \left\{ \sum_{m,n} |m_l^k(m,n)| \right\};$$

其中，SS 是每个镜头的运动强度， $m_l^k(m,n)$ 是当前场景的运动序列图像在第 k 个镜头中的第 i 帧， m 和 n 是运动序列图像的水平 and 垂直分辨率， b 和 e 分别是第 k 个镜头的起始和结束帧号， T 是第 k 个镜头的长度 $T=e-b$ 。

在一种可能的实施方式中，本发明实施例提供的装置中，镜头平均长度等于场景的总时间长度与该场景中的镜头数量之比。

本发明实施例提供的一种视频中暴力内容的检测装置，可以集成在视频软件中，用于对视频中暴力内容的检测，其中，第一处理单元 31 和第二处理单元 33 均可以采用 CPU 处理器等。

如图 4 所示，为本发明实施例提供的另一种视频中暴力内容的检测装置的结构示意图，包括存储器 41 以及一个或多个处理器 42；其中，存储器 41 中存储有一个或多个程序 43；一个或多个处理器 42 在执行存储器 41 中存储的一个或多个程序 43 时，执行下述操作：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；当确定镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容。

本发明实施例中，一个或多个处理器在执行存储器中存储的一个或多个程序时，执行以下操作：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，从而确定待检测的视频中包含暴力内容，与现有技术中基于视频运动和持续时间的检测方法，或者分析音轨的检测方法相比，提取场景中多个元素的特征数据，当确定场景中多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

本发明实施例还提供一种存储介质，存储介质上存储有计算机可执行指令，计算机可执行指令响应于本发明实施例提供的视频中暴力内容的检测装置执行操作，操作包括：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；当确定镜头平均长度小于第一预设阈

值，和/或镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容。

如图 5 所示，51 存储介质上存储有 52 计算机程序产品，52 计算机程序产品可以采用一个或多个可读介质的任意组合，例如，53 信号承载介质，54 可读介质，55 可记录介质，56 通信介质等，53 信号承载介质中存储有至少一个用于确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度的一个或多个指令；以及用于当确定镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容的一个或多个指令。

本发明实施例中，存储介质上存储的计算机可执行指令，响应于本发明实施例提供的视频中暴力内容的检测装置执行操作，操作包括：确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，与现有技术中基于视频运动和持续时间的检测方法，或者分析音轨的检测方法相比，提取场景中多个元素的特征数据，当确定场景中多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

本发明实施例提供的一种视频中暴力内容的检测方法、装置及存储介质，首先确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度，当确定任一场景的镜头平均长度小于第一预设阈值，和/或镜头的平均运动强度大于第二预设阈值时，进一步提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特

征数据，处于预先从特定场景（例如：暴力场景）中提取到的该元素的特征数据范围之内时，确定待检测的视频中包含暴力内容，结合场景中多个元素的特征数据进行检测，提高了对视频中暴力内容检测的准确率。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

权 利 要 求

1、一种视频中暴力内容的检测方法，其特征在于，该方法包括：

确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；

当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

2、根据权利要求1所述的方法，其特征在于，所述多个元素的特征数据，包括：该场景中每帧画面的图像特征数据以及该场景中的音频特征数据。

3、根据权利要求2所述的方法，其特征在于，所述每帧画面的图像特征数据包括：每帧画面的颜色直方图；

当所述多个元素的特征数据包括该场景中每帧画面的图像特征数据时，确定每帧画面的图像特征数据是否处于预先从特定场景中提取到的画面的图像特征数据范围之内，包括：

针对该场景中的每帧画面，提取该帧画面的颜色直方图，当确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

4、根据权利要求3所述的方法，其特征在于，当确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内之后，该方法还包括：

确定该帧画面相邻多帧画面中所述预设数量个颜色的统计数量；

确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内，包括：

当确定该帧画面以及相邻多帧画面中所述预设数量个颜色中每个颜色的统计数量，随着多帧画面的时间顺序逐渐增多时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

5、根据权利要求2所述的方法，其特征在于，所述音频特征数据包括：音频数据的样本向量和协方差矩阵；

当所述多个元素的特征数据包括该场景中的音频特征数据时，确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，包括：

计算该场景中音频数据的样本向量和协方差矩阵，当确定该场景中音频数据的样本向量和协方差矩阵，与预先从特定场景中提取到的音频数据的样本向量和协方差矩阵的相似度大于第三预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

