

(12) 发明专利申请

(10) 申请公布号 CN 101815225 A

(43) 申请公布日 2010. 08. 25

(21) 申请号 200910118218. 6

(22) 申请日 2009. 02. 25

(71) 申请人 三星电子株式会社

地址 韩国京畿道水原市

申请人 北京三星通信技术研究有限公司

(72) 发明人 王西颖 王海涛 马赓宇 金智渊

(74) 专利代理机构 北京铭硕知识产权代理有限公司 11286

代理人 郭鸿禧 杨静

(51) Int. Cl.

H04N 13/00 (2006. 01)

H04N 13/02 (2006. 01)

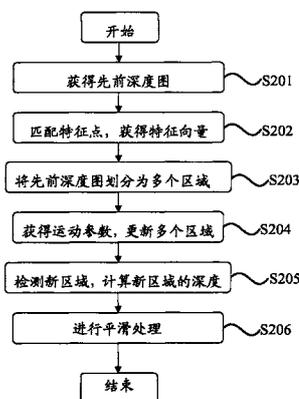
权利要求书 2 页 说明书 6 页 附图 4 页

(54) 发明名称

生成深度图的方法及其设备

(57) 摘要

提供一种生成深度图的方法及其设备, 所述方法包括: 通过稠密匹配方法, 使用单目视频流的第一帧和第二帧获得表示第二帧的像素深度的第一深度图; 检测所述单目视频流的第二帧和第三帧的特征点, 将第三帧的特征点与第二帧的特征点进行匹配以获得特征向量, 并根据所述第二帧的像素深度将第一深度图划分为多个区域; 通过对检测到的特征点进行运动估计来获得运动参数, 并基于所述运动参数和所述特征向量更新所述多个区域的参数, 从而更新所述多个区域以作为表示第三帧的像素深度的第二深度图的一部分; 在第三帧中检测新区域, 并通过计算所述第三帧中的新区域的像素深度作为第二深度图的另一部分, 从而获得第二深度图。



1. 一种从单目视频生成深度图的方法,包括步骤:

(a) 通过稠密匹配方法,使用单目视频流的第一帧和第二帧获得表示第二帧的像素深度的第一深度图;

(b) 检测所述单目视频流的第二帧和第三帧的特征点,将第三帧的特征点与第二帧的特征点进行匹配以获得特征向量,并根据所述第二帧的像素深度将第一深度图划分为多个区域;

(c) 通过对检测到的特征点进行运动估计来获得运动参数,并基于所述运动参数和所述特征向量更新所述多个区域的参数,从而更新所述多个区域以作为表示第三帧的像素深度的第二深度图的一部分;

(d) 在第三帧中检测新区域,所述新区域是在第三帧中出现而未在第二帧中出现的区域,并通过计算所述新区域的像素深度作为第二深度图的另一部分,从而获得第二深度图;以及

(e) 对第二深度图进行平滑处理。

2. 如权利要求 1 所述的方法,其特征在于步骤 (d) 进一步包括步骤:

当所述新区域是与首次出现的场景相应的区域时,通过第三帧和第四帧对该区域进行稠密匹配方法来获得该区域的像素深度,并且当所述新区域是与被覆盖的场景相应的区域时,基于该区域的相邻区域的像素深度通过图像修复方法计算该区域的像素深度。

3. 如权利要求 1 所述的方法,其特征在于所述稠密匹配方法是图分割算法、信任传播算法和动态规划算法中的任何一种。

4. 如权利要求 1 所述的方法,其特征在于步骤 (b) 进一步包括步骤:对所述特征向量进行滤波处理以提高匹配准确度。

5. 如权利要求 1 所述的方法,其特征在于步骤 (c) 中所述多个区域的参数为像素深度、几何形状和位置。

6. 如权利要求 1 所述的方法,其特征在于所述视频流的格式是 AVI 和 MPG 中的任何一种。

7. 一种从单目视频生成深度图的设备,包括:

初始化模块,通过稠密匹配方法,使用单目视频流的第一帧和第二帧获得表示第二帧的像素深度的第一深度图;

处理模块,检测所述单目视频流的第二帧和第三帧的特征点,将第三帧的特征点与第二帧的特征点进行匹配以获得特征向量,根据第二帧的特征点的像素深度将第一深度图划分为多个区域,通过对检测到的特征点进行运动估计来获得运动参数,基于所述运动参数和所述特征向量更新所述多个区域的参数,从而更新所述多个区域以作为表示第三帧的像素深度的第二深度图的一部分,在第三帧中检测新区域,所述新区域是在第三帧中出现而未在第二帧中出现的区域,并计算所述第三区域中的新区域的像素深度作为第二深度图的另一部分,从而获得第二深度图;以及

