(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104809187 B
(45)授权公告日 2017.11.21

(21)申请号 201510187615.4

(22)申请日 2015.04.20

(65)同一申请的已公布的文献号
申请公布号 CN 104809187 A

(43)申请公布日 2015.07.29

(73)专利权人 南京邮电大学
地址 210003 江苏省南京市新模范马路66号

(72)发明人 冯希龙 刘天亮

(74)专利代理机构 南京经纬专利商标代理有限公司 32200
代理人 许方

(51)Int.Cl.
G06F 17/30(2006.01)
G06K 9/62(2006.01)

(56)对比文件
CN 102542302 A,2012.07.04,
CN 102867192 A,2013.01.09,
CN 104077352 A,2014.10.01,
CN 104392228 A,2015.03.04,
US 2010/0040300 A1,2010.02.18,
CN 102436583 A,2012.05.02,
US 2011/0285910 A1,2011.11.24,
Xiaofeng Ren 等.RGB-(D) Scene Labeling: Features and Algorithms.《2012 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)》.2012,2759-2766.

(续)

审查员 李梦颖

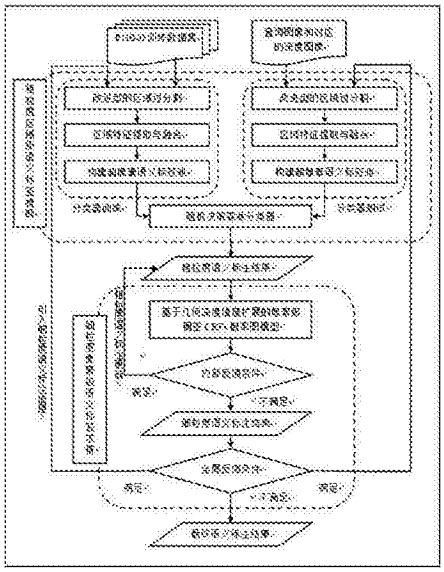
权利要求书4页 说明书14页 附图1页

(54)发明名称

一种基于RGB-D数据的室内场景语义标注方法

(57)摘要

本发明涉及一种基于RGB-D数据的室内场景语义标注方法，构建了一种基于RGB-D数据由粗到精全局递归式反馈的语义标注框架，并将整个语义标注框架划分为粗粒度的区域级语义标签推断与细粒度的像素级语义标签求精两大部分；与传统单一的区域级或像素级语义标注框架不同，该框架重新建立粗粒度区域级语义标注与细粒度像素级语义标注间的联系，通过引入一种合理的全局递归式反馈的机制，使粗粒度区域级的语义标注结果与细粒度像素级的语义标注结果交替迭代更新优化。通过这种方式较好地融合了场景图像中不同区域层次的多模态信息，从一定程度上解决传统室内场景语义标注方案中普遍存在的难以合适地选择标注基元的问题。



CN 104809187 B

[转续页]

[接上页]

(56)对比文件

Philipp 等.Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials.《Advances in Neural Information Processing Systems(NIPS)》.2011,1-9.

Nathan Silberman 等.Indoor Segmentation and Support Inference from RGBD Images.《Computer Vision-ECCV 2012》.2012,第7576卷746-760.

1.一种基于RGB-D数据的室内场景语义标注方法，利用基于RGB-D信息的由粗到精、全局递归式反馈的语义标注框架进行室内场景图像的语义标注，其特征在于：该语义标注框架是由粗粒度的区域级语义标签推断与细粒度的像素级语义标签求精，交替迭代更新构成，包括如下步骤：

步骤001.针对RGB-D训练数据集合中的RGB图像进行过分割，获取该RGB图像中的超像素，形成训练数据的超像素集；

步骤002.根据RGB-D训练数据集合中的RGB图像和对应的深度图像，分别针对该训练数据的超像素集中的各个超像素做如下操作：求取对应超像素的各个区域特征单元，然后对该超像素的各个区域特征单元分别进行归一化处理，获得该超像素的各个归一化区域特征单元，最后将该超像素的各个归一化区域特征单元进行拼接，构成对应于该超像素的多模态特征向量；

步骤003.针对该训练数据的超像素集中的各个超像素，根据RGB-D训练数据集合中包含的基准标注信息，获取该各个超像素分别对应的类别标签；

步骤004.针对该训练数据的超像素集中各个超像素分别对应的类别标签、多模态特征向量，分别整合构成分别对应于各个超像素的各个条目，并整合该训练数据中全部超像素分别所对应的各个条目，构成该训练数据的超像素集对应的语义标签池；

步骤005.将获得的该训练数据的超像素集对应的语义标签池作为训练样本，训练随机决策森林分类器；

步骤006.针对查询图像进行过分割，获取该查询图像中的超像素，形成查询图像的超像素集；并按步骤002中的方法，根据查询图像和对应的深度图像，针对该查询图像的超像素集中的各个超像素，分别求取对应超像素的多模态特征向量，构成该查询图像的超像素集对应的语义标签池；

步骤007.采用已经训练的随机决策森林分类器，针对该查询图像的超像素集中的超像素进行语义标签推断，获得对应该查询图像的区域结构粗粒度级别标注图像；

步骤008.针对获得对应该查询图像的区域结构粗粒度级别标注图像进行标签求精，获得对应该查询图像的细粒度级别标注图像；

步骤009.针对获得对应该查询图像的细粒度级别标注图像，采用内部递归式反馈机制进行标签求精，获得该查询图像的最终细粒度级别标注图像；

步骤010.根据获得该查询图像的最终细粒度级别标注图像，设计获得由粗粒度的区域级语义推断到细粒度的像素级语义求精的全局递归式反馈机制，将该查询图像的最终细粒度级别标注图像作为额外信息引入步骤001和步骤006中分别针对图像的过分割操作中，并根据该全局递归式反馈机制，返回步骤001依次执行各个步骤，且根据全局递归式反馈机制中的终止条件，获得该查询图像的最终标注图像。

2.根据权利要求1所述一种基于RGB-D数据的室内场景语义标注方法，其特征在于：所述步骤001和所述步骤006中分别针对图像进行过分割的操作，采用基于图像分层显著度导引的简单线性迭代聚类过分割算法，其中，该基于图像分层显著度导引的简单线性迭代聚类过分割算法具体包括如下步骤：

步骤A01.初始化各个聚类中心$C_w=[L^*_{cw},a^*_{cw},b^*_{cw},i_{dw},i_{sw},x_w,y_w,A_w]^T$，w＝1，2...W，在原图

像上按照网格大小间隔$S^*=\sqrt{N/W}$均匀分布；其中，G^T表示参数向量G的转置；L^*_{cw} a^*_{cw},b^*_{cw}表示RGB-D室内场景图像在CIELAB颜色空间上的像素值；i_{dw},i_{sw}表示第w个聚类中心的深度值及显著度信息；A_w表示细粒度语义标注图像上某像素所属的标签值；W为期望生成的超像素数目；S^*近似描述每两个邻近超像素中心的距离；N表示图像中包含的像素数目；并且调整聚类中心到预设邻域内梯度最小的点；

同时，设置类标签数组label[i]＝-1,i＝1,2...N,用来记录每个像素点的所属超像素的标签；设置距离数组dis[i]＝M,i＝1,2...N,用来记录每个像素点到最邻近像素中心的距离，M为预设的初始值；

步骤A02.根据如下公式，分别计算各个聚类中心C_w的$2S^**2S^*$邻域内各个像素i到其对应聚类中心C_w的距离D_s；

$$D_s=d_{cds}+\frac{m}{S^*}d_{xyj}+\lambda d_{fb}$$

$$d_{cds}=\sqrt{\begin{matrix}(L^*_{cw}-L^*_{ci})^2+(a^*_{cw}-a^*_{ci})^2+(b^*_{cw}-b^*_{ci})^2\\+(i_{dw}-i_{di})^2+(i_{sw}-i_{si})^2\end{matrix}}$$

$$d_{xy}=\sqrt{(x_w-x_i)^2+(y_w-y_i)^2}$$

$$d_{fb}=\begin{cases}\sqrt{(A_w-A_i)^2} & ，当引入有效的细粒度语义标注信息时\\ 0 & ，当未引入有效的细粒度语义标注信息时\end{cases}$$

$$S^*=\sqrt{N/W}$$

其中，d_{cds}表示图像中任意两个像素点在颜色空间(c)、深度信息(d)、显著度空间(s)上的距离测度；d_{xy}为图像中任意两个像素点在像素位置空间上的距离测度；d_{fb}表示细粒度反馈项，用于在全局反馈阶段引入细粒度语义标注信息；m为紧密系数；λ为细粒度反馈项d_{fb}的平衡系数；

