



(12) 发明专利

(10) 授权公告号 CN 109076229 B

(45) 授权公告日 2022. 04. 15

(21) 申请号 201780028714.5

(22) 申请日 2017.05.09

(65) 同一申请的已公布的文献号
申请公布号 CN 109076229 A

(43) 申请公布日 2018.12.21

(30) 优先权数据
62/339,009 2016.05.19 US
15/589,782 2017.05.08 US

(85) PCT国际申请进入国家阶段日
2018.11.09

(86) PCT国际申请的申请数据
PCT/US2017/031789 2017.05.09

(87) PCT国际申请的公布数据
W02017/200804 EN 2017.11.23

(73) 专利权人 高通股份有限公司
地址 美国加利福尼亚州

(72) 发明人 王业奎 伏努·亨利

(74) 专利代理机构 永新专利商标代理有限公司
72002

代理人 赵腾飞

(51) Int.Cl.
H04N 19/176 (2006.01)
H04N 19/70 (2006.01)
H04N 19/132 (2006.01)
H04N 19/162 (2006.01)
H04N 21/81 (2006.01)
H04N 21/84 (2006.01)
H04N 21/845 (2006.01)
H04N 21/2343 (2006.01)
H04N 21/4728 (2006.01)

(56) 对比文件
WO 2015197815 A1, 2015.12.30
CN 101779465 A, 2010.07.14
CN 101690228 A, 2010.03.31
WO 2013077236 A1, 2013.05.30 (续)

审查员 杨少魁

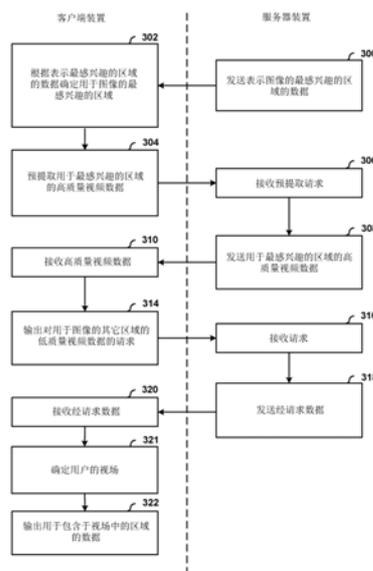
权利要求书7页 说明书26页 附图7页

(54) 发明名称

在图片中最感兴趣的区域

(57) 摘要

本发明提供一种用以确定用于视频数据的信息的装置,所述装置包含实施于电路中的一或多个处理器,所述一或多个处理器经配置以基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域。表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部。所述处理器输出对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求,且在输出对所述相对高质量视频数据的所述请求后,输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求。



CN 109076229 B

[接上页]

(56) 对比文件

M.M. Hannuksela等.Isolated regions in video coding.《IEEE Transactions on

Multimedia》.2004,全文.

刘欣等.基于感兴趣区域的航拍视频压缩编码技术综述.《电视技术》.2015,全文.

1. 一种确定用于视频数据的信息的方法,所述方法包括:

通过客户端装置的处理器基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域,所述处理器实施于电路中,表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部;

通过所述客户端装置的所述处理器输出对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求;

通过所述客户端装置的所述处理器接收用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据;

在输出对所述相对高质量视频数据的所述请求后,通过所述客户端装置的所述处理器输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求,对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示,其中,所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据;及

通过所述客户端装置的所述处理器接收用于所述图像的所述一或多个其它区域的所述相对低质量视频数据。

2. 根据权利要求1所述的方法,其进一步包括:

在输出对用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据的所述请求后,通过所述客户端装置的所述处理器基于视场FOV输出对用于随后图像的一或多个区域的相对高质量视频数据的请求。

3. 根据权利要求1所述的方法,其中确定所述一或多个最感兴趣的区域包括基于包含表示所述一或多个最感兴趣的区域的所述数据的补充增强信息SEI消息来确定所述一或多个最感兴趣的区域。

4. 根据权利要求1所述的方法,其中确定所述一或多个最感兴趣的区域包括基于包含表示所述一或多个最感兴趣的区域的所述数据的样本组逻辑框来确定所述一或多个最感兴趣的区域,所述样本组逻辑框包含于包含所述VCL数据的文件的文件格式标头信息中。

5. 根据权利要求1所述的方法,其中确定所述图像的所述多个区域中的所述一或多个最感兴趣的区域包括基于包含表示所述一或多个最感兴趣的区域的所述数据的计时元数据播放轨来确定所述图像的所述多个区域中的所述一或多个最感兴趣的区域。

6. 根据权利要求1所述的方法,其中确定所述一或多个最感兴趣的区域包括基于包含表示所述一或多个最感兴趣的区域的所述数据的HTTP动态自适应流式发射DASH媒体呈现描述MPD来确定所述一或多个最感兴趣的区域。

7. 根据权利要求1所述的方法,其中确定所述一或多个最感兴趣的区域包括基于表示所述一或多个最感兴趣的区域的所述数据的至少一个语法元素来确定所述一或多个最感兴趣的区域,所述语法元素具有表示所述一或多个最感兴趣的区域的数目的值。

8. 根据权利要求1所述的方法,其中确定所述一或多个最感兴趣的区域包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,确定表示相应最感兴趣的区域的位置及所述相应最感兴趣的区域的大小的多个语法元素的多个值。

9. 根据权利要求8所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的

所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,确定所述相应最感兴趣的区域上表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的循环。

10. 根据权利要求8所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,确定表示所述相应最感兴趣的区域的顶部的第一语法元素、表示所述相应最感兴趣的区域的底部的第二语法元素、表示所述相应最感兴趣的区域的左侧的第三语法元素及表示所述相应最感兴趣的区域的右侧的第四语法元素。

11. 根据权利要求8所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,确定表示所述相应最感兴趣的区域的左上角的第一语法元素及表示所述相应最感兴趣的区域的右下角的第二语法元素。

12. 根据权利要求8所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,确定表示所述相应最感兴趣的区域的左上角的第一语法元素、表示所述相应最感兴趣的区域的宽度的第二语法元素及表示所述相应最感兴趣的区域的高度的第三语法元素。

13. 一种用以确定用于视频数据的信息的装置,所述装置包括实施于电路中之一或多个处理器,所述一或多个处理器经配置以:

基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域,表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部;

输出对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求;

接收用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据;

在输出对所述相对高质量视频数据的所述请求后,输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求,对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示,其中,所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据;及

接收用于所述图像的形成所述图像的第二部分的所述一或多个其它区域的所述相对低质量视频数据。

14. 根据权利要求13所述的装置,其中为确定所述一或多个最感兴趣的区域,所述一或多个处理器经进一步配置以:

基于包含表示所述一或多个最感兴趣的区域的所述数据的样本组逻辑框来确定所述一或多个最感兴趣的区域,所述样本组逻辑框包含于包含所述VCL数据的文件的文件格式标头信息中。

15. 根据权利要求13所述的装置,其中为确定所述一或多个最感兴趣的区域,所述一或

多个处理器经进一步配置以：

基于包含表示所述一或多个最感兴趣的区域的所述数据的计时元数据播放轨来确定所述一或多个最感兴趣的区域。

16. 根据权利要求13的装置，其中为确定所述一或多个最感兴趣的区域，所述一或多个处理器经进一步配置以：

基于表示所述一或多个最感兴趣的区域的所述数据的至少一个语法元素来确定所述一或多个最感兴趣的区域，所述语法元素具有表示所述一或多个最感兴趣的区域的数目的值。

17. 根据权利要求13所述的装置，其中为确定所述一或多个最感兴趣的区域，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，确定所述一或多个最感兴趣的区域、表示相应最感兴趣的区域的位置及所述相应最感兴趣的区域的大小的多个语法元素的多个值。

18. 根据权利要求17所述的装置，其中为针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，确定所述相应最感兴趣的区域上表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的循环。

19. 根据权利要求17所述的装置，其中为针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，确定表示所述相应最感兴趣的区域的顶部的第一语法元素、表示所述相应最感兴趣的区域的底部的第二语法元素、表示所述相应最感兴趣的区域的左侧的第三语法元素及表示所述相应最感兴趣的区域的右侧的第四语法元素。

20. 根据权利要求17所述的装置，其中为针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域确定表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，确定表示所述相应最感兴趣的区域的左上角的第一语法元素、表示所述相应最感兴趣的区域的宽度的第二语法元素及表示所述相应最感兴趣的区域的高度的第三语法元素。

21. 一种用以确定用于视频数据的信息的装置，所述装置包括：

用于基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域的装置，表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部；

用于输出对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求的装置；

用于接收用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据

的装置；

用于在输出对所述相对高质量视频数据的所述请求后，输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求的装置，对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示，其中，所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据；及

用于接收用于所述图像的所述一或多个其它区域的所述相对低质量视频数据的装置。

22. 一种非暂时性计算机可读存储媒体，其上存储有在经执行时使得处理器进行以下操作的指令：

基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域，表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部；

输出对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求；

接收用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据；

在输出对所述相对高质量视频数据的所述请求后，输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求，对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示，其中，所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据；及

接收用于所述图像的形成所述图像的第二部分的所述一或多个其它区域的所述相对低质量视频数据。

23. 一种用信号表示用于视频数据的信息的方法，所述方法包括：

通过源装置的处理器确定视频数据图像的多个区域中的一或多个最感兴趣的区域，所述处理器实施于电路中，所述一或多个最感兴趣的区域包括客户端装置最可能检索的一或多个区域；

通过所述源装置的所述处理器产生表示所述一或多个最感兴趣的区域的数据，其中表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部；

通过所述源装置的所述处理器响应于对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求，输出用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据；及

通过所述源装置的所述处理器响应于在对所述相对高质量视频数据的所述请求之后接收的对用于所述图像的一或多个其它区域的相对低质量视频数据的请求，输出用于所述图像的所述一或多个其它区域的所述相对低质量视频数据，对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示，其中，所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据。

24. 根据权利要求23所述的方法，其中产生表示所述一或多个最感兴趣的区域的所述数据包括产生包含表示所述一或多个最感兴趣的区域的所述数据的补充增强信息SEI消息。

25. 根据权利要求23所述的方法，其中产生表示所述一或多个最感兴趣的区域的所述数据包括产生包含表示所述一或多个最感兴趣的区域的所述数据的样本组逻辑框，所述样

本组逻辑框包含于包含所述VCL数据的文件的文件格式标头信息中。

26. 根据权利要求23所述的方法,产生表示所述一或多个最感兴趣的区域的所述数据包括产生包含表示所述一或多个最感兴趣的区域的所述数据的计时元数据播放轨。

27. 根据权利要求23的方法,其中产生表示所述一或多个最感兴趣的区域的所述数据包括产生包含表示所述一或多个最感兴趣的区域的所述数据的HTTP动态自适应流式发射DASH媒体呈现描述MPD。

28. 根据权利要求23所述的方法,其中产生表示所述一或多个最感兴趣的区域的所述数据包括产生表示所述一或多个最感兴趣的区域的所述数据的至少一个语法元素,所述语法元素具有表示所述一或多个最感兴趣的区域的数目的值。

29. 根据权利要求23所述的方法,其中产生表示所述一或多个最感兴趣的区域的所述数据包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,产生表示相应最感兴趣的区域的位置及所述相应最感兴趣的区域的大小的多个语法元素的多个值。

30. 根据权利要求29所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,产生所述相应最感兴趣的区域上表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的循环。

31. 根据权利要求29所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,产生表示所述相应最感兴趣的区域的顶部的第一语法元素、表示所述相应最感兴趣的区域的底部的第二语法元素、表示所述相应最感兴趣的区域的左侧的第三语法元素及表示所述相应最感兴趣的区域的右侧的第四语法元素。

32. 根据权利要求29所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,产生表示所述相应最感兴趣的区域的左上角的第一语法元素及表示所述相应最感兴趣的区域的右下角的第二语法元素。

33. 根据权利要求29所述的方法,其中针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值包括针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,产生表示所述相应最感兴趣的区域的左上角的第一语法元素、表示所述相应最感兴趣的区域的宽度的第二语法元素及表示所述相应最感兴趣的区域的高度的第三语法元素。

34. 根据权利要求23所述的方法,其中确定所述一或多个最感兴趣的区域包括使用导演的意向或用户统计数据中的一或多个者来确定所述最感兴趣的区域。

35. 一种用以用信号表示用于视频数据的信息的装置,所述装置包括实施于电路中的一或多个处理器,所述一或多个处理器经配置以:

确定视频数据图像的多个区域中的一或多个最感兴趣的区域,所述一或多个最感兴趣

的区域包括客户端装置最可能检索的一或多个区域；

产生表示所述一或多个最感兴趣的区域的数据，其中表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部；

响应于对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求，输出用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据；及

响应于在对所述相对高质量视频数据的所述请求之后接收的对用于所述图像的一或多个其它区域的相对低质量视频数据的请求，输出用于所述图像的所述一或多个其它区域的所述相对低质量视频数据，对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示，其中，所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据。

36. 根据权利要求35所述的装置，其中为产生表示所述一或多个最感兴趣的区域的数据，所述一或多个处理器经进一步配置以：

产生包含表示所述一或多个最感兴趣的区域的所述数据的样本组逻辑框，所述样本组逻辑框包含于包含所述VCL数据的文件的文件格式标头信息中。

37. 根据权利要求35所述的装置，其中为产生表示所述一或多个最感兴趣的区域的数据，所述一或多个处理器经进一步配置以：

产生包含表示所述一或多个最感兴趣的区域的所述数据的计时元数据播放轨。

38. 根据权利要求35所述的装置，其中为产生表示所述一或多个最感兴趣的区域的数据，所述一或多个处理器经进一步配置以：

产生表示所述一或多个最感兴趣的区域的数量值。

39. 根据权利要求35所述的装置，其中为产生表示所述一或多个最感兴趣的区域的数据，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，产生表示相应最感兴趣的区域的位置及所述相应最感兴趣的区域的大小的多个语法元素的多个值。

40. 根据权利要求39所述的装置，其中为针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，产生所述相应最感兴趣的区域上表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的循环。

41. 根据权利要求39所述的装置，其中为针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区域的所述大小的所述多个语法元素的所述多个值，所述一或多个处理器经进一步配置以：

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域，产生表示所述相应最感兴趣的区域的顶部的第一语法元素、表示所述相应最感兴趣的区域的底部的第二语法元素、表示所述相应最感兴趣的区域的左侧的第三语法元素及表示所述相应最感兴趣的区域的右侧的第四语法元素。

42. 根据权利要求39所述的装置，其中为针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域产生表示所述相应最感兴趣的区域的所述位置及所述相应最感兴趣的区

域的所述大小的所述多个语法元素的所述多个值,所述一或多个处理器经进一步配置以:

针对所述一或多个最感兴趣的区域中的每一最感兴趣的区域,产生表示所述相应最感兴趣的区域的左上角的第一语法元素、表示所述相应最感兴趣的区域的宽度的第二语法元素及表示所述相应最感兴趣的区域的高度的第三语法元素。

43. 一种用以用信号表示用于视频数据的信息的装置,所述装置包括:

用于确定视频数据图像的多个区域中的一或多个最感兴趣的区域的装置,所述一或多个最感兴趣的区域包括客户端装置最可能检索的一或多个区域;

用于产生表示所述一或多个最感兴趣的区域的数据的装置,其中表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部;

用于响应于对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求,输出用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据的装置;及

用于响应于在对所述相对高质量视频数据的所述请求之后接收的对用于所述图像的一或多个其它区域的相对低质量视频数据的请求,输出用于所述图像的所述一或多个其它区域的所述相对低质量视频数据的装置,对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示,其中,所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据。

44. 一种非暂时性计算机可读存储媒体,其上存储有在经执行时使得处理器进行以下操作的指令:

确定视频数据图像的多个区域中的一或多个最感兴趣的区域,所述一或多个最感兴趣的区域包括客户端装置最可能检索的一或多个区域;

产生表示所述一或多个最感兴趣的区域的数据,其中表示所述一或多个最感兴趣的区域的所述数据在包含所述视频数据的位流的视频译码层VCL数据外部;

响应于对用于所述图像的所述一或多个最感兴趣的区域的相对高质量视频数据的请求,输出用于所述图像的所述一或多个最感兴趣的区域的所述相对高质量视频数据;及

响应于在对所述相对高质量视频数据的所述请求之后接收的对用于所述图像的一或多个其它区域的相对低质量视频数据的请求,输出用于所述图像的所述一或多个其它区域的所述相对低质量视频数据,对所述相对低质量视频数据的所述请求包括对所述图像的所述一或多个其它区域的指示,其中,所述相对高质量视频数据包括与所述相对低质量视频数据相比更高质量的视频数据。

