

图像与视频自动标注最新进展

尹文杰 韩军伟 郭雷

(西北工业大学自动化学院 西安 710129)

摘要 近年来,图像与视频自动标注技术成为多媒体信息处理领域的一个研究热点并且发展迅速。重点介绍此领域研究的最新进展。将这些新方法分为两类:基于学习的标注方法和基于搜索的标注方法,分别介绍了各个算法的基本思想和优缺点。然后,介绍了一些目前流行的商业的和研究性的图像检索及标注系统。最后,提出了自动标注技术的进一步研究方向。

关键词 图像与视频自动标注,基于学习的标注,基于搜索的标注

中图法分类号 TP37 文献标识码 A

New Survey on Automatic Image and Video Annotation

YIN Wen-jie HAN Jun-wei GUO Lei

(School of Automation, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract In recent years, automatic image and video annotation has been becoming a popular research topic and is advancing very fast. This paper provided a new survey to this technique by reviewing a variety of literatures published lately. The reviewed work was classified into two categories: learning-based approaches and search-based approaches. We firstly summarized their basic ideas, advantages, and drawbacks, respectively. Afterwards, we introduced a few commercial systems and research prototypes of retrieval and annotation. Finally, a potentially promising idea was suggested.

Keywords Automatic image and video annotation, Learning-based approaches, Search-based approaches

1 引言

随着多媒体技术、计算机技术的迅速发展以及互联网的广泛普及,图像和视频成为一种重要的信息资源。如何在海量的信息中快速有效地查找到用户需要的图像和视频,已经成为一个研究热点^[1]。目前,许多商业搜索引擎,如 Google, MSN, Yahoo! 等均利用基于语义的搜索方法来检索相关图像或视频。为了解决低层特征与高层语义概念间的“语义鸿沟”问题^[2],图像与视频自动标注成为关键技术。图像/视频标注就是为其分配相对应的文本关键字,以反映它的内容^[3,4],该技术结合了基于文本的检索与基于内容的检索两者的优点,有效地解决了语义鸿沟问题。标注的方法有两种:人工标注和自动标注。由于数码相机产品的盛行以及网络的迅速发展,图像和视频数据大量增加,人工标注图像和视频的方法需要花费大量时间和人力物力,往往不切实际^[5]。因此,近几年来,出现了大量关于图像和视频自动标注技术的研究。

本文第 2 节将图像和视频自动标注的一些最新方法分为两类,一一介绍它们的基本思路;第 3 节介绍一些自动标注系统;最后总结全文,并指出进一步的研究方向。

2 图像和视频自动标注方法

图像和视频检索的方法主要有基于文本的检索和基于内

容的检索。图像/视频自动标注技术结合了两者的优点,继承了关键字检索的高效率并利用了基于内容检索的自动化处理的优点。如果所有的图像/视频已经被标注,那就能有效并高效地实现语义检索^[6]。因此,图像与视频自动标注成为了当前的一个研究热点。

从是否需要建立模型的角度来看,图像和视频自动标注的方法可以分为基于学习(Learning-Based)和基于搜索(Search-Based)的方法^[6]。

2.1 基于学习的标注

基于学习的标注是将图像或视频标注作为一个学习问题。通常每幅图像的所有关键词或关键词与视觉特征的联合分布需要学习一个统计模型^[6]。常用的学习方法主要分为有监督学习和无监督学习。

2.1.1 有监督学习

有监督学习是在已知类别样本的指导下进行学习和训练,统计出各类训练样本不同的描述量,并利用这些参数进行分类器设计。人们对有监督学习的自动标注进行了大量的研究,下面介绍几种近年来提出的新方法。

Chang 等提出的基于内容的软标注(CBSA)^[7]是比较经典的利用贝叶斯分类器的方法,主要包括 3 个步骤:1)利用预先选择的语义标签人工标注训练图像;2)训练 K 分类器;3)利用分类器自动标注图像。通过 K 个分类器给图像分类,并

