[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06T 1/00 (2006.01)

[12] 发明专利说明书

专利号 ZL 200510082091.9

[45] 授权公告日 2009 年 12 月 9 日

[11] 授权公告号 CN 100568272C

[22] 申请日 2005.6.28

[21] 申请号 200510082091.9

[30] 优先权

[32] 2004. 6.28 [33] US [31] 10/879,235

[73] 专利权人 微软公司

地址 美国华盛顿州

[72] 发明人 C·孜特尼克三世 M·尤特坦戴乐
R·斯泽利司基 S·维恩德
S·B·康

[56] 参考文献

WO2004/015997A1 2004.2.19

A bayesian approach to digital matting. Yung. Yu Chuang, Brian Curless, David H. Salesin, RichardSzeliski. Proceedings 2001 IEEE conference on computer vision and pattern recognition, Vol. 1 . 2001

Compression and transmission of depth maps for image – basedrendering. Ravi Krishnamurthy, Bing. Bing Chai, Hai Tao, SriramSethuraman. Proceedings 2001 international conference on image processing, Vol. 3 . 2001

审查员 孙 艳

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 张政权

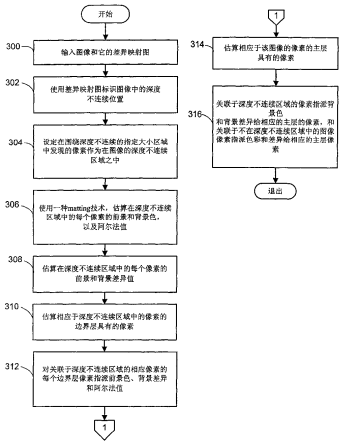
权利要求书 3 页 说明书 13 页 附图 4 页

[54] 发明名称

用于生成场景的两层、3D 表示的系统和过程

[57] 摘要

提出了一种用于从图像和图像的像素视差映射图生成数字或数字化图像的两层、3D 表示的系统和过程。 该两层表示包括一主层，它含有表现与图像中的深度不连续区域的对应位置像素相关联的背景色和背景视差的像素，以及表现与不在这些深度不连续区域中发现的图像的对应位置像素相关联的色彩和视差的像素。 另一层是边界层，它由表现与深度不连续区域的对应位置像素相关联的前景色、前景视差和阿尔法值的像素组成。 该深度不连续区域对应于围绕使用其视差映射图在图像中找到的深度不连续性的指定大小的区域。



1．一种用于从图像和图像的视差映射图生成数字或数字化图像的两层表示的方法，所述方法包括以下动作：

使用所述图像的视差映射图来标识所述图像中的深度不连续性的位置；

标识在围绕所述深度不连续性的指定大小区域中发现的图像的像素，并且将这些像素指定为在所述图像的深度不连续区域中；

为所述深度不连续区域中的每一像素估算前景和背景色以及阿尔法值；

为所述深度不连续区域中的每一像素估算前景和背景视差值；

估算图像的边界层，它包含在对应于所述深度不连续区域的像素的每个相应位置上的像素，其中，每个边界层像素被分配与所述深度不连续区域的对应像素相关联的前景色、前景视差和阿尔法值；以及

建立所述图像的主层，包括，

在对应于所述深度不连续区域的像素的每个相应位置上的像素，其中，每个所述像素被分配与所述深度不连续区域的对应像素相关联的背景色和背景视差值，以及

在不对应于所述深度不连续区域的像素的每个位置上的像素，其中，每个所述像素被分配与所述图像的对应像素相关联的色彩和视差值。

2．如权利要求 1 所述的方法，其特征在于，标识所述图像中的深度不连续性的位置的动作包括把表现相邻像素之间的视差值之差大于指定视差等级数的任何位置标识为深度不连续性的动作。

3．如权利要求 2 所述的方法，其特征在于，所述指定视差等级数是 4。

4．如权利要求 1 所述的方法，其特征在于，围绕所述深度不连续性的指定大小区域被定义为从一被标识的深度不连续性位置在每个方向上扩展 3 个像素的区域。

5．如权利要求 1 所述的方法，其特征在于，为所述深度不连续区域中的每一像素估算前景色和背景色以及阿尔法值的动作包括使用一修边技术来估算所述色彩和阿尔法值的动作。

6．如权利要求 1 所述的方法，其特征在于，为所述深度不连续区域中的每一像素估算前景和背景视差值的动作包括使用所述图像的前景和背景部分中附近视

差的阿尔法加权平均值来估算所述前景和背景视差值的动作。

7．如权利要求 6 所述的方法，其特征在于，使用所述图像的前景和背景部分中附近视差的阿尔法加权平均值来估算所述前景和背景视差值的动作包括以下动作：

通过将所述图像的前景部分中相邻于所考虑的像素的指定大小窗口中每个像素的视差值分别乘以其阿尔法值，并且对所得乘积求平均值，来为每一深度不连续区域中的每一像素计算前景视差值；以及

通过将所述图像的背景部分中相邻于所考虑的像素的指定大小窗口中每个像素的视差值分别乘以 1 减去其阿尔法值后的值，并且对所得的乘积求平均值，来为每一深度不连续区域中的每一像素计算背景视差值。

8．如权利要求 1 所述的方法，其特征在于，还包括将对应于所述边界层像素的区域扩张一指定量，并为每一添加的像素分配与所述主层中的对应像素相同的色彩和视差值以及阿尔法值 1 的动作。

9．如权利要求 8 所述的方法，其特征在于，所述指定的扩张量是一个像素。

10．一种用于从图像和图像的像素深度映射图生成数字或数字化图像的两层表示的系统，包括：

用于使用所述图像的深度映射图来标识所述图像中的深度不连续性的位置的装置；

用于标识在围绕所述深度不连续性的指定大小区域中发现的图像的像素，并且将这些像素指定为在所述图像的深度不连续区域中的装置；

用于为所述深度不连续区域中的每一像素估算前景和背景色以及阿尔法值的装置；

用于为所述深度不连续区域中的每一像素估算前景和背景深度值的装置；

用于估算图像的边界层的装置，所述边界层包含在对应于所述深度不连续区域的像素的每个相应位置上的像素，其中，每个边界层像素被分配与所述深度不连续区域的对应像素相关联的前景色、前景视差和阿尔法值；以及