平滑模块,对第二深度图进行平滑处理。

8. 如权利要求 7 所述的设备,其特征在于:当所述新区域是与首次出现的场景相应的区域时,处理模块通过第三帧和第四帧对该区域进行稠密匹配方法来获得该区域的像素深度,并且当所述新区域是与被覆盖的场景相应的区域时,处理模块基于该区域的相邻区域

的像素深度通过图像修复方法计算该区域的像素深度。

9. 如权利要求 7 所述的设备,其特征在于所述稠密匹配方法是图分割算法、信任传播算法和动态规划算法中的任何一种。

10. 如权利要求 7 所述的设备,其特征在于处理模块对所述特征向量进行滤波处理以提高匹配准确度。

11. 如权利要求 7 所述的设备,其特征在于所述多个区域的参数为像素深度、几何形状和位置。

12. 如权利要求 7 所述的设备,其特征在于所述视频流的格式是 AVI 和 MPG 中的一种。

## 生成深度图的方法及其设备

### 技术领域

[0001] 本发明涉及一种生成深度图的方法及其设备,更具体地,涉及一种从单目视频生成深度图的方法及其设备。

### 背景技术

[0002] 近年来,随着 3D 技术的不断发展,市场上出现了多种配备有 3D 特征的电子产品(诸如,配备有 3D 特征的显示器、电视机、移动终端等)。然而,大多数现有的视频是通过普通的单目摄像机捕捉的,因此这些视频不能直接用于 3D 产品的输入内容。如何从这些单目视频有效地提取 3D 信息成为了主要任务。

[0003] 针对上述任务,目前出现了很多方法。大部分方法主要用于生成单个深度图,而不是深度图序列,因此无法实时生成在线播放的视频的深度图。此外,现有的方法受相机运动的影响和约束很大,例如,使用这些方法需要相机一直保持水平移动。

[0004] 而且,这些方法使用运动估计结构(SFM)来计算相机的参数,这会导致计算的相机参数具有较高的误差,从而会很大程度上影响最终结果。同时,加上平滑处理的时间,这些方法需要大量的时间。

### 发明内容

[0005] 本发明的目的在于提供一种生成深度图的方法,所述方法通过将深度图划分为多个区域并更新所述多个区域,以及检测新区域并仅计算新区域的像素深度来减少了生成深度图的计算量和所花费的时间,并且能够获得清晰的深度图。

[0006] 本发明的另一目的在于提供一种生成深度图的设备,所述设备通过将深度图划分为多个区域并更新所述多个区域,以及检测新区域并仅计算新区域的像素深度来减少了生成深度图的计算量和所花费的时间,并且能够获得清晰的深度图。

[0007] 本发明的一方面在于提供一种从单目视频生成深度图的方法,包括:(a) 通过稠密匹配方法,使用单目视频流的第一帧和第二帧获得表示第二帧的像素深度的第一深度图;(b) 检测所述单目视频流的第二帧和第三帧的特征点,将第三帧的特征点与第二帧的特征点进行匹配以获得特征向量,并根据所述第二帧的像素深度将第一深度图划分为多个区域;(c) 通过对检测到的特征点进行运动估计来获得运动参数,并基于所述运动参数和所述特征向量更新所述多个区域的参数,从而更新所述多个区域以作为表示第三帧的像素深度的第二深度图的一部分;(d) 在第三帧中检测新区域,所述新区域是在第三帧中出现而未在第二帧中出现的区域,并通过计算所述第三帧中的新区域的像素深度作为第二深度图的另一部分,从而获得第二深度图;以及(e) 对第二深度图进行平滑处理。

[0008] 当所述新区域是与首次出现的场景相应的区域时,通过第三帧和第四帧对该区域进行稠密匹配方法来获得该区域的像素深度,并且当所述新区域是与被覆盖的场景相应的区域时,基于该区域的相邻区域的像素深度通过图像修复方法计算该区域的像素深度。

