



(12)发明专利

(10)授权公告号 CN 105654492 B

(45)授权公告日 2018.09.07

(21)申请号 201511029299.4

(22)申请日 2015.12.30

(65)同一申请的已公布的文献号

申请公布号 CN 105654492 A

(43)申请公布日 2016.06.08

(73)专利权人 哈尔滨工业大学

地址 150001 黑龙江省哈尔滨市南岗区西
大直街92号

(72)发明人 王宽全 李兆歆 左旺孟 张磊

(74)专利代理机构 哈尔滨市松花江专利商标事
务所 23109

代理人 杨立超

(51)Int.Cl.

G06T 7/55(2017.01)

G06T 17/30(2006.01)

(56)对比文件

CN 103247075 A,2013.08.14,

CN 103500013 A,2014.01.08,

CN 104933704 A,2015.09.23,

CN 104715504 A,2015.06.17,

CN 103559737 A,2014.02.05,

梅峰 等.基于RGB-D深度相机的室内场景重建.《中国图象图形学报》.2015,第20卷(第2015
年第10期),第1366-1373页.

审查员 罗伟

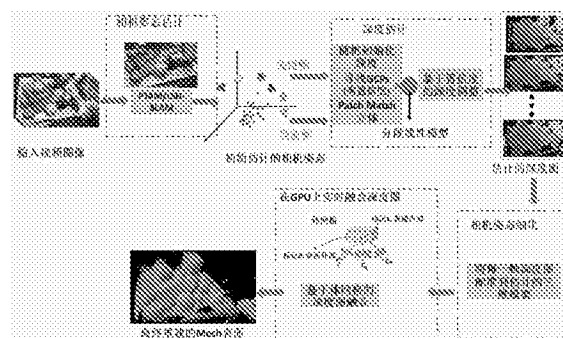
权利要求书3页 说明书7页 附图5页

(54)发明名称

基于消费级摄像头的鲁棒实时三维重建方
法

(57)摘要

基于消费级摄像头的鲁棒实时三维重建方法,本发明涉及基于消费级摄像头的鲁棒实时三维重建方法。本发明是要解决现有方法计算成本高,重建的模型不精确并且不完整的问题。一、在摄像头移动的过程中,基于摄像头当前的视频帧作为输入,估计每一视频帧在场景坐标系下的相机姿态;二、在视频帧中选择最佳的关键帧用于深度估计;三、采用快速鲁棒的深度估计算法估计每一视频帧的深度信息得到每一视频帧的深度图;四、将每一视频帧的深度图转换为截断符号距离场,在体素上并行地执行TSDF的加权平均,增量地融合每一视频帧的深度图,并通过Marching cubes算法构建三角网格表面。本发明应用于图像处理领域。



1. 基于消费级摄像头的鲁棒实时三维重建方法, 其特征在于按以下步骤实现:

一、在摄像头移动的过程中, 基于摄像头当前的视频帧作为输入, 估计每一视频帧在场景坐标系下的相机姿态, 其具体过程为:

(a) 构建一组关键帧集合 \mathcal{K} :

在摄像头移动的过程中, 依据时间距离和空间距离阈值从视频帧中选取关键帧 k , 每一个关键帧对应一个估计的相机姿态, 所有关键帧构成关键帧集合 \mathcal{K} ;

(b) 构建三维图 \mathcal{M} :

三维图 \mathcal{M} 中包含点云数据 $(p_1, p_2, p_3 \dots p_i \dots p_{|\mathcal{M}|})$, 其中 p_i 为点云数据中的某一三维点, $|\mathcal{M}|$ 为 \mathcal{M} 的基, 即 \mathcal{M} 中元素的数量, 当新的关键帧被加入关键帧集合 \mathcal{K} 时, 它与关键帧集合 \mathcal{K} 中其他关键帧执行立体匹配, 产生新的点云数据加入 \mathcal{M} , 点云 \mathcal{M} 中的每一个三维点 p_i 记录着它的三维坐标, 法线方向, 像素特征;

(c) 三维图 \mathcal{M} 和关键帧集合 \mathcal{K} 通过全局捆绑调整进行精细优化:

$$\{\{\hat{\mu}_2 \dots \hat{\mu}_{|\mathcal{K}|}\}, \{\hat{p}_1, \hat{p}_2, \hat{p}_3 \dots \hat{p}_{|\mathcal{M}|}\}\} = \arg \min_{\{\{\mu\}, \{p\}\}} \sum_{k=1}^{|\mathcal{K}|} \sum_{j \in S_k} Obj(e_{jk})$$

其中 e_{jk} 为第 j 个三维点相对于第 k 个关键帧的重投影误差, Obj 为Huber函数, 以增加对噪声和野点的鲁棒性, S_k 为在关键帧 k 上可见的三维图 \mathcal{M} 的子集, $\hat{\mu}_2$ 和 $\hat{\mu}_{|\mathcal{K}|}$ 分别表示第2个和第 $|\mathcal{K}|$ 个关键帧的相机姿态估计值; μ 为待优化的相机姿态, p 为待优化的三维点;

(d) 估计每一视频帧在场景坐标系下的相机姿态:

相机姿态为三维空间的刚体变换, 假设当前视频帧序号为 i , 相机姿态包括了 3×3 的旋转矩阵 R_i 和 3×1 的平移矢量 t_i , 在李群和李代数中, 用一个6维的矢量 μ_i 等价的表示相机姿态; 配准当前视频帧 i 与三维图 \mathcal{M} , 即寻找最优的相机姿态参数 $\hat{\mu}_i$, 使得三维图中相对于视频帧 i 的可见点集 S_i 的重投影误差最小:

$$\hat{\mu}_i = \arg \min_{\mu} \sum_{j \in S_i} Obj(e_j)$$

其中, μ 为待优化相机姿态, e_j 代表 S_i 中第 j 个三维点相对于当前视频帧 i 的重投影误差, Obj 为Huber函数;

二、在视频帧中选择最佳的关键帧用于深度估计, 其具体过程为:

(一) 将关键帧集合 \mathcal{K} 中的关键帧按照和当前帧的基线大小的升序排列, 并选择前 M 帧构成一个子集, 从中选择与当前帧夹角最小的关键帧子集 \mathcal{K}_1 , 假设关键帧集合中的相机中心坐标依次是 $c_1, c_2, c_3 \dots c_n$, 当前帧的相机中心坐标为 c , 当前帧与第 m 个关键帧的基线的计算方法是:

$$D_m = \sqrt{(c - c_m)^2};$$

(二) 根据基线大小, 按照升序进行排序, 根据距离阈值 T 从中选择一个关键帧子集 \mathcal{K}_1 ; T 值定义为相邻关键帧之间距离的平均值的2倍, 当前帧与关键帧之间的夹角的计算如下:

假设当前帧的光学轴为 r , 某一关键帧的光学轴为 r_m , 则二者的夹角为:

$$\arccos\left(\frac{\mathbf{r} \cdot \mathbf{r}_m}{\|\mathbf{r}\| \|\mathbf{r}_m\|}\right)$$

其中 \cdot 为点积操作, $\|\cdot\|$ 为矢量范数;

(三) 根据当前帧与关键帧子集 \mathcal{K}_1 中的关键帧逐个的计算夹角, 从中选择夹角最小的关键帧 K_m 作为最佳关键帧用于下一步的深度估计;

三、采用快速鲁棒的深度估计算法估计每一视频帧的深度信息得到每一视频帧的深度图:

四、将每一视频帧的深度图转化为截断符号距离场, 并在体素上执行增量的深度融合, 最后输出三角网格表面, 即完成了基于消费级摄像头的鲁棒实时三维重建方法。

2. 根据权利要求1所述的基于消费级摄像头的鲁棒实时三维重建方法, 其特征在于所述步骤三采用快速鲁棒的深度估计算法估计每一视频帧的深度信息:

(a) 随机初始化: 对于当前帧 I_t 和最佳关键帧 I_{K_m} , 目标是估计 I_t 上每一像素的深度; 设深度搜索范围为 $[d^{\min}, d^{\max}]$, 对于 I_t 中的每一个像素位置 $p = (x, y)$, 随机初始化一个深度值 d_p , 基于对光照变化的鲁棒性, 使用ZNCC的负值作为度量深度假设的成本, 对于当前帧 I_t 的像素位置 p , 选择一个以它为中心的 $w \times w$ 的窗口, 对于窗口中的每一个像素位置 \mathbf{p}_j^w , 其深度值为 d_p , 计算其在 I_{K_m} 的对应的像素位置, 并计算匹配成本; 其中, 所述 d^{\min} 和 d^{\max} 分别为深度搜索范围中的最小深度和最大深度;

(b) 深度传播: 采用基于扫描线的传播方式, 即:

(i) 行扫描: 逐行从左向右比较相邻像素的成本, 如果 $\text{cost}(d_{x-1,y}) < \text{cost}(d_{x,y})$, 则 $d_{x,y} = d_{x-1,y}$, 然后逐行从右向左扫描, 如果 $\text{cost}(d_{x+1,y}) < \text{cost}(d_{x,y})$, 则 $d_{x,y} = d_{x+1,y}$; 其中, 所述 $\text{cost}(d_{x,y})$ 是像素 (x, y) 的深度为 $d_{x,y}$ 时的图像匹配成本, $\text{cost}(d_{x-1,y})$ 是像素 $(x-1, y)$ 的深度为 $d_{x-1,y}$ 时的图像匹配成本, $\text{cost}(d_{x+1,y})$ 是像素 $(x+1, y)$ 的深度为 $d_{x+1,y}$ 时的图像匹配成本;

(ii) 列扫描: 逐列从上向下比较相邻像素的成本, 如果 $\text{cost}(d_{x,y-1}) < \text{cost}(d_{x,y})$, 则 $d_{x,y} = d_{x,y-1}$; 然后逐列从下向上扫描, 如果 $\text{cost}(d_{x,y+1}) < \text{cost}(d_{x,y})$, 则 $d_{x,y} = d_{x,y+1}$; 其中, 所述 $\text{cost}(d_{x,y-1})$ 是像素 $(x, y-1)$ 的深度为 $d_{x,y-1}$ 时的图像匹配成本, $\text{cost}(d_{x,y+1})$ 是像素 $(x, y+1)$ 的深度为 $d_{x,y+1}$ 时的图像匹配成本;

因为每一行或列的操作相对于其他行或列都是数据独立的, 使用GPU并行实现, 通过上述操作, 正确的深度假设会从图像像素的四邻域区域传播到当前像素;

(c) 置信度调整

(i) 提取高置信度的深度值: 使用深度假设的图像间匹配成本作为置信度的度量指标; 选择一个ZNCC阈值, 选取ZNCC值大或对应的成本较小的像素位置; 去掉大部分的弱纹理区域的像素, 而同时去除一部分在强纹理区域深度假设不正确的像素值;

对于每一个候选点, 进一步执行空间一致性检查操作以过滤可能存在的野点; 在以当前候选点为中心的局部窗口内 ($w_s \times w_s$) 统计与当前候选点的深度假设小于某一阈值 T_d 的像素的数量; 如果满足以上条件的像素数量大于 T_c , 则将当前候选点及其深度值作为地面控制点, 最终, 通过以上步骤记录一组高置信度的点及其深度值作为GCPs;

(ii) 基于GCPs的深度预测: 使用一组GCPs和所对应的深度值预测周边临近区域的深度值; 采用一组分段线性函数, 在GCPs上基于计算Delaunay三角化实现GCPs周围临近区域的

深度值插值, 对于一个GCPs点 (x_{g_j}, y_{g_j}) , 有如下的平面方程 π 的约束:

$$d_{x_{g_j}, y_{g_j}} = a_i x_{g_j} + b_i y_{g_j} + c_i$$

其中 i 是三角形的索引, $d_{x_{g_j}, y_{g_j}}$ 为GCPs点 (x_{g_j}, y_{g_j}) 的深度值, a_i, b_i 和 c_i 为平面方程的系数, 对于每一个三角形, 包含3个GCPs, 获得三个类似的等式, 因此平面参数 (a_i, b_i, c_i) 通过求解线性系统得出; 对于图像中某一像素位置 (x_n, y_n) , 如果位于三角形 i 的投影范围内, 则其经由GCPs预测的深度值为:

$$d_{x_n, y_n}^{g_i} = a_i x_n + b_i y_n + c_i$$

$d_{x_n, y_n}^{g_i}$ 为像素位置 (x_n, y_n) 经由GCPs的预测的深度值, 通过以上的操作, 将GCPs的可靠深度值分段线性的插值到周围的区域, 从而获得对弱纹理区域更平滑可靠的预测;

(d) 后处理: 经由置信度调整处理后, 系统包含一些不连续的小块, 根据邻域深度的一致性将深度图聚成一些连通分量, 去除那些像素数目小于150的连通分量。

3. 根据权利要求2所述的基于消费级摄像头的鲁棒实时三维重建方法, 其特征在于所述步骤四将每一视频帧的深度图转化为截断符号距离场并在体素上增量的融合:

(一) 待重建的物体表面位于解析度为 $N_x \times N_y \times N_z$ 的包围盒中, 其中 N_x, N_y 和 N_z 分别为 X, Y 和 Z 方向的体素的解析度; 视线为从相机中心出发穿过像素的射线; 从体素中心到表面的符号距离定义在视线上, 且增加一个权值因子 $1/\delta$ 并截断至区间 $[-1, 1]$, δ 控制相关的近表面区域的宽度, 反映了在 z -axis方向对深度值期望的不确定性;

(二) 假设在时间 t , 获得深度图 D_t , 将深度图 D_t 转换为TSDF f_{D_t} , $f_{D_t} \in [-1, 1]$, 在逐帧的深度融合的过程中, 期望获得在包围盒的体素上的一个平滑的符号距离场 u , 使得 $u=0$ 为估计的模型表面, $u<0$ 位于模型表面内侧, $u>0$ 位于模型表面外侧, 对于每一个体素 x , 获得如下的增量更新:

$$u_t(\mathbf{x}) = \frac{w_{t-1}(\mathbf{x})u_{t-1}(\mathbf{x}) + w_{D_t}(\mathbf{x})f_{D_t}(\mathbf{x})}{w_{t-1}(\mathbf{x}) + w_{D_t}(\mathbf{x})}$$

$$w_t(\mathbf{x}) = w_{t-1}(\mathbf{x}) + w_{D_t}(\mathbf{x})$$

其中, w_t 是 t 时刻的TSDF的权重, 与表面的不确定性成正比, w_{t-1} 是 $t-1$ 时刻的权重, w_{D_t} 是 t 时刻的权重增量, u_{t-1} 是 $t-1$ 时刻时的符号距离场 u , (x) 代表在体素 x 位置, 权重增量在实际中设置为 $w_{D_t}(\mathbf{x})=1$, f_{D_t} 为深度图 D_t 对应的TSDF; 融合过程可独立的在每一个体素 x 上执行, 因此使用GPU进行并行处理, 融合过程可独立的在每一个体素 x 上执行, 因此使用GPU进行并行处理。