用于建立所述图像的主层的装置，所述主层包括：

在对应于所述深度不连续区域的像素的每个相应位置上的像素，其中，每个所述像素被分配与所述深度不连续区域的对应像素相关联的背景色和背景视差值，以及

在不对应于所述深度不连续区域的像素的每个位置上的像素，其中，每

个所述像素被分配与所述图像的对应像素相关联的色彩和视差值。

11．如权利要求 10 所述的系统，其特征在于，用于标识所述图像中的深度不连续性的位置的装置包括用于把表现相邻像素之间的深度值之差大于指定值的任何位置标识为深度不连续性的装置。

用于生成场景的两层、3D 表示的系统和过程

技术领域

本发明涉及数字或数字化图像的分层表示，尤其涉及用于生成场景的两层、3D 表示的系统和过程。

背景技术

近几年来，电视商业广告和电影长片的观众已经看到用于创建停顿时间和改变照像机视点的幻觉的“冻结帧”效果。最早的商业广告通过使用基于影片的系统来产生，该系统在沿着轨道排列的不同静态照相机之间快速地跳跃以给出穿越冻结时间片移动的幻觉。

当它首次出现时，其效果是新鲜的，且看上去是很壮观的，并且很快在很多产品中被模拟，其中最有名的看来是在名为“The Matrix”的电影中看到的“子弹时间”效果。不幸的是，这种效果是一次性的、预先-计划的事件。视点轨道是提早安排的，而且花费了很多工时来产生所要求的内插场景。较新的系统是基于摄影机阵列，但是仍然依赖于具有很多摄像机以避免软件场景内插。

这样，现有系统不允许用户在观看基于动态图像的场景时交互地改变到任何所期望的视点。在过去，在基于图像的重现（IBR）上的大部分工作涉及重现静态场景，采用了两种著名的技术：光场重现（Light Field Rendering）[11]和光照图（Lumigraph）[7]。它们在高质量再现方面的成功起源于大量采样图像的使用，并且激发了本领域一大群工作。这个奠基工作的一种激动人心的潜在扩展涉及在观看视频时交互地控制视点。用户交互地控制视频视点的能力明显地增强了视觉感受，允许诸如诸如新视点即时重播、改变戏曲中的视点、以及随意地创建“冻结帧”视觉效果等多种应用。

然而，由于同步如此多的摄像机以及采集和存储图像的困难（和成本），将 IBR 扩展到动态场景并不是无足轻重的。不仅仅在从多个视点捕捉、表示和再现动态场景中存在重大的障碍要克服，而且为了能够交互地做此事提供一种相当进一步的复杂性。至今，实现这个目标的努力还不是非常令人满意。

关于交互式视点视频系统的基于视频的再现方面，较一种早期的捕捉和再现动态场景的努力之一是 Kanade 等人的 Virtualized Reality（可视化逼真）系统[10]，它包括围绕一个 5 米网络圆顶排列的 51 个摄像机。每个摄像机的分辨率是 512×512，且捕捉速率 30fps。它们采用基于场景流公式[17]的三维像素着色[14]形式，在每个时间帧提取一个球状表面表示。不幸的是，因为低分辨率、匹配误差和对象边界的不正确处理，使结果看来并不切合实际。

Carranza 等人[3]使用了围绕一个房间分布的 7 个同步摄像机，它们面朝该房间的中心以捕捉 3D 人类运动。每个摄像机为 CIF 分辨率（320×240），且以 15fps 进行捕捉。它们使用一个 3D 人类模型作为在每个时间帧计算 3D 形状的先验。

Yang 等人[18]设计了一个 8×8 的摄像机（每个 320×240）网格，用于捕捉动态场景。它们不再存储和再现数据，而是仅仅发送组成所期望的虚拟视图所必需的光线。在它们的系统中，摄像机没有被同步锁相；相反，它们依赖于跨越 6 个 PC 的内部时钟。摄像机捕捉速率是 15fps，并且交互观看速率是 18fps。

上述系统中常见的是需要大量的图像用于逼真再现，这部分地是因为场景几何结构是未知的或仅仅大约知道。如果几何结构被准确地知道，就可能充分地降低对图像的要求[7]。一种提取场景几何结构的实用方法是通过立体系统，并且为静态场景提出了许多立体算法[13]。然而，对于采用带动态场景的立体技术已经作出了少量的努力。作为 Virtualized Reality 工作[10]的一部分，Vedula 等人[17]提出了一种使用 2D 光流和 3D 场景形状来提取 3D 运动（即，场景形状之间跨越时间的对应性）的算法。在他们的方法中，他们使用一种类似于三维像素着色[14]的投票方案，其中使用的度量是假设的三维像素位置适合该 3D 流等式的良好程度。

Zhang 和 kambhamettu[19]还集成了 3D 场景流和其框架中的结构。其 3D 仿射运动模型被局部地使用，具有空间正则化，并且采用色彩分段以保持不连续性。Tao 等人[16]假设场景是分段平面的。他们还假设每个平面面片的恒定速率，以便约束动态深度映射估算。

在一个更加雄心勃勃的努力中，Carcerroni 和 Kutulakos[2]恢复具有已知光照位置的非刚性运动下的分段连续的几何结构和反射率（Phong 模型）。他们使该空间离散成表面元素（“面元（surfels）”），并且对位置、方向和反射率参数执行搜索，以最大化地与观察到的图像的一致。

在一种对传统的局部窗口匹配的有趣的改变中，Zhang 等人[20]使用跨越空间和时间的匹配窗口。这种方法的优点是随时间变化对亮度恒定性具有较少的依赖

性。

活动测距技术也被应用于移动场景。Hall-Holt 和 Rusinkiewicz[8]使用随时间变化的投影的边缘编码的条纹图案。市场上还有一种以色列 3DVSystems 公司制造的称为 ZCam™的商业系统，它是一种结合广播摄影机使用的范围检测摄影机附加装置。但是，它是一种昂贵的系统，并且只提供单一视点深度，使它较不适用于多视点视频。

然而，不管立体和基于图像的再现方面的所有进步，要再现动态场景的高质量、高分辨率视图仍然是非常困难的。如同在 Light Field Rendering（光场再现）论文[11]中所建议的一种方法是仅仅基于输入和虚拟摄相机的相对位置而简单地对光线进行重新采样。然而，如同在光照图（Lumigraph）[7]和后续工作中所演示的，对场景几何结构使用 3D 顶替器（impostor）或代理能够极大地改进内插视图的质量。另一种方法是创建单个纹理映射的 3D 模型[10]，但是这通常产生使用多个参考视图的较差结果。还有另一种方法采用了需要 3D 代理的几何结构辅助的基于图像的再现方法。一种可能性是使用单个球状多面体模型，如同在 Lumigraph 和 Unstructured Lumigraph（未结构化光照图）论文[1]中所述的。另一种可能性是使用每像素深度，如同在分层深度图像（Layered Depth Images）中[15]、在立面（Façade）中的偏移深度映射[5]、或者带深度的子画面[15]。一般而言，对每个参考视图使用不同的局部几何结构代理[12, 6, 9]将产生高质量的结果。

