



(12)发明专利申请

(10)申请公布号 CN 109902547 A

(43)申请公布日 2019.06.18

(21)申请号 201810533284.9

(22)申请日 2018.05.29

(71)申请人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 乔宇 周磊 王亚立 江立辉 刘健庄

(74)专利代理机构 北京龙双利达知识产权代理有限公司 11329

代理人 张振 张欣

(51)Int.Cl.

G06K 9/00(2006.01)

G06K 9/46(2006.01)

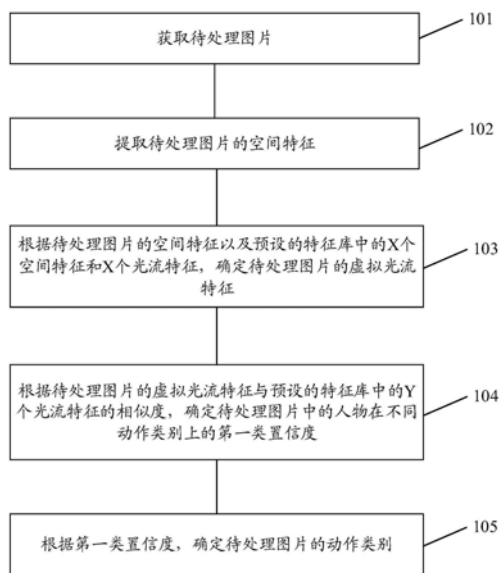
权利要求书3页 说明书19页 附图8页

(54)发明名称

动作识别方法和装置

(57)摘要

本申请提供了动作识别方法和装置。该方法包括：获取待处理图片；提取待处理图片的空间特征；根据待处理图片的空间特征以及预设的特征库中的X个空间特征和X个光流特征，确定待处理图片的虚拟光流特征，其中，预设特征库中的X个空间特征和X个光流特征存在一一对应关系，X为大于1的整数；根据待处理图片的虚拟光流特征与预设特征库中的Y个光流特征的相似度，确定待处理图片在不同动作类别上的第一类置信度，其中，预设特征库中的Y个光流特征每个光流特征对应一种动作类别，Y为大于1的整数；根据第一类置信度确定待处理图片的动作类别。本申请能够提高动作识别的准确率。



1. 一种动作识别方法,其特征在于,包括:

获取待处理图片;

提取所述待处理图片的空间特征;

根据所述待处理图片的空间特征以及预设的特征库中的X个空间特征和X个光流特征,确定所述待处理图片的虚拟光流特征,其中,所述X个空间特征和所述X个光流特征存在一一对应关系,X为大于1的整数;

根据所述待处理图片的虚拟光流特征与所述特征库中的Y个光流特征的相似度,确定所述待处理图片在不同动作类别上的第一类置信度,其中,所述Y个光流特征中的每个光流特征对应一种动作类别,Y为大于1的整数;

根据所述第一类置信度确定所述待处理图片的动作类别。

2. 如权利要求1所述的方法,其特征在于,所述根据所述待处理图片的空间特征以及预设的特征库中的X个空间特征和X个光流特征,确定所述待处理图片的虚拟光流特征,包括:

根据所述待处理图片的空间特征与所述X个空间特征中的每个空间特征的相似度,确定所述X个光流特征中的每个光流特征的权重系数;

根据X个光流特征中的每个光流特征的权重系数,对所述X个光流特征进行加权求和,得到所述待处理图片的虚拟光流特征。

3. 如权利要求2所述的方法,其特征在于,所述X个光流特征中的每个光流特征的权重系数的大小与第一相似度正相关,所述第一相似度为所述X个空间特征中与所述X个光流特征中的每个光流特征相对应的空间特征与所述待处理图片的空间特征的相似度。

4. 如权利要求1-3中任一项所述的方法,其特征在于,所述特征库中的空间特征包括训练视频的空间特征,所述特征库中的光流特征包括所述训练视频的光流特征。

5. 如权利要求4所述的方法,其特征在于,所述特征库中的空间特征还包括训练图片的空间特征,所述特征库的光流特征还包括所述训练图片的虚拟光流特征。

6. 如权利要求5所述的方法,其特征在于,所述训练图片的虚拟光流特征是根据所述训练图片的空间特征与所述训练视频的空间特征的相似度,对所述训练视频的光流特征进行加权求和得到的。

7. 如权利要求1-6中任一项所述的方法,其特征在于,所述方法还包括:

根据所述待处理图片的空间特征与所述特征库中的Z个空间特征的相似度,确定所述待处理图片在不同动作类别上的第二类置信度,其中,所述Z个空间特征中的每个空间特征对应一种动作类别,Z为大于1的整数;

所述根据所述第一类置信度确定所述待处理图片的动作类别,包括:

根据所述第一类置信度和所述第二类置信度,确定所述待处理图片的动作类别。

8. 如权利要求7所述的方法,其特征在于,所述根据所述第一类置信度和第二类置信度,确定所述待处理图片的动作类别,包括:

对所述第一类置信度和所述第二类置信度进行加权求和,得到所述待处理图片在不同动作类别上的最终置信度;

根据所述最终置信度确定所述待处理图片的动作类别。

9. 如权利要求1-8中任一项所述的方法,其特征在于,在确定所述待处理图片的动作类别之后,所述方法还包括:

将所述待处理图片的空间特征和虚拟光流特征,以及所述待处理图片的动作类别信息添加到所述特征库中。

10. 一种动作识别装置,其特征在于,包括:

存储器,用于存储程序;

处理器,用于执行所述存储器中存储的程序,当所述存储器中存储的程序被执行时,所述处理器用于:

获取待处理图片;

提取所述待处理图片的空间特征;

根据所述待处理图片的空间特征以及预设的特征库中的X个空间特征和X个光流特征,确定所述待处理图片的虚拟光流特征,其中,所述X个空间特征和所述X个光流特征存在一一对应关系,X为大于1的整数;

根据所述待处理图片的虚拟光流特征与所述特征库中的Y个光流特征的相似度,确定所述待处理图片在不同动作类别上的第一类置信度,其中,所述Y个光流特征中的每个光流特征对应一种动作类别,Y为大于1的整数;

根据所述第一类置信度确定所述待处理图片的动作类别。

11. 如权利要求10所述的装置,其特征在于,所述处理器用于:

根据所述待处理图片的空间特征与所述X个空间特征中的每个空间特征的相似度,确定所述X个光流特征中的每个光流特征的权重系数;

根据所述特征库中的每个光流特征的权重系数,对所述特征库中的光流特征进行加权求和,得到所述待处理图片的虚拟光流特征。

12. 如权利要求11所述的装置,其特征在于,所述X个光流特征中的每个光流特征的权重系数的大小与第一相似度正相关,所述第一相似度为所述X个空间特征中与所述X个光流特征中的每个光流特征相对应的空间特征与所述待处理图片的空间特征的相似度。

13. 如权利要求10-12中任一项所述的装置,其特征在于,所述特征库中的空间特征包括训练视频的空间特征,所述特征库中的光流特征包括所述训练视频的光流特征。

14. 如权利要求13所述的装置,其特征在于,所述特征库中的空间特征还包括训练图片的空间特征,所述特征库的光流特征还包括所述训练图片的虚拟光流特征。

15. 如权利要求14所述的装置,其特征在于,所述训练图片的虚拟光流特征是根据所述训练图片的空间特征与所述训练视频的空间特征的相似度,对所述训练视频的光流特征进行加权求和得到的。

16. 如权利要求10-15中任一项所述的装置,其特征在于,所述处理器用于:

根据所述待处理图片的空间特征与所述特征库中的Z个空间特征的相似度,确定所述待处理图片在不同动作类别上的第二类置信度,其中,所述Z个空间特征中的每个空间特征对应一种动作类别,Z为大于1的整数;

根据所述第一类置信度和所述第二类置信度,确定所述待处理图片的动作类别。

17. 如权利要求16所述的装置,其特征在于,所述处理器用于:

对所述第一类置信度和所述第二类置信度进行加权求和,得到所述待处理图片在不同动作类别上的最终置信度;

根据所述最终置信度确定所述待处理图片的动作类别。

18. 如权利要求10-17中任一项所述的装置,其特征在于,在确定所述待处理图片的动作类别之后,所述处理器还用于将所述待处理图片的空间特征和虚拟光流特征,以及所述待处理图片的动作类别信息添加到所述特征库中。

动作识别方法和装置

技术领域

[0001] 本申请涉及动作识别技术领域,并且更具体地,涉及一种动作识别方法和装置。

背景技术

[0002] 动作识别包括对视频中的人物的动作识别和对图片中的人物的动作识别,由于视频中包含的信息较多,对视频中的人物的动作识别相对比较容易。与视频相比,图片中包含的信息较少,因此,如何有效地识别图片中的人物的动作类别是一个需要解决的问题。

发明内容

[0003] 本申请提供一种动作识别方法和装置,能够提高动作识别的准确度。

[0004] 第一方面,提供了一种动作识别方法,该方法包括:获取待处理图片;确定待处理图片的空间特征;根据待处理图片的空间特征和特征库中的X个空间特征和X个光流特征,确定待处理图片的虚拟光流特征;根据待处理图片的虚拟光流特征与特征库中的Y个光流特征的相似度,确定待处理图片在不同动作类别上的第一类置信度;根据第一类置信度确定待处理图片动作类别。

[0005] 应理解,上述特征库为预先设置的特征库,该特征库中包含多个空间特征和多个光流特征。特征库中的每个空间特征对应一种动作类别,特征库中的每个光流特征对应一种动作类别。上述X和Y均为大于1的整数。

[0006] 可选地,上述每个空间特征对应的动作类别和每个光流特征对应的动作类别是预先根据卷积神经网络模型训练得到的。

[0007] 可选地,特征库中的多个空间特征和多个光流特征存在一一对应关系,特征库中的每个空间特征对应一个光流特征,特征库中的每个光流特征对应一个空间特征。

[0008] 本申请中,通过待处理图片的空间特征以及特征库中的空间特征和光流特征能够获取待处理图片的虚拟光流特征,从而为图片模拟出与动作密切相关的时序信息,接下来就可以根据待处理图片的虚拟光流特征与特征库中的光流特征的相似度对待处理图片进行动作识别。

[0009] 另外,由于本申请是直接通过对比待处理图片的虚拟光流特征与特征库中的光流特征的方式来进行动作识别,无需建立训练模型来对待处理图片进行动作识别,可以利用较少的光流特征实现对待处理图片的动作识别。

[0010] 可选地,上述X个空间特征和X个光流特征分别是特征库中的全部空间特征和全部光流特征。

[0011] 通过根据待处理图片的空间特征,以及特征库中的全部空间特征和全部光流特征来确定待处理图片的虚拟光流特征,可以更准确地确定待处理图片的虚拟光流特征,进而能够更准确地确定待处理图片的动作类别。

[0012] 可选地,上述X个空间特征和X个光流特征分别是特征库中的部分空间特征和部分光流特征。

[0013] 通过结合特征库中的部分空间特征和部分光流特征以及待处理图片的空间特征来确定待处理图片的虚拟光流特征,能够减少计算待处理图片的虚拟光流特征的计算量,进而提高对待处理图片进行动作识别的速度。

