



(12)发明专利申请

(10)申请公布号 CN 109035329 A

(43)申请公布日 2018.12.18

(21)申请号 201810878967.8

(22)申请日 2018.08.03

(71)申请人 厦门大学

地址 361005 福建省厦门市思明南路422号

(72)发明人 纪荣嵘 郭锋 陈晗

(74)专利代理机构 厦门南强之路专利事务所

(普通合伙) 35200

代理人 马应森

(51)Int.Cl.

G06T 7/70(2017.01)

G06T 7/80(2017.01)

G06T 7/90(2017.01)

权利要求书4页 说明书8页 附图1页

(54)发明名称

基于深度特征的相机姿态估计优化方法

(57)摘要

基于深度特征的相机姿态估计优化方法,涉及基于有监督学习SLAM系统的优化方法。使用基于随机森林的匹配算法,快速计算2D-3D点的相似性以映射2D-3D点信息;使用约束函数和多特征融合的方法评估相机姿态;针对基于深度学习的SLAM算法中存在的稳定性问题,提出一种多特征集束优化算法。使用三维重建数据作为参考,然后使用可见的3D点及来自随机森林的离线数据集的相关关键点映射,并使用多特征融合和约束函数来测量姿态评估分数。以上方法用于优化基于深度学习SLAM的性能。实验结果证明,算法鲁棒性。

1. 基于深度特征的相机姿态估计优化方法,其特征在于包括以下步骤:

- 1) 使用基于随机森林的匹配算法,快速计算2D-3D点的相似性以映射2D-3D点信息;
- 2) 使用约束函数和多特征融合的方法评估相机姿态;

3) 针对基于深度学习的SLAM算法中存在的不稳定性问题,提出一种多特征集束优化算法。

2. 如权利要求1所述基于深度特征的相机姿态估计优化方法,其特征在于在步骤1)中,所述使用基于随机森林的匹配算法,快速计算2D-3D点的相似性以映射2D-3D点信息的具体方法为:

每个决策树由内部节点和叶节点组成,决策树的预测能够计算出2D像素点之间的相似度,再由2D像素点之间的相似度推算到三维空间的相似度,从根节点开始一直到叶子,通过反复修改决策函数而使训练收敛,决策函数表达如下:

$$\text{split}(p; \delta_n) = [f_n(p) > \delta_n]$$

其中, $n$ 表示决策树中节点的索引, $p$ 是代表2D关键点的非叶节点, $[.]$ 是0~1指标函数, $\delta$ 是决策阈值, $f()$ 是决策函数:

$$f(p) = a_1 D_{\text{shape}}(p_1, p_2) + a_2 D_{\text{texture}}(p_1, p_2) + a_3 D_{\text{color}}(p_1, p_2)$$

在特征融合中定义 $a$ 和 $D()$ ,若 $\text{split}(p; \delta_n)$ 求值为0,则为训练路径分支到左边的子树,否则分支到右边, $p_1$ 和 $p_2$ 是关键点周围的像素对点,在训练的过程中,三维重建数据包含相关的相应3D点和2D点的映射;将先验划分为训练数据和验证数据;在算法1中展示训练框架;目标函数 $Q$ 用于使训练数据和验证数据具有相同的学习趋势,其中, $\Theta$ 是验证数据进入与相关训练数据不同路径的数量, $P_{\text{verification}}$ 是验证数据集, $\lambda$ 是折衷相同分支的相似度和不同分支的多样性的参数;目标函数是基于落入相同分支的验证数据和训练数据的比例去衡量点与点之间的相似性:

$$Q = \frac{\Theta}{|P_{\text{verification}}|} \left[ \frac{\lambda}{\sum_{i \in I, j \in J} \|f_i(p) - f_j(p)\|_2} + \sum_{i, j \in \{I, J\}} \|f_i(p) - f_j(p)\|_2 \right]$$

算法1基于随机森林的2D-3D点映射训练过程如表1所示:

表1

---

**Algorithm 1: 训练随机决策树**

---

**Input:** 随机采集包含训练和验证样本的数据集  $P$

**Output:** 得到训练好参数的随机决策树

---

- 1: 初始化随机决策函数  $S = \{f_i(P)\}_{i=1}^n$ ;
- 2: **while** 还剩下非叶子节点或者还剩下训练和验证数据时 **do**;
- 3: 应用每一个决策函数  $f_i(P)$  去决策训练数据和验证数据, 使其分到左子树或者右子树的集合中去;
- 4: 统计验证集的误差  $e_p$ , 如当验证集数据落入到和相对应的训练数据不同的分支时, 出现误差;
- 5: 计算目标函数  $Q$ ;
- 6: 选择使得目标函数最小的最佳决策函数,  $I_{\text{best}} = \operatorname{argmin}_{f \in S} (Q)$ ;
- 7: 接着, 根据最佳决策函数  $f_i(P)$  对剩下的数据进行划分;
- 8: **end while**;
- 9: 返回随机决策树。

