



(12) 发明专利

(10) 授权公告号 CN 118075552 B

(45) 授权公告日 2024. 06. 28

(21) 申请号 202410479259.2

H04N 21/442 (2011.01)

(22) 申请日 2024.04.22

H04N 21/462 (2011.01)

(65) 同一申请的已公布的文献号

H04N 21/845 (2011.01)

申请公布号 CN 118075552 A

G06T 5/50 (2006.01)

(43) 申请公布日 2024.05.24

(56) 对比文件

(73) 专利权人 黑龙江省邦盾科技有限公司

CN 106165393 A, 2016.11.23

地址 150023 黑龙江省哈尔滨市松北区创

CN 108596865 A, 2018.09.28

新二路733号哈尔滨国际金融大厦10

审查员 石璐

层11号

(72) 发明人 胡伟成

(74) 专利代理机构 哈尔滨市阳光惠远知识产权

代理有限公司 23211

专利代理师 刘景祥

(51) Int. Cl.

H04N 21/44 (2011.01)

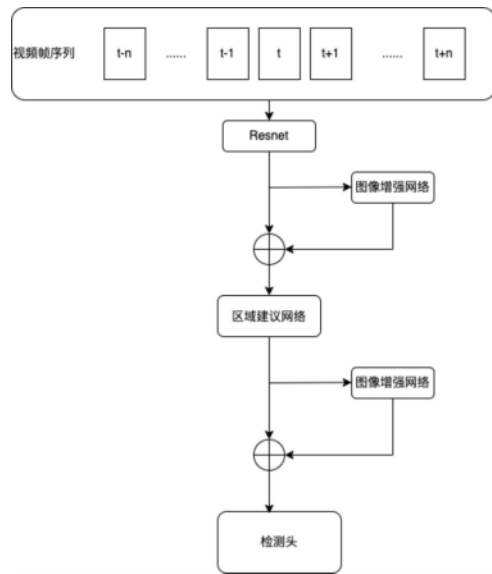
权利要求书4页 说明书6页 附图1页

(54) 发明名称

一种演播室视频特征图像增强处理方法

(57) 摘要

本申请提供了一种演播室视频特征图像增强处理方法,本发明涉及演播室视频图像处理与传输技术领域,该方法应用于演播室环境下采集的视频图像传输与增强。本发明允许多个摄像头采集的视频直接发送给用户,能够识别和跟踪视频中的目标,从而在多个视频流中找到含有相同目标的片段,并将这些片段提供给用户进行切换播放。本发明还包含了一种连续帧融合的图像增强,该方法能够在保证计算精度的同时显著提高计算效率,用以解决视频中目标跟踪和识别过程中可能出现的模糊和遮挡问题。通过选择性地融合关键特征,算法减少了不必要的计算负担,保持了实时性,同时也确保了目标识别和跟踪的准确性。



1. 一种演播室视频特征图像增强处理方法,其特征在于,所述处理方法包括以下步骤:

S1:获取演播室中多个拍摄设备采集得到的视频图像;

S2:客户端接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法;

S3:用户对人像检测结果进行选择,得到目标人像,基于所述目标人像对多视频图像进行所述目标人像追踪;

S4:根据追踪结果进行视频图像切换播放;

所述客户端接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法具体包括:

S21:客户端接收到所述视频图像并在视频图像中选择默认视频进行播放;

所述默认视频可选择主摄像头、副摄像头、特定视角摄像头中任一摄像头所拍摄的视频作为默认视频;

S22:对当前播放的视频进行人像检测,所述人像检测方法首先选定当前检测帧图像;

S23:将ResNet网络作为主干网络对视频帧序列图像进行特征提取,得到视频帧序列浅层特征,并将当前帧浅层特征利用图像增强网络进行特征图像增强;

S24:将其他视频序列特征图和增强后的当前帧特征图输入至区域建议网络中得到带有建议框的当前帧深层特征,并将当前帧深层特征利用图像增强网络进行特征图像增强;