然而，即使是多深度映射图仍然在生成新视图时展现再现的人为因素，即由于前景到背景转移的突然特性而引起的图形失真（锯齿状），以及由于混合像素而引起的污染色彩，当在新背景或对象上合成时，它们变得可见。

这个问题在本发明中通过一种独特的输入图像的两层、3D 表示来解决。注意，该两层、3D 表示不仅仅能够用来解决以上关于交互式视点视频系统中再现新视图的图形失真问题，而且也能够同样有利地用于其它环境中。通常，任何数字或数字化图像能够使用这种两层、3D 表示来表示。

注意，在前面的段落中，以及在本说明书的其余部分中，本描述引用包含在一对方括号中的数字标志符标识的各种独立的出版物。例如，这样的引用可以通过叙述“参考文献[1]”或者更简单地“[1]”来标识。多个参考文献将通过包含一个以上标志符的一对方括号来标识，例如[2，3]。在具体实施方式章节结尾处能够找到包含对应于每个标志符的出版物的参考文献清单。

发明内容

本发明针对一种用于生成数字或数字化图像的两层表示的系统和过程。一般而言，该两层包括一主层，它具有展示背景色和与图像中的深度不连续区域的对应位置像素相关联的背景视差的像素，以及展示色彩和与在这些深度不连续区域中未发现图像的对应位置像素相关联的的视差相关联的像素。另一层是边界层，它由展示前景色、前景视差和与深度不连续区域的对应位置像素相关联的阿尔法值的像素构成。该深度不连续区域对应于围绕在图像中发现的深度不连续性的指定大小的区域。

该两层表示是通过首先使用图像的视差映射来标识所考虑的图像中的深度不连续性的位置来生成的。深度不连续性出现在相邻像素之间的视差值的差大于指定等级数的位置上。然后，围绕标识在围绕深度不连续性的指定大小的区域内发现的图像的像素。下一步，使用一种修边（matting）技术对深度不连续区域中的每个像素估算前景和背景色以及前景阿尔法值。另外，使用图像的前景和背景部分中的邻近视差的阿尔法加权平均，为深度不连续区域中的每个像素估算前景和背景视差值。然后，建立图像的边界层，它包括对应于深度不连续区域的像素的每一位置上一个像素。然后，向每一边界层像素分配前景色、与深度不连续区域的对应像素相关联的前景视差和阿尔法值。另外，建立图像的主层。该主层包括对应于深度不连续区域的像素的每一位置上的一个像素，以及在不对应于深度不连续区域的像素的图像的每一像素位置上的一个像素。与深度不连续区域的对应像素相关联的背景色和背景视差值被分配给主层中对应位置像素的每一个，而与不在深度不连续区域的图像像素相关联的色彩和视差值被分配给主层中对应位置像素的每一个。注意，一旦对深度不连续区域的每一像素建立了前景色、前景视差和阿尔法值，这些区域的大小能够使用传统的扩张技术用一个指定的量来生长，以防止在从层中再现图像期间出现破裂。

除了上文描述的益处之外，当结合附图阅读以下详细描述时，本发明的其它优点将变得显而易见。

附图说明

当参考以下描述、所附权利要求书以及附图时，可以更好地理解本发明的具体特征、方面和优点，附图中：

图 1 是描述构成用于实现本发明的示例性系统的通用计算设备的图示。

图 2 是对照像素位置绘制像素行的视差值的曲线图，其中视差值的突变台阶表示了一个深度不连续性。

图 3A 和 3B 是图示了用于根据本发明生成数字或数字化图像的两层表示的过程的流程图。

图 4(a)-(e)是显示在一组劈裂舞演员的图像上应用图 3A-B 的两层图像表示生成过程的结果的图像。图 4(a)显示主层色彩估算，图 4(b)表示主层视差估算。图 4(c)显示边界层色彩估算，图 4(d)表示边界层视差估算。图 4(e)表示边界层阿尔法值估算。注意，图 4(c)-(e)的图像是求反显示的，从而透明/空像素看上去是白色。

具体实施方式

在以下本发明较佳实施例的描述中，参照了附图，附托形成本发明的一部分，并且在其中作为说明示出了可在其中实施本发明的具体实施例。要理解，可使用其它实施例，并且可以作出结构变化，而不脱离本发明的范围。

1.0 计算环境

在提供本发明的较佳实施例的描述之前，将描述其中能实现本发明的适用的计算环境的简要概括描述。图 1 示出了适用的计算系统环境 100 的例子。计算系统环境 100 仅仅是适用的计算环境的一个例子，并且不打算暗示对本发明的使用范围或功能的任何限制。也不应将计算环境 100 解释成相对于示例性操作环境 100 中示出的组件的任一个或其组合具有任何依赖或要求。

本发明可以用各种其它通用或专用计算系统环境或配置来运行。适用于本发明使用的公知的计算系统、环境和/或配置的例子包括但不限于：个人计算机、服务器计算机、手持或膝上型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费电子设备、网络 PC、小型机、大型机、包括任何以上系统或设备的任一个的分布式计算环境等等。

本发明可以在诸如由计算机执行的程序模块等计算机可执行指令的通用上下文中描述。一般而言，程序模块包括完成特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。本发明还可以在分布式计算环境中实践，其中任务由通过通信网络链接的远程处理设备来完成。在分布式计算环境中，程序模块可以位于本地或远程计算机存储介质中，包括存储器存储设备。

参照图 1，用于实现本发明的示例性系统包括计算机 110 形式的通用计算设

备。计算机 110 的组件可包括但不限于：处理单元 120、系统存储器 130 和将包括系统存储器的各种系统组件耦合到处理单元 120 的系统总线 121。系统总线 121 可以是若干种总线结构的任何一种，包括存储器总线或存储器控制器、外围总线、和使用多种总线体系结构的任何一种的局部总线。作为例子，而非限制，这种体系结构包括工业标准结构（ISA）总线、微通道结构（MCA）、增强型 ISA（EISA）总线、视频电子技术标准协会（VESA）局部总线、以及外围部件互连（PCI）总线（也称为 Mezzanine 总线）。

