



(12)发明专利

(10)授权公告号 CN 105844292 B

(45)授权公告日 2018. 11. 30

(21)申请号 201610158608.6

(22)申请日 2016.03.18

(65)同一申请的已公布的文献号
申请公布号 CN 105844292 A

(43)申请公布日 2016.08.10

(73)专利权人 南京邮电大学
地址 210003 江苏省南京市鼓楼区新模范
马路66号

(72)发明人 刘天亮 徐高帮 戴修斌 罗杰波

(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 朱小兵 朱楨荣

(51)Int. Cl.
G06K 9/62(2006.01)

(56)对比文件

- CN 102486827 A, 2012.06.06,
- CN 104599275 A, 2015.05.06,
- CN 104637045 A, 2015.05.20,
- CN 104809187 A, 2015.07.29,
- US 2013173258 A1, 2013.07.04,
- EP 2820567 A2, 2015.01.07,

审查员 刘坛首

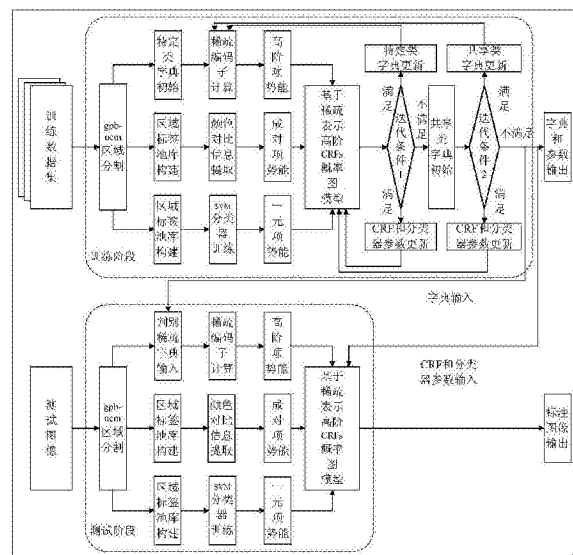
权利要求书2页 说明书10页 附图1页

(54)发明名称

一种基于条件随机场和二次字典学习的图像场景标注方法

(57)摘要

本发明公开了一种基于条件随机场和二次字典学习的图像场景标注方法,针对训练集图像进行超像素区域过分割,获取每幅图像的超像素过分割区域;提取各个超像素过分割区域的特征,并结合基准标注图像构建超像素标签池,利用超像素标签池训练出支持向量机分类器,计算超像素一元势能;计算相邻超像素成对项势能;借助训练集中全局性的过分割超像素区域的类别统计,以构建应用于类别统计直方图分类器作为分类成本,基于每一类超像素区域内关键点特征的稀疏表示的稀疏编码因子之和的直方图统计作为该CRF模型的高阶势能,分别以类字典、共享字典两种判别字典经过二次稀疏表示优化稀疏编码因子,更新字典和CRF参数与分类器参数;本发明提高了标注精度。



1. 一种基于条件随机场和二次字典学习的图像场景标注方法,其特征在于,包括以下步骤:

步骤A、针对训练集图像进行超像素区域过分割,获取每幅图像的超像素过分割区域;

步骤B、根据超像素过分割区域,提取各个超像素过分割区域的特征,并结合基准标注图像,构建超像素标签池,利用超像素标签池训练出支持向量机分类器,计算超像素一元势能;采用邻近区域对比颜色均值特征计算相邻超像素成对项势能;

步骤C、针对训练集图像,提取自顶向下特征,然后根据每一类的关键点训练初始化的每一类字典;针对每一个超像素过分割区域,以过分割区域的类别相对应的字典稀疏表示关键点的特征,并统计区域关键点稀疏系数得到类直方图作为高阶势能项,累加步骤B中的一元势能和成对项势能,并结合CRF参数、分类器参数得到CRF能量函数;对字典、CRF参数与分类器参数进行联合更新;最后用查询图像测试。

2. 根据权利要求1所述的一种基于条件随机场和二次字典学习的图像场景标注方法,其特征在于,所述步骤A具体如下:针对训练集图像,采用全局像素边缘方式检测待标注图像的目标区域边缘轮廓,根据目标边缘轮廓,划分超像素过分割区域,获取每幅图像的超像素过分割区域。

3. 根据权利要求1所述的一种基于条件随机场和二次字典学习的图像场景标注方法,其特征在于,所述步骤B中,计算一元项势能和成对势能的具体包括如下步骤:

步骤B1、根据每幅图像中的各个超像素过分割区域做如下操作:求取某个超像素过分割区域各个特征单元,然后对该超像素过分割区域各个特征单元分别进行归一化处理,获得该超像素过分割的各个归一化区域特征单元,最后将该超像素过分割区域的各个归一化特征单元进行拼接,构成对应于该超像素过分割区域的多模态特征向量;

步骤B2、根据训练集图像的基准真实标签,得到每幅图像中各个超像素过分割区域对应的类别真实标签;

步骤B3、整合每幅图像获得的超像素真实标签与多模态特征向量,构成对应于各个超像素条目,并整合所有训练图像某个标签对应的超像素所有的条目,构成训练集图像的超像素集对应的语义标签池;

步骤B4、将获得的超像素集对应的语义标签池作为训练样本,训练支持向量机分类器;

步骤B5、根据通过训练支持向量机分类器,获得每幅图像每个超像素对应于每个类别的得分,并将其作为一元势能;

步骤B6、根据每幅图像中相邻超像素之间的关系,计算出成对势能。

4. 根据权利要求3所述的一种基于条件随机场和二次字典学习的图像场景标注方法,其特征在于,所述步骤B1中的特征单元包括梯度核描述子、颜色核描述子和局部二值化核描述子。