在图片中最感兴趣的区域

[0001] 本申请案请求2016年5月19日提交的美国临时申请案第62/339,009号的权益,所述申请案在此以其全文引用的方式并入。

技术领域

[0002] 本发明涉及经编码视频数据的存储及输送。

背景技术

[0003] 数字视频能力可并入到广泛范围的装置中,包含数字电视、数字直播系统、无线广播系统、个人数字助理(PDA)、膝上型或台式计算机、数码相机、数字记录装置、数字媒体播放器、视频游戏装置、视频游戏控制台、蜂窝式或卫星无线电电话、视频电话会议装置及类似者。数字视频装置实施视频压缩技术,例如由MPEG-2、MPEG-4、ITU-TH.263或ITU-T H.264/MPEG-4第10部分高级视频译码(AVC)、ITU-T H.265(也称为高效视频译码(HEVC))定义的标准及这些标准的扩展中所描述的那些技术,以更有效地发射及接收数字视频信息。

[0004] 视频压缩技术执行空间预测及/或时间预测以减少或去除视频序列中所固有的冗余。对于基于块的视频译码,可将视频帧或切片分割成宏块或译码树单元(CTU)。在所述文件的剩余部分中,可互换地使用两个术语宏块与CTU。可进一步分割每一宏块。使用相对于相邻宏块的空间预测来编码经帧内译码(I)的帧或切片中的宏块。经帧间译码(P或B)帧或切片中的宏块可使用相对于同一帧或切片中的相邻宏块的空间预测或相对于其它参考帧的时间预测。

[0005] 在已编码视频数据之后,可将视频数据包化以供发射或存储。可将视频数据组译成符合多种标准中的任一者的视频文件,所述标准例如国际标准化组织(ISO)基本媒体文件格式及其扩展,例如AVC文件格式。

[0006] 在已包化视频数据以供发射或存储之后,客户端装置可以适合质量请求经包化视频数据。举例来说,客户端装置可在客户端装置与发射经包化视频的数字视频装置之间的带宽较高时以相对高质量请求经包化视频数据,且在客户端装置与发射经包化视频的数字视频装置之间的带宽较低时以相对低质量请求经包化视频数据。

发明内容

[0007] 一般来说,本发明描述用于产生包含表示视频数据图像的一或多个最感兴趣的区域的数据的位流的技术。图像的一或多个最感兴趣的区域与所述图象的其它区域相比在所述图象的呈现时间更可能呈现给用户。装置可使用一或多个最感兴趣的区域以预提取用于所述一或多个最感兴趣的区域的视频数据。举例来说,所述装置可以比用于所述图象的其它区域的视频数据更低的时延及/或更高的质量提取用于一或多个最感兴趣的区域的视频数据,其中所述装置可指代含有视频解码器的实体、例如内容递送网络(CDN)节点的网络元件或另一实体。

[0008] 在一个实例中,一种确定用于视频数据的信息的方法包含通过客户端装置的处理

器(所述处理器实施于电路中)基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的视频译码层(VCL)数据外部。所述方法进一步包含通过客户端装置的处理器输出对用于图像的一或多个最感兴趣的区域的相对高质量视频数据的请求,以及在输出对所述相对高质量视频数据的请求后,通过客户端装置的处理器输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求。

[0009] 在另一实例中,一种用以确定用于视频数据的信息的装置包含实施于电路中的一或多个处理器,所述一或多个处理器经配置以基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。所述一或多个处理器经进一步配置以输出对用于图像的一或多个最感兴趣的区域的相对高质量视频数据的请求,且在输出对所述相对高质量视频数据的请求后输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求。

[0010] 在另一实例中,一种用以确定用于视频数据的信息的装置包含用于基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域的装置。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。所述装置进一步包含用于输出对用于图像的一或多个最感兴趣的区域的相对高质量视频数据的请求的装置,以及用于在输出对所述相对高质量视频数据的请求后输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求的装置。

[0011] 在另一实例中,一种计算机程序产品包含包括指令的计算机可读存储媒体,所述指令在经执行时使得用以确定用于视频数据的信息的源装置的处理器基于表示视频数据图像的多个区域中的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。所述指令进一步使得所述处理器输出对用于图像的一或多个最感兴趣的区域的相对高质量视频数据的请求,且在输出对用于所述相对高质量视频数据的请求后输出对所述图像的一或多个其它区域的相对低质量视频数据的请求。

[0012] 在另一实例中,一种用信号表示用于视频数据的信息的方法包含通过源装置的处理器(所述处理器实施于电路中)确定视频数据图像的多个区域中一或多个最感兴趣的区域。所述一或多个最感兴趣的区域包括最可能由客户端装置检索的一或多个区域。所述方法进一步包含通过源装置的处理器产生表示所述一或多个最感兴趣的区域的数据。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。

[0013] 在另一实例中,一种用以用信号表示用于视频数据的信息的装置包含实施于电路中的一或多个处理器,所述一或多个处理器经配置以确定视频数据图像的多个区域中的一或多个最感兴趣的区域。所述一或多个最感兴趣的区域包含最可能由客户端装置检索的区域。所述一或多个处理器经进一步配置以产生表示所述一或多个最感兴趣的区域的数据。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。

[0014] 在另一实例中,一种用以用信号表示用于视频数据的信息的装置包含用于确定视频数据图像的多个区域中的一或多个最感兴趣的区域的装置。所述一或多个最感兴趣的区域包括最可能由客户端装置检索的一或多个区域。所述装置进一步包含用于产生表示所述

一或多个最感兴趣的区域的数据的装置。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。

[0015] 在另一实例中,一种计算机程序产品包含包括指令的计算机可读存储媒体,所述指令使得处理器确定视频数据图像的多个区域中的一或多个最感兴趣的区域。所述一或多个最感兴趣的区域包括最可能由客户端装置检索的一或多个区域。所述指令进一步使得所述处理器产生表示所述一或多个最感兴趣的区域的数据。表示所述一或多个最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。

[0016] 在以下附图及描述中阐述一或多个实例的细节。其它特征、目标及优点将根据所述描述及图式以及权利要求书显而易见。

附图说明

[0017] 图1为说明实施用于经由网络来流式发射媒体数据的技术的实例系统的框图。

[0018] 图2为说明实例多媒体内容的元素的概念图。

[0019] 图3为说明实例视频文件的元素的概念图,所述实例视频文件可对应于图2的表示。

[0020] 图4为以立方体投影说明实例图像块译码的概念图。

[0021] 图5为说明根据本发明中所描述的一或多种技术的用以确定用于视频数据的信息的实例方法的流程图。

[0022] 图6为说明根据本发明中所描述的一或多种技术的用以用信号表示用于视频数据的信息的实例方法的流程图。

[0023] 图7为说明根据本发明中所描述的一或多种技术的用于预提取视频数据的实例方法的流程图。

具体实施方式

[0024] 本发明涉及虚拟实境 (VR)。更具体来说,本发明揭示用于创建VR视频中一或多个最感兴趣的区域的集合,用信号表示最感兴趣的区域的信息,以及使用所述信息来编码、流式发射、播放及管理VR视频内容的一些方法。

[0025] 一般来说,本发明描述用于产生包含表示视频数据图像中一或多个最感兴趣的区域的数据的视频内容的技术。导演或制作人可选择一或多个最感兴趣的区域,或计算装置可根据服务提供者或内容提供者提供的用户统计数据推导一或多个最感兴趣的区域,例如在经由流式发射服务提供全向视频内容时用户请求/观看最多的区域的统计数据。在任何情况下,客户端装置可使用一或多个最感兴趣的区域来预提取用于所述一或多个最感兴趣的区域的视频数据以改进视频的播放,所述视频例如虚拟实境视频,其也被称作360视频。

[0026] 具体来说,客户端装置可(例如)使用在用于发射位流的视频译码层数据的标头或消息中接收的信息来确定表示一或多个最感兴趣的区域的数据。在根据表示所述一或多个最感兴趣的区域的数据确定所述一或多个最感兴趣的区域后,客户端装置可以相对高质量请求一或多个最感兴趣的区域。在以相对高质量请求一或多个最感兴趣的区域后,客户端装置可基于用户的视场 (FOV) 以比高质量相对更低的质量请求所述图像的一或多个其它区域。以这种方式,一或多个最感兴趣的区域可经预提取(例如,在接收用户的FOV前经提取)以准许以高质量接收一或多个最感兴趣的区域,借此改进视频(例如虚拟实境视频)的播

放。即,装置可使用一或多个最感兴趣的区域来预提取用于所述一或多个最感兴趣的区域的视频数据。举例来说,所述装置可以比用于所述图像的其它区域的视频数据更低的时延及/或更高的质量提取用于一或多个最感兴趣的区域的视频数据,其中所述装置可指代含有视频解码器的实体、例如内容递送网络(CDN)节点的网络元件或另一实体。

[0027] 本发明的技术可应用于符合根据以下中的任一者封装的视频数据的视频文件:ISO基本媒体文件格式、可调式视频译码(SVC)文件格式、高级视频译码(AVC)文件格式、第三代合作伙伴计划(3GPP)文件格式及/或多视图视频译码(MVC)文件格式或其它类似视频文件格式。

[0028] 在HTTP流式发射中,经常使用的操作包含HEAD、GET及部分GET。HEAD操作检索与给定统一资源定位符(URL)或统一资源名称(URN)相关联的文件的标头,但不检索与所述URL或URN相关联的有效负载。GET操作检索与给定URL或URN相关联的整个文件。部分GET操作接收位组范围作为输入参数且检索文件的连续数目个位组,其中位组数目对应于所接收的位组范围。因此,可提供电影片段以用于HTTP流式发射,这是因为部分GET操作可得到一或多个单独的电影片段。在一电影片段中,可存在具有不同播放轨的若干播放轨片段。在HTTP流式发射中,媒体呈现可为客户端可存取的数据的结构化集合。客户端可请求且下载媒体数据信息以向用户呈现流式发射服务。

[0029] 在使用HTTP流式发射来流式发射3GPP数据的实例中,可能存在用于多媒体内容的视频及/或音频数据的多个表示。如下文所解释,不同表示可对应于不同译码特征(例如,视频译码标准的不同配置文件或层级)、不同译码标准或译码标准的扩展(例如多视图及/或可调式扩展)或不同位速率。这些表示的清单可在媒体呈现描述(MPD)数据结构中加以定义。媒体呈现可对应于HTTP流式发射客户端装置可存取的数据的结构化集合。HTTP流式发射客户端装置可请求且下载媒体数据信息以向客户端装置的用户呈现流式发射服务。媒体呈现可在MPD数据结构中加以描述,所述数据结构可包含MPD的更新。

[0030] 媒体呈现可含有一或多个周期的序列。每一周期可延长到下一周期的开始,或在最后一周期的情况下延长到媒体呈现的结束。每一周期可含有同一媒体内容的一或多个表示。表示可为音频、视频、计时文本或其它这类数据的多种替代性经编码版本中的一者。各表示的区别可在于编码类型,例如,用于视频数据的位速率、分辨率及/或编解码器以及用于音频数据的位速率、语言及/或编解码器。术语表示可用于指经编码音频或视频数据中对应于多媒体内容的特定周期且以特定方式编码的部分。

[0031] 特定周期的表示可经指配到由MPD中指示所述表示所属的适应集合的属性指示的群组。同一适应集合中的表示通常被视为彼此的替代者,这是因为客户端装置可在这些表示之间动态且无缝地切换例如以执行带宽适应。举例来说,特定周期的视频数据的每一表示可经指配到同一适应集合,以使得可选择所述表示中的任一者进行解码以呈现对应周期的多媒体内容的媒体数据(例如视频数据或音频数据)。在一些实例中,一个周期内的媒体内容可由来自群组0(如果存在)的一个表示来表示或由来自每一非零群组的至多一个表示的组合来表示。一周期的每一表示的时序数据可相对于所述周期的开始时间来表达。

[0032] 一表示可包含一或多个区段。每一表示可包含初始化区段,或表示中的每一区段可自初始化。当存在时,初始化区段可含有用于存取表示的初始化信息。一般来说,初始化区段不含有媒体数据。区段可由识别符唯一地参考,所述识别符例如统一资源定位符

(URL)、统一资源名称 (URN) 或统一资源识别符 (URI)。MPD可提供用于每一区段的识别符。在一些实例中,MPD还可提供呈范围属性的形式的位组范围,其可对应于可通过URL、URN或URI存取的文件内的区段的数据。

[0033] 可选择不同表示以用于大体上同时检索不同类型的媒体数据。举例来说,客户端装置可选择从其中检索区段的音频表示、视频表示及计时文本表示。在一些实例中,客户端装置可选择特定适应集合以用于执行带宽适应。即,客户端装置可选择包含视频表示的适应集合、包含音频表示的适应集合及/或包含计时文本的适应集合。替代地,客户端装置可选择用于某些类型的媒体(例如,视频)的适应集合,而直接选择用于其它类型的媒体(例如,音频及/或计时文本)的表示。

[0034] 图1为说明实施用于经由网络来流式发射媒体数据的技术的实例系统10的框图。在这个实例中,系统10包含内容准备装置20、服务器装置60及客户端装置40。客户端装置40与服务器装置60通过网络74以通信方式耦合,所述网络可包括因特网。在一些实例中,内容准备装置20与服务器装置60还可通过网络74或另一网络耦合,或可直接以通信方式耦合。在一些实例中,内容准备装置20与服务器装置60可包括相同装置。

[0035] 在图1的实例中,内容准备装置20包括音频源22及视频源24。音频源22可包括(例如)麦克风,所述麦克风产生表示待通过音频编码器26编码的所获取音频数据的电信号。替代地,音频源22可包括:存储媒体,其存储先前记录的音频数据;音频数据产生器,例如计算机化的合成器;或任何其它音频数据源。视频源24可包括:摄像机,其产生待通过视频编码器28编码的视频数据;存储媒体,其编码有先前记录的视频数据;视频数据产生单元,例如计算机图形源;或任何其它视频数据源。内容准备装置20未必在所有实例中均以通信方式耦合到服务器装置60,而是可将多媒体内容存储到由服务器装置60读取的单独媒体。

[0036] 原始音频及视频数据可包括模拟或数字数据。模拟数据可在通过音频编码器26及/或视频编码器28编码之前被数字化。音频源22可在说话参与者正在说话时从说话参与者获得音频数据,且视频源24可同时获得说话参与者的视频数据。在其它实例中,音频源22可包括有包括所存储音频数据的计算机可读存储媒体,且视频源24可包括有包括所存储视频数据的计算机可读存储媒体。以这种方式,本发明中所描述的技术可应用于实况、流式发射、实时音频及视频数据或所存档的、预先记录的音频及视频数据。

[0037] 对应于视频帧的音频帧通常为含有通过音频源22获取(或产生)的音频数据的音频帧,所述音频数据同时伴随含于视频帧内的通过视频源24获取(或产生)的视频数据。举例来说,当说话参与者通常通过说话而产生音频数据时,音频源22获取音频数据,且视频源24同时(即,在音频源22正获取音频数据的同时)获取说话参与者的视频数据。因此,音频帧在时间上可对应于一或多个特定视频帧。因此,对应于视频帧的音频帧通常对应于同时获取音频数据及视频数据的情形,且因此音频帧及视频帧分别包括同时获取到的音频数据及视频数据。

[0038] 在一些实例中,音频编码器26可对每一经编码音频帧中表示记录经编码音频帧的音频数据的时间的时戳进行编码,且类似地,视频编码器28可对每一经编码视频帧中表示记录经编码视频帧的视频数据的时间的时戳进行编码。在这些实例中,对应于视频帧的音频帧可包括有包括时戳的音频帧及包括相同时戳的视频帧。内容准备装置20可包含内部时钟,音频编码器26及/或视频编码器28可根据所述内部时钟产生时戳,或音频源22及视频源

24可使用所述内部时钟分别使音频数据及视频数据与时戳相关联。

[0039] 在一些实例中,音频源22可向音频编码器26发送对应于记录音频数据的时间的数据,且视频源24可向视频编码器28发送对应于记录视频数据的时间的数据。在一些实例中,音频编码器26可对经编码音频数据中的序列识别符进行编码以指示经编码音频数据的相对时间排序,但未必指示记录音频数据的绝对时间,且相似地,视频编码器28还可使用序列识别符来指示经编码视频数据的相对时间排序。类似地,在一些实例中,序列识别符可映射或以其它方式与时戳相关联。

