(54)发明名称
用于图像融合的方法和设备

(57)摘要
用于融合图像的方法和装置。描述了用于通过融合包含在二级图像(2)中的更丰富的信息来改善主要图像(1)的方法和装置(30)。检索(10)包含在二级图像(2)中的对象的3D结构，并且使用该3D结构生成二级图像(2)的差值修正版本。为此，确定相机位姿，对于该相机位姿，被包含在二级图像(2)中的对象的3D结构的投影最类似于主要图像(1)中的景物，并且，基于所确定的相机位姿来合成二级图像(2)的差值修正版本。然后将二级图像(2)的差值修正版本与主要图像(1)进行融合。
1. 一种用于将二级图像 (2) 与主要图像 (1) 进行融合的方法，所述主要图像是场景的全景视图，并且所述二级图像是所述场景的详细视图，该方法包括：

- 通过使用可用于所述二级图像 (2) 的信息，来检索 (10) 包含在所述二级图像 (2) 中的对象的 3D 结构；
- 通过确定相机位姿并且基于所确定的相位位姿来合成所述二级图像 (2) 的视差修正版；
- 使用所述的3D结构生成所述二级图像 (2) 的视差修正版本，其中，对于所述相机位姿，被包含在所述二级图像 (2) 中的对像的 3D 结构的投影最类似于所述主要图像 (1) 中的景物；以及
- 将所述二级图像 (2) 的视差修正版本与所述主要图像 (1) 进行融合。

2. 根据权利要求 1 所述的方法，其中，从为所述二级图像 (2) 提供的深度信息或光场描述中检索 (10) 所述二级图像 (2) 的3D 结构。

3. 根据权利要求 1 或 2 所述的方法，其中，通过对所述二级图像 (2) 执行照片逼真变形来合成所述二级图像 (2) 的视差修正版本。

4. 根据权利要求 1 或 2 所述的方法，进一步包括：对所述主要图像 (1) 与所述二级图像 (2) 的融合的视差修正版本之间的边界区域执行内插。

5. 根据权利要求 1 或 2 所述的方法，其中，所述主要图像 (1) 是全景图像，所述二级图像 (2) 是所述全景图像中的对象的更加详细的图像。

6. 根据权利要求 1 或 2 所述的方法，其中，所述主要图像 (1) 和所述二级图像 (2) 是一组图像中的两个图像。

7. 一种用于将二级图像 (2) 与主要图像 (1) 进行融合的装置 (20)，所述主要图像是场景的全景视图，并且所述二级图像是所述场景的详细视图，所述装置 (20) 包括至少一个处理器，所述至少一个处理器被配置为：

- 3D 结构检索级 (21)，所述3D结构检索级 (21) 被配置为使用可用于所述二级图像 (2) 的信息，来检索 (10) 包含在所述二级图像 (2) 中的对象的 3D 结构；
- 视差修正级 (22)，所述视差修正级 (22) 被配置为通过确定相机位姿并且基于所确定的相位位姿来进行所述二级图像 (2) 的视差修正版；
- 使用所述3D结构生成所述二级图像 (2) 的视差修正版本，其中，对于所述相机位姿，被包含在所述二级图像 (2) 中的对像的 3D 结构的投影最类似于所述主要图像 (1) 中的景物；以及
- 图像融合级 (23)，所述图像融合级 (23) 被配置为将已所述二级图像 (2) 的视差修正版本与所述主要图像 (1) 进行融合。

8. 根据权利要求 7 所述的装置，其中，所述3D 结构检索级 (21) 被配置为通过为所述二级图像 (2) 提供的深度信息或光场描述中检索 (10) 所述二级图像 (2) 的3D 结构。

9. 根据权利要求 7 或 8 所述的装置，其中，所述视差修正级 (22) 被配置为通过对所述二级图像 (2) 执行照片逼真变形来合成所述二级图像 (2) 的视差修正版本。

10. 根据权利要求 7 或 8 所述的装置，其中，所述图像融合级 (23) 被配置为对所述主要图像 (1) 与所述二级图像 (2) 的融合的视差修正版本之间的边界区域执行内插。

11. 根据权利要求 7 或 8 所述的装置，其中，所述主要图像 (1) 是全景图像，所述二级图像 (2) 是所述全景图像中的对象的更加详细的图像。

12. 根据权利要求 7 或 8 所述的装置，其中，所述主要图像 (1) 和所述二级图像 (2) 是一组图像中的两个图像。
用于图像融合的方法和设备

[0001] 本发明涉及用于融合图像的方法和设备。具体地，本发明涉及通过利用在单独的视图中可获得的感兴趣的对象的更丰富的信息来无缝改善这些对象在大的全景图像中的视觉质量和细节等级的方法和装置。