[0009] 所述稠密匹配方法可以是是图分割算法、信任传播算法和动态规划算法中的任何

一种。

[0010] 可以对所述特征向量进行滤波处理以提高匹配准确度。

[0011] 所述多个区域的参数可以是像素深度、几何形状和位置。

[0012] 所述视频流的格式可以是 AVI 和 MPG 中的任何一种。

[0013] 本发明的另一方面在于提供一种从单目视频生成深度图的设备,包括:初始化模块,通过稠密匹配方法,使用单目视频流的第一帧和第二帧获得表示第二帧的像素深度的第一深度图;处理模块,检测所述单目视频流的第二帧和第三帧的特征点,将第三帧的特征点与第二帧的特征点进行匹配以获得特征向量,根据第二帧的特征点的像素深度将第一深度图划分为多个区域,通过对检测到的特征点进行运动估计来获得运动参数,基于所述运动参数和所述特征向量更新所述多个区域参数,从而更新所述多个区域以作为表示第三帧的像素深度的第二深度图的一部分,在第三帧中检测新区域,所述新区域是在第三帧中出现而未在第二帧中出现的区域,并计算所述第三区域中的新区域的像素深度作为第二深度图的另一部分,从而获得第二深度图;以及平滑模块,对第二深度图进行平滑处理。

[0014] 当所述新区域是与首次出现的场景相应的区域时,处理模块通过第三帧和第四帧对该区域进行稠密匹配方法来获得该区域的像素深度,并且当所述新区域是与被覆盖的场景相应的区域时,处理模块基于该区域的相邻区域的像素深度通过图像修复方法计算该区域的像素深度。

[0015] 所述稠密匹配方法可以是图分割算法、信任传播算法和动态规划算法中的任何一种。

[0016] 处理模块可对所述特征向量进行滤波处理以提高匹配准确度。

[0017] 所述多个区域的参数可以是像素深度、几何形状和位置。

[0018] 所述视频流的格式可以是 AVI 和 MPG 中的一种。

## 附图说明

[0019] 通过下面结合附图进行的详细描述,本发明的上述和其他目的和特点将会变得更加清楚,其中:

[0020] 图 1 是示出根据本发明示例性实施例的生成深度图的设备的框图;

[0021] 图 2 是示出根据本发明示例性实施例的生成深度图的方法的流程图;

[0022] 图 3 是第二帧和根据第一帧和第二帧计算的第一深度图;

[0023] 图 4 是示出通过特征点匹配产生的特征向量的示图;

[0024] 图 5 是示出根据本发明示例性实施例的新区域检测结果和不同新区域的示图;

[0025] 图 6 是示出根据本发明示例性实施例的彩色帧与其深度图的示图;

[0026] 图 7 是通过根据本发明示例性实施例的生成深度图的方法获得的深度图序列;以及

[0027] 图 8 是通过传统方法和本发明的生成深度图的方法获得的深度图的比较结果。

## 具体实施方式

[0028] 以下,参照附图来详细说明本发明的实施例。

[0029] 图 1 是示出根据本发明示例性实施例的生成深度图的设备的框图。参照图 1,根据

本发明示例性实施例的生成深度图的设备 100 包括：初始化模块 101、处理模块 102 和平滑模块 103。

[0030] 初始化模块 101 通过稠密匹配方法，使用单目视频流的第一帧和第二帧获得第一深度图。这里，第一帧表示单目视频流的初始帧，第二帧表示第一帧的下一帧，第一深度图是计算出的第二帧的深度图。图 3 的 (a) 表示第二帧，(b) 表示第一深度图。所述视频流的格式可以是 AVI、MPG 等普通视频流格式，并且所述稠密匹配方法可以是图分割算法 (Graph Cut)、信任传播算法 (Belief Propagation)、动态规划算法 (Dynamic Programming) 等。优选地，可以采用图分割算法来获得第一深度图。所述图分割算法应用于各种不同的场合，例如，图像恢复、图像分割和立体稠密匹配等，该算法是一种使用图形原理的针对立体匹配问题的解决方案，其特点在于将立体相应性匹配问题转换为全局能量功能的最小化，通过在特定图形中找到最优分割来实现该最小化。所述图分割算法、信任传播算法和动态规划算法都是本领域公知技术，这里将不再详细描述。

[0031] 需要注意的是，在以下描述中出现的第三帧表示第二帧的下一帧，第二深度图表示第三帧的深度图。可以依此类推，例如，第四帧表示第三帧的下一帧，第三深度图表示第四帧的深度图。上述第一深度图用作获得第二深度图的依据，第二深度图用作获得第三深度图的依据。也就是，利用先前深度图获得当前深度图，为了描述的方便，将不再重复获得第二深度图之后的多个深度图的过程。

[0032] 处理模块 102，检测单目视频流中的第二帧和第三帧特征点，将第三帧的特征点与第二帧的特征点进行匹配以获得特征向量，根据深度将第一深度图划分为多个区域，通过对检测到的特征点进行运动估计来获得运动参数，基于所述运动参数和所述特征向量更新所述多个区域的参数，从而更新所述多个区域以作为第二深度图的一部分，在第三帧中检测新区域，所述新区域是在第三帧中出现而未在第二帧中出现的区域，并计算所述新区域的像素深度作为第二深度图的另一部分，从而获得第二深度图。下面将对此进行详细描述。

