

图像自动语义标注技术综述^①

孙君顶, 杜娟

(河南理工大学 计算机科学与技术学院, 焦作 454000)

摘要: 近年来, 随着对基于内容图像检索技术研究的深入, 图像自动语义标注已成为了该领域的研究热点。针对目前广泛研究的图像语义标注技术, 从其分类、关键技术、存在问题及发展方向进行了进行了论述, 以期从事该方向研究的人员提供一定的借鉴意义和参考价值。

关键词: 图像语义; 语义标注; 图像检索

Review on Automatic Image Semantic Annotation Techniques

SUN Jun-Ding, DU Juan

(School of Computer Science and technology, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: With the rapid development of the content-based image retrieval technology, automatic semantic annotation has become the challenging task in this field. A wide variety of methods have been proposed during the past decades. The classification, key technology, problems and the further study of semantic annotation were discussed in detail in the paper. It is valuable for the researchers who devote to the research in this field.

Key words: image semantic; semantic annotation; image retrieval

1 引言

随着多媒体技术和网络技术的快速发展, 越来越多的信息以图像的形式呈现出来, 而且图像已广泛应用于医学、通信、工农业生产、航天、教育、军事等多个领域。为了更好地管理和利用这些海量图片信息, 建立有效的分类和检索方式已成为迫切需要解决的问题。前期人们已经在图像检索方面作了大量的研究^[1], 就目前的检索趋势而言, 大致可分为三个不同的着眼点^[2]: 一是基于文本的图像检索; 二是基于内容的图像检索; 三是基于语义的图像检索^[3]。

传统的方法是基于文本的图像检索需要人工对图像添加关键字, 从而将视觉信息检索转换成成熟的文本检索问题。虽然这种方法简单易行, 但是文本描述难以充分表达图像的丰富含义, 而且具有主观性; 另外, 人工标注费时费力, 效率低, 已远不能满足当今图像快速增长的需要。因此, 基于内容的图像检索技术应运而生。区别于基于文本的图像检索它通过自动

提取图像的低层特征, 主要包括图像的颜色、纹理和形状等低层视觉特征及其组合, 该技术解决了基于文本图像检索所存在的一些问题。但对于普通的一般用户而言, 提供实例图像并不是一件容易的事, 而且图像低层视觉特征与图像的高层语义间还存在“语义鸿沟”^[4-6], 所以采用低层特征进行检索并不能充分的表达图像的深层语义。因此, 建立图像语义表示和检索机制势在必行, 解决该问题的关键就是要对图像进行自动语义标注^[7,8]。语义标注的实质是通过对图像视觉特征的分析来提取高层语义用于表示图像的含义, 从而在图像低层特征和高层语义之间建立一座桥梁, 解决低层特征和高层语义间的“鸿沟”问题。其主要思想是从大量图像样本中自动学习语义概念模型, 并用此概念模型标注新的图像。目前, 人们对图像语义标注的研究已经取得了一定的成果, 建立了语义标注的模型, 如翻译模型^[9], 跨媒体相关模型^[10], 还有根据相关模型改进的连续相关模型^[11]和 MBRM(Multiple-

① 基金项目:教育部科学技术研究重点项目(210128);河南省骨干教师资助计划(2010GGJS-059);河南省国际合作项目(104300510066)

收稿时间:2011-10-22;收到修改稿时间:2011-11-16

Bernoulli Relevance Model)^[12]等。

图像语义标注有着重要的理论和应用价值,是目前图像检索领域研究的热点问题。本文论述了图像语义标注现有方法及分类,同时指出了图像语义标注的关键技术、存在的问题及发展趋势。

2 图像自动语义标注方法

目前出现的图像语义标注方法已经很多,本文将相关语义标注方法归结为分类法、基于概率统计模型法和关联文本法三类,并分别对其进行了讨论说明。

2.1 分类法

分类法是将每一类语义概念当作一个类别进行分类,代表性方法有:k-means法、支持向量机方法、信息瓶颈方法等。其中,k-means算法的思想是首先从n个数据对象任意选择k个对象作为初始聚类中心;而对于所剩下对象则根据它们与聚类中心的相似度,分别将它们分配给与其最相似的聚类;然后再计算每个所获新聚类的聚类中心;不断重复这一过程直到标准测度函数开始收敛为止。