到稿日期:2011-01-12 返修日期:2011-04-23 本文受西北工业大学基础研究基金(JC201041),西北工业大学引进高层次人才科研启动费资助项目资助。

尹文杰(1987-),女,硕士,主要研究方向为计算机视觉、模式识别,E-mail:77yinwenjie@163.com;韩军伟(1977-),男,博士,教授,主要研究方向为计算机视觉、模式识别;郭雷(1956-),男,博士,教授,主要研究方向为计算机视觉、模式识别。

对每个标签给出一个可信度,代表某图像被标注为该标签的可能性。利用 25K 图像集和 BPM-OPC 分类,说明软标注的结果,大致分为 3 种情况:1) $P_1 \geq 0.9$ (即可信度最大值大于等于 0.9),表明该标签与图像的相关性非常大,可以忽略其他标签;2) $P_1 < 0.9$ 并且 $P_2 > 0.5$ (即可信度最大值小于 0.9,次大值大于 0.5),说明第二个标签与图像的相关性也比较大,可以提供额外的信息,因此标注时也要考虑该标签;3) $P_1 < 0.9$ 并且 $P_2 < 0.5$,则可信度次大的标签提供的信息往往是错误的,因此不能用其标注图像。此方法的优点是标注的准确性比较高,缺点是需要人工标注训练集。

Carneiro 等提出的有监督多类标注(SML)^[8]方法,把每一个语义概念定义为一个图像类别,让所有的概念类同时竞争图像的标注。SML 将大量用于表示图像的高斯混合模型聚类成一个表示概念类的高斯混合模型,利用 GMM-DCT 方法训练样本集,计算图像与每个类别的相似度,然后用后验概率值最大的 5 个类别标注测试图像。这种方法的优点在于 1) 数据库和词汇量具有可扩展性;2) 在标注时能产生语义标签的自然排序;3) 概念简单,不需要图像的语义分割。缺点是计算时间比较长。

判别式条件随机场(CRF)模型^[9]不需要进行图像分割,而是把图像作为整个观察数据。该模型用组合的方式从语义层次和视觉层次获得语义概念的相互作用。利用图形结构建立语义概念的上下文关系,利用 Contextual Kernel 解决一组独立的二次设计问题来训练模型。实验数据库为 Corel 和 TRECVID。该方法的优点是,不需要图像分割;与 SVM 方法比较,得到的查准率和查全率都有很大提高。

为了解决大规模概念空间学习和语义鸿沟的问题,Wang 等提出了利用局部多标签分类自动标注图像(KMLC)^[10]。通过将局部近邻的相关关键字集中成主题来降维高层次的语义空间,并将这些主题作为分配给待标注图像的词汇。为了处理语义鸿沟,通过找到视觉和语义空间的最优裕度来降低两者的偏差,最后定义一个基于加权 K 近邻分类器的标注框架。利用 Corel5K 数据进行的实验表明,该方法的查准率、查全率以及有效性都比判别式模型 ASVM-MIL 要高。

多标签稀疏编码框架^[11]应用于图像自动标注背景下的特征提取和分类。首先,利用无条理的图像小块的高斯混合模型,将每幅图像编码成所谓的超向量。然后,利用基于稀疏编码的标签子空间学习算法有效地降维多标签信息。最后提出用于多标签数据的稀疏编码方法。与 MT, CRM 等方法比较,该方法得到的平均查准率与查全率更高。

利用相关多标签视频标注(CML)框架^[12]能同时进行概念分类并建立它们之间的关系。该方法主要用于多标签视频处理。首先给出两个模型元素:单独概念模型元素和概念关系元素,然后将这两个类型的元素结合起来构成特征向量。利用特征向量和核表达式,训练分类器模型,利用吉布斯随机场表示多标签模型。利用视频镜头的低层特征的实验显示,该方法得到的标注的平均精度比 IndSVM 和 CBCF 方法都要高。

