



(12) 发明专利

(10) 授权公告号 CN 102272778 B

(45) 授权公告日 2015. 05. 20

(21) 申请号 200980153983. X

(56) 对比文件

(22) 申请日 2009. 12. 16

US 2005/0286759 A1, 2005. 12. 29, 说明书第 4-119 段, 权利要求 1-20 项.

(30) 优先权数据

61/204, 485 2009. 01. 07 US

审查员 刘宇儒

(85) PCT国际申请进入国家阶段日

2011. 07. 07

(86) PCT国际申请的申请数据

PCT/US2009/006568 2009. 12. 16

(87) PCT国际申请的公布数据

W02010/144074 EN 2010. 12. 16

(73) 专利权人 汤姆森特许公司

地址 法国伊西莱穆利诺

(72) 发明人 田栋 P-L. 莱

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 吕晓章

(51) Int. Cl.

G06K 9/36(2006. 01)

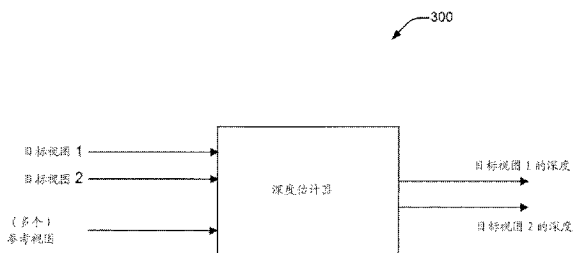
权利要求书3页 说明书12页 附图7页

(54) 发明名称

联合深度估计

(57) 摘要

本发明描述了多种实现方式。几种实现方式涉及对于多深度图的联合深度估计。在一种实现方式中,估计用于第一视图中的位置的第一视图深度指示值,并估计用于第二视图中的对应位置的第二视图深度指示值。基于约束来进行第一视图深度指示值和第二视图深度指示值中的一个或多个的估计。该约束提供了用于对应位置的第一视图深度指示值和第二视图深度指示值之间的关系。



1. 一种联合深度估计方法,包括:
估计用于第一视图中的位置的第一视图深度指示值,以及
估计用于第二视图中的对应位置的第二视图深度指示值,
其中,第一视图深度指示值和第二视图深度指示值中一个或多个的估计基于提供了用于对应位置的第一视图深度指示值和第二视图深度指示值之间的关系的约束,
其中,估计第一视图深度指示值和第二视图深度指示值包括:
依据由该约束提供的关系,确定各深度指示值具有相差多于深度指示值阈值的值;
确定用于第一视图深度指示值的第一视图失真;
确定用于第二视图深度指示值的第二视图失真;
在第一视图失真与第二视图失真之间的差值大于失真阈值的情况下,将第一视图中的位置标记为孔洞;以及
在第二视图失真与第一视图失真之间的差值大于该失真阈值的情况下,将第二视图中的对应位置标记为孔洞。
2. 如权利要求 1 所述的方法,其中,该约束基于一个或多个照相机参数。
3. 如权利要求 2 所述的方法,其中,一个或多个照相机参数包括以下至少一项:第一视图照相机的位置、第二视图照相机的位置、第三视图照相机的位置、第一视图照相机的像主点偏移、第二视图照相机的像主点偏移、或第三视图照相机的像主点偏移。
4. 如权利要求 1 所述的方法,其中,
第一视图深度指示值基于来自第一视图的视频和来自第三视图的视频,以及
第二视图深度指示值基于来自第二视图的视频和来自第三视图的视频。
5. 如权利要求 1 所述的方法,其中,估计第一视图深度指示值和第二视图深度指示值包括:
确定用于第一视图深度指示值的初始值;
确定用于第二视图深度指示值的初始值;以及
在用于第一视图深度指示值的初始值和用于第二视图深度指示值的初始值在满足该约束的给定阈值内的情况下,基于该约束来更新第一视图深度指示值和第二视图深度指示值中的至少一个。
6. 如权利要求 5 所述的方法,其中,确定用于第一视图深度指示值的初始值包括:
执行多个用于第一视图的视频和用于第三视图的视频的基于块的比较,每个基于块的比较基于用于第一视图深度指示值的不同值并导致失真;
确定用于第一视图深度指示值的使失真最小化的值,在多个基于块的比较中,所述使失真最小化的值导致在多个基于块的比较中具有最低失真的基于块的比较;以及
将用于第一视图深度指示值的初始值设置为所确定的使失真最小化的值。
7. 如权利要求 5 所述的方法,其中,更新包括:将第一视图深度指示值设置为基于用于第一视图深度指示值的初始值的幅度和用于第二视图深度指示值的初始值的幅度的平均值而更新的值。
8. 如权利要求 1 所述的方法,还包括:
比较第一视图深度指示值和第二视图深度指示值;以及
基于该比较,向第一视图深度指示值和第二视图深度指示值分配置信度的测量,其中,

在深度指示值相似时置信度的测量较高,而在深度指示值不相似时置信度的测量较低。

9. 如权利要求 1 所述的方法,还包括:使用用于在与所标记的位置相邻的各位置中指示最大深度的、与所标记的位置相邻的位置的深度指示值来填充该孔洞。

10. 如权利要求 1 所述的方法,其中,估计第一视图深度指示值和第二视图深度指示值包括:

确定用于第一视图深度指示值的初始值;以及

基于用于第一视图深度指示值的初始值以及基于该约束,确定用于第二视图深度指示值的初始值。

11. 如权利要求 10 所述的方法,其中,

对于用于第一视图深度指示值的初始值和用于第二视图深度指示值的初始值,确定联合失真;以及

基于该联合失真,确定是否选择用于第一视图深度指示值的初始值作为所估计的第一视图深度指示值、以及是否选择用于第二视图深度指示值的初始值作为所估计的第二视图深度指示值。

12. 如权利要求 11 所述的方法,其中,确定联合失真包括:将用于第一视图深度指示值的初始值的失真和用于第二视图深度指示值的初始值的失真求和。

13. 如权利要求 11 所述的方法,其中,确定联合失真包括:计算用于第一视图深度指示值的初始值的失真和用于第二视图深度指示值的初始值的失真加权,其中,权重基于包括第一视图照相机、第二视图照相机和第三视图照相机的组中各照相机之间的至少一个距离。

14. 如权利要求 1 所述的方法,其中,该方法在视频编码器或视频解码器的至少一个中实现。

15. 如权利要求 1 所述的方法,其中,第一视图深度指示值包括以下的一个或多个:用于第一视图中位置的深度值、或用于第一视图中位置和另一视图中对应位置之间的视差的视差值。

16. 如权利要求 1 所述的方法,还包括:基于第一视图深度指示值或第二视图深度指示值中的一个或多个,呈现附加视图。

17. 如权利要求 1 所述的方法,其中,第一视图中的位置是以下之一:第一视图中的像素、第一视图中的部分、第一视图中的子宏块、第一视图中的宏块、第一视图中的码片、或第一视图中的场。

18. 一种联合深度估计装置,包括:

用于估计用于第一视图中的位置的第一视图深度指示值的部件;以及

用于估计用于第二视图中的对应位置的第二视图深度指示值的部件,

其中,第一视图深度指示值和第二视图深度指示值中的一个或多个的估计基于提供了用于对应位置的第一视图深度指示值和第二视图深度指示值之间的关系的约束,

其中,估计第一视图深度指示值和第二视图深度指示值包括:

依据由该约束提供的关系,确定各深度指示值具有相差多于深度指示值阈值的值;

确定用于第一视图深度指示值的第一视图失真;

确定用于第二视图深度指示值的第二视图失真;

在第一视图失真与第二视图失真之间的差值大于失真阈值的情况下,将第一视图中的位置标记为孔洞;以及

在第二视图失真与第一视图失真之间的差值大于该失真阈值的情况下,将第二视图中的对应位置标记为孔洞。

19. 一种联合深度估计方法,包括:

基于照相机位置,利用一个或多个约束来估计用于给定像素位置的一对深度,其中,使用来自第一视图的视频和来自第三视图的视频来估计用于第一视图中给定像素位置的深度,而使用来自第二视图的视频和来自第三视图的视频来估计用于第二视图中给定像素位置的深度,

其中,第一视图中给定像素位置的深度的估计和第二视图中给定像素位置的深度的估计中的一个或多个基于提供了用于对应位置的第一视图中给定像素位置的深度和第二视图中给定像素位置的深度之间的关系的约束,

其中,估计第一视图深度指示值和第二视图深度指示值包括:

依据由该约束提供的关系,确定各深度指示值具有相差多于深度指示值阈值的值;

确定用于第一视图深度指示值的第一视图失真;

确定用于第二视图深度指示值的第二视图失真;