5. 根据权利要求1所述的一种基于条件随机场和二次字典学习的图像场景标注方法,其特征在于,所述步骤C,具体包括如下子步骤:

步骤C1、针对训练集图像,计算每幅图像中关键点的128维sift自顶向下特征单元,并记录这些关键点图像中所在的位置;

步骤C2、根据训练集图像的基准真实标签,得到每幅图像中每个关键点对应的类别,将同一类别的所有关键点特征单元整合作为字典初始的特征样本,并利用训练集图像中的样

本空间,通过稀疏字典学习获得每个类别初始字典;

步骤C3、通过稀疏表示方法,根据图像超像素过分割区内的特征点对应类别分别采用相应类字典计算得到稀疏编码子,然后计算每一类区域的稀疏编码子绝对值之和构成类直方图,将类直方图作为CRF概率图模型的高阶项;

步骤C4、累加CRF一元势能项、成对势能和高阶势能项,采用以SSVM驱动的BCFW学习算法得到CRF参数与分类器参数,按照梯度下降法迭代更新类字典;

步骤C5、将学习到的每个类字典整合拼接成一个共享字典,保证类字典之间的正交性,并将共享字典正交化,用共享字典计算关键点的稀疏编码子,重新按照步骤C3~C4的方法,进行字典的二次更新和分类器参数与CRF参数的二次学习,得到一个共享字典、分类器参数与CRF参数;

步骤C6、针对查询图像进行超像素过分割,获取该查询图像中的超像素过分割区域,然后提取超像素区域特征,构建该查询图像的超像素对应的语义标签池;采用已经训练的支持向量机分类器,针对该查询图像的超像素进行语义推断,获得该查询图像的一元势能项和该查询图像的成对势能项;然后综合考虑字典的判别性和一致性,以类字典和共享字典整合后的字典作为测试样本的稀疏表示判别字典计算稀疏编码子;将步骤C4~步骤C5学习到的CRF参数求平均,分类器参数整合,作为整个测试过程中的CRF参数和分类器参数,预测最后的标注图像。

一种基于条件随机场和二次字典学习的图像场景标注方法

技术领域

[0001] 本发明涉及图像场景标注技术领域,特别是一种基于条件随机场和二次字典学习的图像场景标注方法。

背景技术

[0002] 视觉场景理解基本问题简单来说就是提取图像中的语义信息。对于提供的一个场景,不仅要通过其视觉上的颜色信息,还要根据先验知识推理出语义场景中存在的目标,及其空间位置关系与依赖关系以及场景布局的应用和场景中的各种复杂活动。对于人类来说,识别这些目标,并与场景中的信息联系在一起并不困难。场景理解的目标在于让机器能够有效的模拟人类特有的先天功能,通过训练样本的先验语义内容知识,提取相关的图像语义信息实现有效图像表示。利用这些图像表示实现诸如基于内容的图像检索,自动导航,机器探险,以及规避视觉模糊的广泛应用。

[0003] 为了提高场景理解能力,为图像场景中的目标进行语义标注无疑是一种很好的解决方案。场景标注基本目标是为给定的查询图像每一个像素或者区域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”处于语义层次的中层,在很多高层次语义推理中起到了承上启下的作用。语义标注相对于简单的图像分类要复杂得多,其在于不仅要为图像区域分配合理的标签,还要考虑图像相邻区域的结构,同时还要捕获长范围区域的交叉关系和高层的目标类别的先验知识。条件随机场或者马尔可夫随机场模型(MRF)在语义建模和标签推断方面取得突出成就。

[0004] 随机场模型广泛应用依据的事实:对应于同样标签的图像区域拥有相同的特征;同理,如果图像区域之间具有相似的位置或者特征空间,那么很有可能属于同一对象标签。文献[Koppula Hema S, Anand Abhishek and Joachims Thorsten. Semantic labeling of 3d point clouds for indoor scenes. Advances in neural information processing systems, 2011: 244-252]中采用Kinect融合产生三维点云,然后利用MRF进行建模;文献[Russell Chris, Kohli Pushmeet, Torr Philip HS. Associative hierarchical crfs for object class image segmentation. 2009 IEEE 12th International Conference on Computer Vision, 2009: 739-746.]中采用了分层MRF模型对像素级和超像素级进行联合推断;利用标准二阶条件随机场,结合SIFT特征, HOG特征, textboost特征, 共生关系构建势能项也是早期研究中经常采用的方式; [Krähenbühl Philipp, Koltun Vladlen. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. Advances in Neural Information Processing Systems, 2011: 109-117.]一文提出了一种成对项势能由高斯核的线性组合而成的稠密全连通CRF模型,该模型在像素级上表示,解决了由于某些超像素分割方法分割精度不高从而影响后续上下文推断的问题。然而这些二阶CRF中,图像特征表示皆基于局部超像素之间的关系,忽略了长范围的区域交叉关系和目标对象层的信息。为了解决这个问题,文献[Ladický L' ubor, Sturges Paul, Alahari Karteek, et

al.What,where and how many?combining object detectors and crfs.Computer Vision-ECCV 2010.Springer,2010:424-437.]结合目标检测结果和基于像素的CRF模型,文献[Yao Jian,Fidler Sanja,Urtasun Raquel.Describing the scene as a whole: Joint object detection,scene classification and semantic segmentation.Computer Vision and Pattern Recognition(CVPR),2012 IEEE Conference on,2012:702-709.]通过检测结果和形状先验知识和场景类别信息进一步提高算法精确度;文献[Russell Chris,Kohli Pushmeet,Torr Philip HS.Associative hierarchical crfs for object class image segmentation.2009 IEEE 12th International Conference on Computer Vision,2009:739-746.]提出基于分段和像素级的分层CRF模型,并引入高阶鲁棒性 P^N 势能,然而这些方法的主要缺陷是CRF高阶团需要预先定义,因此,在未知图像分割情况下,无法得到全局信息。