[0014] 可选地,上述X个空间特征和X个光流特征一一对应,在X个空间特征和X个光流特征中,每个空间特征对应一个光流特征,每个光流特征对应一个空间特征。

[0015] 应理解,上述Y个光流特征可以是特征库中的全部光流特征,也可以是特征库中的部分光流特征,另外,X与Y既可以相同,也可以不同。

[0016] 当Y个光流特征为特征库中的全部光流特征时,是依据待处理图片的虚拟光流特征与特征库中的全部光流特征的相似度来获得待处理图片的动作类别,可以提高第一类置信度的准确程度,进而提高对待处理图片进行动作识别的效果。

[0017] 而当Y个光流特征为特征库中的部分光流特征时,能够减少确定第一类置信度时的运算量,进而可以提高对待处理图片进行动作识别的速度。

[0018] 可选地,上述待处理图片为包含人物的图片,根据第一类置信度确定待处理图片动作类别包括:根据第一类置信度确定待处理图片中人物的动作类别。

[0019] 也就是说,在本申请中,确定待处理图片的动作类别其实是确定待处理图片中的人物或者其它目标物体的动作类别。

[0020] 可选地,上述待处理图片为静态图片。

[0021] 可选地,上述空间特征的具体为空间特征向量,上述光流特征具体为光流特征向量。

[0022] 在某些实现方式中,根据待处理图片的空间特征以及特征库中的X个空间特征和X个光流特征,确定待处理图片的虚拟光流特征,包括:根据待处理图片的空间特征与特征库中X个的空间特征中的每个空间特征的相似度,对X个光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0023] 在某些实现方式中,上述特征库包含训练视频的空间特征和光流特征。

[0024] 本申请中,可以根据训练视频的空间特征和光流特征以及待处理图片的空间特征来确定待处理图片的虚拟光流特征,进而根据该虚拟光流特征确定待处理图片的动作类别。

[0025] 在某些实现方式中,上述特征库还包含训练图片的空间特征和虚拟光流特征。

[0026] 本申请中,可以综合根据训练视频和训练图片的各自的空间特征和光流特征以及待处理图片的空间特征来确定待处理图片的虚拟光流特征,可以得到更准确的虚拟光流特征,能够进一步提高动作识别的准确度。

[0027] 可选地,训练图片的动作类别与训练视频的动作类别不完全相同。

[0028] 由于训练视频的动作类别和训练图片的动作类别不完全相同,可以增加可以识别的动作类别的种类,进而提高动作识别的适用范围。

[0029] 可选地,训练视频中不同动作类别的视频的数目相同。

[0030] 当训练视频中不同类别的视频数目相同时,能够保证不同动作类别的训练视频的数量均衡性,保证动作识别结果的稳定性。

[0031] 可选地,上述方法还包括:从预设的图片库中选择出与需要识别的动作类别相匹配的图片,得到所述训练图片。

[0032] 上述图片库可以是本地的图片数据库,也可以是位于网络服务器中的图片数据库。

[0033] 可选地,上述方法还包括:从预设的视频库中选择出与训练图片的空间特征的相似度满足预设要求的视频,得到所述训练视频。

[0034] 上述视频库可以是本地的视频库,也可以是网络服务器中的视频库。

[0035] 具体地,从预设的视频库中选择出与训练图片的空间特征的相似度满足预设要求的视频,得到训练视频,包括:从预设的视频库中选择出与训练图片的空间特征的相似度大于预设的相似度阈值的视频,得到训练视频。

[0036] 例如,可以将预设的视频库中与训练图片的空间特征的相似度大于0.5的视频都选择出来,组成训练视频。

[0037] 可选地,从预设的视频库中选择出与训练图片的空间特征的相似度满足预设要求的视频,得到训练视频,包括:确定视频库中的视频的空间特征与训练图片的空间特征的相似度;将视频库中与训练图片的空间特征的相似度最大的前J个视频选择出来,得到训练视频,其中,J小于K,J和K均为大于0的整数,K为视频库中视频的总数。

[0038] 例如,视频库中一共包含100个视频,那么,可以将视频库中与训练图片的空间特征的相似度最大的前50个视频选择出来构成训练视频。

[0039] 在某些实现方式中,上述根据待处理图片的空间特征以及特征库中的X个空间特征和X个光流特征,确定待处理图片的虚拟光流特征,具体包括:根据待处理图片的空间特征与X个空间特征中的每个空间特征的相似度,确定特征库中与X个空间特征中的每个空间特征相对应的光流特征的权重系数;根据X个光流特征中的每个光流特征的权重系数,对X个光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0040] 应理解,特征库中相互对应的空间特征和光流特征对应的是同一个视频或者图片,也就是说,特征库中相互对应的空间特征和光流特征属于同一个视频或者同一个图片。

[0041] 在某些实现方式中,上述X个光流特征中的每个光流特征的权重系数的大小与第一相似度是正相关的关系,其中,该第一相似度是X个空间特征中与X个光流特征中的每个光流特征相对应的空间特征与待处理图片的空间特征的相似度。

[0042] 例如,上述X个空间特征中包括第一空间特征,上述X个光流特征中包括第一光流特征,第一空间特征与第一光流特征存在对应关系,第一空间特征与待处理图片的空间特征的相似度为相似度1,那么,第一光流特征的权重系数的大小与相似度1是正相关的关系(具体可以是成正比的关系)。

[0043] 通过合理设置光流特征的权重系数使得根据特征库中的光流特征得到的待处理图片的虚拟光流特征更准确。

[0044] 可选地,特征库包含训练视频的空间特征和光流特征,根据待处理图片的空间特征以及特征库中的空间特征和光流特征,确定待处理图片的虚拟光流特征,具体包括:根据待处理图片的空间特征与训练视频的每个空间特征的相似度,确定与训练视频的每个空间特征相对应的光流特征的权重系数;根据训练视频中的每个光流特征的权重系数,对特征库中的光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0045] 应理解,上述训练视频的空间特征和光流特征均为多个。

[0046] 本申请中,只根据训练视频的空间特征和光流特征来确定待处理图片的虚拟光流

特征,能够减少确定虚拟光流特征的复杂度。

[0047] 可选地,特征库包含训练视频的空间特征和光流特征以及训练图片的空间特征和虚拟光流特征,根据待处理图片的空间特征以及特征库中的空间特征和光流特征,确定待处理图片的虚拟光流特征,具体包括:根据待处理图片的空间特征与训练视频和训练图片的每个空间特征的相似度,确定与训练视频和训练图片的每个空间特征相对应的光流特征的权重系数;根据训练视频和训练图片中的每个光流特征的权重系数,对训练视频和训练图片中的光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0048] 应理解,上述训练图片的空间特征和光流特征均为多个。

[0049] 本申请中,通过训练视频的空间特征和光流特征以及训练图片的空间特征和虚拟光流特征来综合确定待处理图片的虚拟光流特征,能够使得获取的待处理图片的虚拟光流特征更能够反映待处理图片的运动信息。

[0050] 在某些实现方式中,所述训练图片的虚拟光流特征是根据所述训练图片的空间特征与所述训练视频的空间特征的相似度,对所述训练视频的光流特征进行加权求和得到的。

[0051] 在某些实现方式中,上述方法还包括:根据训练图片的空间特征与训练视频的空间特征的相似度,对训练视频的光流特征进行加权求和,得到训练图片的虚拟光流特征。

[0052] 可选地,根据训练图片的空间特征与训练视频的空间特征的相似度,对训练视频的光流特征进行加权求和,得到训练图片的虚拟光流特征,包括:根据训练图片的空间特征与训练视频中的每个空间特征的相似度,确定训练视频中与每个空间特征相对应的光流特征的权重系数;根据训练视频中的每个光流特征的权重系数,对训练视频中的光流特征进行加权求和,得到训练图片的虚拟光流特征。

[0053] 应理解,上述特征库在初始时可以仅包含训练视频的空间特征和光流特征,为了进一步提高最终动作识别的准确性,可以在特征库中再加入训练图片的空间特征和虚拟光流特征,而该训练图片的虚拟光流特征可以根据特征库中包含的训练视频的空间特征和光流特征来确定。

[0054] 因此,本申请中,通过训练视频的空间特征和光流特征确定训练图片的虚拟光流特征,并将训练图片的空间特征和虚拟光流特征并入到特征库中,能够在一定程度上提高动作识别的效果。

[0055] 在某些实现方式中,上述方法还包括:根据待处理图片的空间特征与预设的特征库中的 Z 个空间特征的相似度,确定待处理图片在不同动作类别上的第二类置信度,其中, Z 个空间特征中的每个空间特征对应一种动作类别;根据第一类置信度确定待处理图片的动作类别,包括:根据第一类置信度和第二类置信度,确定待处理图片的动作类别。

[0056] 应理解,上述第一类置信度是通过光流预测过程得到的,上述第二类置信度是通过空间预测流程得到的。 Z 为大于1的整数。 X 、 Y 和 Z 中的任意两个数值可以相同,也可以不同。另外,上述 Z 个空间特征既可以是特征库中的全部空间特征也可以只是特征库中的部分空间特征。

[0057] 本申请中,通过光流预测和空间预测来综合得到待处理图片的置信度,能够更准确地确定待处理图片的动作类别。

[0058] 在某些实现方式中,根据第一类置信度和第二类置信度,确定待处理图片的动作

类别,包括:对第一类置信度和第二类置信度进行加权求和,得到待处理图片在不同动作类别上的最终置信度;根据最终置信度确定待处理图片的动作类别。

[0059] 通过对第一类置信度和第二类置信度进行加权求和,能够得到可以综合反映待处理图片在不同动作类别上的置信度,能够更好地确定待处理图片的动作类别。

[0060] 在某些实现方式中,在确定待处理图片的动作类别之后,上述方法还包括:将待处理图片的空间特征和虚拟光流特征,以及待处理图片的动作类别信息添加到所述特征库中。

[0061] 通过将待处理图片的空间特征和虚拟光流特征,以及对应的动作类别信息添加到特征库中,能够扩充特征库中包含的空间特征和光流特征,便于后续依据特征库中的空间特征和光流特征更好地对图片进行动作识别。

[0062] 第二方面,提供一种动作识别装置,该动作识别装置包括用于执行第一方面中的方法的模块。

[0063] 第三方面,提供一种动作识别装置,该动作识别装置包括:存储器,用于存储程序;处理器,用于执行存储器存储的程序,当存储器存储的程序被执行时,处理器用于执行第一方面中的方法。

[0064] 第四方面,提供一种计算机可读介质,该计算机可读介质存储用于设备执行的程序代码,该程序代码包括用于执行第一方面中的方法。

[0065] 第五方面,提供一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述第一方面中的方法。

[0066] 第六方面,提供一种电子设备,该电子设备包括上述第二方面或第三方面中的动作识别装置。

附图说明

[0067] 图1是本申请实施例的动作识别方法的示意性流程图;

[0068] 图2是本申请实施例的动作识别方法的示意性流程图;

[0069] 图3是根据CNN模型提取空间特征的示意图;

[0070] 图4是提取视频的空间特征和光流特征的示意图;

[0071] 图5是获取输入图片的虚拟光流特征的示意图;

[0072] 图6是对输入图片进行光流预测的示意图;

[0073] 图7是建立光流特征库和空间特征库的示意图;

[0074] 图8是本申请实施例的动作识别装置的示意性框图;

[0075] 图9是本申请实施例的动作识别装置的示意性框图;

[0076] 图10是本申请实施例的动作识别装置的示意性框图;

[0077] 图11是本申请实施例的动作识别装置对输入图片进行动作识别的示意性框图。

具体实施方式

[0078] 下面将结合附图,对本申请中的技术方案进行描述。

[0079] 本申请实施例的动作识别方法能够应用在图片检索、相册管理、平安城市、人机交互以及需要识别进行动作识别的场景。具体而言,本申请实施例的动作识别方法能够应用