S25:将增强后的当前帧深层特征图输入至目标检测头中,得到相应目标检测结果;

所述特征图增强的算法采用图像增强网络中的图像融合网络实现,图像融合网络的输入为候选特征图 $\{F_1, F_2, \dots, F_k\}$ 和当前帧特征图 $\{C_1, C_2, \dots, C_m\}$, 对于第 j 个当前帧特征图 C_j , $1 \leq j \leq m$, m 表示当前帧特征图总数,进行高维特征映射变换:

$$T_j^h = W_T^h C_j$$

其中, T_j^h 表示对第 j 个当前特征图 C_j 进行第 h 次高维特征映射变换, W_T^h 表示对第 j 个当前特征图 C_j 进行第 h 次高维特征映射变换 T 所对应的高维转换矩阵, $1 \leq h \leq H$, H 表示进行变换次数;

对于第 i 个候选特征图 F_i , $1 \leq i \leq k$, k 表示共有 k 个候选特征图,进行高维特征映射变换和高维内容映射变换:

$$S_i^h = W_S^h F_i$$

$$N_i^h = W_N^h F_i$$

其中, S_i^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维特征映射变换, W_S^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维特征映射变换 S 所对应的高维特征变换矩阵, N_i^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维内容映射变换, W_N^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维内容变换 N 所对应的矩阵;

计算高维特征映射变换后的第 j 个当前帧特征图 C_j 与第 i 个候选特征图 F_i 之间的近似度:

$$S_{i,j}^h = \frac{((T_j^h)^T S_i^h)}{\sqrt{K}}$$

其中, $S_{i,j}^h$ 表示第 j 个当前帧特征图 C_j 与第 i 个候选特征图 F_i 之间在第 h 次变换后的相似度; $()^T$ 表示转置计算, $K = (\text{Column}(F_i)) / H$, $\text{Column}(F_i)$ 表示对第 i 个候选特征图 F_i 向量化后的列数;

计算转换权重图;

$$O_j^h = \frac{\exp(S_{1,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_1^h + \frac{\exp(S_{2,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_2^h + \dots + \frac{\exp(S_{k,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_k^h$$

其中, O_j^h 表示在第 h 次变换下第 j 个转换权重图;

第 j 个特征图 C_j 融合后的新特征图 C_j^{\prime} :

$$C_j^{\prime} = \text{LayerNorm}(C_j + \text{Concat}(O_j^1 + O_j^2 + \dots + O_j^H) W_o)$$

W_o 表示输出转换矩阵, $\text{LayerNorm}()$ 表示归一化处理;

所述基于所述目标人像对多视频图像进行所述目标人像追踪包括:

S31: 在终端展示画面检测结果, 当用户选择其中任一检测结果后, 对该检测结果所在的目标区域进行特征提取得到目标结果向量;

S32: 对终端收到的其他视频图像分割成帧, 按照预设频率进行关键帧提取, 并进行特征提取, 得到待确定目标向量;

S33: 将步骤 S31 得到的目标结果向量与步骤 S32 得到的待确定目标向量进行相似度比对, 相似度超过设定阈值的情况下标记该帧包含目标;

S34: 对包含目标的关键帧进行前后连续帧提取, 检测目标结果, 并标记所有包含目标的视频帧。

2. 根据权利要求 1 所述的演播室视频特征图像增强处理方法, 其特征在于, 所述获取演播室中多个拍摄设备采集得到的视频图像具体包括:

S11: 根据演播室内容和场景设置不同类型和数量的摄像头, 包括主摄像机、副摄像机、固定视角摄像机和移动摄像机;

S12: 设置摄像机的分辨率和帧频以保证图像质量, 并对所有摄像机进行时间同步校准;

S13: 将各个摄像头采集得到视频信号进行传输, 所述传输可通过 SDI 或 HDMI 接口进行有线连接方式传输或通过无线视频传输系统进行信号传输, 演播室视频切换台将收到的多个摄像头传输的视频信号发送至客户端。

