

大规模场景图像的情感语义分析 若干关键技术研究

曹建芳 著



知识产权出版社

全国百佳图书出版单位

大规模场景图像的情感语义分析 若干关键技术研究

曹建芳 著

常州大学图书馆
藏书章



知识产权出版社

全国百佳图书出版单位

图书在版编目(CIP)数据

大规模场景图像的情感语义分析若干关键技术研究 / 曹建芳著. —北京: 知识产权出版社, 2018.1

ISBN 978-7-5130-5295-5

I. ①大… II. ①曹… III. ①图象处理软件-研究 IV. ①TP391.413

中国版本图书馆CIP数据核字(2017)第291916号

内容提要

本书围绕场景图像的情感语义理解展开研究,从情感语义数据的获取、自动标注、情感语义类别的预测及大规模场景图像数据的高效检索等方面进行了探讨和研究,系统地阐述了场景图像情感语义分析的关键技术。从理论上,对场景图像蕴含的情感语义进行了抽象、分析和形式化表示;在实践上,搭建了实验平台进行了验证和分析,为各类图像数据的情感语义理解提供了新的思路和途径。全书集理论、技术、方法及实践于一体,具有较强的理论性和实践性,反映了当前该理论的最新研究成果。

本书可作为计算机科学与技术相关专业的本科生及硕士生教材,对相关领域的研究人员和工程技术人员也有较高的参考价值。

责任编辑: 彭喜英

责任出版: 刘译文

大规模场景图像的情感语义分析若干关键技术研究

DAGUIMO CHANGJING TUXIANG DE QINGGAN YUYI FENXI RUOGAN GUANJIAN JISHU YANJIU

曹建芳 著

出版发行: 知识产权出版社有限责任公司 网 址: <http://www.ipph.cn>

电 话: 010-82004826

<http://www.laichushu.com>

社 址: 北京市海淀区气象路50号院

邮 编: 100081

责编电话: 010-82000860 转 8539

责编邮箱: pengxyjane@163.com

发行电话: 010-82000860 转 8101

发行传真: 010-82000893

印 刷: 北京嘉恒彩色印刷有限责任公司

经 销: 各大网上书店、新华书店及相关专业书店

开 本: 720mm×1000mm 1/16

印 张: 8.75

版 次: 2018年1月第1版

印 次: 2018年1月第1次印刷

字 数: 116千字

定 价: 42.00元

ISBN 978-7-5130-5295-5

出版版权专有 侵权必究

如有印装质量问题,本社负责调换。

前 言

随着多媒体技术、互联网技术及社交网络的迅速发展,人们可以访问的信息资源空前丰富。图像数据因其形象直观、蕴含信息综合性强等特点,应用领域逐渐增多,其数量更是以惊人的速度增长。但图像数据本身结构的复杂性、蕴含信息的多样性及时空的多维性导致如何有效组织和管理大规模图像数据、快速检索用户需求的图像成为学术界的研究热点。为此,图像的情感语义分析和检索技术应运而生,它综合人工智能、计算机视觉、模式识别、心理学及数据库管理等领域的相关知识,对图像数据蕴含的高层情感语义进行分析,旨在获得图像蕴含的内在情感语义信息,建立实用性强的图像检索系统。因此,对图像进行情感语义分析和高效检索技术的研究有着广阔的应用前景和实用价值。

本书以场景图像为研究对象,对SUN Database中的各类场景图像进行有针对性的情感语义分析和检索方法研究。从建立开学行为学实验环境下的场景图像情感语义数据获取平台开始,选择并改进了OCC情感模型,分析了场景图像语义理解方面存在的语义模糊性问题,采用粒子群(PSO)算法优化BP神经网络的权值和阈值,并由Adaboost算法组合15个BP神经网络的输出结果,构建强预测器,对场景图像的情感语义类别进行预测,搭建了适合大数据处理的基于MapReduce并行编程模型的场景图像检索平台,系统地研究了场景图像的情感语义分析和检索方法。

全书共分为七章:第1章是绪论,介绍了场景图像情感语义分析的研究现状和本书的研究内容、组织结构;第2章介绍了大数据处理与图像检

索之间的关系；第3章探讨了开放行为学实验环境下的场景图像情感语义分析方法；第4章研究了基于模糊理论的场景图像情感语义标注方法；第5章对基于 Adaboost-PSO-BP 神经网络的场景图像情感语义类别预测算法进行了探讨；第6章研究了基于 MapReduce 并行编程模型的大规模场景图像检索技术；第7章对本书的研究工作进行了总结和展望。

值本书出版之际，我要特别感谢我的博士生导师陈俊杰教授，他不仅在学术上给予我悉心的指导，而且在工作和生活方面也给了我无私的帮助，在这里谨向恩师表示真挚的感谢和诚挚的敬意！同时感谢忻州师范学院计算机系的领导和老师们，在他们的支持、鼓励和帮助下，我顺利地完成了本书的撰写工作！