[0040] 音频编码器26通常产生经编码音频数据的流,而视频编码器28产生经编码视频数据的流。每一个别数据流(不论音频或视频)可被称为基本流。基本流为表示的单一的经数字译码(可能经压缩)的分量。举例来说,表示的经译码视频或音频部分可为基本流。基本流可在被封装于视频文件内之前被转换成包化基本流(PES)。在相同表示内,可使用流ID来区分属于一个基本流的PES包与属于其它基本流的PES包。基本流的数据的基本单位为包化基本流(PES)包。因此,经译码视频数据大体对应于基本视频流。类似地,音频数据对应于一或多个相应基本流。

[0041] 许多视频译码标准(例如,ITU-T H.264/AVC及即将到来的高效视频译码(HEVC)标准)定义无误差位流的语法、语义及解码过程,所述无误差位流中的任一者符合特定配置文件或层级。视频译码标准通常并不指定编码器,但编码器具有保证所产生的位流对于解码器而言是标准相容的任务。在视频译码标准的上下文中,“配置文件”对应于算法、特征或工具及施加到算法、特征或工具的约束的子集。如由(例如)H.264标准所定义,“配置文件”为由H.264标准指定的完整位流语法的子集。“层级”对应于解码器资源消耗(例如,解码器存储器及计算)的限制,所述限制涉及图片分辨率、位速率及块处理速率。配置文件可通过profile_idc(配置文件指示符)值用信号通知,而层级可通过level_idc(层级指示符)值用信号通知。

[0042] 举例来说,H.264标准认为,在给定配置文件的语法所强加的界限内,仍然可能要求编码器及解码器的性能有较大变化,这取决于位流中的语法元素所取的值(例如经解码图片的指定大小)。H.264标准进一步认为,在许多应用中,实施能够处理特定配置文件内的语法的所有假设使用的解码器既不实际也不经济。因此,H.264标准将“层级”定义为强加于位流中的语法元素的值的约束的指定集合。这些约束可为对值的简单限制。替代地,这些约束可采取对值的算术组合约束的形式(例如,图片宽度乘以图片高度乘以每秒解码的图片数目)。H.264标准进一步规定,个别实施对于每一所支持配置文件可支持不同层级。

[0043] 符合配置文件的解码器一般支持配置文件中所定义的所有特征。举例来说,作为译码特征,B图片译码在H.264/AVC的基线配置文件中不被支持,但在H.264/AVC的其它配置文件中被支持。符合一层级的解码器应能够对不需要超出所述层级中所定义的限制的资源的任何位流进行解码。配置文件及层级的定义可对可解译性有帮助。举例来说,在视频发射期间,可针对整个发射工作阶段协商及同意一对配置文件定义及层级定义。更具体来说,在H.264/AVC中,层级可定义对于需要处理的宏块的数目、经解码图片缓冲器(DPB)大小、经译码图片缓冲器(CPB)大小、竖直运动向量范围、每两个连续MB的运动向量的最大数目及B块是否可具有小于 8×8 像素的子宏块分区的限制。以这种方式,解码器可确定解码器是否能够恰当地对位流进行解码。

[0044] 在图1的实例中,内容准备装置20的封装单元30从视频编码器28接收包括经译码视频数据的基本流,且从音频编码器26接收包括经译码音频数据的基本流。在一些实例中,视频编码器28及音频编码器26可各自包含用于从经编码数据形成PES包的包化器。在其它实例中,视频编码器28及音频编码器26可各自与用于从经编码数据形成PES包的相应包化器介接。在另外其它实例中,封装单元30可包含用于从经编码音频及视频数据形成PES包的包化器。

[0045] 视频编码器28可以多种方式对多媒体内容的视频数据进行编码,以产生多媒体内容的呈各种位速率且具有各种特征的不同表示,所述特征例如,像素分辨率、帧速率、对各种译码标准的符合性、对各种译码标准的各种配置文件及/或配置文件层级的符合性、具有一或多个视图的表示(例如,用于二维或三维播放)或其它这些特征。如本发明中所使用,表示可包括音频数据、视频数据、文本数据(例如,用于封闭字幕)或其它这类数据中的一者。表示可包含例如音频基本流或视频基本流的基本流。每一PES包可包含stream_id,其识别PES包所属的基本流。封装单元30负责将基本流组译成各种表示的视频文件(例如,区段)。

[0046] 封装单元30从音频编码器26及视频编码器28接收表示的基本流的PES包且从所述PES包形成对应的网络抽象层(NAL)单元。经译码视频区段可组织为NAL单元,其提供满足例如视频电话、存储、广播或流式发射等应用的“网络友好”视频表示。NAL单元可分类为视频译码层(VCL)NAL单元(例如,VCL数据)及非VCL NAL单元。VCL单元(也被称作VCL数据)可含有核心压缩引擎且可包含块、宏块及/或切片层级数据。其它NAL单元可为非VCL NAL单元。在一些实例中,一个时间实例中的经译码图片(通常呈现为初级经译码图片)可含于存取单元中,所述存取单元可包含一或多个NAL单元。

[0047] 非VCL NAL单元可尤其包含参数集NAL单元及SEI NAL单元。参数集可含有序列层级标头信息(在序列参数集(SPS)中)及不频繁改变的图片层级标头信息(在图片参数集(PPS)中)。就参数集(例如,PPS及SPS)而言,不需要针对每一序列或图片重复不频繁改变的信息,因此可改进译码效率。此外,使用参数集可实现重要标头信息的带外发射,从而避免对于用于抗误码的冗余发射的需要。在带外发射实例中,参数集NAL单元可在与其它NAL单元(例如SEI NAL单元)不同的信道上发射。

[0048] 补充增强信息(SEI)可含有对于对来自VCL NAL单元的经译码图片样本进行解码并非必需的信息,但可辅助与解码、显示、抗误码及其它目的相关的过程。SEI消息可含于非VCL NAL单元中。SEI消息为一些标准规范的标准化部分,且因此对于标准相容的解码器实施并非始终是必选的。SEI消息可为序列层级SEI消息或图片层级SEI消息。一些序列层级信息可含于SEI消息中,例如SVC的实例中的可调式性信息SEI消息,及MVC中的视图可调式性信息SEI消息。这些实例SEI消息可传达关于(例如)操作点的检索及操作点的特征的信息。另外,封装单元30可形成清单文件,例如描述表示的特征的媒体呈现描述符(MPD)。封装单元30可根据可延伸标记语言(XML)来格式化MPD。

[0049] 封装单元30可向输出接口32提供多媒体内容的一或多个表示的数据以及清单文件(例如,MPD)。输出接口32可包括网络接口或用于写入到存储媒体的接口,例如通用流式发射总线(USB)接口、CD或DVD写入器或烧录器、到磁性或快闪存储媒体的接口,或用于存储或发射媒体数据的其它接口。封装单元30可向输出接口32提供多媒体内容的表示中的每一者的数据,所述输出接口可经由网络发射或存储媒体向服务器装置60发送所述数据。在图1

的实例中,服务器装置60包含存储各种多媒体内容64的存储媒体62,每一多媒体内容包含相应清单文件66及一或多个表示68A到68N(表示68)。在一些实例中,输出接口32还可向网络74直接发送数据。

[0050] 在一些实例中,表示68可分成若干适应集合。即,表示68的各种子集可包含相应共同特征集合,例如编解码器、配置文件及层级、分辨率、视图的数目、区段的文件格式、可识别待与待解码及呈现的表示及/或音频数据(例如,由扬声器发出)一起显示的文本的语言或其它特征的文本类型信息、可描述适应集合中的表示的场景的相机角度或现实世界相机视角的相机角度信息、描述对于特定观众的内容适合性的分级信息,或其类似信息。

[0051] 清单文件66可包含指示对应于特定适应集合的表示68的子集以及所述适应集合的共同特征的数据。清单文件66还可包含表示适应集合的个别表示的个别特征(例如,位速率)的数据。以这种方式,适应集合可提供简化的网络带宽调适。适应集合中的表示可使用清单文件66的适应集合元素的子代元素来加以指示。

[0052] 服务器装置60包含请求处理单元70及网络接口72。在一些实例中,服务器装置60可包含多个网络接口。此外,服务器装置60的特征中的任一者或全部可在内容递送网络的其它装置(例如,路由器、桥接器、代理装置、交换器或其它装置)上实施。在一些实例中,内容递送网络的中间装置可高速缓存多媒体内容64的数据,且包含基本上符合服务器装置60的组件的组件。一般来说,网络接口72经配置以经由网络74来发送及接收数据。

[0053] 请求处理单元70经配置以从客户端装置(例如,客户端装置40)接收对存储媒体62的数据的网络请求。举例来说,请求处理单元70可实施超文本传送协议(HTTP)版本1.1,如RFC 2616,R.Fielding等人(网络工作组,IETF,1999年6月)的“超文本传送协议-HTTP/1.1(Hypertext Transfer Protocol-HTTP/1.1)”中所描述。即,请求处理单元70可经配置以接收HTTP GET或部分GET请求,且响应于所述请求而提供多媒体内容64的数据。请求可指定表示68中的一者的区段,例如使用区段的URL。在一些实例中,所述请求还可指定区段的一或多个位组范围,因此包括部分GET请求。请求处理单元70可经进一步配置以服务于HTTP HEAD请求以提供表示68中的一者的区段的标头数据。在任何状况下,请求处理单元70可经配置以处理所述请求以将所请求的数据提供到请求装置(例如客户端装置40)。

[0054] 另外或替代地,请求处理单元70可经配置以经由例如eMBMS的广播或多播协议来传递媒体数据。内容准备装置20可以与所描述大体上相同的方式产生DASH区段及/或子区段,但服务器装置60可使用eMBMS或另一广播或多播网络输送协议来递送这些区段或子区段。举例来说,请求处理单元70可经配置以从客户端装置40接收多播群组加入请求。即,服务器装置60可向与特定媒体内容(例如,实况事件的广播)相关联的客户端装置(包含客户端装置40)通告与多播群组相关联的因特网协议(IP)地址。客户端装置40又可提交加入多播群组的请求。所述请求可在整个网络74(例如,组成网络74的路由器)中传播,以使得致使所述路由器将预定用于与多播群组相关联的IP地址的业务导向到订用的客户端装置(例如客户端装置40)。

[0055] 请求处理单元70可经配置以从请求组译数据。举例来说,请求处理单元70可根据来自客户端装置的每一请求检索已经请求的图像的一或多个区域的指示。举例来说,请求处理单元70可根据来自客户端装置40的请求检索特定图像的用户视场。在检索之后,请求处理单元70可在表或数据库中产生针对每一图像指示客户端装置(例如,客户端装置40)所

请求的用户视场的一或多个区域的条目。

[0056] 请求处理单元70可根据经组译数据推导一或多个最感兴趣的区域。更具体来说,例如,响应于确定图像的一或多个区域与所述图像的其它区域相比为所述图像的最频繁请求区域,请求处理单元70可选择所述一或多个区域作为所述图像的一或多个最感兴趣的区域。在一些实例中,当一或多个区域以满足阈值的频率包含于用户视场中时,请求处理单元70可选择所述一或多个区域作为所述图像的一或多个最感兴趣的区域。举例来说,当经组译数据指示至少一半用户视场包含一或多个区域时,请求处理单元70可选择所述一或多个区域作为所述图像的一或多个最感兴趣的区域。

[0057] 如图1的实例中所说明,多媒体内容64包含清单文件66,其可对应于媒体呈现描述(MPD)。清单文件66可含有不同替代表示68(例如,具有不同质量的视频服务)的描述,且所述描述可包含(例如)编解码器信息、配置文件值、层级值、位速率及表示68的其它描述性特征。客户端装置40可检索媒体呈现的MPD以确定如何存取表示68的区段。

[0058] 特定而言,检索单元52可检索客户端装置40的配置数据(未图示)以确定视频解码器48的解码能力及视频输出44的呈现能力。配置数据还可包含由客户端装置40的用户选择的语言偏好、对应于由客户端装置40的用户设定的深度偏好的一或多个相机视角及/或由客户端装置40的用户选择的分级偏好中的任一者或全部。检索单元52可包括(例如)网页浏览器或媒体客户端,其经配置以提交HTTP GET及部分GET请求。检索单元52可对应于由客户端装置40的一或多个处理器或处理单元(未图示)执行的软件指令。在一些实例中,关于检索单元52所描述的功能性的全部或部分可在硬件或硬件、软件及/或固件的组合中实施,其中可提供必需的硬件以执行软件或固件的指令。

[0059] 检索单元52可将客户端装置40的解码及呈现能力与由清单文件66的信息指示的表示68的特征进行比较。检索单元52可首先检索清单文件66的至少一部分以确定表示68的特征。举例来说,检索单元52可请求描述一或多个适应集合的特征的清单文件66的一部分。检索单元52可选择具有客户端装置40的译码及呈现能力可满足的特征的表示68(例如,适应集合)的子集。检索单元52可接着确定适应集合中的表示的位速率,确定网络带宽的当前可用量,且从具有网络带宽可满足的位速率的表示中的一者检索区段。

[0060] 一般来说,较高位速率表示可产生较高质量的视频播放,而较低位速率表示可在可用网络带宽减少时提供足够质量的视频播放。因此,当可用网络带宽相对较高时,检索单元52可从相对高位速率的表示检索数据,而当可用网络带宽较低时,检索单元52可从相对低位速率的表示检索数据。以这种方式,客户端装置40可经由网络74流式发射多媒体数据,同时也适应网络74的改变的网络带宽可用性。

[0061] 另外或替代地,检索单元52可经配置以根据例如eMBMS或IP多播的广播或多播网络协议来接收数据。在这些实例中,检索单元52可提交加入与特定媒体内容相关联的多播网络群组的请求。在加入多播群组之后,检索单元52可在无发布到服务器装置60或内容准备装置20的其它请求的情况下接收多播群组的数据。当不再需要多播群组的数据时,检索单元52可提交离开所述多播群组的请求,例如以停止播放或将信道改变到不同多播群组。

[0062] 网络接口54可接收所选表示的区段的数据且向检索单元52提供所述数据,所述检索单元又可向解封装单元50提供所述区段。解封装单元50可将视频文件的元素解封装成组成性PES流,解包化所述PES流以检索经编码数据,且取决于经编码数据为音频流或视频流

的部分(例如,如由流的PES包头所指示)而将经编码数据发送到音频解码器46或视频解码器48。音频解码器46对经编码音频数据进行解码,且向音频输出42发送经解码音频数据,而视频解码器48对经编码视频数据进行解码,且向视频输出44发送经解码视频数据,其可包含流的多个视图。

[0063] 视频编码器28、视频解码器48、音频编码器26、音频解码器46、封装单元30、检索单元52及解封装单元50各自可实施为适用的多种合适处理电路中的任一者,合适处理电路例如一或多个微处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)、离散逻辑电路、软件、硬件、固件或其任何组合。视频编码器28及视频解码器48中的每一者可包含于一或多个编码器或解码器中,编码器或解码器中的任一者可经集成为组合式视频编码器/解码器(CODEC)的部分。同样地,音频编码器26及音频解码器46中的每一者可包含于一或多个编码器或解码器中,编码器或解码器中的任一者可经集成为组合式CODEC的部分。包含视频编码器28、视频解码器48、音频编码器26、音频解码器46、封装单元30、检索单元52及/或解封装单元50的装置可包括集成电路、微处理器及/或无线通信装置,例如蜂窝式电话。

[0064] 视频译码标准包含:ITU-T H.261;ISO/IEC MPEG-1Visual;ITU-T H.262或ISO/IEC MPEG-2Visual;ITU-T H.263;ISO/IEC MPEG-4Visual;ITU-T H.264或ISO/IEC MPEG-4 AVC,包含其可调式视频译码(SVC)及多视图视频译码(MVC)扩展;及也被称作ITU-T H.265及ISO/IEC 23008-2的高效视频译码(HEVC),包含其可调式译码扩展(即,可调式高效视频译码SHVC)及多视图扩展(即,多视图高效视频译码MV-HEVC)以及3D扩展(即,3D高效视频译码3D-HEVC)。MVC的一个联合草案描述于2010年3月的“用于一般视听服务的先进视频译码(Advanced video coding for generic audiovisual services)”,ITU-T建议H.264中。