3. 如权利要求1所述基于深度特征的相机姿态估计优化方法, 其特征在于在步骤2) 中, 所述使用约束函数和多特征融合的方法评估相机姿态的具体方法为:

假设姿态估计结果是相机外部参数  $\begin{bmatrix} R & -RC_w \\ 0 & 1 \end{bmatrix}$ , 由  $x, y, z$  和  $w, p, q, r$  转化得到, 假设内参矩阵  $K$  由EXIF标签获得; 然后, 结合内参:

$$K = \begin{bmatrix} fm_x & 0 & p_x & 0 \\ 0 & fm_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

通过图像坐标和世界坐标得到变换矩阵:

$$\begin{bmatrix} wx \\ wy \\ w \end{bmatrix}_{\text{image}} = KR \begin{bmatrix} I & -C_w \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\text{world}}$$

2D点和3D点云通过这个变换矩阵相关联; 提取一个已经评估过相机矩阵的评估图像的FALoG特征点作为查询数据集; 对于查询特征点  $m(x, y)$ , 通过上述映射算法找到查询特征的最近似特征, 搜索已建立的最近的数据集特征, 随机森林通过测试每个决策节点中的查询特征, 最后查询到达叶节点; 随机森林的最终映射特征  $m'$  在  $m$  点是节点概率最大的对应点, 对应点  $m'$  有其相关的3D点; 对于每个3D点, 数据库至少有一个相应的图像, 每个点投影到相关的图像, 结合像素差异作为误差函数, 并使用所述误差函数迭代优化姿态矩阵, 认为不同映射点的误差通过基于颜色的, 基于形状的和基于纹理的像素级别误差表示, 对于每个特征利用评估基于上述特征的偏差率最为的姿态的评分;

有关特征如下:

(1) 基于颜色的像素特征

颜色特征能够表示图像中物体的表面属性, 采用融合单像素和图像区域的融合像素特征方法最大程度地表达单个像素的颜色特征; 使用颜色距离函数来区分两个单个像素颜色变化:

$$d_{\text{point}}(p, p') = \sqrt{\frac{|R(x, y) - R(x', y)|}{256}} + \sqrt{\frac{|G(x, y) - G(x', y)|}{256}} + \sqrt{\frac{|B(x, y) - B(x', y)|}{256}}$$

其中,  $p(x, y)$  和  $p'(x, y)$  是两个目标点,  $R()$ ,  $G()$  和  $B()$  是各自的对应  $R, G,$  和  $B$  值; 对于

区域颜色特征,假设包含在一个区域中的目标点,假设有 $5 \times 5$ 个像素,对于每个RGB通道,区域的中心是目标点;提取目标点 $(x, y)$ 的水平和垂直方向上的梯度值 $G(x, y)$ :

$$G(x, y) = \ln\left(1 + \frac{d_h + d_v}{512}\right)$$

其中, $d_h$ 是左侧两个像素之间的差异目标点的平均值和右侧两个像素的平均值;中心值48水平和垂直区域的其他值是 $d_v$ ,通过 $d_h$ 同样的方式计算得到的;对于其余的四个 $2 \times 2$ 像素块,平均下采样,最后,基于颜色的像素差异性:

$$D_{\text{color}} = \left[ d_{\text{point}}(p, p') + \sum_{i=1}^n \delta d_{\text{region}}^i(p, p') \right] + \eta \|d_G(p, p')\|_2$$

其中, $\delta \in \{0, 1\}$ ,若两个块的下采样值的差异低于 $d_{\text{point}}(p, p')$ ,则 $\delta$ 设为0,否则为1; $d_{\text{region}}(p, p')$ 采用与 $d_{\text{point}}(p, p')$ 相同的计算方法, $d_G(p, p')$ 是梯度值的差值, $\eta$ 是正则因子;

(2) 基于形状和基于纹理的像素特征

作为 $32 \times 32$ 块的中心,提取形状上下文特征,并计算25个采样点的形状上下文,分配块的空间对数坐标分为 $12 \times 3 = 36$ 个部分,表达形状特征差别使用二进制化表示变换两点形状多样性, $\sum_{i=1}^n \text{item}_i$ ,其中, $n = 36$ , $\text{item} \in \{0, 1\}$ , $\text{item} = 1$ 意味着 $i$ 块的列值大于平均值,对于纹理,一个半径为 $R$ 的圆的中心点被分成八个相等的角度区域;计算每个区域的平均图像强度,若平均值大于中心像素值,则该区域的值设置为1,否则,其设置为0;八个二进制序列被转换以十进制数表示纹理特征;最后, $D_{\text{texture}}$ 由海明距离确定;

最后,评估函数 $E$ 是上述特征的组合:

$$E(p_i, p_j) = a_1 D_{\text{shape}}(p_i, p_j) + a_2 D_{\text{texture}}(p_i, p_j) + a_3 \|D_{\text{color}}(p_i, p_j)\|_2$$

其中, $a_i \in (0, 1)$ 表示不同项的权重, $D_{\text{color}}$ 被视为一个这个融合多样性函数中的正则项, $D_{\text{shape}}$ 和 $D_{\text{texture}}$ 被连接起来,通过 $k$ 次迭代训练使得这个函数中的参数能够评估的最好,评估函数描述了姿态评估错误用查询二维特征与先验数据集特征在特征点上的重投影点误差来衡量。