3. 根据权利要求 1 所述的演播室视频特征图像增强处理方法, 其特征在于, 所述图像增强网络还包括图像筛选网络, 图像筛选网络接收了预先设定数量的前后帧特征图, 采用筛选策略对前后帧特征图进行选择, 得到 k 个候选特征图 $\{F_1, F_2, \dots, F_k\}$, 所述筛选策略为单一的筛选策略或多策略融合方法。

4. 根据权利要求 3 所述的演播室视频特征图像增强处理方法, 其特征在于, 所述筛选策略包括随机筛选策略、相似筛选策略以及聚类筛选策略。

5. 一种演播室视频特征图像增强处理系统,其特征在于,该系统包括:

图像采集模块:所述图像采集模块获取演播室中多个视频图像;

客户端检测模块:所述客户端检测模块接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法;

人像追踪模块:基于用户对人像检测结果进行选择,得到目标人像,所述人像追踪模块基于所述目标人像对多视频图像进行所述目标人像追踪;

视频切换模块:所述视频切换模块根据追踪结果进行视频图像切换播放;

所述客户端接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法具体包括:

S21:客户端接收到所述视频图像并在视频图像中选择默认视频进行播放;

所述默认视频可选择主摄像头、副摄像头、特定视角摄像头中任一摄像头所拍摄的视频作为默认视频;

S22:对当前播放的视频进行人像检测,所述人像检测方法首先选定当前检测帧图像;

S23:将ResNet网络作为主干网络对视频帧序列图像进行特征提取,得到视频帧序列浅层特征,并将当前帧浅层特征利用图像增强网络进行特征图像增强;

S24:将其他视频序列特征图和增强后的当前帧特征图输入至区域建议网络中得到带有建议框的当前帧深层特征,并将当前帧深层特征利用图像增强网络进行特征图像增强;

S25:将增强后的当前帧深层特征图输入至目标检测头中,得到相应目标检测结果;

所述特征图增强的算法采用图像增强网络中的图像融合网络实现,图像融合网络的输入为候选特征图 $\{F_1, F_2, \dots, F_k\}$ 和当前帧特征图 $\{C_1, C_2, \dots, C_m\}$, 对于第 j 个当前帧特征图 C_j , $1 \leq j \leq m$, m 表示当前帧特征图总数,进行高维特征映射变换:

$$T_j^h = W_T^h C_j$$

其中, T_j^h 表示对第 j 个当前特征图 C_j 进行第 h 次高维特征映射变换, W_T^h 表示对第 j 个当前特征图 C_j 进行第 h 次高维特征映射变换 T 所对应的高维转换矩阵, $1 \leq h \leq H$, H 表示进行变换次数;

对于第 i 个候选特征图 F_i , $1 \leq i \leq k$, k 表示共有 k 个候选特征图,进行高维特征映射变换和高维内容映射变换:

$$S_i^h = W_S^h F_i$$

$$N_i^h = W_N^h F_i$$

其中, S_i^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维特征映射变换, W_S^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维特征映射变换 S 所对应的高维特征变换矩阵, N_i^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维内容映射变换, W_N^h 表示对第 i 个候选特征图 F_i 进行第 h 次高维内容变换 N 所对应的矩阵;

计算高维特征映射变换后的第 j 个当前帧特征图 C_j 与第 i 个候选特征图 F_i 之间的近似度:

$$S_{i,j}^h = \frac{((T_j^h)^T S_i^h)}{\sqrt{K}}$$

其中, $S_{i,j}^h$ 表示第 j 个当前帧特征图 C_j 与第 i 个候选特征图 F_i 之间在第 h 次变换后的相似度; $()^T$ 表示转置计算, $K = (\text{Column}(F_i)) / H$, $\text{Column}(F_i)$ 表示对第 i 个候选特征图 F_i 向量化后的列数;