[0065] 封装单元30可确定图像的一或多个最感兴趣的区域。如本文中所使用,最感兴趣的区域可指代图像中用户查看的可能性较高的区域。举例来说,最感兴趣的区域可指代图像中导演或制作人选择的区域。在一些实例中,最感兴趣的区域可指代图像中通过计算装置根据服务提供者或内容提供者提供的用户统计数据所推导的区域,例如在经由流式发射服务提供全向视频内容时用户请求/观看最多的区域的统计数据。

[0066] 封装单元30可封装表示一或多个最感兴趣的区域的数据。在一些实例中,封装单元30可封装在VCL数据外部的表示一或多个最感兴趣的区域的数据,所述数据可包含供视频译码引擎使用的实际视频数据。在一些实例中,表示一或多个最感兴趣的区域的数据可包含于位流中。举例来说,封装单元30可封装包含在VCL数据外部的表示一或多个最感兴趣的区域的数据的SEI消息。在一些实例中,封装单元30可封装包含在视频位流外部(且因此也在位流的VCL数据外部)的表示一或多个最感兴趣的区域的数据的文件格式数据。举例来说,封装单元30可封装包含在VCL数据外部的表示一或多个最感兴趣的区域的数据的逻辑框(例如,样本到群组)或计时元数据播放轨。

[0067] 就VR视频应用而言,当考虑利用关于最感兴趣的区域的信息的方式时,我们发现以下问题中的至少一些:

[0068] 1. 如果用户实际上对VR视频中的其它区域感兴趣,那么在没有用户控制的情况下使用关于最感兴趣的区域的信息来显示最感兴趣的区域可降低VR视频体验。替代地,关于最感兴趣的区域的信息可以其它方式用于经改进的VR视频体验。

[0069] 2. 可通过除导演剪辑外的其它方式来确定或产生VR视频的特定图片集合中的最感兴趣的区域。

[0070] 3. 缺乏用信号表示关于最感兴趣的区域的信息的具体方式。

[0071] 本发明描述用于VR视频中的一或多个最感兴趣的区域的信号表示及重建以及VR视频编码、解码、流式发射及播放中的相应处理的若干技术。这些技术中的一或多者可独立地应用,或与其它技术组合应用。

[0072] 在第一实例中,最感兴趣的区域可通过导演或制作人的意图来确定或通过服务提供者或内容提供者提供的用户统计数据来确定,例如在经由流式发射服务提供VR视频内容时用户请求/观看最多的区域。

[0073] 在第二实例中,在视频编码或转码期间,最感兴趣的区域可经有意编码而具有比其它区域更高的质量(例如,通过使用较低量化参数)、更高的分辨率(如果在视频拼接或视频编码期间以不同子采样比率对视频图片的不同区域进行子采样)或随机可存取性的更高频率。

[0074] 在第三实例中,本发明提出使用视频位流中的SEI消息、媒体文件中的文件格式元数据或HTTP动态自适应流式发射(DASH)媒体呈现描述(MPD)元素或属性来用信号表示关于最感兴趣的区域的信息。

[0075] 可使用SEI消息用信号表示关于VR视频位流的特定图片集合的最感兴趣的区域的信息。所述SEI消息的实例通过语法、语义或两者规定适用于(一或多个)图片的集合。SEI消息可含有至少以下语法元素:

[0076] i. 指示最感兴趣的区域的数目的语法元素。

[0077] ii. 语法元素集合的循环,每一集合描述最感兴趣的区域的位置及大小。举例来说,每一集合可包含指示作为最感兴趣的区域的矩形的四个角的呈相对于完整图片的明度样本数目形式的上部、底部、左侧及右侧偏移的四个语法元素。在另一实例中,每一集合可包含指示作为最感兴趣的区域的矩形的左上角的呈相对于完整图片的明度样本数目形式的坐标的两个语法元素及指示矩形的宽度及高度的两个语法元素。

[0078] iii. 对于最感兴趣的区域中的每一者,可通过语法元素指示优先级,其中较高优先级指示区域的较高兴趣。

[0079] iv. 指示是否以解码次序取消所述SEI消息的所有稍早实例的效果的语法元素。

[0080] v. 指示所述SEI消息的当前实例中的信息是否持续到输出次序的下一图片,或新的经译码视频序列的开始,或位流的结束的语法元素。

[0081] SEI消息可经包含作为H.264/AVC、H.265/HEVC或任何其它视频编解码器规范的部分。含有SEI消息的NAL单元可直接地包含于以下中:1) 用于用信号表示文件格式中的信息的文件格式逻辑框或结构;2) 用于用信号表示SDP中的信息的SDP场;3) 用于用信号表示DASH MPD中的信息的DASH MPD属性或元素;及4) 用于用信号表示MPEG-2TS中的信息的MPEG-2输送流(TS)描述符。

[0082] 可(例如)通过使用新样本组或通过包含于计时元数据播放轨中而在文件格式数据结构中直接携带与上文所描述相同的信息。

[0083] 可使用DASH MPD中的一些新元素、属性及/或描述符而直接携带与上文所描述相同的信息。

[0084] 在第四实例中,在例如DASH的受客户端控制的视频流式发射期间,当内容递送网络(CDN)在使用中时,边缘服务器(例如,服务器装置60)或高速缓存存储器可使用关于最感兴趣的区域的信息以获得具有较高质量表示且在临近从客户端装置40接收请求之前覆盖即将出现的图像的最感兴趣的区域的视频数据,从而预测不久后将最可能请求所述视频数据。即将出现的图像可为紧接着最近发送到客户端装置40且通过客户端装置40呈现或将要呈现的一或多个区域的图像之后的图像。

[0085] 在第五实例中,在例如DASH的受客户端控制的视频流式发射期间,客户端装置40可使用关于最感兴趣的区域的信息以请求具有较高质量表示且在临近用户转动其头部并发送请求之前覆盖即将出现的图像的最感兴趣的区域的视频数据,从而预测不久后将最可能请求所述视频数据。同样,即将出现的图像可为紧接着最近发送到客户端装置40且通过客户端装置40呈现或将要呈现的一或多个区域的图像之后的图像。

[0086] 在第六实例中,例如3GPP TS 26.234中的3GPP包交换流式发射(PSS)的受服务器控制的视频流式发射系统(例如,服务器装置60)可经配置具有使得流式发射客户端(例如,客户端装置40)能够在用于服务器的当前FOV的方向上发送反馈以选择覆盖待发送的最感兴趣的区域的适当视频数据的特征。在这类经配置的受服务器控制的视频流式发射系统中,服务器装置60可使用关于最感兴趣的区域的信息以发送具有较高质量表示且在临近接收当前FOV的方向上的流式发射客户端反馈之前覆盖即将出现的图像的最感兴趣的区域的视频数据,从而预测不久后将最可能需要发送所述视频数据。举例来说,响应于确定特定流式发射客户端正输出特定图像,服务器装置60可向特定流式发射客户端输出具有较高质量表示且覆盖紧接在所述特定图像之后的即将出现的图像的最感兴趣的区域的视频数据。

[0087] 在第七实例中,对于涉及高速缓存存储器(例如,在服务器装置60处)的内容(例如,VR内容)的存储管理,服务器装置60可使用表示特定图像中的一或多个最感兴趣的区域的数据来确定所述特定图像中的哪些区域存储于高速缓存存储器中。开发者可设计存储选择算法且服务器装置60可执行所述算法以基于对所述区域的经选择感兴趣程度(例如,基于导演意见)及/或基于对所述区域的经推导感兴趣程度(例如,基于在内容可用一段时间之后的观看行为的用户统计数据)对图像的区域进行分级。服务器装置60可在高速缓存存储器中存储特定图像中分级值满足阈值(例如,大于阈值)的区域的视频数据。举例来说,当仅特定图像的最感兴趣的区域的分级值满足阈值时,服务器装置60可仅在高速缓存存储器中存储用于特定图像的最感兴趣的区域的视频数据。在这个实例中,客户端装置40可从高速缓存存储器检索用于特定图像的最感兴趣的区域的视频数据且可从主要服务器存储装置检索用于特定图像的其他区域的视频数据,所述主要服务器存储装置可放置得比高速缓存存储器更远离用户。另外或替代地,服务器装置60可基于高速缓存存储器中的可用空间的量动态地确定阈值。

[0088] 在第八实例中,对于VR内容的长期存储,服务器装置60可使用表示(例如,VR视频的)特定图像中的一或多个最感兴趣的区域的数据来决定存储所述图像的哪些区域。开发者可设计存储选择算法且服务器装置60可执行所述算法以基于对所述区域的经选择感兴趣程度(例如,基于导演意见)及/或基于对所述区域的经推导感兴趣程度(例如,基于在内容可用一段时间之后的观看行为的用户统计数据)对区域进行分级。服务器装置60可在长期存储装置(例如,经配置以存储视频数据数周、数月、数年等)中存储特定图像中分级值满

足阈值(例如,大于阈值)的区域的视频数据。举例来说,当仅特定图像的最感兴趣的区域的分级值满足阈值时,服务器装置60可仅在长期存储装置中存储用于特定图像的最感兴趣的区域的视频数据且从所述长期存储装置去除用于特定图像的其他区域的视频数据。

[0089] 客户端装置40、服务器装置60及/或内容准备装置20可经配置以根据本发明的技术操作。出于实例的目的,本发明关于客户端装置40及服务器装置60描述这些技术。然而,应理解,代替服务器装置60(或除服务器装置60以外),内容准备装置20可经配置以执行这些技术。

[0090] 封装单元30可形成NAL单元,所述NAL单元包括识别NAL单元所属的程序的标头以及有效负载,例如音频数据、视频数据或描述NAL单元对应于的输送流式发射或程序流的数据。举例来说,在H.264/AVC中,NAL单元包含1位组标头及变化大小的有效负载。在其有效负载中包含视频数据的NAL单元可包括各种粒度级别的视频数据。举例来说,NAL单元可包括视频数据块、多个块、视频数据的切片或视频数据的整个图片。封装单元30可从视频编码器28接收呈基本流(PES包)的形式的经编码视频数据。封装单元30可使每一基本流与对应程序相关联。

[0091] 封装单元30还可组译来自多个NAL单元的存取单元。一般来说,存取单元可包括用于表示视频数据帧以及对应于所述帧的音频数据(当所述音频数据可用时)的一或多个NAL单元。存取单元通常包含针对一个输出时间实例的所有NAL单元,例如针对一个时间实例的所有音频及视频数据。举例来说,如果每一视图具有20帧每秒(fps)的帧速率,那么每一时间实例可对应于0.05秒的时间间隔。在所述时间间隔期间,可同时呈现相同存取单元(相同时间实例)的所有视图的特定帧。在一个实例中,存取单元可包括一个时间实例中的经译码图片,其可呈现为初级经译码图片。

[0092] 因此,存取单元可包括共同时间实例的所有音频帧及视频帧,例如对应于时间X的所有视图。本发明还将特定视图的经编码图片称作“视图分量”。即,视图分量可包括在特定时间针对特定视图的经编码图片(或帧)。因此,存取单元可被定义为包括共同时间实例的所有视图分量。存取单元的解码次序未必与输出或显示次序相同。

[0093] 媒体呈现可包含媒体呈现描述(MPD),所述媒体呈现描述可含有不同替代表示(例如,具有不同质量的视频服务)的描述,且所述描述可包含(例如)编解码器信息、配置文件值及层级值。MPD为清单文件(例如清单文件66)的一个实例。客户端装置40可检索媒体呈现的MPD以确定如何存取各种呈现的电影片段。电影片段可位于视频文件的电影片段逻辑框(moof逻辑框)中。

[0094] 清单文件66(其可包括(例如)MPD)可通告表示68的区段的可用性。即,MPD可包含指示表示68中的一者的第一片段变得可用时的挂钟时间的信息,以及指示表示68内的区段的持续时间的信息。以这种方式,客户端装置40的检索单元52可基于开始时间以及在特定区段之前的区段的持续时间而确定每一区段何时可用。在一些实例中,根据本发明的技术,清单文件66可包含表示特定图像中的一或多个最感兴趣的区域的数据。

[0095] 在封装单元30已基于所接收数据将NAL单元及/或存取单元组译成视频文件之后,封装单元30将视频文件传递到输出接口32以供输出。在一些实例中,封装单元30可将视频文件存储在本地,或经由输出接口32将视频文件发送到远程服务器,而非将视频文件直接发送到客户端装置40。输出接口32可包括(例如)发射器、收发器、用于将数据写入到计算机

可读媒体的装置(例如光盘驱动器、磁性媒体驱动器(例如,软盘驱动器))、通用串行总线(USB)端口、网络接口或其它输出接口。输出接口32将视频文件输出到计算机可读媒体,例如发射信号、磁性媒体、光学媒体、存储器、快闪驱动器或其它计算机可读媒体。

[0096] 网络接口54可经由网络74接收NAL单元或存取单元,且经由检索单元52将NAL单元或存取单元提供到解封装单元50。解封装单元50可将视频文件的元素解封装成组成性PES流,解包化所述PES流以检索经编码数据,且取决于经编码数据为音频流或视频流的部分(例如,如由流的PES包标头所指示)而将经编码数据发送到音频解码器46或视频解码器48。音频解码器46对经编码音频数据进行解码,且向音频输出42发送经解码音频数据,而视频解码器48对经编码视频数据进行解码,且向视频输出44发送经解码视频数据,其可包含流的多个视图。

[0097] 以这种方式,内容准备装置20表示用以用信号表示用于视频数据的信息的装置的一实例,所述装置包含经配置以确定视频数据图像的多个区域中一或多个最感兴趣的区域的处理器。最感兴趣的区域包括最可能由客户端装置检索的一或多个区域。所述处理器可经进一步配置以产生表示最感兴趣的区域的数据。表示最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。

[0098] 另外,客户端装置40表示用以确定用于视频数据的信息的装置的一实例,所述装置包含经配置以基于表示最感兴趣的区域的数据确定视频数据图像的多个区域中一或多个最感兴趣的区域的处理器。表示最感兴趣的区域的数据在包含视频数据的位流的VCL数据外部。所述处理器可经进一步配置以输出对图像的一或多个最感兴趣的区域的高质量视频数据的请求,且在输出对图像的一或多个最感兴趣的区域的高质量视频数据的所述请求后,输出对图像的一或多个其它区域的低质量视频数据的请求。

[0099] 图2为说明实例多媒体内容120的元素的示意图。多媒体内容120可对应于多媒体内容64(图1)或存储于存储媒体62中的另一多媒体内容。在图2的实例中,多媒体内容120包含媒体呈现描述(MPD)122及多个表示124A到124N(表示124)。表示124A包含可选标头数据126及区段128A到128N(区段128),而表示124N包含可选标头数据130及区段132A到132N(区段132)。为了方便起见,使用字母N来指明表示124中的每一者中的最后一个电影片段。在一些实例中,表示124之间可存在不同数目的电影片段。

[0100] 尽管图2描述其中系统10(图1)经配置用于DASH的一实例,但在一些实例中,系统10可经配置用于其它媒体播放控制协议,例如(但不限于)实时流式发射协议(RTSP)、实时传输协议(RTP)、RTP控制协议(RTCP)、会话描述协议(SDP)、另一媒体播放控制协议或其一组合。

[0101] MPD 122可包括与表示124分离的数据结构。MPD 122可对应于图1的清单文件66。同样地,表示124可对应于图2的表示68。一般来说,MPD 122可包含大体上描述表示124的特征的数据,所述特征例如译码及呈现特征、适应集合、MPD 122所对应的配置文件、文本类型信息、相机角度信息、分级信息、窍门模式信息(例如,指示包含时间子序列的表示的信息)及/或用于检索远程周期(例如,用于在播放期间向媒体内容中插入针对性广告)的信息。

[0102] 标头数据126(当存在时)可描述区段128的特征,例如随机存取点(RAP,也称作流存取点(SAP))的时间位置、区段128中的哪一者包含随机存取点、与区段128内的随机存取点的位组偏移、区段128的统一资源定位符(URL)或区段128的其它方面。标头数据130(当存

在时)可描述区段132的类似特征。另外或替代地,这些特征可完全包含于MPD 122内。

[0103] 区段128、132包含一或多个经译码视频样本,其中的每一者可包含视频数据的帧或切片。区段128的经译码视频样本中的每一者可具有类似特征,例如,高度、宽度及带宽要求。这些特征可通过MPD 122的数据来描述,不过所述数据在图2的实例中未说明。MPD 122可包含如由3GPP规范所描述的特征,并且添加了本发明中所描述的用信号表示的信息中的任一者或全部。