在第一视图失真与第二视图失真之间的差值大于失真阈值的情况下,将第一视图中的位置标记为孔洞;以及

在第二视图失真与第一视图失真之间的差值大于该失真阈值的情况下,将第二视图中的对应位置标记为孔洞。

20. 如权利要求 1 所述的方法,其中,第一视图深度指示值和第二视图深度指示值每个都是从深度值、整个深度图、视差值和整个视差图之一中选择的。

21. 如权利要求 1 所述的方法,其中,所述约束基于包括与视图相关联的照相机的每个镜头的焦距长度、测量照相机距离的照相机之间的基线间距、以及照相机的像主点偏移的差的比例等式。

22. 如权利要求 21 所述的方法,其中,所有照相机共享相同的焦距,所有照相机均等地间隔开,每个像主点等于零。

联合深度估计

[0001] 相关申请的交叉引用

[0002] 本申请要求 2009 年 1 月 7 日提交的题为“Joint Depth Estimation for Multiple Depth Maps”的美国临时申请序列号 61/204,485 的权益,通过引用将其全部内容并入于此以用于所有目的。

技术领域

[0003] 描述涉及编码系统的实现方式。多种具体的实现方式涉及联合深度估计。

背景技术

[0004] 三维视频 (3DV) 是针对在接收器侧产生高质量 3D 呈现的框架。这使能了利用自动立体显示器、自由视点应用、和立体显示器进行 3D 视觉体验。在 3DV 应用中,由于在传输带宽中的限制和 / 或存储约束而传送或存储降低数量的视频视图和对应的深度图 (典型地被称为多视图加深度 (MVD))。在接收器侧,使用可用的视图和深度图来呈现附加的视图。

发明内容

[0005] 根据总的方面,估计用于第一视图中的位置的第一视图深度指示值,以及估计用于第二视图中的对应位置的第二视图深度指示值。基于约束来进行第一视图深度指示值和第二视图深度指示值中的一个或多个的估计。该约束提供了用于对应位置的第一视图深度指示值和第二视图深度指示值之间的关系。

[0006] 在附图和以下描述中提出了一种或多种实现方式的细节。即使以一种特定方式描述,也应清楚可以以多种方式来配置或体现多种实现方式。例如,可以将实现方式执行为方法、或者可以将其体现为装置 (诸如,例如被配置为执行一组操作的装置、或存储用于执行一组操作的指令的装置)、或者可以将其体现在信号中。结合附图和权利要求书来考虑以下描述,其它方面和特征将变得明显。

附图说明

[0007] 图 1 是来自左参考视图的左深度图的示例。

[0008] 图 2 是来自右参考视图的右深度图的示例。

[0009] 图 3 是深度估计器的实现方式的图。

[0010] 图 4 是视频传输系统的实现方式的图。

[0011] 图 5 是视频接收系统的实现方式的图。

[0012] 图 6 是视频处理设备的实现方式的图。

[0013] 图 7 包括图 7A 和图 7B,是第一深度估计处理的实现方式的图。

[0014] 图 8 是第二深度估计处理的实现方式的图。

具体实施方式

[0015] 在许多 3DV 应用中,可以产生中间视图(虚拟视图),并且这样的产生典型地被称为呈现。可以使用例如基于深度图像的呈现(DIBR)技术来执行呈现,所述基于深度图像的呈现(DIBR)技术采用所传送/所存储的视图(参考视图)和相关联的每像素深度图作为输入。这样的输入例如可以由 MVD 格式来提供。可以使用多种技术中的任一种来捕获深度。然而,经常仅提供视频,而估计深度。为了获得深度图,经常使用深度估计技术来找到各不同视图之间的对应性。

[0016] 尤其在使用深度估计来获得输入深度图时,生成虚拟视图可能是有挑战的任务。即,由于所估计的深度图可能是有噪声的(不准确的),并且可能没有其它的场景信息(诸如,场景的 3D 表面属性)可用。有噪声的深度图可能是不准确的,这是因为它们不具有正确的深度值。另外,有噪声的深度图的序列可能具有不是一致性地(consistently)不准确的不准确值。例如,处于恒定实际深度处的对象可能在第一画面中利用低深度被不准确地估计,而在第二画面中利用高深度被不准确地估计。

[0017] 如果更多参考视图可用于 DIBR,则典型地可以实现较佳的视图呈现质量。例如,可以使用两个参考视图来呈现中间视图。在多参考视图的情况下,如果不是已经知道每个参考视图的深度图,则估计每个参考视图的深度图。然而,不同视图的多深度图典型地是被彼此独立地估计的。如所提及的,DIBR 方法的挑战之一在于以下事实:在各视图间所估计的深度图经常不一致。在利用两个或多个不一致的深度图来呈现虚拟视图时,这可能导致差的呈现质量。

[0018] 在至少一种实现方式中,我们提出利用联合深度估计来估计参考视图的深度图。联合深度估计在联合处理中而不是独立地执行对于多参考视图的深度估计。联合深度估计获得更一致的不同视图的深度图,导致了在 DIBR 呈现的视图中较佳的质量。

[0019] 在至少一种实现方式中,基于中间视图产生两个深度图。该实现方式发展了以下约束:提供两个深度图中对应位置之间的关系。可以以多种方式使用该约束来提供两个深度图之间的一致性。

[0020] 在至少一种实现方式中,联合深度估计涉及对用于两个或多个视图中对应像素位置之间的视差的各视差估计成本(失真)执行的求和操作。

[0021] 在至少一种实现方式中,联合深度估计涉及使用基于照相机距离的求和操作。在至少一种实现方式中,基于照相机距离来对求和操作进行加权。

[0022] 图 1 示出了根据本原理实施例的为与被已知为“Leaving_Laptop”的 MPEG 测试序列对应的左参考视图产生的示例的左深度图,可以向其应用本原理。图 2 示出了根据本原理实施例的为与被已知为“Leaving_Laptop”的 MPEG 测试序列对应的右参考视图产生的示例的右深度图,可以向其应用本原理。尽管依据所述图不是显而易见的,但是可以观察到:对于一些具体区域,深度图对中的深度级可能有很大区别。具体地,在图 1 和 2 的示例中,区别可能大于 40。在这些情况下,理想的区别是零,但是所观察到的区别为例如大于 40。这是不具有一致的深度图的视图的示例。

[0023] 关于估计一个深度图所需的输入视图的数量,可以典型地将深度估计算法划分为以下种类:单视图;立体视图;以及多视图。所有这三种种类都假设深度图是未知的,并且都使用来自一个或多个视图的视频来产生深度图。

[0024] 在典型实现方式中,在仅一个单视图作为输入的情况下,照相机焦距被考虑为深

度估计的平均。可以使用被称为“依据焦距的深度 (depth from focus)”的方法,基于散焦或模糊的量来估计深度。该方法可能不是非常可靠,这是因为例如焦距估计经常不会提供很好的结果。

[0025] 立体视图方法可以使用一对视图作为输入来估计用于所述视图之一的深度图。对于基于区域的立体匹配方法,这样的方法典型地在两个图像之间匹配一窗口内的相邻像素值。选择适当的窗口大小典型地是关键。在一些应用中,可以基于局部亮度变化和当前深度估计来迭代地改变窗口大小和形状。可以应用一些全局约束来产生稠密深度图,即,具有唯一值并且几乎在任何位置都连续的深度图。

[0026] 随着出现多视图内容,已经探索使用多视图作为输入来进行深度估计。一种可能的方法是使用三个视图作为输入(左视图、中间视图、和右视图)来估计用于中间视图的深度。

[0027] 在本领域的典型方法中,目标是产生单个稠密深度图序列。当需要估计多于一个深度图序列时,无论深度估计算法是什么种类,都独立地对每个目标视图执行深度估计。因此,很有可能出现各视图间的不一致性。

[0028] 图 3 示出了依据本原理实施例的可以向其应用本原理的示例性深度估计器 300。深度估计器 300 接收目标视图 1、目标视图 2、以及一个或多个参考视图作为输入。深度估计器 300 提供目标视图 1 的估计深度和目标视图 2 的估计深度作为输出。下面将更详细地描述深度估计器 300 的操作。

[0029] 更一般地,深度估计器 300 提供用于目标视图 1 的估计深度指示值和用于目标视图 2 的估计深度指示值。深度指示值可以是深度值、或整个深度图。但是深度指示值可以替代地是例如视差值、或整个视差图。在接下来的实现方式和描述中对深度的引用意图包括其它深度指示值,诸如例如视差。深度指示值可以提供用于例如整个目标视图或目标视图中一位置的深度指示。该位置可以是例如具体像素、部分、子宏块、宏块、码片或场。

[0030] 图 4 示出了依据本发明实现方式的可以向其应用本原理的示例视频传输系统 400。视频传输系统 400 可以是例如用于使用多种介质(诸如例如卫星、电缆、电话线、或地面广播)中任一种来传送信号的前端或传输系统。可以在因特网或某些其它网络上提供传输。

[0031] 视频传输系统 400 能够产生并传递具有深度的压缩视频。这通过产生包括深度信息或能够被用来在(可能例如具有解码器的)接收器端合成深度信息的信息的(多个)编码信号来实现。