[0005] 为了解决上述问题,现有文献提出了结合二阶CRF能量函数与全局、基于BoF表示的自顶向下分类势能实现图像分类,由于稀疏表示能够提高基本的BoF模型的分类能力,并且能够从不同类别特征中捕获判别力的特征表示,文献中提出了结合字典学习的语义分割算法,利用稀疏表示理论对每一类的特征稀疏表达。通过每一次的分割,联合更新字典和CRF分类器参数。然而文献[Tao Lingling,Porikli Fatih,Vidal René.Sparse dictionaries for semantic segmentation.Computer Vision-ECCV 2014.Springer, 2014:549-564.]字典是通过所有类别训练得到,无法区分字典基元属于哪一类别,从而影响目标分类的正确性。

发明内容

[0006] 本发明所要解决的技术问题是克服现有技术的不足而提供一种基于条件随机场和二次字典学习的图像场景标注方法,基于条件随机场和二次字典学习的图像语义标注框架,提取自底向上区域级信息构建该条件随机场的一元势能和二元势能和自顶向下的类别级信息构建该条件随机场的高阶势能,对自顶向下类别信息引入类字典和共享字典两种字典优化分类类别信息,借助训练集中全局性的过分割超像素区域的类别统计,以构建应用于类别统计直方图的分类器作为分类成本,基于每一类超像素区域内关键点特征的稀疏表示的稀疏编码之和的直方图统计作为该CRF模型的高阶势能,并联合推断条件随机场CRF参数与分类器参数和字典。

[0007] 本发明为解决上述技术问题采用以下技术方案:

[0008] 根据本发明提出的一种基于条件随机场和二次字典学习的图像场景标注方法,包括以下步骤:

[0009] 步骤A、针对训练集图像进行超像素区域过分割,获取每幅图像的超像素过分割区域;

[0010] 步骤B、根据超像素过分割区域,提取各个超像素过分割区域的特征,并结合基准标注图像,构建超像素标签池,利用超像素标签池训练出支持向量机分类器,计算超像素一元势能;采用邻近区域对比颜色均值特征计算相邻超像素成对项势能;

[0011] 步骤C、针对训练集图像,提取自顶向下特征,然后根据每一类的关键点训练初始化的每一类字典;针对每一个超像素过分割区域,以过分割区域的类别相对应的字典稀疏

表示关键点的特征,并统计区域关键点稀疏系数得到类直方图作为高阶势能项,累加步骤B中的一元势能和成对项势能,并结合CRF参数、分类器参数得到CRF能量函数;对字典、CRF参数与分类器参数进行联合更新;最后用查询图像测试。

[0012] 作为本发明所述的一种基于条件随机场和二次字典学习的图像场景标注方法进一步优化方案,所述步骤A具体如下:针对训练集图像,采用全局像素边缘方式检测待标注图像的目标区域边缘轮廓,根据目标边缘轮廓,划分超像素过分割区域,获取每幅图像的超像素过分割区域。

[0013] 作为本发明所述的一种基于条件随机场和二次字典学习的图像场景标注方法进一步优化方案,所述步骤B中,计算一元项势能和成对势能的具体包括如下步骤:

[0014] 步骤B1、根据每幅图像中的各个超像素过分割区域做如下操作:求取某个超像素过分割区域各个特征单元,然后对该超像素过分割区域各个特征单元分别进行归一化处理,获得该超像素过分割的各个归一化区域特征单元,最后将该超像素过分割区域的各个归一化特征单元进行拼接,构成对应于该超像素过分割区域的多模态特征向量;

[0015] 步骤B2、根据训练集图像的基准真实标签,得到每幅图像中各个超像素过分割区域对应的类别真实标签;

[0016] 步骤B3、整合每幅图像获得的超像素真实标签与多模态特征向量,构成对应于各个超像素条目,并整合所有训练图像某个标签对应的超像素所有的条目,构成训练集图像的超像素集对应的语义标签池;

[0017] 步骤B4、将获得的超像素集对应的语义标签池作为训练样本,训练支持向量机分类器;

[0018] 步骤B5、根据通过训支持向量机分类器,获得每幅图像每个超像素对应于每个类别的得分,并将其作为一元势能;

[0019] 步骤B6、根据每幅图像中相邻超像素之间的关系,计算出成对势能。

[0020] 作为本发明所述的一种基于条件随机场和二次字典学习的图像场景标注方法进一步优化方案,所述步骤B1中的特征单元包括梯度核描述子、颜色核描述子和局部二值化核描述子。

[0021] 作为本发明所述的一种基于条件随机场和二次字典学习的图像场景标注方法进一步优化方案,所述步骤C,具体包括如下子步骤:

[0022] 步骤C1、针对训练集图像,计算每幅图像中关键点的128维sift自顶向下特征单元,并记录这些关键点图像中所在的位置;

[0023] 步骤C2、根据训练集图像的基准真实标签,得到每幅图像中每个关键点对应的类别,将同一类别的所有关键点特征单元整合作为字典初始的特征样本,并利用该样本空间通过稀疏字典学习获得每个类别初始字典;

[0024] 步骤C3、通过稀疏表示方法,根据图像超像素过分割区内的特征点对应类别分别采用相应类字典计算得到稀疏编码子,然后计算每一类区域的稀疏编码子绝对值之和构成类直方图,将类直方图作为CRF概率图模型的高阶项;