[0104] 区段128、132中的每一者可与唯一的统一资源定位符(URL)相关联。因此,区段128、132中的每一者可为使用流式发射网络协议(例如DASH)而可独立地检索的,但可使用其它流式发射网络协议。以这种方式,例如客户端装置40的目的地址装置可使用HTTP GET请求以检索区段128或132。在一些实例中,客户端装置40可使用HTTP部分GET请求以检索区段128或132的特定组范围。举例来说,客户端装置40可使用第一HTTP部分GET请求以首先从区段128、132中的一者检索表示最感兴趣的区域的数据,确定所述最感兴趣的区域,接着使用第二HTTP部分GET请求以从区段128、132中的一者预提取用于最感兴趣的区域的视频数据的高质量版本,假定最感兴趣的区域数据包含于(例如)文件格式信息中的区段128、132内。

[0105] 另外或替代地,根据本发明的技术,MPD 122可包含指示一或多个最感兴趣的区域的元素或属性。举例来说,对于区段128、132中的每一者,MPD 122可包含表示最感兴趣的区域的相应位置及所述最感兴趣的区域的相应大小的元素或属性。

[0106] 更具体来说,例如,对于区段128、132中的每一者(或区段128、132内的一或多个图片),MPD 122可包含在最感兴趣的区域上循环的元素或属性,循环的每一迭代表示最感兴趣的区域中的一者且包含表示最感兴趣的区域的位置及最感兴趣的区域的大小的语法元素。在一些实例中,对于区段128、132中的每一者,MPD 122可包含针对最感兴趣的区域中的每一者包含表示最感兴趣的区域的顶部的第一语法元素、表示最感兴趣的区域的底部的第二语法元素、最感兴趣的区域的左侧及最感兴趣的区域的右侧的元素或属性。在一些实例中,对于区段128、132中的每一者,MPD 122可包含针对最感兴趣的区域中的每一者包含表示最感兴趣的区域的左上角的第一语法元素及表示最感兴趣的区域的右下角的第二语法元素的元素或属性。在一些实例中,对于区段128、132中的每一者,MPD 122可包含针对最感兴趣的区域中的每一者包含表示最感兴趣的区域的左上角的第一语法元素、表示最感兴趣的区域的宽度的第二语法元素及表示最感兴趣的区域的高度的第三语法元素的元素或属性。在一些实例中,MPD 122可表示亮度样本中最感兴趣的区域的相应位置及最感兴趣的区域的相应大小。

[0107] 图3为说明实例视频文件150的元素的框图,实例视频文件可对应于表示的区段,例如图2的区段128、132中的一者。区段128、132中的每一者可包含大体上符合图3的实例中所说明的数据布置的数据。据称视频文件150可封装一区段。如上所述,根据ISO基本媒体文件格式及其扩展的视频文件将数据存储于一系列对象(被称为“逻辑框”)中。在图3的实例中,视频文件150包含文件类型(FTYP)逻辑框152、电影(MOOV)逻辑框154、一或多对电影片段(MOOF)逻辑框164及媒体数据(MDAT)逻辑框167,以及电影片段随机存取(MFRA)逻辑框166。尽管图3表示视频文件的一实例,但应理解,根据ISO基本媒体文件格式及其扩展,其它媒体文件可包含在结构上类似于视频文件150的数据的其它类型的媒体数据(例如,音频数

据、计时文本数据或其类似者)。

[0108] 文件类型 (FTYP) 逻辑框152通常描述视频文件150的文件类型。文件类型逻辑框152可包含识别描述视频文件150的最佳用途的规范的数据。文件类型逻辑框152可替代地放置在MOOV逻辑框154、MOOF逻辑框164及/或MFRA逻辑框166之前。

[0109] 在图3的实例中,MOOV逻辑框154包含电影标头 (MVHD) 逻辑框156及播放轨 (TRAK) 逻辑框158。一般来说,MVHD逻辑框156可描述视频文件150的一般特征。举例来说,MVHD逻辑框156可包含描述视频文件150何时最初建立、视频文件150何时经最后修改、视频文件150的时间标度、视频文件150的播放持续时间的数据,或大体上描述视频文件150的其它数据。

[0110] MOOV逻辑框154可包含多个TRAK逻辑框,其等于视频文件150中的播放轨数目。TRAK逻辑框158可描述视频文件150的对应播放轨的特征。举例来说,TRAK逻辑框158可描述对应播放轨的时间及/或空间信息。TRAK逻辑框158可包含视频文件150的播放轨的数据。TRAK逻辑框158可包含描述对应于TRAK逻辑框158的播放轨的特征的播放轨标头 (TKHD) 逻辑框。在一些实例中,TRAK逻辑框158可包含经译码视频图片,而在其它实例中,播放轨的经译码视频图片可包含于MDAT逻辑框167中,其可由TRAK逻辑框158的数据参考。

[0111] 视频文件150可包含计时元数据播放轨。举例来说,视频文件150可用于包含一或多个计时元数据播放轨的表示。举例来说,视频文件150可用于包含一或多个计时元数据播放轨且不包含相关联媒体播放轨的表示124N。在一些实例中,视频文件150可用于包含与相关联媒体播放轨复用在一起的一或多个计时元数据播放轨的表示。举例来说,视频文件150可用于包含与相关联媒体播放轨复用在一起的一或多个计时元数据播放轨的表示124A。在任何情况下,计时元数据播放轨可含有针对视频文件150的播放轨的描述性或注释性元数据。

[0112] 视频文件150可包含针对包含于视频文件150中的每一相应播放轨的相应对MOOF逻辑框164及MDAT逻辑框167。如所展示,每一MOOF逻辑框164可包含播放轨片段 (TRAF) 逻辑框165。在图3的实例中,TRAF逻辑框165包含样本到群组 (SBGP) 162。SBGP逻辑框162可描述样本或播放轨所属的群组及所述样本组的相关联描述。当封装单元30 (图1) 包含视频文件 (例如视频文件150) 中的参数集播放轨时,TRAF逻辑框165可描述参数集播放轨的特征。封装单元30可在描述参数集播放轨的TRAK逻辑框158内用信号表示序列层级SEI消息存在于参数集播放轨中。

[0113] MDAT逻辑框167可包含用于相应播放轨的SEI消息172及VCL NAL单元170。VCL NAL单元170可包含一或多个经译码视频图片。在一些实例中,VCL NAL单元170可包含一或多个图片组 (GOP), 其中的每一者可包含多个经译码视频图片,例如帧或图片。

[0114] 如上文所提及,封装单元30可存储视频样本中不包含实际经译码视频数据的序列数据集。视频样本可大体上对应于存取单元,所述存取单元为特定时间实例处的经译码图片的表示。在AVC的上下文中,存取单元包含含有用以构建存取单元的所有像素的信息的至少一个VCL NAL单元170及其它相关联的非VCL NAL单元 (例如SEI消息172)。因此,封装单元30可包含序列数据集,所述序列数据集可包含SEI消息172及VCL NAL单元170。封装单元30可进一步将序列数据集及/或SEI消息172的存在用信号表示为存在于MDAT逻辑框167内。

[0115] 视频文件150可包含表示根据本发明的技术的一或多个最感兴趣的区域的数据。更具体来说,视频文件150可包含表示视频文件150的文件格式标头信息中的一或多个最感

兴趣的区域的数据。举例来说,视频文件150可包含用信号表示最感兴趣的区域信息的SEI消息及/或文件格式信息。举例来说,计时元数据播放轨可包含表示一或多个最感兴趣的区域的数据。在一些个例中,SBGP逻辑框162可包含表示一或多个最感兴趣的区域的数据。在一些个例中,SEI消息172可包含表示一或多个最感兴趣的区域的数据。

[0116] SEI消息172可经包含作为H.264/AVC、H.265/HEVC或其它视频编解码器规范的部分。在图3的实例中,SEI消息包含表示用于(例如)VCL NAL单元170的对应图片的最感兴趣的区域的数据。在其它实例中,这类最感兴趣的区域数据可包含于以下中:1)用于用信号表示文件格式中的信息的文件格式逻辑框或结构;2)用于用信号表示SDP中的信息的SDP场;3)用于用信号表示DASH MPD(例如,图2的MPD 122)中的信息的DASH MPD属性或元素;4)用于在另一逻辑框或结构中用信号表示MPEG-2TS中的信息的MPEG-2输送流(TS)描述符;或其一组合。

[0117] MFRA逻辑框166可描述视频文件150的电影片段内的随机存取点。这可帮助执行窍门模式,例如执行对由视频文件150封装的区段内的特定时间位置(即,播放时间)的搜索。在一些实例中,MFRA逻辑框166通常可选的且不必包含于视频文件中。同样,客户端装置(例如客户端装置40)未必需要参考MFRA逻辑框166以正确解码及显示视频文件150的视频数据。MFRA逻辑框166可包含多个播放轨片段随机存取(TFRA)逻辑框(未展示),其等于视频文件150的播放轨数目或在一些实例中等于视频文件150的媒体播放轨(例如,非提示播放轨)数目。

[0118] 在一些实例中,视频文件150的电影片段可包含一或多个流存取点(SAP),例如IDR图片。同样地,MFRA逻辑框166可提供对SAP在视频文件150内的位置的指示。因此,视频文件150的时间子序列可由视频文件150的SAP形成。时间子序列还可包含其它图片,例如取决于SAP的P帧及/或B帧。时间子序列的帧及/或切片可布置于区段内,使得时间子序列的取决于子序列的其它帧/切片的帧/切片可被正确地解码。举例来说,在数据的阶层式布置中,用于其它数据的预测的数据还可包含于时间子序列中。

[0119] 视频文件150可包含指示最感兴趣的区域的数目的语法元素。举例来说,计时元数据播放轨可包含指示最感兴趣的区域的数目的语法元素。在一些实例中,SBGP逻辑框162可包含指示最感兴趣的区域的数目的语法元素。在一些实例中,SEI消息172可包含指示最感兴趣的区域的数目的语法元素。

[0120] 视频文件150可包含指示当前实例中的信息持续到输出次序的下一图像、或新的经译码视频序列的开始、或位流的结束的语法元素。举例来说,计时元数据播放轨可包含指示计时元数据播放轨的当前实例中的信息持续到输出次序的下一图像、或新的经译码视频序列的开始、或位流的结束的语法元素。在一些实例中,SBGP逻辑框162可包含指示SBGP逻辑框162的当前实例中的信息持续到输出次序的下一图像、或新的经译码视频序列的开始、或位流的结束的语法元素。在一些实例中,SEI消息172可包含指示SEI消息172的当前实例中的信息持续到输出次序的下一图像、或新的经译码视频序列的开始、或位流的结束的语法元素。

[0121] 视频文件150可包含指示是否以解码次序取消所有稍早实例的效果的语法元素。换句话说,在表示一或多个最感兴趣的区域的数据指示一或多个最感兴趣的区域持续到输出次序的下一图像、或新的经译码视频序列的开始、或位流的结束的实例中,指示是否取消

效果的语法元素可去除一或多个最感兴趣的区域的持续。举例来说,计时元数据播放轨可包含指示是否以解码次序取消计时元数据播放轨的所有稍早实例的效果的语法元素。在一些实例中,SBGP逻辑框162可包含指示是否以解码次序取消SBGP逻辑框162的所有稍早实例的效果的语法元素。在一些实例中,SEI消息172可包含指示是否以解码次序取消SEI消息172的所有稍早实例的效果的语法元素。

[0122] 视频文件150可针对最感兴趣的区域中的每一者通过语法元素指示优先级,其中较高优先级指示所述区域的较高兴趣。举例来说,计时元数据播放轨可针对最感兴趣的区域中的每一者通过语法元素指示优先级,其中较高优先级指示所述区域的较高兴趣。在一些实例中,SBGP逻辑框162可针对最感兴趣的区域中的每一者通过语法元素指示优先级,其中较高优先级指示所述区域的较高兴趣。在一些实例中,SEI消息172可针对最感兴趣的区域中的每一者通过语法元素指示优先级,其中较高优先级指示所述区域的较高兴趣。

[0123] 根据本发明的技术,包含VCL NAL单元170中所含有的视频数据的视频文件150的文件格式信息中的数据可包含最感兴趣的区域的相应位置及最感兴趣的区域的相应大小的指示。

[0124] 举例来说,计时元数据播放轨可包含表示最感兴趣的区域的数据。举例来说,计时元数据播放轨可针对每一电影片段包含最感兴趣的区域的相应位置及最感兴趣的区域的相应大小的指示。

[0125] 更具体来说,在一些实例中,计时元数据播放轨可针对形成图像的一或多个电影片段中的每一者包含最感兴趣的区域上的循环的指示,循环的每一迭代表示最感兴趣的区域中的一者且包含表示相对于所述图像的多个明度样本中的位置及相对于所述图像的多个明度样本中的最感兴趣的区域的大小的语法元素。

[0126] 在一些实例中,计时元数据播放轨可针对一或多个电影片段中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的顶部的第一语法元素、表示相对于所述图像的多个明度样本中的最感兴趣的区域的底部的第二语法元素、相对于所述图像的多个明度样本中的最感兴趣的区域的左侧及相对于所述图像的多个明度样本中的最感兴趣的区域的右侧的指示。

[0127] 在一些实例中,计时元数据播放轨可针对每一电影片段包含元素或属性,所述元素或属性针对最感兴趣的区域中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的左上角的第一语法元素及表示相对于所述图像的多个明度样本中的最感兴趣的区域的右下角的第二语法元素。

[0128] 在一些实例中,计时元数据播放轨可针对每一电影片段包含元素或属性,所述元素或属性针对最感兴趣的区域中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的左上角的第一语法元素、表示相对于所述图像的多个明度样本中的最感兴趣的区域的宽度的第二语法元素及表示相对于所述图像的多个明度样本中的最感兴趣的区域的高度的第三语法元素。

[0129] 在另一实例中,SBGP逻辑框162可包含表示最感兴趣的区域的数据。举例来说,SBGP逻辑框162可针对每一电影片段包含最感兴趣的区域的相应位置及最感兴趣的区域的相应大小的指示。

[0130] 更具体来说,在一些实例中,SBGP逻辑框162可针对形成图像的每一电影片段包含

最感兴趣的区域上的循环的指示,循环的每一迭代表示最感兴趣的区域中的一者且包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的位置及相对于所述图像的多个明度样本中的最感兴趣的区域的大小的语法元素。

[0131] 在一些实例中,SBGP逻辑框162可针对形成图像的每一电影片段包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的顶部的第一语法元素、表示相对于所述图像的多个明度样本中的最感兴趣的区域的底部的第二语法元素、相对于所述图像的多个明度样本中的最感兴趣的区域的左侧及相对于所述图像的多个明度样本中的最感兴趣的区域的右侧的指示。

[0132] 在一些实例中,SBGP逻辑框162可针对形成图像的每一电影片段包含元素或属性,所述元素或属性针对最感兴趣的区域中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的左上角的第一语法元素及表示相对于所述图像的多个明度样本中的最感兴趣的区域的右下角的第二语法元素。

[0133] 在一些实例中,SBGP逻辑框162可针对形成图像的每一电影片段包含元素或属性,所述元素或属性针对最感兴趣的区域中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的左上角的第一语法元素、表示相对于所述图像的多个明度样本中的最感兴趣的区域的宽度的第二语法元素及表示相对于所述图像的多个明度样本中的最感兴趣的区域的高度的第三语法元素。

[0134] 在另一实例中,SEI消息172可包含表示最感兴趣的区域的数据。举例来说,SEI消息172可针对形成图像的每一电影片段包含最感兴趣的区域的相应位置及最感兴趣的区域的相应大小的指示。

[0135] 更具体来说,例如,SEI消息172可针对形成图像的每一电影片段包含最感兴趣的区域上的循环的指示,循环的每一迭代表示最感兴趣的区域中的一者且包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的位置及相对于所述图像的多个明度样本中的最感兴趣的区域的大小的语法元素。

[0136] 在一些实例中,SEI消息172可针对形成图像的每一电影片段包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的顶部的第一语法元素、表示相对于所述图像的多个明度样本中的最感兴趣的区域的底部的第二语法元素、相对于所述图像的多个明度样本中的最感兴趣的区域的左侧及相对于所述图像的多个明度样本中的最感兴趣的区域的右侧的指示。

[0137] 在一些实例中,SEI消息172可针对形成图像的每一电影片段包含元素或属性,所述元素或属性针对最感兴趣的区域中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的左上角的第一语法元素及表示相对于所述图像的多个明度样本中的最感兴趣的区域的右下角的第二语法元素。