[0032] 视频传输系统 400 包括编码器 410 和能够传送编码信号的发射器 420。编码器 410 接收视频信息并且产生具有深度的(多个)编码信号。编码器 410 可以包括子模块,包括例如用于接收并组装多个信息为用于存储或传输的构造格式的组装单元。多个信息可以包括例如编码或未编码的视频、编码或未编码的深度信息、以及编码或未编码的元素(诸如例如运动向量、编码模式 指示值、和语法元素)。

[0033] 发射器 420 可以例如被适配为传送具有一个或多个比特流的节目信号,其表示编码画面和/或与其有关的信息。典型的发射器执行以下功能中的一个或多个,诸如例如提供纠错编码、将数据交织在信号中、将信号中的能量随机化、以及将信号调制到一个或多个载波上。发射器可以包括天线(未示出)、或者与天线(未示出)接口。相应地,发射器 420

的实现方式可以包括或被限于调制器。

[0034] 视频传输系统 400 还可以被全部或部分地包括在多种用户设备中。这样的设备包括例如蜂窝电话、膝上型或其它计算机、和摄像机。

[0035] 图 5 示出了依据本原理实施例的可以向其应用本原理的示例性视频接收系统 500。视频接收系统 500 可以被配置为在多种介质（诸如例如卫星、电缆、电话线、或地面广播）上接收信号。可以在因特网或某些其它网络上接收信号。

[0036] 视频接收系统 500 可以是例如蜂窝电话、计算机、机顶盒、电视机、或接收编码视频并提供例如解码视频以供向用户显示或存储的其它设备。因此，视频接收系统 500 可以将其输出提供至例如电视机的屏幕、计算机监视器、计算机（以供存储、处理、或显示）、或者某些其它存储、处理或显示设备。

[0037] 视频接收系统 500 能够接收并处理包括视频信息的视频内容。视频接收系统 500 包括能够接收编码信号（诸如例如在本申请实现方式中描述的信号）的接收器 510、以及能够解码所接收的信号的解码器 520。

[0038] 接收器 510 可以例如被配置为接收具有多个比特流的节目信号，其表示编码画面。典型的接收器执行以下功能中的一个或多个，诸如例如接收调制且编码的数据信号、从一个或多个载波中解调数据信号、将信号中的能量去随机化、将信号中的数据去交织、以及对信号进行纠错解码。接收器 510 可以包括天线（未示出）、或者可以与天线（未示出）接口。接收器 510 的实现方式可以包括或被限于解调器。解码器 520 输出包括例如视频信息和深度信息的视频信号。

[0039] 图 6 示出了依据本原理实施例的可以向其应用本原理的示例性视频处理设备 600。视频处理设备 600 可以是例如机顶盒、或接收编码视频并提供例如解码视频以供向用户显示或存储的其它设备。因此，视频处理设备 600 可以将其输出提供到电视机、计算机监视器、或者计算机或其它处理设备。

[0040] 视频处理设备 600 包括前端 (FE) 设备 605 和解码器 610。前端设备 605 可以是例如被适配为接收具有多个表示编码画面的比特流的节目信号、并从多个比特流中选择要解码的一个或多个比特流的接收器。典型的接收器执行以下功能中的一个或多个，诸如例如接收调制且编码的数据信号、解调数据信号、对数据信号的一个或多个编码（例如，信道编码和 / 或源编码）进行解码、以及 / 或者对信号进行纠错。前端设备 605 可以从例如天线（未示出）接收节目信号。前端设备 605 将所接收的数据信号提供至解码器 610。

[0041] 解码器 610 接收数据信号 620。数据信号 620 可以包括例如一个或多个兼容高级视频编码 (AVC)、可缩放视频编码 (SVC)、或多视图视频编码 (MVC) 的流。

[0042] AVC 更具体地指代现有的国际标准化组织 / 国际电工委员会 (ISO/IEC) 运动画面专家组 -4 (MPEG-4) 第 10 部分高级视频编码 (AVC) 标准 / 国际电信联盟电信分部 (ITU-T) H. 264 建议（以下称为“H. 264/MPEG-4AVC 标准”或其变型，诸如“AVC 标准”或简单地“AVC”）。

[0043] MVC 更具体地指代 AVC 标准的多视图视频编码（“MVC”）扩展（附录 H），被称为 H. 264/MPEG-4AVC MVC 扩展（“MVC 扩展”或简单地“MVC”）。

[0044] SVC 更具体地指代 AVC 标准的可缩放视频编码（“SVC”）扩展（附录 G），被称为 H. 264/MPEG-4AVC SVC 扩展（“SVC 扩展”或简单地“SVC”）。

[0045] 解码器 610 对所接收的信号 620 的全部或部分进行解码,并且提供解码视频信号 630 作为输出。解码视频 630 被提供给选择器 650。设备 600 还包括接收用户输入 670 的用户接口 660。用户接口 660 基于用户输入 670 将画面选择信号 680 提供给选择器 650。画面选择信号 680 和用户输入 670 指示用户期望显示多画面、序列、可缩放版本、视图、或可用解码数据的其它选择中的哪个。选择器 650 提供所选择的画面作为输出 690。选择器 650 使用画面选择信号 680 来选择提供解码视频 630 中的哪个画面作为输出 690。

[0046] 在各种实现方式中,选择器 650 包括用户接口 660,而在其它实现方式中,由于选择器 650 直接接收用户输入 670 而不执行单独的接口功能,因而不需要用户接口 660。例如,选择器 650 可以以软件形式实现,或者可以实现为集成电路。在一种实现方式中,选择器 650 与解码器 610 合并,而在另一实现方式中,解码器 610、选择器 650 和用户接口 660 全部被集成。

[0047] 在一个应用中,前端 605 接收各种电视节目 (show) 的广播,并选择一个用于处理。一个节目的选择基于期望观看频道的用户输入。尽管在图 6 中未示出对前端设备 605 的用户输入,但前端设备 605 接收用户输入 670。前端设备 605 接收广播,并且通过解调广播频谱的相关部分并对解调节目的任何外部编码进行解码来处理期望节目。前端设备 605 将解码节目提供给解码器 610。解码器 610 是包括设备 660 和 650 的集成单元。因此,解码器 610 接收用户输入,该用户输入是用户提供的对节目中期望观看的视图的指示。解码器 610 对所选择的视图进行解码,以及从其它视图中解码任何需要的参考画面,并且提供解码视图 690 以供在电视 (未示出) 上显示。

[0048] 继续以上应用,用户可能期望切换被显示的视图,并且然后可能向解码器 610 提供新的输入。在从用户接收到“视图改变”之后,解码器 610 解码老视图和新视图两者、以及介于老视图和新视图之间的任何视图。即,解码器 610 对从物理地位于拍摄老视图的照相机和拍摄新视图的照相机之间的照相机拍摄的任何视图进行解码。前端设备 605 还接收标识老视图、新视图、以及中间视图的信息。这种信息可以由例如具有关于各视图的位置的信息的控制器 (图 6 中未示出)、或者解码器 610 来提供。其它实现方式可以使用具有与前端设备集成在一起的控制器的前端设备。

[0049] 解码器 610 提供所有这些解码视图作为输出 690。后处理器 (图 6 中未示出) 在各视图之间进行内插以便提供从老视图到新视图的平滑过渡,并且向用户显示该过渡。在过渡到新视图之后,后处理器 (通过未示出的一个或多个通信链路) 向解码器 610 和前端设备 605 通知仅需要新视图。此后,解码器 610 仅提供新视图作为输出 690。

[0050] 可以使用系统 600 来接收图像序列的多视图,呈现用于显示的单个视图,并且以平滑的方式在多个视图之间进行切换。平滑的方式可以包含在各视图之间进行内插以便移到另一视图。另外,系统 600 可以允许用户旋转对象或场景、或者不同地看到对象或场景的三维表示。对象的旋转例如可以对应于从一视图移到另一视图、以及在各视图之间进行内插以便获得各视图之间的平滑过渡或简单地获得三维表示。即,用户可以“选择”内插视图作为要显示的“视图”。

[0051] 在至少一种所描述的实现方式中,提出了以联合方式而不是独立地执行用于目标视图的深度估计。

[0052] 对于照相机几何学,使用以下公式可转换两个视图之间的视差 (d) 和对象深度

(z) :

$$[0053] \quad d = \frac{f \cdot l}{z} + du \quad (1)$$

[0054] 在公式 (1) 中, f 是照相机镜头的焦距, l 是基线间距 (也已知为照相机距离), du 是像主点 (principal point) 偏移的差。

[0055] 为了简要地解释 du , 对于给定照相机 i , 设 u_i 指示图像的中心和光学系统的中心之间的差。即, 对于照相机 i , u_i 是从光学中心的中心的图像的中心的偏移。则 du 或更具体地 $du_{i,j}$ 为 u_i 和 u_j 之间的差, 其中 i 和 j 是两个照相机 / 视图。