基于消费级摄像头的鲁棒实时三维重建方法

技术领域

[0001] 本发明涉及基于消费级摄像头的鲁棒实时三维重建方法。

背景技术

[0002] 随着手机,数码相机的普及,获取高质量的图像越来越便捷,一个迫切的需求是利用这些图片数据重建我们所生活的三维世界,包括:物体、场景甚至整个环境。已有的基于图像的三维重建方法,所采用的工业相机设备价格昂贵,且计算成本较高,对于一个小场景的重建一般需要一台高性能计算机几个小时的处理时间。然而,传感器噪声,遮挡和光照变化经常导致三维重建任务失败,这些问题仅通过观察图像往往难以预测。由于这些问题,花费几个小时重建的模型往往是不精确的,不完整的,且包含瑕疵。在很多场合,几乎不可能重新返回待重建场景重新捕获图像数据。即便可以对待重建场景反复的拍摄和重建,造成的计算负担和工作量仍然是巨大的。因此,期望有一个重建方法能够在拍摄过程中就实时的反馈重建结果的质量,并帮助用户评估重建质量,规划可提高重建精度和完整度的相机移动轨迹。

发明内容

[0003] 本发明是要解决现有方法计算成本高,重建的模型不精确并且不完整的问题,而提供了基于消费级摄像头的鲁棒实时三维重建方法。

[0004] 基于消费级摄像头的鲁棒实时三维重建方法,按以下步骤实现:

[0005] 一、在摄像头移动的过程中,基于摄像头当前的视频帧作为输入,估计每一视频帧在场景坐标系下的相机姿态:

[0006] 二、在视频帧中选择最佳的关键帧用于深度估计;

[0007] 三、采用快速鲁棒的深度估计算法估计每一视频帧的深度信息得到每一视频帧的深度图:

[0008] 四、将每一视频帧的深度图转化为截断符号距离场,并在体素上增量的融合,最后输出三角网格表面,即完成了基于消费级摄像头的鲁棒实时三维重建方法。

[0009] 发明效果:

[0010] 本发明技术方案提出一个实时、低成本的三维重建方法,可以实时地且高质量地重建并可可视化真实场景。如图1所示,提出的方法使用低成本消费级摄像头的视频作为输入,如网络摄像头或是手机摄像头。由于这些消费级摄像头广泛用于日常生活,方法实际上可允许重建相当广泛的场景。三维重建的模型可以直接用于增强现实,3D打印,计算机游戏等应用,如图2。在实验中,使用一枚视频分辨率为 640×360 的网络摄像头,帧率为30Hz。算法在一台配置为CPU i7,32G RAM,显卡NVIDIA Titan black,6G显存的台式电脑上运行。深度融合中体素的解析度为 $256 \times 256 \times 256$ 。算法平均运行速度是每秒5帧。研究内容主要包括了视频帧相机姿态的实时估计,基于置信度的鲁棒深度估计算,以及基于GPU的深度融合。虽然基于消费级摄像头的实时三维重建非常有价值,提出的方法需要解决几个挑战问

题：首先，相机在每一个视频帧中的姿态是未知的；第二，消费级相机的成像质量比工业级相机要低很多；第三，弱纹理或无纹理区域广泛存在，这些区域的三维重建的质量和可靠性较差，比如人脸重建。本发明有效解决上述难题。

[0011] (1) 通过随机策略初始化深度搜索范围，深度估计算法具有尺度可扩展性；

[0012] (2) 提出的深度估计算法对于弱纹理区域比较鲁棒。算法为每一个深度估计一个置信度值，并基于高置信度区域构建一个分段线性模型用于预测弱纹理区域的深度。

[0013] (3) 大多数重建方法使用稀疏采样的关键帧进行重建，提出的专利使用视频的每一帧进行模型的估计，充分的利用了视频图像的冗余信息，能够生成更稠密的重建，以及处理遮挡问题。

[0014] (4) 在深度融合阶段，我们采用了基于GPU和体网格的融合，可以高效的抑制噪声并重建复杂的场景结构。通过对体数据进行可视化，可以实时可视化重建过程。

附图说明

[0015] 图1是提出的实时三维重建方法的概述图；

[0016] 图2是提出的重建方法在3D打印上的应用实例图；

[0017] 图3是提出的单目相机三维重建的算法流程图；

[0018] 图4是深度传播示意图；

[0019] 图5是基于置信度的深度调整；

[0020] 图6是截断符号距离场的图示；

[0021] 图7单目实时三维重建方法对静态场景三维重建结果图；其中，第一列和第三列是输入的视频的某一帧，第二列和第四列是可视化的三维重建结果；

[0022] 图8是提出的单目相机三维重建方法对人脸的重建结果以及与LSD-SLAM算法的对比图；其中，第一行，从左到右：输入视频的一帧，基于图像间直接匹配的LSD-SLAM算法估计的深度图，基于随机初始化和成本传播生成的深度图，以及使用置信度调整后的深度图；

[0023] 第二行，从左到右：基于图像间直接匹配的LSD-SLAM的稀疏点云重建结果，不使用基于置信度的深度调整的稠密重建结果，使用所提出的基于置信度的深度调整的稠密重建结果；

[0024] 图9是提出的单目相机三维重建方法对人脸的重建结果以及与LSD-SLAM算法的对比图；其中，第一行，从左到右：输入视频的一帧，基于图像间直接匹配的LSD-SLAM算法估计的深度图，基于随机初始化和成本传播生成的深度图，以及使用置信度调整后的深度图；

[0025] 第二行，从左到右：基于图像间直接匹配的LSD-SLAM的稀疏点云重建结果，不使用基于置信度的深度调整的稠密重建结果，使用所提出的基于置信度的深度调整的稠密重建结果；

[0026] 图10是提出的单目相机三维重建方法对人脸的重建结果以及与LSD-SLAM算法的对比图；第一行，从左到右：输入视频的一帧，基于图像间直接匹配的LSD-SLAM算法估计的深度图，基于随机初始化和成本传播生成的深度图，以及使用置信度调整后的深度图；

[0027] 第二行，从左到右：基于图像间直接匹配的LSD-SLAM的稀疏点云重建结果，不使用基于置信度的深度调整的稠密重建结果，使用所提出的基于置信度的深度调整的稠密重建结果。

具体实施方式

[0028] 具体实施方式一：结合图1~图10说明本实施方式，本实施方式的基于消费级摄像头的鲁棒实时三维重建方法，按以下步骤实现：

[0029] 一、在摄像头移动的过程中，基于摄像头当前的视频帧作为输入，估计每一视频帧在场景坐标系下的相机姿态：

[0030] 二、在视频帧中选择最佳的关键帧用于深度估计；

[0031] 三、采用快速鲁棒的深度估计算法估计每一视频帧的深度信息得到每一视频帧的深度图：

[0032] 四、将每一视频帧的深度图转化为截断符号距离场，并在体素上增量的融合，最终初始三角网格表面，即完成了基于消费级摄像头的鲁棒实时三维重建方法。

[0033] 具体实施方式二：本实施方式与具体实施方式一不同的是：所述步骤一具体为：

[0034] (a) 构建一组关键帧集合 \mathcal{K} ：

[0035] 在摄像头移动的过程中，依据时间距离和空间距离阈值从视频帧中选取关键帧 k ，每一个关键帧对应一个估计的相机姿态，所有关键帧构成关键帧集合 \mathcal{K} ；