计算转换权重图;

$$O_j^h = \frac{\exp(S_{1,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_1^h + \frac{\exp(S_{2,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_2^h + \dots + \frac{\exp(S_{k,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_k^h$$

其中, O_j^h 表示在第 h 次变换下第 j 个转换权重图;

第 j 个特征图 C_j 融合后的新特征图 C_j' :

$$C_j' = \text{LayerNorm}(C_j + \text{Concat}(O_j^1 + O_j^2 + \dots + O_j^H) W_o)$$

W_o 表示输出转换矩阵, $\text{LayerNorm}()$ 表示归一化处理;

所述基于所述目标人像对多视频图像进行所述目标人像追踪包括:

S31: 在终端展示画面检测结果, 当用户选择其中任一检测结果后, 对该检测结果所在的目标区域进行特征提取得到目标结果向量;

S32: 对终端收到的其他视频图像分割成帧, 按照预设频率进行关键帧提取, 并进行特征提取, 得到待确定目标向量;

S33: 将步骤 S31 得到的目标结果向量与步骤 S32 得到的待确定目标向量进行相似度比对, 相似度超过设定阈值的情况下标记该帧包含目标;

S34: 对包含目标的关键帧进行前后连续帧提取, 检测目标结果, 并标记所有包含目标的视频帧。

6. 一种电子设备, 包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序, 所述处理器执行所述计算机程序时实现如权利要求 1-4 任一项所述的一种演播室视频特征图像增强处理方法。

7. 一种计算机可读存储介质, 所述计算机可读存储介质存储有计算机程序, 所述计算机程序被处理器执行时实现如权利要求 1-4 任一项所述的一种演播室视频特征图像增强处理方法。

一种演播室视频特征图像增强处理方法

技术领域

[0001] 本发明涉及演播室视频图像处理与传输领域,尤其是关于一种演播室视频特征图像增强处理方法,该方法应用于演播室环境下采集的视频图像传输与增强。

背景技术

[0002] 现有技术中在赛事或演出的视频制作和传输主要依赖导播控制模式,这一模式在实际应用中显现了诸多不足。首先,观众的观看体验受限于导播的选择,他们无法按照个人偏好观看特定角度或焦点的内容,导致个性化需求难以得到满足。其次,导播在选择摄像机画面时需要快速做出决策,这不仅限制了内容的多样性,也可能导致错过实时的精彩瞬间。此外,传统的观看模式缺乏互动性,观众仅能被动接受内容,与赛事或演出的互动层面相对薄弱。

[0003] 观众参与度的低下进一步凸显了内容个性化的不足。尽管可能有多个摄像头同时捕捉赛事或演出,但传统模式下一次只能展示一个摄像头的的内容,造成了大量摄像资源的闲置和浪费。从技术层面来看,导播系统的操作复杂性要求专业人员控制,这不仅增加了人力成本,同时也增加了技术成本。而且,在现有的模式中,视频播放缺乏与社交网络的集成,限制了观众之间的互动及社交体验。

[0004] 现有技术在提供个性化观看体验、内容多样性、观众互动、资源利用效率以及成本控制等方面存在显著的局限性。随着用户需求的不断发展和技术的进步,这些局限性日益成为制约赛事和演出视频内容制作与分发创新的瓶颈。

发明内容

[0005] 有鉴于此,本发明提供一种演播室视频特征图像增强处理方法,该方法可包括以下步骤:

[0006] S1:获取演播室中多个拍摄设备采集得到的视频图像;

[0007] S2:客户端接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法;

[0008] S3:用户对人像检测结果进行选择,得到目标人像,基于所述目标人像对多视频图像进行所述目标人像追踪;

[0009] S4:根据追踪结果进行视频图像切换播放。

[0010] 本发明还提供一种演播室视频特征图像增强处理系统,该系统包括:

[0011] 图像采集模块:所述图像采集模块获取演播室中多个视频图像;