同时采用图像分类和标注的方法^[13]将类别标签看作图像的全局描述,把关键字标签看作图像的局部描述。该方法建立了两个模型:多类 sLDA 和有标签多类 sLDA。用两个包含类别标签和标注词语的数据集来测试模型。图像特征使

用 128 维的 SIFT 区域描述符。与 corr-LDA 模型对比,在所有的测试中有标签多类 sLDA 模型略好于前者。该模型的优点在于分类和标注可以使用相同的潜在空间,而且标注的结果比较好,缺点是训练集样本需要同时包含类别标签和关键字。

主题回归模型 LDA(tr-mmLDA)^[14]是一个关于图像或视频标注的统计主题模型。用 SIFT 描述符得到 Bag-of-Word 来表示图像和文本。这个模型的主要创新之处在于用潜在变量回归方法获得图像特征和标注文本的相互关系,可以得到图像和标签相互关系的变化程度,允许图像和标注文本主题数目不同。该方法比 cLDA 的标注和检索结果更好。

利用类别稀疏的标注方法^[3]可以解决图像标注中的特征相关问题。训练部分利用稀疏度和聚类先验知识来选择低层特征,选用颜色以及显著点作为图像特征,用迭代方法得到正负图像对的分布。测试部分通过从相似图像中传递关键字来自动标注输入图像。由于特征之间可能有很大的相关性,该方法的优点是稀疏先验知识可以在相当程度上减少运算量,聚类先验知识有利于模型的稳定性。

基于规则(Rule-based)^[15]的方法利用图像参数维数和元数据以及颜色、形状、纹理共同作为特征进行语义概念处理,用语义概念人工标注图像。利用训练集构建一个决策树,然后引入一些规则将属性向量(特征变量)映射到类别标签(输出-因变量)。实验结果表明它能达到很理想的性能,查准率和查全率甚至有时超过 80%。

针对在线视频门户的标注方法^[16]主要研究是否可以利用用户标注的在线视频来学习基于内容的标注。首先去除视频冗余,然后提取关键帧,利用关键帧的视觉特征计算某标签出现的后验概率,判断是否包含该标签。所用的视觉特征包括全局和局部特征。计算标签出现的后验概率时,不同特征利用不同的算法。实验表明利用在线视频进行训练是可行的,优点是不需要人工标注训练集,缺点是去除冗余不彻底,仍然存在重复内容。

Socher 等提出的半监督模型^[17]利用极少量已标注图像和大量未组织的文本语料库将图像区域与文本标签相关联。利用核典型相关分析(kCCA)学习文本词汇和视觉词汇,kCCA 将这两种词汇都映射到低维空间。为了得到视觉词汇,首先进行图像分割,然后提取颜色、纹理、形状和位置特征,再用 K 均值方法聚类。这种方法的优点在于只需要很少的训练样本图像,缺点是需要进行图像分割,增加了算法复杂度。

人类行为的自动标注^[18]利用最小人工监督处理视频中人类行为的时间自动标注问题,将电影脚本作为弱监督的途径。该方法利用基于核的可判别聚类算法在弱标注的训练数据中定位人类行为来解决脚本信息不精确的问题,利用得到的行为样本训练时间行为检测器,并应用于原始视频数据的行为定位。用 4 个电影中的 3 个动作集合做实验,结果表明这种行为模型的弱监督学习方法可以改善行为探测。优点是不需要人工标注,缺点是对于行为的时间定位不够精确。

有监督学习是目前最常用的标注方法,近几年来人们通过大量研究,提出了各种模型学习方法。文献[7-9]是多类标注,将图像标注看作是分类问题,不需要图像分割。文献[7]提出的是比较经典的分类方法,其主要利用 SVM 和 BPM 分类器。文献[8]利用高斯混合模型进行学习,概念简单。文献