并且，分别针对各个像素点，判断像素点的D_s是否小于像素点的dis[i]，是则更新该像素点dis[i]的数据为其D_s的数据，并更新该像素点label[i]的数据为该像素点所对应聚类中心的次序w；否则不做任何操作；

步骤A03.计算更新各个聚类中心，并分别判断新各个聚类中心对应的类标签变化的像素数目是否不足其对应全部像素个数的1％，是则结束；否则返回步骤A02。

3.根据权利要求2所述一种基于RGB-D数据的室内场景语义标注方法，其特征在于：所述步骤010中，所述像素级语义求精的全局递归式反馈机制的实现包括如下步骤：

步骤D01.将获得查询图像的最终细粒度级别标注图像，作为一种额外信息，针对步骤001和步骤006中分别对图像进行过分割操作的简单线性迭代聚类过分割算法，引入细粒度语义标注信息，将简单线性迭代聚类过分割算法的聚类中心扩充至8维；

步骤D02.根据全局递归式反馈机制，返回步骤001依次执行各个步骤，更新获得查询图像的最终细粒度级别标注图像，并根据全局递归式反馈机制中的终止条件，判断更新后查询图像的最终细粒度级别标注图像与更新前查询图像的最终细粒度级别标注图像是否至多有5％的像素标签不同，是则将该更新后查询图像的最终细粒度级别标注图像作为该查询图像的最终标注图像；否则返回步骤D01。

4.根据权利要求1所述一种基于RGB-D数据的室内场景语义标注方法，其特征在于：所述步骤002中，所述区域特征单元包括超像素质心、色彩HSV分量均值及其相应直方图、基于彩色RGB图像的梯度方向直方图、基于深度图像的梯度方向直方图、基于表面法线向量图像的梯度方向直方图。

5.根据权利要求1所述一种基于RGB-D数据的室内场景语义标注方法，其特征在于：所述步骤008中，所述针对获得对应该查询图像的区域结构粗粒度级别标注图像进行标签求精的操作采用改进型像素级稠密CRFs概率图模型，该改进型像素级稠密CRFs概率图模型的具体构建包括如下步骤：

步骤B01.利用深度图像和PCL点云库，计算图像中每个像素点的法线向量信息，并将法线向量信息转换存储为法线向量图像；

步骤B02.根据已有深度图像及法线向量图像，针对稠密CRFs概率图模型，以像素为图模型节点进行成对项势能的修正拓展，获得像素级稠密CRFs概率图模型，并获得该像素级稠密CRFs概率图模型的能量函数表达式，如下所示：

$$E(\boldsymbol{X}\mid\boldsymbol{I})=\sum_{i}\psi_u(x_i^*)+\sum_{(i,j)}\psi_p(x_i,x_j)$$

$$\psi_u(x_i^*)=\exp(\varphi(e_i,x_i^*))/(1+\exp(\varphi(e_i,x_i^*)))$$

$$\varphi(e_i,x_i^*)=\log P(e_i\mid x_i^*)-\log P(e_i\mid \overline{x_i^*})$$

$$\psi_p(x_i,x_j)=\mu(x_i,x_j)K(f_i,f_j)$$

$$\boldsymbol{K}(\boldsymbol{f}_i,\boldsymbol{f}_j)=w^{(1)}\exp\left(-\frac{|\boldsymbol{p}_i-\boldsymbol{p}_j|^2}{2\theta_\alpha^2}-\frac{|\boldsymbol{I}_i-\boldsymbol{I}_j|^2}{2\theta_\beta^2}-\frac{|d_i-d_j|^2}{2\theta_\gamma^2}-\frac{|\boldsymbol{n}_i-\boldsymbol{n}_j|^2}{2\theta_\delta^2}\right)+w^{(2)}\exp\left(-\frac{|\boldsymbol{p}_i-\boldsymbol{p}_j|^2}{2\theta_\chi^2}\right)$$

其中，E(X|I)表示能量函数；X表示随机场；I表示给定的图像；$\psi_u(x_i^*)$为图像中第i个像素点的一元项势能；$\varphi(e_i,x_i^*)$表示对数似然比；e_i表示图像中第i个像素；x_i^*表示图像中第i个像素最有可能属于的类别标签；$\overline{x_i^*}$表示除x_i^*以外的所有类别标签；$P(e_i\mid x_i^*)$、$P(e_i\mid\overline{x_i^*})$均为普通的条件概率表达式，其概率值通过所述随机决策森林分类器获取；x_i和x_j分别表示图像中第i个像素点、第j个像素点的类别标签；$\psi_p(x_i,x_j)$为图像中任意两像素点i,j间的成对项势能；$\mu(x_i,x_j)$为标签兼容性函数；$K(f_i,f_j)$为高斯核的线性组合，f_i和f_j分别表示第i个像素点、第j个像素点的特征向量；p_i和p_j表示图像中任意两像素点i,j的坐标位置向量；I_i和I_j表示图像中任意两像素点i,j的RGB色彩通道向量；d_i和d_j表示图像中任意两像素点i,j的深度值；ni和n_j表示任意两像素点i,j相应的表面法线向量；$w^{(1)}$和$w^{(2)}$为两个高斯核的权值系数；$\theta_\alpha,\theta_\beta,\theta_\gamma$和$\theta_\delta$用来控制任意两像素点i,j属于一类的可能性，其所在项被称为外观核；θ_χ所控制的项称为平滑核，θ_χ用于控制衡量孤立区域大小。

6.根据权利要求5所述一种基于RGB-D数据的室内场景语义标注方法，其特征在于：所述步骤009中，所述内部递归式反馈机制为基于所述深度图像与法线向量图像修正拓展后的像素级稠密CRFs概率图模型，获得的由输入到输出的内部递归式反馈机制，该内部递归式反馈机制的实现包括如下步骤：

步骤C01.根据获得的像素级稠密CRFs概率图模型，针对查询图像对应粗粒度级别区域级语义标签推断部分生成的粗粒度语义标注进行细粒度级别的求精，更新获得该查询图像的细粒度级别标注图像；

步骤C02.根据该查询图像对应区域级语义标签推断部分产生的过分割信息，将获得的该查询图像的细粒度级别标注图像中的类别标签映射回该查询图像的超像素集中，更新该查询图像的超像素集；

步骤C03.根据该查询图像的超像素集中的类别标签和该查询图像对应区域级语义标签推断部分产生的过分割信息，更新该查询图像的区域结构粗粒度级别标注图像，并判断更新后的该查询图像的区域结构粗粒度级别标注图像与未更新前该查询图像的区域结构粗粒度级别标注图像对应的超像素语义标签是否一致，是则获得的该查询图像的细粒度级别标注图像作为该查询图像的最终细粒度级别标注图像；否则返回步骤C01。

一种基于RGB-D数据的室内场景语义标注方法

技术领域

[0001] 本发明涉及一种图像语义标注方法，尤其涉及一种基于RGB-D数据的室内场景语义标注方法，属于计算机视觉的语义标签分类技术领域。

背景技术

[0002] 图像语义标注是计算机视觉中场景理解工作的核心单元，其基本目标是为给定的查询图像中的每一个像素稠密地提供一个预定义的语义类别标签。考虑到图像语义的模糊性、复杂性和抽象性，一般建立的图像语义模型都是分层次的。其中，“目标语义”处于语义层次中的中层，在很多高层次语义推理过程中起到了承上启下的作用。根据图像语义标注问题中标注基元的量化级别，可将当前多数图像语义标注方案大致分为两类，包括：像素级的语义标注方案和区域级的语义标注方案。两种方案在实现效率、标注精度和视觉效果上各有其优劣势。

[0003] 一方面，相较于区域级表达，像素级表达的确不失为是一种简易直观的图像表达层次，像素级语义标注方案将单一像素作为标注的基本单元，免除了对数据集中的图像进行区域级分割的繁复操作。此外，像素级特征的获取一般较为简单，故相较于区域级标注方案，其在整体实现效率上存在较大优势，而且由于其表达层次较低，不易出现错误标签分布密集的问题。但由于像素自身有效载荷相对有限，如何针对像素级方案构建更为鲁棒且更具辨识力的像素级特征，成为了像素级语义标注方案发展的难点和瓶颈。典型的像素级语义标注方案包括：[KR HENB HL P,KOLTUN V.Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials[C]//Advances in Neural Information Processing Systems(NIPS),2011.]通过探讨一种成对项势能由高斯核线性组合而成的像素级稠密全连通Conditional Random Fields(CRFs)概率图模型的推断算法，一定程度上提升了像素级标注方案在上下文推断期间的效率。

