



(12) 发明专利

(10) 授权公告号 CN 112749613 B

(45) 授权公告日 2024.03.26

(21) 申请号 202010880221.8

CN 109671097 A, 2019.04.23

(22) 申请日 2020.08.27

CN 110786002 A, 2020.02.11

(65) 同一申请的已公布的文献号

CN 111464716 A, 2020.07.28

申请公布号 CN 112749613 A

CN 111556336 A, 2020.08.18

(43) 申请公布日 2021.05.04

DE 102019121570 A1, 2020.02.13

(73) 专利权人 腾讯科技(深圳)有限公司

US 2017357860 A1, 2017.12.14

地址 518057 广东省深圳市南山区高新区

US 2019089923 A1, 2019.03.21

科技中一路腾讯大厦35层

US 2019130585 A1, 2019.05.02

(72) 发明人 余自强

US 5566251 A, 1996.10.15

(74) 专利代理机构 北京三高永信知识产权代理

WO 2009107487 A1, 2009.09.03

有限责任公司 11138

徐知劼,等.针对部分景观运动模糊的还原算法.电子设计工程.2019,第27卷(第24期),全文.

专利代理师 张所明

Jiaya Jia, et al. Single Image Motion Deblurring Using Transparency. 2007 IEEE Conference on Computer Vision and Pattern Recognition. 2007, 全文.

(51) Int. Cl.

G06V 20/40 (2022.01)

H04N 5/265 (2006.01)

(56) 对比文件

审查员 于海侠

CN 102073652 A, 2011.05.25

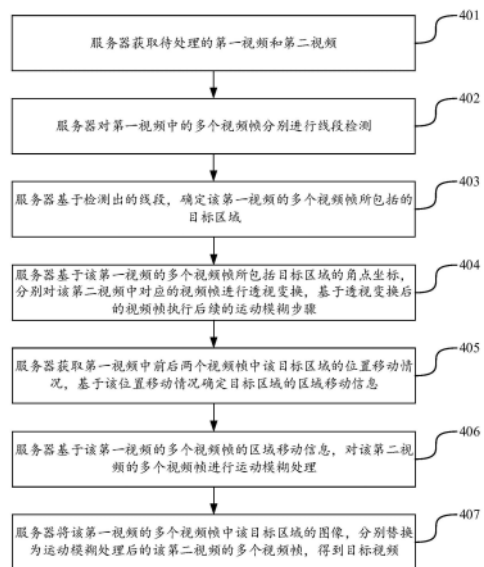
权利要求书4页 说明书18页 附图12页

(54) 发明名称

视频数据处理方法、装置、计算机设备及存储介质

(57) 摘要

本申请公开了一种视频数据处理方法、装置、计算机设备及存储介质,属于视频处理技术领域。本申请通过在将第二视频嵌入到第一视频的目标区域时,基于人工智能技术来获取第一视频中目标区域的位置移动情况,根据该位置移动情况相应地对第二视频的画面进行运动模糊处理,使第二视频和第一视频具有相同的运动模糊效果,从而使第一视频和第二视频能够更好的融合,得到更好的视频过渡效果。



1. 一种视频数据处理方法,其特征在于,所述方法包括:

获取待处理的第一视频和第二视频;

确定所述第一视频的多个视频帧的区域移动信息,所述区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况,所述目标区域用于显示所述第二视频的视频帧,所述区域移动信息包括所述目标区域的至少三个角点的移动距离和移动方向;

基于前后两个视频帧中所述目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数,所述卷积参数用于指示一次卷积运算中的卷积区域大小以及卷积核的移动方向;

通过第三卷积核分别基于所述至少三组卷积参数对所述第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧;

对所述至少三个中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧;

将所述第一视频的多个视频帧中所述目标区域的图像,分别替换为运动模糊处理后的所述第二视频的多个视频帧,得到目标视频。

2. 根据权利要求1所述的方法,其特征在于,所述确定所述第一视频的多个视频帧的区域移动信息,包括:

确定所述第一视频的多个视频帧中的目标区域;

基于所述第一视频中前后两个视频帧中所述目标区域的位置移动情况,确定所述区域移动信息。

3. 根据权利要求2所述的方法,其特征在于,所述确定所述第一视频的多个视频帧中的目标区域,包括:

对所述第一视频中的多个视频帧分别进行线段检测;

基于检测出的线段,确定所述第一视频的多个视频帧所包括的目标区域。

4. 根据权利要求3所述的方法,其特征在于,所述对所述第一视频中的多个视频帧分别进行线段检测,包括:

对所述第一视频中的多个视频帧分别进行特征提取,得到视频帧对应的特征图;

基于所述视频帧对应的特征图,确定所述视频帧所包括的连接点;

基于所述连接点确定所述视频帧所包括的线段。

5. 根据权利要求4所述的方法,其特征在于,所述基于所述视频帧对应的特征图,确定所述视频帧所包括的连接点,包括:

将所述特征图划分为多个子区域;

基于第一卷积核对所述特征图中的子区域进行卷积运算,得到概率特征图,所述概率特征图中的一个元素用于指示一个子区域中存在连接点的概率;

基于第二卷积核对所述概率特征图进行卷积运算,得到位置特征图,所述位置特征图中的一个元素用于指示一个子区域中连接点的位置信息;

基于所述概率特征图以及所述位置特征图,确定所述视频帧所包括的连接点。

6. 根据权利要求3所述的方法,其特征在于,所述基于检测出的线段,确定所述第一视频的多个视频帧所包括的目标区域,包括:

对于所述第一视频中的第一帧,从检测出的所述线段中,获取与参考区域的边框之间距离最近的线段作为目标线段,将所述目标线段围成的区域确定为所述目标区域;

对于所述第一视频中除第一帧以外的任一帧,从检测出的所述线段中,获取与前一帧中的目标线段之间距离最近的线段,作为所述任一帧中的目标线段,将所述任一帧中的目标线段围成的区域确定为所述目标区域。

7. 根据权利要求1所述的方法,其特征在于,所述通过第三卷积核分别基于所述至少三组卷积参数对所述第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧,包括:

对于任一组卷积参数,从多个不同大小的候选卷积核中,获取尺寸与所述任一组卷积参数中所述卷积区域大小相匹配的候选卷积核,作为第三卷积核;

通过所述第三卷积核按照所述任一组卷积参数中的移动方向,对所述第二视频中对应的视频帧进行卷积运算,得到中间视频帧。

8. 根据权利要求1所述的方法,其特征在于,所述对所述至少三个中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧,包括:

对于所述第二视频的视频帧中的任一像素点,基于所述任一像素点相对于各个角点的距离,确定在任一像素点位置,各个所述中间帧对应的透明度参数;

对于所述第二视频的视频帧中的任一像素点位置,基于各个所述中间帧对应的透明度参数,对所述中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧。

9. 根据权利要求1所述的方法,其特征在于,所述基于前后两个视频帧中所述目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数之前,所述方法还包括:

基于所述第一视频的多个视频帧所包括目标区域的角点坐标,分别对所述第二视频中对应的视频帧进行透视变换,基于透视变换后的所述第二视频的视频帧,执行以下步骤:

基于前后两个视频帧中所述目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数;通过第三卷积核分别基于所述至少三组卷积参数对所述第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧;对所述至少三个中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧。

10. 根据权利要求9所述的方法,其特征在于,所述基于所述第一视频的多个视频帧所包括目标区域的角点坐标,分别对所述第二视频中对应的视频帧进行透视变换,包括:

基于所述第一视频的多个视频帧所包括目标区域的角点坐标之间的相对位置,确定所述第一视频中各个视频帧对应的透视变换矩阵;

基于所述第一视频中各个视频帧对应的透视变换矩阵,分别对所述第二视频中对应的视频帧进行透视变换。

11. 一种视频数据处理装置,其特征在于,所述装置包括:

获取模块,用于获取待处理的第一视频和第二视频;

确定模块,用于确定所述第一视频的多个视频帧的区域移动信息,所述区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况,所述目标区域用于显示所述第二视频的视频帧,所述区域移动信息包括所述目标区域的至少三个角点的移动距离和移动方向;

运动模糊处理模块包括参数确定单元、卷积单元和叠加单元;

所述参数确定单元,用于基于前后两个视频帧中所述目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数,所述卷积参数用于指示一次卷积运算中的

卷积区域大小以及卷积核的移动方向；

所述卷积单元,用于通过第三卷积核分别基于所述至少三组卷积参数对所述第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧；

所述叠加单元,用于对所述至少三个中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧；

替换模块,用于将所述第一视频的多个视频帧中所述目标区域的图像,分别替换为运动模糊处理后的所述第二视频的多个视频帧,得到目标视频。

12. 根据权利要求11所述的装置,其特征在于,所述确定模块包括:

第一确定单元,用于确定所述第一视频的多个视频帧中的目标区域；

第二确定单元,用于基于所述第一视频中前后两个视频帧中所述目标区域的位置移动情况,确定所述区域移动信息。

13. 根据权利要求12所述的装置,其特征在于,所述第一确定单元,包括:

检测子单元,用于对所述第一视频中的多个视频帧分别进行线段检测；

区域确定子单元,用于基于检测出的线段,确定所述第一视频的多个视频帧所包括的目标区域。

14. 根据权利要求13所述的装置,其特征在于,所述检测子单元包括:

特征提取子单元,用于对所述第一视频中的多个视频帧分别进行特征提取,得到视频帧对应的特征图；

连接点确定子单元,用于基于所述视频帧对应的特征图,确定所述视频帧所包括的连接点；

线段确定子单元,用于基于所述连接点确定所述视频帧所包括的线段。

15. 根据权利要求14所述的装置,其特征在于,所述连接点确定子单元用于:

将所述特征图划分为多个子区域；

基于第一卷积核对所述特征图中的子区域进行卷积运算,得到概率特征图,所述概率特征图中的一个元素用于指示一个子区域中存在连接点的概率；

基于第二卷积核对所述概率特征图进行卷积运算,得到位置特征图,所述位置特征图中的一个元素用于指示一个子区域中连接点的位置信息；

基于所述概率特征图以及所述位置特征图,确定所述视频帧所包括的连接点。

16. 根据权利要求13所述的装置,其特征在于,所述区域确定子单元用于:

对于所述第一视频中的第一帧,从检测出的所述线段中,获取与参考区域的边框之间距离最近的线段作为目标线段,将所述目标线段围成的区域确定为所述目标区域；

对于所述第一视频中除第一帧以外的任一帧,从检测出的所述线段中,获取与前一帧中的目标线段之间距离最近的线段,作为所述任一帧中的目标线段,将所述任一帧中的目标线段围成的区域确定为所述目标区域。

17. 根据权利要求11所述的装置,其特征在于,所述卷积单元用于:

对于任一组卷积参数,从多个不同大小的候选卷积核中,获取尺寸与所述任一组卷积参数中所述卷积区域大小相匹配的候选卷积核,作为第三卷积核；

通过所述第三卷积核按照所述任一组卷积参数中的移动方向,对所述第二视频中对应的视频帧进行卷积运算,得到中间视频帧。

18. 根据权利要求11所述的装置,其特征在於,所述叠加单元用于:

对于所述第二视频的视频帧中的任一像素点,基于所述任一像素点相对于各个角点的距离,确定在任一像素点位置,各个所述中间帧对应的透明度参数;

对于所述第二视频的视频帧中的任一像素点位置,基于各个所述中间帧对应的透明度参数,对所述中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧。

19. 根据权利要求11所述的装置,其特征在於,所述装置还包括:

透视变换模块,用于基于所述第一视频的多个视频帧所包括目标区域的角点坐标,分别对所述第二视频中对应的视频帧进行透视变换,基于透视变换后的所述第二视频的视频帧,执行以下步骤:

基于前后两个视频帧中所述目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数;通过第三卷积核分别基于所述至少三组卷积参数对所述第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧;对所述至少三个中间帧进行叠加,得到运动模糊处理后的所述第二视频的视频帧。

20. 根据权利要求19所述的装置,其特征在於,所述透视变换模块用于:

基于所述第一视频的多个视频帧所包括目标区域的角点坐标之间的相对位置,确定所述第一视频中各个视频帧对应的透视变换矩阵;

基于所述第一视频中各个视频帧对应的透视变换矩阵,分别对所述第二视频中对应的视频帧进行透视变换。

21. 一种计算机设备,其特征在於,所述计算机设备包括一个或多个处理器和一个或多个存储器,所述一个或多个存储器中存储有至少一条程序代码,所述至少一条程序代码由所述一个或多个处理器加载并执行以实现如权利要求1至权利要求10任一项所述的视频数据处理方法所执行的操作。

22. 一种计算机可读存储介质,其特征在於,所述计算机可读存储介质中存储有至少一条程序代码,所述至少一条程序代码由处理器加载并执行以实现如权利要求1至权利要求10任一项所述的视频数据处理方法所执行的操作。