[0012] 客户端检测模块:所述客户端检测模块接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法;

[0013] 人像追踪模块:基于用户对人像检测结果进行选择,得到目标人像,所述人像追踪模块基于所述目标人像对多视频图像进行所述目标人像追踪;

[0014] 视频切换模块:所述视频切换模块根据追踪结果进行视频图像切换播放。

[0015] 本发明还提供一种电子设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现上述一种演播室视频特征图像增强处理方法。

[0016] 本发明还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现上述一种演播室视频特征图像增强处理方法。

[0017] 与现有技术相比,本发明公开了一种演播室视频特征图像增强处理方法,与传统导播剪辑和切换不同,本发明允许多个摄像头采集的视频直接发送给用户,不进行中间的编辑或切换处理。本发明采用了能够识别和跟踪视频中的目标,从而在多个视频流中找到含有相同目标的片段,并将这些片段提供给用户进行实时切换播放。观众可以根据自身的偏好直接选择观看赛事或演出中特定目标的视频片段。

[0018] 此外,针对视频流的处理,本发明还包含了一种连续帧融合的筛选机制。这一机制能够在保证计算精度的同时显著提高计算效率,用以解决视频中目标跟踪和识别过程中可能出现的模糊和遮挡问题。通过选择性地融合关键帧,算法减少了不必要的计算负担,保持了实时性,同时也确保了目标识别和跟踪的准确性。

附图说明

[0019] 为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,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它的附图。

[0020] 图1是本申请所采用的网络结构图。

具体实施方式

[0021] 下面结合附图对本申请实施例进行详细描述。

[0022] 以下通过特定的具体实例说明本申请的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本申请的其他优点与功效。显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。本申请还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本申请的精神下进行各种修饰或改变。需说明的是,在不冲突的情况下,以下实施例及实施例中的特征可以相互组合。基于本申请中的实施例,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,都属于本申请保护的范围。

[0023] 要说明的是,下文描述在所附权利要求书的范围内的实施例的各种方面。应显而易见,本文中所描述的方面可体现于广泛多种形式中,且本文中所描述的任何特定结构及/或功能仅为说明性的。基于本申请,所属领域的技术人员应了解,本文中所描述的一个方面可与任何其它方面独立地实施,且可以各种方式组合这些方面中的两者或两者以上。举例来说,可使用本文中所阐述的任何数目和方面来实施设备及/或实践方法。另外,可使用除了本文中所阐述的方面中的一或多者之外的其它结构及/或功能性实施此设备及/或实践此方法。