[0004] 另一方面，区域级表达在特征构建层面上较像素级表达具有显著优势，这主要是因为分割区域一般被定义为像素的集合，相较于单一像素，具有更为丰富的纹理及上下文信息。以往利用区域级表达进行图像语义标注的经典范例很多：[REN Xiaofeng,BO Liefeng,FOX D.RGB-(D)scene labeling:Features and algorithms[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2012:2759-2766.]在区域级表达层面上成功构建了一种称为核描述子的特征，在一定程度上提高了室内场景语义标注的性能；[SILBERMAN N,HOIEM D,KOHLI P,et al.Indoor segmentation and support inference from RGBD images[M].Computer Vision-ECCV 2012.Springer Berlin Heidelberg,2012:746-760.]则不仅在区域级上解析了室内场景中的主要平面和物体，还利用上述信息对室内场景中物体的支撑关系进行了推断；[TIGHE J,LAZEBNIK S.Superparsing:scalable nonparametric image parsing with superpixels[C]//European Conference on Computer Vision(ECCV),2010:352-365.]提出了一种简单有效的非参数语义标注算法，其基于懒惰学习的思想，实现中涉及区域级匹配等降低系统复杂

度的操作。但基于区域的语义标注方案仍存在一个致命的缺陷，即其大大依赖于区域分割算法的性能。一旦区域分割算法将所属两种或两种以上类别的像素划分至同一个分割区域下，则无论后续采用的分类器性能如何优异，即便通过有效的区域级上下文推断，皆无法改变该区域误标注的结果(仅能在误标注的前提下，尽可能地提升像素标签正确率)，严重限制了区域级语义标注方案的准确度和视觉表现能力。

[0005] 鉴于深度传感器能够获取更加丰富的外观和几何结构信息，深度摄像机在计算机视觉领域越来越受到广泛的关注。近年来，越来越多的公司开发出了RGB-D摄像机，该类设备的特点是能够实时地为所摄场景捕获得到相应的RGB图像和深度图像，比如2010年微软发布了能够实时采集RGB-D图像的摄像头(即Kinect)；2011年华硕发布了Xtion PRO；2013年体感控制器制造公司Leap发布的Leap Motion。

[0006] 室内场景语义标注，有其内在固有属性(即存在大量的语义类别及类别间存在遮挡、视觉特征缺乏较强辨识能力以及室内光照不可控等问题)，已成为了图像语义标注技术中棘手且极富挑战性的研究方向之一。在室内条件下，相较于单一的RGB数据，RGB-D数据的优势在于：其提供了普通摄像机在三维投影过程中丢失的3D几何信息，包含的深度信息可作为一个独立于室内环境照明条件的通道，这为解决室内场景语义标注问题带来了新思路。作为利用深度信息进行室内场景语义标注的先驱，[SILBERMAN N,FERGUS R.Indoor scene segmentation using a structured light sensor[C]//IEEE International Conference on Computer Vision(ICCV),2011:601-608.]在NYU RGB-D数据集中获得了56.6%的准确率，展示了基于RGB-D数据语义感知室内场景的巨大潜力。但目前大多数语义标注工作仅仅将深度信息用于构造区域级特征，却忽略了其在上下文推断中的作用，而且所采用的深度信息也较为单一。

[0007] 综上所述，现有室内场景语义标注方案普遍存在着标注基元量化级别难于选择的问题，且几何深度信息在上下文推理过程中的作用也未获得足够的重视。

发明内容

[0008] 针对上述技术问题，本发明所要解决的技术问题是提供一种基于RGB-D数据的室内场景语义标注方法，基于RGB-D数据，采用区域级语义标签推断与像素级语义标签求精，两部分交替迭代更新优化的方式，能够在一定程度上解决传统语义标注工作中难以合适地选择标注基元的问题。

[0009] 本发明为了解决上述技术问题采用以下技术方案：本发明设计了一种基于RGB-D数据的室内场景语义标注方法，利用基于RGB-D信息的由粗到精、全局递归式反馈的语义标注框架进行室内场景图像的语义标注，该语义标注框架是由粗粒度的区域级语义标签推断与细粒度的像素级语义标签求精，交替迭代更新构成，包括如下步骤：

[0010] 步骤001.针对RGB-D训练数据集合中的RGB图像进行过分割，获取该RGB图像中的超像素，形成训练数据的超像素集；

[0011] 步骤002.根据RGB-D训练数据集合中的RGB图像和对应的深度图像，分别针对该训练数据的超像素集中的各个超像素做如下操作：求取对应超像素的各个区域特征单元，然后对该超像素的各个区域特征单元分别进行归一化处理，获得该超像素的各个归一化区域特征单元，最后将该超像素的各个归一化区域特征单元进行拼接，构成对应于该超像素的

多模态特征向量；

[0012] 步骤003.针对该训练数据的超像素集中的各个超像素，根据RGB-D训练数据集合中包含的基准标注信息，获取该各个超像素分别对应的类别标签；

[0013] 步骤004.针对该训练数据的超像素集中各个超像素分别对应的类别标签、多模态特征向量，分别整合构成分别对应于各个超像素的各个条目，并整合该训练数据中全部超像素分别所对应的各个条目，构成该训练数据的超像素集对应的语义标签池；

[0014] 步骤005.将获得的该训练数据的超像素集对应的语义标签池作为训练样本，训练随机决策森林分类器；

[0015] 步骤006.针对查询图像进行过分割，获取该查询图像中的超像素，形成查询图像的超像素集；并按步骤002中的方法，根据查询图像和对应的深度图像，针对该查询图像的超像素集中的各个超像素，分别求取对应超像素的多模态特征向量，构成该查询图像的超像素集对应的语义标签池；

[0016] 步骤007.采用已经训练的随机决策森林分类器，针对该查询图像的超像素集中的超像素进行语义标签推断，获得对应该查询图像的区域结构粗粒度级别标注图像；

[0017] 步骤008.针对获得对应该查询图像的区域结构粗粒度级别标注图像进行标签求精，获得对应该查询图像的细粒度级别标注图像；

[0018] 步骤009.针对获得对应该查询图像的细粒度级别标注图像，采用内部递归式反馈机制进行标签求精，获得该查询图像的最终细粒度级别标注图像；

[0019] 步骤010.根据获得该查询图像的最终细粒度级别标注图像，设计获得由粗粒度的区域级语义推断到细粒度的像素级语义求精的全局递归式反馈机制，将该查询图像的最终细粒度级别标注图像作为额外信息引入步骤001和步骤006中分别针对图像的过分割操作中，并根据该全局递归式反馈机制，返回步骤001依次执行各个步骤，且根据全局递归式反馈机制中的终止条件，获得该查询图像的最终标注图像。

[0020] 作为本发明的一种优选技术方案：所述步骤001和所述步骤006中分别针对图像进行过分割的操作，采用基于图像分层显著度导引的简单线性迭代聚类过分割算法，其中，该基于图像分层显著度导引的简单线性迭代聚类过分割算法具体包括如下步骤：

[0021] 步骤A01.初始化各个聚类中心$C_w=[L^*_{cw},a^*_{cw},b^*_{cw},i_{dw},i_{sw},x_w,y_w,A_w]^T$，$w=1,2...W$，在原图像上按照网格大小间隔$S^*=\sqrt{N/W}$均匀分布；其中，$G^T$表示参数向量G的转置；$L^*_{cw}$ a^*_{cw},b^*_{cw}表示RGB-D室内场景图像在CIELAB颜色空间上的像素值；i_{dw}，i_{sw}表示第w个聚类中心的深度值及显著度信息；A_w表示细粒度语义标注图像上某像素所属的标签值；W为期望生成的超像素数目；S^*近似描述每两个邻近超像素中心的距离；N表示图像中包含的像素数目；并且调整聚类中心到预设邻域内梯度最小的点；

[0022] 同时，设置类标签数组label[i]=-1，i=1，2...N，用来记录每个像素点的所属超像素的标签；设置距离数组dis[i]=M，i=1，2...N，用来记录每个像素点到最邻近像素中心的距离，M为预设的初始值；

[0023] 步骤A02.根据如下公式，分别计算各个聚类中心C_w的$2S^*$*$2S^*$邻域内各个像素i到其对应聚类中心C_w的距离D_s；

[0024] $$D_s = d_{cds} + \frac{m}{S^*} d_{xy} + \lambda d_{fb}$$

[0025] $$d_{cds} = \sqrt{\begin{matrix}(L^*_{cw} - L^*_{ci})^2 + (a^*_{cw} - a^*_{ci})^2 + (b^*_{cw} - b^*_{ci})^2 \\ + (i_{dw} - i_{di})^2 + (i_{sw} - i_{si})^2\end{matrix}}$$

[0026] $$d_{xy} = \sqrt{(x_w - x_i)^2 + (y_w - y_i)^2}$$

[0027] $$d_{fb} = \begin{cases} \sqrt{(A_w - A_i)^2} & ，当引入有效的细粒度语义标注信息时 \\ 0 & ，当未引入有效的细粒度语义标注信息时 \end{cases}$$