[0025] 步骤C4、累加CRF一元势能项、成对势能和高阶势能项,采用以SSVM驱动的BCFW学习算法得到CRF参数与分类器参数,按照梯度下降法迭代更新类字典;

[0026] 步骤C5、将学习到的每个类字典整合拼接成一个共享字典,保证类字典之间的正

交性,并将共享字典正交化,用共享字典计算关键点的稀疏编码子,重新按照步骤C3~C4的方法,进行字典的二次更新和分类器参数与CRF参数的二次学习,得到一个共享字典、分类器参数与CRF参数;

[0027] 步骤C6、针对查询图像进行超像素过分割,获取该查询图像中的超像素过分割区域,然后提取超像素区域特征,构建该查询图像的超像素对应的语义标签池;采用已经训练的支持向量机分类器,针对该查询图像的超像素进行语义推断,获得该查询图像的一元势能项和该查询图像的成对势能项;然后综合考虑字典的判别性和一致性,以类字典和共享字典整合后的字典作为测试样本的稀疏表示判别字典计算稀疏编码子;将步骤C4~步骤C5学习到的CRF参数求平均,分类器参数整合,作为整个测试过程中的CRF参数和分类器参数,预测最后的标注图像。

[0028] 本发明采用以上技术方案与现有技术相比,具有以下技术效果:

[0029] (1) 本发明结合条件随机场和判别的稀疏字典学习成本的语义标注框架,相比于传统的基于K-Means字典学习的BoF模型,其更能捕获判别力的类别信息;

[0030] (2) 本发明同时引入类别字典和共享字典的方式,将每一类的字典学习建立在特定类的样本特征空间,并考虑特定类字典之间的共性,使每一类的分割更加具有判别力,同时为了能更好的区别类别之间的影响,引入共享字典,使得所有类别之间的一致性得以保留,提高准确性。

附图说明

[0031] 图1是基于条件随机场和二次字典学习的场景语义标注方法的流程示意图。

具体实施方式

[0032] 下面结合附图对本发明的技术方案做进一步的详细说明:

[0033] 如图1所示,本发明设计基于条件随机场和二次字典学习的场景语义标注框架,在实际应用过程当中,利用基本二阶CRF的语义标注框架,并基于字典学习得到的稀疏编码子构成的直方图作为CRF扩展的高阶项进行场景图像的语义标注。该语义标注框架是由自底向上区域级组成的二阶势能和由自顶向下类别级信息组成的高阶势能构成,包括如下步骤:

[0034] 步骤A. 针对训练集图像进行超像素过分割,获取每幅图像的超像素过分割区域;

[0035] 步骤A1. 对于每一个像素,以其所在位置 (x, y) 为圆心画圆,并以倾斜角 θ 的直径分成两个半圆,并将两个半圆区域中的亮度值分别做直方图统计,最终以由两个亮度直方图 g 和 h 计算到的 x^2 距离表示像素的梯度:

$$[0036] \quad G(x, y, \theta) = \chi^2(g, h) = \frac{1}{2} \sum_k \frac{(g_\theta(k) - h_\theta(k))^2}{g_\theta(k) + h_\theta(k)} \quad (1)$$

[0037] 其中, $G(x, y, \theta)$ 为坐标 (x, y) 以角度 θ 切割半圆得到的方向梯度, $g_\theta(k)$ 表示其中一个半圆中亮度值为 k 的直方图统计, $h_\theta(k)$ 为另一半圆中亮度值为 k 的直方图统计, $\chi^2(g, h)$ 表示直方图 g 和 h 的 x^2 距离。

[0038] 步骤A2. 将每一副图像分解为4个不同的通道,分别为CIE Lab空间中的亮度通道、颜色a通道、颜色b通道和纹理通道,并且结合多个尺度的直径 σ ,获得多尺度像素梯度幅度:

$$[0039] \quad gPb(x, y, \theta) = mPb(x, y, \theta) = \sum_s \sum_t \alpha_{i,s} G_{i,\sigma(i,s)}(x, y, \theta) \quad (2)$$

[0040] 其中, i 表示每一个通道, s 表示每一个尺度, 系数 α 表示每一个不同直径条件下的每一个特征通道的权重, 是针对F-measure梯度下降得到。最后对于该gPb进行sigmoid函数转换, 使其值处在0-1之间, 作为该像素是否为边缘的概率。

[0041] 步骤A3. 对于每一个像素, $E(x, y) = \max_{\theta} gPb(x, y, \theta)$, 计算8个设定的角度 $\theta \in [0, \pi]$ 对应的边缘概率取其最大值; 接着使用分水岭转换技术, 将输入转换成一系列区域和边界。

[0042] 步骤A4. 通过检测到的边缘生成的区域几何组合成超度量边界图 (Ultrametric Contour Map, UCM), 然后通过阈值 k 调节输出不同细节的图像轮廓, 同时可以得到不同数量的目标超像素过分割区域。

[0043] 步骤B. 根据超像素分割后的区域, 提取各个超像素过分割区域的特征单元, 并结合基准标注图像, 构建超像素标签池, 采用支持向量机计算超像素一元势能; 采用邻近区域对比颜色均值特征计算相邻超像素成对项势能, 以约束相邻区域之间的平滑关系;

[0044] 步骤B1. 根据训练集图像每幅图像中的各个超像素过分割区域做如下操作: 求取某个超像素过分割区域各个特征单元, 然后对该超像素过分割区域各个特征单元分别进行归一化处理, 获得该超像素过分割的各个归一化区域特征单元, 最后将该超像素过分割区域的各个归一化特征单元进行拼接, 构成对应于该超像素过分割区域的多模态特征向量。