23. 一种计算机程序产品,其特征在於,所述计算机程序产品包括至少一条程序代码,所述至少一条程序代码存储在计算机可读存储介质中,由计算机设备的处理器从所述计算机可读存储介质中读取并执行以实现如权利要求1至权利要求10任一项所述的视频数据处理方法所执行的操作。

视频数据处理方法、装置、计算机设备及存储介质

技术领域

[0001] 本申请涉及视频处理技术领域,特别涉及一种视频数据处理方法、装置、计算机设备及存储介质。

背景技术

[0002] 视频转场特效是指在对两个视频片段进行合并时添加特效,使两个视频之间能够自然过渡。例如,相框转场就是一种常用的视频转场特效,相框转场是通过对第一视频的画面中所出现的相框区域进行识别,将相框区域中的画面替换为第二视频的画面,从而实现从第一视频到第二视频的过渡。

[0003] 在上述视频数据处理过程中,当第一视频的画面出现运动模糊的情况时,直接将相框区域中的画面替换为第二视频的画面,会导致第一视频画面和第二视频画面的融合效果较差,严重影响视频过渡效果。

发明内容

[0004] 本申请实施例提供了一种视频数据处理方法、装置、计算机设备及存储介质,可以优化两个视频之间的过渡效果。该技术方案如下:

[0005] 一方面,提供了一种视频数据处理方法,该方法包括:

[0006] 获取待处理的第一视频和第二视频;

[0007] 确定该第一视频的多个视频帧的区域移动信息,该区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况,该目标区域用于显示该第二视频的视频帧;

[0008] 基于该第一视频的多个视频帧的区域移动信息,对该第二视频的多个视频帧进行运动模糊处理;

[0009] 将该第一视频的多个视频帧中该目标区域的图像,分别替换为运动模糊处理后的该第二视频的多个视频帧,得到目标视频。

[0010] 一方面,提供了一种视频数据处理装置,该装置包括:

[0011] 获取模块,用于获取待处理的第一视频和第二视频;

[0012] 确定模块,用于确定该第一视频的多个视频帧的区域移动信息,该区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况,该目标区域用于显示该第二视频的视频帧;

[0013] 运动模糊处理模块,用于基于该第一视频的多个视频帧的区域移动信息,对该第二视频的多个视频帧进行运动模糊处理;

[0014] 替换模块,用于将该第一视频的多个视频帧中该目标区域的图像,分别替换为运动模糊处理后的该第二视频的多个视频帧,得到目标视频。

[0015] 在一种可能实现方式中,该确定模块包括:

[0016] 第一确定单元,用于确定该第一视频的多个视频帧中的目标区域;

[0017] 第二确定单元,用于基于该第一视频中前后两个视频帧中该目标区域的位置移动

情况,确定该区域移动信息。

[0018] 在一种可能实现方式中,该第一确定单元包括:

[0019] 检测子单元,用于对该第一视频中的多个视频帧分别进行线段检测;

[0020] 区域确定子单元,用于基于检测出的线段,确定该第一视频的多个视频帧所包括的目标区域。

[0021] 在一种可能实现方式中,该检测子单元包括:

[0022] 特征提取子单元,用于对该第一视频中的多个视频帧分别进行特征提取,得到视频帧对应的特征图;

[0023] 连接点确定子单元,用于基于该视频帧对应的特征图,确定该视频帧所包括的连接点;

[0024] 线段确定子单元,用于基于该连接点确定该视频帧所包括的线段。

[0025] 在一种可能实现方式中,该连接点确定子单元用于:

[0026] 将该特征图划分为多个子区域;

[0027] 基于第一卷积核对该特征图中的子区域进行卷积运算,得到概率特征图,该概率特征图中的一个元素用于指示一个子区域中存在连接点的概率;

[0028] 基于第二卷积核对该概率特征图进行卷积运算,得到位置特征图,该位置特征图中的一个元素用于指示一个子区域中连接点的位置信息;

[0029] 基于该概率特征图以及该位置特征图,确定该视频帧所包括的连接点。

[0030] 在一种可能实现方式中,该区域确定子单元用于:

[0031] 对于该第一视频中的第一帧,从检测出的该线段中,获取与参考区域的边框之间距离最近的线段作为目标线段,将该目标线段围成的区域确定为该目标区域;

[0032] 对于该第一视频中除第一帧以外的任一帧,从检测出的该线段中,获取与前一帧中的目标线段之间距离最近的线段,作为该任一帧中的目标线段,将该任一帧中的目标线段围成的区域确定为该目标区域。

[0033] 在一种可能实现方式中,该区域移动信息包括该目标区域的至少三个角点的移动距离和移动方向。

[0034] 在一种可能实现方式中,该运动模糊处理模块包括:

[0035] 参数确定单元,用于基于前后两个视频帧中目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数,该卷积参数用于指示第三卷积核在一次卷积运算中的卷积区域大小以及该第三卷积核的移动方向;

[0036] 卷积单元,用于通过该第三卷积核分别基于该至少三组卷积参数对该第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧;

[0037] 叠加单元,用于对该至少三个中间帧进行叠加,得到运动模糊处理后的该第二视频的视频帧。

[0038] 在一种可能实现方式中,该卷积单元用于:

[0039] 对于任一组卷积参数,从多个不同大小的候选卷积核中,获取尺寸与该任一组卷积参数中该卷积区域大小相匹配的候选卷积核,作为第三卷积核;

[0040] 通过该第三卷积核按照该任一组卷积参数中的移动方向,对该第二视频中对应的视频帧进行卷积运算,得到中间视频帧。

[0041] 在一种可能实现方式中,该叠加单元用于:

[0042] 对于该第二视频的视频帧中的任一像素点,基于该任一像素点相对于各个角点的距离,确定在任一像素点位置,各个该中间帧对应的透明度参数;

[0043] 对于该第二视频的视频帧中的任一像素点位置,基于各个该中间帧对应的透明度参数,对该中间帧进行叠加,得到运动模糊处理后的该第二视频的视频帧。

[0044] 在一种可能实现方式中,该装置还包括:

[0045] 透视变换模块,用于基于该第一视频的多个视频帧所包括目标区域的角点坐标,分别对该第二视频中对应的视频帧进行透视变换,基于透视变换后的该第二视频的视频帧,执行基于该第一视频的多个视频帧的区域移动信息,对该第二视频的多个视频帧进行运动模糊处理的步骤。

[0046] 在一种可能实现方式中,该透视变换模块用于:

[0047] 基于该第一视频的多个视频帧所包括目标区域的角点坐标之间的相对位置,确定该第一视频中各个视频帧对应的透视变换矩阵;

[0048] 基于该第一视频中各个视频帧对应的透视变换矩阵,分别对该第二视频中对应的视频帧进行透视变换。

[0049] 一方面,提供了一种计算机设备,该计算机设备包括一个或多个处理器和一个或多个存储器,该一个或多个存储器中存储有至少一条程序代码,该至少一条程序代码由该一个或多个处理器加载并执行以实现该视频数据处理方法所执行的操作。

[0050] 一方面,提供了一种计算机可读存储介质,该计算机可读存储介质中存储有至少一条程序代码,该至少一条程序代码由处理器加载并执行以实现该视频数据处理方法所执行的操作。

[0051] 一方面,提供了一种计算机程序产品,该计算机程序产品包括至少一条程序代码,该至少一条程序代码存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该至少一条程序代码,处理器执行该计算机指令,使得该计算机设备实现该视频数据处理方法所执行的操作。

[0052] 本申请实施例提供的技术方案,通过在将第二视频嵌入到第一视频的目标区域时,基于人工智能技术来获取第一视频中目标区域的位置移动情况,根据该位置移动情况相应地对第二视频的画面进行运动模糊处理,使第二视频和第一视频具有相同的运动模糊效果,从而使第一视频和第二视频能够更好的融合,得到更好的视频过渡效果。

附图说明

[0053] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述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他的附图。

[0054] 图1是本申请实施例提供的一种视频数据处理方法的实施环境示意图;

[0055] 图2是本申请实施例提供的一种视频数据处理方法的流程图;

[0056] 图3是本申请实施例提供的一种第一视频的视频帧画面示意图;

[0057] 图4是本申请实施例提供的一种视频数据处理方法的流程图;

- [0058] 图5是本申请实施例提供的一种线段检测模型的示意图；
- [0059] 图6是本申请实施例提供的一种参考区域标注的示意图；
- [0060] 图7是本申请实施例提供的一种目标区域确定方法示意图；
- [0061] 图8是本申请实施例提供的一种运动模糊效果的视频帧示意图；
- [0062] 图9是本申请实施例提供的一种目标区域移动情况的示意图；
- [0063] 图10是本申请实施例提供的一种运动模糊处理的示意图；
- [0064] 图11是本申请实施例提供的一种运动模糊效果示意图；
- [0065] 图12是本申请实施例提供的一种目标视频的视频帧示意图；
- [0066] 图13是本申请实施例提供的一种相框转场效果的实现方法流程图；
- [0067] 图14是本申请实施例提供的一种目标区域跟踪过程示意图；
- [0068] 图15是本申请实施例提供的一种视频数据处理装置的结构示意图；
- [0069] 图16是本申请实施例提供的一种终端的结构示意图；
- [0070] 图17是本申请实施例提供的一种服务器的结构示意图。

具体实施方式

[0071] 为使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请实施方式作进一步地详细描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做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。

[0072] 本申请中术语“第一”“第二”等字样用于对作用和功能基本相同的相同项或相似项进行区分,应理解,“第一”、“第二”、“第n”之间不具有逻辑或时序上的依赖关系,也不对数量和执行顺序进行限定。

[0073] 人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。本申请实施例涉及人工智能技术中的计算机视觉技术。

[0074] 计算机视觉技术(Computer Vision, CV)计算机视觉是一门研究如何使机器“看”的科学,更进一步的说,就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉,并进一步做图形处理,使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科,计算机视觉研究相关的理论和技术,试图建立能够从图像或者多维数据中获取信息的人工智能系统。计算机视觉技术通常包括图像处理、图像识别、图像语义理解、图像检索、OCR、视频处理、视频语义理解、视频内容/行为识别、三维物体重建、3D技术、虚拟现实、增强现实、同步定位与地图构建等技术,还包括常见的人脸识别、指纹识别等生

物特征识别技术。本申请涉及计算机视觉技术中的视频语义理解技术,例如,通过对视频帧进行线段检测,对视频帧中的目标区域进行跟踪,来准确确定出各个视频帧中的相框的位置。

[0075] 图1是本申请实施例提供的一种视频数据处理方法的实施环境示意图,参见图1,该实施环境包括终端110和视频编辑平台140。

[0076] 终端110可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等,但并不局限于此。终端110安装和运行有支持视频编辑的目标应用程序等。示例性的,终端110是用户使用的终端,终端110中运行的应用程序内登录有用户账号。终端110可以泛指多个终端中的一个,本实施例仅以终端110来举例说明。

[0077] 视频编辑平台140用于为上述目标应用程序提供后台服务。可选地,视频编辑平台140承担主要视频数据处理工作,终端110承担次要视频数据处理工作;或者,视频编辑平台140承担次要视频数据处理工作,终端110承担主要视频数据处理工作;或者,视频编辑平台140或终端110分别可以单独承担视频数据处理工作。可选地,视频编辑平台140包括:接入服务器、视频数据处理服务器和数据库。接入服务器用于为终端110提供接入服务。视频数据处理服务器用于提供视频合成、视频特效添加等视频编辑相关的后台服务。视频数据处理服务器可以是一台或多台。当视频数据处理服务器是多台时,存在至少两台视频数据处理服务器用于提供不同的服务,和/或,存在至少两台视频数据处理服务器用于提供相同的服务,比如以负载均衡方式提供同一种服务,本申请实施例对此不加以限定。视频数据处理服务器中可以设置有线段检测模型,用于对视频帧中的目标区域进行识别、跟踪。其中,上述服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN(Content Delivery Network,内容分发网络)、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0078] 上述终端110与视频编辑平台140可以通过有线或无线通信方式进行直接或间接地连接,本申请实施例对此不作限定。

[0079] 本领域技术人员可以知晓,上述终端的数量可以更多或更少。比如上述终端可以仅为一个,或者上述终端为几十个或几百个,或者更多数量。本申请实施例对终端的数量和设备类型不加以限定。

[0080] 本申请实施例提供的视频数据处理方法,应用于多种类型的应用程序中。例如,在视频编辑类应用程序中,为两个视频添加相框转场特效时,即将第二视频嵌入到第一视频中相框所框出的区域进行显示时,应用基于神经网络构建的线段检测模型,来识别第一视频的视频帧中的线段,从而实现由这些线段所围成的目标区域,即相框所框出的区域进行跟踪,应用这种目标区域跟踪方法,无需用户手动调整每一视频帧中的跟踪点,提高目标区域跟踪的准确度;在视频合成阶段,基于第一视频中相框的位置移动情况,对第二视频的视频帧进行运动模糊处理,使第二视频能够具有与第一视频相同的运动模糊效果,使合成后的视频更加逼真,过渡效果更加自然。