[0138] 在一些实例中,SEI消息172可针对形成图像的每一电影片段包含元素或属性,所述元素或属性针对最感兴趣的区域中的每一者包含表示相对于所述图像的多个明度样本中的最感兴趣的区域的左上角的第一语法元素、表示相对于所述图像的多个明度样本中的最感兴趣的区域的宽度的第二语法元素及表示相对于所述图像的多个明度样本中的最感兴趣的区域的高度的第三语法元素。

[0139] 图4为以立方体投影230说明实例图像块译码的概念图。立方体投影230包含如所

展示的立方体面232A、232B、232C,且包含由于被遮挡而未展示的额外立方体面。在这个实例中,将立方体投影230的立方体面中的每一者划分成四个图像块:将立方体面232A划分成四个图像块234A到234D,将立方体面232B划分成四个图像块236A到236D,且将立方体面232C划分成四个图像块238A到238D。在这个实例中,其它未展示的立方体面也将划分成四个相应图像块,总共24个较小图像块。一般来说,内容准备装置20可提供用于图像块(例如,图像块234A到234D、图像块236A到236D、图像块238A到238D及未展示立方体面的图像块)中的每一者的视频数据。

[0140] 另外,内容准备装置20可提供用于较大图像块的视频数据,所述较大图像块可覆盖立方体面中的每一者。举例来说,内容准备装置20可提供用于覆盖立方体面232A的图像块的视频数据、用于覆盖立方体面232B的图像块的视频数据、用于覆盖立方体面232C的图像块的视频数据及用于覆盖未展示立方体面的图像块的视频数据。在一些实例中,视频编码器28可仅针对更可能被观看的那些视点编码较大图像块大小。因此,如果用户将不大可能向上看或向下看,那么视频编码器28可仅针对覆盖立方体面232B及232C的较大图像块及针对立方体面232B及232C的相对立方体面但不针对(例如)立方体面232A译码视频数据。任选地,图像块中的每一者可具有用于立体视频数据的左侧视图及右侧视图。

[0141] VR为虚拟存在于通过呈现由于沉浸用户的移动而相互关联的天然及/或合成图像与声音所产生的非实际世界中,从而允许用户与所述世界交互的能力。随着在呈现装置(例如头戴式显示器(HMD))及VR视频创建(通常也被称作360度视频)中所取得的最新进展,可提供显著的体验质量。VR应用包含游戏、训练、教育、体育视频、在线购物、成人娱乐等等。

[0142] 典型VR系统可包含以下组件及技术:

[0143] 1) 相机集合,其通常由指向不同方向且理想地共同覆盖所述相机集合周围的所有视点的多个单独相机组成。

[0144] 2) 图像拼接,其中通过多个单独相机拍摄的视频图片在时域中同步且在空间域中拼接为球形视频,但映射到矩形格式,例如等矩形映射(如世界映射)或立方体映射。

[0145] 3) 使用视频编解码器(例如,H.265/HEVC或H.264/AVC)编码/压缩呈经映射矩形格式的视频。

[0146] 4) 经压缩视频位流可以媒体格式存储及/或封装且经由网络发射(可能只有仅覆盖用户看见的区域的子集)到接收器。

[0147] 5) 接收器接收可能以一格式封装的视频位流或其部分,且将经解码视频信号或其部分发送到呈现装置。

[0148] 6) 呈现装置可为(例如)HMD,其可跟踪头部移动且甚至眼睛移动时刻且呈现视频的对应部分以使得将沉浸体验传递给用户。

[0149] 与普通视频相比,VR视频的特征为在VR中,通常仅显示视频图片呈现的整个视频区域的对应于当前FOV(也被称作检视区,例如,用户当前正观看的区域)的子集,而在普通视频应用中,通常显示整个视频区域。所述特征可用于(例如)通过使用检视区依赖性投影映射或检视区依赖性视频译码来改进VR视频系统的性能。在呈现给用户的视频部分的分辨率/质量相同的情况下,与常规VR视频系统相比,性能改进可为降低发射带宽及降低解码复杂性中的任一者或两者。

[0150] 检视区依赖性投影映射也可被称作不对称投影映射。一个实例为经子采样立方体

映射,例如立方体投影230。典型立方体映射可包含六个相同大小的面。在经子采样立方体映射的一个实例中,所述面中的一者可保持不变,而相对侧上的面可经子采样或向下按比例缩放到位于初始面形状的中心区域处的较小大小,且相应地在几何形状上按比例缩放其它面(仍然保持六个面)。极端情况可为将相对侧上的面向下按比例缩放到单个点,且因此立方体变为角锥形。在经子采样立方体映射的另一实例中,例如以 2×2 的比例(即,在与所述面的边缘平行的每一方向上2:1)按比例减小一些面的大小。

[0151] 这种大小减小的按比例缩放还可应用于例如等矩形的其它投影映射的不同区域。一个实例为使上部区域及底部区域两者(即,极点)大小减小。

[0152] 检视区依赖性视频译码也可被称作基于检视区的部分视频解码,以使得能够仅部分地解码整个经编码视频区域以为当前FOV或检视区的显示提供足够的信息。

[0153] 在检视区依赖性视频译码的一个实例中,使用图像块的受运动约束集合来译码VR视频,使得覆盖FOV或检视区的每一潜在区域可独立于其它区域地解码。对于特定当前FOV或检视区,将覆盖当前FOV或检视区的可独立解码图像块的最小集合的经译码视频数据发送到客户端,解码并显示。所述方法的缺点为当用户将其头部快速地转动到所发送视频数据未覆盖的新FOV时,在覆盖新FOV的数据到达前,无法看见视频数据的图像块未覆盖的区域中的任何内容。所述情形可容易地发生,除非网络的往返时间极其短(例如,在10毫秒的量值下),其为不可行的或至少为当今或不远未来的巨大挑战。

[0154] 检视区依赖性视频译码的另一实例在2016年6月28日提交的美国专利申请案15/195,439中提出,所述申请案的全部内容以引用的方式并入本文中,命名为独立多分辨率译码(IMRC),其中球形/全景视频以独立于彼此的多个不同分辨率加以编码,且每一表示使用图像块的受运动约束集合加以译码。接收器选择以不同分辨率对球形/全景视频的不同部分进行解码。通常,球形视频的用户当前正观察的部分(即,当前FOV或当前检视区)为视频位流的以最高分辨率译码的部分。当前FOV周围的区域为视频位流的使用略微较低分辨率译码的部分,诸如此类。全景的在观测者头部正后方的部分为视频位流的以最低分辨率译码的部分。所主张的是,在用户头部移动的情况下,用户体验在大多数情况下将仅少量降低,且质量降低仅在极大的突然头部移动的情况下最严重,这极少发生。

[0155] 在检视区依赖性视频译码的另一实例中,使用例如SHVC的可调式视频编解码器可调式地译码具有多分辨率的视频位流。另外,始终发送最低分辨率视频(或基本层)的整个位流。不需要使用图像块或图像块的受运动约束集合来译码最低分辨率视频,但其在使用图像块或图像块的受运动约束集合加以译码的情况下也将起作用。对于其它方面,应用如上文所描述的策略。所述方法允许最低分辨率视频的更高效译码,这是因为译码图像块或图像块的受图像块运动约束集合降低译码效率,且还由于使用层间预测而允许较高分辨率的流的更高效译码。另外,所述方法还允许更高效FOV切换,这是因为一旦FOV切换开始,服务器或边缘服务器可立即停止发送来自具有更高分辨率(相比于基本层)的视频流且未覆盖新FOV的任何区域的视频数据。如果不发送覆盖当前FOV的具有最低分辨率的视频数据(即,仅发送覆盖当前FOV的具有最高分辨率的视频数据),那么在FOV切换期间,服务器有时将必须继续发送来自具有最高分辨率的视频流且仅覆盖旧FOV或其部分的视频数据,准备好以使得用户在其转回到旧FOV或其部分的情况下可至少看到最低质量的视频。

[0156] 在MPEG提案m37819中,论述关于用信号表示及使用导演剪辑的信息的使用情况,

使得即使在用户不转动其头部或经由其它UI改变检视区时,VR播放还可显示导演希望观众聚焦的动态改变的检视区。其提及所述检视区可逐场景具备全向视频。

[0157] 本文所描述的一或多种技术准许客户端装置40及/或服务器装置60预提取用于包含一或多个最感兴趣的区域的一或多个图像块的高质量数据,而非必需仅依赖于用户的视场来选择质量以请求图像块(例如,图像块234A到234D、图像块236A到236D、图像块238A到238D及未展示立方体面的图像块)。举例来说,当表示一或多个最感兴趣的区域的数据指示图像块234A到234D包含一或多个最感兴趣的区域时,客户端装置40可在接收用户的FOV之前请求图像块234A到234D。在一实例中,其中FOV包含图像的不包含于一或多个最感兴趣的区域中的区域,在接收用户的FOV后,客户端装置40可请求另外的图像块(例如,236A到236D、238A到238D)以在较低质量下满足用户的当前FOV。以这种方式,一或多个最感兴趣的区域可经预提取(例如,在接收用户的FOV前经提取)以准许以高质量接收一或多个最感兴趣的区域,借此改进视频(例如虚拟实境视频)的播放。

[0158] 图5为说明根据本发明中所描述的一或多种技术的用以确定用于视频数据的信息的实例方法的流程图。关于图1的客户端装置40及服务器装置60描述图5的方法。然而,应理解,其它装置可经配置以执行这些技术。

[0159] 首先,服务器装置60可发送表示图像的最感兴趣的区域的数据(300)。表示最感兴趣的区域的数据可在包含视频数据的位流的VCL数据外部。举例来说,服务器装置60发送包含表示最感兴趣的区域的数据的补充增强信息(SEI)消息。在一些实例中,服务器装置60发送包含表示最感兴趣的区域的数据的样本组逻辑框(例如,图3的SBGP逻辑框162)。在一些实例中,服务器装置60发送包含表示最感兴趣的区域的数据的计时元数据播放轨。

[0160] 在任何情况下,客户端装置40可基于表示图像的最感兴趣的区域的数据确定所述最感兴趣的区域(302)。举例来说,针对一或多个最感兴趣的区域中的每一者,客户端装置40可使用表示最感兴趣的区域的数据来确定表示相应最感兴趣的区域的位置及相应最感兴趣的区域的大小的语法元素的值。

[0161] 客户端装置40可预提取用于一或多个最感兴趣的区域的高质量视频数据(304)。举例来说,客户端装置40可使用与图像的其它区域相比相对较高的位速率来请求用于一或多个最感兴趣的区域的视频数据。服务器装置60可接收预提取请求(306)且可发送用于一或多个最感兴趣的区域的高质量视频数据(308)。客户端装置40可接收所述高质量视频数据(310)。举例来说,客户端装置40可以比用于图像的其它区域的视频数据更低的时延及/或更高的质量提取用于一或多个最感兴趣的区域的视频数据。

[0162] 客户端装置40可输出对用于图像的一或多个其它区域的低质量视频数据的请求(314)。举例来说,当一或多个最感兴趣的区域包含图4的图像238A到238D时且当用户的FOV包含图4的图像238A到238D及234C到234D时,客户端装置40可请求与用于一或多个最感兴趣的区域的视频质量相比相对低质量的用于图4的图像234C到234D的视频数据。应理解,可不请求或可以与使用FOV确定的图像(例如,图4的图像238A到238D及234C到234D)相比相对更低的质量请求FOV外的区域(例如,图4的图像234A到234B)。

[0163] 服务器装置60可接收请求(316)且可发送用于图像的一或多个其它区域的经请求数据(318)。客户端装置40可接收所述经请求数据(320)。客户端装置40可确定用户的FOV(321)且向用户输出用于包含于FOV中的区域的数据(322)。举例来说,客户端装置40可确定

用户的FOV包含图4的图像238A到238D及234C到234D且使用高质量视频数据及低质量视频数据输出图4的图像238A到238D及234C到234D以用于显示(322)。

[0164] 以这种方式,图5的方法表示方法的一实例,所述方法包含:通过客户端装置的处理器(所述处理器实施于电路中)基于表示视频数据图像的多个区域的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域,表示所述一或多个最感兴趣的区域的数据在包含所述视频数据的位流的视频译码层(VCL)数据外部;通过客户端装置的处理器输出对用于所述图像的一或多个最感兴趣的区域的相对高质量视频数据的请求;以及在输出对所述相对高质量视频数据的请求后,通过客户端装置的处理器输出对用于所述图像的一或多个其它区域的相对低质量视频数据的请求。

[0165] 同样地,图5的方法还表示方法的一实例,所述方法包含:通过源装置的处理器(所述处理器实施于电路中)确定视频数据图像的多个区域的一或多个最感兴趣的区域,所述一或多个最感兴趣的区域包括客户端装置最可能检索的一或多个区域;以及通过源装置的处理器产生表示所述一或多个最感兴趣的区域的数据,其中表示所述一或多个最感兴趣的区域的数据在包含所述视频数据的位流的视频译码层(VCL)数据外部。

[0166] 图6为说明根据本发明中所描述的一或多种技术的用以用信号表示用于视频数据的信息的实例方法的流程图。关于图1的客户端装置40及服务器装置60描述图6的方法。然而,应理解,其它装置可经配置以执行这些技术。

[0167] 首先,服务器装置60可确定图像的一或多个最感兴趣的区域(330)。举例来说,服务器装置60可接收导演或制作人对一或多个最感兴趣的区域的选择的指示或服务器装置60(或另一装置)可根据用户统计数据推导一或多个最感兴趣的区域。在任何情况下,服务器装置60可产生表示最感兴趣的区域的数据(332)。表示最感兴趣的区域的数据可在包含视频数据的位流的VCL数据外部。举例来说,服务器装置60可产生包含表示最感兴趣的区域的数据的补充增强信息(SEI)消息。在一些实例中,服务器装置60可产生包含表示最感兴趣的区域的数据的文件格式数据,例如样本组逻辑框(例如,图3的SBGP逻辑框162)。在一些实例中,服务器装置60可产生包含表示最感兴趣的区域的数据的计时元数据播放轨。

[0168] 在任何情况下,服务器装置60可发送表示最感兴趣的区域的数据(334)且客户端装置40可接收表示最感兴趣的区域的数据(336)。响应于接收到表示最感兴趣的区域的数据,客户端装置40可输出预提取用于一或多个最感兴趣的区域的高质量视频数据的请求(338)。服务器装置60可接收预提取用于一或多个最感兴趣的区域的高质量视频数据的请求(340)且发送用于一或多个最感兴趣的区域的高质量视频数据(342)。客户端装置40可接收所述高质量视频数据(344)。

[0169] 客户端装置40可输出对用于图像的一或多个其它区域的低质量视频数据的请求(346)。服务器装置60可接收对用于图像的一或多个其它区域的低质量视频数据的请求(348)。响应于接收到对用于图像的一或多个其它区域的低质量视频数据的请求,服务器装置60可发送用于图像的一或多个其它区域的低质量视频数据(350)。客户端装置40可接收所述低质量视频数据(352)。尽管图6的实例中未展示,但客户端装置40接着可继续进行到(例如)基于用户的视场向用户呈现所接收视频数据的至少部分,例如,如上文关于图5所论述。

[0170] 以这种方式,图6的方法表示方法的一实例,所述方法包含:通过客户端装置的处

理器(所述处理器实施于电路中)基于表示视频数据图像的多个区域的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域,表示所述一或多个最感兴趣的区域的数据在包含所述视频数据的位流的视频译码层(VCL)数据外部;通过客户端装置的处理器输出对用于图像的一或多个最感兴趣的区域的相对高质量视频数据的请求;以及在输出对所述相对高质量视频数据的请求后,通过客户端装置的处理器输出对用于图像的一或多个其它区域的相对低质量视频数据的请求。

[0171] 同样地,图6的方法还表示方法的一实例,所述方法包含:通过源装置的处理器(所述处理器实施于电路中)确定视频数据图像的多个区域的一或多个最感兴趣的区域,所述一或多个最感兴趣的区域包括客户端装置最可能检索的一或多个区域;以及通过源装置的处理器产生表示所述一或多个最感兴趣的区域的数据,其中表示所述一或多个最感兴趣的区域的数据在包含所述视频数据的位流的视频译码层(VCL)数据外部。