[9]用 CRF 对语义概念建模,考虑了背景联系。文献[10-12]都需要图像分割:文献[10,11]是图像标注,需要降维多标签语义信息;文献[12]是视频标注,利用吉布斯随机场建模。文献[13,14]都是对主题模型 LDA 的扩展。文献[3,11]利用类别稀疏方法简化计算。文献[15]加入图像参数维数和元数据作为特征,利用决策树和一些规则进行学习,得到很高的查准率和查全率。文献[16]主要研究是否可以利用用户标注的在线视频来学习基于内容的标注,提出了多种特征组合以及不同的融合规则。文献[17,18]属于半监督方法,利用少量训练集学习:文献[17]使用文本语料库;文献[18]使用电影脚本作为训练集。半监督方法的优点在于不需要大量学习样本,减少了计算复杂度。

2.1.2 无监督学习

无监督学习是不知道样本类别,只知道样本的某些信息,然后利用这些信息进行估计。近几年关于无监督学习的标注方法的研究主要有下面一些方法。

利用语言学习结构外观模型的图像标注方法^[19],目的是标注背景零乱的图像中的典型目标。除了视觉模式外,还需利用语言提示,从噪声图像集合-标题中学习目标外观模型。学习的结果是一组典型目标外观模型与它们名字的成对组合,可以用来在新图像中标注类似目标。这种方法最适合学习结构鲜明的目标(例如商标或标志),优点是不需要感兴趣的先验知识,缺点是检测目标类别或总体环境对于效果不理想。

传统的自动标注方法通常有两个问题:不能保证标注词语的语义连贯性;语义相似性严重依赖于视觉相似性。为了解决这两个问题,Mei 等提出了基于 SDF(语义距离函数)的方法^[6],即通过获取标注的先验知识学习语义距离,同时将图像作为整个实体进行标注。为了标注一幅新图像,首先聚类训练图像,根据每个集群的训练图像与该图像的 SDF 值划分等级,这样将相应的标签作为整个实体传递给新图像,保证了语义连贯性。实验利用 Corel 图像,与基于 SVM 的方法对比,基于 SDF 的方法在语义连贯性方面性能更佳,特别是在训练图像与多个词语相关的情况下。缺点主要是现有的语义距离学习不能有效处理大量相关的约束。

TagProp^[20]是一种通过计算近邻中某标签不出现/出现的加权组合来预测某图像标注的最近邻模型,利用已标注图像的加权最近邻模型找出测试图像的标签,近邻权重取决于近邻的等级或距离。TagProp 通过最大化训练集标注的对数似然函数进行学习。为了提高低频词语的检索率,引入一个加权近邻标注预测的特定词语反曲模型。实验表明,该方法的查准率和查全率比 CRM 等方法都有很大提高。

Neela 等提出了一个完全自动化和公众分类的标签传递方法^[21],利用某用户的本地交互网(LIN)中其他用户的标注建立模型,不需要该用户标注偏好的详细信息。标签传递模型生成基于内容的标签,然后利用朴素贝叶斯公式将这些标签转化为一组从本地交互网用户的标签中选择的公众分类标签,从而得到合适的标签。

基于背景的多标签标注方法^[22]是将关键字传递和图像标注看作多标签学习问题。这种方法可以利用关键字的背景内容传播关键字,同时可利用相似性从训练图像传递多个关键字给测试图像。文中提出了一个称为空间谱内核的 2D 线

性核,将视觉关键字的背景加入图像(即 2D 序列)相似性中,可以用于关键字传递。用 MSRC 和 Corel5K 图片库做实验,结果表明,与 SPM,PLSA 等方法比较,该方法能得到更理想的效果。

语义域自适应扩散(DASD)^[23]针对大范围视频概念标注问题,考虑背景转移时利用语义背景的标注方法。DASD 利用图扩散技术优化初始标注,维持标签在语义图表的连贯性和平滑性。语义图表将概念看作节点,将概念相似性看作边界权重。特别地,DASD 方法能改善标注结果,同时把概念相似性用于新的测试数据。实验结果显示了该方法标注的连贯性和较高的精确度,而且标注所用时间很短。