[0033] 现有技术中有很多方法可以检测多个帧中的特征点，这里采用 SURF (升速稳健特征, Speed Up Robust Features) 来检测多个帧中的特征点，SURF 是一种新的尺度和旋转不变兴趣点检测器和描述器。具体说来，通过依靠完整图像进行图像旋转来实现 SURF，基于来自兴趣点周围的圆形区域的信息固定可再现的方向，构建调整为所述方向的方形区域，并提取 SURF 描述符。实验证明，SURF 可更加有效且快速地被计算。此外，处理模块 102 对特征点进行的匹配是一种不同于上述的稠密匹配方法的稀疏匹配方法，因此相比于稠密匹配算法，显著减少计算量和计算时间。

[0034] 图 4 表示通过第二帧和第三帧获得的特征向量。之后，处理模块 102 可通过对特征向量进行滤波以提高匹配准确度。具体地，可计算特征向量的平均长度和方向，并将每个特征向量与平均值进行比较，如果相差大于预定阈值，则过滤出该特征向量。

[0035] 处理模块 102 检测到的新区域是指第二帧中未出现而在第三帧中出现的区域。可以对所述新区域进行上述稠密匹配方法来计算新区域的像素深度。具体地，通过利用单目视频流的第三帧和随后的第四帧，如获得第一深度图一样，按照上述稠密匹配方法获得新区域的像素深度作为第二深度图的另一部分。

[0036] 或者，处理模块 102 可首先判断上述新区域的类型。新区域可以分为与首次出现的场景相应的区域和与被覆盖的场景相应的区域这两种类型。当所述新区域是与首次出

现的场景相应的区域时,处理模块 102 通过第三帧和第四帧对该区域进行稠密匹配方法来获得该区域的像素深度,并且当所述新区域是与被覆盖的场景相应的区域时,处理模块 102 基于该区域的相邻区域的深度通过图像修复方法计算该区域的像素深度。由于与被覆盖的区域相邻的区域的像素深度通常与被覆盖的区域的像素深度相似,因此,通过该区域的信息来获得被覆盖的区域的深度,从而获得被覆盖的区域的像素深度。上述图像修复方法将在下面进一步描述。

[0037] 平滑模块 103,对第二深度图进行平滑处理。平滑处理的目的在于不同帧中的相同区域应该保持相似的像素深度以避免显著的深度变化。将在下面进一步描述所述平滑处理。

[0038] 以下,将详细描述生成深度图的步骤。如上所述,在以下描述中出现的帧表示视频流的初始帧,第二帧表示第一帧的下一帧,第一深度图是计算出的第二帧的深度图,第三帧表示第二帧的下一帧,第二深度图表示第三帧的深度图。可以依此类推,例如,第四帧表示第三帧的下一帧,第三深度图表示第四帧的深度图。第一深度图用作获得第二深度图的依据,第二深度图用作获得第三深度图的依据。也就是,利用先前深度图获得当前深度图,为了描述的方便,将不再重复生成第二深度图之后的多个深度图的过程。

[0039] 图 2 是示出根据本发明示例性实施例的生成深度图的方法的流程图。参照图 2,在步骤 S201,初始化模块 101 通过稠密匹配方法,使用视频流的第一帧和第二帧获得第一深度图。实际上,该步骤所花费的时间最长,因为第一深度图的准确度在随后的计算中非常重要,需要在该步骤中尽量准确地获得第一深度图。这里,所述视频流的格式可以是 AVI、MPG 等普通视频流格式,并且所述稠密匹配方法可以是图分割算法 (Graph Cut)、信任传播算法 (Belief Propagation)、动态规划算法 (Dynamic Programming) 等。优选地,可以选择结果较好的图分割算法来获得第一深度图。由于这些算法都是本领域公知技术,因此将省略对此的详细描述。

[0040] 在步骤 S202,处理模块 102 检测单目视频流的第二帧和第三帧的特征点,将第三帧的特征点与第二帧中的特征点进行匹配以获得特征向量。处理模块 102 还可对所述特征向量进行滤波处理以提高匹配准确度。具体地,可计算特征向量的平均长度和方向,并将每个特征向量与平均值进行比较,如果相差大于预定阈值,则过滤出该特征向量。

[0041] 在步骤 S203,处理模块 102 根据像素深度将第一深度图划分为多个区域,其中,每个区域的像素具有相似的深度。这样,每个区域中包含在步骤 S202 获得的特征向量。