4. 如权利要求1所述基于深度特征的相机姿态估计优化方法,其特征在于在步骤3)中,所述针对基于深度学习的SLAM算法中存在的不稳定性问题,提出一种多特征集束优化算法的具体方法如下:

基于先验知识的深度学习SLAM系统通过打开一个新线程来存储2D-3D映射和关键帧,选择那些满足 $E(p_i, p_j) < \hat{\sigma}$ 的关键帧作为集束优化(bundle adjustment,简称BA)的关键帧集合;局部BA优化所有被关键帧看到的点;观测点有助于约束函数和最终姿势的优化,全局BA优化和局部BA优化类似,只不过系统将在30帧之间执行局部BA,300帧之间执行全局BA;

表2

场景	错误率（错误估计帧（阈值为室内：5cm5度，室外：2m10度）/总测试帧）（%）						
	特征融合	最近邻	Score Forest 算法	Posenet 算法	ORB-SLAM 算法	本发明（无集束优化）	本发明（有集束优化）
棋盘	72.6	81.2	91.8	93.2	95.4	95.0	96.2
火炉	65.7	69.6	81.6	81.6	90.3	90.1	90.6
桌头	41.7	47.7	53.7	63.0	65.5	63.5	64.9
办公室	63.6	72.3	80.2	86.6	88.9	88.3	89.5
颂恩楼	57.3	64.3	N/A	83.3	87.4	87.0	88.4
玛丽教堂	56.7	62.3	N/A	82.1	85.7	85.1	85.5
店铺	63.4	67.3	N/A	87.2	88.8	88.1	89.7
国王学院	64.5	66.3	N/A	85.6	87.3	87.1	88.0

使用优化算法的SLAM系统后与其他先进SLAM系统的比较如表2所示。

## 基于深度特征的相机姿态估计优化方法

### 技术领域

[0001] 本发明涉及基于有监督学习SLAM系统的优化方法,尤其是涉及用于有监督学习SLAM算法的基于深度特征的相机姿态估计优化方法。

### 背景技术

[0002] SLAM技术在机器人、自动驾驶、虚拟与增强现实领域有着很好的应用前景,在众多的计算机视觉与人工智能技术中,SLAM的研究持续火热。近几年,越来越多的机器人出现在人们的视野当中,给人们生活带来很多便利,通过自身的摄像头、陀螺仪、激光传感器等来得到具体场景的环境,并对自身进行定位,在满足实时性的条件下完成特定任务。近年来,国内外多家公司投入大量人力、物力来进行无人驾驶车辆的研究。无人驾驶的核心技术也是SLAM技术,鲁棒而快速的环境识别和语义分割是该无人驾驶的关键所在。在增强现实领域,目前在市面上投入到商业场景中的AR应用大多是基于特定模板,从模板识别到模板跟踪匹配,结合三维注册和模型渲染来进行虚实交互。而真正的增强现实需要对应用场景所处的环境进行识别与语义理解,这时候仍需SLAM技术来作为核心技术。

[0003] 相机重新定位一直是SLAM中的一项重要任务。基于图像的重新定位方法在SLAM中是一个强大而有效的线程,基于成像的相机姿态估计的常用技术是图像检索和基于3D场景重建的方法,但是,图像检索错误大于GPS位置传感器。此外,估计精度也取决于数据集,AR Zamir等人([1]Zamir,A.R.,Shah,M.:Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs.IEEE Trans.Pattern Anal.Mach.Intell.(2014))提出多个最近邻的地理定位特征匹配框架,但如果查询图像(Query)和数据集(Database)不匹配那么此方法受到严重限制。这种方法为了实现更好的定位结果,需要三维先验信息提供有价值的空间关系。但是,现有的基于三维先验的图像定位方法只关注局部的定位精度,而忽视后续的优化。对于局部优化,如FastSLAM([2]Parsons,S.:Probabilistic robotics by Sebastian Thrun,Wolfram Burgard and Dieter Fox.Knowledge Eng.Review(2006).<https://doi.org/10.1017/S0269888906210993>),通过传感器估计噪声来测量估计姿态误差,并对其优化。ORB-SLAM([3]Mur-Artal,R.,Montiel,J.M.M.,Tard'os,J.D.:ORB-SLAM:a versatile and accurate monocular SLAM System.IEEE Trans.Robot.(2015))是一种优秀视觉SLAM系统,使用orb特征和局部及全局优化,但是它没有使用深度特征和先验知识。([4]Kendall,A.,Cipolla,R.:Modelling uncertainty in deep learning for camera relocalization.In:IEEE International Conference on Robotics and Automation, ICRA 2016,Stockholm,Sweden,16-21May 2016.<https://doi.org/10.1109/ICRA.2016.7487679>)提出了一个使用贝叶斯的视觉重定位系统,并用卷积神经网络来回归相机姿态,重用了三维信息并接近实时性能。我们提出的方法不利用任何传感器,而是基于先验三维点的相似度定位模型,但是,目前基于有监督学习的方法都没有相当适用的优化方法。虽然一些基于深卷积神经网络的视觉定位算法被认为是容忍大基线的端到端定位方