本书的出版得到了山西省自然科学基金（No. 201701D121059）和山西省教育科学”十三五“规划课题（No. GH-17059）的资助，在此一并表示感谢！

本书的内容中有一部分内容反映了场景图像情感语义分析和检索的最新研究成果、研究方法和研究动向，在理论体系和方法上均有创新，构建了场景图像情感语义关键技术分析的平台。本书可作为计算机应用技术、信息科学、工程技术等专业高年级本科生和研究生的教材，对相关领域的研究人员和工程技术人员也有重要的参考和使用价值。

由于作者才疏学浅，书中疏漏在所难免，恳请各位专家学者批评指正，并提出宝贵意见。

曹建芳

2017年8月

目 录

第1章 绪 论	001
1.1 研究背景与意义	001
1.1.1 研究背景	001
1.1.2 研究意义	005
1.2 国内外研究现状	007
1.2.1 图像的情感语义分析	007
1.2.2 图像检索技术研究现状	014
1.3 本书主要工作	017
1.3.1 场景图像的情感语义特征提取研究	018
1.3.2 基于模糊理论的场景图像自动标注方法研究	018
1.3.3 场景图像的情感类别预测研究	019
1.3.4 基于MapReduce的大规模场景图像检索技术研究	020
1.4 本书组织结构	020
1.5 本章小结	022
第2章 大数据处理与图像检索	023
2.1 大数据的种类、特点和应用	023
2.1.1 大数据的种类	023
2.1.2 大数据的特点	025
2.1.3 大数据的应用	026
2.2 大数据处理面临的问题	029
2.3 大数据处理与图像检索的关系	030

2.4 本章小结	032
第3章 开放行为学实验环境下的场景图像情感语义分析	033
3.1 图像情感语义理解的相关概念	033
3.1.1 情感与情感计算	033
3.1.2 情感建模	034
3.1.3 图像情感语义分析	035
3.1.4 图像情感语义标注	036
3.1.5 图像情感语义检索	037
3.2 图像语义层次模型	038
3.3 开放行为学实验环境下场景图像的情感语义分析	040
3.3.1 情感模型的选择	040
3.3.2 被试的选取	042
3.3.3 实验数据和方案设计	043
3.3.4 场景图像情感语义数据分析方法	044
3.3.5 场景图像的情感语义数据实验分析	045
3.4 图像分析和检索的性能评测	049
3.5 本章小结	051
第4章 基于模糊理论的场景图像情感语义标注方法	052
4.1 模糊理论	052
4.1.1 概述	052
4.1.2 基本定义	054
4.2 颜色视觉特征提取	055
4.2.1 颜色空间的选取	055
4.2.2 HSV 颜色空间量化	058
4.2.3 基于权重的不规则分块场景图像颜色特征提取	059
4.3 场景图像的模糊语义描述	061

4.3.1	情感值的确定	061
4.3.2	情感规则 G	062
4.3.3	情感规则映射 T	063
4.4	基于T-S模糊神经网络的场景图像的情感语义特征映射	063
4.4.1	T-S模糊神经网络(T-S FNN)	064
4.4.2	语义特征映射	066
4.5	场景图像的情感语义自动标注	069
4.6	实验结果及分析	070
4.7	本章小结	072
第5章	基于Adaboost-PSO-BP神经网络的场景图像情感语义类别预测算法	074
5.1	融合情绪、性格因素的OCC情感模型	074
5.1.1	情绪因素描述	075
5.1.2	性格因素描述	078
5.1.3	融合情绪、性格因素的情感建模方法	079
5.2	BP神经网络权值和阈值的优化	080
5.2.1	BP神经网络	080
5.2.2	粒子群优化算法(PSO)优化BP神经网络	082
5.3	Adaboost-PSO-BP神经网络预测算法	085
5.3.1	Adaboost算法	085
5.3.2	Adaboost-PSO-BP神经网络算法	085
5.4	场景图像情感语义类别预测	087
5.5	实验结果与讨论	088
5.6	本章小结	092
第6章	基于MapReduce的大规模场景图像检索技术	093
6.1	Hadoop平台相关技术介绍	093

6.1.1 Hadoop的起源和背景	093
6.1.2 HDFS体系结构	095
6.1.3 MapReduce编程模型	096
6.1.4 Mahout算法库	098
6.2 基于MapReduce的大规模场景图像检索方案	101
6.2.1 场景图像检索整体框架	101
6.2.2 大规模场景图像及其特征的存储	104
6.2.3 场景图像的特征提取	106
6.2.4 基于分布式Mean Shift的场景图像特征聚类算法	106
6.3 实验与结果分析	110
6.3.1 实验环境与测试数据	110
6.3.2 系统性能测试与分析	110
6.4 本章小结	115
第7章 总结与展望	116
7.1 本书工作总结	116
7.2 研究展望	118
参考文献	120