[0081] 图2是本申请实施例提供的一种视频数据处理方法的流程图。该方法可以应用于上述终端或者服务器,在本申请实施例中,以服务器作为执行主体,对该视频数据处理方法进行介绍,参见图2,该实施例具体可以包括以下步骤:

[0082] 201、服务器获取待处理的第一视频和第二视频。

[0083] 其中,该第一视频、第二视频为待合成的视频,该第一视频的视频帧中显示有相框,该相框所在区域的画面需要替换为第二视频的视频帧。例如,图3是本申请实施例提供的一种第一视频的视频帧画面示意图,如图3所示,该视频帧画面中显示有用户手持相框301。当然,上述相框也可以替换为其他物品,例如纸板等可以确定出一个区域的物品,以图3所示的画面为例,该第一视频也可以包括用户手持纸板的镜头,则在后续视频合成时,该纸板区域的画面需要替换为第二视频的视频帧。需要说明的是,在本申请实施例中,仅以第一视频中显示有相框为例进行说明。

[0084] 在一种可能实现方式中,服务器可以响应于视频编辑请求,获取该第一视频和第二视频。该视频编辑请求可以是任一终端发送的,本申请实施例对此不作限定。

[0085] 202、服务器确定该第一视频的多个视频帧的区域移动信息,该区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况,该目标区域用于显示该第二视频的视频帧。

[0086] 在本申请实施例中,以该第一视频的视频帧中包括相框为例,该目标区域即为相框所在区域。在一种可能实现方式中,服务器可以基于线段检测的方式,识别出相框的边框,进而确定出相框所在区域,即目标区域,实现对目标区域的跟踪。在第一视频中,相框位置的快速移动会导致相框所在区域,即目标区域出现运动模糊的效果,在本申请实施例中,服务器需要获取相邻两个视频帧中的目标区域的位置移动情况,该位置移动情况用于指示第一视频的各个视频帧中目标区域的运动模糊程度。

[0087] 203、服务器基于该第一视频的多个视频帧的区域移动信息,对该第二视频的多个视频帧进行运动模糊处理。

[0088] 在一种可能实现方式中,服务器可以通过滤波器对第二视频的多个视频帧进行运动模糊处理,该滤波器可以包括至少一个卷积层。例如,对于一个视频帧中的任一像素,服务器可以通过卷积运算,对该任一像素及其周围的像素取平均值,得到该任一像素对应的平均像素,由视频帧中各个像素对应的平均像素所构成的新的画面,即为运动模糊处理后的视频帧。

[0089] 204、服务器将该第一视频的多个视频帧中目标区域的图像,分别替换为运动模糊处理后的该第二视频的多个视频帧,得到目标视频。

[0090] 例如,第一视频的第*i*帧和第二视频的第*j*帧相对应,则该服务器可以将第一视频的第*i*帧中目标区域的图像,替换为第二视频的第*j*帧,该第*j*帧是经过运动模糊处理后的视频帧。其中,*i*和*j*均为正整数。服务器完成视频帧替换后,得到目标视频,将该目标视频发送至终端。

[0091] 本申请实施例提供的技术方案,通过在将第二视频嵌入到第一视频的目标区域时,基于人工智能技术来获取第一视频中目标区域的位置移动情况,根据该位置移动情况相应地对第二视频的画面进行运动模糊处理,使第二视频和第一视频具有相同的运动模糊效果,从而使第一视频和第二视频能够更好的融合,得到更好的视频过渡效果。

[0092] 上述实施例是对本申请实施方式的一个简要介绍,图4是本申请实施例提供的一种视频数据处理方法的流程图,结合图4,对上述视频数据处理过程进行具体说明。

[0093] 401、服务器获取待处理的第一视频和第二视频。

[0094] 在一种可能实现方式中,服务器基于终端的视频编辑请求,获取该第一视频和第二视频。例如,该视频编辑请求中携带第一视频和第二视频的视频标识,服务器响应于该视频编辑请求,从数据库中获取视频标识所指示的视频。其中,该数据库可以用于存储用户预先上传的视频。可选地,终端也可以将视频编辑请求以及第一视频、第二视频一起发送至服务器。需要说明的是,本申请实施例对具体采用哪种方法获取第一视频、第二视频不作限定。

[0095] 402、服务器对第一视频中的多个视频帧分别进行线段检测。

[0096] 在本申请示例中,该第一视频中显示有相框,该相框所框出的区域用于显示第二视频的视频帧。在一种可能实现方式中,该第一视频的每一视频帧中都显示有相框,则服务器对该第一视频的每一帧均进行线段检测。可选的,该第一视频的部分视频帧中显示有相框,即第一视频中的某个视频片段显示有相框,则服务器获取到第一视频后可以确定出第一视频中显示有相框的视频片段,对该视频片段中的视频帧进行线段检测。需要说明的是,本申请实施例对服务器对第一视频中的哪些视频帧进行线段检测不作限定。

[0097] 在一种可能实现方式中,服务器可以基于线段检测模型对视频帧进行线段检测,其中,该线段检测模型是基于深度神经网络构建的模型,例如,该线段检测模型是L-CNN模型。需要说明的是,本申请实施例对具体采用哪种线段检测模型不作限定,在本申请实施例中,仅以L-CNN模型为例进行说明。图5是本申请实施例提供的一种线段检测模型的示意图,参见图5,该线段检测模型可以包括主干网络501、连接点预测单元502、线段采样单元503以及线段校正单元504。其中,该主干网络501用于对视频帧进行特征提取,得到视频帧对应的特征图;该连接点预测单元502用于基于各个视频帧的特征图来预测各个视频帧中所包括的连接点;线段采样单元503用于基于连接点来预测线段,确定预测出的各个线段对应的置信度;线段校正单元504基于线段采样单元503的输出结果进行线段筛选。以下,结合图5,对线段检测的方法进行说明,该方法可以包括以下步骤:

[0098] 步骤一、服务器对该第一视频中的多个视频帧分别进行特征提取,得到视频帧对应的特征图。

[0099] 在一种可能实现方式中,服务器将第一视频的各个视频帧输入线段检测模型,由线段检测模型的主干网络分别对各个视频帧进行特征提取。其中,该主干网络可以是基于Stacked Hourglass Networks(堆叠沙漏网络)构建的网络,服务器通过主干网络中多个不同尺度的卷积层,依次对视频帧对应的像素矩阵进行卷积运算,来提取视频帧的语义特征,从而得到视频帧对应的特征图。需要说明的是,上述对特征图获取方法的说明,仅是一种示例性说明,本申请实施例对具体采用哪种方法获取特征图不作限定。

[0100] 步骤二、服务器基于各个视频帧对应的特征图,确定各个视频帧所包括的连接点。

[0101] 其中,该连接点用于在后续线段识别过程中确定线段的端点,视频帧中线段的端点、线段的交叉点等均可以被识别为连接点。

[0102] 在一种可能实现方式中,服务器将主干网络输出的特征图输入到连接点预测单元,通过连接点预测单元中的多个卷积层对特征图进行进一步卷积运算,来预测出视频帧所包括的连接点。例如,首先,服务器基于将该特征图划分为多个子区域。然后,服务器基于第一卷积核对该特征图中的各个子区域进行卷积运算,得到概率特征图,基于第二卷积核对该概率特征图进行卷积运算,得到位置特征图。其中,该概率特征图中的一个元素用于指

示一个子区域中存在连接点的概率,该位置特征图中的一个元素用于指示一个子区域中连接点的位置信息,在一种可能实现方式中,该位置信息可以表示为连接点的位置相对于子区域中心点的偏移量。最后,服务器基于该概率特征图以及该位置特征图,确定该视频帧所包括的连接点。例如,概率特征图中的每个元素以数值0或1表示,元素为1时,表示该元素对应的子区域中包括连接点,元素为0时,表示该元素对应的子区域中不包括连接点。服务器基于概率特征图确定出包括连接点的目标子区域,再基于位置特征图确定各个目标子区域中连接点的位置。需要说明的是,本申请实施例对上述第一卷积核、第二卷积核中参数的具体数值不作限定。

[0103] 步骤三、服务器基于该连接点确定该视频帧所包括的线段。

[0104] 在一种可能实现方式中,服务器将连接点预测单元的输出结果输入到线段采样单元,得到候选线段列表,即连接点对列表,也即是两个连接点分别作为线段的端点;再将候选线段列表和主干网络输出的特征图,输入线段校正单元,由线段校正单元的LoI (line of interest) 池化层提取各个候选线段的线段特征,再由至少一个全连接基于线段特征确定各个候选线段的分类结果,例如,全连接层可以输出各个候选线段对应的置信度,服务器获取置信度大于目标阈值的候选线段作为视频帧所包括的线段。其中,该目标阈值可以由开发人员进行设置,本申请实施例对此不作限定。

[0105] 需要说明的是,上述对线段检测方法的说明,仅是一种示例性说明,本申请实施例对具体采用哪种方法进行线段检测不作限定。

[0106] 403、服务器基于检测出的线段,确定该第一视频的多个视频帧所包括的目标区域。

[0107] 在本申请实施例中,对于该第一视频中的第一帧,服务器从检测出的该线段中,获取与参考区域的边框之间距离最近的线段作为目标线段,将目标线段围成的区域确定为该目标区域。其中,该参考区域是用户指定的区域。在一种可能实现方式中,用户在进行视频编辑,即对两个视频应用相框转场特效时,可以现在第一视频的第一帧中对相框框出的区域进行标注,即标出参考区域。例如,用户的终端运行有用于进行视频编辑的目标应用程序,用户可以在该目标应用程序中进行参考区域标注,图6是本申请实施例提供的一种参考区域标注的示意图,目标应用程序的界面中显示有第一视频的第一帧601,用户以矩形框的形式标注出参考区域602。在一种可能实现方式中,终端可以将参考区域在第一帧中的位置信息发送至服务器,服务器在检测出第一帧所包括的线段后,确定各个线段与参考区域的边框线段之间的距离,例如,确定线段的端点与参考区域的边框线段的端点之间的距离,获取与参考区域的边框之间距离最近的线段作为目标线段,将目标线段或目标延段的延长线所围成的区域确定为目标区域。在本申请实施例中,基于用户标注的参考区域以及线段检测模型的预测结果进行目标区域确定,一方面,参考区域可以对线段检测模型的数据结果进行筛选,去除大量的干扰线段,准确确定出目标区域;另一方面,由于是基于线段检测模型识别出的线段,来确定目标区域的边框,用户在进行参考区域标注时,只需标注出大致位置即可,降低了用户的标注难度。

[0108] 在本申请实施例中,对于该第一视频中除第一帧以外的任一帧,服务器从检测出的该线段中,获取与前一帧中的目标线段之间距离最近的线段,作为该任一帧中的目标线段,将该任一帧中的目标线段围成的区域确定为该目标区域。在一种可能实现方式中,线段

之间的距离可以由线段端点之间的距离表示,也即是,服务器获取当前帧所包括线段的线段端点坐标、前一帧所包括线段的线段端点坐标,基于线段端点之间的横、纵坐标差值,确定线段之间的距离。例如,上述确定线段之间距离的方法可以表示为下述公式(1):

$$[0109] \quad d = \sum_{i=1}^2 \sqrt{(L_{\text{current}}^{x_i} - L_{\text{pre}}^{x_i})^2 + (L_{\text{current}}^{y_i} - L_{\text{pre}}^{y_i})^2} \quad (1)$$

[0110] 其中,d表示线段之间的距离;pre用于标识前一帧,current用于标识当前帧,i用于表示线段端点的序号, $L_{\text{current}}^{x_i}$ 表示当前帧中线段端点i的横坐标, $L_{\text{current}}^{y_i}$ 表示当前帧中线段端点i的纵坐标, $L_{\text{pre}}^{x_i}$ 表示前一帧中线段端点i的横坐标, $L_{\text{pre}}^{y_i}$ 表示前一帧中线段端点i的纵坐标。