[0045] 传统的直方图特征如HOG和SIFT特征在图像分类或者图像分割中得到广泛应用。HOG特征是将单个像素属性值离散成bin然后计算某个图像块中的离散直方图, 两个块之间的相似性就通过直方图计算得到。SIFT特征计算关键点的像素特征, 然后通过Bag of Words方式将匹配到直方图中。然而两种特征都存在量化误差, 不能准确的衡量局部特征的相似性。目前有学者提出的采用高效匹配核 (EMK) 的方式将局部特征匹配到相应的低维空间并均值化形成区域级特征, 能够有效提高识别精度; 同时从核视觉角度重新定义了SIFT、HOG特征, 核描述子避免了直方图特征提取过程中像素的离散化, 并且能够生成不同类型丰富的基于块层面的像素属性特征, 两个块之间的相似性基于一个匹配核, 其优点在于两个块之间的像素属性的连续性。本发明首先对图像分割以2个像素差, 大小为 16×16 的块, 然后在块基础上提取梯度核描述子, 颜色核描述子, 局部二值化核描述子, 然后通过EMK匹配方式分别将块描述子映射到200个words中, 共计600维的特征。以梯度核描述子为例, 基于核的特征提取如下:

[0046] 首先将图像转换成灰度图, 然后计算每个像素点的梯度, 梯度核描述子 F_{grad} 通过每个像素梯度相似性函数 k_0 组成:

$$[0047] \quad F_{grad}^t(Z) = \sum_{i=1}^{d_0} \sum_{j=1}^{d_s} \alpha_{ij}^t \left\{ \sum_{z \in Z} \tilde{m}_z k_0(\tilde{\theta}_z, p_i) k_s(z, q_j) \right\} \quad (3)$$

[0048] 其中, Z 表示图像块, $z \in Z$ 表示图像块中的像素之间的2维相互关系(归一化到 $[0, 1]$) \tilde{m} 和 $\tilde{\theta}_z$ 是归一化的像素梯度的幅度和方向值。方向核 $k(\tilde{\theta}_z, \tilde{\theta}_x) = \exp(-\gamma_0 \|\tilde{\theta}_z - \tilde{\theta}_x\|)$ 计算梯度方向的相似性, 位置高斯核 $k_s(z, x) = \exp(-\gamma_s \|z - x\|^2)$ 衡量两个像素空间的相近关系。

$\{p_i\}_{i=1}^{d_0}$ 和 $\{q_j\}_{j=1}^{d_s}$ 为从其支撑区域一致采样结果, d_0 和 d_s 是采样基准向量的数量。 α_{ij}^t 是通

过核主成分分析计算得到的投影系数,其余的核描述子也是通过类似的形式从像素级相似性函数计算得到。

[0049] 步骤B2.根据训练集图像的基准真实标签,得到每幅图像中每个超像素对应的类别真实标签;

[0050] 针对该训练集图像的超像素集中的各个超像素过分割,根据训练集图像集合中包含的基准标注信息,获取各个超像素分别对应的类别标签。

[0051] 本发明中采用的数据集是图像标注中经典的数据集GRAZ,其都包含人工稠密地标注的基准标注信息,为了后续构造语义标签池,即获取训练SVM分类器的训练样本,需要将基准标注信息的类别标签映射至训练集图像的超像素中。本发明采用的映射原则是确保每个超像素标签的唯一性和准确性,然而由于前期的超像素过分割结果不能保证完全按照目标边缘进行精确分割,导致所生成的超像素区域中可能包含多个标签类别,违背映射原则,为了最大化标签分配正确性,因此本发明采用折中方案:对于某一个超像素,若其中包含的某一类别像素最大,则认定该超像素为此类别标签。具体步骤如下:

[0052] 步骤01.对基准标注图像按照像素RGB值的不同,分配一个不同的标签(如1,2,3,⋯,n,无效类为0);

[0053] 步骤02.将过分割图像与步骤01得到的图像像素一一对应;

[0054] 步骤03.将每一个超像素过分割区域的标签按照类别统计像素个数,像素个数最多的类别即为该超像素过分割区域的类别标签。

[0055] 步骤B3.整合每幅图像获得的超像素过分割区域真实标签与多模态特征向量,构成对应于各个超像素条目,并整合所有训练图像某个标签对应的超像素所有的条目,构成训练集图像的超像素过分割区域集对应的语义标签池;

[0056] 步骤B4.将获得的超像素过分割区域集语义标签池作为训练样本,训练支持向量机分类器;

[0057] 支持向量机方法通过核函数的方式,不需要非线性映射的显示表达方式将样本空间映射到一个高维乃至无穷维的特征空间中,使得在原来样本空间中非线性可分的问题转换为特征空间中的线性可分问题。本发明中,实验采用的是LIBSVM模式识别软件包,主要参数设置为:采用自定义核函数,实现方法由vl_feat工具包中的vl_alldist函数提供,惩罚系数c设置为1。

[0058] 步骤B5.根据通过训练得到的支持向量机分类器,利用libsvm中的预测函数,计算得到获得每幅训练图像每个超像素对应于属于每一类的概率和超像素类别,然后通过每一类的概率的对数似然得到属于每一类的得分,将得分作为每一类别的势能值。

[0059] 步骤B6所述成对项的势能用于度量相邻超像素过分割区域之间的平滑关系,具体势能计算公式:

$$[0060] \quad \psi_{ij}^p(x_i, x_j; D) = \frac{L(i, j)}{1 + \|f(i) - f(j)\|} \delta(x_i \neq x_j) \quad (4)$$