在相册管理系统和图片查找系统中,下面分别对相册管理系统和图片查找系统进行简单的介绍。

[0080] 相册管理系统:

[0081] 当用户在手机或者云盘上存储了大量的图片时,为了方便地查找不同类型的图片,可以对相册中的图片进行分类。例如,可以利用本申请实施例的动作识别方法对相册中的图片进行动作识别,得到每个图片的动作类别,使得用户能够对不同动作类别的图片进行分类管理,从而方便用户查找图片,能够节省管理时间,提高相册管理的效率。

[0082] 图片查找系统:

[0083] 互联网上有海量的图片,目前查询图片时主要是根据图片所在的网页的文字进行查找,而图片所在的网页的文字并不能完全反映图片本身的特征。采用本申请实施例的动作识别方法能够识别出图片中的动作类型,进而可以根据图片的动作类别从互联网或者数据库中查找出自己感兴趣的图片。

[0084] 为了实现对图片的动作识别,传统方案是从大量的训练图片中提取人物图像,然后根据从训练图片中提取到的人物图像以及人物图像对应的动作类别对卷积神经网络(Convolutional Neural Networks,CNN)模型进行训练,得到CNN模型的各个参数。当需要对待处理图片中的人物进行动作识别时,可以从该待处理图片中提取人物图像,并将从待处理图片中提取出来的人物图像输入到已经训练好的CNN模型进行动作识别,从而确定待处理图片的动作类别。传统方案在进行动作识别时仅考虑到了图片的空间特征,而未进一步探索待处理图片中的人物的动作的时间属性,导致进行动作识别的准确率仍然较低。

[0085] 因此,本申请提出了一种新的动作识别方法,通过已知的光流特征来模拟待处理图片的光流特征,以得到待处理图片的虚拟光流特征,接下来再根据待处理图片的虚拟光流特征来对待处理图片的动作进行识别。

[0086] 下面结合图1对本申请实施例的动作识别方法进行详细的介绍。图1所示的方法可以由动作识别装置执行,该动作识别装置具体可以是监控设备、终端设备、网络服务器以及网络云平台等具有图片处理功能的设备。图1所示的方法包括步骤101至步骤105,下面分别对步骤101至步骤105分别进行详细的描述。

[0087] 101、获取待处理图片。

[0088] 上述待处理图片(也可以称为待处理图像)可以是包含人物的图片。对待处理图片进行动作识别实质上是对待处理图片中的人物的动作进行识别,确定待处理图片的动作类别。

[0089] 上述待处理图片可以通过电子设备拍摄的照片,也可以是从视频中截取的图片。上述待处理图片可以存储在本地图片数据库中,也可以存储在网络中。

[0090] 在获取待处理图片时可以从本地图片数据库中直接调取,也可以从网络中在线获取。

[0091] 在本申请中,动作类别可以是指待处理图片中的人物具体在做什么动作,例如,动作类别可以包括:跑步(run)、行走(walk)、棒球投掷(Baseball_pitch)、棒球击球(baseball_swing)、投球(bowl)、挺举(clean_and_jerk)、打高尔夫球(golf_swing)、跳绳(jump_rope)、引体向上(pullup)、俯卧撑(pushup)、端坐(situp)、蹲坐(squat)、弹吉他(strum_guitar)以及游泳(swim)等等。

[0092] 应理解,上述例子只是动作类型的一些具体例子(主要是体育运动方面的动作类别),事实上,本申请实施例包含的动作类型不限于此,本申请的动作类别还可以包含体育运动之外的其它动作,例如,看手机、人机交互系统中人的姿势等等。另外,上述动作类别还可以称为动作种类、动作类型等等。

[0093] 102、提取待处理图片的空间特征。

[0094] 在确定待处理图片的空间特征时,可以采用卷积神经网络(Convolutional Neural Networks, CNN)模型(该CNN模型可以是预先训练好的模型)对待处理图片进行卷积运算,从而得到待处理图片的空间特征。

[0095] 103、根据待处理图片的空间特征以及预设的特征库中的X个空间特征和Y个光流特征,确定待处理图片的虚拟光流特征。

[0096] 其中,上述X和Y均为大于1的整数。

[0097] 应理解,上述特征库可以是预先设置的特征库,该特征库中包含多个空间特征和多个光流特征,该多个空间特征中的每个空间特征对应一种动作类别,该多个光流特征中的每个光流特征也对应一种动作类别。其中,每个空间特征对应的动作类别以及每个光流特征对应的动作类别可以是预先根据卷积神经网络模型训练得到的。

[0098] 另外,上述特征库中的多个空间特征和多个光流特征存在一一对应关系,多个空间特征中的任意一个空间特征对应多个光流特征中的一个光流特征,多个光流特征中的任意一个光流特征对应多个空间特征中的一个特征。特征库中的空间特征和光流特征的数目一般是相同的。

[0099] 上述X个空间特征和X个光流特征也可以是一一对应的,也就是说,在X个空间特征和X个光流特征中,每个空间特征对应一个光流特征,每个光流特征对应一个空间特征。

[0100] 上述X个空间特征可以是特征库中的全部空间特征或者部分空间特征。

[0101] 上述X个光流特征可以是特征库中的全部光流特征或者部分光流特征。

[0102] 当上述X个空间特征和X个光流特征分别为特征库中的全部空间特征和全部光流特征时,能够根据待处理图片的空间特征,以及特征库中的全部空间特征和全部光流特征来确定待处理图片的虚拟光流特征,可以更准确地确定待处理图片的虚拟光流特征,进而能够更准确地确定待处理图片的动作类别。

[0103] 当上述X个空间特征和X个光流特征分别为特征库中的部分空间特征和部分光流特征时,通过结合特征库中的部分空间特征和部分光流特征以及待处理图片的空间特征来确定待处理图片的虚拟光流特征,能够减少计算待处理图片的虚拟光流特征的计算量,进而提高对待处理图片进行动作识别的速度。

[0104] 另外,上述特征库中存在对应关系的空间特征和光流特征对应的是同一个视频或者图片,也就是说,特征库中对应关系的空间特征和光流特征属于同一个视频或者同一张图片。另外,本申请中提及的空间特征具体表现形式可以为空间特征向量,光流特征或者虚拟光流特征具体表现形式可以为光流特征向量或者虚拟光流特征向量。

[0105] 具体地,由于特征库中的每个空间特征对应一个光流特征(特征库中的空间特征和光流特征是一一对应关系),因此,在根据待处理图片的空间特征以及预设的特征库中的空间特征和光流特征,确定待处理图片的虚拟光流特征时可以根据待处理图片的空间特征与特征库中的空间特征的相似度,对特征库中与特征库中的空间特征相对应的光流特征进

行加权求和,得到待处理图片的虚拟光流特征。

[0106] 因此,上述步骤103中的根据待处理图片的空间特征以及预设的特征库中的X个空间特征和X个光流特征,确定待处理图片的虚拟光流特征,包含以下具体过程:

[0107] (1)、根据待处理图片的空间特征与X个空间特征中的每个空间特征的相似度,确定X个光流特征中的每个光流特征的权重系数(也可以称为加权系数);

[0108] (2)、根据X个光流特征中的每个光流特征的权重系数,对X个光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0109] 可选地,上述X个光流特征中的每个光流特征的权重系数的大小与第一相似度是正相关的关系,其中,该第一相似度是X个空间特征中与X个光流特征中的每个光流特征相对应的空间特征与待处理图片的空间特征的相似度。

[0110] 例如,上述X个空间特征中包括第一空间特征和第二空间特征,上述Y个光流特征中包括第一光流特征和第二光流特征,第一空间特征对应第一光流特征,第二空间特征对应第二光流特征,待处理图片的空间特征与第一空间特征的相似度为相似度1,待处理图片的空间特征与第二空间特征的相似度为相似度2,相似度1大于相似度2,那么,在对第一光流特征和第二光流特征以及X个光流特征中的其它光流特征进行加权求和时,第一光流特征的权重系数大于第二光流特征的权重系数。

[0111] 本申请中,通过合理设置光流特征的权重系数使得根据特征库中的光流特征得到的待处理图片的虚拟光流特征更准确。

[0112] 104、根据待处理图片的虚拟光流特征与预设的特征库中的Y个光流特征的相似度,确定待处理图片在不同动作类别上的第一类置信度。

[0113] 其中,上述Y个光流特征中的每个光流特征对应一种动作类别,Y为大于1的整数。

[0114] 应理解,上述Y个光流特征可以是特征库中的全部光流特征,也可以是特征库中的部分光流特征,另外,Y与X既可以相同,也可以不同。

[0115] 当Y个光流特征为特征库中的全部光流特征时,是依据待处理图片的虚拟光流特征与特征库中的全部光流特征的相似度来获得待处理图片的动作类别,可以提高第一类置信度的准确程度,进而提高对待处理图片进行动作识别的效果。

[0116] 而当Y个光流特征为特征库中的部分光流特征时,能够减少确定第一类置信度时的运算量,进而可以提高对待处理图片进行动作识别的速度。

[0117] 105、根据第一类置信度确定待处理图片的动作类别。

[0118] 本申请中,通过待处理图片的空间特征以及预设特征库中的空间特征和光流特征能够获取待处理图片的虚拟光流特征,从而为图片模拟出与动作密切相关的时序信息,接下来就可以根据待处理图片的虚拟光流特征与预设特征库中的光流特征的相似度对待处理图片进行动作识别。

[0119] 进一步地,由于本申请是直接通过对比待处理图片的虚拟光流特征与特征库中的光流特征的方式来进行动作识别,无需建立训练模型来对待处理图片进行动作识别,可以利用较少的光流特征实现对待处理图片的动作识别。

[0120] 可选地,上述特征库中的空间特征包含训练视频的空间特征,特征库中的光流特征包含训练视频的光流特征。

[0121] 其中,上述训练视频的空间特征可以为多个,上述训练视频的光流特征也可以为

多个。

[0122] 本申请中,根据训练视频的空间特征和光流特征能够模拟出待处理图片的虚拟光流特征,进而可以综合待处理图片的空间特征和虚拟光流特征进行动作识别,提高动作识别的准确性。

[0123] 可选地,上述特征库的空间特征还包含训练图片的空间特征,上述特征库中的光流特征还包括训练图片的虚拟光流特征。

[0124] 其中,上述训练图片的空间特征可以为多个,上述训练图片的光流特征也可以为多个。

[0125] 上述特征库中不仅包含训练视频的空间特征和光流特征,还包含训练图片的空间特征和光流特征,能够综合训练视频和训练图片的空间特征和光流特征来确定待处理图片的虚拟光流特征,可以进一步提高最终动作识别的准确度。

[0126] 通过训练视频的空间特征和光流特征以及训练图片的空间特征和虚拟光流特征来综合确定待处理图片的虚拟光流特征,能够得到更准确的虚拟光流特征。

[0127] 应理解,在根据特征库中空间特征和光流特征确定待处理图片的虚拟光流特征时,既可以只根据特征库中训练视频的空间特征和光流特征来确定待处理图片的虚拟光流特征,也可以结合特征库中训练视频的空间特征和光流特征以及训练图片的空间特征和虚拟光流特征来综合确定待处理图片的虚拟光流特征。