计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是可以由计算机 110 访问的任何可用介质，包括易失性和非易失性介质、可移动和不可移动介质。作为例子，但非限制，计算机可读介质可包含计算机储存介质或通信介质。计算机储存介质包括以任何方法和技术实现来存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的易失性和非易失性、可移动和不可移动介质。计算机储存介质包括，但不限于：RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘（DVD）或其它光盘存储、磁带盒、磁带、磁盘储存或其它磁储存设备，或者能够用来存储所要求的信息并能够由计算机 110 访问的任何其它介质。通信介质通常在载波或其它传输机制等已调制数据信号中具体化计算机可读指令、数据结构、程序模块或其它数据，并且包括任何信息递送介质。术语“已调制数据信号”指其一个或多个特征以在信号中编码信息的方式而设置或改变的信号。作为例子，但非限制，通信介质包括有线介质，如有线网络或直接线路连接，以及无线介质，如声学、RF、红外和其它无线介质。以上各种组合也应该被包括在计算机可读介质的范围之内。

系统存储器 130 包括易失性和/或非易失性存储器形式的计算机储存介质，例如只读存储器（ROM）131 和随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）包含如在启动时帮助在计算机 110 中的元件之间传输信息的基本例程，通常储存在 ROM 131 中。RAM 132 通常包含处理单元 120 可直接访问和/或当前正在操作的数据和/或程序模块。作为例子，但非限制，图 1 示出操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 还可包括其它可移动/不可移动、易失性/非易失性计算机储存介质。仅仅作为例子，图 1 示出读取或写入不可移动、非易失性磁介质的硬盘驱动器 141、读取或写入可移动、非易失性磁盘 152 的磁盘驱动器 151、以及读取或写入可移动、非易失性光盘 156，例如 CD-ROM 或其它光介质的光盘驱动器 155。可用于示例性

操作环境中的其它可移动/不可移动、易失性/非易失性计算机储存介质包括但不限于：磁带盒、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 通常通过不可移动存储器接口（如接口 140）连接到系统总线 121，而磁盘驱动器 151 和光盘驱动器 155 通常由可移动存储器接口（例如接口 150）连接到系统总线 121。

以上讨论并且在图 1 中示出的驱动器及其相关联的计算机储存介质为计算机 110 提供了计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在图 1 中，硬盘驱动器 141 被示出为储存操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意，这些组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同或不同。操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 在此被给以不同的标号以表示至少它们是不同的副本。用户可以通过输入设备，如键盘 162 和定点设备 161（通常指鼠标、跟踪球或触摸垫）输入命令和信息到计算机 110。其它输入设备（未示出）可包括话筒、操纵杆、游戏垫、圆盘式卫星天线、扫描仪等等。这些和其它输入设备经常通过耦合到系统总线 121 的用户输入接口 160 连接到处理单元 120，但是也可以由其它接口和总线结构，如并行端口、游戏端口或通用串行总线（USB）连接。监视器 191 或其它类型显示设备也通过接口，如视频接口 190 连接到系统总线 121。除了监视器以外，计算机还可包括其它外围输入设备，如扬声器 197 和打印机 196，它们可以通过输出外围接口 195 连接。能够捕捉图像序列 193 的摄像机 192（如数字/电子静态或视频摄像机，或者胶卷/照片扫描仪）也能够作为个人计算机 110 的输入设备被包括在内。此外，尽管仅仅描述了一台摄像机，然而也可包括多台摄像机，作为个人计算机 110 的输入设备。来自一台或多台摄像机的图像 193 通过适当的摄像机接口 194 输入到计算机 110。该接口 194 连接到系统总线 121，因此允许图像被路由到并储存在 RAM 132 中，或者与计算机 110 相关联的其它数据储存设备之一中。然而，要注意，图像数据也能够从上述任一计算机可读介质输入到计算机 110，而不要求使用摄像机 192。

计算机 110 可以使用到一个或多个远程计算机（如远程计算机 180）的逻辑连接在网络环境中操作。远程计算机 180 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其它普通网络结点，并且通常包括许多或所有关于计算机 110 所描述的元件，尽管在图 1 中仅仅示出了存储器储存设备 181。在图 1 中描述的逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但是还可以包括其它网络。

这样的网络环境普遍存在于办公室、企业范围计算机网络、内联网和因特网中。

当在 LAN 网络环境中使用时，计算机 110 通过网络接口或适配器 170 连接到 LAN 171。当在 WAN 网络环境中使用时，计算机 110 通常包括调制解调器 172 或通过 WAN 173（例如因特网）建立通信的其它装置。调制解调器 172 可以是内置或者外置的，它可以通过用户输入接口 160 或者其它适当的机制连接到系统总线 121。在网络环境中，相对于计算机 110 所描述的程序模块或其部分可以存储在远程存储器储存设备中。作为例子，但非限制，图 1 示出远程应用程序 185 驻留在存储器设备 181 上。将会明白，示出的网络连接是示例性的，并且可以使用在计算机之间建立通信链路的其它手段。

2.0 两层图像表示

现在已经讨论了示例性操作环境，本描述章节的其余部分将专门致力于对实施本发明的程序模块的描述。一般而言，本发明涉及生成图像的唯一两层、3D 表示，它便于其压缩、传输和储存。该表示在图像是动态场景的视频帧并且帧数据正被编码以进行实时再现时特别有用。它还包括像素视差或者深度信息，由此提供了该表示的 3D 方面。图像或帧是数字图像，它或者由数字摄像机捕捉，或者如果不是，则在进一步处理之前被数字化。还要注意，数字图像数据在它是通过使用摄像机捕捉场景的实际图像而获得的这一点上能够是基于图像的，或者是合成的图像数据。

两层表示是通过首先定位所考虑的图像或帧的视差映射图中的深度不连续来生成的。这些深度不连续被定义为大于指定视差等级数（例如，在本发明的测试实施例中为 4 级）的跳跃。所考虑的图像的视差映射图能够以任何传统的方式来获得。然而，本发明的测试实施例采用一种新方法，它是本申请的发明人的题目为“Color Segmentation-Based Stereo Reconstruction System And Process（基于色彩分段的立体重建系统和过程）”的共同提交的待决申请的主题，并且被转让给同一受让人。该共同提交的待决申请提交于 2004 年 6 月 28 日并被分配序列号 10/879,327。