[0024] 另外,在以下描述中,提供具体细节是为了便于透彻理解实例。然而,所属领域的技术人员将理解,可在没有这些特定细节的情况下实践。

- [0025] 以下结合附图,说明本申请各实施例提供的技术方案。
- [0026] 本说明书实施例提供演播室视频特征图像增强处理方法,该方法可包括以下步骤:
- [0027] S1:获取演播室中多个拍摄设备采集得到的视频图像具体包括:
- [0028] S11:根据演播室内容和场景设置不同类型的摄像头的数量,包括主摄像机、副摄像机、固定视角摄像机和移动摄像机;
- [0029] S12:设置摄像机的分辨率和帧频以保证图像质量,并对所有摄像机进行时间同步校准;
- [0030] S13:将各个摄像头采集得到视频信号进行传输,所述传输可通过SDI或HDMI接口进行有线连接方式稳定传输或通过无线视频传输系统进行信号传输,演播室视频切换台将收到的多个摄像头传输的视频信号发送至客户端。
- [0031] 本发明采用将演播室中摄像头的采集的所有视频图像均发送给用户的模式,这种模式下,用户可以根据自己的喜好进行个性化的镜头切换,例如,在一场体育赛事中,一位对技术分析感兴趣的观众可能会选择更多地关注全场视角,而热衷于紧跟球星动态的观众则可能更频繁地切换到跟随特定球员的摄像头。
- [0032] S2:客户端接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法具体包括:
- [0033] S21:客户端接收到所述视频图像并在视频图像中选择默认视频进行播放;
- [0034] 所述默认视频可选择主摄像头、副摄像头、特定视角摄像头中任一摄像头所拍摄的视频作为默认视频;
- [0035] S22:对当前播放的视频进行人像检测,所述人像检测方法首先选定当前检测帧图像;
- [0036] S23:将ResNet网络作为主干网络对视频帧序列图像进行特征提取,得到视频帧序列浅层特征,并将当前帧浅层特征利用图像增强网络进行特征图像增强;
- [0037] S24:将其他视频序列特征图和增强后的当前帧特征图输入至区域建议网络中得到带有建议框的当深层特征,并将当前帧深层特征利用图像增强网络进行特征图像增强;
- [0038] S25:将增强后的当前帧深层特征图输入至目标检测头中,得到相应目标检测结果。
- [0039] 本发明对当前播放的视频进行目标检测,这种方法允许观众根据个人偏好选择他们感兴趣的镜头或人物,比如在大型演出演播室或体育赛事,观众可以直接选择他们最感兴趣的人物。
- [0040] 本发明对当前播放的视频进行目标检测,还可以增强视频内容的可搜索性。比如在一个长时间的体育赛事中,观众可能只对特定的瞬间或人物感兴趣。通过目标识别和视频切换功能,本发明可以快速定位到这些时刻,而不是手动跳过大量不相关的内容。此外,本发明中基于用户交互而实现的目标识别、自动视频切换技术还提高了节目制作的效率,节省了制作人员的时间和劳力。
- [0041] 所述图像增强网络包括图像筛选网络和图像融合网络;
- [0042] 图像筛选网络接收了预先设定数量的前后帧特征图,采用筛选策略对前后帧特征图进行选择,得到k个候选特征图 $\{F_1, F_2, \dots, F_k\}$,所述筛选策略为单一的筛选策略或多

策略融合方法；

[0043] 所述筛选策略包括随机筛选策略、相似筛选策略以及聚类筛选策略；

[0044] 所述随机筛选策略对每个前后帧特征图进行信息统计,计算特征图的平均激活强度,选取前K个前后帧特征图作为候选特征图；

[0045] 所述相似筛选策略对所有前后帧特征图和当前帧特征图进行相似度计算,选取前K个前后帧特征图作为候选帧特征图；

[0046] 所述聚类筛选策略为对所有前后帧特征图进行频谱图像转换,并对频谱图像进行聚类,选取最大类别中K个特征图作为候选帧特征图。

[0047] 图像融合网络的输入为候选特征图 $\{F_1, F_2, \dots, F_k\}$ 和当前帧特征图 $\{C_1, C_2, \dots, C_m\}$, 对于第j个当前帧特征图 C_j , $1 \leq j \leq m$, m表示共有m个当前帧特征图,进行高维特征映射变换：

$$[0048] \quad T_j^h = W_T^h C_j$$

[0049] 其中, T_j^h 表示对第j个当前特征图 C_j 进行第h次高维特征映射变换, W_T^h 表示对第j个当前特征图 C_j 进行第h次高维特征映射变换所对应的高维转换矩阵, $1 \leq h \leq H$, H表示进行变换次数；

[0050] 对于第i个候选特征图 F_i , $1 \leq i \leq k$, k表示共有k个候选特征图,进行高维特征映射变换和高维内容映射变换：

$$[0051] \quad S_i^h = W_S^h F_i$$

$$[0052] \quad N_i^h = W_N^h F_i$$

[0053] 其中, S_i^h 表示对第i个候选特征图 F_i 进行第h次高维特征映射变换, W_S^h 表示对第i个候选特征图 F_i 进行第h次高维特征映射变换所对应的高维特征变换矩阵, N_i^h 表示对第i个候选特征图 F_i 进行第h次高维内容映射变换, W_N^h 表示对第i个候选特征图 F_i 进行第h次高维内容变换矩阵；