法,但是,估计姿态时往往出现平均误差较大的情况。这一切归结为没有后端优化(例如局部集束优化),基于上述问题,展示了如何使用随机森林(Cutler, A., Cutler, D.R., Stevens, J.R.: Random forests. In: Machine Learning (2012))做2D-3D的点匹配,以及如何改进及利用约束函数优化基于深度学习的SLAM系统。

### 发明内容

[0004] 本发明的目的在于提供基于深度特征的相机姿态估计优化方法。

[0005] 本发明包括以下步骤:

[0006] 1) 使用基于随机森林的匹配算法,快速计算2D-3D点的相似性以映射2D-3D点信息;

[0007] 2) 使用约束函数和多特征融合的方法评估相机姿态;

[0008] 3) 针对基于深度学习的SLAM算法中存在的稳定性问题,提出一种多特征集束优化算法。

[0009] 在步骤1)中,所述使用基于随机森林的匹配算法,快速计算2D-3D点的相似性以映射2D-3D点信息的具体方法可为:

[0010] 每个决策树由内部节点和叶节点组成,决策树的预测能够计算出2D像素点之间的相似度,再由2D像素点之间的相似度推算到三维空间的相似度,从根节点开始一直到叶子,通过反复修改决策函数而使训练收敛,决策函数表达如下:

[0011]  $\text{split}(p; \delta_n) = [f_n(p) > \delta_n]$

[0012] 其中, $n$ 表示决策树中节点的索引, $p$ 是代表2D关键点的非叶节点, $[.]$ 是0~1指标函数, $\delta$ 是决策阈值, $f()$ 是决策函数:

[0013]  $f(p) = a_1 D_{\text{shape}}(p_1, p_2) + a_2 D_{\text{texture}}(p_1, p_2) + a_3 D_{\text{color}}(p_1, p_2)$

[0014] 在特征融合中定义 $a$ 和 $D()$ ,若 $\text{split}(p; \delta_n)$ 求值为0,则为训练路径分支到左边的子树,否则分支到右边, $p_1$ 和 $p_2$ 是关键点周围的像素对点,在训练的过程中,三维重建数据包含相关的相应3D点和2D点的映射;将先验划分为训练数据和验证数据;在算法1中展示训练框架;目标函数 $Q$ 用于使训练数据和验证数据具有相同的学习趋势,其中, $\Theta$ 是验证数据进入与相关训练数据不同路径的数量, $P_{\text{verification}}$ 是验证数据集, $\lambda$ 是折衷相同分支的相似度和不同分支的多样性的参数;目标函数是基于落入相同分支的验证数据和训练数据的比例去衡量点与点之间的相似性:

[0015] 
$$Q = \frac{\Theta}{|P_{\text{verification}}|} \left[ \frac{\lambda}{\sum_{i \in I, j \in I, i \neq j} \|f_i(p) - f_j(p)\|_2} + \sum_{i, j \in \{L, R\}} \|f_i(p) - f_j(p)\|_2 \right]$$

[0016] 算法1基于随机森林的2D-3D点映射训练过程如表1所示:

[0017] 表1

[0018]

**Algorithm 1: 训练随机决策树****Input:** 随机采集包含训练和验证样本的数据集  $P$ **Output:** 得到训练好参数的随机决策树1: 初始化随机决策函数  $S = \{f_i(P)\}_{i=1}^n$ ;2: **while** 还剩下非叶子节点或者还剩下训练和验证数据时 **do**;3: 应用每一个决策函数  $f_i(P)$  去决策训练数据和验证数据, 使其分到左子树或者右子树的集合中去;4: 统计验证集的误差  $e_p$ , 如当验证集数据落入到和相对应的训练数据不同的分支时, 出现误差;5: 计算目标函数  $Q$ ;6: 选择使得目标函数最小的最佳决策函数,  $I_{\text{best}} = \operatorname{argmin}_{f \in S} (Q)$ ;7: 接着, 根据最佳决策函数  $f_i(P)$  对剩下的数据进行划分;8: **end while**;

9: 返回随机决策树。

[0019] 在步骤2)中,所述使用约束函数和多特征融合的方法评估相机姿态的具体方法可为:

[0020] 假设姿态估计结果是相机外部参数  $\begin{bmatrix} R & -RC_w \\ 0 & 1 \end{bmatrix}$ , 由  $x, y, z$  (位置信息) 和  $w, p, q, r$  (四元素) 转化得到, 假设内参矩阵  $K$  由 EXIF 标签获得 (假设没有径向变形); 然后, 结合内参:

$$[0021] \quad K = \begin{bmatrix} fm_x & 0 & p_x & 0 \\ 0 & fm_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