下一步，在所考虑的图像中发现的深度不连续性的附近标识小区域。这些小区域被定义为包括该深度不连续性的位置的 3 个像素内的所有像素。这在图 2 中示出，其中像素行的视差值对照像素位置来绘制。视差值中的突变台阶表示一个深度不连续性，假设它大于指定的视差等级数。上述小区域被称作为深度不连续区域，它是围绕该不连续性建立的。在图 2 的图中表示的概况中，该区域具有跨越图像中

深度不一致性位置的宽度。如果该深度不连续性跟随在图像中一个对象的轮廓之后（往往是典型的情况），则该深度不连续区域将合并以形成该轮廓之后的条纹。

沿着对象边界的某些像素将接收来自前景和背景区域的影响。然而，如果在再现期间使用原始的混合像素色彩，则导致可见的人为因素。由此，重要的是分离这两种影响。因此，下一步建立深度不连续区域中的每个像素的前景和背景色，如同是像素的不透明性（以阿尔法值的形式）。这是使用一种修边技术来实现的。通常，修边技术涉及通过估算起源于每个像素的前景和背景元素的色彩和不透明性，来提取图像的前景元素和背景元素。像素的不透明性由范围在 0 到 1 之间的阿尔法值定义。事实上，该阿尔法值定义了像素色彩可归因于前景元素的百分比。一种典型的修边操作的最终结果是为每个所考虑的像素标识前景色、背景色和阿尔法值。虽然任何修边过程能够用于本发明，然而测试的实施例采用在参考文献[4]中描述的贝叶斯图像修边技术。要注意，参考文献[4]还包含许多其它现有修边技术的描述。

如上所述，本发明的一种主要应用涉及从与两个实际场景图像相关联的视点中间的视点再现场景的虚拟图像。虽然这是如何实现的细节并不在本发明的范围之内，然而要注意，该过程通常涉及知道与每个像素相关联的深度。为此，根据本发明的图像表示包括像素深度（或者视差值，它能容易地被转换成深度值）。虽然这些值对于来自前述的视差映射图的大多数像素是可用的，然而要注意，存在与被发现为与深度不连续性相邻的混合像素相关联的两种深度，即与前景元素相关联的深度和与背景相关联的深度。这样，在下一步，对深度不连续区域中的每一个像素估算前景和背景的深度（或视差）值。通常，这是分别通过使用图像的前景和背景部分中的附近深度的阿尔法加权平均值来实现的。更具体地，前景视差通过使用来自深度不连续区域的原始前景区域内的像素的视差值的窗口（例如 7×7 像素）内的阿尔法加权平均值来找到。在图 2 中，原始前景区域指视差不连续性左方的像素。背景视差通过深度不连续区域的原始背景区域内的视差值的窗口内的加权平均值来找到。用于对背景视差求平均值的权重被设置为等于 1 减去阿尔法值。

一旦建立了深度不连续区域的像素色彩、视差和不透明性，就形成了该图像表示的两个层。更具体地，与在深度不连续区域中发现的每个相应像素相关联的先前计算的前景色、前景视差和阿尔法值被分配给该图像的边界层的对应位置像素。类似地，与在深度不连续区域中的每个相应像素相关联的先前计算的背景色和背景视差值，连同图像中不在深度不连续区域中的所有像素的色彩和视差值一起（从图

像和其视差映射图中取得），被分配给该图像的主层的对应位置像素。

因此，所得的两层图像表示包括一边界层，它为该层中的每一个像素标识前景色、前景视差和阿尔法值。另外，表示包含一主层，它为该层中与深度不连续区域相关联的像素标识背景色和背景深度，并且为该层的每个其它像素标识色彩和视差值。这样，能够看到，主层将包括关于图像中每个像素的像素信息，然而该数据将在该场景的一个图像与下一个图像之间相当类似。如果图像是视频帧，并且空间上如果要传输或存储同一场景的多个图像，则两种场合在时间上都是真实的。这就造成使用标准图像压缩方法能够高度压缩主层。另外，边界层将包含相对较少的数据，通常仅仅该图像中的全部数量的像素一个小的百分比将包含在该层中。这样，即使当随着时间变化拍摄时，该数据可能在动态场景的一个图像与下一个图像之间发生显著的改变，然而并没有如此多的数据需要传输或存储。因此，即使没有压缩，该数据也能够被容易地传输和存储。使用恰当的压缩方案，能够进一步减少数据量和能够传输它的速度。按这种方式，上述两层图像表示提供了数据传输和存储的所要求的简易化，并使得实时再现变得切实可行。

现在，将参考图 3A-B 中所示的流程图来略述上述两层图像表示生成过程。首先，输入将被表示的数字或数字化的图像及其视差映射图（处理动作 300）。然后使用视差映射图在图像中标识深度不连续性位置（处理动作 302）。下一步，标识在围绕该深度不连续性的指定大小区域中发现的像素，并且将其指定为是在该图像的深度不连续区域中（处理动作 304）。使用修边技术，下一步建立深度不连续区域中的每个像素的前景和背景色，及其阿尔法值（处理动作 306）。另外，在处理动作 308，对深度不连续区域中的每一像素估算前景和背景视差值。然后建立边界层，它具有对应于深度不连续区域像素的像素（处理动作 310）。与深度不连续区域中发现的每个相应像素相关联的前景色、前景视差以及阿尔法值被分配给边界层的对应像素（处理动作 312）。也建立主层，它具有对应于该图像的所有像素的像素（处理动作 314）。在此情况下，与深度不连续区域中的像素相关联的背景色和背景视差被分配给主层的对应像素，并且与不在深度不连续区域中的图像像素相关联的色彩和视差被分配给该主层的对应像素（处理动作 316）。

能够被结合到用于根据本发明的生成两层图像表示的上述过程一种改进涉及到在形成该图像表示的边界层和主层之前，扩张深度不连续区域的前景衬边（matte）。这具有防止在从层再现图像期间出现由于处理中的不准确度而发生的破裂的优点。更具体地，一旦为深度不连续区域的每一像素建立了前景色、前景视

差和阿尔法值，这些区域的大小被增长指定量（例如 1 个像素）。在扩张期间添加到边界层的新像素被分配与主层中的对应像素相同的色彩和视差值，以及阿尔法值 1。

图 4(a)-(e)示出了在一组霹雳舞演员的图像上应用根据本发明的两层图像表示生成过程的结果。图 4(a)示出了主层色彩估算，图 4(b)示出了主层视差估算。类似地，图 4(c)示出了边界层色彩估算，图 4(d)示出了边界层视差估算。最后，图 4(e)示出了边界层的阿尔法估算。注意，图 4(c)-(e)的图像被求反显示，从而透明/空像素看上去为白色。注意仅仅少量信息是如何需要被发送以解决软对象边缘的，以及边界层的不透明性和两层中的色彩是如何被干净地恢复的。