在文献[13]中,Wang等人采用基于子空间聚类算法用K-means算法生成blob-token,并用统计的方法在token和key-word之间建立关联,实现图像的标注。但该算法过分依赖距离函数和聚类中心的选择,聚类数K和目标函数也很难选择。为此,文献[14]中提出改进的k-means算法,采用遗传聚类算法确定聚类数K,通过已标注的图像求出语义概念和聚类区域的关联度,用它作为待标注图像的先验知识,然后结合区域的低层特征,对未标注的图像进行标注。Li Wei等人^[15]利用K-means的聚类算法结合语义约束在visual terms和keyword之间建立关联,以此构建分类器,将分类器用于后续图像的标注。这种方法当语义概念相当多时会遇到困难。为了解完善分类器和复杂的语义概念之间的对映关系,文献[16]提出一种基于Boosting学习的图片自动语义标注方法,首先构造很多的模型,然后在模型和概念之间建立联系,保持一种多对多的关系,但是性能容易受Boosting的迭代次数的影响。

支持向量机是一种有监督的分类方法,已经被证明在高位数据分类上的高效性。文献[17]采用基于SVM来进行图像分类器的构造,为每一个图像类别构造一个分类器,把图像的组合特征作为支持向量机的输入向量。文献[18]提出一种概念索引方法,采用SVM

的多类分类器的空间映射方法,将图像的低层特征映射为具有一定高层语义的模型特征以实现概念索引。这种方法在一定程度上提高了标注的精确度,但SVM的参数固定不变,系统的鲁棒性不强。信息瓶颈方法是由Slonim等人^[19]提出的一种新颖的无监督聚类方法。该方法首先采用改进的K均值无监督图像分割算法将图像分割,然后采用信息瓶颈算法对分割后的图像进行聚类,并确定语义概念和区域之间的关系,然后采用条件概率最大的语义关键字对图像分割区域进行标注。但由于信息瓶颈算法是无监督的聚类算法本身存在一定的缺陷。为了得到更好的聚类结果,钟洪等^[20]在聚类过程中通过加入语义约束来对信息瓶颈算法进行改进,在一定程度上提高了标注的准确度。

上述语义标注方法,往往会将具有相同视觉特征的区域归为一类,即使区域语义完全不同,也会用相同的关键字标注,如“sky”与“water”具有类似的图像的低层特征,但它们所蕴涵的语义却完全不同。

2.2 基于概率统计模型的方法

基于概率模型的方法主要是从已标注好的训练图像集中直接或间接地学习图像视觉特征和语义概念之间的关系,实现低层特征空间到模型空间的映射,估计出待标注图像具有某个属性的概率,然后利用该概率分布对待标注图像进行语义标注,选取那些具有较大概率的模型对应的一组关键词中出现做多的词作为图像的最终标注。

Duygalu等人^[11]提出的翻译模型,对分割后的图像区域特征进行聚类,将连续特征变成离散视觉关键字单词表,从而将图像的标注问题可以看作从图像视觉关键字到语义关键字的翻译过程。Jeon等人^[12]提出了一种相关模型,用视觉关键字与语义关键字的联合概率进行标注。上述模型虽然考虑了对象和区域的语义含义,但采用对视觉特征的离散处理方法会造成视觉特征内容的损失,由于事先很难确定一个理想的聚类粒度,标注结果受离散化影响较大。为解决上述问题,文献[14]提出了连续相关模型和多伯努力相关模型,使用非参数高斯核进行特征生成概率的连续估计。与离散模型相比,其标注性能得到了显著提高。

最近人们利用词间关系来改善图像的标注效果,代表性工作包括CLM(Coherent Language Model)模型^[21]、结合WordNet的方法^[22]以及AGAnn(Adaptive Graph for Automatic Annotation)^[23]方法。CLM模型使