[0111] 在上述确定目标区域的过程中,采用先获取与目标区域的边框距离最近的目标线段,再基于目标线段的延长线的相交位置来确定目标区域的至少三个角点,根据确定出的至少三个角点来定位目标区域。图7是本申请实施例提供的一种目标区域确定方法示意图,以下结合图7,对上述基于目标线段确定出目标区域的过程进行说明,在相框转场特效的应用过程中,会出现因用户手持相框,导致手部对相框边缘、角点造成遮挡的情况,也会出现因相框晃动而导致相框区域运动模糊的情况,也即是,在视频帧画面中无法清晰地显示相框。在这种情况下,对相框区域,也即是目标区域的边界进行识别时,仅能识别到目标区域边界上的部分线段,如图7所示,在相框发生运动模糊和被遮挡的情况下,在对该视频帧进行线段检测时,仅能检测出线段701、702、703和704,也即是,仅能识别出相框的部分边缘区域,在本申请实施例中,可以基于线段701、702、703和704的延长线,来确定出角点705、706、707和708,再根据角点定位出目标区域709,例如对各个角点进行连接,围成目标区域。在相框转场特效的应用过程中,采用先确定线段,再基于线段的延长线来确定目标区域角点的方法,能够确定出目标区域的全部角点,从而避免在相框转场过程中出现因目标区域的角点被遮挡或显示模糊,而导致角点无法识别、目标区域定位失败的情况。在后续对两个视频进行合并时,避免因识别出的目标区域不完整,而导致第二视频的视频画面不能完全显示的情况,确保第二视频的视频帧画面有良好的显示效果。

[0112] 需要说明的是,上述步骤402和步骤403,是确定该第一视频的多个视频帧中的目标区域的步骤。在本申请实施例中,基于前一帧的所确定出的目标区域,来预测当前帧中的目标区域,逐帧检测目标区域,实现对目标区域的准确跟踪,无需用户手动逐帧标注目标区域。当第一视频出现运动模糊时,对第一视频的视频帧进行线段检测,仅能识别出相框边缘的一部分,在这种情况下,基于前一帧的识别结果,来预测当前帧中的目标区域,可以有效提高目标区域识别的准确率,即在运动模糊或遮挡的情况下,也能准确跟踪到目标区域。例如,在运动模糊或者有遮挡的情况下,目标区域的一条边可能会被遮挡,以图8为例,图8是本申请实施例提供的一种运动模糊效果的视频帧示意图,由于该视频帧的相框产生的运动模糊,在对该视频帧进行线段检测时,仅能检测出线段801、802、和803,相框的一条边无法识别出,从而仅根据当获取到的线段无法识别出目标区域,在这种情况下,结合前一帧所识别出目标区域,以及当前帧中识别出的部分线段,就可以对当前帧中的目标区域进行预测。

[0113] 404、服务器基于该第一视频的多个视频帧所包括目标区域的角点坐标,分别对该

第二视频中对应的视频帧进行透视变换,基于透视变换后的视频帧执行后续的运动模糊步骤。

[0114] 在一种可能实现方式中,服务器基于该第一视频的多个视频帧所包括目标区域的角点坐标之间的相对位置,确定该第一视频中各个视频帧对应的透视变换矩阵;再基于该第一视频中各个视频帧对应的透视变换矩阵,分别对该第二视频中对应的视频帧进行透视变换。以第一视频中的第*i*帧与第二视频中的第*j*帧相对应为例,上述透视变换过程可以表示为下述公式(2)和(3):

$$[0115] \quad [x', y', w'] = [u, v, w] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (2)$$

$$[0116] \quad \begin{aligned} x &= \frac{x'}{w'} = \frac{a_{11}u + a_{21}v + a_{31}}{a_{13}u + a_{23}v + a_{33}} \\ y &= \frac{y'}{w'} = \frac{a_{12}u + a_{22}v + a_{32}}{a_{13}u + a_{23}v + a_{33}} \end{aligned} \quad (3)$$

[0117] 其中, a_{11} 、 a_{12} 、 a_{13} 、 a_{21} 、 a_{22} 、 a_{23} 、 a_{31} 、 a_{32} 、 a_{33} 是透视变换矩阵的参数,其数值基于第一视频中第*i*帧所包括目标区域的各个角点坐标确定,本申请实施例对此不作限定。 u 、 v 表示第二视频的第*j*帧中各个像素点的原始横、纵坐标, w 可以设置为1; x 、 y 表示第二视频的第*j*帧中各个像素点在透视变换后的横、纵坐标, x' 、 y' 表示第二视频的第*j*帧中各个像素点在透视变换过程中的中间横、纵坐标, w' 表示像素点在三维空间的竖坐标。在上述过程中,服务器先将二维坐标(u, v)先映射到三维空间得到(x', y', w'),再映射到二维空间得到(x, y),以完成透视变换。需要说明的是上述对透视变换方法的说明,仅是一种示例性说明,本申请实施例对具体采用哪种方法对第二视频的视频帧进行透视变换不作限定。

[0118] 在本申请时实施例中,第一视频中的相框在移动时,因各个角点距离镜头位置的不同,相框会出现透视的情况,在这种情况下,基于识别出的目标区域中各个角点之间的相对位置关系,来确定相框的透视情况,在将第二视频嵌入到第一视频之前,基于第一视频中相框的透视情况,对第二视频的视频帧进行透视变换,使第二视频能够更好地与第一视频进行融合,使合成后的视频更加逼真,呈现出良好的视频过渡效果。

[0119] 405、服务器获取第一视频中前后两个视频帧中该目标区域的位置移动情况,基于该位置移动情况确定目标区域的区域移动信息。

[0120] 在本申请实施例中,由于目标区域在移动时,目标区域的各个角点的移动距离和移动方向均不同,则目标区域的区域移动信息包括目标区域的各个角点对应的移动信息,即该区域移动信息包括目标区域的各个角点的移动距离和移动方向。

[0121] 在一种可能实现方式中,服务器可以基于各个角点在前后两个视频帧中位置坐标的变化情况,来确定该角点的移动距离和移动方向。图9是本申请实施例提供的一种目标区域移动情况的示意图,图9示出了第一视频的第*i*帧中目标区域的位置901和第一视频的第*i*-1帧中目标区域的位置902,在目标区域移动过程中,各个角点的移动幅度是不同的,例如,第*i*-1帧中*p*1位置的角点移动到了第*i*帧的*p*1'位置。以获取一个角点的移动距离和移动方向为例,上述过程可以表示为下述公式(4)和(5):

$$[0122] \quad k = p\sqrt{(X_{\text{current}} - X_{\text{pre}})^2 + (Y_{\text{current}} - Y_{\text{pre}})^2} \quad (4)$$

$$[0123] \quad \cos\theta = \frac{|X_{\text{current}} - X_{\text{pre}}|}{\sqrt{(X_{\text{current}} - X_{\text{pre}})^2 + (Y_{\text{current}} - Y_{\text{pre}})^2}} \quad (5)$$

[0124] 其中, X_{current} 、 Y_{current} 分别表示当前帧中角点的横、纵坐标, X_{pre} 、 Y_{pre} 分别表示前一帧中角点的横、纵坐标; k 表示角点的移动距离; $\cos\theta$ 表示角点的移动方向; p 表示调节因子, 其具体数值由开发人员进行设置, 本申请实施例对此不作限定。

[0125] 需要说明的是, 上述步骤402、步骤404, 是确定该第一视频的多个视频帧的区域移动信息, 该区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况, 该目标区域用于显示该第二视频的视频帧的步骤, 在本申请实施例中, 基于相邻两个视频帧中相框的位置移动情况, 来对第二视频进行后续的运动模糊处理, 可以使第二视频和第一视频的融合效果更自然。

[0126] 406、服务器基于该第一视频的多个视频帧的区域移动信息, 对该第二视频的多个视频帧进行运动模糊处理。

[0127] 在一种可能实现方式中, 服务器可以通过卷积模块, 对第二视频的视频帧进行卷积运算, 来实现运动模糊处理, 该方法可以包括以下步骤:

[0128] 步骤一、服务器基于前后两个视频帧中目标区域的至少三个角点的移动距离和移动方向, 分别确定至少三组卷积参数。

[0129] 其中, 基于一个角点的移动距离和移动方向, 可以确定出一组卷积参数, 该卷积参数用于指示在一次卷积运算中的卷积区域大小以及卷积核的移动方向。角点的移动距离可以表示卷积核的卷积区域大小, 卷积区域的大小与运动模糊半径正相关; 角点的移动方向可以表示卷积核的移动方向。

[0130] 步骤二、服务器通过该第三卷积核分别基于该至少三组卷积参数对该第二视频中对应的视频帧进行卷积运算, 得到至少三个中间帧。

[0131] 在一种可能实现方式中, 卷积模块中可以设置有多个不同尺度的候选卷积核, 不同尺度的候选卷积核对应于不同尺度的卷积区域, 也即是对应于不同的移动距离取值范围。对于任一组卷积参数, 服务器可以从多个不同大小的候选卷积核中, 获取尺寸与该任一组卷积参数中该卷积区域大小相匹配的候选卷积核, 作为第三卷积核; 通过该第三卷积核按照该任一组卷积参数中的移动方向, 对该第二视频中对应的视频帧进行卷积运算, 得到中间视频帧。例如, 一组卷积参数中, 卷积区域大小为 3×3 , 则服务器可以选择 3×3 的候选卷积核作为第三卷积核。在一种可能实现方式中, 卷积模块中也可以仅设置一种尺度的第三卷积核, 服务器在获取到卷积参数后, 对该第三卷积核的卷积区域进行调整。以卷积运算是获取卷积区域内各个元素的平均值为例, 若一组卷积参数中的卷积区域是 5×5 时, 服务器在应用第三卷积核对目标像素点进行卷积运算时, 可以以目标像素点为中心, 对该目标像素点周围 5×5 区域内的元素取平均值。需要说明的是, 本申请实施例对具体采用哪种方式调整卷积运算过程中的卷积区域不作限定。

[0132] 以第一视频的第 i 帧和第二视频的第 j 帧相对应为例, 对上述中间帧获取过程进行说明, 根据目标区域在第一视频的第 $i-1$ 帧和第 i 帧之间的区域移动信息, 确定出至少三组卷积参数后, 服务器通过第三卷积核分别基于各组卷积参数对第二视频的第 j 帧进行卷积

运算,得到第j帧对应的至少三个中间帧。图10是本申请实施例提供的一种运动模糊处理的示意图,以图10为例对上述中间帧获取过程进行说明,参见图10,若目标区域中的角点A对应于第一组卷积参数,该第一组卷积参数中卷积区域的大小是 $4*4$,卷积方向是1001方向,则服务器在基于该第一组卷积参数进行卷积运算时,第三卷积核从第二视频的视频帧1002的左上方开始,根据第一组卷积参数所指示的卷积区域大小和卷积方向进行卷积运算,得到一个中间帧。若目标区域中的角点B,对应于第二组卷积参数,该第二组卷积参数中卷积区域的大小是 $3*3$,卷积方向是1003方向,则服务器在基于该第二组卷积参数进行卷积运算时,第三卷积核从视频帧1002的左上方开始,根据第二组卷积参数所指示的卷积区域大小和卷积方向进行卷积运算,得到另一个中间帧。图11是本申请实施例提供的一种运动模糊效果示意图,目标区域为有四个角点时,根据每个角点对应的一组卷积参数分别对第二视频的视频帧进行卷积运算后,所得到的四个中间帧有不同的运动模糊效果,如图11中的(a)图、(b)图、(c)图和(d)图所示。

[0133] 步骤三、服务器对该至少三个中间帧进行叠加,得到运动模糊处理后的该第二视频的视频帧。

[0134] 在一种可能实现方式中,对于该第二视频的视频帧中的任一像素点,服务器基于该任一像素点相对于各个角点的距离,确定在任一像素点位置,各个中间帧对应的透明度参数;对于该第二视频的视频帧中的任一像素点位置,基于各个中间帧对应的透明度参数,对各个中间帧进行叠加,得到运动模糊处理后的第二视频的视频帧。例如,对于第二视频中第j帧的像素点t,服务器获取像素点t与各个角点之间的距离,例如,第二视频的第j帧对应于第一视频的第i帧,服务器可以将第二视频的第j帧置于第一视频中第i帧的目标区域内,再确定出像素点t与各个角点之间的距离。服务器基于像素点t与各个角点之间的距离比例,确定像素点t位置处,各个中间帧对应的透明度参数。例如,服务器分别基于角点A对应的卷积参数进行卷积运算,得到中间帧a,则中间帧a对应的透明度参数与像素点t到角点A的距离正相关,即像素点t到角点A的距离越远,在像素点t处,中间帧a的透明度越大,像素点t到角点A的距离越近,在像素点t处,中间帧a的透明度越小。在像素点t处,服务器基于各个中间帧对应的透明度参数对中间帧进行叠加,得到一个像素点。基于上述过程,得到第二视频的第j帧中每个像素点所对应的新像素点,也就得到了经过运动模糊处理的第j帧。

[0135] 需要说明的是,上述对视频帧进行运动模糊处理方法的说明,仅是一种示例性说明,本申请实施例对具体采用哪种方式对视频帧进行运动模糊处理不作限定。在本申请实施例中,对第二视频的视频帧进行运动模糊处理,使第二视频的显示效果与第一视频的显示效果相匹配,以获得更自然的视频转场效果。且应用上述运动模糊处理方法,无需对视频帧中的各个像素点均进行不同程度的运动模糊处理,只需调整中间帧的透明度,对不同透明度的中间帧进行叠加即可,大大降低了运动模糊处理过程中的运算量。