注意，贯穿前面的描述，使用了图像像素的视差值。然而，在生成过程的任一点上，这些值可以使用标准方法被转换成深度值。在此情况下，主层和边界层将包括深度信息，而不是包含在其中的像素的视差值。

3.0 参考文献

[1]Buehler,C.、Bosse,M.、McMillan,L.、Gortler,S.J.和 Cohen,M.F.，2001，Unstructured Lumigraph Rendering，*Proceeding of SIGGRAPH 2001*（8 月），425-432。

[2]Carceroni,R.L.和 Kutulakos,K.N.，2001，Multi-view scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape and reflectance，*Eigth International Conference on Computer Vision（ICCV 2001）*，第 II 卷，60-67。

[3]Carranza,J.、Theobalt,C.、Magnor,M.A.和 Seidel,H.-P，2003，*Free-viewpoint video of human actors，ACM Transactions on Graphics 22,3*（7 月），569-577。

[4]Chuang,Y.-Y.等人，2001，Bayesian Approach to digital matting，*Conference on Computer Vision and Pattern Recognition（CVPR'2001）*，第 II 卷，264-271。

[5]Debevec,P.E、Taylor,C.J.和 Malik,J.，1996，Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach，*Computer Graphics(SIGGRAPH'96)*（8 月），11-20。

[6]Debevec,P.E.、Yu,Y.和 Borshukov,G.D.，1998，Efficient view-dependent image-based rendering with projective texture-mapping，*Eurographics Rendering Workshop 1998*，105-116。

[7]Gortler,S.J.、Grzeszczuk,R.、Szeliski,R.和 Cohen,M.F.，1996，The Lumigraph. *Computer Graphics (SIGGRAPH'96) Pro-ceedings*，ACM SIGGRAPH，43–54。

[8]Hall-Holt,O.、和 Rusinkiewicz,S.，2001，Stripe boundary codes for real-time structured-light range scanning of moving objects，*Eighth International Conference on Computer Vision (ICCV 2001)*，第 II 卷，359–366。

[9]Heigl,B.等人，1999，Plenoptic modeling and rendering from image sequences taken by hand-held camera，*DAGM'99*，94–101。

[10]Kanade,T.、Rander,P.W.、和 Narayanan,P.J.，1997，Virtualized reality: constructing virtual worlds from real scenes，*IEEE Mul-tiMedia Magazine* 1, 1 (1 月—3 月)，34–47。

[11]Levoy,M.、和 Hanrahan, P.，1996，Light field rendering，*In Computer Graphics (SIGGRAPH'96) Proceedings*，ACM SIG-GRAPH，31–42。

[12]Pulli,K.等人，1997，View-based rendering: Visualizing real objects from scanned range and color data，*In Proceedings of the 8th Eurographics Workshop on Rendering*。

[13]Scharstein,D.和 Szeliski,R.，2002，A taxonomy and evaluation of dense two-frame stereo correspondence algorithms，*International Journal of Computer Vision* 47, 1 (5 月)，7–42。

[14]Seitz,S.M.和 Dyer,C.M.，1997，Photorealistic scene reconstruction by voxel coloring，*In Conference on Computer Vision and Pattern Recognition (CVPR'97)*，1067–1073。

[15]Shade,J.、Gortler,S.、He,L.-W.和 Szeliski,R.，1998，Layered depth images，*Computer Graphics (SIGGRAPH'98) Proceedings*，ACM SIGGRAPH，Orlando，231–242。

[16]Tao,H.、Sawhney,H.和 Kumar,R.，2001，A global matching framework for stereo computation，*In Eighth International Conference on Computer Vision (ICCV 2001)*，第 I 卷，532–539。

[17]Vedula,S.、Baker,S.、Seitz,S.和 Kanade,T.，2000，Shape and motion carving in 6D，*Conference on Computer Vision and Pattern Recognition (CVPR'2000)*，第 II 卷，592–598。

[18]Yang,J.C.、Everett,M.、Buehler,C.和 McMillan,L.，2002，A real-time distributed light field camera，*Eurographics Workshop on Rendering*，P.Debevec 和 S.Gibson 编辑，77–85。

[19]Zhang,Y.、和 Kambhamettu,C.，2001，On 3D scene flow and structure estimation，*Conference on Computer Vision and Pattern Recognition (CVPR'2001)*，第 II 卷，778–785。

[20]Zhang,L.、Curless,B.和 Seitz,S.M.，2003，Spacetime stereo: Shape recovery for dynamic scenes，*Conference on Computer Vision and Pattern Recognition*，367–374。

[21]Zhang,Z.，2000，A flexible new technique for camera calibration，*IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 11, 1330–1334。