[0128] 可选地,特征库包含训练视频的空间特征和光流特征,根据待处理图片的空间特征以及特征库中的空间特征和光流特征,确定待处理图片的虚拟光流特征,具体包括:根据待处理图片的空间特征与训练视频的每个空间特征的相似度,确定与训练视频的每个空间特征相对应的光流特征的权重系数;根据训练视频中的每个光流特征的权重系数,对特征库中的光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0129] 本申请中,只根据训练视频的空间特征和光流特征来确定待处理图片的虚拟光流特征,能够减少确定虚拟光流特征的复杂度。

[0130] 可选地,特征库包含训练视频的空间特征和光流特征以及训练图片的空间特征和虚拟光流特征,根据待处理图片的空间特征以及特征库中的空间特征和光流特征,确定待处理图片的虚拟光流特征,具体包括:根据待处理图片的空间特征与训练视频和训练图片的每个空间特征的相似度,确定与训练视频和训练图片的每个空间特征相对应的光流特征的权重系数;根据训练视频和训练图片中的每个光流特征的权重系数,对训练视频和训练图片中的光流特征进行加权求和,得到待处理图片的虚拟光流特征。

[0131] 本申请中,通过训练视频的空间特征和光流特征以及训练图片的空间特征和虚拟光流特征来综合确定待处理图片的虚拟光流特征,能够使得获取的待处理图片的虚拟光流特征更能够反映待处理图片的运动信息。

[0132] 可选地,上述特征库中的训练图片的虚拟光流特征可以是根据训练视频的空间特征和光流特征以及训练图片的空间特征得到的。也就是说,训练图片的虚拟光流特征是根据训练图片的空间特征与训练视频的空间特征的相似度,对训练视频的光流特征进行加权求和得到的。

[0133] 具体地,可以在对待处理图片进行动作识别之前先确定训练图片的虚拟光流特征。

[0134] 可选地,作为一个实施例,图1所示的方法还包括:根据训练图片的空间特征与训练视频的空间特征的相似度,对训练视频的光流特征进行加权求和,得到训练图片的虚拟光流特征。

[0135] 上述根据训练图片的空间特征与训练视频的空间特征的相似度,对训练视频的光流特征进行加权求和,得到训练图片的虚拟光流特征,具体包括:

[0136] 根据训练图片的空间特征与训练视频中的每个空间特征的相似度,确定训练视频中与每个空间特征相对应的光流特征的权重系数;根据训练视频中的每个光流特征的权重系数,对训练视频中的光流特征进行加权求和,得到训练图片的虚拟光流特征。

[0137] 除了根据待处理图片的虚拟光流特征计算待处理图片在不同动作类别上置信度之外,还可以再根据待处理图片的空间特征来计算待处理图片在不同动作类别上的置信度,然后根据这两类置信度来综合判断待处理图片的动作类别。

[0138] 具体地,在步骤102中提取到待处理图片的空间特征之后,可以根据待处理图片的空间特征与预设的特征库中的Z个空间特征的相似度,确定待处理图片在不同动作类别上的第二类置信度,其中,该Z个空间特征中的每个空间特征对应一种动作类别。

[0139] 上述Z为大于1的整数,Z与上述X或者Y的数值既可以相同,也可以不同,上述Z个空间特征既可以是特征库中的全部空间特征也可以只是特征库中的部分空间特征。

[0140] 在得到了上述第二类置信度之后,可以综合根据第一类置信度和第二类置信度来确定待处理图片的动作类别。

[0141] 本申请中,通过光流预测和空间预测来综合得到待处理图片的置信度,能够更准确地确定待处理图片的动作类别。

[0142] 具体地,在根据第一类置信度和第二类置信度来确定待处理图片的动作类别时,可以先对第一类置信度和第二类置信度进行加权求和,得到待处理图片在不同动作类别上的最终置信度,然后再根据最终置信度确定待处理图片的动作类别。

[0143] 通过对第一类置信度和第二类置信度进行加权求和,能够得到可以综合反映待处理图片在不同动作类别上的置信度,能够更好地确定待处理图片的动作类别。

[0144] 应理解,还可以根据第一类置信度和第二类置信度分别确定待处理图片的动作类别,然后再确定待处理图片的动作类别。

[0145] 可选地,根据第一类置信度和第二类置信度,确定待处理图片的动作类别,包括:根据第一类置信度确定待处理图片的动作类别为第一动作类别;根据第二类置信度确定待处理图片的动作类别为第二动作类别;在第一动作类别和第二动作类别相同的情况下,确定待处理图片的动作类别为第一动作类别。

[0146] 为了增加特征库中包含的空间特征和光流特征,便于后续对图片进行更好的动作识别,在确定了待处理图片的动作类别之后,还可以将待处理图片的空间特征和光流特征以及该待处理图片的动作类别等信息添加到特征库中。

[0147] 可选地,作为一个实施例,在确定了待处理图片的动作类别之后,图1所示的方法还包括:将待处理图片的空间特征和虚拟光流特征,以及待处理图片的动作类别信息添加到所述特征库中。

[0148] 通过将待处理图片的空间特征和虚拟光流特征,以及对应的动作类别信息添加到特征库中,能够扩充特征库中包含的空间特征和光流特征,便于后续依据特征库中的空间

特征和光流特征更好地对图片进行动作识别。

[0149] 下面结合图2对本申请实施例的动作识别方法的过程进行详细的描述，

[0150] 图2是本申请实施例的动作识别方法的示意图。图2所示的动作识别方法的具体过程包括：

[0151] 201、获取输入图片。

[0152] 输入图片相当于上文中的待处理图片。

[0153] 202、提取输入图片的空间特征。

[0154] 具体地，在步骤202中，可以采用卷积神经网络CNN模型来提取输入图片的空间特征。如图3所示，通过CNN模型对输入图片进行卷积处理，得到输入图像的卷积特征图 $C_t \in R^{K_1 \times K_2 \times d_c}$ ，接下来再将 $C_t \in R^{K_1 \times K_2 \times d_c}$ 拉成一维向量，得到向量 u^{rgb} 。向量 u^{rgb} 就是输入图片的空间特征。CNN模块可以采用多种构架实现，例如，VGG16、TSN网络等。另外，CNN模块的系数需要在动作识别的数据集上进行预训练。

[0155] 203、生成输入图片的虚拟光流特征。

[0156] 在步骤203中，可以根据视频仓库中存储的视频的光流特征来模拟或者生成输入图像的虚拟光流特征。

[0157] 具体地，假设视频仓库中存储了N个视频，那么，可以根据该N个视频的虚拟光流特征来生成输入图片的虚拟光流特征。

[0158] 在生成输入图片的虚拟光流特征之前需要先获取N个视频的空间特征和光流特征。对于N个视频中的每个视频来说，都可以按照图4所示的过程来提取该视频的空间特征和光流特征。

[0159] 如图4所示，提取视频的空间特征和光流特征的具体过程包括：

[0160] 首先，抽取视频中间的RGB图像和中间的光流图（该中间的光流图包括光流x和光流y）；

[0161] 其次，将视频中间的RGB图像送入到预先训练好的空间特征CNN模型，得到视频的空间特征；

[0162] 再次，将视频中间的光流图送入到预先训练好的光流特征CNN模型，得到视频的光流特征；

[0163] 最后，将每个视频的空间特征和光流特征送入到视频仓库中。

[0164] 应理解，视频中间的光流图可以由视频中间时刻前后若干帧图片产生。另外，提取视频的空间特征与提取视频的光流特征可以是相互独立的，两者既可以同时进行，也可以依次进行。

[0165] 上述图4所示的过程中提取到的视频的空间特征和光流特征具体可以是空间特征向量和光流特征向量，其中，每个视频的空间特征向量和光流特征向量的长度可以均为M，那么，N个视频空间特征向量就可以用一个矩阵 $V^{rgb} \in N * M$ 表示，N个视频光流特征向量可以用一个矩阵 $V^{flow} \in N * M$ 表示。这样就获取到了N个视频的空间特征向量 $V^{rgb} \in N * M$ 和光流特征向量 $V^{flow} \in N * M$ 。

[0166] 在得到了视频仓库中的N个视频的空间特征和光流特征之后，就可以根据输入图片的空间特征与N个视频中的每个视频的空间特征的相似度，对N个视频的光流特征进行加权求和，从而得到输入图片的虚拟光流特征。

[0167] 如图5所示,根据输入图片的空间特征与N个视频中的每个视频的空间特征的相似度,对N个视频的光流特征进行加权求和,得到输入图片的虚拟光流特征的具体过程包括:

[0168] 首先,将输入图片的空间特征和视频仓库中的视频的空间特征进行比较,得到输入图片的空间特征与视频仓库中的每个视频的空间特征的相似度。

[0169] 其次,根据输入图片的空间特征与视频仓库中的每个视频的空间特征的相似度,对视频仓库中的每个视频的光流特征进行加权求和,得到输入图片的虚拟光流特征。

[0170] 具体地,可以采用高斯过程来计算输入图片的空间特征和视频仓库中的视频的空间特征的相似度。例如,可以采用公式(3)来确定输入图片的空间特征和视频仓库中的视频的空间特征的相似度。

$$[0171] \quad w_h = K_h(u^{rgb}, V^{rgb}) [K_h(V^{rgb}, V^{rgb}) + \sigma_h^2 \mathbf{I}]^{-1} \in 1 * N \quad (3)$$

[0172] 其中, u^{rgb} 为输入图片的空间特征, V^{rgb} 为视频仓库中的视频的空间特征, $K_h(u^{rgb}, V^{rgb}) \in 1 * N$ 中的每个元素是 u^{rgb} 与 V^{rgb} 每行的点积, $K_h(V^{rgb}, V^{rgb})$ 是 V^{rgb} 的协方差矩阵, σ_h^2 是一个噪声参数, \mathbf{I} 为单位矩阵, w_h 为输入图片的空间特征和视频仓库中的视频的空间特征的相似度。其中, w_h 是一个长度为N的一维向量, w_h 中的第i个元素表示输入图片的空间特征与第i个视频的空间特征的相似度, w_h 中的第i个元素的数值越大,输入图片的空间特征与第i个视频的空间特征的相似度越大。

[0173] 在得到了输入图片的空间特征与视频仓库中的每个视频的空间特征的相似度之后,具体可以采用公式(4)来计算输入图片的虚拟光流特征。

$$[0174] \quad u^{flow} = w_h * V^{flow} \in 1 * M \quad (4)$$

[0175] 其中, w_h 表示输入图片的空间特征与视频仓库中的视频的空间特征的相似度, V^{flow} 表示视频仓库中的视频的光流特征, u^{flow} 表示输入图片的虚拟光流特征, u^{flow} 也是一个长度为M的特征向量。

[0176] 204、对输入图片进行光流预测,得到输入图片中的人物在各个动作类别上的第一类置信度。

[0177] 如图6所示,对输入图片进行光流预测,得到第一类置信度的具体过程如下:

[0178] 首先,确定输入图片的虚拟光流特征与光流特征库中的光流特征的相似度;

[0179] 其次,根据输入图片的虚拟光流特征与光流特征库中的光流特征的相似度,对光流特征库中的光流特征对应的动作类别标签进行加权求和,得到光流预测的置信度(相当于步骤204中的第一类置信度)。

[0180] 其中,动作类别标签(One-hot-label)用来表示每个视频或图片的动作类别标签。动作类别标签可以采用一个向量来表示,该向量的长度与动作类型的总数相同,动作类别标签中的每个位置对应一个动作类别。该向量有且仅有一个位置的值为1,其余位置为0,其中,值为1的位置对应的动作类别即为该视频或图片的动作类别。