[0056] 视差指代一个视图中的一个位置与另一视图中的对应位置之间的移位或转移 (translation)。深度指代从照相机平面到视频中对象的距离。视差和深度两者可能逐像素地变化。深度是应用于单个视图的值。另外, 焦距应用于单个视图。然而, 视差和 du 是描述两个视图内的对应像素之间的关系的相对值。另外, l 描述两个照相机之间的距离, 或者更具体地 l 描述两个视图 (处于视图位置的照相机) 之间的关系 (距离)。

[0057] 注意, 在许多深度估计方法中, 首先估计视差 (d), 然后将其转换为深度 (z), 这是因为视差是两个视图之间的对应性、并且可以通过利用诸如均方差 (MSE) 或绝对差之和 (SAD) 等之类的成本函数 (也被称为失真) 来搜索其它 (多个) 视图中的最相似像素来识别。

[0058] 不失一般性, 我们提供了可以如何基于所捕获的视图来产生深度图的示例。我们假设: 具有用于 5 个所捕获的视图 (即, 视图 1 到视图 5) 的视频, 并且希望估计用于视图 2 和视图 4 的深度图。在一个可能处理中, 使用来自视图 1、2 和 3 的视频作为输入, 以便估计视图 2 的深度图。类似地, 使用来自视图 3、4 和 5 的视频作为输入, 以便估计视图 4 的深度图。以此方式, 独立地产生视图 2 和视图 4 的深度图。值得注意, 在刚刚描述的两个深度图估计处理中共享视图 3。但用于视图 2 和视图 4 的深度估计仍是独立的。

[0059] 我们提出了以以下方式使用来自视图 3 的视频来链接两个估计处理的至少一个实施例。对于视图 3 中的每个像素, 设 d_{32} 表示从视图 3 到视图 2 的视差, 并且设 d_{34} 表示从视图 3 到视图 4 的视差。当该像素在三个视图中的任一个中没有被遮挡时, 即, d_{32} 和 d_{34} 是有效的视差, 在公式 (1) 中的 du 为零并且各照相机被均等地间隔 (相同的 l) 并且所有照相机共享相同的焦距 (f) 的情况下, 它们将满足公式 (2)。公式 (2) 如下:

$$[0060] \quad d_{32} = -d_{34} \quad (2)$$

[0061] 在没有进行 du 为零、均等间隔的照相机和相同的焦距的简化的情况下, 可以利用更复杂的形式来替代公式 (2), 如下面所解释的。换句话说, 仍可以确定 d_{32} 和 d_{34} 之间的关系 (即, 存在 d_{32} 和 d_{34} 应满足的约束)。

[0062] 可以将 d_{32} 和 d_{34} 应满足某种约束的事实用于对视图 2 的深度图和视图 4 的深度图的联合深度估计。

[0063] 现在形成更一般的约束。

[0064] 依据公式 (1), 我们具有:

$$[0065] \quad d_{32} = \frac{f_3 l_{32}}{z_3} + du_{32} \quad (3)$$

[0066] 可以将公式 (3) 重新排列为:

$$[0067] \quad z_3 = \frac{f_3 l_{32}}{d_{32} - du_{32}} \quad (4)$$

[0068] 类似地,依据公式 (1),我们具有:

$$[0069] \quad d_{34} = \frac{f_3 l_{34}}{z_3} + du_{34} \quad (5)$$

[0070] 可以将公式 (5) 重新排列为:

$$[0071] \quad z_3 = \frac{f_3 l_{34}}{d_{34} - du_{34}} \quad (6)$$

[0072] 公式 (4) 和 (6) 的右手侧可以被等同为:

$$[0073] \quad \frac{f_3 l_{32}}{d_{32} - du_{32}} = \frac{f_3 l_{34}}{d_{34} - du_{34}} \quad (7)$$

[0074] 可以将公式 (7) 重新排列为:

$$[0075] \quad (d_{34} - du_{34}) = \frac{f_3 l_{34}}{f_3 l_{32}} (d_{32} - du_{32}) \quad (8)$$

[0076] 公式 (8) 简化为:

[0077]

$$(d_{34} - du_{34}) = \frac{l_{34}}{l_{32}} (d_{32} - du_{32}) \quad (9)$$

[0078] 如果各照相机被均等地间隔开,则 $l_{34} = -l_{32}$ 。相应地,如果 du 值为零并且各照相机被均等地间隔开,则公式 (9) 简化为公式 (2)。然而,公式 (9) 例示了 d_{32} 和 d_{34} 之间稍微更一般的约束。当然,其它约束也是可能的,并且可以设想其它约束。

[0079] 公式 (2) 和 (9) 两者都描述了提供对于两个视图中的对应位置的、用于一个视图的深度指示值 (视差) (例如, d_{32}) 和用于另一视图的深度指示值 (例如, d_{34}) 之间的关系的约束。所述位置典型地指代实际像素。公式 (2) 和 (9) 中的深度指示值可以等同地被表示为深度。

[0080] 该约束基于一个或多个照相机参数,诸如例如 du_{34} 、 du_{32} 、 l_{34} 和 l_{32} 。通过取两个视图的像主点偏移的差,可以从照相机参数中导出 du_{32} 和 du_{34} 。像主点偏移是固有的照相机参数之一。另外, f (照相机镜头的焦距) 也是固有的照相机参数,并且在其它实现方式中可以作为约束的一部分。可以依据非固有的照相机参数来计算 l_{34} 和 l_{32} (基线照相机间距)。

[0081] 实施例 1

[0082] 在该实施例中,提出了四个步骤。然而,变型可以包括不同数量的步骤。

[0083] 步骤 1. 对于视图 3 中的当前像素,分别估计并存储 d_{32} 和 d_{34} 。如果两个视差 d_{32} 和 d_{34} 满足 (或者在给定阈值内几乎满足) 公式 (2) 中的关系,指示两者是可靠的,则将例如使用如下面所提出的公式 (10) 来更新两个视差。然后,处理前进到下一像素,并且从步骤 1 执行。公式 (10) 如下:

$$[0084] \quad d'_{32} = (d_{32} - d_{34}) / 2 \quad (10)$$

$$[0085] \quad d'_{34} = -(d_{32} - d_{34}) / 2$$

[0086] 否则,如果 d_{32} 和 d_{34} 不在给定阈值内满足公式 (2),则进行到步骤 2。

[0087] 可以以各种方式来执行估计视差 (诸如例如 d_{32})。在一种实现方式中,在视图 3 中以当前像素为中心的块和视图 2 中以所选像素为中心的相似大小的块之间执行基于块的

比较。所选像素反映了具体视差。例如,基于块的比较可以是 SAD 的 MSE 的计算,对于该具体视差确定失真(或成本)。然后,可以改变所选像素(以及因此具体视差),并且可以执行新的基于块的比较并且可以确定新的失真。可以评估一系列可能的视差,并且可以选择导致最低失真的视差作为用于 d_{32} 的估计。

[0088] 步骤 2. 步骤 2 涉及当 d_{32} 和 d_{34} 不在给定阈值内满足公式 (2) 的情况。给定阈值可以是例如偏移或缩放因子。在一种实现方式中, d_{32} 的估计值必须在 d_{32} 的约束值的四个像素(阈值偏移)内(如例如由公式 (2) 或公式 (9) 预测的)。在另一实现方式中, d_{32} 的估计值必须在 d_{32} 的约束值的百分之五(缩放因子阈值)内(如例如由公式 (2) 或公式 (9) 预测的)。

[0089] 设 E_{32} 和 E_{34} 分别表示用于 d_{32} 和 d_{34} 的估计成本(失真)。例如,像素之间的 MSE 是典型的用于视差估计的估计成本。即,由于估计不是完美的,因此在基于所估计的视差呈现新视图时存在成本(损失)。当然,在保持本原理的精神的同时,还可以利用其它的成本函数。由于我们已经知道当 d_{32} 和 d_{34} 的估计中的至少一个不满足 (2) 时它或它们是不可靠的,如果 E_{32} 和 E_{34} 之一显著地大于另一个,则很有可能具有较大估计成本的视差估计未能找到良好匹配的像素。我们将在视差图中将该视差估计标记为孔洞(hole),并且仅保留具有较小估计成本的视差,然后进行到步骤 3。在各种实现方式中,通过确定 E_{32} 和 E_{34} 之间的差是否大于给定阈值来作出对 E_{32} 和 E_{34} 之一是否显著地大于另一个的确定。

[0090] 如果 E_{32} 和 E_{34} 不是显著地不同(例如, E_{32} 和 E_{34} 之间的差的绝对值不大于给定阈值),则对于我们而言确定哪一个更可靠变得含糊不清。我们提出,对于这种情形,返回到步骤 1 并利用更新后的搜索参数来执行视差估计。例如,可以增加搜索范围以便找到更好的匹配。还可以改变块大小。将再次检查公式 (2) 以及 E_{32} 和 E_{34} , 以便确定是否可以前进到步骤 3。应该监视迭代的数量以便避免无限的估计循环。如果已经执行了给定数量的迭代,但处理仍未能找到可靠的视差值,则将对于每个视差图(视图 2 和视图 4) 标记孔洞像素,并且该方法进行到步骤 3。