[0136] 407、服务器将该第一视频的多个视频帧中该目标区域的图像,分别替换为运动模糊处理后的该第二视频的多个视频帧,得到目标视频。

[0137] 以第一视频的第i帧和第二视频的第j帧相对应为例,服务器可以将第一视频的第i帧中目标区域的图像,替换为第二视频的第j帧,得到目标视频的一个视频帧。图12是本申请实施例提供的一种目标视频的视频帧示意图,如图12所示,目标区域1201中显示有运动模糊处理后的图像。在对目标区域的图像替换完成后,即可得到目标视频,即添加了相框转

场效果的目标视频。在一种可能实现方式中,服务器可以将目标视频发送至终端,由终端进行播放。

[0138] 图13是本申请实施例提供的一种相框转场效果的实现方法流程图,结合图13,对上述视频数据处理过程进行说明。在一种可能实现方式中,服务器在获取到待处理的第一视频、第二视频后,先执行对第一视频进行线段检测,提取相框边缘直线的步骤1301;再根据检测出的线段,执行目标跟踪步骤1302,即分析第一视频的相邻两帧中线段的位置变化,来对相框进行跟踪。该目标区域的确定过程如图14所示,图14是本申请实施例提供的一种目标区域跟踪过程示意图,具体地,在线段检测阶段,服务器检测出第一视频的当前帧中的全部线段,如图14中的(a)图所示,再根据前一帧的目标区域确定结果,筛选出当前帧中的目标线段,根据目标线段或目标线段的延长线确定目标区域,如图14中的(b)图所示,得到的目标区域如图14中的(c)图所示。上述目标区域跟踪的过程即步骤402、403的过程。在确定出目标区域,即相框框出的区域后,服务器提取相框的透视信息,执行对第二视频的视频帧进行透视变换的步骤1303,也即是,执行上述步骤404中的内容;对于透视变换后的第二视频中的视频帧,执行运动模糊处理的步骤1304,使第二视频和第一视频保持相同的运动模糊效果,也即是,执行上述步骤405和步骤406;再将运动模糊处理后的第二视频和第一视频进行合成,即执行渲染特效视频的步骤1305,得到添加了相框转场特效的目标视频。

[0139] 本申请实施例提供的技术方案,通过在将第二视频嵌入到第一视频的目标区域时,基于人工智能技术来获取第一视频中目标区域的位置移动情况,根据该位置移动情况相应地对第二视频的画面进行运动模糊处理,使第二视频和第一视频具有相同的运动模糊效果,从而使第一视频和第二视频能够更好的融合,得到更好的视频过渡效果。

[0140] 上述所有可选技术方案,可以采用任意结合形成本申请的可选实施例,在此不再一一赘述。

[0141] 图15是本申请实施例提供的一种视频数据处理装置的结构示意图,参见图15,该装置包括:

[0142] 获取模块1501,用于获取待处理的第一视频和第二视频;

[0143] 确定模块1502,用于确定该第一视频的多个视频帧的区域移动信息,该区域移动信息用于表示前后两个视频帧中的目标区域的位置移动情况,该目标区域用于显示该第二视频的视频帧;

[0144] 运动模糊处理模块1503,用于基于该第一视频的多个视频帧的区域移动信息,对该第二视频的多个视频帧进行运动模糊处理;

[0145] 替换模块1504,用于将该第一视频的多个视频帧中该目标区域的图像,分别替换为运动模糊处理后的该第二视频的多个视频帧,得到目标视频。

[0146] 在一种可能实现方式中,该确定模块1502包括:

[0147] 第一确定单元,用于确定该第一视频的多个视频帧中的目标区域;

[0148] 第二确定单元,用于基于该第一视频中前后两个视频帧中该目标区域的位置移动情况,确定该区域移动信息。

[0149] 在一种可能实现方式中,该第一确定单元包括:

[0150] 检测子单元,用于对该第一视频中的多个视频帧分别进行线段检测;

[0151] 区域确定子单元,用于基于检测出的线段,确定该第一视频的多个视频帧所包括

的目标区域。

[0152] 在一种可能实现方式中,该检测子单元包括:

[0153] 特征提取子单元,用于对该第一视频中的多个视频帧分别进行特征提取,得到视频帧对应的特征图;

[0154] 连接点确定子单元,用于基于该视频帧对应的特征图,确定该视频帧所包括的连接点;

[0155] 线段确定子单元,用于基于该连接点确定该视频帧所包括的线段。

[0156] 在一种可能实现方式中,该连接点确定子单元用于:

[0157] 将该特征图划分为多个子区域;

[0158] 基于第一卷积核对该特征图中的子区域进行卷积运算,得到概率特征图,该概率特征图中的一个元素用于指示一个子区域中存在连接点的概率;

[0159] 基于第二卷积核对该概率特征图进行卷积运算,得到位置特征图,该位置特征图中的一个元素用于指示一个子区域中连接点的位置信息;

[0160] 基于该概率特征图以及该位置特征图,确定该视频帧所包括的连接点。

[0161] 在一种可能实现方式中,该区域确定子单元用于:

[0162] 对于该第一视频中的第一帧,从检测出的该线段中,获取与参考区域的边框之间距离最近的线段作为目标线段,将该目标线段围成的区域确定为该目标区域;

[0163] 对于该第一视频中除第一帧以外的任一帧,从检测出的该线段中,获取与前一帧中的目标线段之间距离最近的线段,作为该任一帧中的目标线段,将该任一帧中的目标线段围成的区域确定为该目标区域。

[0164] 在一种可能实现方式中,该区域移动信息包括该目标区域的至少三个角点的移动距离和移动方向。

[0165] 在一种可能实现方式中,该运动模糊处理模块1503包括:

[0166] 参数确定单元,用于基于前后两个视频帧中目标区域的至少三个角点的移动距离和移动方向,分别确定至少三组卷积参数,该卷积参数用于指示第三卷积核在一次卷积运算中的卷积区域大小以及该第三卷积核的移动方向;

[0167] 卷积单元,用于通过该第三卷积核分别基于该至少三组卷积参数对该第二视频中对应的视频帧进行卷积运算,得到至少三个中间帧;

[0168] 叠加单元,用于对该至少三个中间帧进行叠加,得到运动模糊处理后的该第二视频的视频帧。

[0169] 在一种可能实现方式中,该卷积单元用于:

[0170] 对于任一组卷积参数,从多个不同大小的候选卷积核中,获取尺寸与该任一组卷积参数中该卷积区域大小相匹配的候选卷积核,作为第三卷积核;

[0171] 通过该第三卷积核按照该任一组卷积参数中的移动方向,对该第二视频中对应的视频帧进行卷积运算,得到中间视频帧。

[0172] 在一种可能实现方式中,该叠加单元用于:

[0173] 对于该第二视频的视频帧中的任一像素点,基于该任一像素点相对于各个角点的距离,确定在任一像素点位置,各个该中间帧对应的透明度参数;

[0174] 对于该第二视频的视频帧中的任一像素点位置,基于各个该中间帧对应的透明度

参数,对该中间帧进行叠加,得到运动模糊处理后的该第二视频的视频帧。

[0175] 在一种可能实现方式中,该装置还包括:

[0176] 透视变换模块,用于基于该第一视频的多个视频帧所包括目标区域的角点坐标,分别对该第二视频中对应的视频帧进行透视变换,基于透视变换后的该第二视频的视频帧,执行基于该第一视频的多个视频帧的区域移动信息,对该第二视频的多个视频帧进行运动模糊处理的步骤。

[0177] 在一种可能实现方式中,该透视变换模块用于:

[0178] 基于该第一视频的多个视频帧所包括目标区域的角点坐标之间的相对位置,确定该第一视频中各个视频帧对应的透视变换矩阵;

[0179] 基于该第一视频中各个视频帧对应的透视变换矩阵,分别对该第二视频中对应的视频帧进行透视变换。

[0180] 本申请实施例提供的装置,通过在将第二视频嵌入到第一视频的目标区域时,基于人工智能技术来获取第一视频中目标区域的位置移动情况,根据该位置移动情况相应地对第二视频的画面进行运动模糊处理,使第二视频和第一视频具有相同的运动模糊效果,从而使第一视频和第二视频能够更好的融合,得到更好的视频过渡效果。

[0181] 需要说明的是:上述实施例提供的视频数据处理装置在视频数据处理时,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将装置的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。另外,上述实施例提供的视频数据处理装置与视频数据处理方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0182] 图16是本申请实施例提供的一种终端的结构示意图。该终端1600可以是:智能手机、平板电脑、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、笔记本电脑或台式电脑。终端1600还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

[0183] 通常,终端1600包括有:一个或多个处理器1601和一个或多个存储器1602。

[0184] 处理器1601可以包括一个或多个处理核心,比如4核心处理器、8核心处理器等。处理器1601可以采用DSP(Digital Signal Processing,数字信号处理)、FPGA(Field-Programmable Gate Array,现场可编程门阵列)、PLA(Programmable Logic Array,可编程逻辑阵列)中的至少一种硬件形式来实现。处理器1601也可以包括主处理器和协处理器,主处理器是用于对在唤醒状态下的数据进行处理的处理器,也称CPU(Central Processing Unit,中央处理器);协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中,处理器1601可以在集成有GPU(Graphics Processing Unit,图像处理器),GPU用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中,处理器1601还可以包括AI(Artificial Intelligence,人工智能)处理器,该AI处理器用于处理有关机器学习的计算操作。

[0185] 存储器1602可以包括一个或多个计算机可读存储介质,该计算机可读存储介质可以是非暂态的。存储器1602还可包括高速随机存取存储器,以及非易失性存储器,比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中,存储器1602中的非暂态的计算机可

读存储介质用于存储至少一条程序代码,该至少一条程序代码用于被处理器1601所执行以实现本申请中方法实施例提供的视频数据处理方法。

[0186] 在一些实施例中,终端1600还可选包括有:外围设备接口1603和至少一个外围设备。处理器1601、存储器1602和外围设备接口1603之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口1603相连。具体地,外围设备包括:射频电路1604、显示屏1605、摄像头组件1606、音频电路1607和电源1609中的至少一种。

[0187] 外围设备接口1603可被用于将I/O(Input/Output,输入/输出)相关的至少一个外围设备连接到处理器1601和存储器1602。在一些实施例中,处理器1601、存储器1602和外围设备接口1603被集成在同一芯片或电路板上;在一些其他实施例中,处理器1601、存储器1602和外围设备接口1603中的任意一个或两个可以在单独的芯片或电路板上实现,本实施例对此不加以限定。

[0188] 射频电路1604用于接收和发射RF(Radio Frequency,射频)信号,也称电磁信号。射频电路1604通过电磁信号与通信网络以及其他通信设备进行通信。射频电路1604将电信号转换为电磁信号进行发送,或者,将接收到的电磁信号转换为电信号。可选地,射频电路1604包括:天线系统、RF收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。射频电路1604可以通过至少一种无线通信协议来与其它终端进行通信。该无线通信协议包括但不限于:城域网、各代移动通信网络(2G、3G、4G及5G)、无线局域网和/或WiFi(Wireless Fidelity,无线保真)网络。在一些实施例中,射频电路1604还可以包括NFC(Near Field Communication,近距离无线通信)有关的电路,本申请对此不加以限定。

[0189] 显示屏1605用于显示UI(User Interface,用户界面)。该UI可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏1605是触摸显示屏时,显示屏1605还具有采集在显示屏1605的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器1601进行处理。此时,显示屏1605还可以用于提供虚拟按钮和/或虚拟键盘,也称软按钮和/或软键盘。在一些实施例中,显示屏1605可以为一个,设置终端1600的前面板;在另一些实施例中,显示屏1605可以为至少两个,分别设置在终端1600的不同表面或呈折叠设计;在一些实施例中,显示屏1605可以是柔性显示屏,设置在终端1600的弯曲表面上或折叠面上。甚至,显示屏1605还可以设置成非矩形的不规则图形,也即异形屏。显示屏1605可以采用LCD(Liquid Crystal Display,液晶显示屏)、OLED(Organic Light-Emitting Diode,有机发光二极管)等材质制备。

[0190] 摄像头组件1606用于采集图像或视频。可选地,摄像头组件1606包括前置摄像头和后置摄像头。通常,前置摄像头设置在终端的前面板,后置摄像头设置在终端的背面。在一些实施例中,后置摄像头为至少两个,分别为主摄像头、景深摄像头、广角摄像头、长焦摄像头中的任意一种,以实现主摄像头和景深摄像头融合实现背景虚化功能、主摄像头和广角摄像头融合实现全景拍摄以及VR(Virtual Reality,虚拟现实)拍摄功能或者其它融合拍摄功能。在一些实施例中,摄像头组件1606还可以包括闪光灯。闪光灯可以是单色温闪光灯,也可以是双色温闪光灯。双色温闪光灯是指暖光闪光灯和冷光闪光灯的组合,可以用于不同色温下的光线补偿。