[0028] $$S^* = \sqrt{N/W}$$

[0029] 其中，d_{cds}表示图像中任意两个像素点在颜色空间(c)、深度信息(d)、显著度空间(s)上的距离测度；d_{xy}为图像中任意两个像素点在像素位置空间上的距离测度；d_{fb}表示细粒度反馈项，用于在全局反馈阶段引入细粒度语义标注信息；m为紧密系数；λ为细粒度反馈项d_{fb}的平衡系数；

[0030] 并且，分别针对各个像素点，判断像素点的D_s是否小于像素点的dis[i]，是则更新该像素点dis[i]的数据为其D_s的数据，并更新该像素点label[i]的数据为该像素点所对应聚类中心的次序w；否则不做任何操作；

[0031] 步骤A03.计算更新各个聚类中心，并分别判断新各个聚类中心对应的类标签变化的像素数目是否不足其对应全部像素个数的1%，是则结束；否则返回步骤A02。

[0032] 作为本发明的一种优选技术方案：所述步骤010中，所述像素级语义求精的全局递归式反馈机制的实现包括如下步骤：

[0033] 步骤D01.将获得查询图像的最终细粒度级别标注图像，作为一种额外信息通道，针对步骤001和步骤006中分别对图像进行过分割操作的简单线性迭代聚类过分割算法，引入细粒度语义标注信息，将简单线性迭代聚类过分割算法的聚类中心扩充至8维；

[0034] 步骤D02.根据全局递归式反馈机制，返回步骤001依次执行各个步骤，更新获得查询图像的最终细粒度级别标注图像，并根据全局递归式反馈机制中的终止条件，判断更新后查询图像的最终细粒度级别标注图像与更新前查询图像的最终细粒度级别标注图像是否至多有5%的像素标签不同，是则将该更新后查询图像的最终细粒度级别标注图像作为该查询图像的最终标注图像；否则返回步骤D01。

[0035] 作为本发明的一种优选技术方案：所述步骤002中，所述区域特征单元包括超像素质心、色彩HSV分量均值及其相应直方图、基于彩色RGB图像的梯度方向直方图、基于深度图像的梯度方向直方图、基于表面法线向量图像的梯度方向直方图。

[0036] 作为本发明的一种优选技术方案：所述步骤008中，所述针对获得对应该查询图像的区域结构粗粒度级别标注图像进行标签求精的操作采用改进型像素级稠密CRFs概率图模型，该改进型像素级稠密CRFs概率图模型的具体构建包括如下步骤：

[0037] 步骤B01.利用深度图像和PCL点云库，计算图像中每个像素点的法线向量信息，并将法线向量信息转换存储为法线向量图像；

[0038] 步骤B02.根据已有深度图像及法线向量图像，针对稠密CRFs概率图模型，以像素为图模型节点进行成对项势能的修正拓展，获得像素级稠密CRFs概率图模型，并获得该像

素级稠密CRFs概率图模型的能量函数表达式，如下所示：

[0039] $$E(\boldsymbol{X}|\boldsymbol{I})=\sum_{i}\psi_{u}(x_{i}^{*})+\sum_{(i,j)}\psi_{p}(x_{i},x_{j})$$

[0040] $$\psi_{u}(x_{i}^{*})=\exp(\varphi(e_{i},x_{i}^{*}))/(1+\exp(\varphi(e_{i},x_{i}^{*})))$$

[0041] $$\varphi(e_{i},x_{i}^{*})=\log P(e_{i}|x_{i}^{*})-\log P(e_{i}|\overline{x_{i}^{*}})$$

[0042] $\psi_{p}(x_{i},x_{j})=\mu(x_{i},x_{j})K(f_{i},f_{j})$

[0043] $$\boldsymbol{K}(\boldsymbol{f}_{i},\boldsymbol{f}_{j})=w^{(1)}\exp(-\frac{|\boldsymbol{p}_{i}-\boldsymbol{p}_{j}|^{2}}{2\theta_{\alpha}^{2}}-\frac{|\boldsymbol{I}_{i}-\boldsymbol{I}_{j}|^{2}}{2\theta_{\beta}^{2}}-\frac{|d_{i}-d_{j}|^{2}}{2\theta_{\gamma}^{2}}-\frac{|\boldsymbol{n}_{i}-\boldsymbol{n}_{j}|^{2}}{2\theta_{\delta}^{2}})+w^{(2)}\exp(-\frac{|\boldsymbol{p}_{i}-\boldsymbol{p}_{j}|^{2}}{2\theta_{\chi}^{2}})$$

[0044] 其中，$E(X|I)$表示能量函数；X表示随机场；I表示给定的图像；$\psi_{u}(x_{i}^{*})$为图像中第i个像素点的一元项势能；$\varphi(e_{i},x_{i}^{*})$表示对数似然比；e_{i}表示图像中第i个像素；x_{i}^{*}表示图像中第i个像素最有可能属于的类别标签；$\overline{x_{i}^{*}}$表示除x_{i}^{*}以外的所有类别标签；$P(e_{i}|x_{i}^{*})$、$P(e_{i}|\overline{x_{i}^{*}})$均为普通的条件概率表达式，其概率值通过所述随机决策森林分类器获取；x_{i}和x_{j}分别表示图像中第i个像素点、第j个像素点的类别标签；$\psi_{p}(x_{i},x_{j})$为图像中任意两像素点i，j间的成对项势能；$\mu(x_{i},x_{j})$为标签兼容性函数；$K(f_{i},f_{j})$为高斯核的线性组合，f_{i}和f_{j}分别表示第i个像素点、第j个像素点的特征向量；p_{i}和p_{j}表示图像中任意两像素点i，j的坐标位置向量；I_{i}和I_{j}表示图像中任意两像素点i，j的RGB色彩通道向量；d_{i}和d_{j}表示图像中任意两像素点i，j的深度值；n_{i}和n_{j}表示任意两像素点i，j相应的表面法线向量；$w^{(1)}$和$w^{(2)}$为两个高斯核的权值系数；θ_{α}，θ_{β}，θ_{γ}和θ_{δ}用来控制任意两像素点i，j属于一类的可能性，其所在项被称为外观核；θ_{x}所控制的项称为平滑核，θ_{x}用于控制衡量孤立区域大小。

[0045] 作为本发明的一种优选技术方案：所述步骤009中，所述内部递归式反馈机制为基于所述深度图像与法线向量图像修正拓展后的像素级稠密CRFs概率图模型，获得的由输入到输出的内部递归式反馈机制，该内部递归式反馈机制的实现包括如下步骤：

[0046] 步骤C01.根据获得的像素级稠密CRFs概率图模型，针对查询图像对应粗粒度级别区域级语义标签推断部分生成的粗粒度语义标注进行细粒度级别的求精，更新获得该查询图像的细粒度级别标注图像；

[0047] 步骤C02.根据该查询图像对应区域级语义标签推断部分产生的过分割信息，将获得的该查询图像的细粒度级别标注图像中的类别标签映射回该查询图像的超像素集中，更新该查询图像的超像素集；

[0048] 步骤C03.根据该查询图像的超像素集中的类别标签和该查询图像对应区域级语义标签推断部分产生的过分割信息，更新该查询图像的区域结构粗粒度级别标注图像，并判断更新后的该查询图像的区域结构粗粒度级别标注图像与未更新前该查询图像的区域结构粗粒度级别标注图像对应的超像素语义标签是否一致，是则获得的该查询图像的细粒度级别标注图像作为该查询图像的最终细粒度级别标注图像；否则返回步骤C01。

[0049] 本发明所述一种基于RGB-D数据的室内场景语义标注方法采用以上技术方案与现有技术相比，具有以下技术效果：

[0050] 首先构建了一种基于RGB-D数据由粗到精全局递归式反馈的语义标注框架，并将整个语义标注框架划分为粗粒度的区域级语义标签推断与细粒度的像素级语义标签求精两大部分；与传统单一的区域级或像素级语义标注框架不同，该框架重新建立粗粒度区域

级语义标注与细粒度像素级语义标注间的联系，通过引入一种合理的全局递归式反馈的机制，使粗粒度区域级的语义标注结果与细粒度像素级的语义标注结果交替迭代更新优化。通过这种方式较好地融合了场景图像中不同区域层次的多模态信息，从一定程度上解决传统室内场景语义标注方案中普遍存在的难以合适地选择标注基元的问题。