6、根据权利要求2所述的方法，其特征在于，所述音频特征数据包括：音频数据的能量熵；

当所述多个元素的特征数据包括该场景中的音频特征数据时，确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，包括：

将该场景中的音频数据分为多段，计算每段音频数据的能量熵，当多段音频数据的能量熵中至少一段音频数据的能量熵小于第四预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

7、根据权利要求6所述的方法，其特征在于，所述每段音频数据的能量熵通过如下公式计算得到：

$$I = -\sum_{i=1}^J \sigma_i^2 \log_2 \sigma_i^2 ;$$

其中，I为每段音频的能量熵，J是将场景中的音频数据分为多段的总段数， σ^2 是第i段音频数据的归一化的能量值。

8、根据权利要求1-7中任一项所述的方法，其特征在于，所述镜头的平均运动强度等于场景中所有镜头的运动强度之和与场景中的镜头数量之比，其中，场景中每个镜头的运动强度通过如下公式计算得到：

$$SS = \frac{1}{T} \sum_{i=b+1}^e \left\{ \sum_{m,n} |m_l^k(m,n)| \right\};$$

其中，SS是每个镜头的运动强度， $m_l^k(m,n)$ 是当前场景的运动序列图像在第k个镜头中的第i帧，m和n是所述运动序列图像的水平 and 垂

直分辨率，b和e分别是第k个镜头的起始和结束帧号，T是第k个镜头的长度 $T=e-b$ 。

9、根据权利要求1-7中任一项所述的方法，其特征在于，所述镜头平均长度等于场景的总时间长度与该场景中的镜头数量之比。

10、一种视频中暴力内容的检测装置，其特征在于，该装置包括：

第一处理单元，用于确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度；

第二处理单元，用于当确定所述镜头平均长度小于第一预设阈值，和/或所述镜头的平均运动强度大于第二预设阈值时，提取该场景中多个元素的特征数据，当确定提取到的多个元素的特征数据中至少一个元素的特征数据，处于预先从特定场景中提取到的该元素的特征数据范围之内时，确定所述待检测的视频中包含暴力内容。

11、根据权利要求10所述的装置，其特征在于，所述多个元素的特征数据，包括：该场景中每帧画面的图像特征数据以及该场景中的音频特征数据。

12、根据权利要求11所述的装置，其特征在于，所述每帧画面的图像特征数据包括：每帧画面的颜色直方图；

当所述多个元素的特征数据包括该场景中每帧画面的图像特征数据时，所述第二处理单元确定每帧画面的图像特征数据是否处于预先从特定场景中提取到的画面的图像特征数据范围之内，具体用于：

针对该场景中的每帧画面，提取该帧画面的颜色直方图，当确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

13、根据权利要求12所述的装置，其特征在于，当所述第二处理单元确定该帧画面的颜色直方图中预设数量个颜色的统计数量，处于预先从特定场景中提取到的画面的颜色直方图中对应颜色的统计数量范围之内之后，所述第二处理单元还用于：

确定该帧画面相邻多帧画面中所述预设数量个颜色的统计数量；

所述第二处理单元确定该帧画面的图像特征数据处于预先从特定场景

中提取到的画面的图像特征数据范围之内，具体用于：

当确定该帧画面以及相邻多帧画面中所述预设数量个颜色中每个颜色的统计数量，随着多帧画面的时间顺序逐渐增多时，确定该帧画面的图像特征数据处于预先从特定场景中提取到的画面的图像特征数据范围之内。

14、根据权利要求 11 所述的装置，其特征在于，所述音频特征数据包括：音频数据的样本向量和协方差矩阵；

当所述多个元素的特征数据包括该场景中的音频特征数据时，所述第二处理单元确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，具体用于：

计算该场景中音频数据的样本向量和协方差矩阵，当确定该场景中音频数据的样本向量和协方差矩阵，与预先从特定场景中提取到的音频数据的样本向量和协方差矩阵的相似度大于第三预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

15、根据权利要求 11 所述的装置，其特征在于，所述音频特征数据包括：音频数据的能量熵；

当所述多个元素的特征数据包括该场景中的音频特征数据时，所述第二处理单元确定该场景中的音频特征数据是否处于预先从特定场景中提取到的音频特征数据范围之内，具体用于：

将该场景中的音频数据分为多段，计算每段音频数据的能量熵，当多段音频数据的能量熵中至少一段音频数据的能量熵小于第四预设阈值时，确定该场景中的音频特征数据处于预先从特定场景中提取到的音频特征数据范围之内。