与有监督学习相比,无监督方法不需要标注对象类别的先验知识。文献[6,19]研究利用聚类算法进行学习:文献[19]使用传统的 K 均值聚类算法,它不仅利用视觉特征,同时利用语言提示进行学习;文献[6]对 K 均值算法进行改进,用基于 SDF 的方法标注图像。文献[20]提出一种最近邻模型,它利用近邻中某标签不出现/出现的加权组合来预测某图像。文献[21]提到的利用 LIN 的标注方法是一种利用公众分类的方法。文献[22,23]利用图像或视频背景进行标注,优点是标注结果具有很好的语义连贯性。

2.2 基于搜索的标注

基于搜索的方法是利用搜索技术寻找相似图像或视频,然后从这些相关图像或视频中挖掘相关文字^[6]。人们对于基于搜索的标注的研究相对来说是比较少的,主要有以下一些工作。

Wang 等提出的无模型图像标注^[24]是通过挖掘搜索结果来标注图像的数据驱动方法。这种方法包括 3 个步骤:1)搜索过程,发现视觉和语义相似的搜索结果;2)挖掘过程,从搜索结果的文本描述中识别显著词汇;3)标注拒绝处理,滤除第二步得到的噪声词汇。为了保证标注的实时性,需要两种关键技术:一是将高维图像视觉特征映射成散列码,二是作为分布式系统实现标注。该方法不需要训练集,能够用无限制的词汇标注图像,对异常值具有很好的尺度性和鲁棒性以及有效性和高效性。

传统的基于机器学习或计算机视觉的方法只能处理几百种对象类别,而 ARISTA^[25]的目的是建立一个处理真实世界的流行概念的图像标注引擎。主要思想是通过挖掘类似图像的关键字来标注图像。通过图像检测方法搜索图像的相似值,挖掘算法主要用 SRC 和 MV 方法。实验表明,随着数据库范围增大,能发现更多的正确标注。虽然标注精度在数亿规模的图像中保持不变,但是检索率持续提高。可以利用 VisualRank^[26]分析图像间的视觉联系,改善搜索结果。

Moxley 等提出的视频标注方法^[27],目的是利用新闻视频中的内容重叠部分,通过挖掘类似视频进行自动标注来增加、滤除和改善原有标签。该算法利用两步过程:搜索和挖掘。给出一个包含视觉内容和语音识别文本的视频,首先在多模型搜索结果中划分相似视频的等级,然后挖掘这些类似视频的相关文本,为查询提取关键字。常用的搜索方法有基于图像、文本、概念以及它们的组合模型。挖掘方法则利用关键词的词频向量,找出最优的关键词。实验利用 TREC-VID2005 库,对比了不同搜索方法及挖掘方法的优劣。优点在于该方法是完全非监督的,不受预先定义的词汇的限制,有

很好的鲁棒性,缺点是标注的精确度比较低。

冗余可以提供视频间关系的有用信息,利用基于内容的视频分析技术发现这些联系并用它们产生新的标签。Siersdorfer 等提出了几种不同的自动获取更多视频标签的标注传递的方法^[28]:1)基于近邻的方法,只考虑有直接重叠的视频;2)TagRank 方法,同时考虑有上下文联系的视频。用基于CBCR 的方法查找近似视频,分别做分类(SVM)和聚类(k-Means)实验,对比几种标注方法的性能。优点是该方法不需要去除冗余和提取关键帧。

总的来说,基于搜索的标注方法是完全非监督的,不需要训练集,不受预先定义的词汇的限制。这种方法主要包含两个关键步骤:搜索和挖掘。文献[24,25]针对图像自动标注,两者都应用 SRC 方法进行挖掘,但侧重点不同:文献[24]提出的方法具有很好的实时文献[25]主要是为了处理真实世界的流行概念。文献[27,28]目的是标注视频:文献[27]是利用新闻视频,提出了多种搜索的方法,挖掘过程利用关键帧的词汇向量;文献[28]利用视频间的冗余信息,不需要提取关键帧。

