

学校编码: 10384

分类号 _____ 密级 _____

学号: 31520081153337

UDC _____

厦 门 大 学

硕 士 学 位 论 文

基于内容的图像标注改善算法研究

Research on Content-based Automatic Image Annotation
Refinement

李 东 艳

指导教师姓名: 李 绍 滋 教授

专 业 名 称: 计算机应用技术

论文提交日期: 2011 年 月

论文答辩时间: 2011 年 月

学位授予日期: 2011 年 月

答辩委员会主席: _____

评 阅 人: _____

2011 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

摘要

图像自动标注隶属于计算机视觉、模式识别、信息检索以及机器学习等领域，在学术界和工业界均得到高度的关注，但是由于图像自动标注存在数据集的不平衡性，底层视觉特征与用户高层语义之间的鸿沟等问题，使得图像自动标注仍然面临着许多困难，针对上述情况，本文开展了如下四个方面的研究工作：

1、针对数据库词频分布不平衡问题，提出了一种基于平衡数据库的图像自动标注改善的方法。这个方法主要是通过自动平衡模式找出弱频点，并依据基于外部数据库自动平衡模式对弱频点以下的弱频词，从外部数据库中追加相应数量的图片。通过找到数据库的最佳平衡状态，以及词频的最佳分布使得数据库达到最佳平衡状态，并使得最终标注结果的精确率、召回率、F 度量和至少正确标注一次的关键词数等指标得到一定提高。

2、提出了基于关键词同现的图像自动标注方法。现有图像自动标注模型为了计算方便，假设标注词之间都是相互独立的，忽略了关键词间相关性对最终标注结果带来的影响，针对以上问题，我们提出了基于关键词同现的图像自动标注方法。该方法根据关键词同现（Co-occurrence）的频数计算关键词间的相关性，并根据词频间的相关性建立关联度衡量模式。为了充分考虑关键词间的关联性，采用 k-1 层关联算法计算关键词之间的关联度，并将该关联度与已经得到的标注模型的标注结果相结合，使得最终的图像标注结果得到提高。

3、提出了基于 Wordnet 的图像自动标注模型。该模型先通过 CMRM 方法对图像进行标注，根据得到的关键词标注概率，选取标注词概率较大的前 n 个关键词作为候选标注词。通过结合 JNC 方法与 BNP 方法，建立了关键词间相关度的度量模式，通过此模式来度量 n 个候选关键词间的相关度，筛选出相关度最大的前五个关键词作为最终的标注词，该标注方法可以在一定程度上更好的解决语义鸿沟问题，并最终提升图像标注结果。

4、提出了基于语义相似的图像自动标注模型。该模型将前面提出的关键词同现方法与 Wordnet 相结合来计算关键词间的相关度，并实现最终的图像自动标注改善模型。通过最后的实验，表明建立平衡数据集以及考虑关键词间的各种关

系（如词频同现、词频注释集、上位词与下位词等），均对标注结果有一定的改善。我们的模型在一定程度上解决了语义鸿沟问题，并最终提升图像标注结果。

本文提出的模型均以通过实验验证，图像自动标注的精确率，召回率，F 度量以及至少被正确标注一次以上的关键词数量均得到了一定程度的提高，尤其是召回率及至少被正确标注一次以上的关键词数量得到了较大的提高。

关键词：图像自动标注；语义鸿沟；Wordnet；词频同现；标注改善

Abstract

Automatic image annotation is a part of computer vision, pattern recognition, information retrieval and machine learning and other fields, which has a high degree of attention in academia and industry, but because of the image data set of automatic image annotation existing the problem of imbalance, the gap problem between the underlying high-level visual features and high-level user semantic and so on, which make the automatic image annotation still facing many difficulties. For the above, this paper carries out research in the following four aspects:

1. A approach of based on balance database to improve the automatic image annotation is proposed, because of frequency distribution for the imbalance in the database. Firstly, the weak frequency points are found based on word frequency distribution of the original database, and an appropriate amount of image is added from an external database under the self-balancing mode for each low-frequency word. By finding the best balance of the database, and the best word frequency distribution making the database to achieve the best balance, which making the precision, recall, F and the number of words which can be correctly labeled at least once are improved.