[0022] 可以通过图像坐标和世界坐标得到变换矩阵:

$$[0023] \quad \begin{bmatrix} wx \\ wy \\ w \end{bmatrix}_{\text{image}} = KR \begin{bmatrix} I & -C_w \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\text{world}}$$

[0024] 2D点和3D点云通过这个变换矩阵相关联; 如提取一个已经评估过相机矩阵的评估图像的FALoG特征点作为查询数据集。对于查询特征点  $m(x, y)$ , 可以通过上述映射算法找到查询特征的最近似特征, 搜索已建立的最近的数据集特征, 随机森林通过测试每个决策节点中的查询特征, 最后查询到达叶节点; 随机森林的最终映射特征  $m'$  在  $m$  点是节点概率最大的对应点, 对应点  $m'$  有其相关的3D点。对于每个3D点, 数据库至少有一个相应的图像, 每个点可以投影到相关的图像, 结合像素差异作为误差函数, 并使用所述误差函数迭代优化姿态矩阵, 认为不同映射点的误差可以通过基于颜色的, 基于形状的和基于纹理的像素级别误差表示, 对于每个特征利用评估基于上述特征的偏差率最为的姿态的评分;

[0025] 有关特征如下:

[0026] (1) 基于颜色的像素特征。

[0027] 颜色特征能够表示图像中物体的表面属性, 采用融合单像素和图像区域的融合像素特征方法可以最大程度地表达单个像素的颜色特征; 使用颜色距离函数来区分两个单个像素颜色变化:



$$[0028] \quad d_{\text{point}}(p, p') = \sqrt{\frac{|R(x, y) - R(x', y)|}{256}} + \sqrt{\frac{|G(x, y) - G(x', y)|}{256}} + \sqrt{\frac{|B(x, y) - B(x', y)|}{256}}$$

[0029] 其中,  $p(x, y)$  和  $p'(x, y)$  是两个目标点,  $R()$ ,  $G()$  和  $B()$  是各自的对应R, G, 和B值; 对于区域颜色特征, 假设包含在一个区域中的目标点, 假设有  $5 \times 5$  个像素, 对于每个RGB通道, 区域的中心是目标点; 提取目标点  $(x, y)$  的水平和垂直方向上的梯度值  $G(x, y)$ :

$$[0030] \quad G(x, y) = \ln\left(1 + \frac{d_h + d_v}{512}\right)$$

[0031] 其中,  $d_h$  是左侧两个像素之间的差异目标点的平均值和右侧两个像素的平均值; 中心值48水平和垂直区域的其他值是  $d_v$ , 通过  $d_h$  同样的方式计算得到的。对于其余的四个  $2 \times 2$  像素块, 平均下采样, 最后, 基于颜色的像素差异性:

$$[0032] \quad D_{\text{color}} = \left[ d_{\text{point}}(p, p') + \sum_{i=1}^n \delta d_{\text{region}}^i(p, p') \right] + \eta \|d_G(p, p')\|_2$$

[0033] 其中,  $\delta \in \{0, 1\}$ , 若两个块的下采样值的差异低于  $d_{\text{point}}(p, p')$ , 则  $\delta$  设为0, 否则为1;  $d_{\text{region}}(p, p')$  采用与  $d_{\text{point}}(p, p')$  相同的计算方法,  $d_G(p, p')$  是梯度值的差值,  $\eta$  是正则因子;

[0034] (2) 基于形状和基于纹理的像素特征

[0035] 作为  $32 \times 32$  块的中心, 提取形状上下文特征, 并计算25个采样点的形状上下文, 分配块的空间对数坐标分为  $12 \times 3 = 36$  个部分, 表达形状特征差别使用二进制化表示变换两点形状多样性,  $\sum_{i=1}^n \text{item}_i$ , 其中,  $n = 36$ ,  $\text{item}_i \in \{0, 1\}$ ,  $\text{item}_i = 1$  意味着  $i$  块的列值大于平均值, 对于纹理, 一个半径为  $R$  的圆的中心点被分成八个相等的角度区域; 计算每个区域的平均图像强度, 若平均值大于中心像素值, 则该区域的值设置为1, 否则, 其设置为0; 八个二进制序列被转换以十进制数表示纹理特征; 最后,  $D_{\text{texture}}$  由海明距离确定;

[0036] 最后, 评估函数  $E$  是上述特征的组合:

$$[0037] \quad E(p_i, p_j) = a_1 D_{\text{shape}}(p_i, p_j) + a_2 D_{\text{texture}}(p_i, p_j) + a_3 \|D_{\text{color}}(p_i, p_j)\|_2$$

[0038] 其中,  $a_i \in (0, 1)$  表示不同项的权重,  $D_{\text{color}}$  被视为一个这个融合多样性函数中的正则项,  $D_{\text{shape}}$  和  $D_{\text{texture}}$  被连接起来, 通过  $k$  次迭代训练使得这个函数中的参数能够评估的最好, 评估函数描述了姿态评估错误可以用查询二维特征与先验数据集特征在特征点上的重投影点误差来衡量。

[0039] 在步骤3)中, 所述针对基于深度学习的SLAM算法中存在的不稳定性问题, 提出一种多特征集束优化算法的具体方法如下:

[0040] 基于先验知识的深度学习SLAM系统通过打开一个新线程来存储2D-3D映射和关键帧, 选择那些满足  $E(p_i, p_j) < \delta$  的关键帧作为集束优化 (bundle adjustment, 简称BA) 的关键帧集合; 局部BA优化所有可以被关键帧看到的点; 观测点有助于约束函数和最终姿势的优化, 全局BA优化和局部BA优化类似, 只不过系统将在30帧之间执行局部BA, 300帧之间执行全局BA; 使用优化算法的SLAM系统后与其他先进SLAM系统的比较如表2所示。

[0041] 表2

[0042]

场景	错误率（错误估计帧（阈值为室内：5cm5度，室外：2m10度）/总测试帧）（%）						
	特征融合	最近邻	Score Forest 算法	Posenet 算法	ORB-SLAM 算法	本发明（无集束优化）	本发明（有集束优化）
棋盘	72.6	81.2	91.8	93.2	95.4	95.0	96.2
火炉	65.7	69.6	81.6	81.6	90.3	90.1	90.6
桌头	41.7	47.7	53.7	63.0	65.5	63.5	64.9
办公室	63.6	72.3	80.2	86.6	88.9	88.3	89.5
颂恩楼	57.3	64.3	N/A	83.3	87.4	87.0	88.4
玛丽教堂	56.7	62.3	N/A	82.1	85.7	85.1	85.5
店铺	63.4	67.3	N/A	87.2	88.8	88.1	89.7
国王学院	64.5	66.3	N/A	85.6	87.3	87.1	88.0

[0043] 本发明结合基于深度学习的SLAM技术、基于随机森林的二维图像点与三维点云的映射,设计了一个基于深度学习SLAM系统优化算法,很好地解决了深度学习SLAM系统优化算法的空白。该系统集低计算量的SLAM环境构建、图像2D点与3D点云匹配于一体,在PC端和移动端都可以对场景进行实时重建和优化,保持了比较高的重建精度,对机器人、无人驾驶、增强现实等领域都有着重要的实用价值和意义。

[0044] 大多数现有的定位方法仅基于参考点距离近似姿态置信度。与其他的方法不同,本发明使用三维重建数据作为参考,然后使用可见的3D点及来自随机森林的离线数据集的相关关键点映射,并使用多特征融合和约束函数来测量姿态评估分数。以上方法用于优化基于深度学习SLAM的性能。实验结果证明,本发明的算法鲁棒性。

## 附图说明

[0045] 图1为2D-3D点匹配的系统概述图。

[0046] 图2为基于颜色特征的5×5像素区域的图示。

[0047] 图3为基于形状的特征的图示。

## 具体实施方式

[0048] 以下实施例将结合附图对本发明作进一步的说明。

[0049] 一、基本概念

[0050] 1) 基于随机森林的2D-3D点映射

[0051] 构造随机森林的目的在于找到查询图像和三维重建先验数据之间的关系。能够通过该映射架构评估不同的定位算法的姿态。由于随机森林算法有着较低复杂性和较高鲁棒的优点,使得映射算法满足实时有效的评估。实质上,决策树只是一个从2D到3D检索三角几何关系的映射工具,能够用其他方法替代。

[0052] 决策树的训练是映射算法的关键。随机森林的性能是由各个不同的决策树的整合

性能来决定的。每个决策树由内部节点和叶节点组成。决策树的预测能够计算出2D像素点之间的相似度,再由2D像素点之间的相似度推算到三维空间的相似度。从根节点开始一直到叶子,通过反复修改决策函数而使训练收敛,决策函数表达如下:

$$[0053] \quad \text{split}(p; \delta_n) = [f_n(p) > \delta_n]$$

[0054] 其中, $n$ 表示决策树中节点的索引, $p$ 是代表2D关键点的非叶节点, $[.]$ 是0-1指标函数, $\delta$ 是决策阈值, $f()$ 是决策函数:

$$[0055] \quad f(p) = a_1 D_{\text{shape}}(p_1, p_2) + a_2 D_{\text{texture}}(p_1, p_2) + a_3 D_{\text{color}}(p_1, p_2)$$

[0056] 在特征融合中定义 $a$ 和 $D()$ 。若 $\text{split}(p; \delta_n)$ 求值为0,则为训练路径分支到左边的子树,否则分支到右边。 $p_1$ 和 $p_2$ 是关键点周围的像素对点。在训练的过程中,三维重建数据(先验数据)包含相关的相应3D点和2D点的映射。将先验划分为训练数据和验证数据。在算法1中展示了训练框架。要注意的是使用了快速的FALoG特征点(Wang, Z., Fan, B., Wu, F.: FRIF: fast robust invariant feature. In: British Machine Vision Conference, BMVC 2013, Bristol, UK, 9-13 September 2013. <https://doi.org/10.5244/C.27.16>)能够快速检测出相应的二值特征点。目标函数 $Q$ 用于使训练数据和验证数据尽可能具有相同的学习趋势。其中 $\Theta$ 是验证数据进入与相关训练数据不同路径的数量, $P_{\text{verification}}$ 是验证数据集。 $\lambda$ 是折衷相同分支的相似度和不同分支的多样性的参数。目标函数是基于落入相同分支的验证数据和训练数据的比例去衡量点与点之间的相似性。

$$[0057] \quad Q = \frac{\Theta}{P_{\text{verification}}} \left[ \frac{\lambda}{\sum_{i \in I, j \in I} \|f_i(p) - f_j(p)\|_2} + \sum_{i, j \in I} \|f_i(p) - f_j(p)\|_2 \right]$$