目前,对基于搜索的标注方法的研究还不是很多,大部分研究集中在基于学习的标注,尤其是有监督的方法。此外,还有少量关于标签优化方法的研究,例如文献[29]提到的将标注优化作为一个马尔可夫过程,把候选标签看作马尔可夫链,然后重新划分候选标签的等级,以达到优化的目的。

3 图像检索与标注系统介绍

上传和共享图片已经成为一种常见的活动,利用一些在线图片共享工具和社交网络平台,例如 Yahoo! Flickr, Google, Picasa, Photobucket, Facebook, Orkut 等,大大方便了这项活动^[21]。在上传图片时,用户同时提供标签,以便于检索。在此基础上,出现了大量的图片搜索网站,例如 Yahoo! Image Search, Google Image Search 等。

Behold Image Search^[30]是一个利用自动产生的标签检索 Flickr 图片的搜索引擎。与其它图片搜索网站不同,Behold 能识别图片的多种视觉概念,用户可以在检索结果的基础上让 Behold 返回,使其看起来像某个概念的图像,例如可以让 Behold 返回标注为‘London’并且看起来像‘buildings’的图片。这样可以通过图像视觉近似来过滤基于文本的检索结果。同时,Behold 能够分析用户查询的词语,主动提供视觉过滤概念。在网页右上方它会展示使用某个建议的概念得到的搜索结果,可以节省用户的查找时间。例如,当搜索‘eagle’时,Behold 会建议用视觉过滤概念‘animal’,‘face’和‘bird’来提炼检索结果。Behold 的视觉概念搜索是基于图像自动标注技术的。

目前已经有一些自动标注图像的网站。例如一个实时标注系统 ALIPR^[31],它可以标注用户上传的图像,也可以通过指定的 URL 标注任意在线图像。用户上传图像或给出在线图像的网址后,系统会给出 15 个标注的关键字。ALIPR 希望用户从中选择正确的关键字,或提供其他关键字以及图像的一些相关信息,以完善图像标注结果^[32]。

SpiritTagger Global Photograph Annotation^[33]是一个图

像标注系统,该系统通过挖掘以地理坐标为参数的世界各地拍摄的照片来提取信息。处理的问题包括图像数据的快速检索、位置估计、描述符用法、标注方法以及自动学习标签语义。SpiritTagger 可以利用包含 1,400,000 图片的数据库标注世界各地拍摄的图像。首先用户上传一张照片,在地图上标出拍摄的位置,然后 SpiritTagger 通过考虑视觉相似性和地理位置产生一系列建议标签,用户点击选择某些标签,最后将结果上传到 flickr。

不同的人对同一幅图像会有不同的认识,人工标注图像代价很大而且可信度不高。为了解决这个问题,Ahn 发明了 game ESP, Google 获得了这个游戏许可并发明了自己的版本 Google Image Labeler^[2]。这两个网站都具有很强的趣味性,吸引用户去标注图像。

ESP game^[34]是一个非常有趣的网站,它具有娱乐和帮助训练计算机解决问题的双重功效。它包括 7 种游戏: ESP Game, Tag a Tune, Verbosity, Squigl, Matchin, Fliplt, PopVideo。例如,在进行 ESP Game 游戏时,用户和他的搭档同时观看并标注同一幅图像,当两个人的标签相同时会继续进行游戏并赢得一个分数。在一分钟内两人往往可以得到六七个相同的标签,网页会记录这些标签并把它们与图像相关联,这样搜索引擎就可以更好地认识图像的内容。Squigl 游戏是两个人观察同一幅图像和词语,按下鼠标并跟踪该词语描述的对象,最后提交结果。如果两人的跟踪相符,就会得到相应的分数。PopVideo 游戏是两个用户观看同一个视频剪辑,并提交每个人对所看到和听到的描述。