用迭代 EM 算法隐含考虑词与词的相关性;结合 WordNet 的方法需借助外部数据源 WordNet,对训练集中有用的信息未能充分利用;AGAnn 对自适应标注的结果应用词与词的相关性。Jin 等人^[21]在解决词间关系问题时放松为估计语言模型生成图像标注词的概率,隐含考虑了词和词的相关性,与使用外部知识源 WordNet 进行词与词相关性不同,该方法在标注过程中利用训练集中词和词的关系进行标注,改善了标注性能。

2.3 关联文本法

近年来利用 Web 图像的关联文本来提高标注性能成为另外一个热门的研究话题。Web 图像通常关联着丰富的文本信息,如图像文件名、替代文本(ALT)、周边文本、所属页面的标题等,图像的语义或多或少地都与这些关联文本相关,所以抽取这些关联文本作为 web 数据对象的一部分或者训练集的数据对象会有有效的提高标注效果。

wang 等人^[24]提出了一个基于搜索和数据挖掘技术的 Web 图像语义自动标注系统,由两步来完成,首先在网络上搜索语义和视觉上相似的图像;第二步在有一个有准确标注的在语义上类似的图像的基础上进行基于内容的图像检索,挖掘周围的关联文本信息标注图像。但该系统要求至少一个准确的初始关键词作为种子来执行基于文本的图像搜索,如果没有初始关键词或初始关键词是不正确,标注的性能将大大降低。文献[25]提出了一种搜索和挖掘相结合的 Web 图像语义自动标注方法,但其标注性能受图像搜索阶段的影响很大。Feng 等人^[26]和文献[27]提出分别基于 Web 图像的视觉特征和文本特征建立两个分类器,并且假定它们是正交的,然后利用 Cotraining 的方法来学习 Web 图像的语义标注词,然后将两个分类器加权起来得到最终的语义标注模型。然而,很难确定两个分类器各自所占的比重,而且利用决策树模型对 Web 图像语义在其关联文本上的分布进行建模,训练完成后,决策树模型将不再改变,但训练出的固定决策树模型并不能够准确地表示所有待标注图像的语义分布,从而影响了 Web 图像语义标注的性能。多数已有的利用 Web 图像关联文本进行标注的方法大都把所有关联文本作为一个整体,或者根据先验知识与启发想法对各类关联文本赋予同定的权重。然而,不同的关联文本对预测图像语义的重要性是不同的,而且随着图像和语义

关键词的改变而改变。为此,文献[28]综合考虑了 Web 图像的视觉特征和关联文本对预测图像语义的贡献,通过带约束的分段惩罚加权回归模型将关联文本权重分布估计和先验知识约束有机地结合在一起,自适应地对图像语义在其关联文本上的分布进行建模,提高了 Web 图像标注的性能。

3 语义标注关键技术

图像自动语义标注首先需要确定提取哪些低层视觉特征作为图像的语义描述,然后考虑选择哪种匹配模型或方法进行近似匹配,最后建立低层视觉特征和高层语义之间的联系。其关键技术主要体现在两个方面:

(1) 图像的视觉特征的提取和匹配。图像特征的提取是自动标注的前提,视觉特征的匹配也就是图像的内容相似性计算是图像能不能准确标注图像的关键。

(2) 图像低层特征和高层语义之间的关联策略。要建立二者之间的联系需要解决二者间的四种映射关系,即图像间的关系、词间的关系以及图像到词和词到图像的映射关系,处理好这四种关系是解决问题的关键。

4 语义标注存在的问题

虽然图像自动语义标注技术将图像的低层视觉特征转化为语义标注字,解决了手工标注费时费力的问题,缩小了图像低层视觉特征和高层语义之间的语义鸿沟,但是语义标注的准确度仍然是今后要深入研究的问题,语义标注尚存在一系列的问题。

(1) 标注效率和准确率低。现有的图像标注系统一般只是用于某一个领域或者类别的图像,没有普适性。

(2) 对象语义识别技术还不完善,由于图像数据在建立的模型数量上庞大而复杂,要实现真正语义层次上的智能图像检索、让机器懂得人类是怎样观察和理解图像还是一个具有很大挑战性的工作。