[0054] 计算高维特征映射变换后的第j个当前帧特征图 C_j 与第i个候选特征图 F_i 之间的近似度：

$$[0055] \quad S_{i,j}^h = \frac{((T_j^h)^T S_i^h)}{\sqrt{K}}$$

[0056] 其中, $S_{i,j}^h$ 表示第j个当前帧特征图 C_j 与第i个候选特征图 F_i 之间在第h次变换后的相似度；T表示转置计算, $K = (\text{Column}(F_i)) / H$, $\text{Column}(F_i)$ 表示对第i个候选特征图 F_i 向量化后的列数；

[0057] 计算转换权重图；

$$[0058] \quad O_j^h = \frac{\exp(S_{1,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_1^h + \frac{\exp(S_{2,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_2^h + \dots + \frac{\exp(S_{k,j}^h)}{\sum_{n=1}^k \exp(S_{n,j}^h)} N_k^h$$

[0059] 其中, O_j^h 表示在第 h 次变换下第 j 个转换权重图;

[0060] 第 j 个特征图 C_j 融合后的新特征图 C_j' :

[0061] $C_j' = \text{LayerNorm}(C_j + \text{Concat}(O_j^1 + O_j^2 + \dots + O_j^H))W_o$

[0062] W_o 表示输出转换矩阵, $\text{LayerNorm}()$ 表示归一化处理。

[0063] 静态图像目标识别主要依赖于单一图像中的空间信息, 通过对图像特征的学习和识别, 算法可以有效地定位和识别图像中的目标。这种识别过程相对简单, 因为它不需要考虑时间维度上的变化, 而是专注于提取静态图像的细节特征。这种方法在处理高质量、清晰度高图像时尤为有效。

[0064] 然而, 在视频目标识别的场景下, 面临的挑战显著增加。视频不仅包含了静态图像的空间信息, 还包括了时间维度上的动态信息。视频中的目标可能会由于运动模糊、快速移动、遮挡、光照变化以及画面质量等问题而难以识别。为了提高识别精度, 需要利用视频中的时间序列信息。这意味着将连续的多帧图像进行融合, 以便利用前后帧中的信息来弥补某一帧中目标的不清晰或不完整。通过这种方式, 算法可以对目标的运动轨迹和形态变化有更好的理解, 从而提高检测的准确性。

[0065] 传统的视频帧目标检测技术依赖于融合连续帧的全部特征图来提升对当前帧的时空信息理解, 这种方法虽然能够提高检测的准确性, 但却因为其算法设计中必须处理大量的数据而导致计算负担重, 时间和空间复杂度高, 对计算资源的要求较高。此外, 该方法处理了大量可能并不会对最终识别结果产生显著影响的冗余信息, 造成了效率低下。

[0066] 相对而言, 本发明在目标检测中采取了更为灵活和高效的策略。通过筛选前后多帧中的关键特征图而非全部特征图, 本发明集中处理那些对于增强当前帧的时空信息最为重要的特征即可。这样的选择性特征融合能够显著减少不必要的计算量, 降低了时间和空间复杂度, 从而提高了检测的效率。通过对筛选出的重要特征图进行加权融合, 本发明不仅保证了重要信息的有效利用, 也在保持了相对较高的检测精度的同时提高了处理速度。

[0067] 本发明在具体筛选规则时, 设定了预设数量 K , 其设定的目的在于其能够在精度和效率之间找到一个更优的平衡点。它摒弃了传统技术中对所有特征同等处理的做法, 转而采用更为精细化的特征筛选机制, 这种方法的实质是一种资源的优化配置, 它将计算资源集中于那些对提升检测性能最为关键的部分。因此, 相较于传统的特征融合技术, 本发明能够以更低的资源消耗实现相似甚至更好的检测结果, 特别是在处理大规模视频数据时, 它的优势更为明显。