Google Image Labeler^[35]是一个允许用户标注图像并帮助改善 Google Image Search 结果的网站。当用户点击开始按键后,系统会自动分配另一个在线用户作为其搭档,然后两人在两分钟时间内观察同一组图像,分别提供尽可能多的描述图像的标签。当两个用户提供的标签相符合时,会得到一个分数,描述得越详细,得分越高。时间用完后,网页会显示你标注的图像、图像来源、得分分数以及你的搭档提供的标签。这样就帮助了 Google 改善图像搜索结果的相关性,帮助 Google 用户快速方便地找到想要的结果。

结束语 本文介绍了近几年来提出的一些关于图像与视频自动标注的方法以及一些图像检索与标注的网站。目前人们的研究主要集中于有监督的标注方法,而无监督的方法和基于搜索的方法还有待发展。近年来对于图像和视频的自动标注的研究已经取得了很大进展,但仍存在很多不完善的地方,需要进一步研究。由于网络的快速发展,利用大量的在线图像和视频以及互联网用户的反馈信息将是一个很有前途的研究方向。此外,为了提高标注的有效性和高效性,可以结合提取图像或视频的显著区域,针对显著区域进行标注。

参 考 文 献

- [1] Wang J Z, Geman D, Luo J, et al. Real-world Image Annotation and Retrieval: An Introduction to the Special Section[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11):1873-1876
- [2] 栾悉道, 谢毓湘, 谭义红, 等. 多媒体语义模型研究进展[J]. 计算机科学, 2010, 37(11):1-6