[0181] 例如,现有3个视频和3张图片:视频1(Video1)、视频2(Video2)、视频3(Video3)、图片1(Image1)、图片2(Image2)和图片3(Image3)。动作类别依次为跑步(Run)、跳舞(Dance)、跑步(Run)、跳跃(Jump)、跳舞(Dance)和跑步(Run)。

[0182] 表1

[0183]

	视频1	视频2	视频3	图片1	图片2	图片3
跑步	1	0	1	0	0	1
跳舞	0	1	0	0	1	0
跳跃	0	0	0	1	0	0

[0184] 如表1所示,这3个视频和3张图片的动作类别标签依次为 $[1, 0, 0]$, $[0, 1, 0]$, $[1, 0, 0]$, $[0, 0, 1]$, $[0, 1, 0]$, $[1, 0, 0]$ 。那么,根据这些动作类别标签可知这3个视频和3张图片依次对应的动作类别分别是跑步、跳舞、跑步、跳跃、跳舞和跑步。

[0185] 上述光流特征库中可以包含 N_v 个视频的光流特征和 N_i 张图片的虚拟光流特征,其中, N_v 个视频的光流特征为 $V_{train}^{flow} \in N_v * M$, N_i 张图片的虚拟光流特征为 $U_{train}^{flow} \in N_i * M$,两者

共同构成了光流特征仓库的光流特征 $M^{flow} = \begin{bmatrix} V_{train}^{flow} \\ U_{train}^{flow} \end{bmatrix} \in (N_v + N_i) * M$ 。

[0186] 在计算输入图片的虚拟光流特征与光流特征库中的光流特征的相似度时,仍然可以采用高斯过程进行计算。例如,可以采用公式(5)来计算输入图片的虚拟光流特征与光流特征库中的光流特征的相似度。

$$[0187] \quad w_{flow} = K_p(u^{flow}, M^{flow}) [K_p(M^{flow}, M^{flow}) + \Sigma_p]^{-1} \in 1 * (N_v + N_i) \quad (5)$$

[0188] 其中, u^{flow} 表示输入图片的虚拟光流特征, M^{flow} 表示光流特征仓库的光流特征, $K_p(u^{flow}, M^{flow}) \in 1 * (N_v + N_i)$ 中每个元素是 u^{flow} 与 M^{flow} 每行的点积, $K_p(M^{flow}, M^{flow})$ 是 M^{flow} 的协方差矩阵, $\Sigma_p = \begin{bmatrix} \sigma_v^2 \mathbf{I} & 0 \\ 0 & \sigma_i^2 \mathbf{I} \end{bmatrix}$ 是一个噪声参数矩阵, w_{flow} 表示输入图片的虚拟光流特征与光流

特征库中的光流特征的相似度, w_{flow} 是一个长度为 $N_v + N_i$ 的一维向量,其中第 i 个元素表示输入图片的虚拟光流特征与第 i 个光流特征的相似度,这个值越大,说明输入图片的光流特征与该第 i 个光流特征相越接近。

[0189] 在得到输入图片的虚拟光流特征与光流特征库中的光流特征的相似度之后,可以根据公式(6)来计算输入图片在各个动作类别上的第一类置信度。

$$[0190] \quad L_{flow} = (w_{flow} L) \in 1 * P \quad (6)$$

[0191] 其中 $L \in (N_v + N_i) * P$ 中的每一行表示光流特征仓库中每个光流特征对应的动作类别标签, P 是动作类别的总数,对于每个动作类别标签来说,只有在其所属类别上为1,其余位置为0。其中, L_{flow} 也就是输入图片中的人物在各个动作类别上的第一类置信度。

[0192] 下面结合表2对计算输入图片中的人物在各个动作类别上的第一类置信度的过程进行说明。假设光流特征仓库中包含3个视频和3张图片:视频1 (Video1)、视频2 (Video2)、视频3 (Video3)、图片1 (Image1)、图片2 (Image2) 和图片3 (Image3)。动作类别依次为跑步 (Run)、跳舞 (Dance)、跑步 (Run)、跳跃 (Jump)、跳舞 (Dance) 和跑步 (Run)。这3个视频和3张图片各自对应的动作类别标签如表2的第2列至第7列(不包含最后一行)所示,输入图片与光流特征库中的3个视频/图片的光流特征的相似度如表2最后一行所示,最终得到的输入图片在各个动作类别上的第一类置信度如表2最后一列所示。

[0193] 表2

[0194]

	视频1	视频2	视频3	图片1	图片2	图片3	置信度
跑步	1	0	1	0	0	1	0.3
跳舞	0	1	0	0	1	0	0.45
跳跃	0	0	0	1	0	0	0.25
	0.1	0.2	0.01	0.25	0.25	0.19	

[0195] 205、对输入图片进行空间预测,得到输入图片中的人物在各个动作上的第二类置信度。

[0196] 对输入图片进行空间预测的过程与对输入图片进行光流预测的过程基本相同,首先比较输入图片的空间特征 u^{rgb} 与空间特征库中的空间特征 M^{rgb} 的相似度 w_{rgb} ,然后利用 w_{rgb} 去加权动作类别得到在空间上每个类别的置信度 L_{rgb} ,最终得到的输入图片中的人物在各个动作类别上的第二类置信度。

[0197] 上述步骤204和步骤205中采用的光流特征库和空间特征库可以是预先建立好的特征库。

[0198] 具体地,可以采用图7所示的过程来建立光流特征库(也可以称为光流特征仓库)和空间特征库(也可以称为空间特征仓库)。

[0199] 如图7所示,从训练视频集合和训练图片集合中提取空间特征和光流特征,并将提取到的训练视频集合的空间特征和训练图片集合中的空间特征送入到空间特征库中,将提取到的训练视频集合的光流特征和训练图片集合中的光流特征送入到光流特征库中。

[0200] 假设,最终得到的空间特征库包含 N_v 个训练视频的空间特征和 N_i 张图片的空间特征,最终得到的光流特征库中包含 N_v 个训练视频的光流特征和 N_i 张图片的虚拟光流特征。那

么,空间特征库中的空间特征可以表示为 $M^{rgb} = \begin{bmatrix} V_{train}^{rgb} \\ U_{train}^{rgb} \end{bmatrix} \in (N_v + N_i) * M$,光流特征库中的光

流特征可以表示为 $M^{flow} = \begin{bmatrix} V_{train}^{flow} \\ U_{train}^{flow} \end{bmatrix} \in (N_v + N_i) * M$ 。

[0201] 另外,上述训练视频集合和训练图片集合可以是存储在本地的数据库中的视频集合和图片集合。

[0202] 206、对第一类置信度和第二类置信度进行融合处理,得到目标置信度。

[0203] 在对第一类置信度和第二类置信度进行融合时,可以采用但不限于等比例融合,从而得到输入图片在每个动作类别上的目标置信度(也就是输入图片在每个动作类别上的最终置信度)。

[0204] 207、根据目标置信度确定输入图片中的人物的动作类别。

[0205] 应理解,上述目标置信度包含输入图片在各个动作类别上的置信度,因此,在根据目标置信度确定输入图片中的人物的动作类别时,可以将目标置信度中最大的置信度对应的动作类别确定为输入图片中的人物的动作类别。

[0206] 进一步地,在根据目标置信度确定输入图片中的人物的动作类别时,还可以先从目标置信度中选择出大于预设阈值的置信度,然后再从该置信度中选择出最大置信度,并将该最大的置信度对应的动作类别确定为输入图片中的人物的动作类别。

[0207] 当目标置信度中不存在大于预设阈值的置信度时,说明在进行动作识别时,没有能够识别出与该输入图片的准确动作类别。

[0208] 应理解,在上述步骤201至步骤207之前,还可以先从本地视频库中选择出训练图片相关度较高的视频放入到视频仓库中。

[0209] 假设,现有的视频库中的视频共对应 P_v 个动作类别,不同动作类别的视频个数不一致。为了避免动作类别的不均衡性,要从每个动作类别选择相同个数(比如 K , K 为大于0的整数)个候选视频,组成 P_v 个视频包(Video Bag),其中,每个视频包里面有 K 个候选视频。现有的训练图片集合共有 P_i 个动作类别,并且这些动作类别与视频的动作类别不完全相同。

[0210] 那么,根据训练图片选择与训练图片相关度较高的视频的具体过程如下:

[0211] (1)、提取视频库中每个视频包中的视频的空间特征和光流特征;

[0212] (2)、提取训练图片集合中所有训练图片的空间特征,并根据视频库中每个视频包中的视频的空间特征和光流特征确定每个训练图片的虚拟光流特征;

[0213] (3)、根据训练图片的空间特征和虚拟光流特征建立本地的图片空间特征库和图片光流特征库;

[0214] (4)、将视频包中的每个视频的空间特征和光流特征分别与图片空间特征库和图片光流特征库进行相似度比较,最后得到视频包中的每个视频在不同动作类别置信度,然后将置信度的最大值作为每个视频与训练图片的相似度度量;

[0215] (5)、在每个视频包中,选择相似度度量最大的前 J ($J < K$, J 和 K 均为大于0的整数,)个视频作为最终入库的视频。

[0216] 与视频相比,图片中的动作缺乏时间上下文关系,进行动作识别时难度较大。本申请提出一种基于虚拟光流和特征库的动作识别方法。通过为单张图片产生与动作密切相关的光流特征,进而结合图片的空间特征和动作特征进行动作识别,能够提高动作识别准确率。

[0217] 另外,本申请利用训练视频和训练图片的空间特征和(虚拟)光流特征建立特征库,并通过输入图片的空间特征和虚拟光流特征与特征库进行比较得到动作类别,进而可以在训练数据较为稀少的情况下,取得较高的动作识别准确率。

[0218] 为了与现有的动作识别方法的性能进行对比,下面结合表3对本申请实施例的动作识别方法的识别效果进行说明。表3示出了本申请实施例的动作识别方法和现有动作识别方法在不同动作识别数据集上的识别准确率。为了凸显本申请适合于训练图片稀少的情况,表2所示的训练集中,每个类别的训练图片仅用1张图片作为训练集。

[0219] 表3

[0220]

动作识别方法	WEB101数据集	VOC数据集	DIFF20数据集
KNN算法	26.1	38.3	55.7
SVM算法	22.3	32.0	54.2
TSN算法	26.1	40.3	56.3
RCNN算法	n/a	28.3	n/a
本申请	35.4	42.2	60.2

[0221] 从表3中可以看出,在不同数据集上,本申请的动作识别方法的识别准确率均高其

它现有方案的识别准确率,因此,本申请实施例的动作识别方法具有较高的识别准确率。

[0222] 上文结合图1至图7对本申请实施例的动作识别方法进行了详细的描述。下文结合图8至图11对本申请实施例的动作识别装置进行描述,应理解,图8至图11所示的动作识别装置具体可以是监控设备、终端设备、网络服务器以及网络云平台等具有图片处理功能的设备。图8至图11所示的动作识别装置可以执行本申请实施例的动作识别方法的各个步骤,为了简洁,下面适当省略重复的描述。

[0223] 图8是本申请实施例的动作识别装置的示意性框图。图8所示的动作识别装置800包括:

[0224] 获取模块801,用于获取待处理图片;

[0225] 提取模块802,用于提取所述待处理图片的空间特征;