2. A approach of automatic image annotation of based on keywords co-occurrence is proposed. Existing of some automatic image annotation model to calculate conveniently, which suppose independent of each other between the words, ignoring the correlation between the keywords to result in the impact of the final mark. For the above problem, we propose a approach of based on keywords co-occurrence to achive automatic image annotation. The method is based on the frequency of the keywords co-occurrence to calculate the correlation of keywords, and according to the correlation of keywords to establish the measured model of correlation. To consider the correlation between keywords, by k-1 level algorithm to compute the correlation between the keywords, and making a combination of the correlation and the

annotation result of annotation model, which making the final image annotation results increased.

3. A model of automatic image annotation of based on Wordnet is proposed. The model acchives the image annotation through CMRM method, at the same time, according to keywords tagging probability to choose some the larger probability of the keywords as the candidate label. Through a combination of BNP method and JNC method, to estabilishe a measured model of the keywords correlation. By this model to measure the correlation between the candidate keywords, choosing the top five largest correlation as the final mark word keywords. To a certain extent, the semantic gap problem is better soluted by the annotation method and ultimately to enhance the image annotation results.

4. A model of automatic image annotation of based on semantic similarity is proposed. This model makes a combination of the method of keywords co-occurrence and wordnet proposed by the previous to compute the corelation between keywords, and to achieve the final automatic image annotation to improve the model. By the final experiment, indicating that to establish the balance data set and to consider the various relationships between keywords are good to improve the result of the marked. To some extent, our model solves the semantic gap problem, and ultimately enhance the image annotation results.

The proposed model are verified by experiments, automatic image annotation precision, recall, F and the number of words which can be correctly labeled at least once are improved, especially the recall rate and the number of words which can be correctly labeled at least once have been greatly improved.

Key Words: automatic image annotation; semantic gap; Wordnet; word co-occurrence; annotation refinement

目 录

摘要.....	I
Abstract.....	III
第一章 绪论	1
1.1 选题背景及研究意义	1
1.2 研究现状及存在的问题	3
1.2.1 研究现状.....	3
1.2.2 存在的问题.....	11
1.3 本文主要研究工作	12
1.4 本文组织结构	13
第二章 基于平衡数据库的图像自动标注改善模型.....	15
2.1 目前图像库存在的问题	15
2.2 CMRM 图像自动标注算法	16
2.3 基于平衡数据库的图像自动标注	18
2.3.1 基于平衡数据库的图像自动标注原理.....	18
2.3.2 基于平衡数据库的图像自动标注改善模型的实现.....	20
2.4 实验结果与分析	22
2.5 本章小结	26
第三章 基于关键词同现的图像自动标注	27
3.1 关键词同现(Co-occurrence).....	27
3.2 基于关键词同现的图像自动标注模型	30
3.2.1 标注词关联度的衡量标准.....	30
3.2.2 基于关键词同现的图像自动标注模型的构建.....	32
3.2.3 基于关键词同现的标注算法.....	35
3.3 实验结果与分析	35
3.4 本章小结	37
第四章 基于语义相似的图像自动标注模型	39
4.1 基于 Wordnet 的度量模式	39
4.1.1 JNC 方法定义	39
4.1.2 BNP 方法定义.....	41
4.2 基于 Wordnet 的图像自动标注模型	45
4.2.1 标注词关联的度量.....	45

4.2.2 基于 Wordnet 的图像自动标注算法及模型的实现	47
4.3 基于语义相似的图像自动标注模型	49
4.4 实验结果与分析	51
4.5 本章小结	55
第五章 总结与展望	57
5.1 工作总结	57
5.2 对将来研究工作的展望	58
参考文献	61
致谢	67
附录 攻读硕士学位期间发表的论文	69