16、根据权利要求 15 所述的装置，其特征在于，所述第二处理单元通过如下公式计算每段音频数据的能量熵：

$$I = -\sum_{i=1}^J \sigma_i^2 \log_2 \sigma_i^2 ;$$

其中，I 为每段音频的能量熵，J 是将场景中的音频数据分为多段的总段数， σ_i^2 是第 i 段音频数据的归一化的能量值。

17、根据权利要求 10-16 中任一项所述的装置，其特征在于，所述镜头的平均运动强度等于场景中所有镜头的运动强度之和与场景中的镜头数量之比，其中，所述第一处理单元通过如下公式计算场景中每个镜头的运

动强度:

$$SS = \frac{1}{T} \sum_{i=b+1}^e \left\{ \sum_{m,n} |m_l^k(m,n)| \right\};$$

其中, SS 是每个镜头的运动强度, $m_l^k(m,n)$ 是当前场景的运动序列图像在第 k 个镜头中的第 i 帧, m 和 n 是所述运动序列图像的水平 and 垂直分辨率, b 和 e 分别是第 k 个镜头的起始和结束帧号, T 是第 k 个镜头的长度 $T=e-b$ 。

18、根据权利要求 10-16 中任一项所述的装置, 其特征在于, 所述镜头平均长度等于场景的总时间长度与该场景中的镜头数量之比。

19、一种视频中暴力内容的检测装置, 其特征在于, 包括: 存储器以及一个或多个处理器; 其中,

所述存储器中存储有一个或多个程序;

所述一个或多个处理器在执行所述存储器中存储的一个或多个程序时, 执行下述操作:

确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度;

当确定所述镜头平均长度小于第一预设阈值, 和/或所述镜头的平均运动强度大于第二预设阈值时, 提取该场景中多个元素的特征数据, 当确定提取到的多个元素的特征数据中至少一个元素的特征数据, 处于预先从特定场景中提取到的该元素的特征数据范围之内时, 确定所述待检测的视频中包含暴力内容。

20、一种存储介质, 其特征在于, 所述存储介质上存储有计算机可执行指令, 所述计算机可执行指令响应于如权利要求 10-18 所述的视频中暴力内容的检测装置执行操作, 所述操作包括: 确定待检测视频中任一场景的镜头平均长度以及该场景中镜头的平均运动强度; 当确定所述镜头平均长度小于第一预设阈值, 和/或所述镜头的平均运动强度大于第二预设阈值时, 提取该场景中多个元素的特征数据, 当确定提取到的多个元素的特征数据中至少一个元素的特征数据, 处于预先从特定场景中提取到的该元素的特征数据范围之内时, 确定所述待检测的视频中包含暴力内容。