[0226] 处理模块803,用于根据所述待处理图片的空间特征以及预设的特征库中的X个空间特征和X个光流特征,确定所述待处理图片的虚拟光流特征,其中,所述X个空间特征和所述X个光流特征存在一一对应关系,X为大于1的整数;

[0227] 所述处理模块803还用于根据所述待处理图片的虚拟光流特征与所述特征库中的Y个光流特征的相似度,确定所述待处理图片在不同动作类别上的第一置信度,其中,所述Y个光流特征中的每个光流特征对应一种动作类别,Y为大于1的整数;

[0228] 所述处理模块803还用于根据所述第一置信度确定所述待处理图片的动作类别。

[0229] 本申请中,通过待处理图片的空间特征以及特征库中的空间特征和光流特征能够获取待处理图片的虚拟光流特征,从而为图片模拟出与动作密切相关的时序信息,接下来就可以根据待处理图片的虚拟光流特征与特征库中的光流特征的相似度对待处理图片进行动作识别。

[0230] 图9是本申请实施例的动作识别的装置的硬件结构示意图。图9所示的动作识别装置900(该动作识别装置900具体可以是一种计算机设备)包括存储器901、处理器902、通信接口903以及总线904。其中,存储器901、处理器902、通信接口903通过总线904实现彼此之间的通信连接。

[0231] 存储器901可以是只读存储器(Read Only Memory,ROM),静态存储设备,动态存储设备或者随机存取存储器(Random Access Memory,RAM)。存储器901可以存储程序,当存储器901中存储的程序被处理器902执行时,处理器902和通信接口903用于执行本申请实施例的动作识别方法的各个步骤。

[0232] 处理器902可以采用通用的中央处理器(Central Processing Unit,CPU),微处理器,应用专用集成电路(Application Specific Integrated Circuit,ASIC),图形处理器(graphics processing unit,GPU)或者一个或多个集成电路,用于执行相关程序,以实现本申请实施例的动作识别的装置中的模块所需执行的功能,或者执行本申请方法实施例的动作识别方法。

[0233] 处理器902还可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,本申请的动作识别方法的各个步骤可以通过处理器902中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器902还可以是通用处理器、数字信号处理器(Digital Signal Processing,DSP)、专用集成电路(ASIC)、现成可编程门阵列(Field Programmable Gate

Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器901,处理器902读取存储器901中的信息,结合其硬件完成本申请实施例的动作识别的装置中包括的模块所需执行的功能,或者执行本申请方法实施例的动作识别方法。

[0234] 通信接口903使用例如但不限于收发器一类的收发装置,来实现装置900与其他设备或通信网络之间的通信。例如,可以通过通信接口903获取待处理图片。

[0235] 总线904可包括在装置900各个部件(例如,存储器901、处理器902、通信接口903)之间传送信息的通路。

[0236] 应注意,尽管图9所示的装置900仅仅示出了存储器901、处理器902、通信接口903,但是在具体实现过程中,本领域的技术人员应当理解,装置900还包括实现正常运行所必须的其他器件。同时,根据具体需要,本领域的技术人员应当理解,装置900还可包括实现其他附加功能的硬件器件。此外,本领域的技术人员应当理解,装置900也可仅仅包括实现本申请实施例所必须的器件,而不必包括图9中所示的全部器件。

[0237] 应理解,动作识别装置800中的获取模块801相当于动作识别装置900中的通信接口903,提取模块802和处理模块803相当于处理器902。

[0238] 图10是本申请实施例的动作识别装置的示意性框图。图10所示的动作识别装置1000包括:CNN模块1001、虚拟光流模块1002、空间预测模块1003、光流预测模块1004、融合输出模块1005、视频仓库1006、空间特征仓库1007和光流特征仓库1008。

[0239] 其中,视频仓库1006、空间特征仓库1007和光流特征仓库1008既可以位于动作识别装置1000的内部,也可以位于动作识别装置1000的外部,例如,视频仓库1006、空间特征仓库1007和光流特征仓库1008可以位于服务器中或者本地的数据库中,动作识别装置可以从服务器或者本地的数据库中调取视频仓库1006、空间特征仓库1007和光流特征仓库1008中包含的特征。

[0240] 应理解,动作识别装置1000可以执行本申请实施例的动作识别方法的各个步骤。各个模块的具体作用如下:

[0241] CNN模块1001,用于对待处理图片进行卷积运算,得到图像的空间特征 u^{rgb} 。

[0242] 视频仓库1006,用于存储训练视频的空间特征 V^{rgb} 和光流特征 V^{flow} 。

[0243] 虚拟光流模块1002,用于利用待处理图片的空间特征 u^{rgb} 以及视频仓库1006中存储的训练视频的空间特征 V^{rgb} 和光流特征 V^{flow} ,产生待处理图片的虚拟光流特征 u^{flow} 。

[0244] 空间特征仓库1007,用于存储训练视频和图像的空间特征以及类标签。

[0245] 光流特征仓库1008,用于存储训练视频的光流特征、训练图片的虚拟光流特征以及它们的动作类别标签。

[0246] 空间预测模块1003,用于把待处理图片的空间特征和空间特征仓库中的特征进行比较,得到待处理图片在各个动作类别上的置信度。

[0247] 光流预测模块1004,用于把待处理图片的虚拟光流特征和光流特征仓库中的特征

进行比较,得到待处理图片在各个类别上的置信度。

[0248] 融合输出模块1005,用于把空间预测模块1003和光流预测模块1004得到的待处理图片在各个动作类别的置信度进行融合,得到最终每个类别的置信度,选取置信度最大的动作类别输出。

[0249] 可选地,图10所示的动作识别装置1000还可以包括获取模块1009和视频选择模块1010,其中,获取模块1009用于获取待处理图片,视频选择模块1010用于为视频仓库1006选择视频。

[0250] 应理解,动作识别装置1000中的获取模块1009相当于动作识别装置900中的通信接口903,动作识别装置1000中的其它模块相当于动作识别装置900中的处理器902。

[0251] 上述动作识别装置1000中的CNN模块1001、虚拟光流模块1002、空间预测模块1003、光流预测模块1004、融合输出模块1005、获取模块1009以及视频选择模块1010在具体实现时可以采用硬件或者软件,或者硬件和软件相结合的方式来实现。

[0252] 为了更好地理解动作识别装置1000中各个模块的工作流程,下面结合图11对动作识别装置1000进行动作识别的过程进行大致的描述(详细过程可以参见图2所示的方法中的各个步骤,这里不再详细描述)。

[0253] 如图11所示,动作识别装置1000对输入图片进行动作识别的过程具体如下:

[0254] CNN模块1001得到输入图片,提取该输入图片的空间特征;

[0255] 虚拟光流模块1002根据视频仓库1006中的光流特征和空间特征以及输入图片的空间特征确定输入图片的虚拟光流特征;

[0256] 接下来,光流预测模块1004根据提取到的输入图片的虚拟光流特征以及光流特征仓库1008中的光流特征进行光流预测,得到第一类置信度;

[0257] 空间预测模块1003根据提取到的输入图片的空间特征以及空间特征仓库1007中的空间特征进行空间预测,得到第二类置信度;

[0258] 融合输出模块1005对第一类置信度和第二类置信度进行融合,得到目标置信度,然后根据目标置信度确定输入图片中的人物的动作类别。

[0259] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0260] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0261] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0262] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显

示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0263] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0264] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0265] 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所述权利要求的保护范围为准。