[0091] 注意,我们已经使用阈值来确定视差是否足够地满足约束。我们也已经使用阈值来确定失真是否显著不同。然而,其它实现方式在这些情形的一个或多个下为一对视差值分配置信度(confidence)。基于置信度的测量,可以采取多个动作中的任一个。一些实现方式向用户或呈现设备提供置信度的测量。

[0092] 步骤 3. 如果存在剩余的要处理的像素,则前进到下一像素并且转到步骤 1。否则,转到步骤 4。

[0093] 步骤 4. 孔洞填补:如果在视图 2 和 / 或视图 4 的视差图中存在孔洞像素,则在该步骤中填补孔洞像素。在一种实现方式中,可以基于相邻的视差像素,使用一些内插算法来填补孔洞。在其它实现方式中,可以仅仅使用相邻的距照相机更远(优选的是背景)的深度级(即,简单地选择更小的视差 / 更大的深度)来填补孔洞。

[0094] 还注意,以上构造视差(例如,用于视图 2) 的处理确定用于视图 2 中与视图 3 中的位置相对应的位置的视差。然而,对视图 3 中的位置采取循环。相应地,在视图 2 中可能存在视图 3 中没有对应位置的位置。视图 2 中的 这些位置被简单地在视图 2 的视差图中保持为孔洞。

[0095] 相反,视图 3 中的多个位置可能产生映射到视图 2 中的同一位置的视差。在此情

况下,较大的视差(较小的深度)是优选的,这是因为其指示前景对象。

[0096] 在图7中示出了实施例1的例程。图7示出了依据本原理实施例的用于估计深度的示例性方法700。方法700可以由例如图3的深度估计器300执行。注意,图7中的视图编号以及下面的文字仅仅用于说明/解释目的。总的构思可以被应用来联合估计两个视图的深度图。在步骤705,像素*i*被设置为视图3中的下一像素。在步骤710,在目标视图2上执行视差估计以便得到视差 d_{32} 。在步骤715,在目标视图4上执行视差估计以便得到视差 d_{34} 。在步骤720,确定 d_{32} 和 d_{34} 是否在给定阈值内满足公式(2)。如果是,则方法进行到步骤725。否则,方法进行到步骤730。在步骤725,按照公式(10)更新 d_{32} 和 d_{34} 。注意,该步骤725提供了用于视图2和4的深度图的对应视差值的一致性。在步骤730,确定估计成本 E_{32} 和 E_{34} 是否彼此接近(例如,在阈值量内)。如果是,则方法进行到步骤735。否则,方法进行到步骤760。在步骤735,更新视差搜索参数。在步骤740,确定迭代的数量是否大于阈值(迭代的数量)。如果是,则方法进行到步骤745。否则,该方法返回步骤710和715。在步骤745,标记两个视差图视图2和视图4中的孔洞像素。在步骤760,标记具有较大视差误差的视差图中的孔洞像素。在步骤750,确定在视图3中是否存在更多像素要被处理。如果是,则该方法返回步骤705。否则,该方法进行到步骤755。在步骤755,填充视差图中的孔洞像素(如果存在任何孔洞像素的话)。

[0097] 实施例2

[0098] 与实施例1不同,我们提出联合估计成本(失真)函数,使得两个视差(d_{32} 和 d_{34})搜索例程被组合到单个例程中。使用公式(2)中示出的关系来设计联合视差搜索处理。对于每个候选视差 d_{32} ,基于公式(2)具有对应视差 d_{34} ,由此可以计算来自 d_{32} 和 d_{34} 的联合估计成本函数。例如,联合估计成本函数可以简单地为 E_{32} 和 E_{34} 之和, $E_{\text{joint}} = E_{32} + E_{34}$,其中, d_{32} 和 d_{34} 满足公式(2)。可以使用高级度量(诸如例如,基于照相机距离的加权和)作为替代度量。评估不同的视差对,并且选择导致最小估计成本 E_{joint} 的视差对。以与实施例1相似的方式产生不同的视差对,其对于 d_{32} 在可能视差值的预定范围上循环。对于每个像素,该实施例中的所选视差将产生满足公式(2)的视差向量,形成了两个视差图中的一致性。

[0099] 在图8中示出了实施例2的例程。图8示出了依据本原理实施例的用于估计深度的另一示例性方法800。方法800可以由例如图3的深度估计器300执行。注意,图8中的视图编号以及下面的文字仅仅用于说明/解释目的。总的构思可以被应用来联合估计两个视图的深度图。在步骤805,像素*i*被设置为视图3中的下一像素。在步骤810, E_{min} 被设置为INT_MAX,其是最大可用整数。在步骤815,对于下一候选视差 d_{32} ,基于公式(2)计算 d_{34} 。在步骤820,基于 d_{32} 和 d_{34} 计算联合估计误差 E 。在步骤825,确定 E 是否小于 E_{min} 。如果是,则方法进行到步骤830。否则,方法进行到步骤835。在步骤830,将 $d_{32\text{best}}$ 设置为 d_{32} ,将 $d_{34\text{best}}$ 设置为 d_{34} ,并且将 E_{min} 设置为 E 。该方法进行到步骤835。在步骤835,确定是否存在任何更多候选视差要被评估。如果是,则该方法返回步骤815。否则,该方法进行到步骤840。在步骤840,确定在视图3中是否存在任何更多像素。如果是,则该方法返回步骤805。否则,该方法终止。

[0100] 在一些实现方式中实施例1和2被一起使用。例如,尽管实施例1和2可能在许多或所有情况下都同样工作良好,但在存在很少孔洞的情况下实施例2可能比实施例1更好地工作。相应地,一种实现方式在预计有孔洞的区域中使用实施例1,而在其它区域使用

实施例 2。

[0101] 应清楚,可以在编码器或解码器处执行各种实现方式。在一种实现方式中,在编码器(或例如处理器)处执行深度估计,而然后与或不与视频一起传送所估计的深度。在另一实现方式中,无任何深度地传送视频,而接收器执行深度估计。

[0102] 在传输视频之前不估计深度并且因而不传送深度的情况下,各种实现方式传送关于照相机参数的信息。在几种实现方式中,传送一个或多个约束,使得接收器知道在联合地估计多深度图时要使用的约束。在一种这样的实现方式中,在传输时使用标准化格式来编码并传送照相机数量、照相机之间的间距、 d_u 值、以及焦距。在另一实现方式中,由于假设焦距对于每个照相机相同以及 d_u 值对于每个照相机为零,因此该系统仅编码并传送照相机之间的间距来确定约束。

[0103] 我们注意到,可以执行用于多深度图的联合深度估计,其中,多深度图是来自不同视图、或来自相同视图。例如,多深度图可以在不同时刻来自相同视图。

[0104] 进一步,各种实现方式可以联合估计相对于来自给定视图的公共视频画面处于各种相对位置的深度图。例如,几种实现方式使用来自视图 3 的视频来联合估计用于视图 1 和 2 的深度图。在一种这样的实现方式中,视图 3 位于视图 1 和 2 之间。在另一种这样的实现方式中,定位视图 3 作为视图 1-3 的最左视图。而在又一种这样的实现方式中,定位视图 3 作为视图 1-3 的最右视图。

[0105] 另外,其它实现方式联合估计三个或更多深度图。在这样的实现方式中,确定提供所述三个或更多深度图之间的关系的一个或多个约束。例如,在上面的示例中视图 3 可以被用来确定不仅用于视图 2 和 4 的深度图、而且还有用于视图 5 的深度图。使用上面导出的公式,直接确定 d_{32} 、 d_{34} 和 d_{35} 之间的关系。

[0106] 在本申请中描述的几种实现方式和特征可能被用于 H. 264/MPEG-4AVC (AVC) 标准、或带有 MVC 扩展的 AVC 标准、或带有 SVC 扩展的 AVC 标准的背景下。然而,这些实现方式和特征可能被用于其它标准(现有的或将来的)的背景下、或者不涉及标准的背景下。

[0107] 我们因此提供一种或多种具有特定特征和方面的实现方式。然而,所描述的实现方式的特征和方面也可能被适配用于其它实现方式。

[0108] 多种实现方式可以使用各种技术(包括但不限于 SEI 消息、码片首标、其它高级语法、非高级语法、带外信息、数据流数据、以及隐式信令)来发信号通知信息。相应地,尽管可能在特定的背景下描述了这里描述的实现方式,但是这样的描述不应被当作将特征和构思限制于这样的实现方式或背景。

[0109] 在本说明书中对本原理的“一个实施例”或“实施例”或“一种实现方式”以及它们的其它变型的引用意味着结合所述实施例描述的特定特征、结构、特性等被包括在本原理的至少一个实施例中。因此,在说明书多个位置出现的短语“在一个实施例中”或“在实施例中”或者“在一种实现方式中”或“在实现方式中”以及所有其它变型的出现不一定都指代相同的实施例。