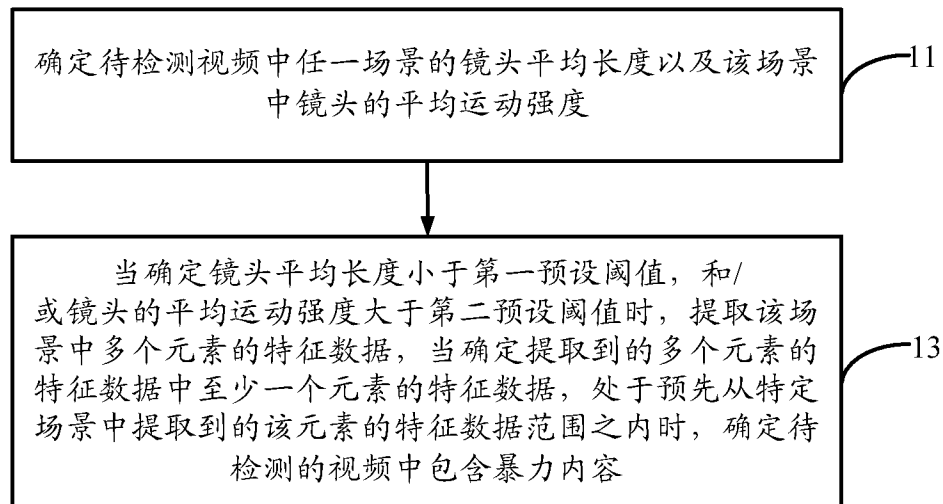


图 1

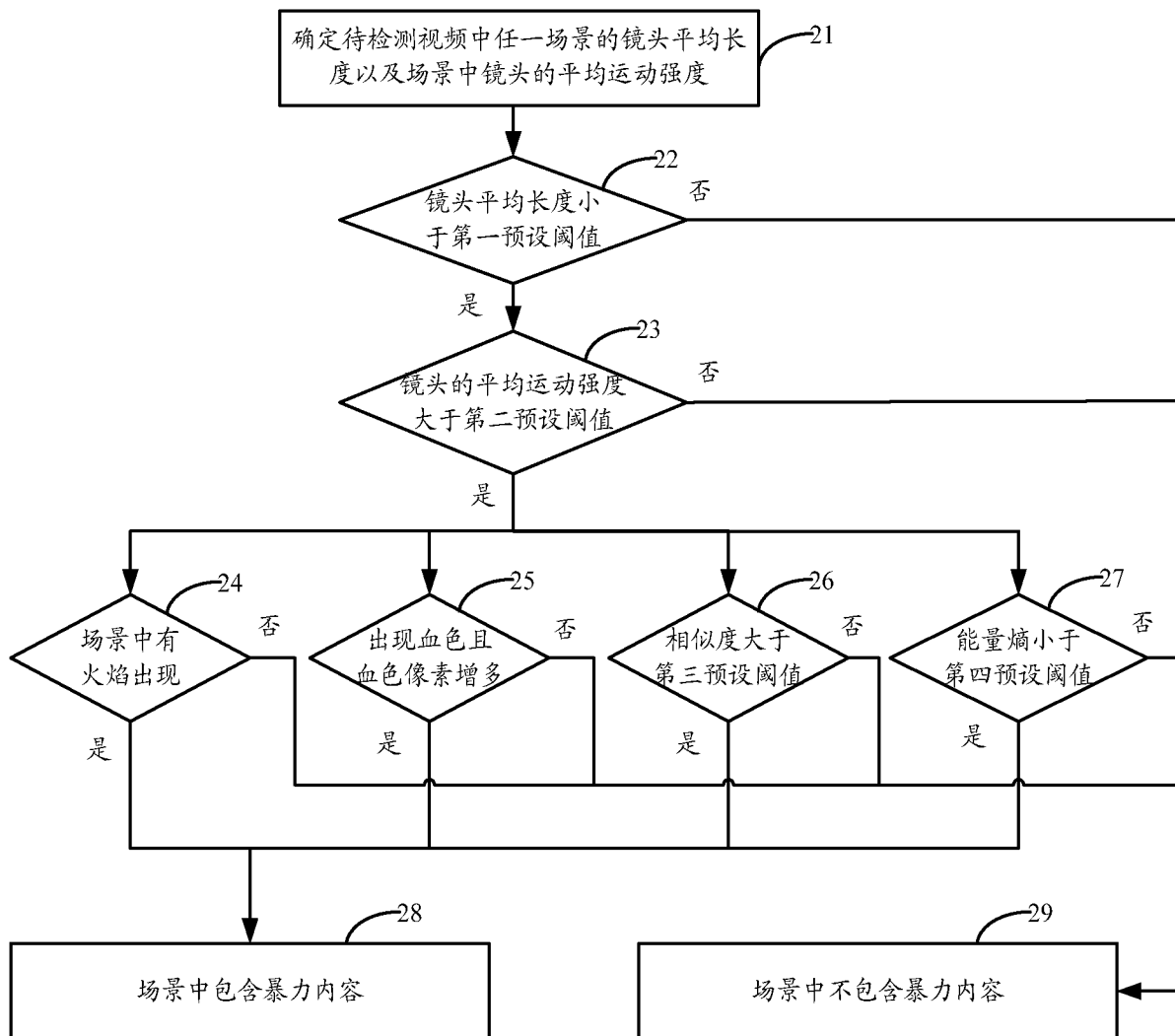


图 2

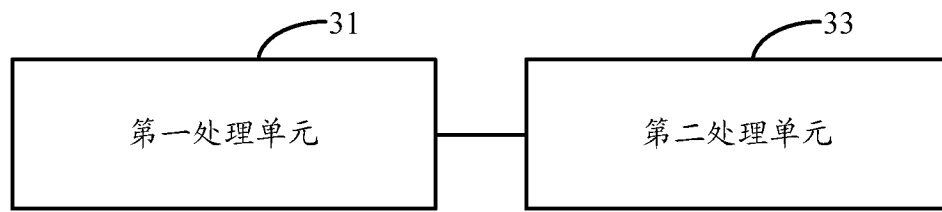


图 3

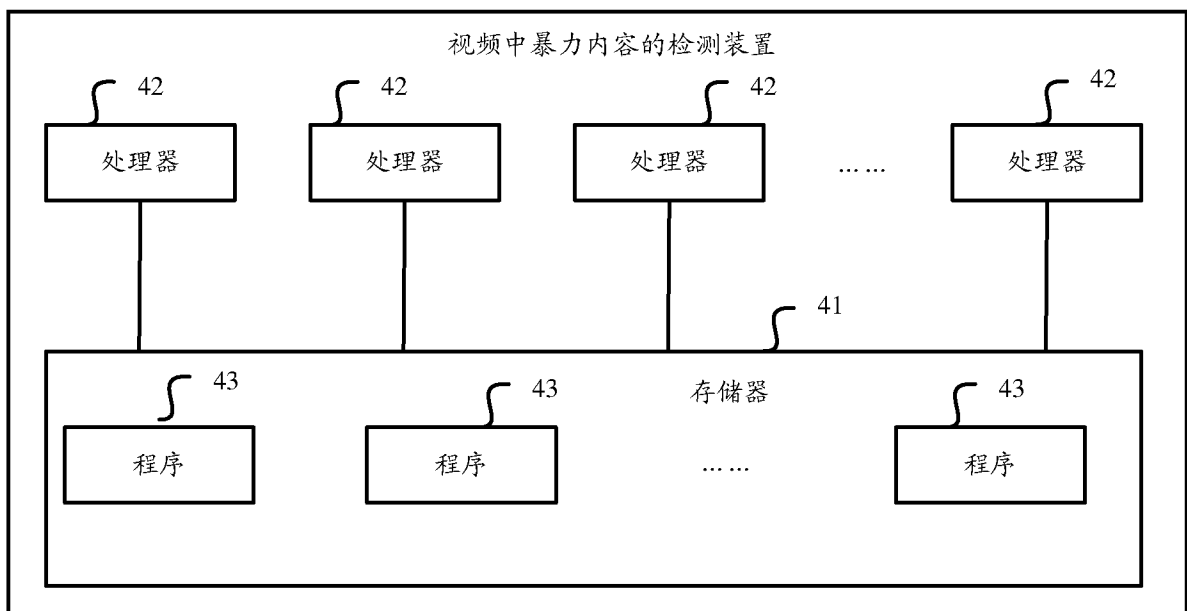


图 4

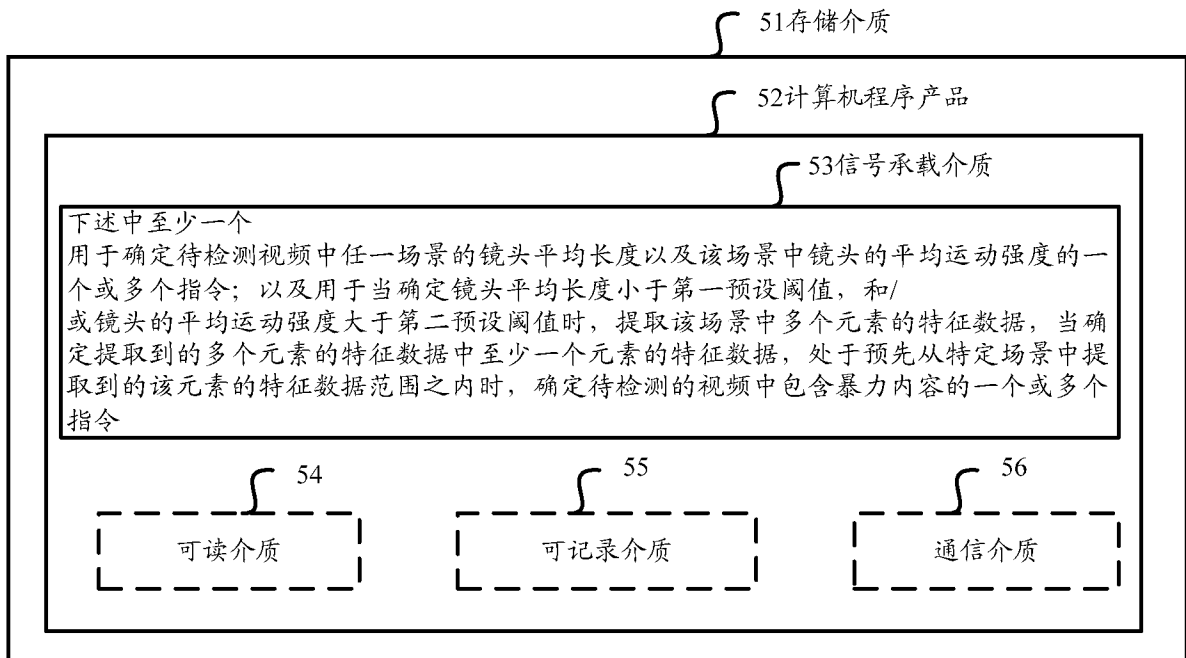


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/088980

A. CLASSIFICATION OF SUBJECT MATTER

H04N 21/234 (2011.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC: violence, force, fight+, bus tup, beat+, blood, gun, shoot, video, film, program, detect+, filt+, lens, scene, number, motion, movement, threshold, average, audio, video, sound, image, frame, picture, character+, color w histogram, vector, covariance w matrix, energy w entropy, identify, setting value, preset value

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103218608 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 24 July 2013 (24.07.2013) description, paragraphs [0015]-[0061] and [0082], and figure 1	1-20
A	CN 101834982 A (SHANGHAI JIAOTONG UNIVERSITY) 15 September 2010 (15.09.2010) the whole document	1-20
A	CN 102930553 A (CHINA MOBILE GROUP SHANGHAI CO., LTD.) 13 February 2013 (13.02.2013) the whole document	1-20