[0042] 在步骤 S204,处理模块 102 通过对检测到的特征点进行运动估计来获得运动参数,并基于所述运动参数和所述特征向量更新所述多个区域的参数,从而更新所述多个区域以作为第二深度图的一部分。可通过在 2D 空间中的欧氏变换来更新所述多个区域的参数,即,可通过以下等式 1 来计算运动参数:

$$[0043] \quad \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} r \cos \theta & -r \sin \theta & dx \\ r \sin \theta & r \cos \theta & dy \\ 0 & 0 & 1 \end{bmatrix} * (x \ y \ 1)^T \text{等式 1}$$

[0044] 其中, r 指示缩放因子,即,表示区域沿着 z 轴 (垂直于图像平面) 的运动,如果 r 大于 1,则认为区域朝相机运动,否则认为区域朝相机的相反方向移动。θ 表示运动的角度, dx 和 dy 分别表示位移向量, x 和 y 表示原来的坐标, x' 和 y' 表示运动后的新坐标。

[0045] 优选地,所述多个区域的参数可以是像素深度、几何形状和位置。

[0046] 在步骤 S205,处理模块 102 在第三帧中检测新区域,所述新区域是在第三帧中出现而未在第二帧中出现的区域,并计算该区域的像素深度作为第二深度图的另一部分,从而获得第二深度图。这里,处理模块 102 检测到的新区域是指第二帧中未出现而在第三帧中出现的区域。在步骤 S205,处理模块 102 可使用上述稠密匹配方法,如计算第一深度图一样,通过单目视频流的第三帧和随后的第四帧计算该新区域的像素深度。或者,处理模块 102 也可以将检测到的新区域分类为两种类型的区域。第一种类型的区域为与首次出现的场景相应的区域,第二种类型的区域为与被覆盖的场景相应的区域。

[0047] 图 5 是示出根据本发明示例性实施例的新区域检测结果和不同新区域的示图。其中,图 5 中的 (a) 是第二帧,(b) 是第三帧,(c) 是检测到的新区域。如图 5(c) 所示,区域 1 是与首次出现的场景相应的区域,区域 2 是与被覆盖的场景相应的区域。

[0048] 对区域 1,处理模块 102 通过第三帧和第四帧对该区域进行稠密匹配方法来获得该区域的像素深度。所述稠密匹配方法与在计算第一深度图时使用的稠密匹配方法相同。由于与首次出现的场景相应的区域较小,因此所花费的时间也较短。对区域 2,处理模块基于该区域的相邻区域的像素深度通过图像修复方法计算该区域的像素深度。图像修复方法可由以下等式 2 表示:

$$[0049] \quad d(q) = \frac{\sum_{p \in B(q)} \omega(q, p) [d(p) + \nabla d(p)(q - p)]}{\sum_{p \in B(q)} \omega(q, p)} \quad \text{等式 2}$$

[0050] 其中,q 表示与被覆盖的场景相应的区域的像素,d(q) 表示 q 的像素深度,p 表示该区域的相邻区域的像素,d(p) 表示 p 的像素深度, $\omega(q, p)$  是像素 q 和 p 之间的权重,q 和 p 越靠近权重就越大, $\nabla d(p)$  是像素 p 的梯度,(q-p) 是像素 q 和 p 之间的距离。

[0051] 在步骤 S206,平滑模块 103 对第二深度图进行平滑处理。平滑处理的目的在于不同帧中的相同区域应该保持相似的像素深度以避免显著的深度变化。平滑模块 103 可通过已经获得的在不同帧中的相同区域的像素深度对第二深度图进行平滑。可通过以下等式 3 对第二深度图进行平滑处理:

$$[0052] \quad d_t(q_i) = \frac{\omega_{(t-m)} d_{(t-m)}(q_i) + \dots + \omega_{(t-1)} d_{(t-1)}(q_i) + \dots + \omega_{(t)} d_{(t)}(q_i)}{m+1} \quad \text{等式 3}$$

[0053] 其中, $d_t(q_i)$  表示在第 t 帧中的区域  $q_i$  的像素深度, $\omega$  是权重,距离第 t 帧越近,权重就越大,即  $\omega_{(t-1)}$  比  $\omega_{(t-m)}$  大,m 是小于 t 的正整数。

[0054] 图 6 示出根据本发明示例性实施例的彩色帧与其深度图。图 6 中,上面一排是彩色帧,下面一排是根据本发明示例性实施例获得的彩色帧的深度图。图 7 示出通过根据本发明示例性实施例的生成深度图的方法获得的视频流的深度图序列。该深度图序列的顺序为从左到右,从上到下。

[0055] 如图 6 和图 7 所示,通过根据本发明示例性实施例的生成深度图的方法,可实时提供视频的深度图,从而不仅能够提供已经下载的视频的深度图,而且还能够提供在线播放的视频流的深度图。