(3) 训练集图像标注的准确性在后来的机器学习中起着重要的作用,外样本的合理性参数的科学性都会影响自动标注的结果。

(4) 目前图像标注的语义层级没有明确的区别,无法准确识别图像的场景语义、行为语义、情感语义信息。

5 语义标注研究趋势

语义标注是基于内容图像检索技术的发展趋势,为了提高语义标注的准确度及基于语义图像检索的效率,需要在以下几个方面展开更深入的研究:

(1) 人机交互技术。把人纳入到系统中去,在标注和检索过程中介入人机交互,充分考虑用户的反馈信息,利用用户的丰富的先验知识。

(2) 完善图像分割技术。图像分割往往是图像区域特征提取的第一步,目前许多技术只是将图像简单的分成几个区域,仍缺乏简单易行而又准确可靠的通用办法,所以完善图像分割技术,准确分割图像,是以后势必要研究的内容。

(3) 提高语义对象识别技术。随着计算机技术和模式识别技术的发展,图像集识别技术也被纳入其研究范围,虽然在某些特定领域里识别技术已经取得了不错的成果,但图像识别技术还不成熟,有待进一步提高,减少对象识别误差带来的影响。

6 结语

本文对目前图像检索方面的研究热点图像语义标注进行了分析,介绍了图像标注和检索的应用领域、发展历程、研究现状、研究成果及研究趋势,并重点从现有的标注技术进行了详细论述。通过本文的总结,希望对从事图像检索研究的人员提供借鉴和参考。

参考文献

- 1 石军,常义林.图像检索技术综述,西安电子科技大学报. 2003,30(4):486-491.
- 2 孙君顶,赵珊.图像低层特征提取与检索技术.北京:电子工业出版社,2009.
- 3 张毅.基于高层语义的图像检索技术研究[硕士学位论文].西安:西安电子科技大学,2007.
- 4 温超,耿国华.基于图像内容的图像检索问题中的‘语义鸿沟’问题,西北大学学报(自然科学版),2005,35(5):536-540.
- 5 Wang C, Zhang L, Zhang HJ. Learning to reduce the semantic gap in web image retrieval and annotation. ACM SIGIR Conference on Research and Development in Information Retrieval. Singapore, 2008. 355-362.
- 6 Lu Y, Zhang L, Tian Q, et al. What are the high-level concepts with small semantic gaps, CVPR'08. 2008.1-8.
- 7 Cavus O, Aksoy S. Semantic scene classification for image annotation and retrieval. Proc. of the Joint IAPR International Workshop on Structural, Syntactic and Statistical Pattern Recognition. Orlando, 2008. 402-410.
- 8 Carneiro G, Chan AB, Moreno PJ, et al. Supervised learning of semantic classes for image annotation and retrieval. IEEE PAMI, 2007,29(3):394-410.
- 9 Duygnlu P, Bamard K, Freitas N. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. Proc. of the 7th European Conference on Computer Vision-Part IV, 2002. 97-112.
- 10 Jeon J, Lavrenko V, Manmatha R. Automatic Image Annotation and Retrieval Using Cross-media Relevance Models. ACM SIGIR'03. 2003. 119-126.
- 11 Lavrenko V. A Model for Learning the Semantics of Pictures. Proc. of the Neural Information Processing Systems. Vancouver, Whistler: MIT Press, 2004. 553-560.
- 12 Feng SL, Manmatha R, Lavrenko V. Multiple bernoulli relevance models for image and video annotation. CVPR'04. 2004,2:1002-1009.
- 13 Wang L, Liu L, Latifu K. Automatic image annotation and retrieval using subspace clustering algorithm. Proc. of the 2nd ACM International workshop on Multimedia Databases. USA. 2004. 100-108.
- 14 潘崇,朱红斌.改进 k-means 算法在图像标注和检索中的应用.计算机工程与应用,2010,46(4):183-185.
- 15 Li W, Sun MS. Automatic image annotation based on WordNet and hierarchical ensembles. Proc. of the 7th International Conference on Computational Linguistics and Intelligent Text Processing. Mexico, 2006. 417-428.
- 16 茹立云,马少平.基于 Boosting 学习的图片自动语义标注,中国图象图形学报,2006,11(4):486-491.
- 17 胡全,丘兆文,王霓虹.基于多特征融合的图像语义标注.东北林业大学学报,2008,36(10):88-89.
- 18 路晶,马少平.基于概念索引的图像自动标注.计算机研究与发展,2007,44(3):452-459.
- 19 Slonim N, Tishby N. Agglomerative information bottleneck. Neural Information Processing Systems. Cambridge, USA: MIT Press, 1999. 617-623.
- 20 钟洪,夏利民.基于互信息约束聚类的图像标注.中国图象