[0036] (b) 构建三维图 \mathcal{M} ：

[0037] 三维图 \mathcal{M} 中包含点云数据 $(p_1, p_2, p_3 \dots p_i \dots p_{|\mathcal{M}|})$ ，其中 p_i 为点云数据中的某一三维点， $|\mathcal{M}|$ 为 \mathcal{M} 的基，即 \mathcal{M} 中元素的数量，当新的关键帧被加入关键帧集合 \mathcal{K} 时，它与关键帧集合 \mathcal{K} 中其他关键帧执行立体匹配，产生新的点云数据加入 \mathcal{M} ，点云 \mathcal{M} 中的每一个三维点 p_i 记录着它的三维坐标，法线方向，像素特征。当关键帧加入关键帧集合 \mathcal{K} 时，它与关键帧集合 \mathcal{K} 中其他关键帧进行匹配，产生点云数据；

[0038] (c) 三维图 \mathcal{M} 和关键帧集合 \mathcal{K} 通过全局捆绑调整进行精细优化：

$$[0039] \quad \{\{\hat{\mu}_2, \dots, \hat{\mu}_{|\mathcal{K}|}\}, \{\hat{p}_1, \hat{p}_2, \hat{p}_3, \dots, \hat{p}_{|\mathcal{M}|}\}\} = \arg \min_{\{\{\mu\}, \{p\}\}} \sum_{k=1}^{|\mathcal{K}|} \sum_{j \in S_k} \text{Obj}(e_{jk})$$

[0040] 其中 e_{jk} 为第 j 个三维点相对于第 k 个关键帧的重投影误差， Obj 为Huber函数，以增加对噪声和野点的鲁棒性， S_k 为在关键帧 k 上可见的三维图 \mathcal{M} 的子集， $\hat{\mu}_2$ 和 $\hat{\mu}_{|\mathcal{K}|}$ 分别表示第2个和第 $|\mathcal{K}|$ 个关键帧的相机姿态； μ 为待估计的相机姿态， p 为待估计的三维点；

[0041] (d) 估计每一视频帧在场景坐标系下的相机姿态：

[0042] 相机姿态为三维空间的刚体变换，假设当前视频帧序号为 i ，相机姿态包括了 3×3 的旋转矩阵 R_i 和 3×1 的平移矢量 t_i ，在李群和李代数中，用一个6维的矢量 μ_i 等价的表示相机姿态；配准当前视频帧 i 与三维图 \mathcal{M} ，即寻找最优的相机姿态参数 $\hat{\mu}_i$ ，使得三维图中相对于视频帧 i 的可见点集 S_i 的重投影误差最小：

$$[0043] \quad \hat{\mu}_i = \arg \min_{\mu} \sum_{j \in S_i} \text{Obj}(e_j)$$

[0044] 其中, μ 为待估计的相机姿态向量, e_j 代表 S_i 中第 j 个三维点相对于当前视频帧 i 的重投影误差, Obj 为 Huber 函数。

[0045] 其它步骤及参数与具体实施方式一相同。

[0046] 具体实施方式三: 本实施方式与具体实施方式一或二不同的是: 所述步骤二具体为:

[0047] (一) 将关键帧集合 \mathcal{K} 中的关键帧按照和当前帧的基线大小的升序排列, 并选择前 M 帧构成一个子集, 从中选择与当前帧夹角最小的关键帧子集 \mathcal{K}_1 , 假设关键帧集合中的相机中心坐标依次是 $c_1, c_2, c_3 \dots c_n$, 当前帧的相机中心坐标为 c , 当前帧与第 m 个关键帧的基线的计算方法是:

$$[0048] \quad D_m = \sqrt{(c - c_m)^2};$$

[0049] (二) 根据基线大小, 按照升序进行排序, 根据距离阈值 T 从中选择一个关键帧子集 \mathcal{K}_1 ; T 值定义为相邻关键帧之间距离的平均值的 2 倍, 当前帧与关键帧之间的夹角的计算如下:

[0050] 假设当前帧的光学轴为 r , 某一关键帧的光学轴为 r_m , 则二者的夹角为:

$$[0051] \quad \arccos\left(\frac{r \cdot r_m}{|r| |r_m|}\right)$$

[0052] 其中 \cdot 为点积操作, $||$ 为矢量范数;

[0053] (三) 根据当前帧与关键帧子集 \mathcal{K}_1 中的关键帧逐个的计算夹角, 从中选择夹角最小的关键帧 K_m 作为最佳关键帧用于下一步的深度估计。

[0054] 其它步骤及参数与具体实施方式一或二相同。

[0055] 具体实施方式四: 本实施方式与具体实施方式一至三之一不同的是: 所述步骤三采用快速鲁棒的深度估计算法估计每一视频帧的深度信息:

[0056] (a) 随机初始化: 对于当前立体图像当前帧 I_t 和最佳关键帧 I_{K_m} , 目标是估计 I_t 上每一像素的深度; 设深度搜索范围为 $[d^{\min}, d^{\max}]$, 对于 I_t 中的每一个像素位置 $p = (x, y)$, 随机初始化一个深度值 d_p , 基于对光照变化的鲁棒性, 使用 ZNCC 的负值作为度量深度假设的成本, 对于当前帧 I_t 的像素位置 p , 选择一个以它为中心的 $w \times w$ 的窗口, 对于窗口中的每一个像素位置 p_j^w , 其深度值为 d_p , 计算其在 I_{K_m} 的对应的像素位置, 并计算匹配成本; 其中, 所述 d^{\min} 和 d^{\max} 分别为深度搜索范围中的最小深度和最大深度;

[0057] (b) 如图 4 所示, 深度传播: 采用基于扫描线的传播方式, 即:

[0058] (i) 行扫描: 逐行从左向右比较相邻像素的成本, 如果 $\text{cost}(d_{x-1,y}) < \text{cost}(d_{x,y})$, 则 $d_{x,y} = d_{x-1,y}$, 然后逐行从右向左扫描, 如果 $\text{cost}(d_{x+1,y}) < \text{cost}(d_{x,y})$, 则 $d_{x,y} = d_{x+1,y}$; 其中, 所述 $\text{cost}(d_{x,y})$ 是像素 (x, y) 的深度为 $d_{x,y}$ 时的图像匹配成本, $\text{cost}(d_{x-1,y})$ 是像素 $(x-1, y)$ 的深度为 $d_{x-1,y}$ 时的图像匹配成本, $\text{cost}(d_{x+1,y})$ 是像素 $(x+1, y)$ 的深度为 $d_{x+1,y}$ 时的图像匹配成本;

[0059] (ii) 列扫描: 逐列从上向下比较相邻像素的成本, 如果 $\text{cost}(d_{x,y-1}) < \text{cost}(d_{x,y})$, 则

$d_{x,y}=d_{x,y-1}$;然后逐列从下向上扫描,如果 $\text{cost}(d_{x,y+1}) < \text{cost}(d_{x,y})$,则 $d_{x,y}=d_{x,y+1}$;其中,所述 $\text{cost}(d_{x,y-1})$ 是像素 $(x,y-1)$ 的深度为 $d_{x,y-1}$ 时的图像匹配成本, $\text{cost}(d_{x,y+1})$ 是像素 $(x,y+1)$ 的深度为 $d_{x,y+1}$ 时的图像匹配成本;