[0056] 图 8 是通过传统方法和本发明的生成深度图的方法获得的深度图的比较结果。图 8 中的 (a) 是基于传统的深度图生成方法生成的深度图,图 8 中的 (b) 是根据本发明的生成深度图的方法生成的深度图。从图 8 可以看出与通过传统方法生成的深度图比较,根据本

发明生成的深度图显著提高了深度图的准确度,从而能够给用户提供更清晰的立体效果。

[0057] 根据本发明示例性实施例的生成深度图的方法及其系统,通过将深度图划分为多个区域并更新所述多个区域,以及仅对新区域分类并计算深度图来减少了生成深度图的计算量和所花费的时间。同时,通过保证先前深度图的准确率来提高了随后的深度图的准确率。此外,通过参照先前深度图来计算当前深度图,能够实时生成在线播放的视频流的深度图。

[0058] 本发明不限于上述实施例,在不脱离本发明的精神和范围的情况下,可以对本发明进行形式和细节上的各种改变。

[0059] 产业上的可利用性

[0060] 根据本发明示例性实施例的生成深度图的方法及其系统可应用于将普通视频流作为输入内容以输出立体视频的立体电视机、多视点显示器、移动终端和个人数字助理。

100

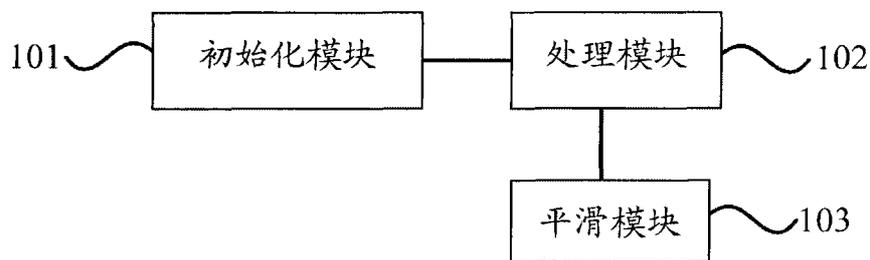


图 1

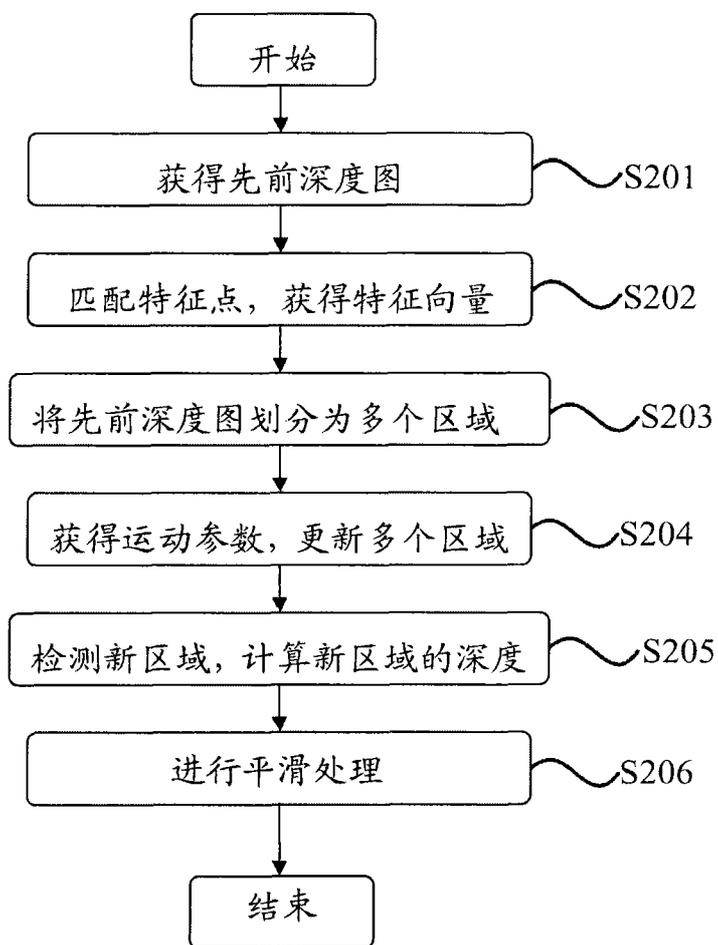


图 2



(a)

(b)