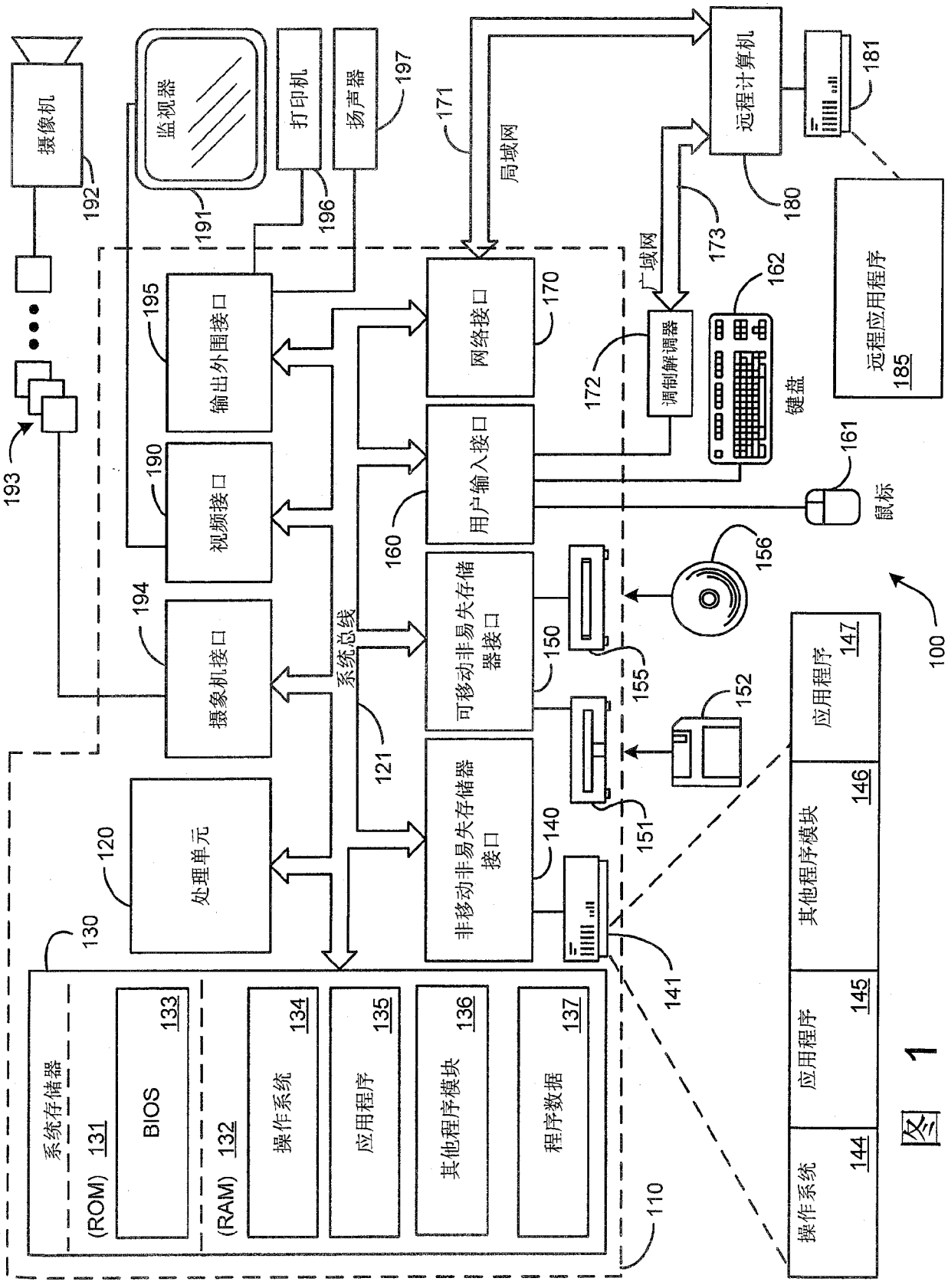

摄像机
192
监视器
191
打印机
196
扬声器
197
171
局域网
远程计算机
181
180
173
广域网
162
键盘
161
鼠标
远程应用程序
185
193
195
输出外围接口
190
视频接口
194
摄象机接口
网络接口
170
172
调制解调器
用户输入接口
160
156
系统总线
121
可移动非易失存储器接口
150
155
152
非移动非易失存储器接口
140
151
141
处理单元
120
130
系统存储器
(ROM) 131
BIOS 133
(RAM) 132
操作系统 134
应用程序 135
其他程序模块 136
程序数据 137
110
操作系统 144
应用程序 145
其他程序模块 146
应用程序 147
100

图 1

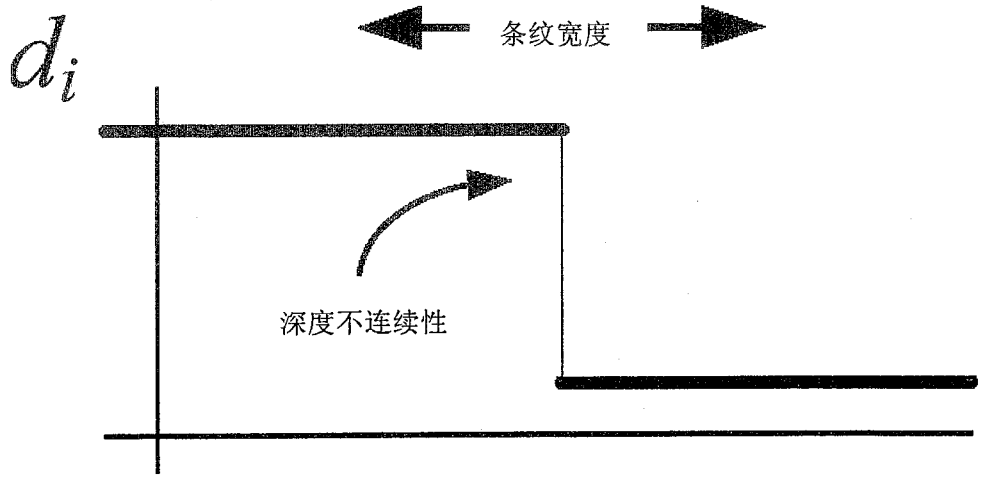


图 2

图 4(a)

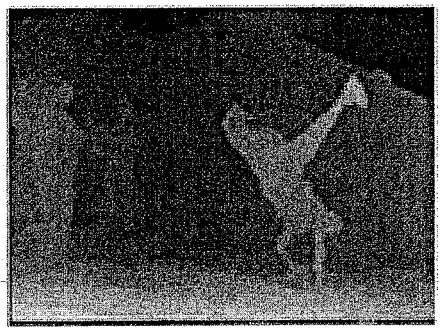


图 4(b)

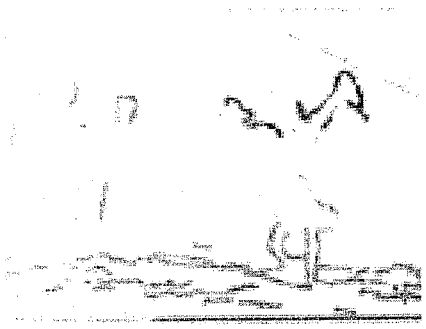


图 4(c)

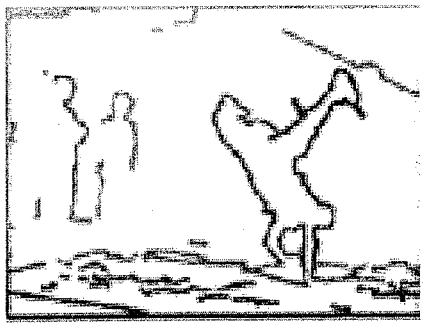


图 4(d)