[0060] 因为每一行或列的操作相对于其他行或列都是数据独立的,使用GPU并行实现,通过上述操作,正确的深度假设会从图像像素的四邻域区域传播到当前像素;

[0061] (c) 置信度调整

[0062] 为了鲁棒的估计弱纹理区域的深度,提出一种基于置信度的深度估计调整算法(confidence-based depthadjustment)。置信度是深度假设正确性的评价指标,置信度越高,深度假设正确的可能性就越高。算法首先提取当前高置信度的深度假设所对应的像素位置,而后根据这些当前高置信度的深度值和像素位置预测那些当前深度假设的置信度较低的像素点的深度值。算法的具体步骤如下:

[0063] (i) 提取高置信度的深度值:使用深度假设的图像间匹配成本作为置信度的度量指标;选择一个ZNCC阈值,选取ZNCC值大或对应的成本较小的像素位置;去掉大部分的弱纹理区域的像素,而同时去除一部分在强纹理区域深度假设不正确的像素值;

[0064] 对于每一个候选点,进一步执行空间一致性检查操作以过滤可能存在的野点;在以当前候选点为中心的局部窗口内($w_s \times w_s$)统计与当前候选点的深度假设小于某一阈值 T_d 的像素的数量;如果满足以上条件的像素数量大于 T_c ,则将当前候选点及其深度值作为地面控制点,最终,通过以上步骤记录一组高置信度的点和其深度值作为GCPs;

[0065] 通过这一阈值操作,去掉大部分的弱纹理区域的像素,而同时去除一部分在强纹理区域深度假设不正确的像素值;在实际中,ZNCC的阈值设置为0.5,并将ZNCC大于0.5的深度值和对应的像素位置作为候选点。在实验中, w_s 设置为5, T_d 设置为11,而 T_c 设置为6;最终,通过以上步骤可以记录一组高置信度的点和其深度值作为GCPs;过于密集的GPCs会影响后续基于GPCs的深度预测步骤的效率,并且可能包含噪声,出于效率和对噪声鲁棒的考虑,以

上提取GPCs的操作,仅在输入图像中每隔5个像素的位置执行,即对原始输入图像进行 $\frac{1}{5}$ 大小的降采样;

[0066] (ii) 基于GPCs的深度预测:使用一组GPCs和所对应的深度值预测周边临近区域的深度值;采用一组分段线性函数,在GPCs上基于计算Delaunay三角化实现GPCs周围临近区域的深度值插值,对于一个GPCs点 (x_{g_j}, y_{g_j}) ,有如下的平面方程 π 的约束:其中,所述区域为ZNCC阈值小于0.5的区域;

$$[0067] \quad d_{x_{g_j}, y_{g_j}} = a_i x_{g_j} + b_i y_{g_j} + c_i$$

[0068] 其中 i 是三角形的索引, $d_{x_{g_j}, y_{g_j}}$ 为GPCs点 (x_{g_j}, y_{g_j}) 的深度值, a_i, b_i 和 c_i 为平面方程的系数,对于每一个三角形,包含3个GPCs,获得三个类似的等式,因此平面参数 (a_i, b_i, c_i) 通过求解线性系统得出;对于图像中某一像素位置 (x_n, y_n) ,如果位于三角形 i 的投影范围内,则其经由GPCs预测的深度值为:

$$[0069] \quad d_{x_n, y_n}^{g_i} = a_i x_n + b_i y_n + c_i$$

[0070] $d_{x_n, y_n}^{g_i}$ 为像素位置 (x_n, y_n) 经由GCPs的预测的深度值,通过以上的操作,将GCPs的可靠深度值分段线性的插值到周围的区域,从而获得对弱纹理区域更平滑可靠的预测;

[0071] 通过以上的操作,将GCPs的可靠深度值分段线性的插值到周围的区域,从而获得对弱纹理区域更平滑可靠的预测,如图5所示。方法的优势在于可在非均匀的区域进行插值;选取的GCPs不能保证是彼此等距离的网格点,通过Delaunay三角化,不同区域间三角形的大小可以是不一样的,保证了在强纹理区域会有更多的GCPs,构建的三角形较密集且面积较小,通过插值保证了局部的细节恢复,而位于弱纹理区域GCPs点的数量较少,构建的三角形较稀疏且面积较大,通过插值可以产生局部较平滑的结果。

[0072] 图5中虚线链接的像素点为GCPs,结合其深度信息构建空间的三角面 P_{gi} 。图像中某一像素位置 (x_n, y_n) 如恰好位于GCPs所构成的三角形 i 的投影范围内,则 (x_n, y_n) 经由GCPs预测的深度值为其在三角面 P_{gi} 上的投影 $d_{x_n, y_n}^{g_i}$ 。

[0073] (d) 后处理:经由置信度调整处理后,系统包含一些不连续的小块,根据邻域深度的一致性将深度图聚成一些连通分量,去除那些像素数目小于150的连通分量。

[0074] 其它步骤及参数与具体实施方式一至三之一相同。

[0075] 具体实施方式五:本实施方式与具体实施方式一至四之一不同的是:所述步骤四基于TSDF和体素的深度在线增量融合,每一视频帧的深度图转化为截断符号距离场(Truncated signed distance field, TSDF):

[0076] (一) 如图6所示,待重建的物体表面位于解析度为 $N_x \times N_y \times N_z$ 的包围盒中,其中 N_x , N_y 和 N_z 分别为X,Y和Z方向的体素的解析度,视线为从相机中心出发穿过像素的射线;从体素中心到表面的符号距离定义在视线上,且增加一个权值因子 $1/\delta$ 并截断至区间 $[-1, 1]$, δ 控制相关的近表面区域的宽度,反映了在z-axis方向对深度值期望的不确定性;

[0077] (二) 假设在时间 t , 获得深度图 D_t , 将深度图 D_t 转换为TSDF f_{D_t} , $f_{D_t} \in [-1, 1]$, 在逐帧的深度融合的过程中,期望获得在包围盒的体素上的一个平滑的符号距离场 u , 使得 $u=0$ 为估计的模型表面, $u<0$ 位于模型表面内侧, $u>0$ 位于模型表面外侧,对于每一个体素 x , 获得如下的增量更新:

$$[0078] \quad u_t(\mathbf{x}) = \frac{w_{t-1}(\mathbf{x})u_{t-1}(\mathbf{x}) + w_{D_t}(\mathbf{x})f_{D_t}(\mathbf{x})}{w_{t-1}(\mathbf{x}) + w_{D_t}(\mathbf{x})}$$

$$w_t(\mathbf{x}) = w_{t-1}(\mathbf{x}) + w_{D_t}(\mathbf{x})$$

[0079] 其中, w_t 是 t 时刻的TSDF的权重,与表面的不确定性成正比, w_{t-1} 是 $t-1$ 时刻的权重, w_{D_t} 是 t 时刻的权重增量, u_{t-1} 是 $t-1$ 时刻时的符号距离场 u , (x) 代表在体素 x 位置,权重增量在实际中设置为 $w_{D_t}(\mathbf{x})=1$, f_{D_t} 为深度图 D_t 对应的TSDF;融合过程可独立的在每一个体素 x 上执行,因此使用GPU进行并行处理,融合过程可独立的在每一个体素 x 上执行,因此使用GPU进行并行处理。

[0080] 其它步骤及参数与具体实施方式一至四之一相同。

[0081] 实施例:

[0082] 本发明对于弱纹理物体的重建特别有效,比如人脸的重建,与LSD-SLAM算法(Engel J,SchöpsT,Cremers D.LSD-SLAM:Large-Scale Direct MonocularSLAM[C]//Proc.European Conf.Comput.Vis.Zurich,Switzerland:Springer,2014:834-849)的对比实验列在图7-10中,在这些图中,强调了所提出的深度估计算法的有效性。

[0083] 使用提出的方法在apples,pear,boot,dinosaur,earphone和Mario&house上的运行结果如图7所示。从apples,pear和dinosaur的重建效果可以看出,算法可以有效的恢复表面的形状,同时有效的去除噪声,从boot的重建结果可以看出,鞋子上的细微的表面特征也可以有效的恢复。earphone场景包含一个由弱纹理覆盖的耳机,提出的基于置信度的深度调整算法可以较好的重建耳机的形状。Mario&house场景包含一个Mario模型,一个房屋微缩模型和一个苹果。其中Mario模型和苹果的纹理较弱,而房屋模型的纹理较丰富,从重建结果可以看出,本文基于置信度的深度调整算法,在弱纹理区域可以有效的平滑噪声,产生光滑的重建效果,而同时,对于强纹理区域,也可以有效的保留表面的边缘特征。相比于其它的数据集,人脸数据集包含大量的弱纹理区域。LSD-SLAM算法在估计相机位置的同时,也可以输出一组半稠密的点云数据,对一位志愿者的人脸重建如图8所示。这些点云数据依赖于表面的纹理信息,对于展示三维物体,特别是包含大量弱纹理区域的人脸,仍然过于稀疏且包含较多的噪声。本文使用LSD-SLAM算法估计相机姿态后执行基于随机初始化和深度传播的深度估计,产生稠密的深度图,然后从中选择高置信度的像素位置构成GCPs,并使用分段线性插值重新调整弱纹理区域的深度。可以看出,提出的深度调整算法可以有效的抑制由于弱纹理导致的噪声和空洞现象。对另外两个志愿者的人脸重建结果,分别如图9和10所示。