[0061] 其中, $f(i)$ 是超像素过分割区域LUV空间的颜色平均值, $L(i, j)$ 是超像素 i 和 j 公共边界的长度。

[0062] 步骤C.针对训练图像,提取自顶向下特征,将同一类别的所有关键点特征单元整合作为每一类初始字典的特征样本,并通过稀疏表示方式学习得到初始化的每一类字典;

针对每一个超像素过分割区域,以过分割区域的类别相对应的字典稀疏表示关键点的特征,并统计区域关键点稀疏系数得到类直方图作为高阶势能项,以超像素过分割区域为CRF概率图模型节点,以相邻超像素过分割区域的边缘作为CRF模型的边,结合CRF参数、分类器参数和各势能项得到能量函数,使其能量最小化得到预测标签。为了学习CRF能量函数中的字典和参数,针对经典结构化SVM (SSVM) 的不适用参数和字典的都未知的问题,分别采用梯度下降法、以结构化支持向量机 (SSVM) 驱动的BCFW学习算法联合更新字典和CRF参数与分类器参数。为了提取到类字典中的相关性,以类字典正交化构成初始化共享字典,按照上述方式,重新训练参数和迭代计算共享字典,最后以类字典、共享字典结合方式作为测试数据集编码子的字典。包括如下子步骤:

[0063] 步骤C1. 针对训练集图像,计算每幅图像中关键点的128维sift自顶向下特征单元,并记录这些关键点图像中所在的位置;

[0064] 步骤C2. 根据训练集图像的基准真实标签,得到每幅图像中每个关键点对应的类别,将同一类别的所有关键点特征单元整合作为字典初始的特征样本,并利用该样本空间通过稀疏字典学习获得每个类别初始字典;

[0065] 步骤C3. 通过稀疏表示方法,根据图像超像素过分割区内的特征点对应类别分别采用相应类字典计算得到稀疏编码子,然后计算每一类区域的稀疏编码子绝对值之和构成类直方图,将类直方图作为CRF概率图模型的高阶项,具体做法如下:

[0066] 步骤1,假设 $D_l \in R^{F \times K_l}$ 是稀疏字典学习得到的由 K_l 个视觉单词组成的类别 l 的类字典,每个视觉单词已经归一化。对应于每个属于 l 类的特征描述子都采用类别 l 的类字典进行稀疏编码,也即求如下问题:

$$[0067] \quad \alpha_j(D_l) = \arg \min_a \left\{ \frac{1}{2} \|f_j - D_l a\| + \lambda \|a\|_1 \right\} \quad (5)$$

[0068] 式中, f_j 表示关键点 j 的特征描述子, D_l 表示关键点 j 所属类别对应的类字典, a 表示稀疏编码子。 l 的取值为1至 L 类, L 为训练集图像总的类别数。

[0069] 步骤2. 对所有的特征描述子都经过步骤1的稀疏编码,然后构建每一类别稀疏编码子直方图:

$$[0070] \quad h_l(X, D_l) = \sum_{j \in S} \alpha_j(D_l) \delta(x_j = l) = \sum_{i \in V} \sum_{j \in S_i} \alpha_j(D_l) \delta(x_i = l) \quad (6)$$

[0071] 式中, S_i 是超像素 i 中特征关键点集合。由于每一类直方图跟分割图像 X 和每一类字典都相关,假设线性分类器每一类的参数 $w_l \in R^{K_l}$, 最终CRF能量函数由一元势能项,二元势能项和高阶势能项及对应参数组成:

$$[0072] \quad E(X, I) = \lambda_1 \sum_{i \in V} \phi_i^U(x_i, I) + \lambda_2 \sum_{(i,j) \in E} \phi_{ij}^P(x_i, x_j, I) + \sum_{l=1}^L w_l^T h_l(X, D_l) \quad (7)$$

[0073] 步骤C4. 累加CRF各个势能项,由于能量函数中,参数和字典都是未知,经典的SSVM算法并不适用此问题,为了学习参数和字典,采用以SSVM驱动的BCFW算法学习得到CRF参数与分类器参数,按照梯度下降法迭代更新类字典,具体做法如下:

[0074] 步骤1. 给定图像 I , CRF参数 λ_1, λ_2 , 分类器参数 $\{w_l\}_{l=1}^L$, 类字典 D_l , 我们的目标是计算标签 X^* , 使得能够最大化条件概率, 即,

$$[0075] \quad X^* = \arg \max_X P(X|I) = \arg \min_X E(X, I)$$

[0076] 由于每个超像素过分割区域内各个类别稀疏编码子累加与每个类别各个超像素过分割区域中的稀疏编码子累加先后顺序的调整对结果没有影响,由L种类别组成的图像由每一个类别直方图组成,可以表示成:

$$[0077] \quad \sum_{l=1}^L \mathbf{w}_l^T h_l(X, \mathbf{D}_l) = \sum_{l=1}^L \mathbf{w}_l^T \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{S}_l} \alpha_j(\mathbf{D}_l) \delta(x_i = l) = \sum_{i \in \mathcal{V}} \sum_{l=1}^L \mathbf{w}_l^T \alpha_l(\mathbf{D}_l) \delta(x_i = l) \quad (8)$$

[0078] 因此能量函数可以合并成只有一元项和二元项的形式:

$$[0079] \quad E(X, I) = \sum_{i \in \mathcal{V}} (\lambda_1 \phi_i^U(x_i, I) + \sum_{l=1}^L \mathbf{w}_l^T \alpha_l(\mathbf{D}_l)) + \lambda_2 \sum_{(i,j) \in \mathcal{E}} \phi_{ij}^P(x_i, x_j, I) \quad (9)$$