Table of Contents

Abstract in Chinese	I
Abstract in English	III
Chapter 1 Introduction	1
1.1 Background and Significance	1
1.2 Research Status and Problems	3
1.2.1 Research Status	3
1.2.2 Problems	11
1.3 Main Research and Innovations	12
1.4 Structure of This Paper	13
Chapter 2 Based on Balance Database to Improve the Automatic Image Annotation	15
2.1 Problems of Database	15
2.2 Algorithm of CMRM Annotation	16
2.3 Automatic Image Annotation Based on Balanced Database	18
2.3.1 Principle of Automatic Image Annotation Based on Balanced Database	18
2.3.2 Acchiving Automatic Image Annotation Based on Balanced Database	20
2.4 Experimental Results and Analysis	22
2.5 Conclusion	26
Chapter 3 Automatic Image Annotation Based on Keywords Co-occurrence	27
3.1 Co-occurrence	27
3.2 Automatic Image Annotation Based on Keywords Co-occurrence	30
3.2.1 Measure of Keywords Correlation.....	30
3.2.2 Automatic Image Annotation of Based on Keywords Co-occrrence	32
3.2.3 Algorithm of Based on Keywords Co-occurrence	35
3.3 Experimental Results and Analysis	35
3.4 Conclusion	37
Chapter 4 Automatic Image Annotation Based on Semantic Similarity	39
4.1 Measurement Mode Based on Wordnet	39

4.1.1 Method Definition of JNC	39
4.1.2 Method Definition of BNP.....	41
4.2 Automatic Image Annotation Based on Wordnet	45
4.2.1 Measurement Model of Keywords Correlation	45
4.2.2 Algorithm of Automatic Image Annotation Based on Wordnet.....	47
4.3 Automatic Image Annotation Based on Semantic Similarity.....	49
4.4 Experimental Results and Analysis	51
4.5 Conclusion	55
Chapter 5 Conclusion and Future Work.....	57
5.1 Conclusion	57
5.2 Future Work	58
References	61
Acknowledgements	67
Appendix: My Published Papers.....	69

第一章 绪论

1.1 选题背景及研究意义

随着多媒体数字化技术的发展和推广、存储成本的降低以及网络传输带宽的增长,数字图像成为一种获得广泛利用的媒体。最近几十年来,数码相机和可拍照移动设备的迅速普及,使数字图像更加易于获取,其应用范围也大大扩展。而计算机技术以及互联网的兴起使各种信息(包括图像)的传播速度大大加快,人们可以访问和获取的图像数据呈现出爆炸式的增长。但是,海量的图像数据在带给人们各种便利的同时,也带来了极大的问题:人们容易迷失在浩如烟海的图像数据中而难以找到自己真正需要的图像信息。从而,图像检索技术成为当前相关领域的热点研究课题,在学术界和工业界均得到高度的关注。

按照检索方式的不同,目前的图像检索主要可以分为两类:一种是基于图像视觉内容的图像检索(Content-Based Image Retrieval, CBIR)^[1,2,3,4],而另一种是基于文本的图像检索(Text-Based Image Retrieval, TBIR)^[5,6,7,8]。

基于图像视觉内容的图像检索(Content-Based Image Retrieval, CBIR)是近年来得到许多关注和重点研究的一个领域。在许多应用环境下, CBIR 有着不可替代的作用。根据用户输入目标图像描述的方式,基于图像视觉内容的图像检索又可以分为基于图例的查询(Query by Example, QbE)和基于草图的查询(Query by Sketch, Qbs)两种。在 QbE 系统中,用户提交一幅实际的示例图像,系统根据该图像的颜色、纹理、形状等等信息在数据库中进行查询。但是如何获取该示例图像在许多情况下并不易解决,所以人们提出了利用草图的查询方式 QbS。系统向用户提供一幅画板,用户可以用笔或鼠标在该画板上描绘出期望图像的主要内容和结构特征。在利用 QbS 的检索中,从草图中提取出来的各个元素之间的相对位置等结构特征具有更加重要的作用。