[0058] 算法1基于随机森林的2D-3D点映射训练过程如表1所示。

[0059] 2) 特征融合算法

[0060] 虽然2D与3D的关系很容易通过上述映射获得。由于没有图像姿态标签,要判断相机姿态的好坏是一个很难进行的度量。设计出一套相机姿态评估算法(评估相机姿态的分数),这种算法不需要进行人工的标注,只需要通过迁移学习获得的三维重建数据。要预测的相机姿态可以使通过任何相机姿态预测算法来计算的。假设姿态估计结果是相机外部参数 $\begin{bmatrix} R & -RC_w \\ 0 & 1 \end{bmatrix}$ ,由 $x, y, z$ (位置信息)和 $w, p, q, r$ (四元素)转化得到。假设内参矩阵 $K$ 由EXIF标签获得(假设没有径向变形)。然后,结合内参:

$$[0061] \quad K = \begin{bmatrix} fm_x & 0 & p_x & 0 \\ 0 & fm_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

[0062] 可以通过图像坐标和世界坐标得到变换矩阵:

$$[0063] \quad \begin{bmatrix} wx \\ wy \\ w \end{bmatrix}_{\text{image}} = KR \begin{bmatrix} I & -C_w \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\text{world}}$$

[0064] 2D点和3D点云通过这个变换矩阵相关联。如提取一个已经评估过相机矩阵的评估图像的FALoG特征点作为查询数据集。对于查询特征点 $m(x, y)$ ,可以通过上述映射算法找到

查询特征的最近似特征。如图1所示,搜索已建立的最近的数据集特征,随机森林通过测试每个决策节点中的查询特征,最后查询到达叶节点。随机森林的最终映射特征m'在m点是节点概率最大的对应点。对应点m'有其相关的3D点。对于每个3D点,数据库至少有一个相应的图像,每个点可以投影到相关的图像。结合像素差异来作为误差函数,并使用这个误差函数迭代优化姿态矩阵。认为不同映射点的误差可以通过基于颜色的,基于形状的和基于纹理的像素级别误差表示。对于每个特征,我们在下面进行介绍。最后,利用评估基于上述特征的偏差率最为的姿态的评分。

[0065] 1. 基于颜色的像素特征。

[0066] 颜色特征能够表示了图像中物体的表面属性。在本发明中,采用融合单像素和图像区域的融合像素特征方法可以最大程度地表达单个像素的颜色特征。使用颜色距离函数来区分两个单个像素颜色变化:

[0067]

$$d_{point}(p, p') = \sqrt{\frac{|R(x, y) - R(x', y')|}{256}} + \sqrt{\frac{|G(x, y) - G(x', y')|}{256}} + \sqrt{\frac{|B(x, y) - B(x', y')|}{256}}$$

[0068] 其中p(x, y)和p'(x, y)是两个目标点,R(),G()和B()是各自的对应R,G,和B值。对于区域颜色特征,假设包含在一个区域中的目标点,假设有5×5个像素,如图2所示,对于每个RGB通道,区域的中心是目标点。提取目标点(x, y)的水平 and 垂直方向上的梯度值G(x, y)

$$[0069] \quad G(x, y) = \ln\left(1 + \frac{d_h + d_v}{512}\right)$$

[0070] 其中d<sub>h</sub>是左侧两个像素之间的差异目标点的平均值和右侧两个像素的平均值。在图2中心值48水平和垂直区域的其他值是d<sub>v</sub>,是通过图2中d<sub>h</sub>同样的方式计算得到的。对于其余的四个2×2像素块(图2中除中心点垂直和水平区域之外的点),平均他们的下采样。最后,基于颜色的像素差异性:

$$[0071] \quad D_{color} = \left[ d_{point}(p, p') + \sum_{i=1}^n \delta d_{region}^i(p, p') \right] + \eta \|d_G(p, p')\|_2$$

[0072] 其中,δ∈{0,1},若两个块的下采样值的差异低于d<sub>point</sub>(p, p'),则δ设为0,否则为1。d<sub>region</sub>(p, p')采用与d<sub>point</sub>(p, p')相同的计算方法。d<sub>G</sub>(p, p')是梯度值的差值,η是正则