[0051] 其次，本发明具体设计了基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法，相较于传统的简单线性迭代聚类(SLIC)过分割算法，一定程度上解决了目前非监督过分割算法在杂乱的室内场景中难以得到具有较高边缘一致性超像素的现状。并且利用分层显著度具备的抗小范围高对比度模式的特点，将图像分层显著度引入简单线性迭代聚类(SLIC)过分割算法，即扩展简单线性迭代聚类(SLIC)过分割算法的聚类空间，有助于改善小范围高对比度模式对划分简单线性迭代聚类(SLIC)超像素过程中的不利影响，该类模式对超像素大小近似均匀的简单线性迭代聚类(SLIC)过分割算法及类似过分割算法影响很大。

[0052] 最后，本发明具体设计了像素级稠密CRFs概率图模型，并在像素级稠密CRFs概率图模型中引入几何深度信息与内部递归式反馈机制。其中，具体设计的像素级稠密CRFs概率图模型深入挖掘了几何深度信息在室内场景语义标签上下文优化求精中的潜力，并且实验表明，通过在概率图模型中引入有效且可靠的几何深度信息，一定程度上抑制了室内光源对室内场景语义标注视觉效果的影响，并提升了语义标签的准确性。而内部递归式反馈机制，则通过引入稠密CRFs概率图模型输入到输出间的关系，用于基于稠密CRFs概率图模型改善细粒度像素级语义标签，同时也令细粒度像素级语义标签求精部分的标注结果更加稳定，最终产生视觉表现力更强、标注准确率更高的语义标注图像。

附图说明

[0053] 图1是基于RGB-D数据的室内场景语义标注方法的流程示意图。

具体实施方式

[0054] 下面结合说明书附图针对本发明的具体实施方式作进一步详细的说明。

[0055] 如图1所示，本发明设计基于RGB-D数据的室内场景语义标注方法在实际应用过程当中，利用基于RGB-D信息的由粗到精、全局递归式反馈的语义标注框架进行室内场景图像的语义标注，其特征在于：该语义标注框架是由粗粒度的区域级语义标签推断与细粒度的像素级语义标签求精，交替迭代更新构成，包括如下步骤：

[0056] 步骤001.采用基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法，针对RGB-D训练数据集合中的RGB图像进行过分割，获取该RGB图像中的超像素，形成训练数据的超像素集。

[0057] 本发明针对RGB-D室内场景图像数据可采用现有的各类RGB-D摄像器材获取。例如微软公司的Kinect，该装置利用内置的RGB摄像头和红外传感器收集RGB图像和深度图像。亦可直接采用某些权威计算机视觉研究社区所提供的室内场景图像数据集。本发明在具体实施过程中选用的是NYU Depth v2[SILBERMAN N,HOIEM D,KOHLI P,et al.Indoor segmentation and support inference from RGBD images[M].Computer Vision-ECCV 2012.Springer Berlin Heidelberg,2012:746-760.]及SUN3D数据集[XIAO Jianxiong,

OWENS A,TORRALBA A.SUN3D:A database of big spaces reconstructed using sfm and object labels[C]//IEEE International Conference on Computer Vision(ICCV),2013:1625-1632.]。NYU Depth系列数据集是国际上首个特别为大规模语义标注工作构建的RGB-D室内场景图像数据集。近期由Princeton University&MIT联合推出的SUN3D数据集则具有很多其他传统的基于视角的2D数据集所不具备的特性,该数据集为数据集中包含的任意室内场景提供了连续的视频帧。其它一些深度数据集,包括伯克利大学的3D目标数据集等,很多并不适合用以训练语义标注系统,主要是由于这些室内场景图像数据集中缺乏较为稠密的语义标注信息。然而NYU Depth系列数据集与SUN3D室内场景图像数据集在涵盖大量室内场景图像的同时,均包含较为稠密且可用的语义类别标签。由于上述两个数据集均是使用Kinect或类似的深度传感设备于室内场景记录并生成的,且同时提供了RGB图像及深度图像,因此被统称为RGB-D(depth)室内场景图像数据集。

[0058] 由图1所示,本发明由粗粒度的区域级语义标签推断和细粒度的像素级语义标签求精两部分交替迭代组成。由于区域级的语义标签推断处理阶段所产生的粗粒度标注的视觉效果极度依赖于过分割算法性能,如何划分更具一致性且能够较好的覆盖目标真实边缘的过分割区域一直都是图像处理研究中的热点和难点之一。按照综合性能考虑,目前国际上性能较为显著的过分割算法为SEEDS算法[BERGH V D,BOIX X,ROIG G,et al.SEEDS:Superpixels extracted via energy-driven sampling[C]//European Conference on Computer Vision(ECCV),2012:13-26.]及简单线性迭代聚类(SLIC)算法[ACHANTA R,SHAJI A,SMITH K,et al.SLIC superpixels compared to state-of-the-art superpixel methods[J].Pattern Analysis and Machine Intelligence(PAMI),2012,34(11):2274-2281.],两者的性能非常的接近。其中简单线性迭代聚类(SLIC)算法,是K-means聚类算法的一种快速逼近,能比较好地覆盖目标真实边缘,超像素大小也较均匀且具备计算复杂度为线性的优势,可满足目前很多计算机视觉实际应用的需求。但即便简单线性迭代聚类(SLIC)过分割算法具备生成较高质量超像素的能力,但其在面对结构混乱、目标交叠且光照条件复杂的室内场景图像时,仍可能出现错误划分超像素的问题。

[0059] 为使简单线性迭代聚类(SLIC)过分割算法更适于解决室内场景语义标注问题,设法提升粗粒度区域级语义标签推断部分的性能,本发明提出了一种基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法,将传统简单线性迭代聚类(SLIC)过分割算法的5维(3维彩色通道+2维位置信息通道)的聚类空间扩展至8维(当引入有效的细粒度语义标注信息时将扩展至8维:3维彩色RGB图像通道+2维位置信息通道+1维图像分层显著度信息通道+1维深度信息通道+1维细粒度语义标注信息通道)。所述图像分层显著度是一种从多层结构中分析显著度信息的方案,本发明采用[YAN Qiong,XU Li,SHI Jianping,et al.Hierarchical saliency detection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2013:1155-1162.]中所述方法获取图像分层显著度信息。其关键在于通过采用上述方法所述多层的分析与分层的推断过程获取的图像分层显著度信息具有弱化复杂结构中常出现的小规模高对比度模式的干扰的特点。

[0060] 室内场景图像作为一种复杂结构,常包含某些小规模高对比度的模式。由于这类小规模模式具有高对比度的特点,传统简单线性迭代聚类(SLIC)过分割算法执行结果常会以此类模式的边界作为室内场景中重要目标的边界。由于传统简单线性迭代聚类(SLIC)过

分割算法具有所划分超像素大小基本一致的特点，故在此类模式边界高对比度的影响下，常导致语义标注系统所重点关注的目标的边界遭到忽略，从而导致经区域级语义标签推断后的粗粒度语义标注结果的视觉效果并不理想。故本发明利用图像分层显著度信息对传统简单线性迭代聚类(SLIC)过分割算法进行了扩充与修正。

[0061] 深度信息来自于独立于室内较远的光源对场景光照影响的通道，由于室内场景光照情况复杂，某些重要的目标常会被强烈的光照所掩盖，但深度信息具备独立于光源的特性，有助于弱化室内光照对过分割算法的影响，以获取真实的物体边界，故本发明也同时使用有效可靠的深度信息对传统简单线性迭代聚类(SLIC)过分割算法进行了修正。

[0062] 其中，该基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法具体包括如下步骤：

[0063] 步骤A01.初始化各个聚类中心$C_w=[L^*_{cw},a^*_{cw},b^*_{cw},i_{dw},i_{sw},x_w,y_w,A_w]^T$，$w=1,2...W$，在原图像上按照网格大小间隔$S^*=\sqrt{N/W}$均匀分布；其中，$G^T$表示参数向量G的转置；$L^*_{cw}$ a^*_{cw},b^*_{cw}表示RGB-D室内场景图像在CIELAB颜色空间上的像素值；i_{dw}，i_{sw}表示第w个聚类中心的深度值及显著度信息；A_w表示细粒度语义标注图像上某像素所属的标签值(若语义标注系统执行至所述基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割部分时并未引入有效的细粒度语义标注信息，则A_w等于0)；W为期望生成的超像素数目；S^*近似描述每两个邻近超像素中心的距离；N表示图像中包含的像素数目；并且调整聚类中心到3*3邻域内梯度最小的点。

[0064] 同时，设置类标签数组label[i]=-1，i=1，2...N，用来记录每个像素点的所属超像素的标签；设置距离数组dis[i]=M，i=1，2...N，用来记录每个像素点到最邻近像素中心的距离，M为预设的初始值。