通常 CBIR 要求用户提交一幅图像来作为查询,采用图像的视觉特征(如颜色、纹理和形状等)建立索引,然后根据图像与查询间的视觉相似性度量来实现检索。但是高维索引是非常困难、需要进行更多研究的课题^[9]。此外,基于内容的图像检索要求用户每次都输入示例图片,这给用户在使用上带来了很大困难,也不符合广大用户的检索习惯。

基于文本的图像检索(Text-Based Image Retrieval, TBIR)系统易于使用,用户只需输入词汇就可以进行检索并得到相应结果。这种便捷特性使基于文本的图像检索成为当前最重要的图像检索方式,并且得到了广泛的应用^[10],例如,目前的网络图像搜索引擎基本都采用了基于文本的图像检索。基于文本的图像检索系统按照标注信息的来源方式可以分为以下几类:

(1)人工标注。一些商业图像网站或专业图像供应者依赖于手工标注信息对图像进行索引。然而,当图像数据量快速增长时,手工标注的代价难以接受。目前,Google、Yahoo、MSN等搜索引擎都提供了图像搜索的功能,其索引的图像数目均达到上十亿张。对所有这些图像进行手工标注所需人工花费将十分巨大,耗时很长,人们难以接受。而且,新图像不断出现,手工标注的方式无法满足及时产生和更新对应标注的要求,所以试图一劳永逸对所有图像进行标注是不可能的。此外,由于一幅图像本身往往包含着非常丰富的信息,所以不同的人对同一幅图像会出现不同的理解。对于同一语义信息,不同的人可能使用不同的词汇进行描述。例如,用户查询时使用的词汇就可能与标注时的词汇不同,这进一步加重了标注的不一致性带来的影响。

(2)基于网页上下文的图像自动标注。目前的网络图像搜索引擎及相当数量的网络图像库采用了一种半自动获取图像标注的方法。网络图像的一个重要特点是它是被包含在某一文本网页当中的,因此总是可以获得一些对应的文本信息,例如,图像的文件名及URL、ALT标签文本(ALT tag)、锚文本以及图像周围的环绕文本等信息,其中的ALT文本是当图像不能被正确显示时用以替代图像的文本信息^[11],所以它往往是图像内容的直接描述。由于网络上的网页是由千千万万不同的用户自由建立的,其形式、内容都有着非常大的变化,对应文本质量参差不齐,而且其中与图像相关的文本信息经常是非常有限的。此外,基于网页上下文的图像自动标注并没有考虑到图像的视觉特征,当图像周围的网页文本信息缺失以及网页内容与图像视觉内容无关时,该类方法无法得到或只能得到错误的标注。除了上述问题,这类标注方法还无法对离线图像进行标注。

(3)基于内容的图像自动标注。许多研究工作着眼于采用完全自动的方法从图像的视觉信息本身来获得图像对应的文本信息。这种方法自动化程度最高,也是目前最具有研究意义的方法。近些年来,人们尝试着从各种不同的方向进行探

索，做出了大量的尝试如基于内容的图像自动标注方法。这种图像标注方法可以更好地构建图像视觉内容与语义关键词之间的联系，这也是本文研究的重点。

1.2 研究现状及存在的问题

本章将介绍图像标注的基本框架，并对其中涉及到的图像分割和图像视觉特征提取方法进行归纳总结，最后将对具有代表性的图像标注模型进行阐述。

1.2.1 研究现状

图像自动标注技术，是由底层视觉特征向高层语义转化的过程，因此，图像自动标注的一般步骤如图 1-1 所示，首先给定训练图像集，将训练图像集中的图像进行图像分割，对分割后的图像进行视觉特征提取，并结合图像语义标注算法和已标注信息实现图像的自动标注。