[0073] 因子。

[0074] 2. 基于形状和基于纹理的像素特征。

[0075] 为了更快,更高效地提取像素的形状特征点,作为32×32块的中心,提取形状上下文特征,并计算25个采样点的形状上下文。如图3所示,分配块的空间对数坐标分为12×3=36个部分。直方图用于表示形状特征向量。为了表达形状特征差别,使用二进制化来表示变换两点形状直方图的多样性,∑<sub>i=1</sub><sup>n</sup> item<sub>i</sub>,其中n=36, item∈{0,1}, item=1意味着i块的列值大于直方图的平均值。对于纹理,一个半径为R的圆的中心点被分成八个相等的角度区域。计算每个区域的平均图像强度,若平均值大于中心像素值,则该区域的值设置为1,否则,其设置为0。八个二进制序列被转换以十进制数表示纹理特征。最后,D<sub>texture</sub>由海明距离确定。

[0076] 最后,评估函数E是上述特征的组合:

$$[0077] \quad E(p_i, p_j) = a_1 D_{\text{shape}}(p_i, p_j) + a_2 D_{\text{texture}}(p_i, p_j) + a_3 \| D_{\text{color}}(p_i, p_j) \|_2$$

[0078] 其中,  $a_i \in (0, 1)$  表示不同项的权重。 $D_{\text{color}}$  被视为一个这个融合多样性函数中的正则项。 $D_{\text{shape}}$  和  $D_{\text{texture}}$  被连接起来。通过  $k$  次迭代训练使得这个函数中的参数能够评估的最好。评估函数描述了姿态评估错误可以用查询二维特征与先验数据集特征在特征点上的重投影点误差来衡量。

[0079] 3) 基于深度SLAM的后端优化

[0080] 局部定位错误将反映在不同关键帧之间的姿态预测和三维重建中。关键帧姿态估计错误的积累导致了全局性错误增长,这样,整个系统的精度会受到限制。基于先验知识的深度学习SLAM系统通过打开一个新线程来存储2D-3D映射和关键帧。选择那些满足  $E(p_i, p_j) < \hat{\sigma}$  的关键帧作为集束优化(bundle adjustment, 简称BA)的关键帧集合。局部BA优化所有可以被关键帧看到的点。观测点有助于约束函数和最终姿势的优化。全局BA优化和局部BA优化类似,只不过系统将在30帧之间执行局部BA,300帧之间执行全局BA。

[0081] 使用本发明的优化算法的SLAM系统后与其他先进SLAM系统的比较如表2所示。

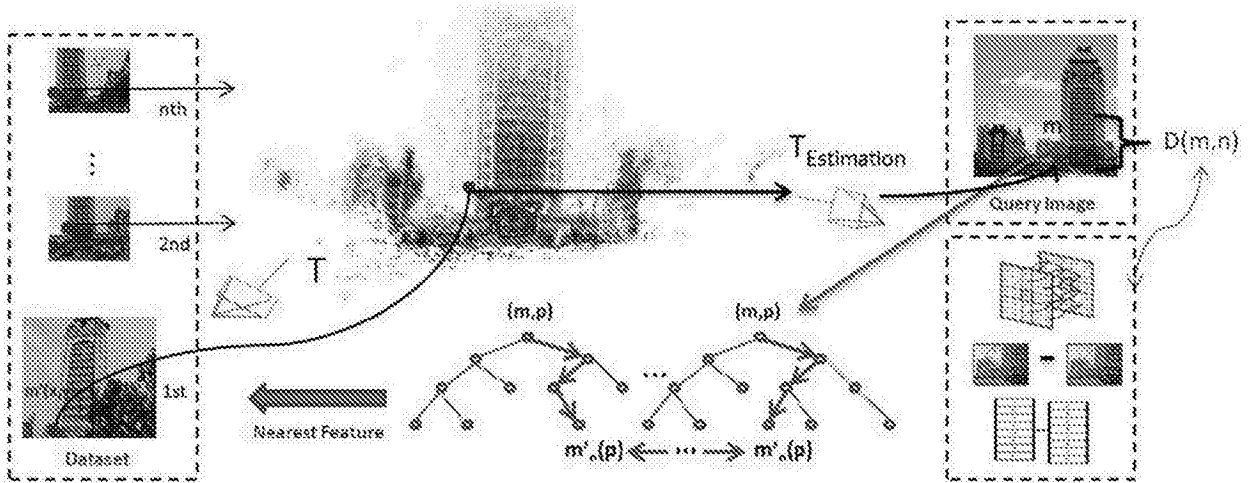


图1

				136	222	63	45	203
236	212	16	45	36	230	23	133	145
45	232	23	121	212	13	68	126	133
63	12	48	122	123	54	5	56	200
23	54	5	42	111	65	79	132	190
45	54	87	136	213				

图2

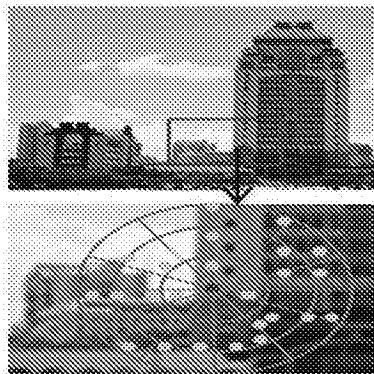


图3