[0065] 步骤A02.根据如下公式，分别计算各个聚类中心C_w的$2S^*$*$2S^*$邻域内各个像素i到其对应聚类中心C_w的距离D_s；

[0066] $$D_s=d_{cds}+\frac{m}{S^*}d_{xy}+\lambda d_{fb}$$

[0067] $$d_{cds}=\sqrt{\begin{matrix}(L^*_{cw}-L^*_{ci})^2+(a^*_{cw}-a^*_{ci})^2+(b^*_{cw}-b^*_{ci})^2\\+(i_{dw}-i_{di})^2+(i_{sw}-i_{si})^2\end{matrix}}$$

[0068] $$d_{xy}=\sqrt{(x_w-x_i)^2+(y_w-y_i)^2}$$

[0069] $$d_{fb}=\begin{cases}\sqrt{(A_w-A_i)^2}, & \text{当引入有效的细粒度语义标注信息时}\\ 0, & \text{当未引入有效的细粒度语义标注信息时}\end{cases}$$

[0070] $$S^*=\sqrt{N/W}$$

[0071] 其中，d_{cds}表示图像中任意两个像素点在颜色空间(c)、深度信息(d)、显著度空间(s)上的距离测度；d_{xy}为图像中任意两个像素点在像素位置空间上的距离测度(常用x、y分别表示某像素在图像中的横纵坐标)；d_{fb}表示细粒度反馈项，用于在全局反馈阶段引入细粒度语义标注信息；λ为细粒度反馈项d_{fb}的平衡系数；m为紧密系数，实验表明在CIELAB颜色空间中[1，80]都是可行的，本发明设计中，根据经验针对m取值为20。若m值取值越小，聚类结果是超像素形状越不规整，但边界和物体真实边缘重叠较好；m取值越大，超像素越紧凑规

整，但边界性能会下降。

[0072] 并且，分别针对各个像素点，判断像素点的D_s是否小于像素点的dis[i]，是则更新该像素点dis[i]的数据为其D_s的数据，并更新该像素点label[i]的数据为该像素点所对应聚类中心的次序w；否则不做任何操作。

[0073] 步骤A03.计算更新各个聚类中心，并分别判断新各个聚类中心对应的类标签变化的像素数目是否不足其对应全部像素个数的1％，是则结束；否则返回步骤A02。

[0074] 步骤002.根据RGB-D训练数据集合中的RGB图像和对应的深度图像，分别针对该训练数据的超像素集中的各个超像素做如下操作：求取对应超像素的各个区域特征单元，然后对该超像素的各个区域特征单元分别进行归一化处理，获得该超像素的各个归一化区域特征单元，最后将该超像素的各个归一化区域特征单元进行拼接，构成对应于该超像素的多模态特征向量。其中，区域特征单元包括超像素质心、色彩HSV分量均值及其相应直方图、基于彩色RGB图像的梯度方向直方图(HOG)、基于深度图像的梯度方向直方图(HOG)、基于表面法线向量图像的梯度方向直方图(HOG)等6个区域特征单元。

[0075] 超像素质心和色彩HSV分量均值是室内场景图像语义标注方案中常用的特征描述子，由于类别标签在场景图像中的分布往往呈一定规律性，例如："Ground"这个类别标签大多数情况位于室内场景图像中部偏下的位置，所以引入超像素质心这一特征描述具有一定意义，共2维；而色彩HSV分量均值则用于表示场景图像整体的纹理信息分布情况，共3维。

[0076] 利用几何深度信息构造区域级特征描述对提高特征辨识力非常有效，尤其针对本发明所讨论的复杂的室内场景。Kinect及其它专业深度传感设备不仅提供了在普通彩色摄像机投影中无法获取的3D几何信息，包含的深度信息可作为一个独立于室内环境照明条件的通道。来自RGB-D数据的区域特征将降低室内照明在RGB场景图像中掩盖重要目标的可能性。通常，过于强烈的室内光源掩盖场景图像中重要物体的可能性非常高。

[0077] 经所述基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割方法改进方案划分的超像素是一种紧致但边缘并不规则的分割区域，每个超像素所包含的像素数目相近却不一定相同。故在特征描述的选择上，主要考虑与过分割区域像素数目无关的特征描述子，如直方图。为了有效利用RGB图像中纹理信息和深度图像中几何信息，本发明选取了四种与直方图相关的特征描述子：1)色彩HSV分量直方图(6/2/2bins)，共10维；2)基于彩色RGB图像的梯度方向直方图(HOG)(有方向梯度占18bins，无方向梯度占9bins)，共27维；3)基于深度图像的梯度方向直方图(HOG)(有方向梯度占18bins，无方向梯度占9bins)，共27维；4)基于表面法线向量图像的梯度方向直方图(HOG)(有方向梯度占18bins，无方向梯度占9bins)，共27维。

[0078] 梯度方向直方图(HOG)特征计算关键在于计算图像中像素点梯度的幅值和方向，并依据预划分的bins对各类图像进行直方图统计，所述图像中像素点梯度的幅值和方向的数学表达为：

[0079] $G_x(x,y)=P(x+1,y)-P(x-1,y)$

[0080] $G_y(x,y)=P(x,y+1)-P(x,y-1)$

[0081] $$G(x,y)=\sqrt{G_x(x,y)^2+G_y(x,y)^2}$$

[0082] $$\alpha(x,y)=\tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right)$$

[0083] 式中$G_x(x,y)$、$G_y(x,y)$和$P(x,y)$分别表示输入图像中像素点(x,y)处的水平方向梯度、垂直方向梯度和强度值，$G(x,y)$和$\alpha(x,y)$则表示像素点(x,y)处的梯度幅值和梯度方向。

[0084] 步骤003.针对该训练数据的超像素集中的各个超像素，根据RGB-D训练数据集合中包含的基准标注信息(Ground Truth)，获取该各个超像素分别对应的类别标签。

[0085] NYU Depth v2及SUN3D作为国际上权威的RGB-D室内场景数据集，其都包含有经人工稠密地标注的基准标注信息(Ground Truth)，为了后续构造语义标签池，即获取训练随机决策森林的训练样本，需将基准标注信息(Ground Truth)的类别标签映射至训练数据的超像素集中。映射过程的核心宗旨是保证映射至每个超像素的类别标签具备正确性和唯一性。但即便采用较为适应室内复杂环境的基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法改进方案，仍无法保证过分割区域的边缘完全吻合于室内场景图像中物体的真实边缘，这将导致所生成的某些超像素涵盖了两类甚至更多类别的目标，由基准标注信息(Ground Truth)映射后将导致这些超像素所对应的类别标签不唯一，违背了唯一性的准则。对此类由现有过分割方法无法避免的误差，秉承最大限度地降低误分割对区域级语义标签推断部分影响的原则，本发明在映射过程中采取如下折中方案：经基准标注图像(Ground Truth)映射后，若某超像素包含多种既定的类别标签，语义标注系统将判定该超像素的正确标签为占该超像素中像素数目比例最大的类别标签。

[0086] 步骤004.针对该训练数据的超像素集中各个超像素分别对应的类别标签、多模态特征向量，分别整合构成分别对应于各个超像素的各个条目，并整合该训练数据中全部超像素分别所对应的各个条目，构成该训练数据的超像素集对应的语义标签池。

[0087] 步骤005.将获得的该训练数据的超像素集对应的语义标签池作为训练样本，训练随机决策森林分类器。

[0088] 随机决策森林分类器是一种被广泛应用于各类计算机视觉任务的分类器，其通过建立很多的决策树，组成一个决策树的森林，通过多棵树的判决结果进行决策。随机决策森林分类器的构建包括三个基本步骤：

[0089] 步骤E01.随机决策森林分类器对样本数据进行自举重重采样，组成多个样本集，自举重重采样指每次从原有的所有训练样本中有放回地随机抽取同等数量的样本。

[0090] 步骤E02.用每个重采样样本集作为训练样本构造一个决策树，在构造决策树的过程中，每次从所有候选特征中随机抽取一定数量的特征，作为当前节点下决策的备选特征，从这些特征中选择最好地划分训练样本的特征。

[0091] 步骤E03.得到所需数目的决策树后，随机决策森林分类器对这些树的输出进行投票，以得票最多的类作为随机决策森林分类器的决策结果。

[0092] 并且，本发明中的随机决策森林分类器采用OpenCV计算机视觉库中的开源代码实现。对于随机决策森林分类器参数在NYU Depth v2与SUN3D室内场景数据集中的设置，本发明中采用如下方案：对于NYU Depth v2与SUN3D室内场景数据集，最大决策树深度分别设置为100和50，最大决策树数目分别设置为1000和500，决策树每个非叶子节点可选择备选特征维数均设置为10。

[0093] 步骤006.采用步骤001中基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法，针对查询图像进行过分割，获取该查询图像中的超像素，形成查询图像的超像素

集;并按步骤002中的方法,根据查询图像和对应的深度图像,针对该查询图像的超像素集中的各个超像素,分别求取对应超像素的多模态特征向量,构成该查询图像的超像素集对应的语义标签池。