[0110] 应当认识到,例如在“A/B”、“A 和 / 或 B”以及“A 和 B 中的至少一个”的情况中对于以下的“/”、“和 / 或”和“中的至少一个”的任一个的使用旨在包括仅对于第一个列出的选项 (A) 的选择、或者仅对于第二个列出的选项 (B) 的选择、或者对于两个选项 (A 和 B) 的选择。作为另一示例,在“A、B 和 / 或 C”和“A、B 和 C 中的至少一个”的情况中,这种措

辞旨在包括仅对于第一个列出的选项 (A) 的选择、或者仅对于第二个列出的选项 (B) 的选择、或者仅对于第三个列出的选项 (C) 的选择、或者仅对于第一个和第二个列出的选项 (A 和 B) 的选择、或者仅对于第一个和第三个列出的选项 (A 和 C) 的选择、或者仅对于第二个和第三个列出的选项 (B 和 C) 的选择、或者对于全部三个选项 (A 和 B 和 C) 的选择。如本领域和相关领域普通技术人员容易看到的,这可以被扩展用于很多列出的项目。

[0111] 而且,如这里所使用的,词语“画面”和“图像”被可互换地使用,并且指代来自视频序列的静止图像或画面。如已知的,画面可以是一帧或一场。

[0112] 可以以例如方法或处理、装置、软件程序、数据流或信号来实现这里描述的实现方式。即使仅在单种形式的实现方式的背景下进行讨论(例如,仅仅作为方法讨论),所讨论的特征的实现方式仍可能以其它形式(例如,装置或程序)来实现。可以以例如适当硬件、软件、和固件来实现装置。可以在例如诸如例如处理器的装置(其总地指代包括例如计算机、微处理器、集成电路、或可编程逻辑器件的处理设备)中实现方法。处理器还包括通信设备,诸如例如计算机、蜂窝电话、便携式/个人数字助理(“PDA”)、和便利终端用户之间的信息通信的其它设备。

[0113] 可以在各种不同的装备或应用(具体地,例如与数据编码和解码相关联的装备或应用)中体现这里描述的各种处理和特征的实现方式。这样的装备的示例包括编码器、解码器、处理来自解码器的输出的后处理器、向编码器提供输入的预处理器、视频编码器、视频解码器、视频编解码器、网络服务器、机顶盒、膝上型计算机、个人计算机、蜂窝电话、PDA、以及其它通信设备。如应当清楚的,该装备可以是移动的并且甚至可以安装在移动的车辆中。

[0114] 另外,可以通过由处理器执行的指令来实现各种方法,并且这样的指令(和/或由实现方式产生的数据值)可以存储在处理器可读介质上,诸如例如,集成电路、软件载体或其它存储设备(诸如例如硬盘、紧致盘(compact diskette)、随机存取存储器(“RAM”)或只读存储器(“ROM”)。所述指令可以形成在处理器可读介质上有形体现的应用程序。指令可以例如处于硬件、固件、软件或组合中。指令可以在例如操作系统、分离的应用、或两者的组合中找到。因此,处理器可以被特征化为例如被配置为执行处理的设备、以及包括具有用于执行处理的指令的处理器可读介质的设备(诸如存储设备)两者。此外,处理器可读介质除了可以存储指令外或者替代可以存储指令,还可以存储由实现方式产生的数据值。

[0115] 如对于本领域技术人员将显而易见的,各种实现方式可以产生被格式化为可以被例如存储或传送的承载信息的各种信号。该信息可以包括例如用于执行方法的指令、或由所描述的实现方式之一产生的数据。例如,可以将信号格式化为承载所描述的实施例的用于写入或读取语法的规则作为数据、或者承载由所描述的实施例写入的实际语法值作为数据。这样的信号可以被格式化为例如电磁波(例如,使用频谱的射频部分)或者为基带信号。格式化可以包括例如编码数据流并利用编码后的数据流来调制载波。信号承载的信息可以是例如模拟或数字信息。如已知的,可以在各种不同的有线或无线链路上传送该信号。该信号可以被存储在处理器可读介质上。

[0116] 已经描述了多种实现方式。然而,将理解可以作出各种修改。例如,可以组合、补充、修改或移除不同实现方式中的元素,以便产生其它的实现方式。另外,本领域技术人员将理解,其它结构和处理可以替换所公开的那些结构和处理,并且作为结果的实现方式将

以与所公开的实现方式至少基本相同的方式执行与所公开的实现方式至少基本相同的功能,以便实现与所公开的实现方式至少基本相同的结果。相应地,这些和其它实现方式是由本申请所预期的,并且在所附权利要求的范围之内。