(下转第 257 页)

标签群的选择是由 select 命令来实现的。Select 命令可以确认或取消确认标签的 SL 标记, 或者可以在四个通道的其中一个通道中将标签的已盘标记设置为 A 或 B。此时码用十六进制表示为 0x0C。

⑤ 空中接口协议 query 参数设置

Query 命令用于启动和规定盘存周期。码设定为 0x0E。

⑥ 反向散射链接频率设置

反向散射链接频率设置为 40kHz, 用十六进制表示的码设定为 0xBF。

⑦ 读标签命令设置

读取标签的设置可以读取多标签的 ID 信息。参数: RC(重复计数): 读写执行盘存循环的次数, 此处设置的为 100, 码用十六进制表示为 0x27,RC 此时为 0x64。

⑧ 停止读标签命令设置

本命令用于停止读取标签的 ID, 此时将码设置为 0x28 即可。

4 结语

文章使用 PR9000 射频芯片设计了一种支持 ISO 18000-6C 协议的超高频 RFID 读写器, 使用频谱分析仪实测读写器输出功率可达 28dBm, 使用驻波比为 1.2 增益为 5dBi 的陶瓷天线, 经测试, 在 2m 范围内可以稳定的读取 Allen 标签, 并且能够连续读取标签, 取得了较好的效果, 有效的解决了防冲突问题。该读写器尺寸比较小, 开发成本相对较低, 可应用在距离要求在 2m 范围内的场合, 也可为需要使用和设计超高频读写器的技术人员提供一个低价位的开发平台。

参考文献

- Phychips, Inc. PR9000-FD12 Datasheet 2010. http://www.phychips.com/new/eng/products/rfid_s1.html.
- ISO/IEC JTC/SC31N. ISO/IEC 18000-6C. Information technology-Radio frequency identification for item management-part 6:Parameters for air interface communications at 860 MHz to 960 MHz.2006.
- Li XR, Chen L, Zhang L, et al. Image annotation by large-scale content-based image retrieval. Proc. of the 14th Annual ACM International Conference on Multimedia. 2006. 607-610.
- Feng HM, ShiChua TS. A bootstrapping framework for annotating and retrieving WEB images. Proc. of the 12th Annual ACM International Conference on Multimedia. 2004. 960-967.
- Tseng VS, Su JH, Wang BW, et al. WEB image annotation by fusing visual features and textual information. Proc. of the 2007 ACM Symposium on Applied Computing. 2007. 1056-1060.
- 许红涛,周向东,向宇,等.一种自适应的 Web 图像语义自动标注方法.软件学报,2010,21(9):2183-2194.
- Jin R, Chai JY, Si L. Effective automatic image annotation via a coherent language model and active learning. Proc. of the International Conference on ACM Multimedia. New York: ACM Press, 2004. 892-899.
- Jin Y, Khan L, Wang L. Image annotations by combining multiple evidence WordNet. ACM International Conference on Multimedia. 2005. 706-715.
- Liu J, Li M, Ma WY. et al. An adaptive graph model for autofmtic image annotation. Proc. of the 8th ACM International Workshop on Multimedia Information Retrieval, 2006. 61-70.
- Wang XJ, Zhang L, Jing F, et al. AnnoSearch: Image auto-annotation by search. CVPR'06. 2006. 1483-1490.

(上接第 261 页)

图像报,2009,14(6):1199-1205.