[0094] 步骤007.采用已经训练的随机决策森林分类器,针对该查询图像的超像素集中的超像素进行语义标签推断,获得对应该查询图像的区域结构粗粒度级别标注图像。

[0095] 步骤008.采用改进型像素级稠密CRFs(Conditional Random Fields)概率图模型,针对获得对应该查询图像的区域结构粗粒度级别标注图像进行标签求精,获得对应该查询图像的细粒度级别标注图像。

[0096] 细粒度级别像素级语义标签求精部分本质上是一个全局求精的过程,目的在于为从粗粒度的区域级语义标签推断中获得的粗粒度语义标注引入全局上下文约束。随着近年稠密CRFs概率图模型的兴起,在此基础上亦涌现出了很多优秀的图模型构造方案。至于上下文推断则是指为已构建的能量函数进行优化求解,推断的效率和准确性是评判推断算法是否优秀的主要标准。现如今,随着图模型构造的日趋复杂化,某些传统的推断算法在效率上已显得难以满足实际的需要。

[0097] 本发明中细粒度像素级语义标签求精部分继承并发展了Krähenbühl等提出的基于高斯边界势能的稠密CRFs概率图模型的构建和推断策略[KR HENB HL P,KOLTUN V.Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials [C]//Advances in Neural Information Processing Systems(NIPS),2011.]。后者主要探讨的是一种稠密全连通CRFs概率图模型(其成对项势能是由高斯核线性组合)的建立及对应的高效推断算法。本发明中细粒度的像素级语义标签求精部分在上述稠密CRFs模型中引入几何深度信息以辅助优化求精,其本质目的在于为上下文推断过程引入更为丰富且不受室内光照条件影响的上下文信息,最终使得通过几何深度信息优化后的细粒度语义标注结果相比较于仅借助原始RGB图像求精的方案,在视觉效果上显得更为细腻且标注准确率更高。

[0098] 其中,该改进型像素级稠密CRFs概率图模型的具体构建包括如下步骤:

[0099] 步骤B01.利用深度图像和PCL(Point Cloud Library)点云库,计算图像中每个像素点的法线向量信息,并将法线向量信息转换存储为法线向量图像。

[0100] 步骤B02.根据已有深度图像及法线向量图像,针对稠密CRFs概率图模型,以像素为图模型节点进行成对项势能的修正拓展,获得像素级稠密CRFs概率图模型,并获得该像素级稠密CRFs概率图模型的能量函数表达式,如下所示:

[0101] $$E(\boldsymbol{X}\mid\boldsymbol{I})=\sum_{i}\psi_{u}(x_i^{*})+\sum_{(i,j)}\psi_{p}(x_i,x_j)$$

[0102] $$\psi_{u}(x_i^{*})=\exp(\varphi(e_i,x_i^{*}))/(1+\exp(\varphi(e_i,x_i^{*})))$$

[0103] $$\varphi(e_i,x_i^{*})=\log P(e_i\mid x_i^{*})-\log P(e_i\mid\overline{x_i^{*}})$$

[0104] $\psi_p(x_i,x_j)=\mu(x_i,x_j)K(f_i,f_j)$

[0105] $$\boldsymbol{K}(\boldsymbol{f}_i,\boldsymbol{f}_j)=w^{(1)}\exp\left(-\frac{|\boldsymbol{p}_i-\boldsymbol{p}_j|^2}{2\theta_\alpha^2}-\frac{|\boldsymbol{I}_i-\boldsymbol{I}_j|^2}{2\theta_\beta^2}-\frac{|d_i-d_j|^2}{2\theta_\gamma^2}-\frac{|\boldsymbol{n}_i-\boldsymbol{n}_j|^2}{2\theta_\delta^2}\right)+w^{(2)}\exp\left(-\frac{|\boldsymbol{p}_i-\boldsymbol{p}_j|^2}{2\theta_\chi^2}\right)$$

[0106] 其中,E(X|I)表示能量函数;X表示随机场;I表示给定的图像;$\psi_u(x_i^{*})$为图像中第i

个像素点的一元项势能；$\varphi(e_i,x_i^*)$表示对数似然比；e_i表示图像中第i个像素；x_i^*表示图像中第i个像素最有可能属于的类别标签；$\overline{x_i^*}$表示除x_i^*以外的所有类别标签；$P(e_i|x_i^*)$、$P(e_i|\overline{x_i^*})$均为普通的条件概率表达式，其概率值通过所述随机决策森林分类器获取；x_i和x_j分别表示图像中第i个像素点、第j个像素点的类别标签；$\psi_p(x_i,x_j)$为图像中任意两像素点i,j间的成对项势能；$\mu(x_i,x_j)$为标签兼容性函数，本发明中使用波茨模型(Potts model)表示，即$\mu(x_i,x_j)=1,x_i\neq x_j$；$K(f_i,f_j)$为高斯核的线性组合，本发明中使用对比度敏感的两个高斯核的线性形式；f_i和f_j分别表示第i个像素点、第j个像素点的特征向量；p_i和p_j表示图像中任意两像素点i,j的坐标位置向量；I_i和I_j表示图像中任意两像素点i,j的RGB色彩通道向量；d_i和d_j表示图像中任意两像素点i,j的深度值；n_i和n_j表示任意两像素点i,j相应的表面法线向量；$w^{(1)}$和$w^{(2)}$为两个高斯核的权值系数；$\theta_\alpha,\theta_\beta,\theta_\gamma$和$\theta_\delta$用来控制任意两像素点i,j属于一类的可能性，其所在项被称为外观核；θ_x所控制的项称为平滑核，其目的在于消除粗粒度标注结果中的孤立区域，θ_x用于控制衡量孤立区域大小。

[0107] 外观核与平滑核对区域级语义标签推断部分产生的粗粒度标注结果的贡献较微妙。首先，在本发明中外观核负责平滑标注图像的纹理，但并不会使目标的边界变得模糊，反而使其更加拟合于实际的目标边界，这一方面要归功于RGB图像中的纹理特征贡献。由于室内光源的影响，RGB纹理特征常显得并不非常可靠；而几何深度信息凭借其独立于室内光照条件的特点，能在一定程度降低室内光照变化的影响。其次，平滑核利用其可消除粗粒度标注结果中孤立区域的能力，平滑处理粗粒度语义标注结果，同时消除噪声。

[0108] 本发明中的改进型像素级稠密CRFs概率图模型由像素节点的颜色向量、位置信息、深度值以及表面法线向量等信息定义。为了获取更具实际物理意义的表面法线向量信息，本发明根据摄像机内部参数，利用PCL点云库将深度图转换成所拍摄场景的三维点云结构表达，并将从三维点云获取的法线向量信息映射至二维平面，进而与RGB图像、深度图像一并形成具有更强判别能力的视觉特征，引导成对项依赖CRFs概率图模型的推断。

[0109] 步骤009.针对获得对应该查询图像的细粒度级别标注图像，采用内部递归式反馈机制进行标签求精，获得该查询图像的最终细粒度级别标注图像。

[0110] 该内部反馈机制是一种由模型输入到模型输出的递归式反馈，以改善细粒度像素级语义标签，提升系统稳定性；本发明所述内部递归式反馈机制为基于所述深度图像与法线向量图像修正拓展后的像素级稠密CRFs概率图模型，获得由输入到输出的内部递归式反馈机制，主要是由于仅通过一次细粒度的像素级语义标签求精步骤难以使所得细粒度语义标注结果达到最优。该机制的设置可确保在像素级的语义标签优化阶段对粗标注的求精效果达到较高水平，也使得细粒度语义标注结果趋于稳定，该内部递归式反馈机制的实现包括如下步骤：

[0111] 步骤C01.根据获得的像素级稠密CRFs概率图模型，针对查询图像对应粗粒度级别区域级语义标签推断部分生成的粗粒度语义标注进行细粒度级别的求精，更新获得该查询图像的细粒度级别标注图像。

[0112] 步骤C02.根据该查询图像对应区域级语义标签推断部分产生的过分割信息，将获得的该查询图像的细粒度级别标注图像中的类别标签映射回该查询图像的超像素集中，更新该查询图像的超像素集。

[0113] 步骤C03.根据该查询图像的超像素集中的类别标签和该查询图像对应区域级语义标签推断部分产生的过分割信息，更新该查询图像的区域结构粗粒度级别标注图像，并判断更新后的该查询图像的区域结构粗粒度级别标注图像与未更新前该查询图像的区域结构粗粒度级别标注图像对应的超像素语义标签是否一致，是则获得的该查询图像的细粒度级别标注图像作为该查询图像的最终细粒度级别标注图像；否则返回步骤C01。