A	CN 102509084 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 20 June 2012 (20.06.2012) the whole document	1-20

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 30 November 2016	Date of mailing of the international search report 26 December 2016
---	--

Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer LI, Jing Telephone No. (86-10) 62413233
---	--

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CN2016/088980

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2005226524 A1 (TAMA-TLO LTD.) 13 October 2005 (13.10.2005) the whole document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/088980

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103218608 A	24 July 2013	None	
CN 101834982 A	15 September 2010	None	
CN 102930553 A	13 February 2013	None	
CN 102509084 A	20 June 2012	None	
US 2005226524 A1	13 October 2005	JP 2005303566 A	27 October 2005

A. 主题的分类 H04N 21/234(2011.01) i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNKI, CNPAT, WPI, EPDOC: 暴力, 敏感, 不良, 打架, 斗殴, 血腥, 枪战, 视频, 节目, 电影, 筛查, 筛选, 识别, 场景, 镜头, 数目, 数量, 运动, 平均, 均值, 阈值, 阙值, 设定值, 预设值, 音频, 声音, 视频, 图像, 特征, 颜色直方图, 向量, 矢量, 协方差矩阵, 能量熵, violence, force, fight+, bustup, beat+, blood, gun, shoot, video, film, program, detect+, filt+, lens, scene, number, motion, movement, threshold, average, audio, video, sound, image, frame, picture, character+, color w histogram, vector, covariance w matrix, energy w entropy		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 103218608 A (中国科学院自动化研究所) 2013年 7月 24日 (2013 - 07 - 24) 说明书第[0015]-[0061]、[0082]段, 图1	1-20
A	CN 101834982 A (上海交通大学) 2010年 9月 15日 (2010 - 09 - 15) 全文	1-20
A	CN 102930553 A (中国移动通信集团上海有限公司) 2013年 2月 13日 (2013 - 02 - 13) 全文	1-20
A	CN 102509084 A (中国科学院自动化研究所) 2012年 6月 20日 (2012 - 06 - 20) 全文	1-20
A	US 2005226524 A1 (TAMA-TLO LTD.) 2005年 10月 13日 (2005 - 10 - 13) 全文	1-20
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期		国际检索报告邮寄日期
2016年 11月 30日		2016年 12月 26日
ISA/CN的名称和邮寄地址		受权官员
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		李晶 电话号码 (86-10)62413233

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/088980

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	103218608	A	2013年 7月 24日	无	
CN	101834982	A	2010年 9月 15日	无	
CN	102930553	A	2013年 2月 13日	无	
CN	102509084	A	2012年 6月 20日	无	
US	2005226524	A1	2005年 10月 13日	JP 2005303566 A	2005年 10月 27日