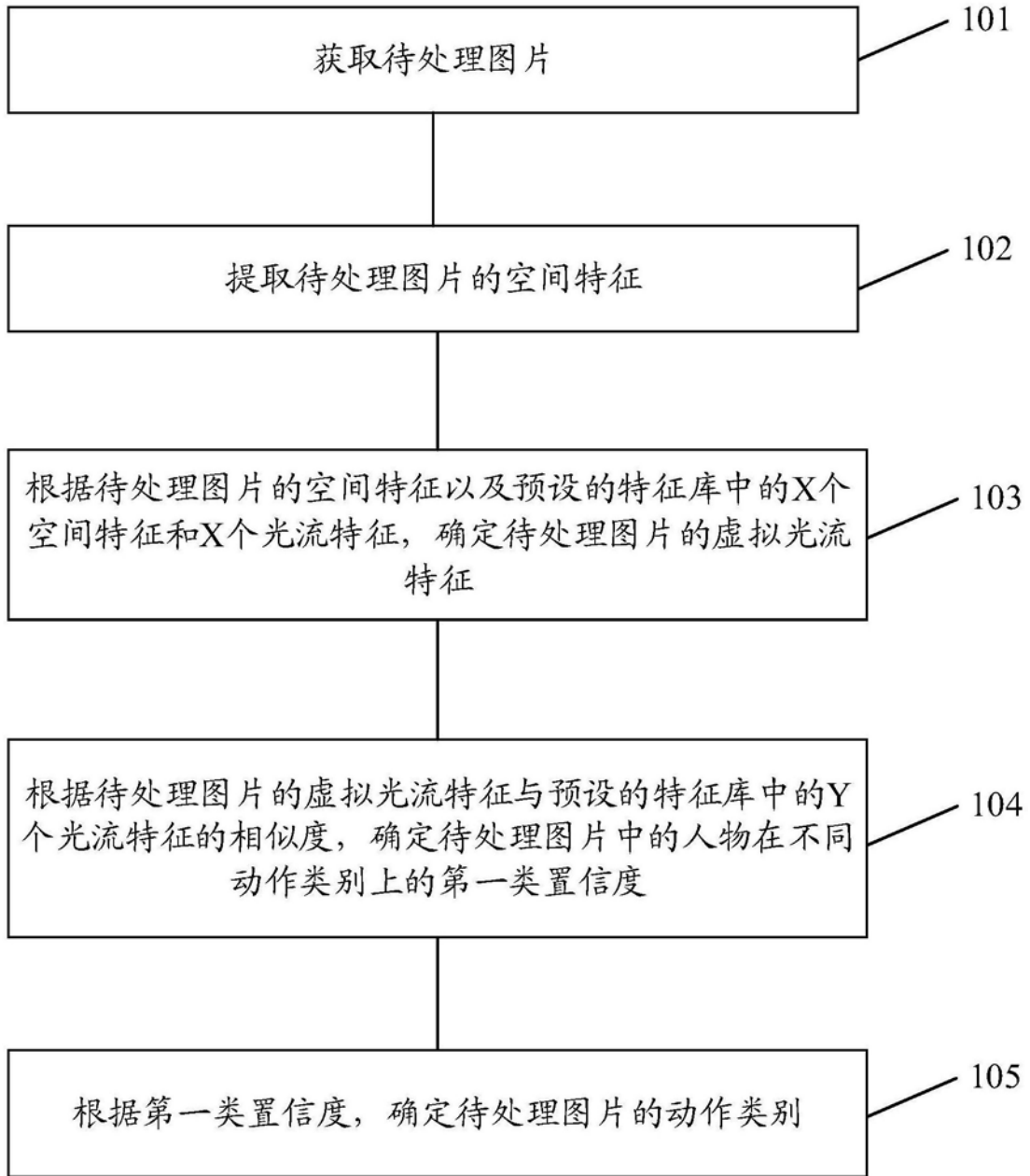


图1



图2

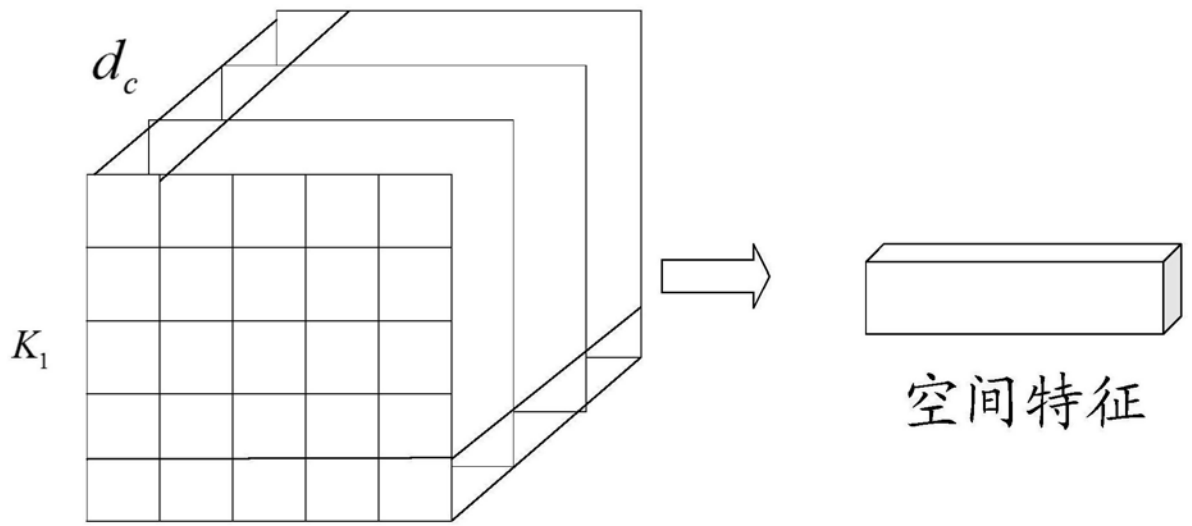


图3

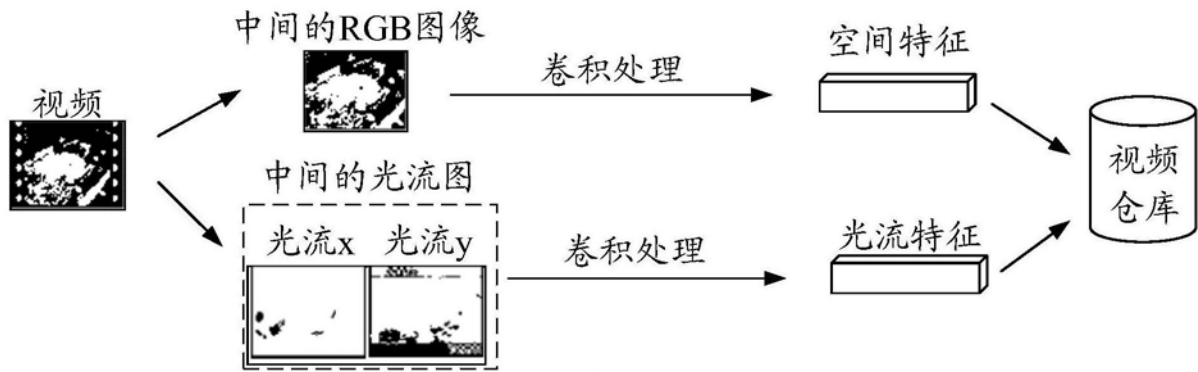


图4

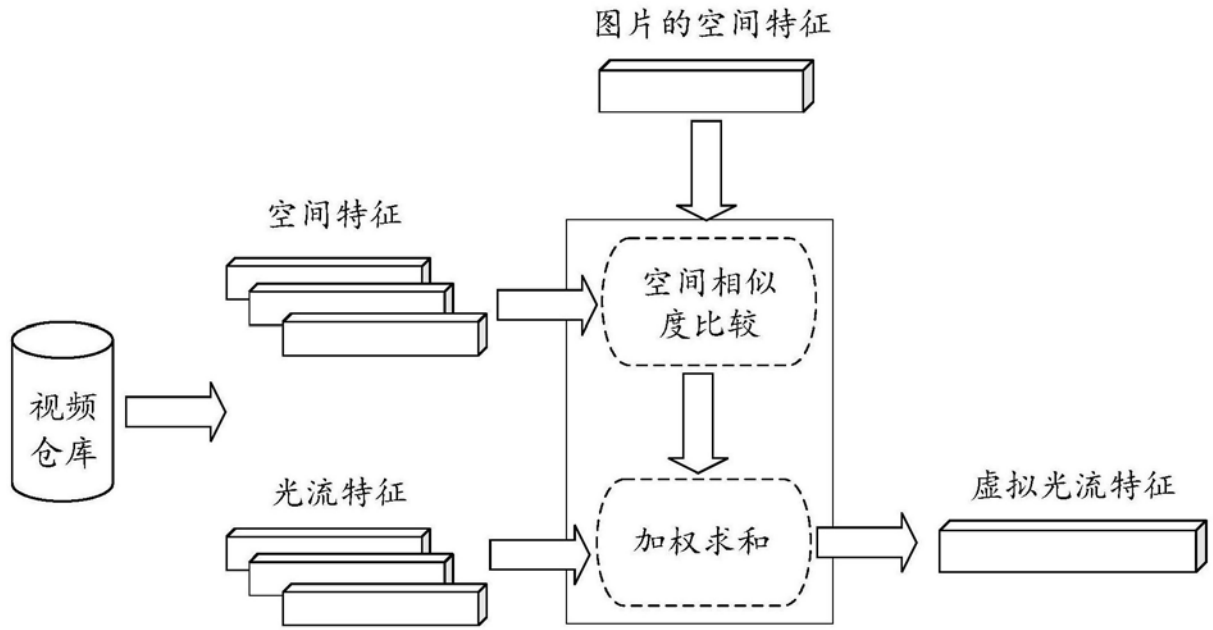


图5

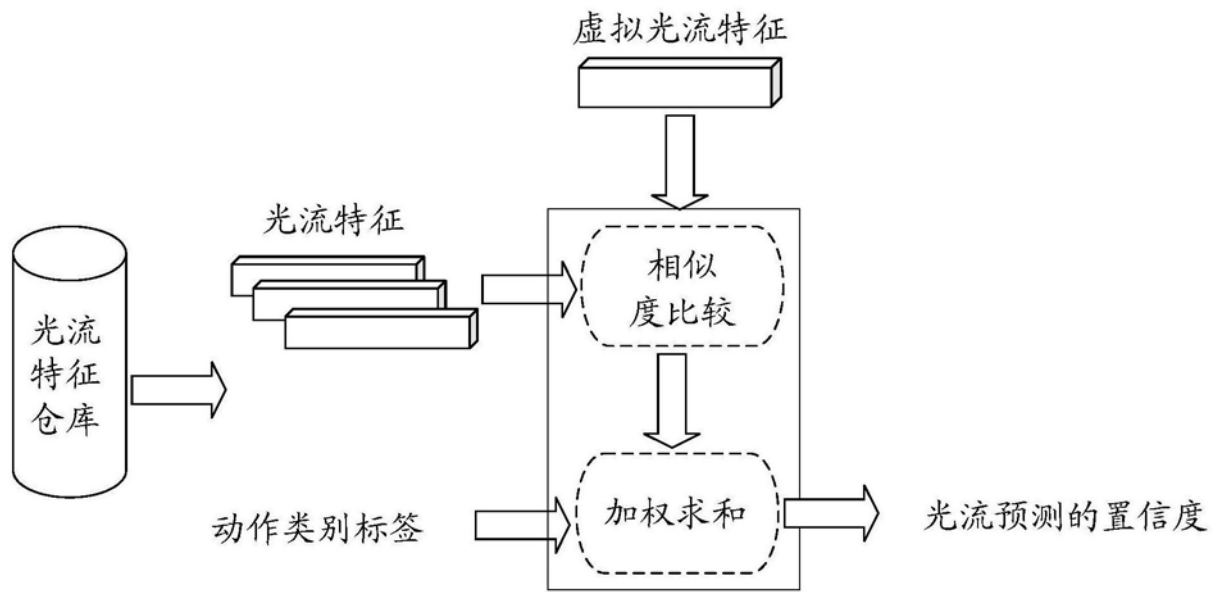