[0002] 有时希望仔细看看包含某些感兴趣的对象的特定区域。在这样的情形中，当这些对象离相机较远时，全景图像提供的分辨率通常是不足的。针对该问题的直接解决方案将是增加分辨率并且缩短全景捕捉设置的曝光时间。然而，该解决方案将需要巨大的传输带宽或存储容量，这可能使得该解决方案在实践中无法使用。而且，在当前的图像传感器的能力上存在技术限制。

[0003] 可替代地，当单独的相机跟踪感兴趣的物体并捕捉具有更高的细节等级的感兴趣的对象时，该补充图像能够与全景图像进行融合，以便在该全景图中提供感兴趣的对象的增强版本。相关技术能够被分类为多图像融合或混合。例如，文件CA 2386347描述了用于将低分辨率视频流和高分辨率视频流准确地组合为对于人类观察者而言的单个无缝显示的处理。全景视觉传感器捕捉全景图，而进一步的传感器捕捉详细的图像。两个传感器均被安排为固定系统。为了减小低分辨率视频和高分辨率视频之间的边界的干扰，在这两个视频之间的环形区域内将这两个视频平滑地混合在一起。不同分辨率图像的融合只能很好地适用于离传感器为中间距离。在其他深度处会发生图像失配（misregistration）。

[0004] 通常，尝试融合来自任意视点的信息一般会导致不正确插入附加的高质量视觉数据。该问题在图1和图2中被示出。在该示例中，详细视图2已被插入到（全景的）图像1中以匹配显示机的位置。由此可见，所插入的矩形的顶部和底部的景物偏移（perspective misalignment）导致了显著的伪影。

[0005] 为了解决该问题，US 2003/0235341公开了一种用于接合两个或多个图像的方法。通过使用平面扫描算法来计算重叠图像中的像素的相对深度。该深度用于图像校正。然后结果图像被接合。该文件中所描述的解决方案只适用于如下所述的情形的有限子集：所有的图像（即，详细视图和全景捕捉视图）都至少大概对齐，以使矫正实际上有用。对于融合图像，不同差距等级（即，深度等级）的像素的条纹仅仅被混合，这可能导致图像失真。此外，需要逆向变形阶段，这可能是进一步图像失真的根源。

[0006] 因此，本发明的目的在于为无缝融合图像提出改善的解决方案。

[0007] 本发明，一种用于将二级图像与主要图像进行融合的方法包括以下步骤：

[0008] - 检索包含在该二级图像中的对象的3D结构；

[0009] - 通过确定相机位姿并且基于所确定的相机位姿来合成二级图像的视差修正版本，来使用3D结构生成二级图像的视差修正版本，其中，对于该相机位姿，被包含在二级图像中的对象的3D结构的投影最类似于主要图像中的景物（）；

[0010] - 将该二级图像的视差修正版本与主要图像进行融合。

[0011] 本发明的一个目标情景是伴随有多个独立的高清相机的静态全景图像捕捉设置。这些独立的高清相机是立体相机，或者最为期待的是全光相机或光场相机。这样的相机允许在某一界限内最少引入误差的情况下任意更改景物视差（perspective parallax），其
中，景物视差由相机的物理实现来限定。通过将这些相机放置于离主要景物设置合理的距离处，感兴趣的对对象的特征可以被补为类似于全景捕捉的场景视差（scene parallax）。通过对3D结构的明确利用，将感兴趣的对象的详细视图与全景图像进行无缝融合是可能的。当然，全景图像没必要是静态图像。全景图像同样可以包含运动对象。与此的一个示例是运动竞技场，在运动竞技场中，运动员相对于本质上是静止的背景而言是运动的。

【0012】对上述情景的应用按照如下来工作。用户交互地选择全景图中用于显示的区域，该区域中包含一个或多个感兴趣的对象。这些对对象被自动或交互地从背景中分割。然后，第二图像（包含感兴趣的对象的高细节捕捉）被视差修正以适合于全景图像（panoramic perspective）。最后，全景图中感兴趣的对象通过变换（旋转、平移、缩放）的高细节版本所代替，并且边界被合适地混合在该全景图的背景中提供平滑的整合。

【0013】主流立体相机和光场相机的日益传播以及即将到来的3D内容的广泛使用使得存在本发明的进一步的目标情景。例如，从数据库（例如，互联网）收集到的多组立体图像可以被接合以构造高质量的全景图。另外，可以使用来自包含上述对象的不同图像的具有更好地视觉质量的类似内容来对图像中感兴趣的内容进行改善，而不考虑这些图像之间的视差差别。

【0014】为了更好的理解，现将参照附图在以下描述中更详细地解释本发明。应当理解，本发明不限于所述示例性实施例，并且在不背离所附权利要求定义的本发明的范围的情况下也可以方便地组合或修改特定特征。在这些附图中：