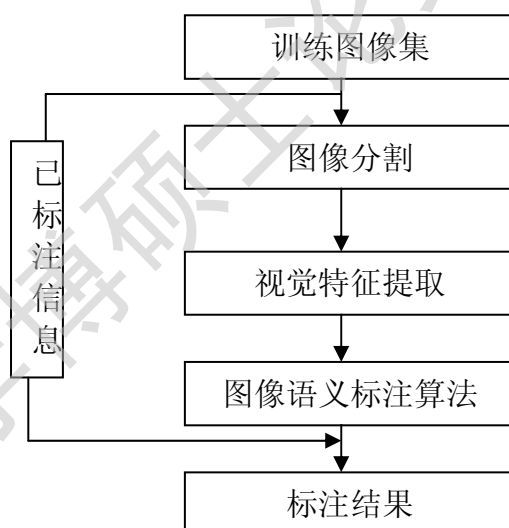


图1-1 图像自动标注流程图

本小节的后面部分将对图像分割、图像视觉特征提取以及现有的图像标注模型进行介绍。

1.2.1.1 图像分割

图像分割(Image Segmentation)是计算机视觉领域中最基本的一项技术，是图像分析、图像理解、模式识别以及图像检索等领域的关键步骤。所谓图像分割就是根据图像某些特征集合的相似性准则，对图像像素进行聚类分组，把图像划分成许多有特殊意义的不同区域，区域之间互不相交，并且满足区域属性一致性的

条件。经过分割后的图像更能有效的在对象层中描述图像的特征，使图像检索更加符合图像的语义。分割的结果直接影响之后进行的图像区域特征提取和匹配，进而影响到图像标注结果的准确性。目前为止，人们在图像分割领域中已经提出了很多方法^[12,13,14,15,16]，但都没有很好地解决这个问题，图像分割仍然是计算机视觉和图像处理领域中的一个热点和难点问题。

直观上讲，图像分割就是将一幅图像分割成视觉相对独立的若干区域，每个区域代表着某个对象。图 1-2 为图像 a 比较理想的分割结果，图像 a 被分割成为图像区域 b 和图像区域 c，其中，b 区域为天空部分，c 为飞机。

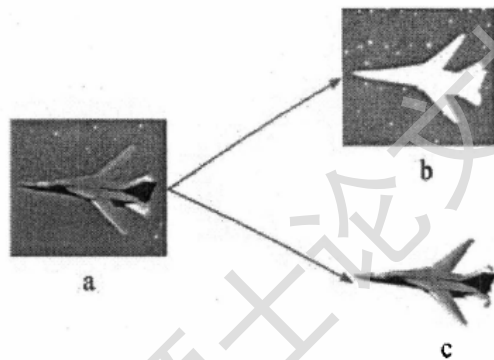


图1-2 图像分割示意图

图像分割的方法可以归结为 4 类典型的方法^[17]包括阈值法、边缘检测法、区域法和聚类法。

1、阈值法

阈值(Threshold)法是最为常用的图像分割方法，其实质是利用图像的灰度直方图信息得到用于分割的阈值。阈值法又包括全局阈值法和局部阈值法。在阈值检测方面，使用比较普遍的阈值法包括 Otsu 利用类别方差作为判别依据，提出的类别方差法^[18]，以及 Pun 提出的利用最大先验熵估计分类和理性来选取阈值的方法^[19]等。基于阈值算法的优点是实现简单，运算速度快，缺点是难以处理包含多个前景物体的情况。

2、边缘检测法

边缘是指图像周围像素灰度有跳跃变化的那些像素的集合。它存在于目标与背景之间，是图像分割所依赖的最重要的特征。基于边缘的分割就是利用边缘上像素灰度值的不连续性，试图通过检测不同区域间的边缘来解决图像分割问题。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库