[0191] 音频电路1607可以包括麦克风和扬声器。麦克风用于采集用户及环境的声波,并

将声波转换为电信号输入至处理器1601进行处理,或者输入至射频电路1604以实现语音通信。出于立体声采集或降噪的目的,麦克风可以为多个,分别设置在终端1600的不同部位。麦克风还可以是阵列麦克风或全向采集型麦克风。扬声器则用于将来自处理器1601或射频电路1604的电信号转换为声波。扬声器可以是传统的薄膜扬声器,也可以是压电陶瓷扬声器。当扬声器是压电陶瓷扬声器时,不仅可以将电信号转换为人类可听见的声波,也可以将电信号转换为人类听不见的声波以进行测距等用途。在一些实施例中,音频电路1607还可以包括耳机插孔。

[0192] 电源1609用于为终端1600中的各个组件进行供电。电源1609可以是交流电、直流电、一次性电池或可充电电池。当电源1609包括可充电电池时,该可充电电池可以支持有线充电或无线充电。该可充电电池还可以用于支持快充技术。

[0193] 在一些实施例中,终端1600还包括有一个或多个传感器1610。该一个或多个传感器1610包括但不限于:加速度传感器1611、陀螺仪传感器1612、压力传感器1613、光学传感器1615以及接近传感器1616。

[0194] 加速度传感器1611可以检测以终端1600建立的坐标系的三个坐标轴上的加速度大小。比如,加速度传感器1611可以用于检测重力加速度在三个坐标轴上的分量。处理器1601可以根据加速度传感器1611采集的重力加速度信号,控制显示屏1605以横向视图或纵向视图进行用户界面的显示。加速度传感器1611还可以用于游戏或者用户的运动数据的采集。

[0195] 陀螺仪传感器1612可以检测终端1600的机体方向及转动角度,陀螺仪传感器1612可以与加速度传感器1611协同采集用户对终端1600的3D动作。处理器1601根据陀螺仪传感器1612采集的数据,可以实现如下功能:动作感应(比如根据用户的倾斜操作来改变UI)、拍摄时的图像稳定、游戏控制以及惯性导航。

[0196] 压力传感器1613可以设置在终端1600的侧边框和/或显示屏1605的下层。当压力传感器1613设置在终端1600的侧边框时,可以检测用户对终端1600的握持信号,由处理器1601根据压力传感器1613采集的握持信号进行左右手识别或快捷操作。当压力传感器1613设置在显示屏1605的下层时,由处理器1601根据用户对显示屏1605的压力操作,实现对UI界面上的可操作性控件进行控制。可操作性控件包括按钮控件、滚动条控件、图标控件、菜单控件中的至少一种。

[0197] 光学传感器1615用于采集环境光强度。在一个实施例中,处理器1601可以根据光学传感器1615采集的环境光强度,控制显示屏1605的显示亮度。具体地,当环境光强度较高时,调高显示屏1605的显示亮度;当环境光强度较低时,调低显示屏1605的显示亮度。在另一个实施例中,处理器1601还可以根据光学传感器1615采集的环境光强度,动态调整摄像头组件1606的拍摄参数。

[0198] 接近传感器1616,也称距离传感器,通常设置在终端1600的前面板。接近传感器1616用于采集用户与终端1600的正面之间的距离。在一个实施例中,当接近传感器1616检测到用户与终端1600的正面之间的距离逐渐变小时,由处理器1601控制显示屏1605从亮屏状态切换为息屏状态;当接近传感器1616检测到用户与终端1600的正面之间的距离逐渐变大时,由处理器1601控制显示屏1605从息屏状态切换为亮屏状态。

[0199] 本领域技术人员可以理解,图16中示出的结构并不构成对终端1600的限定,可以

包括比图示更多或更少的组件,或者组合某些组件,或者采用不同的组件布置。

[0200] 图17是本申请实施例提供的一种服务器的结构示意图,该服务器1700可因配置或性能不同而产生比较大的差异,可以包括一个或多个处理器(Central Processing Units, CPU) 1701和一个或多个的存储器1702,其中,该一个或多个存储器1702中存储有至少一条程序代码,该至少一条程序代码由该一个或多个处理器1701加载并执行以实现上述各个方法实施例提供的方法。当然,该服务器1700还可以具有有线或无线网络接口、键盘以及输入输出接口等部件,以便进行输入输出,该服务器1700还可以包括其他用于实现设备功能的部件,在此不做赘述。

[0201] 在示例性实施例中,还提供了一种计算机可读存储介质,例如包括至少一条程序代码的存储器,上述至少一条程序代码可由处理器执行以完成上述实施例中的视频数据处理方法。例如,该计算机可读存储介质可以是只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、只读光盘(Compact Disc Read-Only Memory,CD-ROM)、磁带、软盘和光数据存储设备等。

[0202] 在示例性实施例中,还提供了一种计算机程序产品,该计算机程序产品包括至少一条程序代码,该至少一条程序代码存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该至少一条程序代码,处理器执行该计算机指令,使得该计算机设备实现该视频数据处理方法所执行的操作。

[0203] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来至少一条程序代码相关的硬件完成,该程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0204] 上述仅为本申请的可选实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