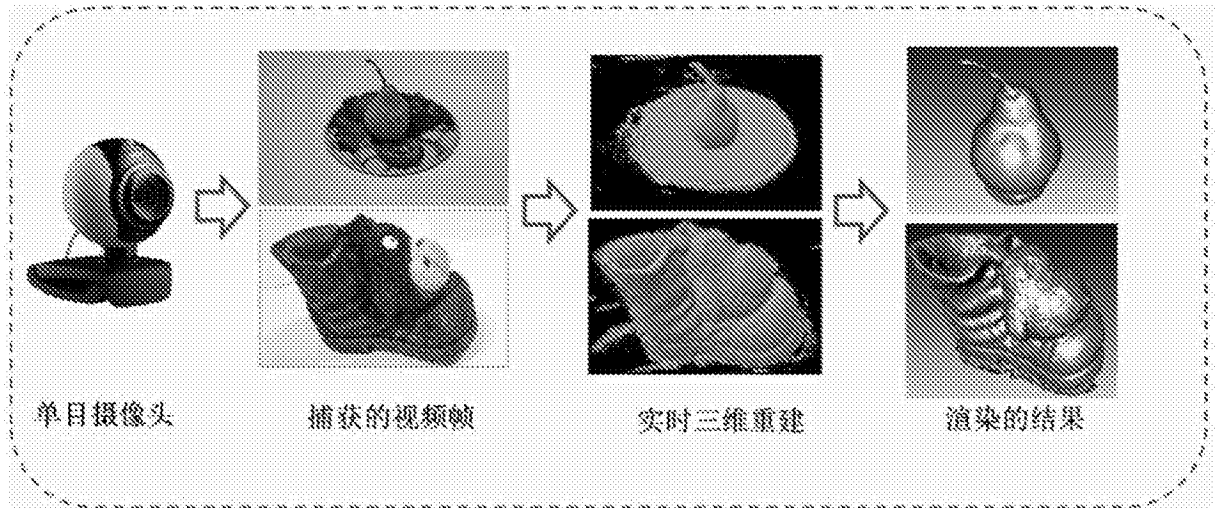


图1

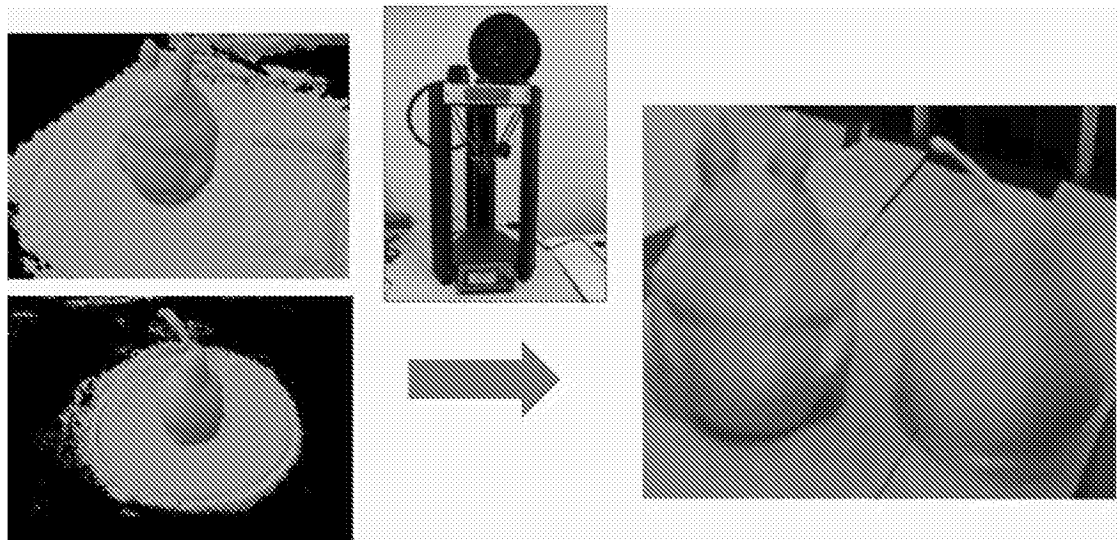


图2

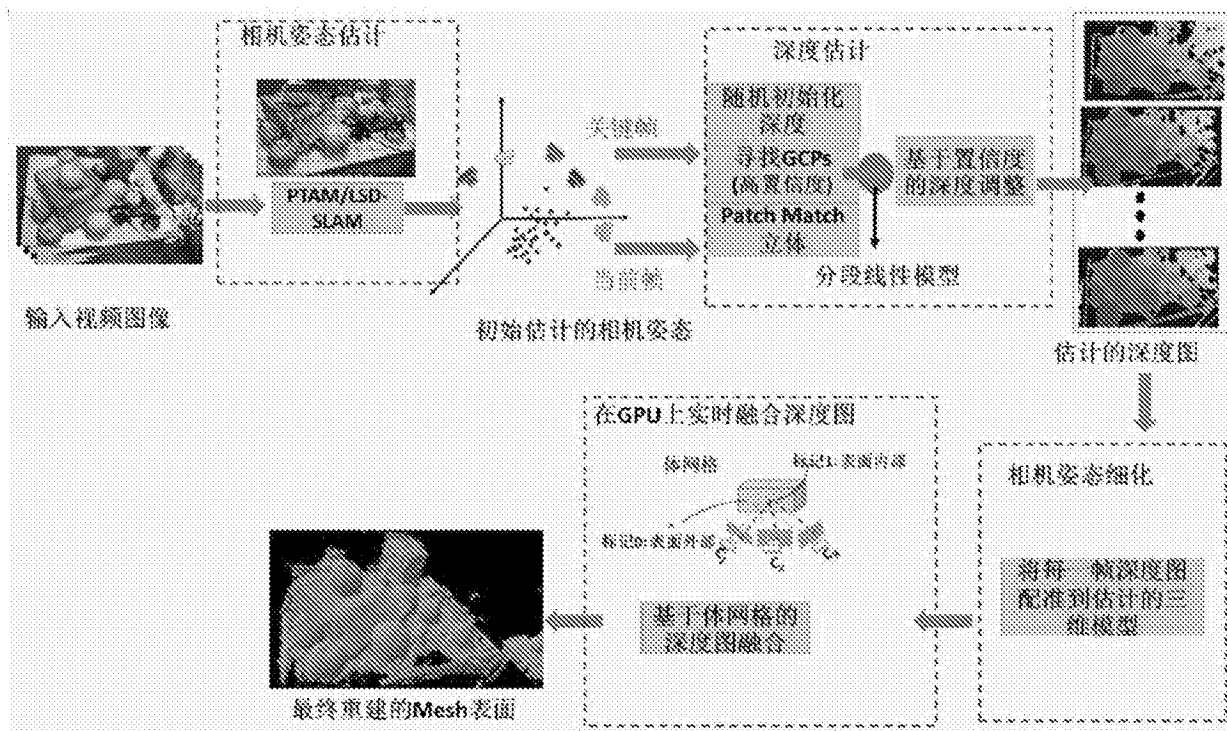


图3

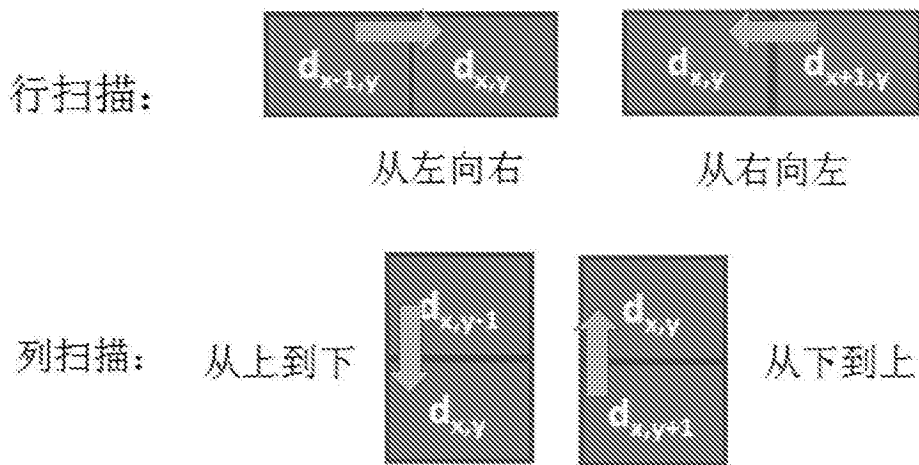


图4

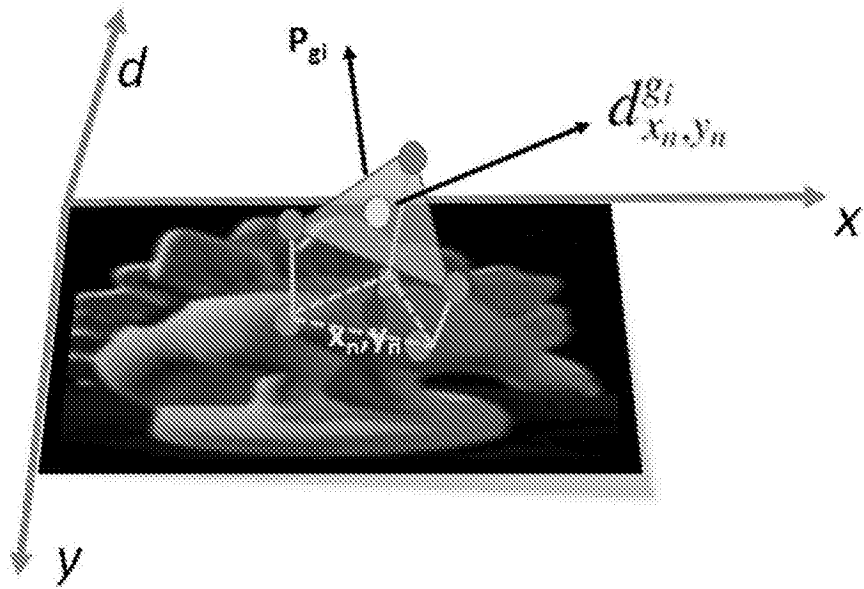