图6

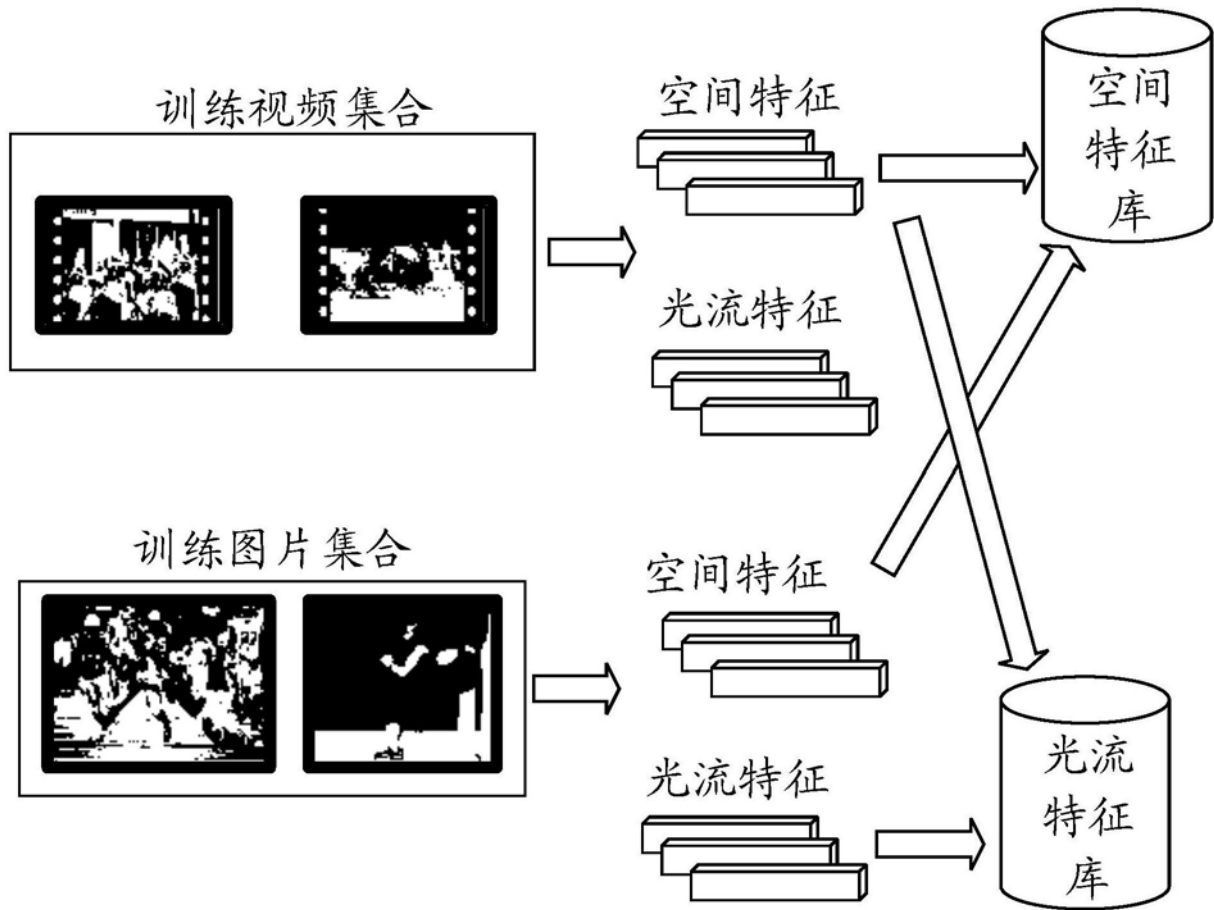


图7

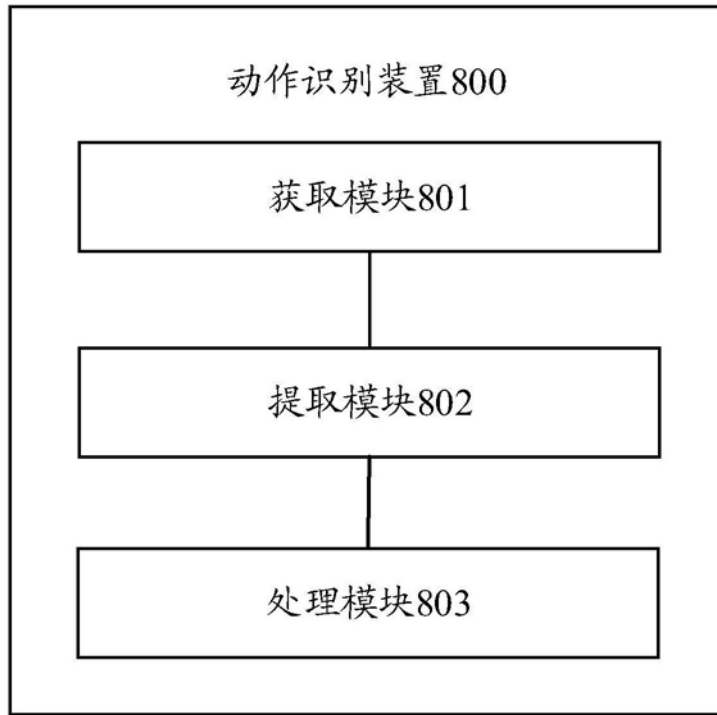


图8

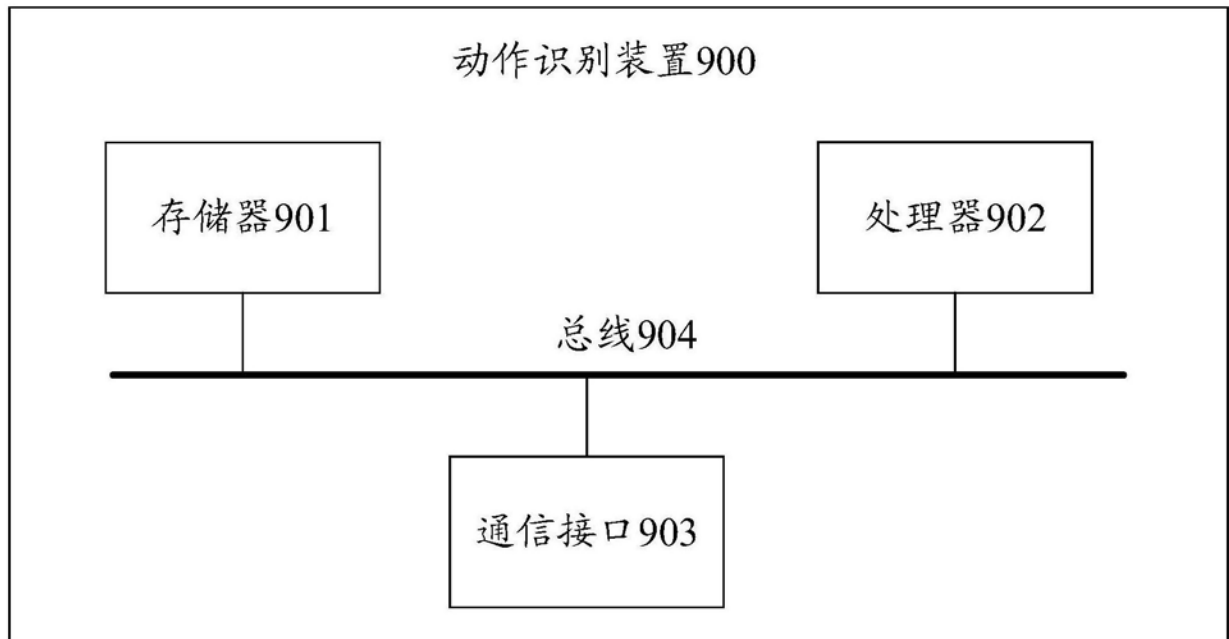


图9

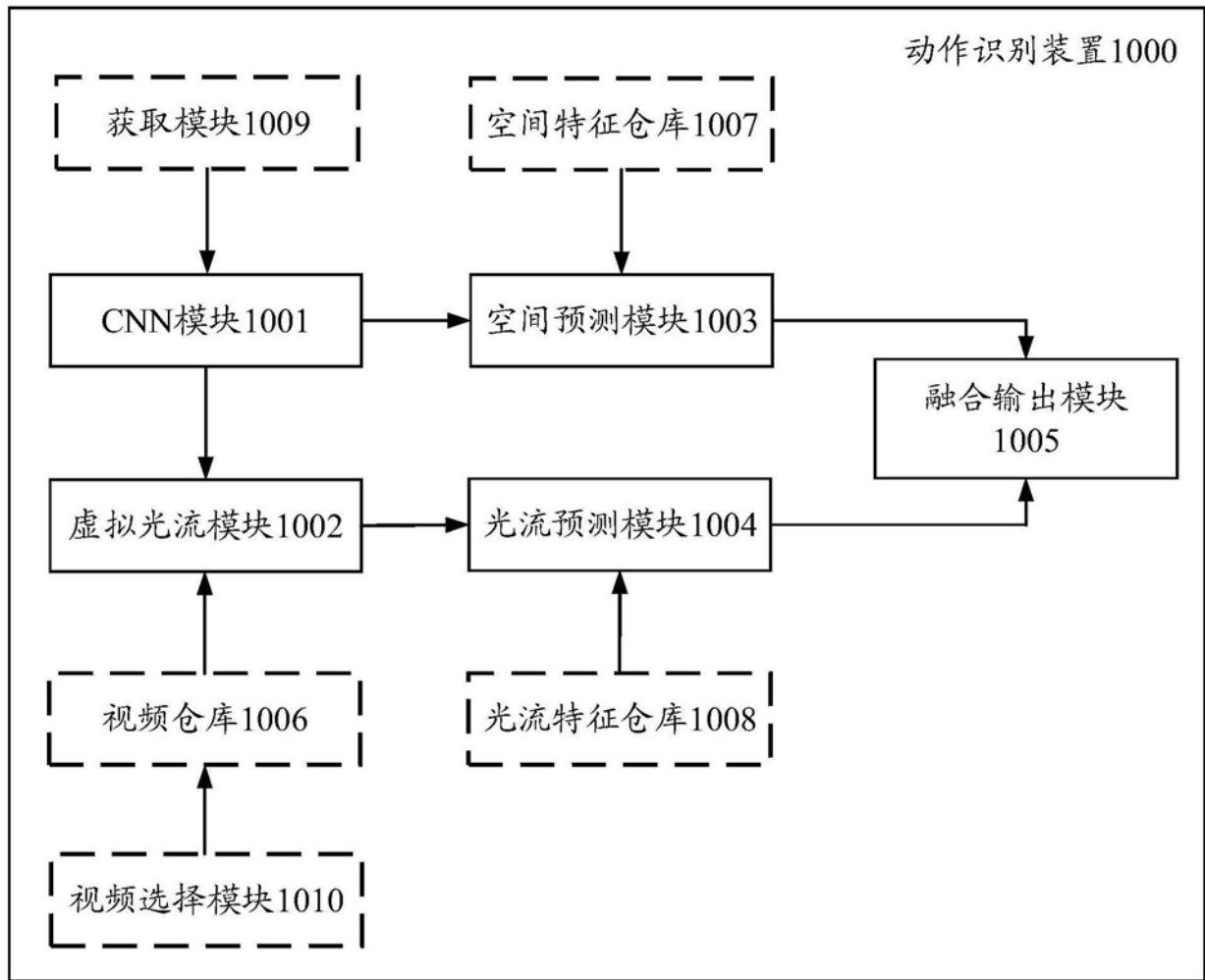


图10

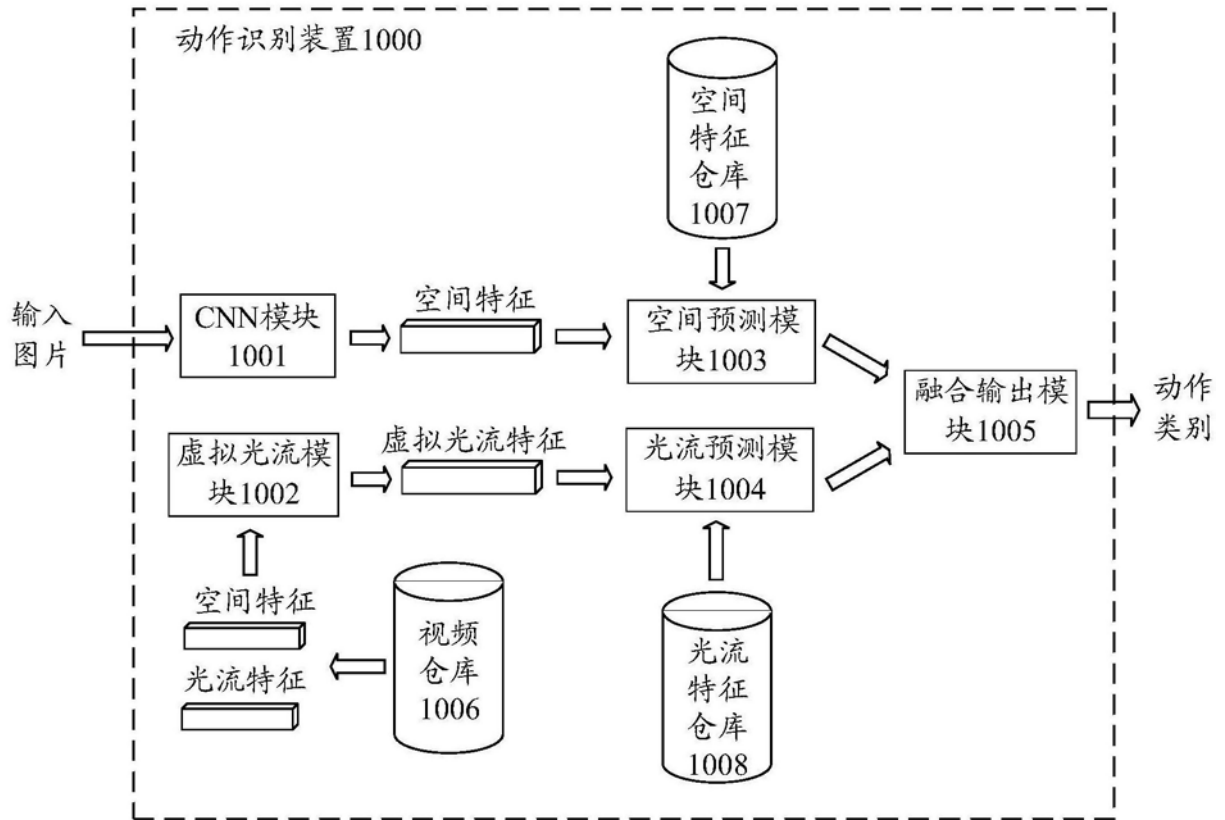


图11