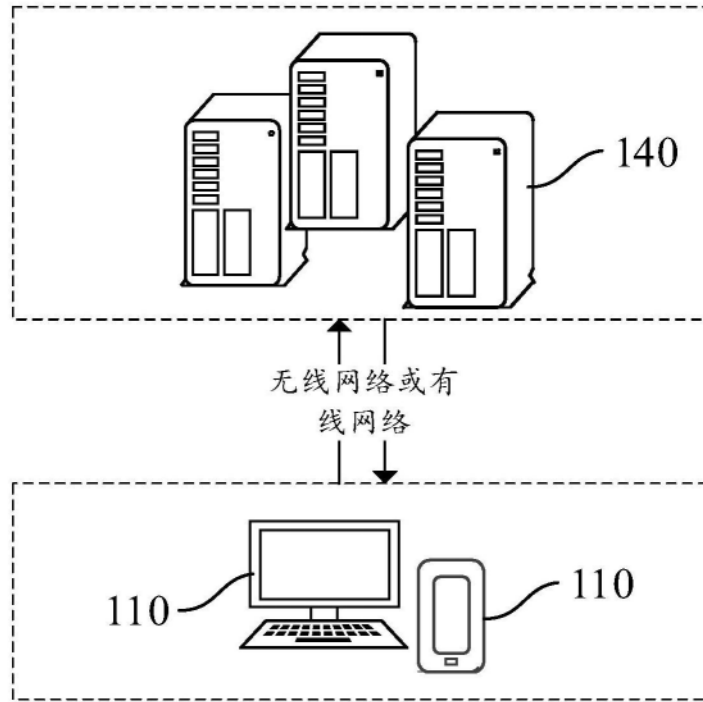


图1

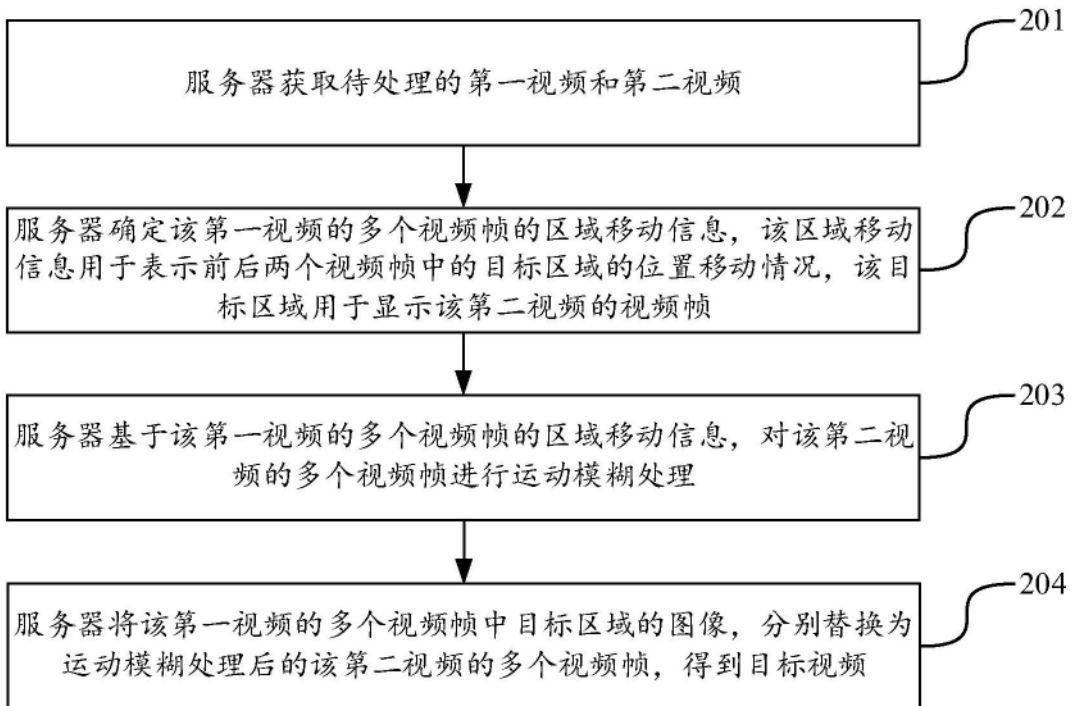


图2

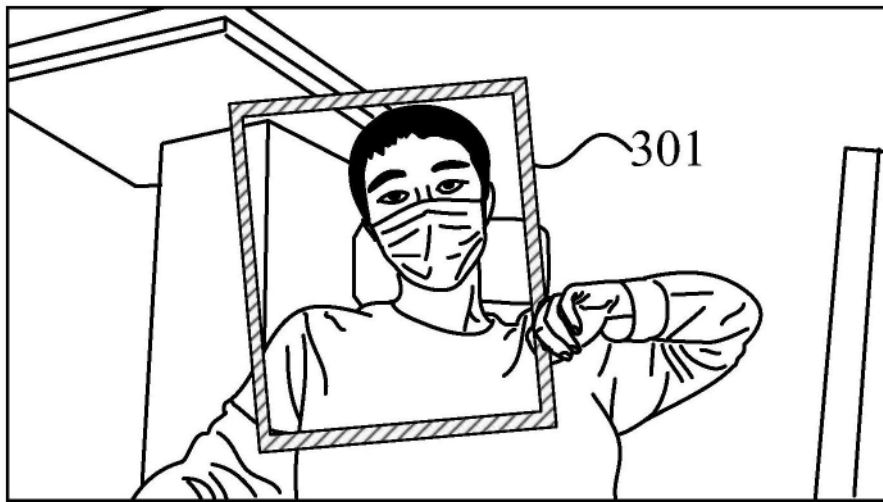


图3

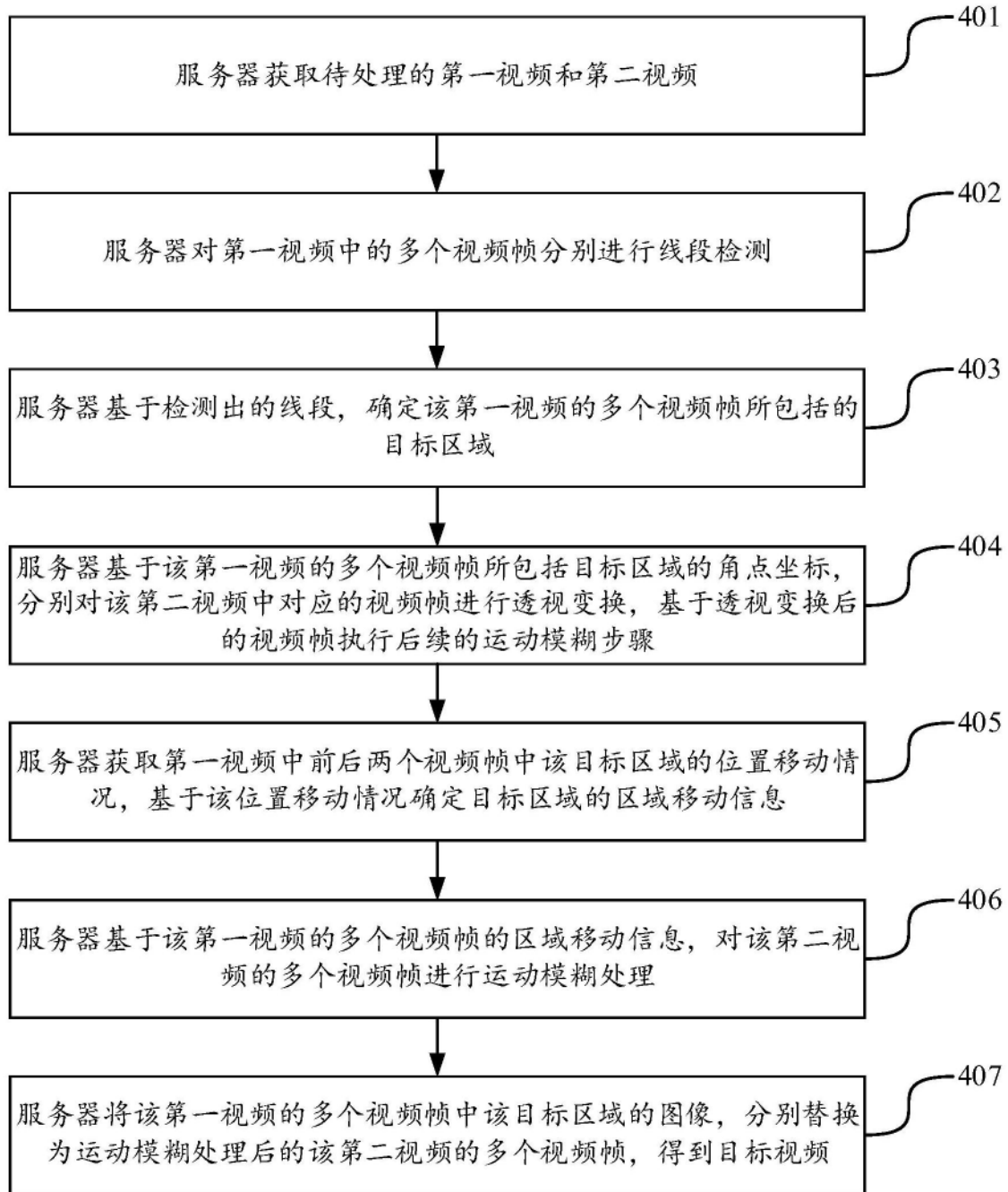


图4

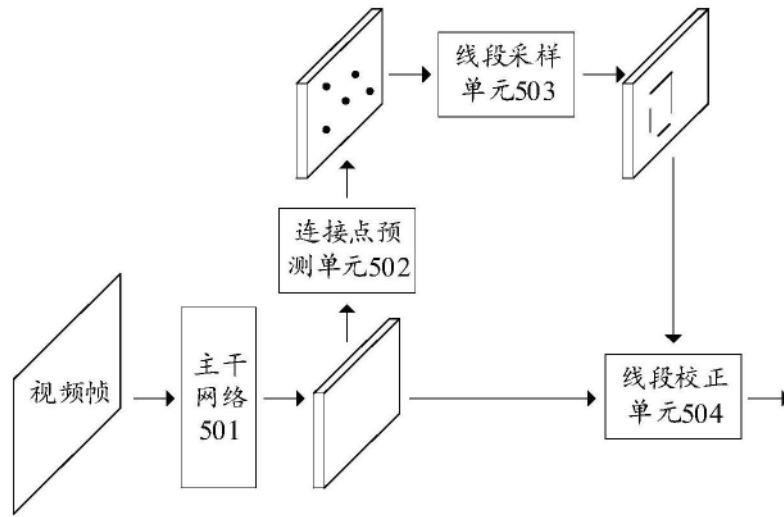


图5

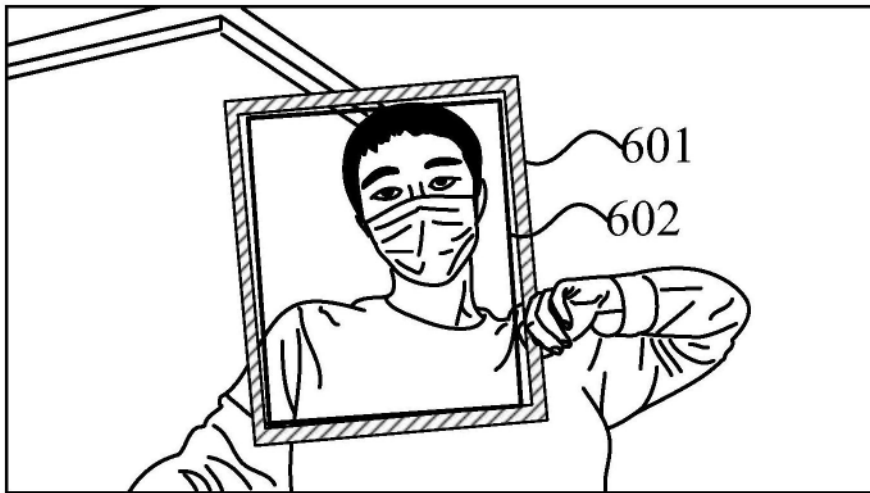


图6

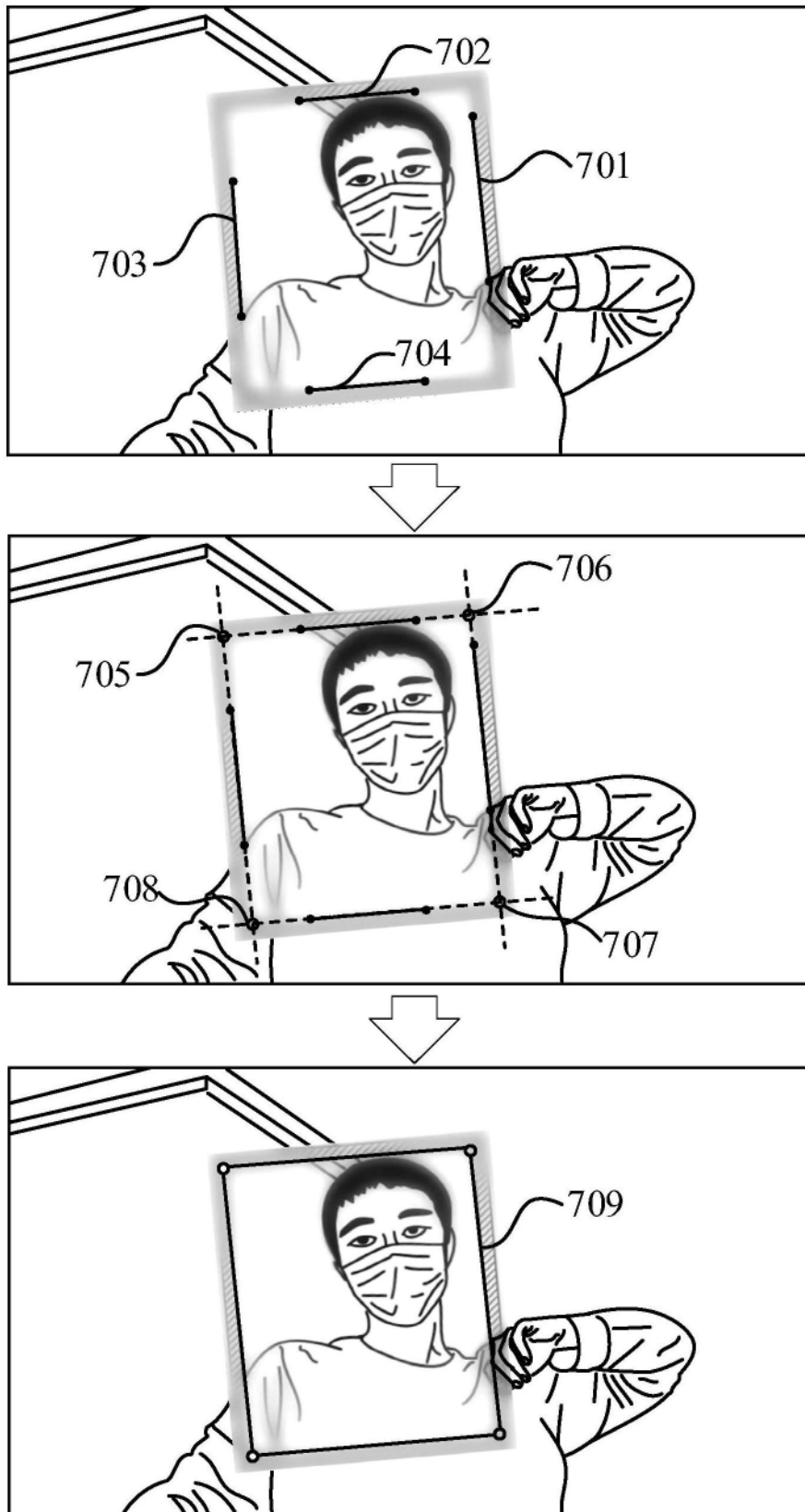


图7

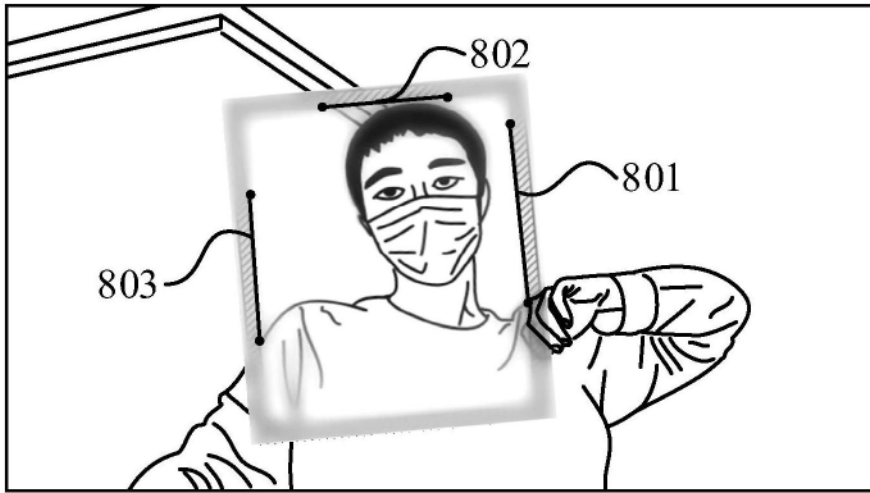


图8

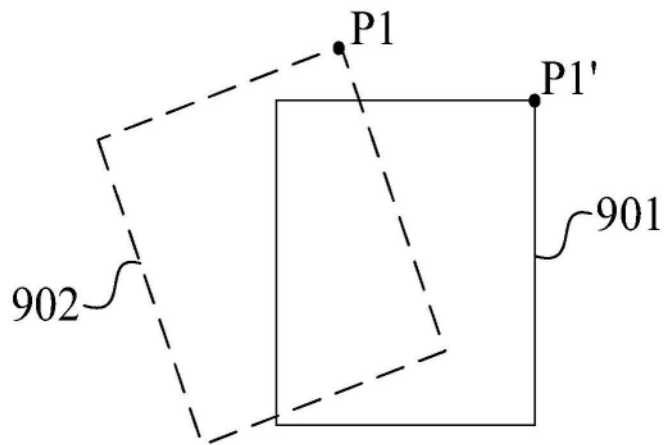