图 3

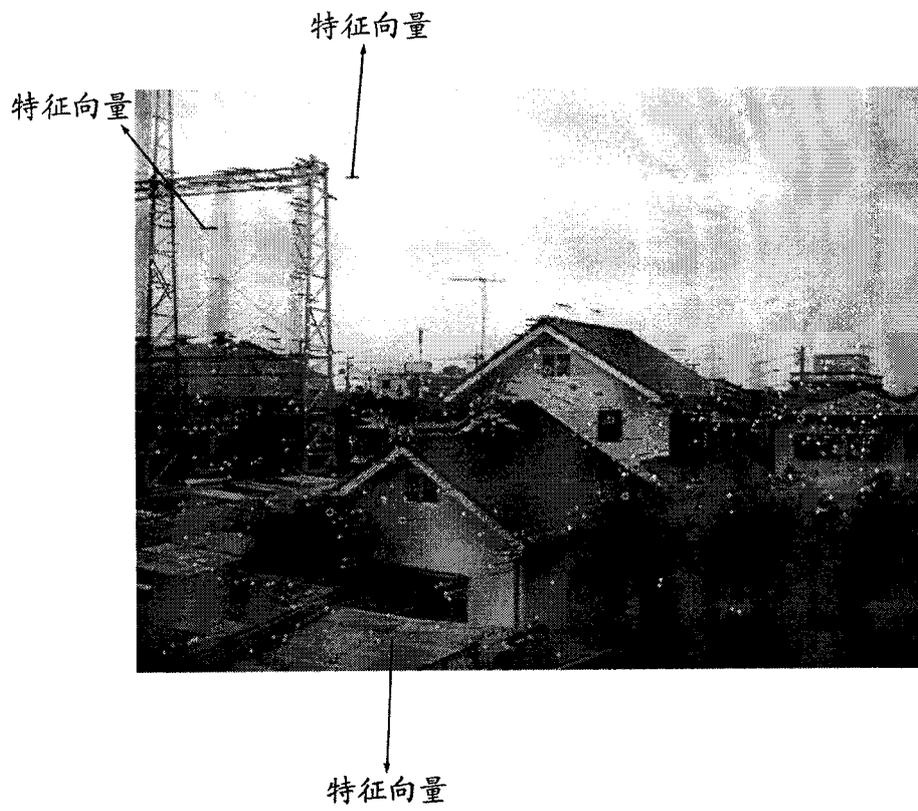


图 4

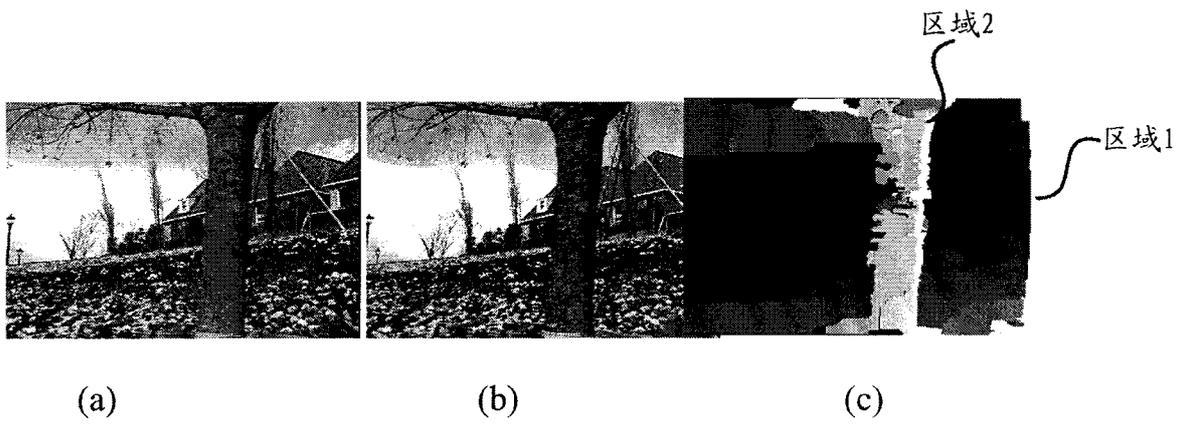


图 5

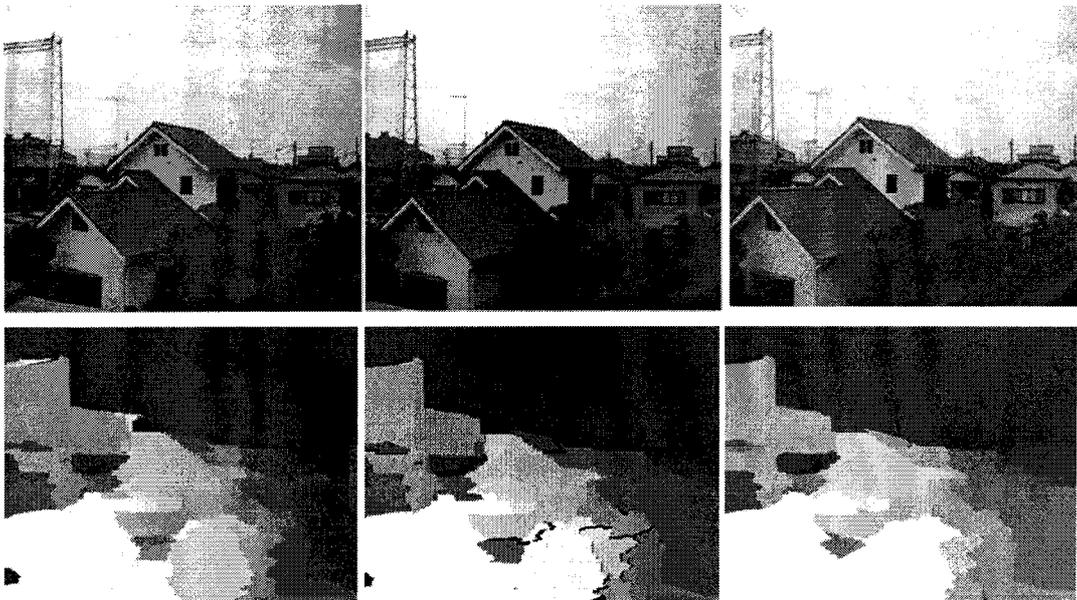