[0114] 步骤010.根据获得该查询图像的最终细粒度级别标注图像，设计获得由粗粒度的区域级语义推断到细粒度的像素级语义求精的全局递归式反馈机制，将该查询图像的最终细粒度级别标注图像作为额外信息引入步骤001和步骤006中分别针对图像的过分割操作中，并根据该全局递归式反馈机制，返回步骤001依次执行各个步骤，且根据全局递归式反馈机制中的终止条件，获得该查询图像的最终标注图像。

[0115] 由粗粒度的区域级语义推断到细粒度的像素级语义求精的全局递归式反馈机制是联系区域级语义推断与像素级语义求精的核心。通过交替迭代的方式融合场景图像中不同区域层次的多模态信息，从一定程度上解决传统语义标注工作中难以合适地选择标注基元的问题。该像素级语义求精的全局递归式反馈机制的实现包括如下步骤：

[0116] 步骤D01.将获得查询图像的最终细粒度级别标注图像，作为一种三维彩色通道，针对步骤001和步骤006中分别对图像进行过分割操作的简单线性迭代聚类(SLIC)过分割算法，引入细粒度语义标注信息，将简单线性迭代聚类(SLIC)过分割算法的聚类中心扩充至8维(3维彩色RGB图像通道+2维位置信息通道+1维图像分层显著度信息通道+1维深度信息通道+1维细粒度语义标注信息通道)。

[0117] 步骤D02.根据全局递归式反馈机制，返回步骤001依次执行各个步骤，更新获得查询图像的最终细粒度级别标注图像，并根据全局递归式反馈机制中的终止条件，判断更新后查询图像的最终细粒度级别标注图像与更新前查询图像的最终细粒度级别标注图像是否至多有5%的像素标签不同，是则将该更新后查询图像的最终细粒度级别标注图像作为该查询图像的最终标注图像；否则返回步骤D01。

[0118] 本发明设计的基于RGB-D数据的室内场景语义标注方法，首先构建了一种基于RGB-D数据由粗到精全局递归式反馈的语义标注框架，并将整个语义标注框架划分为粗粒度的区域级语义标签推断与细粒度的像素级语义标签求精两大部分；与传统单一的区域级或像素级语义标注框架不同，该框架重新建立粗粒度区域级语义标注与细粒度像素级语义标注间的联系，通过引入一种合理的全局递归式反馈的机制，使粗粒度区域级的语义标注结果与细粒度像素级的语义标注结果交替迭代更新优化。通过这种方式较好地融合了场景图像中不同区域层次的多模态信息，从一定程度上解决传统室内场景语义标注方案中普遍存在的难以合适地选择标注基元的问题。其次，本发明具体设计了基于图像分层显著度导引的简单线性迭代聚类(SLIC)过分割算法，相较于传统的简单线性迭代聚类(SLIC)过分割算法，一定程度上解决了目前非监督过分割算法在杂乱的室内场景中难以得到具有较高边缘一致性超像素的现状。并且利用分层显著度具备的抗小范围高对比度模式的特点，将图像分层显著度引入简单线性迭代聚类(SLIC)过分割算法，即扩展简单线性迭代聚类(SLIC)过分割算法的聚类空间，有助于改善小范围高对比度模式对划分简单线性迭代聚类(SLIC)超像素过程中的不利影响，该类模式对超像素大小近似均匀的简单线性迭代聚类(SLIC)过分割算法及类似过分割算法影响很大。最后，本发明具体设计了像素级稠密CRFs概率图模

型，并在像素级稠密CRFs概率图模型中引入几何深度信息与内部递归式反馈机制。其中，具体设计的像素级稠密CRFs概率图模型深入挖掘了几何深度信息在室内场景语义标签上下文优化求精中的潜力，并且实验表明，通过在概率图模型中引入有效且可靠的几何深度信息，一定程度上抑制了室内光源对室内场景语义标注视觉效果的影响，并提升了语义标签的准确性。而内部递归式反馈机制，则通过引入稠密CRFs概率图模型输入到输出间的关系，用于基于稠密CRFs概率图模型改善细粒度像素级语义标签，同时也令细粒度像素级语义标签求精部分的标注结果更加稳定，最终产生视觉表现力更强、标注准确率更高的语义标注图像。

[0119] 上面结合附图对本发明的实施方式作了详细说明，但是本发明并不限于上述实施方式，在本领域普通技术人员所具备的知识范围内，还可以在不脱离本发明宗旨的前提下做出各种变化。

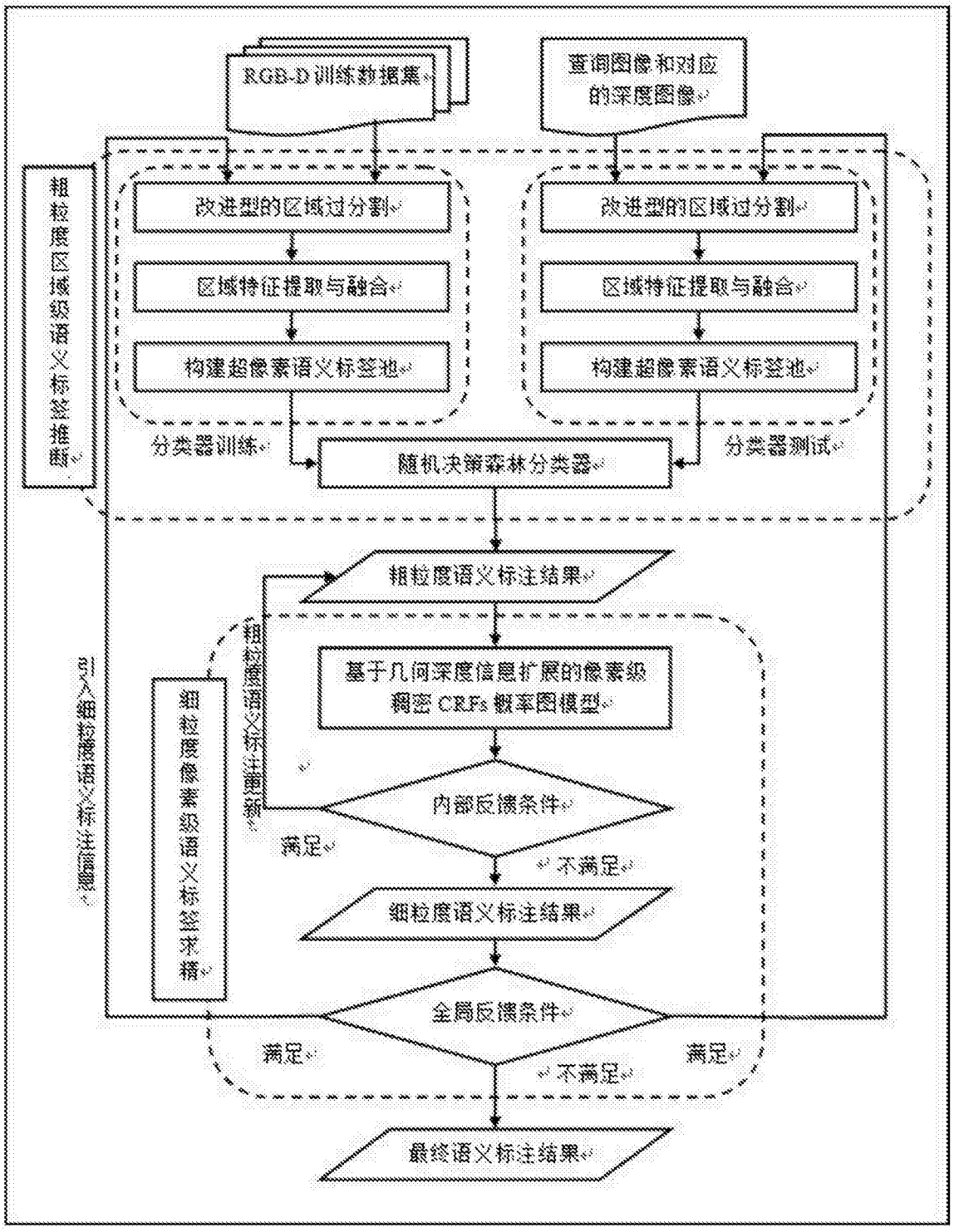

RGB-D 训练数据集
查询图像和对应的深度图像
改进型的区域过分割
区域特征提取与融合
构建超像素语义标签池
改进型的区域过分割
区域特征提取与融合
构建超像素语义标签池
粗粒度区域级语义标签推断
分类器训练
随机决策森林分类器
分类器测试
粗粒度语义标注结果
基于几何深度信息扩展的像素级稠密 CRFs 概率图模型
内部反馈条件
满足
不满足
细粒度语义标注更新
细粒度像素级语义标签求精
细粒度语义标注结果
全局反馈条件
满足
不满足
满足
最终语义标注结果

图1