[0080] 因此可以通过近似推断算法如 α -expansion算法最小化能量函数。

[0081] 步骤2. 给定训练集图像集 $\{I^n\}_{n=1}^N$ 和ground truth分割图 $\{X^n\}_{n=1}^N$,当能量函数中的类字典已知固定,只有CRF参数变化时,可以采用SSVM框架学习得到CRF参数 λ_1, λ_2 ,分类器参数 $\{\mathbf{w}_l\}_{l=1}^L$ 。首先重写能量函数:

$$[0082] \quad E(X, I) = \mathbf{W}^T \phi(X, I, \mathbf{D}_l)$$

[0083] 其中

$$[0084] \quad \mathbf{W} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mathbf{w}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{w}_L \end{bmatrix} \quad \phi(X, I, \mathbf{D}_l) = \begin{bmatrix} \sum_{i \in \mathcal{V}} \phi_i^U(x_i, I) \\ \sum_{(i,j) \in \mathcal{E}} \phi_{ij}^P(x_i, x_j, I) \\ \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{S}_1} \alpha_j \delta(x_i = 1) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{S}_L} \alpha_j \delta(x_i = L) \end{bmatrix}$$

[0085] 然后,可以寻找一组范数较小的参数 \mathbf{W} 使得基准标注图像与任何推断出的分割相差 $\Delta(\hat{X}^n, X^n)$,也就是解决下述问题:

$$\min \frac{1}{2} \|\mathbf{W}\|^2 + \frac{C}{N} \sum_{n=1}^N \xi_n$$

[0086]

$$\text{s.t.} \quad \forall n \in \{1, \dots, N\}, \forall \hat{X}^n \\ \mathbf{W}^T \phi(\hat{X}^n, I^n, \mathbf{D}_l) - \mathbf{W}^T \phi(X^n, I^n, \mathbf{D}_l) \geq \Delta(\hat{X}^n, X^n) - \xi_n,$$

[0087] 这里 $\{\xi_n\}$ 是松弛变量以约束意外变化 $\Delta(\hat{X}^n, X^n) = \gamma \sum_{l=1}^L \frac{1}{N_l} \sum_{i \in \mathcal{V}} \delta(\hat{x}_i^n = x_i^n) \delta(x_i^n = l)$ 。

[0088] 上式是一个二次规划问题,如果当字典是固定已知的,可以用经典的SSVM方法分两步解决上述问题:

[0089] 步骤1. 给定 \mathbf{W} 求得最意外的分割 $\bar{X}^n = \arg \min_{\hat{X}} \{\mathbf{W}^T \phi(\hat{X}, I^n, \mathbf{D}_l) - \Delta(\hat{X}, X^n)\}$;

[0090] 步骤2. 给定约束的 \bar{X}^n , 更新W值。

[0091] 然而实际情况中, 我们的模型问题是字典和W都是未知, 更重要的是能量函数与字典之间的关系并非线性, 并且能量函数与字典的关系是非显性的。因此传统的SSVM方法不适用于我们的问题。因此我们分开更新W和类字典, 重写上述问题:

$$[0092] \quad J(W, D_l) = \frac{1}{2} \|W\|^2 + \frac{C}{N} \sum_{i=1}^N [W^T \phi(X^n, I^n, D_l) - \min\{W^T \phi(\hat{X}^n, I^n, D_l) - \Delta(\hat{X}^n, X^n)\}]$$

[0093] 当固定W的时候, 采用梯度下降法计算类字典, 关键是要求出J对D_l的导数, 由于J和字典是隐性关系, 可以通过链式关系, 求出 $\partial J / \partial \alpha$ 和 $\partial \alpha / \partial D_l$ 得到 $\partial J / \partial D_l$

$$[0094] \quad \frac{\partial J^n}{\partial D_l} = \sum_{i \in V} \sum_{j \in S_i} \left(\frac{\partial J^n}{\partial \alpha_i} \right) \frac{\partial \alpha_i}{\partial D_l}$$

[0095] 针对满足 $\alpha_j(D_l) = \arg \min_a \left\{ \frac{1}{2} \|f_j - D_l a\| + \lambda \|a\| \right\}$ 式子, 需要满足梯度为0的条件, 因此得到

$$[0096] \quad D_l^T (D_l a - f_j) = -\lambda \text{sign}(a)$$

[0097] 展开得到

$$[0098] \quad D_l^T D_l a - D_l^T f_j = -\lambda \text{sign}(a)$$

[0099] 两边对D_l求导

$$[0100] \quad D_{l(\wedge)}^T D_{l(\wedge)} \frac{\partial(a)}{\partial D_l} + \frac{D_{l(\wedge)}^T D_{l(\wedge)}}{\partial D_l} a - \frac{\partial(D_{l(\wedge)}^T f_j)}{\partial D_l} = 0$$

[0101] 其中 \wedge 表示编码不为0的列, $\bar{\wedge}$ 表示编码为0的列, 化简求得

$$[0102] \quad \frac{\partial(a_{\wedge})}{\partial D_l} = (D_{l(\wedge)}^T D_{l(\wedge)})^{-1} \left(\frac{\partial(D_{l(\wedge)}^T f_j)}{\partial D_l} - \frac{D_{l(\wedge)}^T D_{l(\wedge)}}{\partial D_l} a_{\wedge} \right)$$

$$[0103] \quad \frac{\partial J^n}{\partial a_{\wedge}} = w_{x_j} - w_{\bar{x}_j}$$

[0104] 为了简化计算, 引入辅助向量, 令,

$$[0105] \quad z_{\wedge} = 0, \quad z_{\bar{\wedge}} = (D_{l(\bar{\wedge})}^T D_{l(\bar{\wedge})})^{-1} \frac{\partial J^n}{\partial a_{\bar{\wedge}}}$$