[0172] 图7为说明根据本发明中所描述的一或多种技术的用于预提取视频数据的实例方法的流程图。关于图1的客户端装置40及服务器装置60描述图7的方法。然而,应理解,其它装置可经配置以执行这些技术。在图7的实例中,服务器装置60可配置为首先从源服务器检索数据且暂时存储所述数据的CDN服务器。应理解,图7的实例证实客户端装置40及服务器装置60中的任一者或两者可使用表示最感兴趣的区域的数据。

[0173] 首先,服务器装置60可接收表示图像的最感兴趣的区域的数据(360)。表示最感兴趣的区域的数据可在包含视频数据的位流的VCL数据外部。举例来说,服务器装置60接收包含表示最感兴趣的区域的数据的补充增强信息(SEI)消息。在一些实例中,服务器装置60接收包含表示最感兴趣的区域的数据的样本组逻辑框(例如,图3的SBGP逻辑框162)。在一些实例中,服务器装置60接收包含表示最感兴趣的区域的数据的计时元数据播放轨。在任何情况下,服务器装置60可预提取用于一或多个感兴趣的区域的高质量视频数据(362)。举例来说,服务器装置60可使用一或多个最感兴趣的区域来预提取用于所述一或多个最感兴趣的区域的视频数据。举例来说,服务器装置60可以比用于图像的其它区域的视频数据更低的时延及/或更高的质量提取用于所述一或多个最感兴趣的区域的视频数据。

[0174] 在预提取用于一或多个感兴趣的区域的高质量视频数据后,服务器装置60可存储用于所述一或多个最感兴趣的区域的经预提取的高质量视频数据(364)。在服务器装置60为经配置以存储(例如,高速缓存)视频数据的边缘服务器的一实例中,服务器装置60可存储所述高质量视频数据。在一些实例中,服务器装置60可基于表示图像的最感兴趣的区域的数据来确定用于一或多个最感兴趣的区域的高质量视频数据的存储持续时间。举例来说,当服务器装置60确定高质量视频数据对于最感兴趣的区域较高时,服务器装置60可将所述高质量视频数据分级为较高,使得用于最感兴趣的区域的高质量视频数据保持长期存储,同时去除其它视频数据(例如,用于其它区域),所述其它视频数据在与用于最感兴趣的区域的高质量视频数据类似的时间处检索。

[0175] 客户端装置40可输出预提取用于一或多个最感兴趣的区域的高质量视频数据的请求(366)。举例来说,客户端装置40可使用与图像的其它区域相比相对较高的位速率来请求用于一或多个最感兴趣的区域的视频数据。服务器装置60可接收所述预提取请求(368)且可发送用于一或多个最感兴趣的区域的高质量视频数据(370)。客户端装置40可接收所述高质量视频数据(372)。

[0176] 客户端装置40可输出对用于图像的一或多个其它区域的低质量视频数据的请求(374)。服务器装置60可接收对用于图像的一或多个其它区域的低质量视频数据的请求(376)。响应于接收到对用于图像的一或多个其它区域的低质量视频数据的请求,服务器装置60可提取用于图像的一或多个其它区域的低质量视频数据(378)且发送用于图像的一或多个其它区域的低质量视频数据(380)。客户端装置40可接收所述低质量视频数据(382)。尽管图7的实例中未展示,但客户端装置40接着可继续进行到(例如)基于用户的视场向用户呈现所接收视频数据的至少部分,例如,如上文关于图5所论述。

[0177] 以这种方式,图7的方法表示方法的一实例,所述方法包含:通过客户端装置的处理器的(所述处理器实施于电路中)基于表示视频数据图像的多个区域的一或多个最感兴趣的区域的数据来确定所述一或多个最感兴趣的区域,表示所述一或多个最感兴趣的区域的数据在包含所述视频数据的位流的视频译码层(VCL)数据外部;通过客户端装置的处理器的输出对用于图像的一或多个最感兴趣的区域的相对高质量视频数据的请求;以及在输出对所述相对高质量视频数据的请求后,通过客户端装置的处理器的输出对用于图像的一或多个其它区域的相对低质量视频数据的请求。

[0178] 同样地,图7的方法还表示方法的一实例,所述方法包含:通过源装置的处理器的(所述处理器实施于电路中)确定视频数据图像的多个区域的一或多个最感兴趣的区域,所述一或多个最感兴趣的区域包括客户端装置最可能检索的一或多个区域;以及通过源装置的处理器的产生表示所述一或多个最感兴趣的区域的数据,其中表示所述一或多个最感兴趣的区域的数据在包含所述视频数据的位流的视频译码层(VCL)数据外部。

[0179] 在一或多个实例中,所描述功能可以硬件、软件、固件或其任何组合来实施。如果以软件实施,那么所述功能可作为一或多个指令或代码而存储于计算机可读媒体上或经由计算机可读媒体进行发射,且通过基于硬件的处理单元来执行。计算机可读媒体可包含:计算机可读存储媒体,其对应于例如数据存储媒体的有形媒体;或通信媒体,其包含例如根据通信协议促进计算机程序从一处到另一处的传送的任何媒体。以这种方式,计算机可读媒体大体可对应于(1)非暂时性的有形计算机可读存储媒体,或(2)通信媒体,例如信号或载波。数据存储媒体可为可由一或多个计算机或一或多个处理器存取以检索指令、代码及/或数据结构以用于实施本发明中所描述的技术的任何可用媒体。计算机程序产品可包含计算机可读媒体。

[0180] 借助于实例而非限制,这类计算机可读存储媒体可包括RAM、ROM、EEPROM、CD-ROM或其它光盘存储装置、磁盘存储装置或其它磁性存储装置、快闪存储器或可用以存储呈指令或数据结构形式的所要程序码且可由计算机存取的任何其它媒体。而且,将任何连接恰当地称为计算机可读媒体。举例来说,如果使用同轴缆线、光纤缆线、双绞线、数字订户线(DSL)或例如红外线、无线电及微波的无线技术从网站、服务器或其它远程源发射指令,那么同轴缆线、光纤缆线、双绞线、DSL或例如红外线、无线电及微波的无线技术包含于媒体的定义中。然而,应理解,计算机可读存储媒体及数据存储媒体不包含连接、载波、信号或其它暂时性媒体,而是针对非暂时性有形存储媒体。如本文中所使用,磁盘及光盘包含压缩光盘(CD)、激光光盘、光学光盘、数字影音光盘(DVD)、软盘及蓝光光盘,其中磁盘通常以磁性方式再现数据,而光盘通过激光以光学方式再现数据。以上各者的组合还应包含于计算机可读媒体的范围内。

[0181] 指令可由一或多个处理器执行,所述一或多个处理器例如一或多个数字信号处理器(DSP)、通用微处理器、专用集成电路(ASIC)、场可编程逻辑阵列(FPGA)或其它等效集成式或离散逻辑电路。因此,如本文中所使用的术语“处理器”可指上述结构或适用于实施本文中所描述的技术的任何其它结构中的任一者。另外,在一些方面中,本文中所描述的功能性可提供于经配置用于编码及解码的专用硬件及/或软件模块内,或并入组合式编解码器中。而且,所述技术可完全实施于一或多个电路或逻辑元件中。

[0182] 本发明的技术可在广泛多种装置或设备中实施,所述装置或设备包含无线手机、集成电路(IC)或IC集合(例如,芯片组)。在本发明中描述各种组件、模块或单元以强调经配置以执行所揭示技术的装置的功能方面,但未必要由不同硬件单元来实现。确切而言,如上文所描述,可将各种单元组合于编解码器硬件单元中,或通过互操作性硬件单元(包含如上文所描述的一或多个处理器)的集合结合合适的软件及/或固件来提供所述单元。