图9

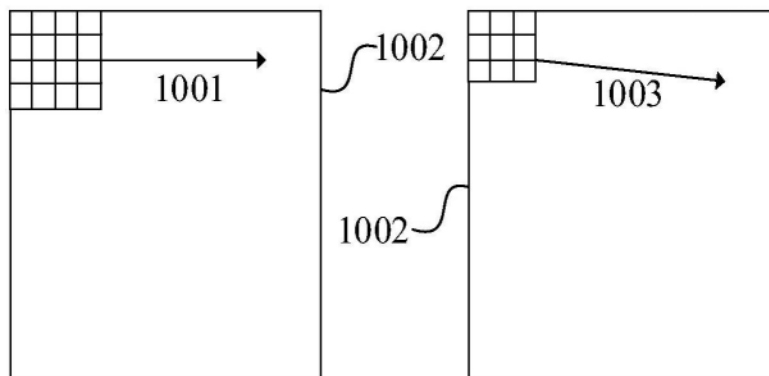


图10

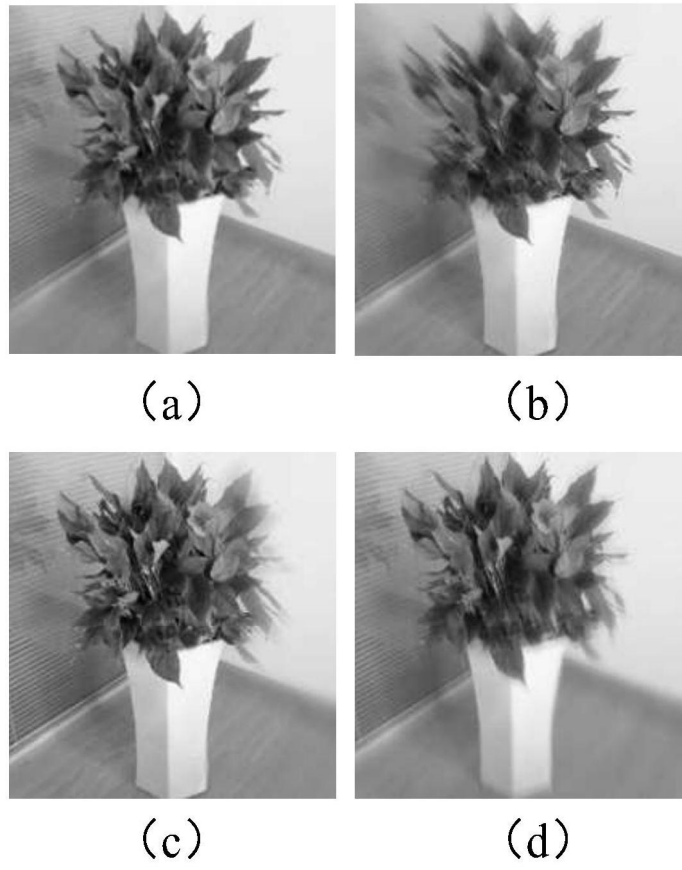


图11



图12

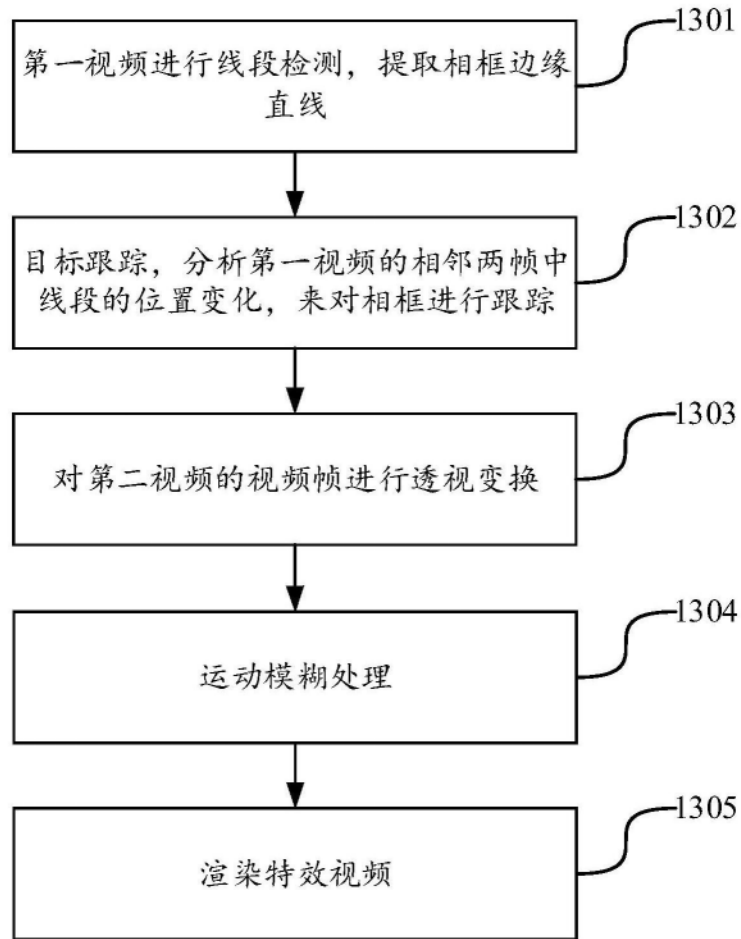
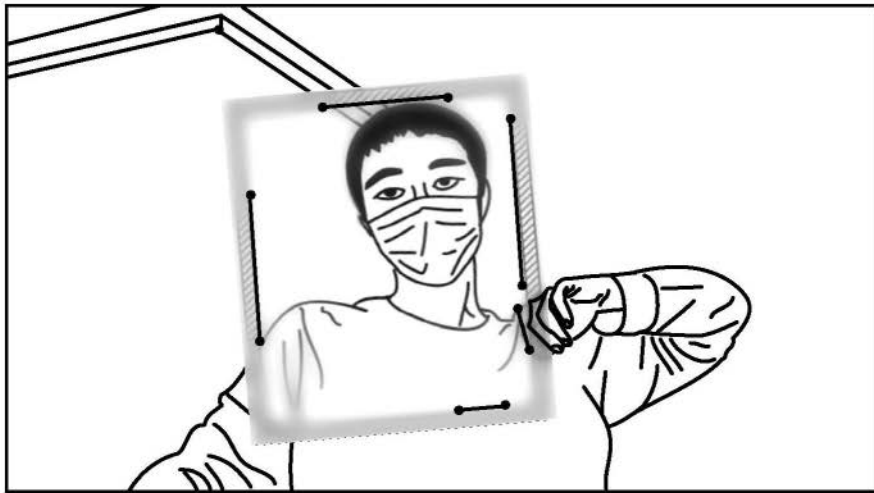
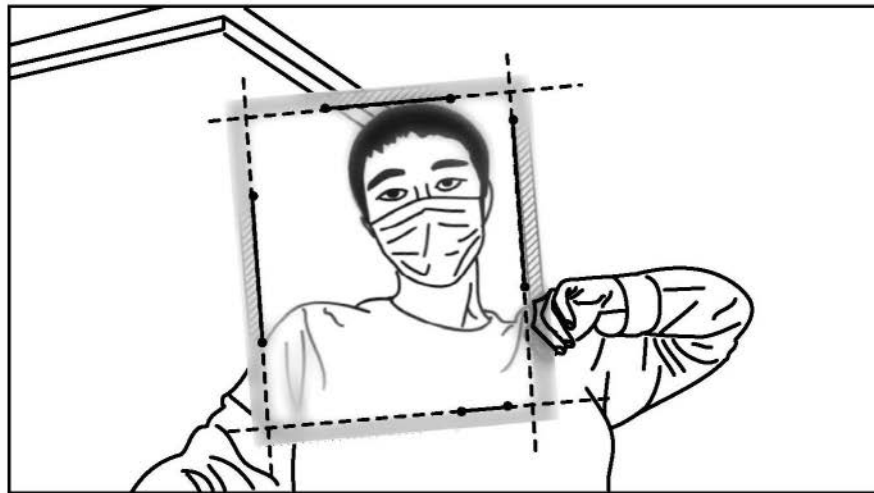


图13



(a)



(b)



(c)

图14

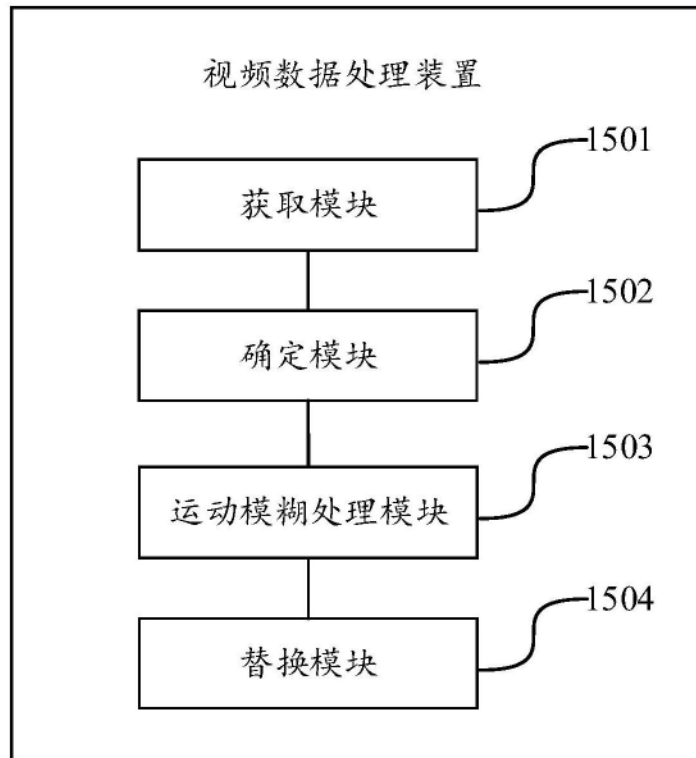


图15

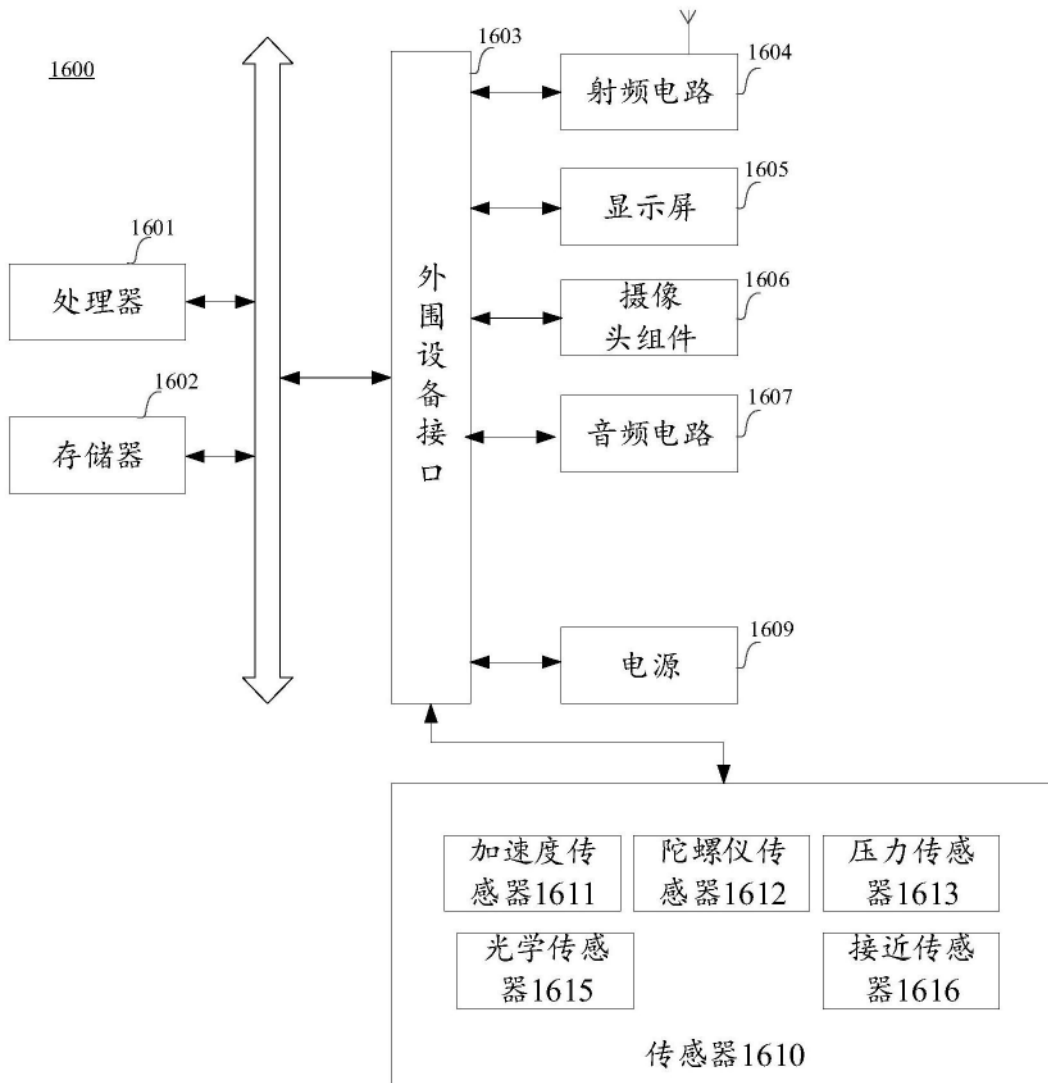


图16

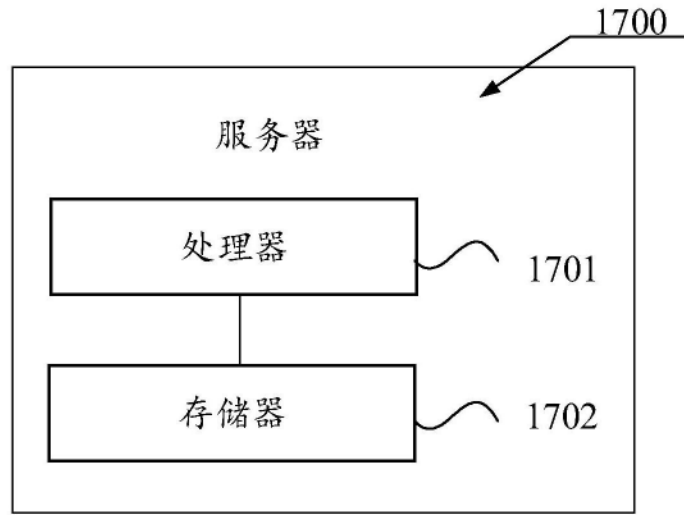


图17