[0068] S3: 用户对人像检测结果进行选择, 得到目标人像, 基于所述目标人像对多视频图像进行所述目标人像追踪具体包括:

[0069] S31: 在终端展示画面检测结果, 当用户选择其中任一检测结果后, 对该检测结果所在的目标区域进行特征提取得到目标结果向量;

[0070] S32: 对终端收到的其他视频图像分割成帧, 并按照预设频率进行关键帧提取, 并进行特征提取, 得到待确定目标向量;

[0071] S33: 将步骤 S31 得到的目标结果向量与步骤 S32 得到的待确定目标向量进行相似度比对, 相似度超过设定阈值的情况下标记该帧包含目标;

[0072] S34: 对包含目标的关键帧进行前后连续帧提取, 检测目标结果, 并标记所有包含

目标的视频帧。

[0073] S4:根据追踪结果进行视频图像切换播放包括,提取包含目标的视频帧得到视频片段,利用图形用户界面向用户展示其他视频中包含目标的视频片段,在用户确认后,进行图像切换播放。

[0074] 本发明允许用户根据个性化选择从多个摄像头拍摄的视频片段中切换播放,使观众可以根据个人兴趣和偏好深度参与到赛事或演出之中。例如体育和艺术表演训练中,学习者可以通过切换目标对象的不同视频,观察从不同角度拍摄的表演,从而、获得更全面的理解和学习。在医疗、紧急响应培训或其他技能提升的培训节目中,本发明提供更为真实和多角度的情境模拟,帮助专业人员更好地准备实际情况。

[0075] 本发明还提供一种演播室视频特征图像增强处理系统,该系统包括:

[0076] 图像采集模块:所述图像采集模块获取演播室中多个视频图像;

[0077] 客户端检测模块:所述客户端检测模块接收所述视频图像,并对播放视频进行人像检测,所述人像检测采用特征图增强的算法;

[0078] 人像追踪模块:基于用户对人像检测结果进行选择,得到目标人像,所述人像追踪模块基于所述目标人像对多视频图像进行所述目标人像追踪;

[0079] 视频切换模块:所述视频切换模块根据追踪结果进行视频图像切换播放。

[0080] 本发明还提供一种电子设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现上述一种演播室视频特征图像增强处理方法。

[0081] 本发明还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现上述一种演播室视频特征图像增强处理方法。

[0082] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0083] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范围为准。

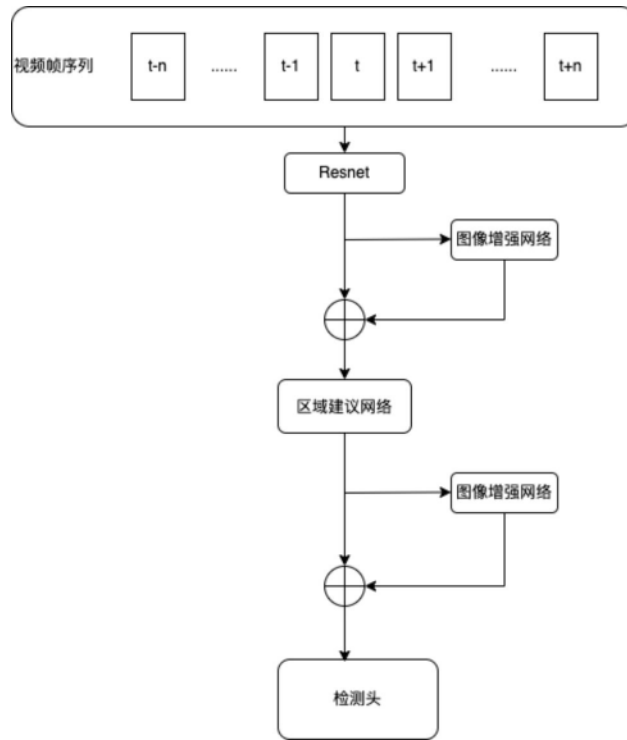


图1