100

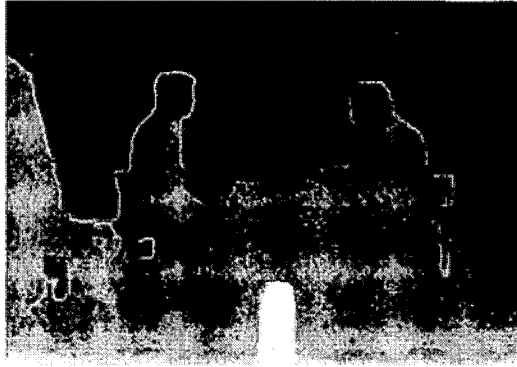


图 1

200

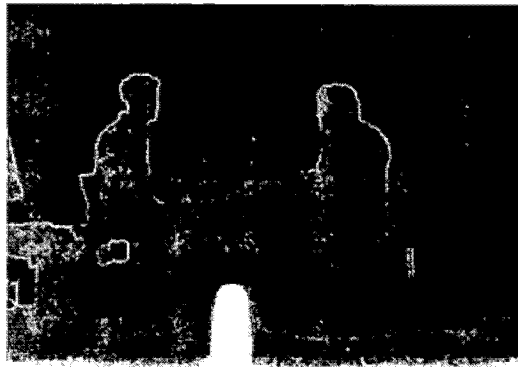


图 2

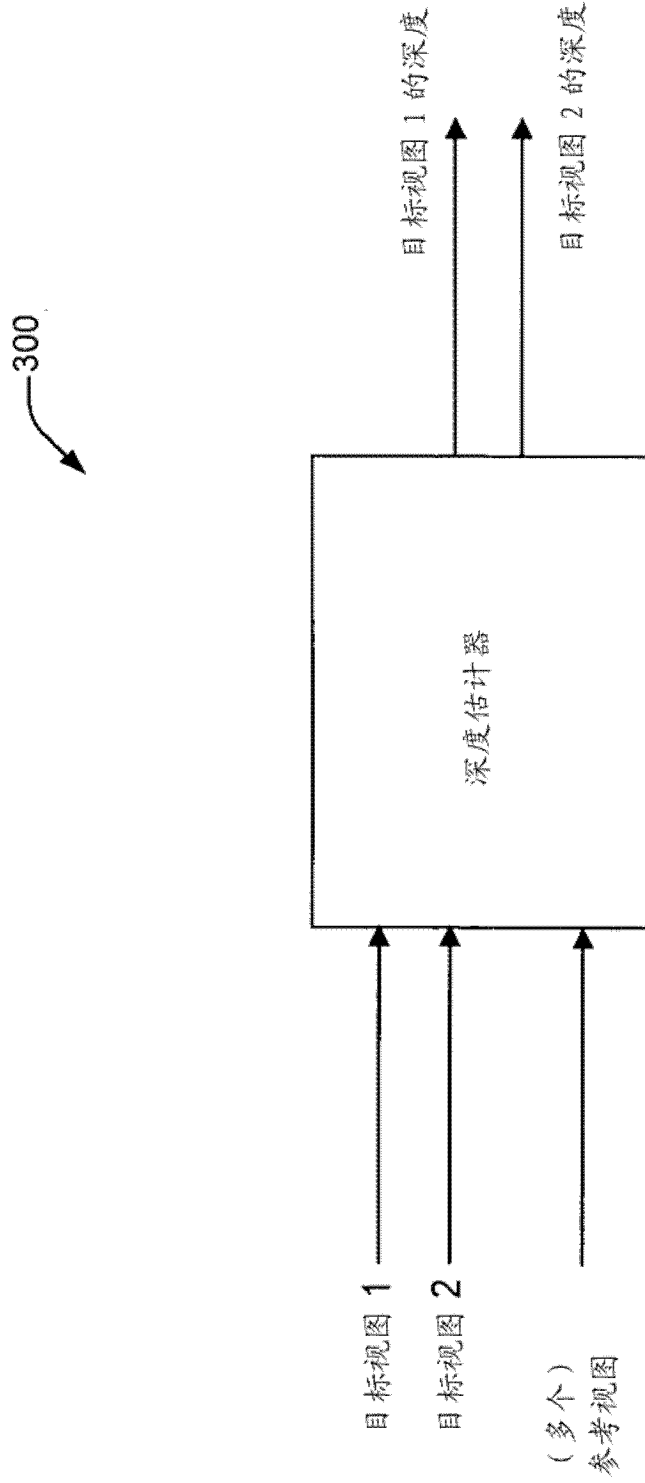


图 3

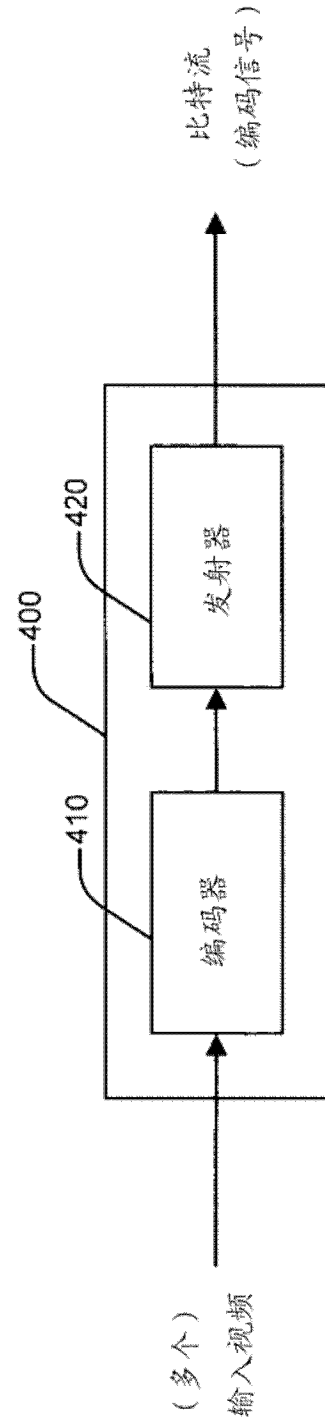


图 4