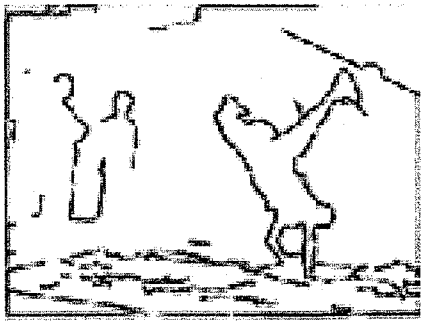


图 4(e)

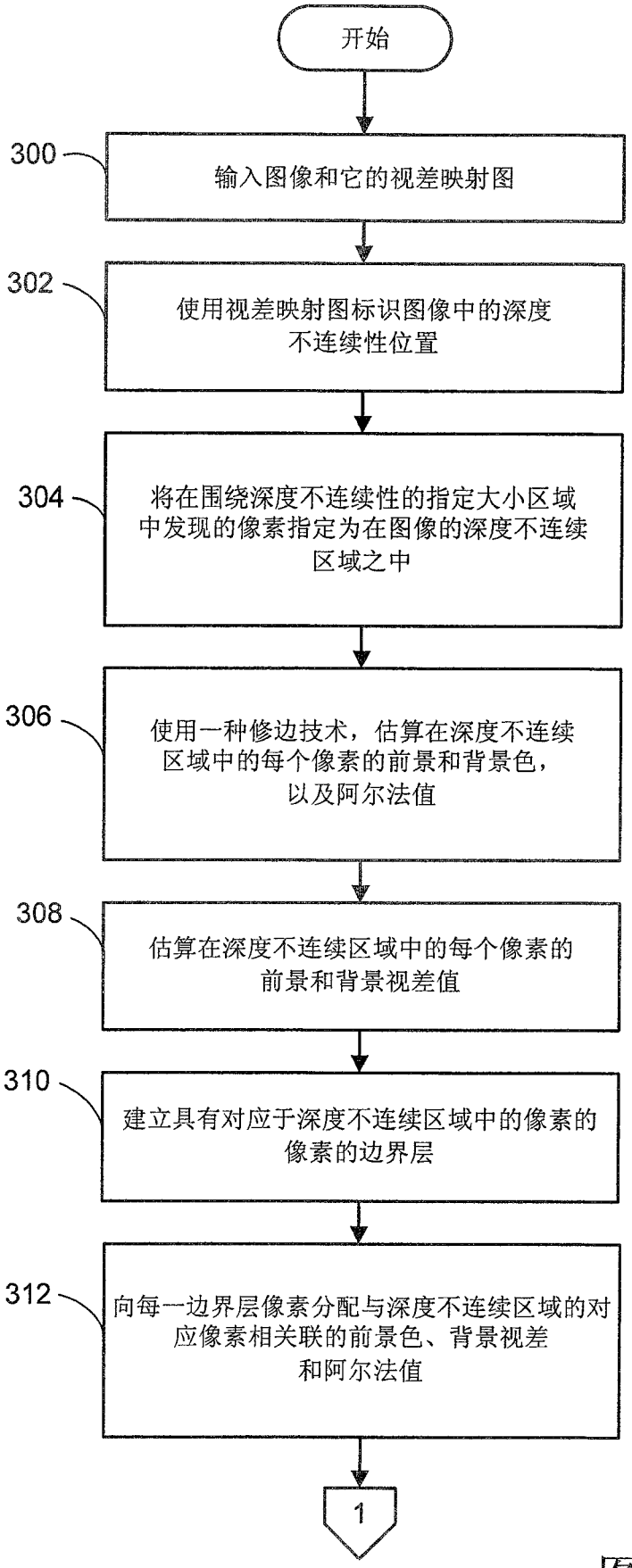

开始
300
输入图像和它的视差映射图
302
使用视差映射图标识图像中的深度不连续性位置
304
将在围绕深度不连续性的指定大小区域中发现的像素指定为在图像的深度不连续区域之中
306
使用一种修边技术，估算在深度不连续区域中的每个像素的前景和背景色，以及阿尔法值
308
估算在深度不连续区域中的每个像素的前景和背景视差值
310
建立具有对应于深度不连续区域中的像素的像素的边界层
312
向每一边界层像素分配与深度不连续区域的对应像素相关联的前景色、背景视差和阿尔法值
1

图 3A

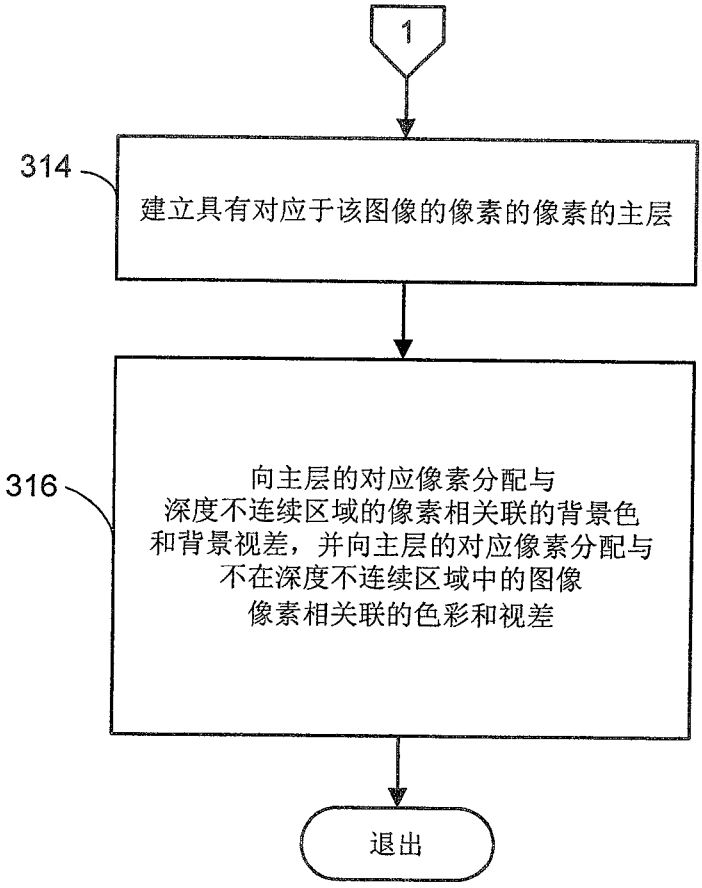

1
314
建立具有对应于该图像的像素的像素的主层
316
向主层的对应像素分配与
深度不连续区域的像素相关联的背景色
和背景视差，并向主层的对应像素分配与
不在深度不连续区域中的图像
像素相关联的色彩和视差
退出

图 3B