【0015】图1示出从全景图像中选择出的原始的模糊区域，
【0016】图2示出在插入来自详细视图的更丰富数据之后的图1的区域，
【0017】图3示出根据本发明的用于融合图像的方法，
【0018】图4示例性地描绘出实现图3的方法的装置，
【0019】图5更详细地示出图4的装置的视差修正级，以及
【0020】图6描绘出使用图4的装置所获得的平滑融合的图像。
【0021】图1示出从全景图像中选择出的原始的稍微模糊的区域1。在图2中，更详细的图像2已被插入至区域1中。由此可见，所插入的详细图像2的顶部和底部的景物偏移导致了显著的伪影。

【0022】图3示出根据本发明的用于融合图像的方法，该方法避免或至少减弱了这样的伪影。在第一步中，使用可用于详细视图的信息来检索该详细视图的3D结构（10）。例如，使用一对立体相机提供的深度通道、光场相机或光场相机提供的光场描述等来检索3D结构。然后，基于由观察者在大的全景视图中交互选择出的区域以及从详细视图中提取到的3D结构，产生详细视图的视差修正版本（11），该视差修正版本与从全景的视点所看到的景物相匹配。于是，来自视差修正视图的数据能够被用来将来自详细视图的更丰富数据无缝地整合到全景图中（12）。

【0023】在图4中示例性地描述了实现上述解决方案的装置20。在3D结构检索级12，从详细的辅助视图中检索场景的3D结构。对于立体设置，可以通过任意现有的立体重建技术来完成。在由光场相机所捕获的数据的情形中，优选地使用更精确的过程（例如，N.Snavely等描述的来自运动的结构（Structure from Motion）: "Modeling the World from Internet

[0024] 视差修正级22被提供给用于生成视差修正的详细视图。视差修正级22包括两个子级,如图5中所示。第一子级,位姿估计器30,获得相机位姿(即,相机位置和方向),对于该相机位姿,详细视图的3D结构的投影类似于全景视图中的景物。然后视图合成器37用第一子级30中获得的相机位姿对详细视图执行照片逼真变形。

[0025] 位姿估计器30优选地实现像RANSAC(RANdom SAMple Consensus,随机采样一致性)的鲁棒技术,即,通过检查变换的详细视图和全景视图之间的对齐点的投影误差来完成位姿估计。用于最优化的位姿模型由10个参数组成:焦距(1:f),光学中心偏移(2:c_x和c_y),旋转(4:参见下文)以及平移(3:t_x,t_y和t_z)。有些假设包括简化位姿模型。假设全景图像不存在偏斜误差,即,偏斜系数等于零。而且,假设垂直焦距和平行焦距相等,并且镜头失真可以忽略不计,这对于由于大的变焦设置而焦距较大的感兴趣的情形而言是合理的。

[0026] 结果位姿模型能够借助于标定矩阵K进行描述,K为:

\[
P = K \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}
\]

[0027] 其中,R为旋转矩阵,T为平移量向量。旋转矩阵R能够通过使用罗德里格斯(Rodrigues)的旋转公式由四个参数进行参数化:给定单位向量(\(u_x,u_y,u_z\) (其中,\(u_x^2+u_y^2+u_z^2=1\)),以围绕沿所述方向方向的轴线以角度\(\theta\)旋转的矩阵R为:

\[
R = \begin{bmatrix}
\cos \theta + u_x^2(1-\cos \theta) & u_x u_y (1-\cos \theta) - u_z \sin \theta & u_x u_z (1-\cos \theta) + u_y \sin \theta \\
u_x u_y (1-\cos \theta) + u_z \sin \theta & \cos \theta + u_y^2(1-\cos \theta) & u_y u_z (1-\cos \theta) - u_x \sin \theta \\
u_x u_z (1-\cos \theta) - u_y \sin \theta & u_y u_z (1-\cos \theta) + u_x \sin \theta & \cos \theta + u_z^2(1-\cos \theta)
\end{bmatrix}
\]

[0029] 通过位姿变换P的3D点X的像素位置x通过以下两个运算来获得:

\[
x' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = PX
\]

[0031] 且

\[
x = \left(\frac{x'}{z'}\right)
\]

[0033] 视图合成器31优选地执行光场渲染(当辅助视图为该类型时),或者在更传统的立体捕捉设置下执行基于图像的渲染。视图合成器31的输出为视差修正的详细视图,该视差修正的详细视图现在能够容易地被融合在全景视图中。

[0034] 为此,图像融合级23通过由视差修正级22生成的高细节变形的视图来替代全景视图中感兴趣的对象的轮廓内的内容。然后,该轮廓的边界周围的小区域(例如,沿法线方向的±2个像素)被用来平滑地对变形的视图的精细细节与来自全景视图的模糊背景之间的
不连续进行内插。在更有利的情景中（即，全景视图和详细视图二者具有类似视点），该处理能够针对感兴趣的对象和其周围的背景二者来完成。在这样的情形中，轮廓对应于详细视图的图像区域，平滑的内插被实施于这样的矩形的周线。这样的示例在图6中被示出，该示例通过使用图4的装置来获得。
图5

图6