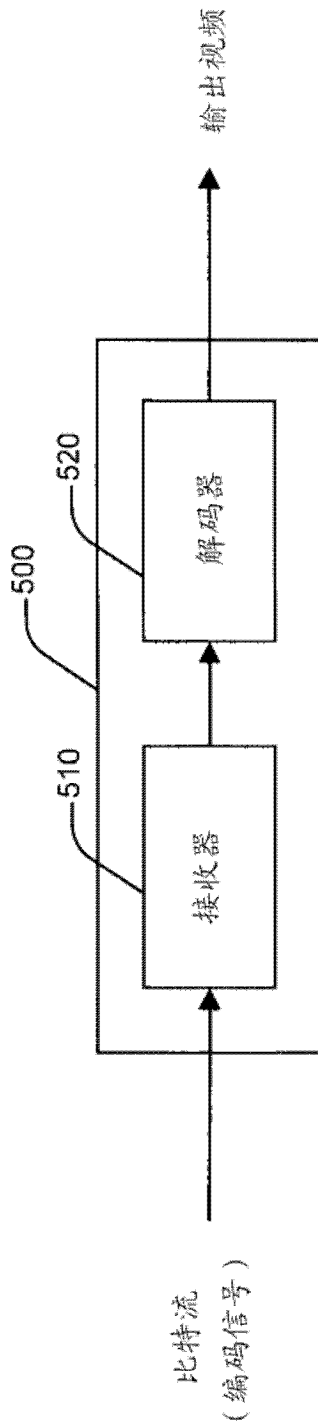


图 5

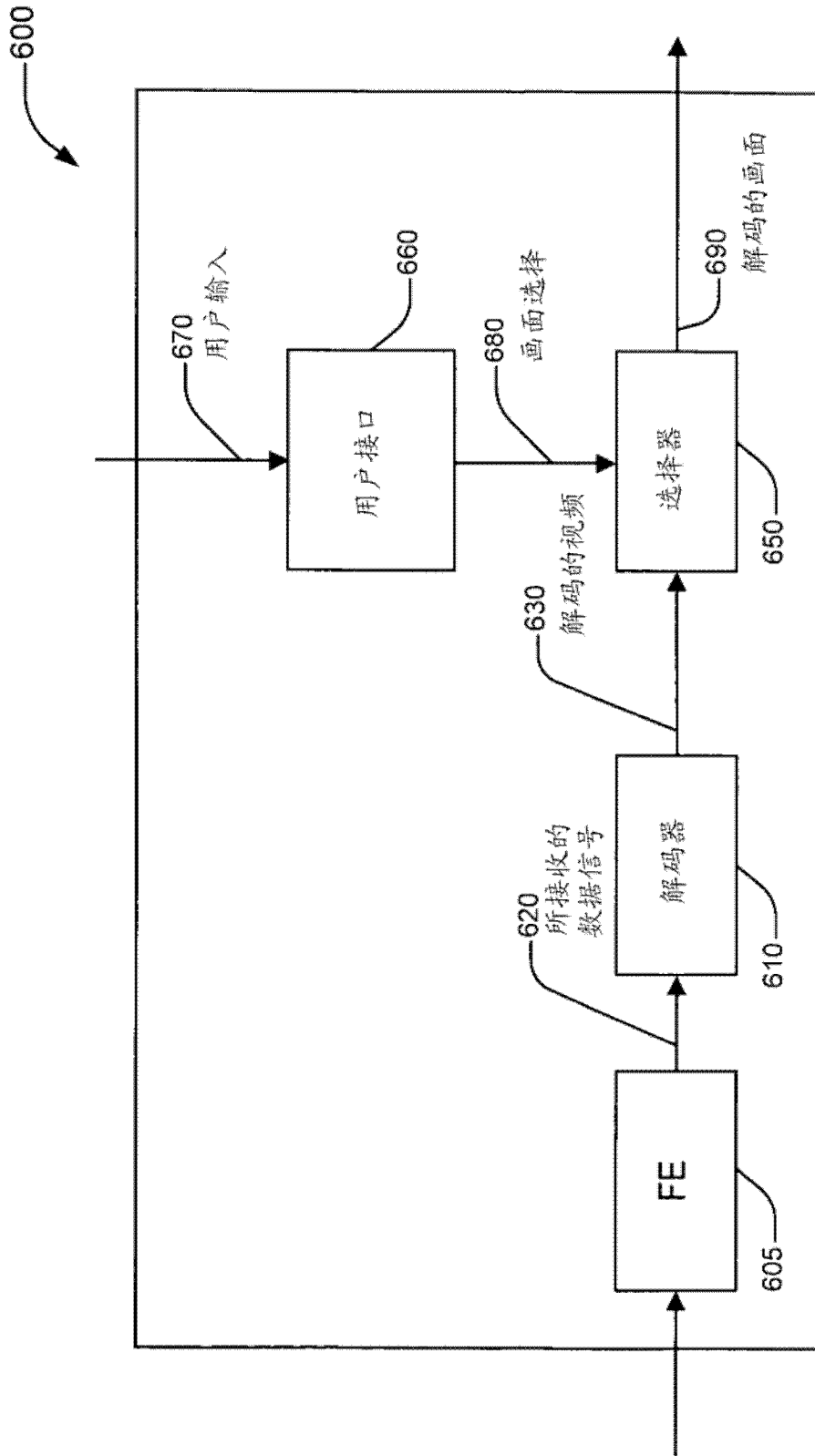


图 6

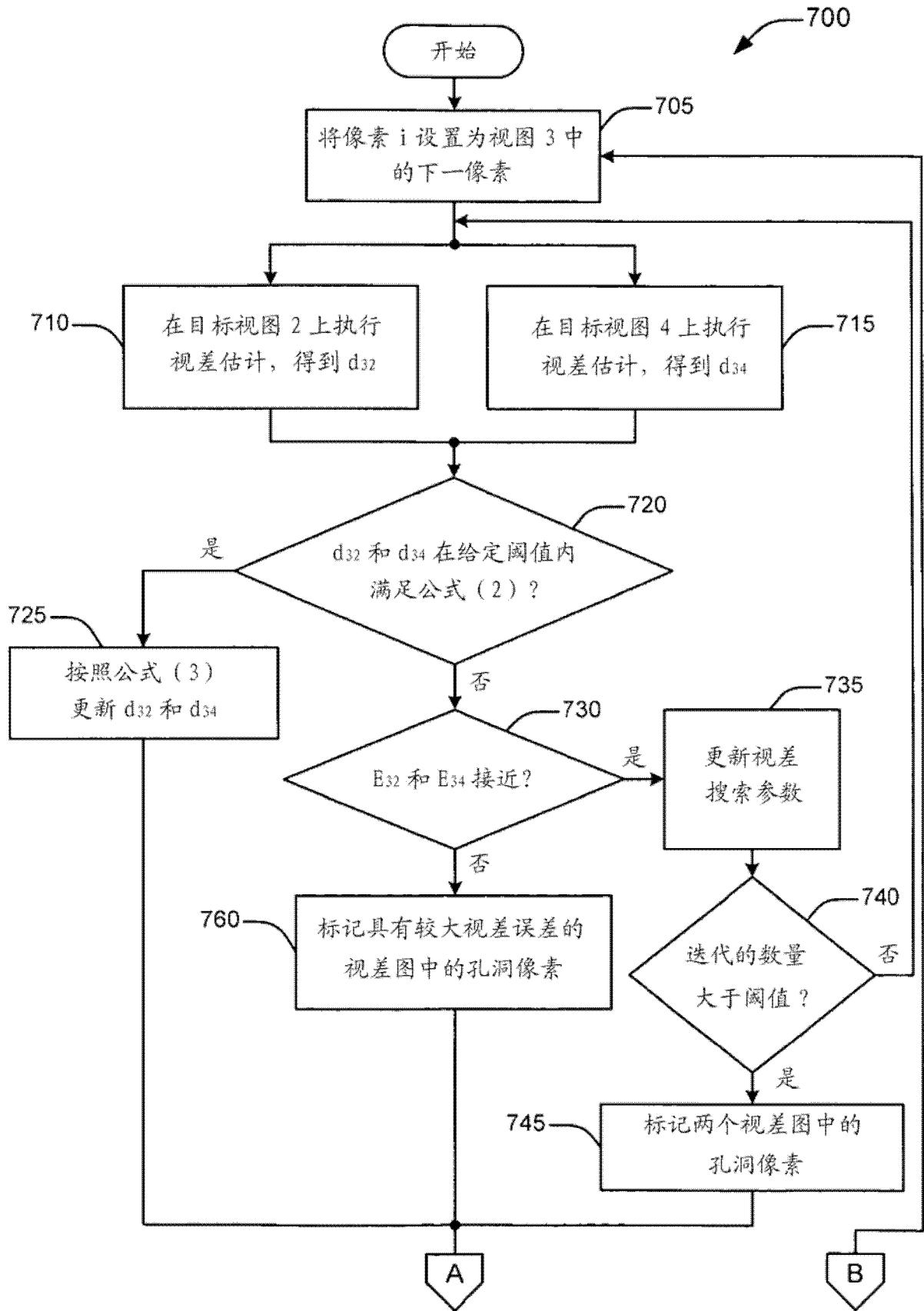


图 7A

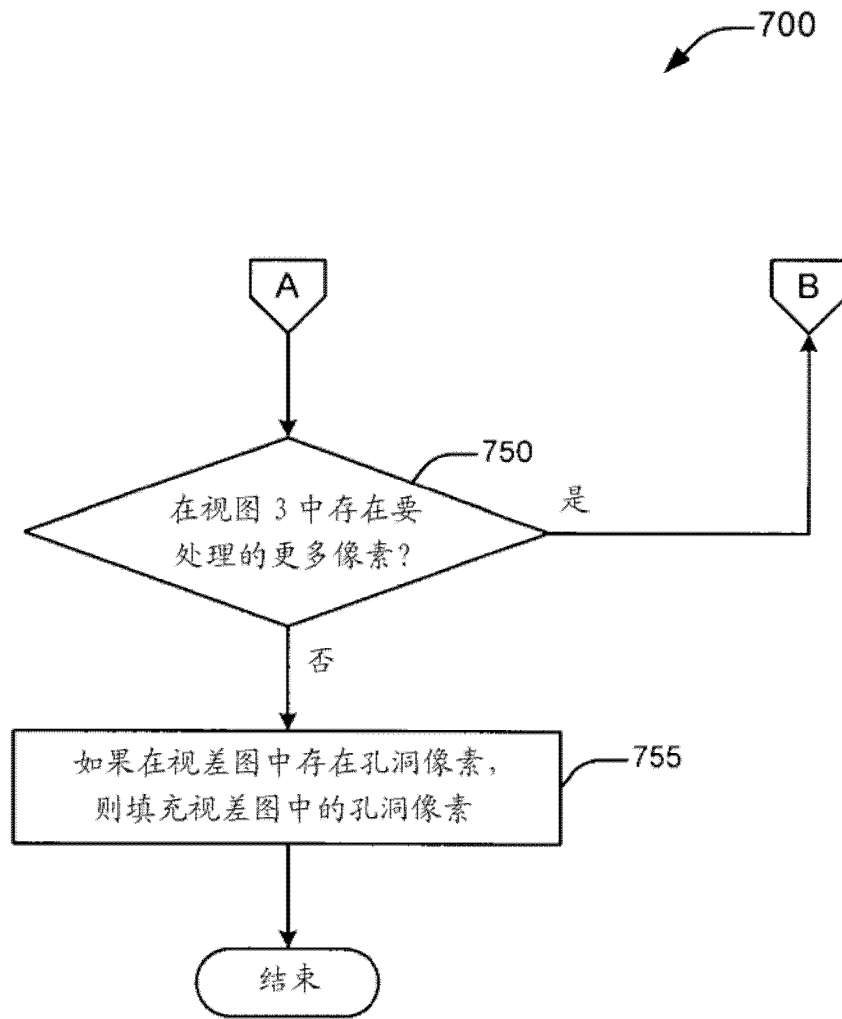


图 7B

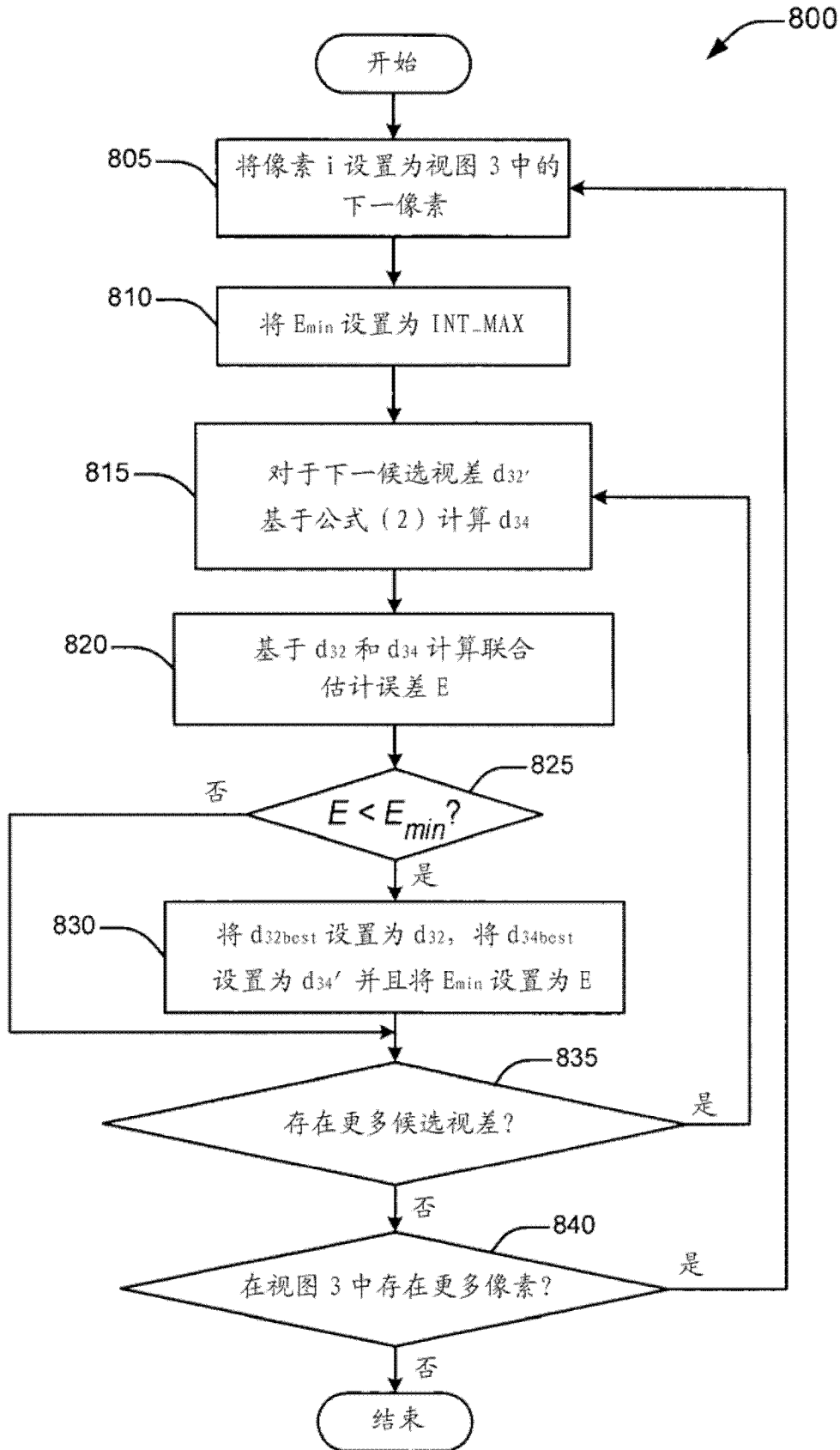


图 8