[0106] 同时表示 $Z = [z_1, z_2, \dots, z_m]$, m为超像素的个数, 最后Jⁿ相对于类字典的导数为

$$[0107] \quad \frac{\partial J^n}{\partial D_l} = -D_l Z A^T + (F - D_l A) Z^T \quad (10)$$

[0108] 然后 $D_l = D_l - \tau \frac{\partial J}{\partial D_l}$ 更新字典, 其中 τ 取值 $1e-6$ 。其他类别字典也如同上式所示, 每次迭代更新完类字典后, 固定每一个类字典, 同时采用BCFW算法更新W参数, 最后求得每一类字典和W参数。

[0109] 步骤C5. 实际过程中, 通过学习得到每一类字典, 能够有效表示每一类信息, 但是每个类字典的相关约束并没有考虑, 使得每个类字典之间的相关性丢失, 为了提取类字典之间的相关性, 将学习到的每个类字典整合拼接成一个共享字典, 并保证类字典之间的

正交性,即式子(5)中的 $D_1=D=[D_1, D_2, \dots, D_L]$,并将 D 正交化,用共享字典计算关键点的稀疏编码,重新并按照步骤C3~C4方法,进行字典的二次更新和分类器参数与CRF参数的二次学习,得到一个共享字典 D 和分类器参数与CRF参数;

[0110] 步骤C6. 针对查询图像进行超像素过分割,获取该查询图像中的超像素过分割区域,然后按照步骤B方法提取超像素区域特征,构建该查询图像的超像素对应的语义标签池;采用已经训练的支持向量机分类器,针对该查询图像的超像素进行语义推断,获得该查询图像的一元势能项和该查询图像的成对势能项;然后综合考虑字典的判别性和一致性,以类字典和共享字典整合后的字典 $D_c=[D_1, D]$ 作为测试样本的稀疏表示判别字典计算稀疏编码。测试过程中,字典整合后的维度与模型中应有的字典维度不一致,使得学习到分类器系数与稀疏编码子的实际维度不一致,实际实验过程中,我们将两次学习到的CRF参数求平均,分类器参数整合,作为整个测试过程中的CRF参数和分类器参数,预测最后的标注图像。

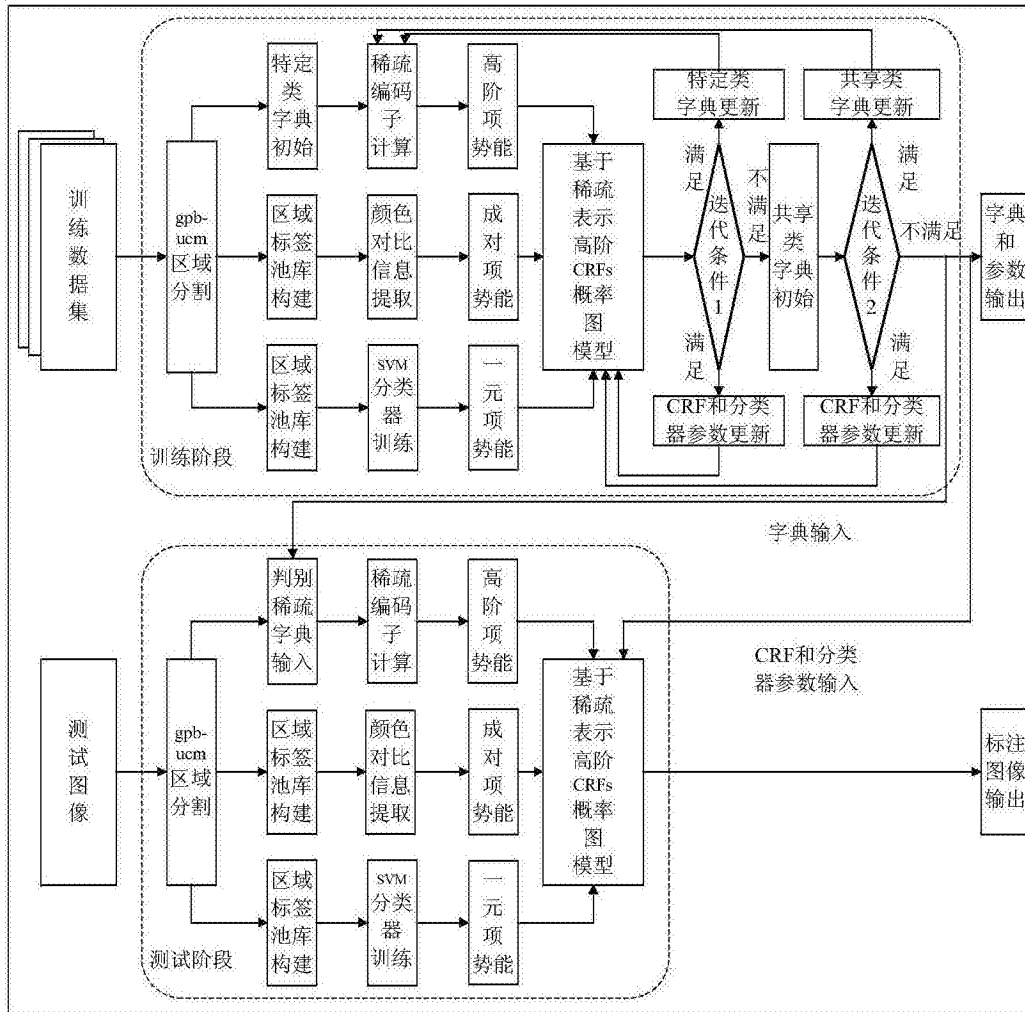


图1