- [3] Zhang Shao-ting, Huang Jun-zhou, Huang Yu-chi, et al. Automatic Image Annotation Using Group Sparsity[C]//IEEE Conference on CVPR. San Francisco; IEEE, 2010; 3312-3319
- [4] Yan Rong, Natsev A, Campbell M. A Learning-based Hybrid Tagging and Browsing Approach for Efficient Manual Image Annotation[C] // IEEE Conference on CVPR. Anchorage; IEEE, 2008; 1-8
- [5] Sankar K P, Jawahar C V. Probabilistic Reverse Annotation for Large Scale Image Retrieval[C]//IEEE Conference on CVPR. Minneapolis; IEEE, 2007; 1-6
- [6] Mei Tao, Wang Yong, Hua Xian-sheng, et al. Coherent Image Annotation by Learning Semantic Distance[C]//IEEE Conference on CVPR. Anchorage; IEEE, 2008; 1-8
- [7] Chang E, Goh K, Sychay G, et al. CBSA: Content-Based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 26-38
- [8] Carneiro G, Chan A B, Moreno P J, et al. Supervised Learning of Semantic Classes for Image Annotation and Retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410
- [9] Xiang Yu, Zhou Xiang-dong, Liu Zuo-tao, et al. Semantic Context Modeling with Maximal Margin Conditional Random Fields for Automatic Image Annotation[C] // IEEE Conference on CVPR. San Francisco; IEEE, 2010; 3368-3375
- [10] Wang Mei, Zhou Xiang-dong, Chua T S. Automatic Image Annotation via Local Multi-label Classification[C]//Proceedings of the International Conference on CIVR. New York; ACM, 2008; 17-26
- [11] Wang Chang-hu, Yan Shui-cheng, Zhang Lei, et al. Multi-label Sparse Coding for Automatic Image Annotation[C] // IEEE Conference on CVPR. Miami; IEEE, 2009; 1643-1650
- [12] Qi Guo-jun, Hua Xian-sheng, Rui Yong, et al. Correlative Multi-label Video Annotation[C]//International Multimedia Conference. New York; ACM, 2007; 17-26
- [13] Wang Chong, Blei D, LI Fei-fei. Simultaneous Image Classification and Annotation[C]//IEEE Conference on CVPR. Miami; IEEE, 2009
- [14] Putthividhya D, Attias H T, Nagarajan S S. Topic Regression Multi-modal Latent Dirichlet Allocation for Image Annotation [C]//IEEE Conference on CVPR. San Francisco; IEEE, 2010; 3408-3415
- [15] Wong R C F, Leung C H C. Automatic Semantic Annotation of Real-World Web Images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1933-1944
- [16] Ulges A, Schulze C, Keysers D, et al. Content-based Video Tagging for Online Video Portals[C]//3rd MUSCLE ImageCLEF Workshop on Image and Video Retrieval Evaluation. Budapest; MUSCLE, 2007
- [17] Socher R, Li Fei-fei. Connecting Modalities: Semi-supervised Segmentation and Annotation of Images Using Unaligned Text Corpora[C]//IEEE Conference on CVPR. San Francisco; IEEE, 2010; 966-973
- [18] Duchenne O, Laptev I, Sivic J, et al. Automatic Annotation of Human Actions in Video[C]//Proceedings of IEEE International Conference on Computer Vision. Kyoto; IEEE, 2009; 1491-1498
- [19] Jamieson M, Fazly A, Stevenson S, et al. Using Language to Learn Structured Appearance Models for Image Annotation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(1): 148-164
- [20] Guillaumin M, Mensink T, Verbeek J, et al. TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation[C] // IEEE 12th ICCV. Kyoto; IEEE, 2009; 309-316
- [21] Sawant N, Datta R, Li Jia, et al. Quest for Relevant Tags Using Local Interaction Networks and Visual Content[C] // Proceedings of the International Conference on Multimedia Information Retrieval. New York; ACM, 2010; 231-240
- [22] Lu Zhi-wu, Ip H H S, He Q. Context-based Multi-label Image Annotation[C]//ACM International Conference on Image and Video Retrieval. New York; ACM, 2009; 1-7
- [23] Jiang Yu-gang, Wang Jun, Chang Shih-fu. Domain Adaptive Semantic Diffusion for Large Scale Context-based Video Annotation[C]//IEEE 12th International Conference on Computer Vision. Kyoto; IEEE, 2009; 1420-1427
- [24] Wang Xin-jing, Zhang Lei, Li Xi-rong, et al. Annotating Images by Mining Image Search Results[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1919-1932
- [25] Wang Xin-jing, Zhang Lei, Liu Ming, et al. ARISTA-Image Search to Annotation on Billions of Web Photos[C]//IEEE Conference on CVPR. San Francisco; IEEE, 2010; 2987-2994
- [26] Jing Yu-shi, Baluja S. VisualRank: Applying PageRank to Large-Scale Image Search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1877-1890
- [27] Moxley E, Mei Tao, Hua Xian-sheng, et al. Automatic Video Annotation Through Search and Mining[C] // IEEE ICME. Hannover; IEEE, 2008; 685-688
- [28] Siersdorfer S, Pedro J S, Sanderson M. Automatic Video Tagging Using Content Redundancy[C]//Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. New York; ACM, 2009; 395-402
- [29] Wang Chang-hu, Jing Feng, Zhang Lei, et al. Content-based Image Annotation Refinement[C] // IEEE Conference on CVPR. Minneapolis; IEEE, 2007; 1-8
- [30] Behold Image Search[OL]. <http://behold.cc>
- [31] ALIPR[OL]. <http://alipr.com>
- [32] Li Jia, Wang J Z. Real-time Computerized Annotation of Pictures [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(6): 985-1002
- [33] SpiritTagger Global Photograph Annotation[OL]. <http://cortina.ece.ucsb.edu/index.php>
- [34] ESP game web site[OL]. <http://www.espgame.org/gwap/>
- [35] Google Image Labeler Web site [OL]. <http://images.google.com/imagelabeler/>