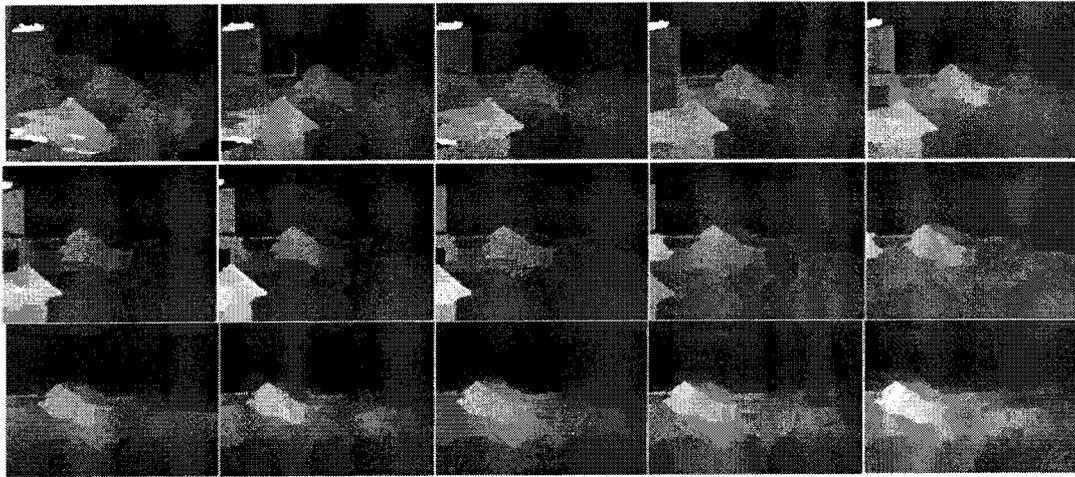
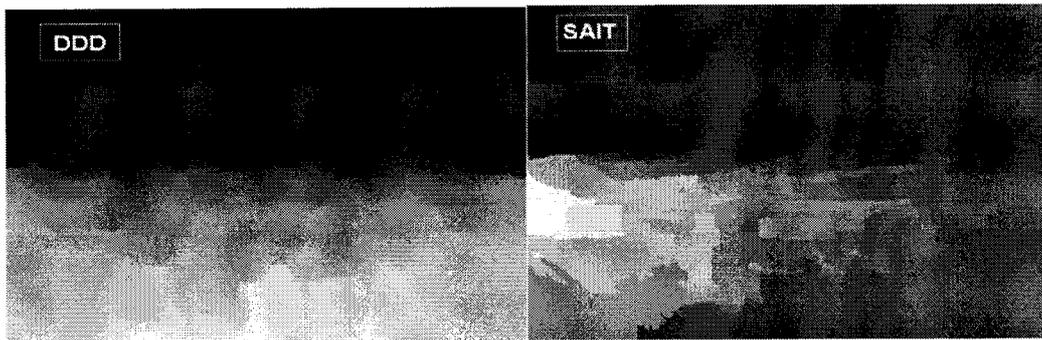


图 6



深度图序列（从左到右，从上到下）

图 7



传统的方法

(a)

本发明

(b)

图 8