[0183] 已描述各种实例。这些及其它实例在以下权利要求书的范围内。

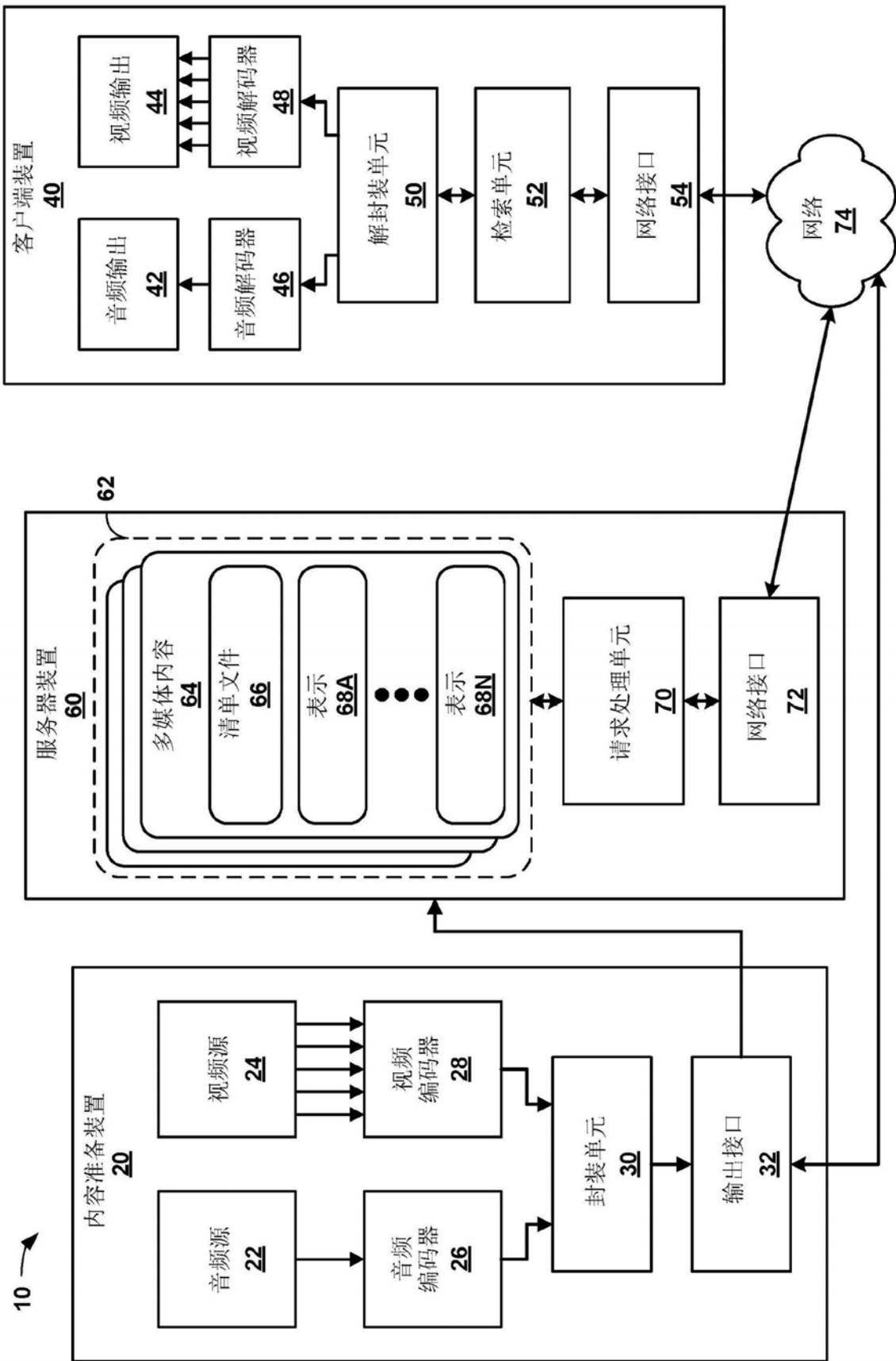


图1

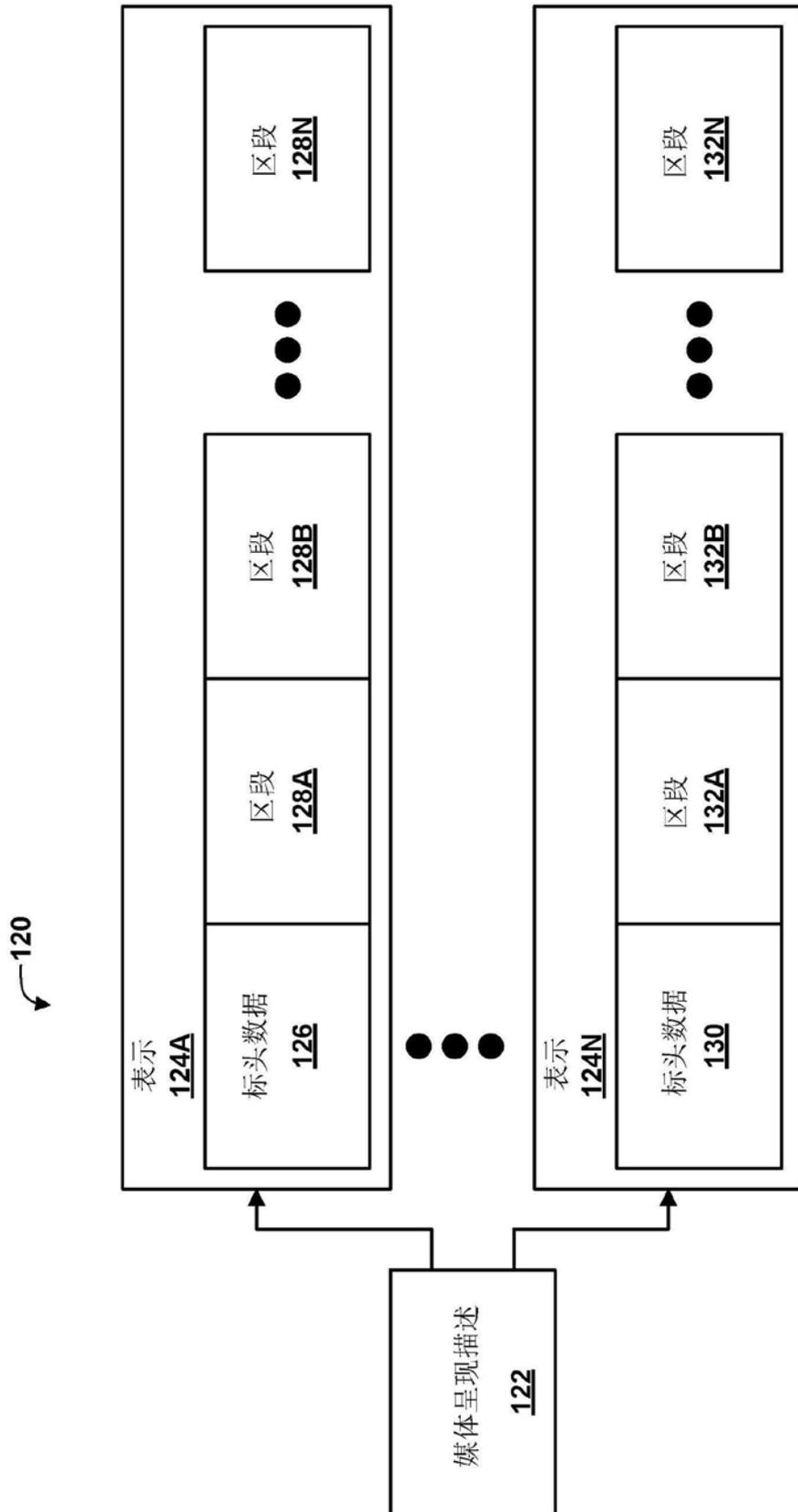


图2

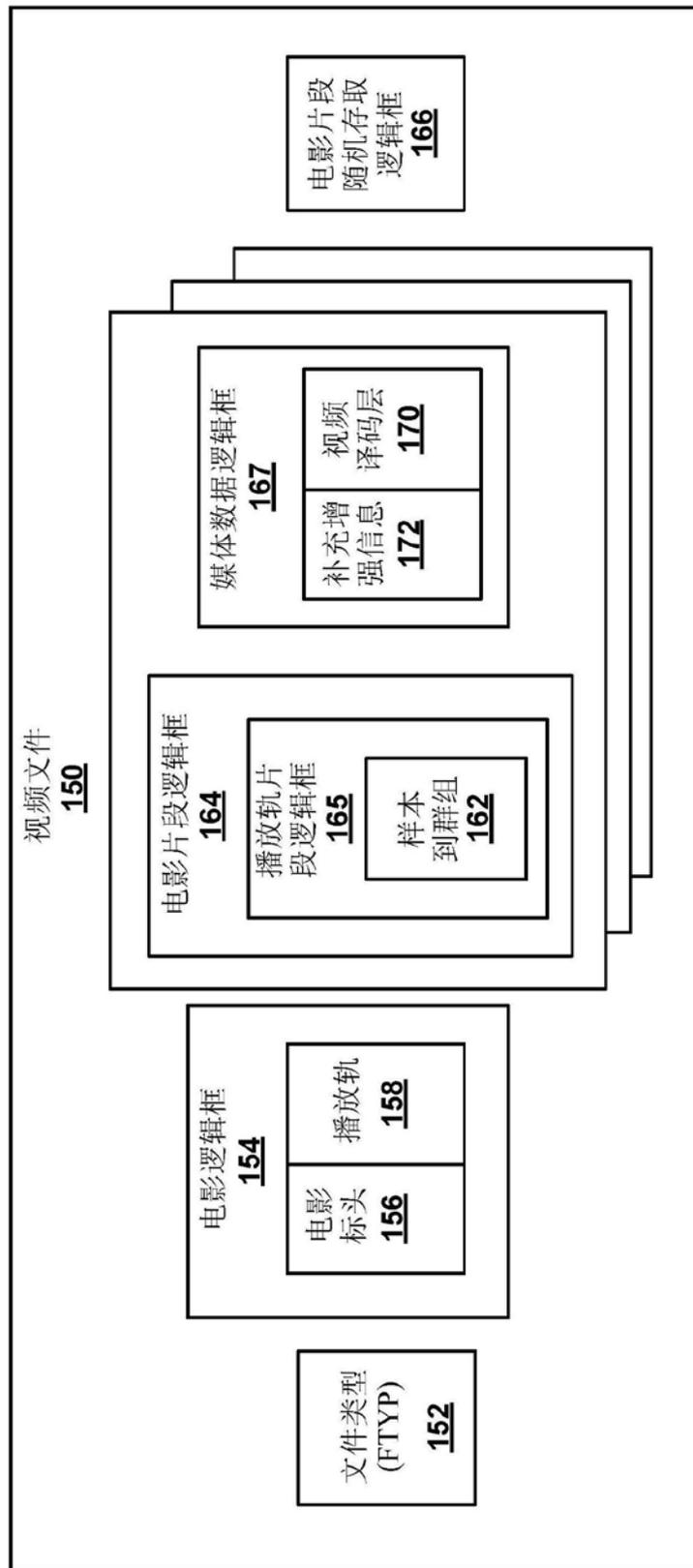


图3

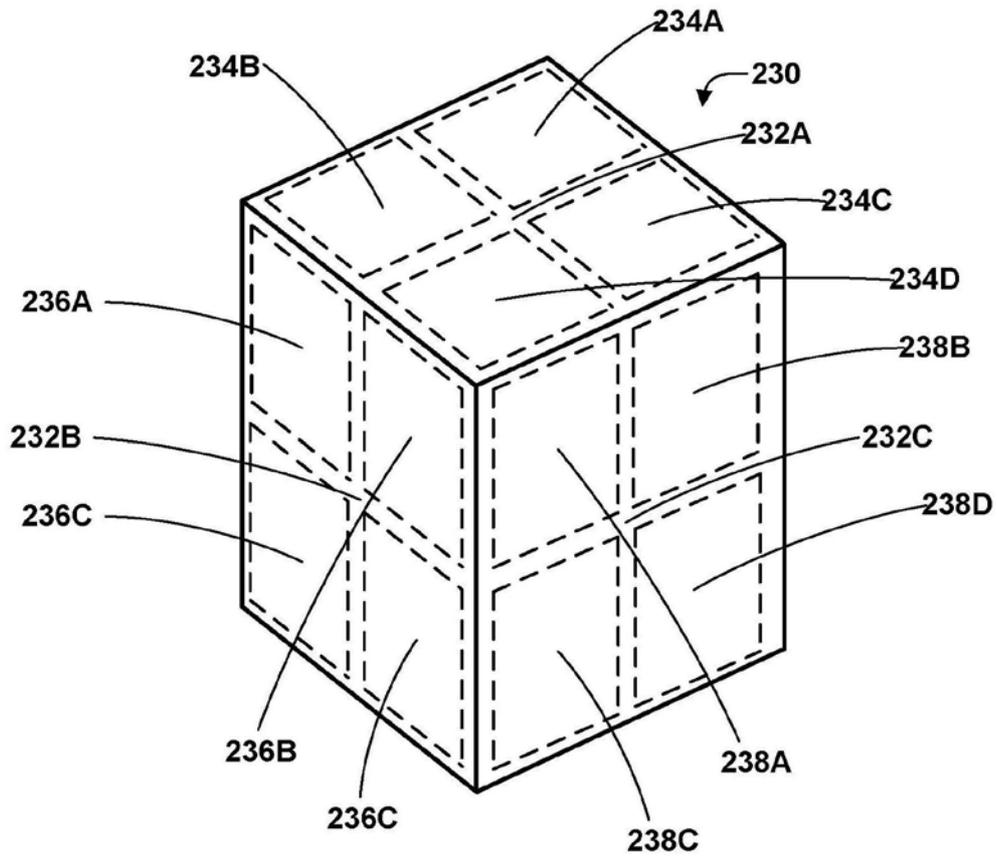


图4

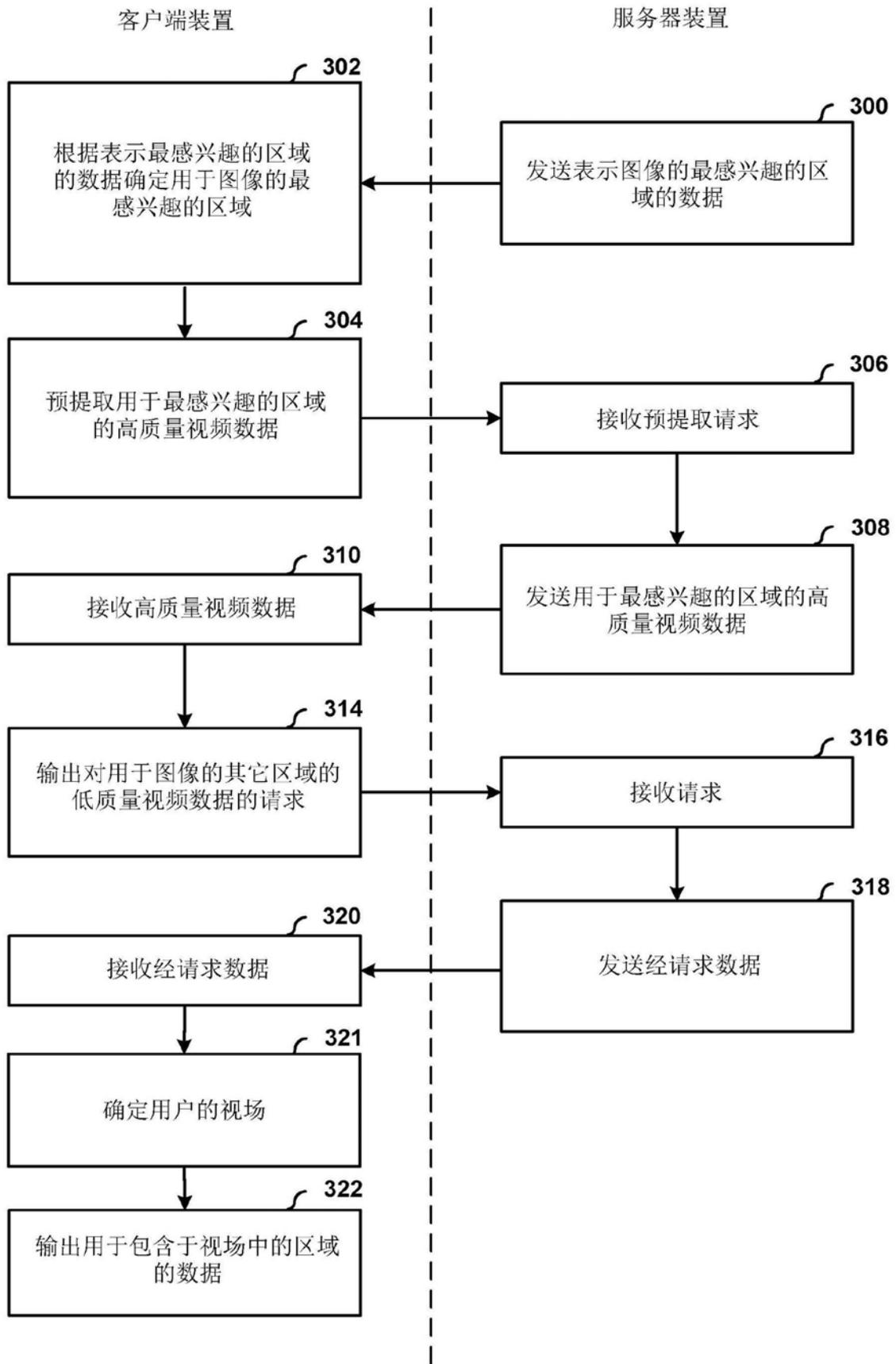


图5

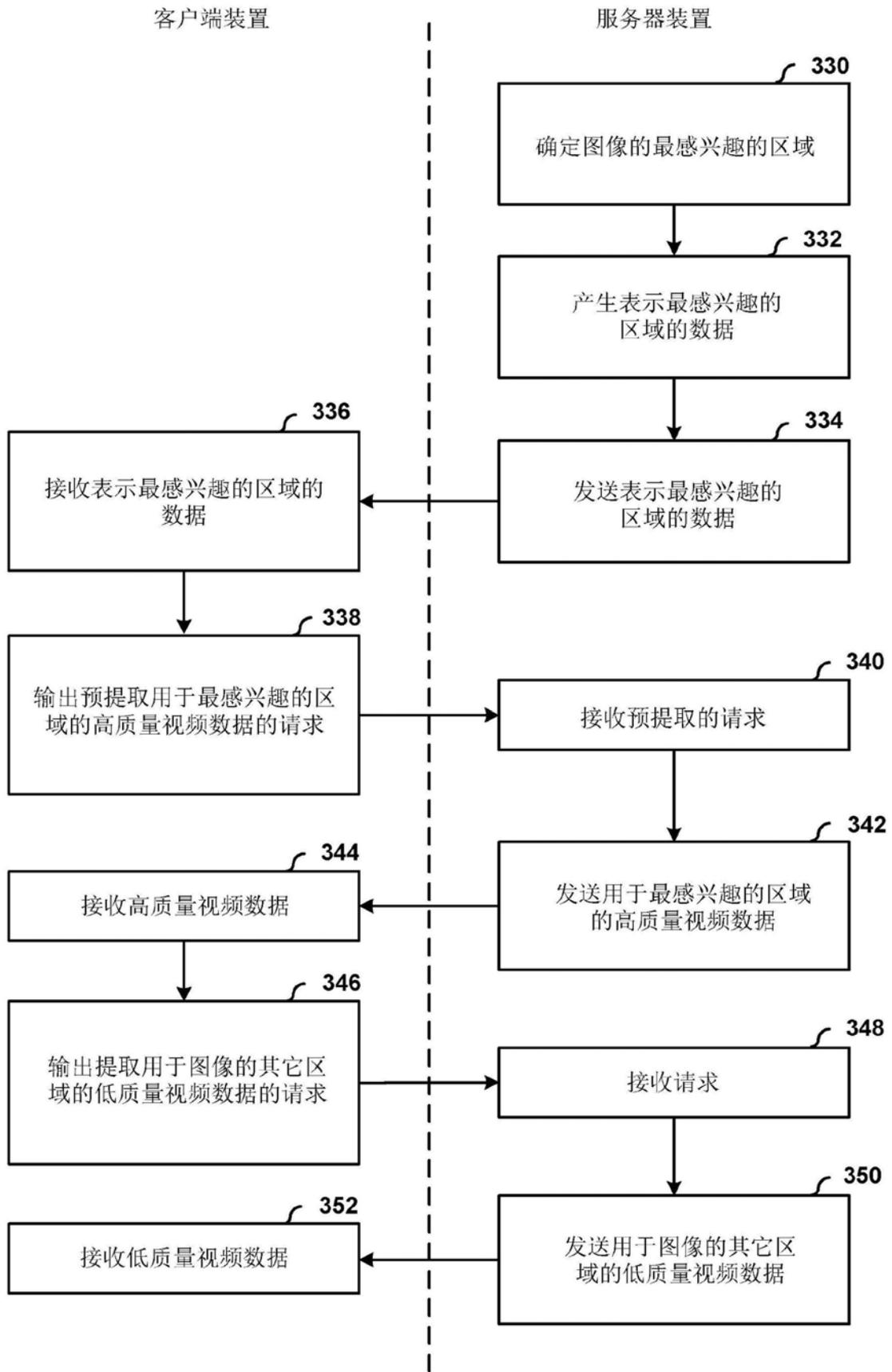


图6

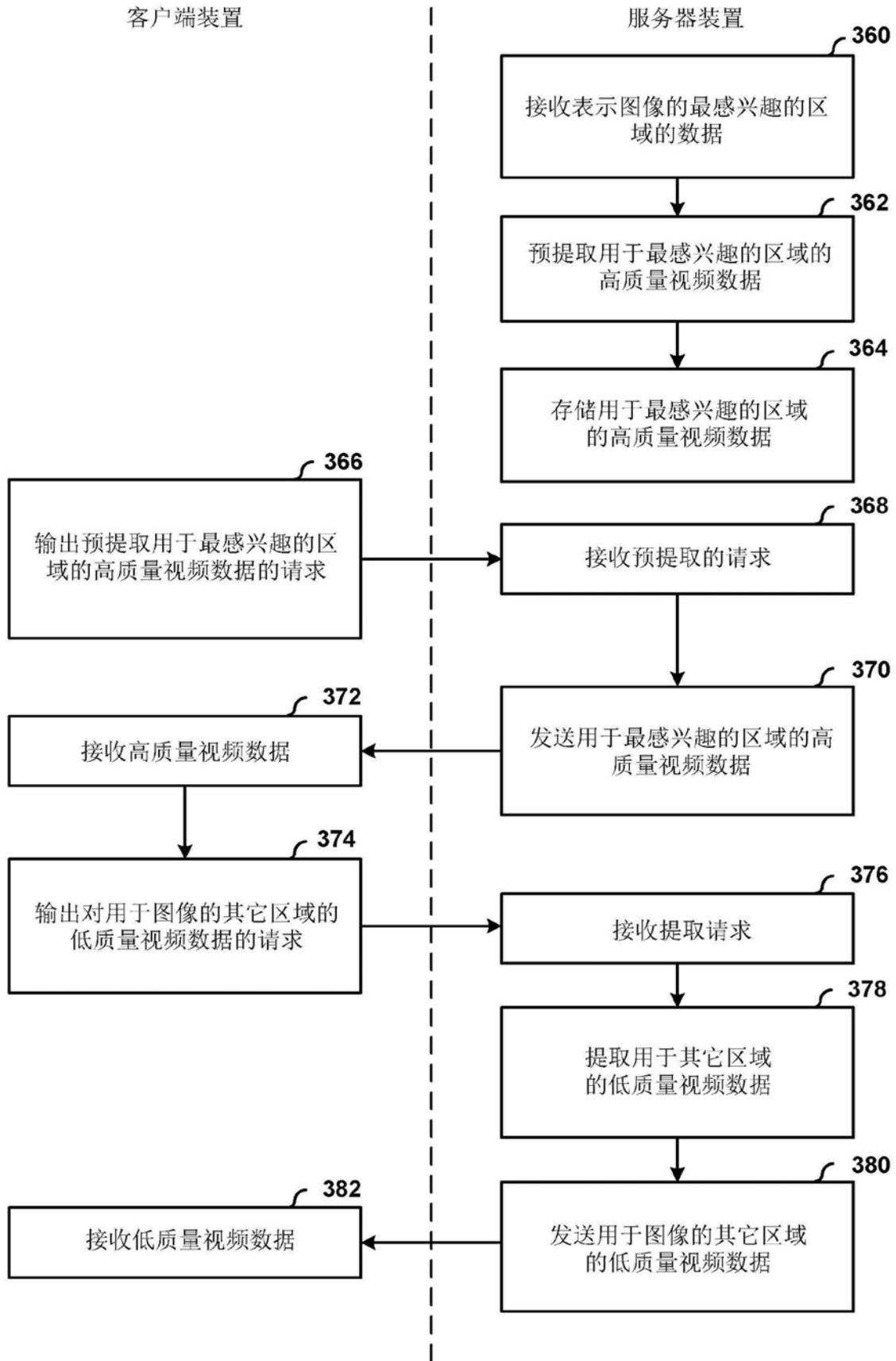


图7