图5

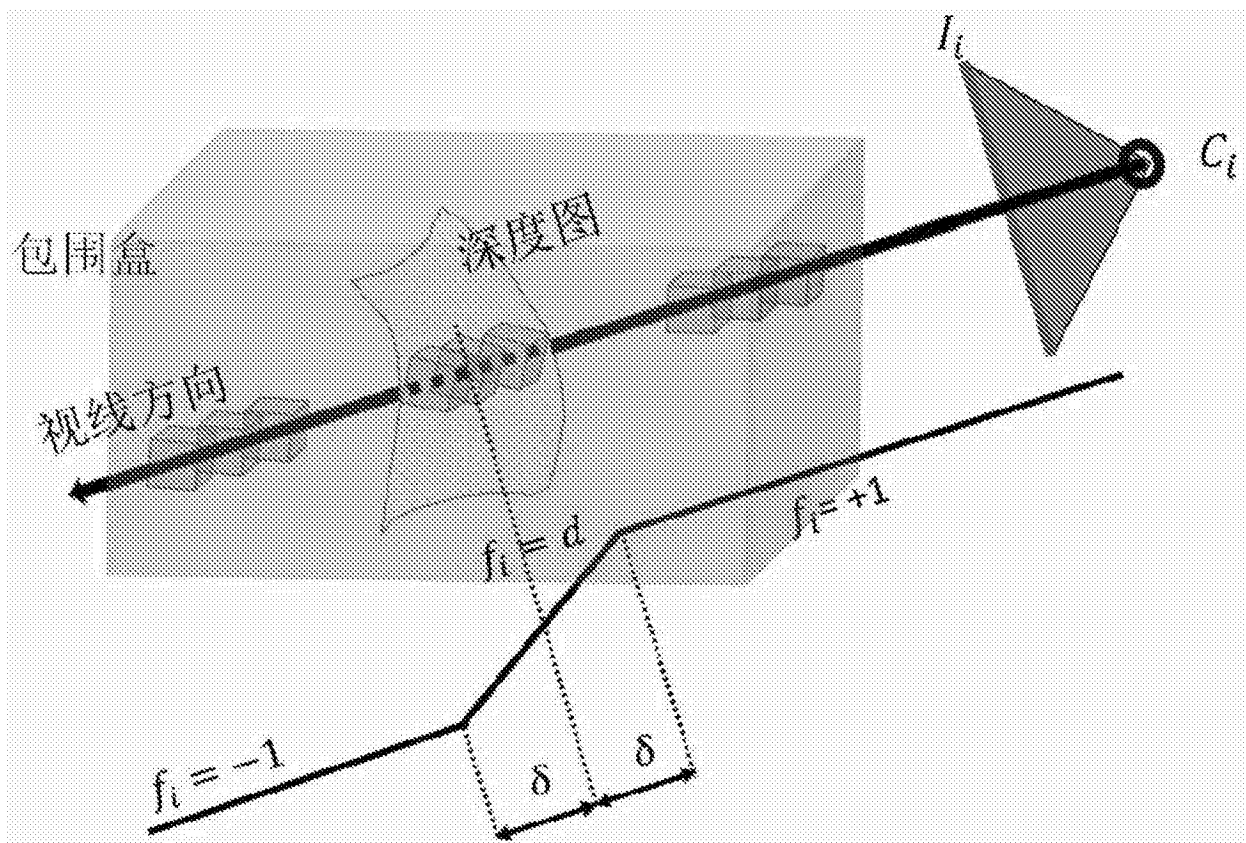


图6

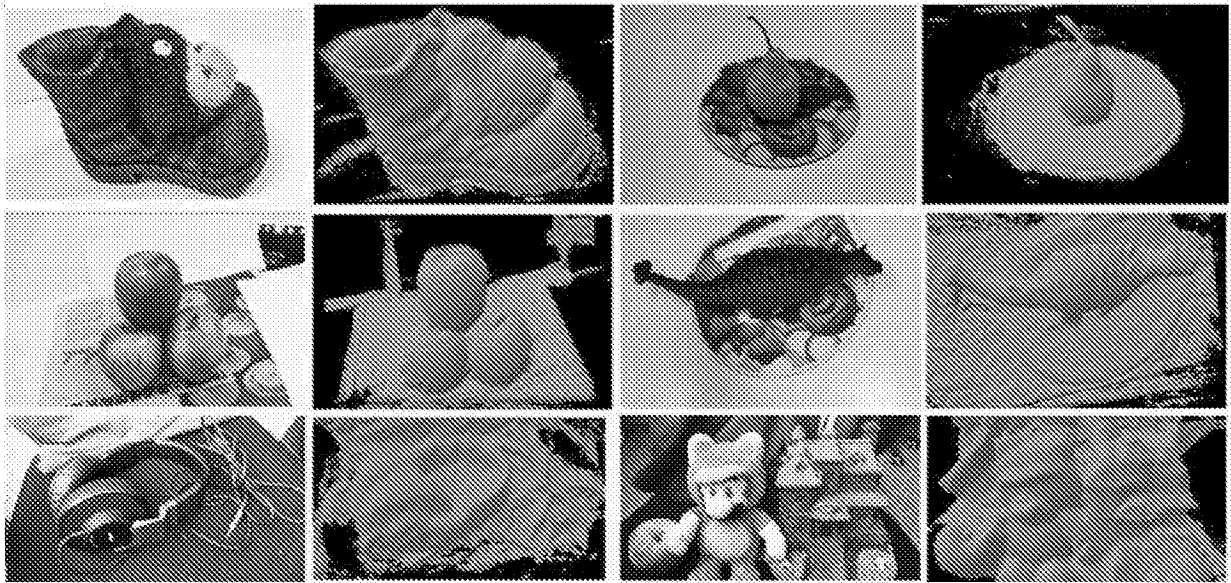


图7



图8

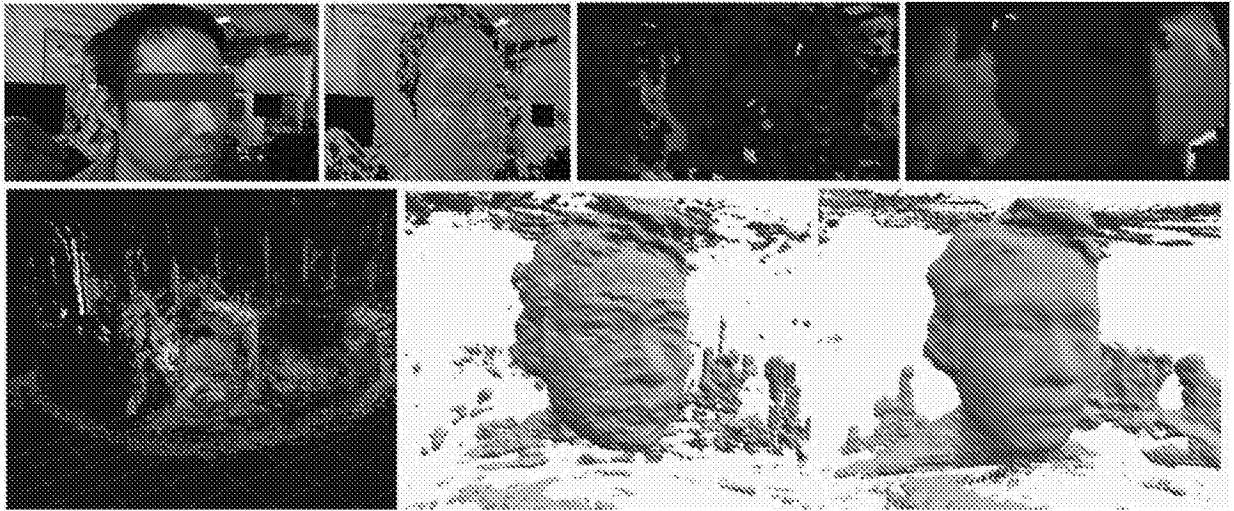


图9

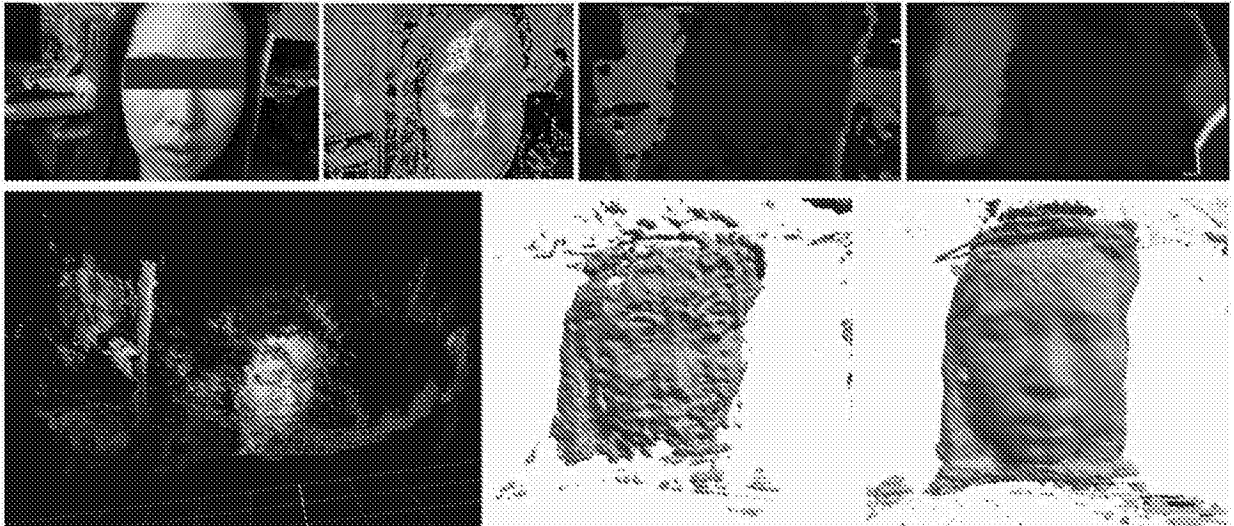


图10