第1章 绪 论

场景图像是指在真实环境中,由具有合理空间分布构成的背景和一些分散的物体对象组成的连贯图像,是一类非常常见的图像。伴随着大数据时代的到来,场景图像作为蕴含丰富语义信息的载体,其数量呈爆炸式增长,传统的图像特征提取及检索方法已经显得力不从心,我们迫切需要从情感语义的层面分析场景图像并对其进行情感语义检索。本章主要介绍了场景图像情感语义分析及检索的研究背景和意义、国内外研究现状及本书的主要工作。

1.1 研究背景与意义

1.1.1 研究背景

近年来,随着互联网和多媒体技术的飞速发展,大数据时代悄然而至。全球数字媒体资源数量正以惊人的速度增长,每天都会产生数以万兆字节的图像。作为一种蕴含丰富语义的信息载体,图像蕴藏着比文本更丰富的信息,其本身易于超越文化、种族和时间障碍,传递更丰富的情感和意境。因此,如何使用计算机提取图像的情感语义信息因其直接影响图像的检索效率而在众多的研究和应用领域引起了广泛关注。为了有效地组织和管理这些海量的图像,人们急需获取图像的情感语义检索各类图像库,

从而使得检索结果更加符合人们对图像的实际理解。场景图像是人们在日常生活中最常见的一类图像,对其进行情感语义分析和检索技术研究是实现各类图像情感语义检索的基础,其目标是从用户对图像理解的角度出发,即从情感语义层面,迅速、准确地从海量图像数据中找到所需要的图像数据,最终达到满足用户需求的目的。一直以来,对场景图像没有一个统一的定义,比较经典的是Henderson和Hollingworth 1999年在*High-level scene perception*一文中将场景图像定义为“由空间分布合理的背景和离散的物体构成真实环境的连贯图像”^[1]。从这个意义上讲,场景图像一般由背景和物体两部分组成。图1-1是SUN Database^[2]中已经标注的部分场景图像,可以看出,该数据库只对场景图像中的背景和物体进行了标注,从而也只能按照图像中的背景和物体进行检索。事实上,“触景生情”,当人们看到一幅图像时,就会油然而生不同的情感,从情感层面去理解看到的图像。但目前SUN Database以及其他一些常用图像库、还有一些自建图像库大多是按照图像的背景及图像中包括的物体对象进行标注、分类、检索的,这样导致检索的结果有时会与人们的需求大相径庭。例如,图1-1中的(8)和(9)、(19)和(20),它们的标注结果完全一样,但人们观察这两组图像时,产生的情感和对它们的理解可能是完全不一样的。面对越来越多的场景图像,人们更多地需要分析场景图像的情感语义,实现高效的情感语义检索。

20世纪70年代,主要使用的是基于文本的图像检索(Text-based Image Retrieval, TBIR)方法,利用文本描述的方式表示图像的特征^[3]。其本质是手工对图像进行标注,然后利用数据库管理系统的查询机制,用查询关键字与图像库中标注的词语进行匹配完成检索。在早期Internet环境下,百度、Google、Yahoo等搜索引擎采用的都是TBIR技术,优点是使用成熟的文本检索和搜索引擎技术,实现简单;缺点是手工标注的准确性差,不能满足用户对图像原始特征信息的检索,更不会满足用户从情感语义的角

度理解图像的需求。其检索流程^[3]如图1-2所示。



图 1-1 场景图像示例

Fig. 1-1 The Examples of Scene Images

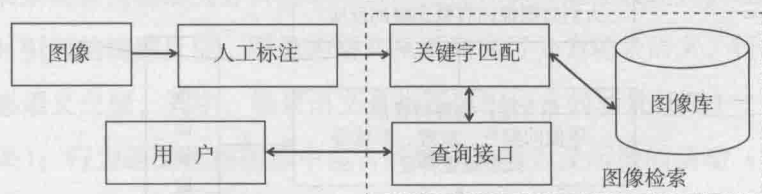


图 1-2 基于文本的图像检索流程

Fig. 1-2 The Flow of Text-based Image Retrieval

这种传统的检索技术强调的是文本注解，但文本注解主观性较强，也无法涵盖图像的全部内容，而且随着大数据的发展，人工标注文本的工作

量日益增大,因此,TBIR技术逐渐成为其他图像检索技术的辅助手段,而不再是图像检索的主流技术。

为了解决TBIR方法因人工标注带来的问题,1998年10月,国际标准化组织ISO/IEC提出了MPEG-7国际标准——多媒体内容描述接口(Multimedia Content Description Interface)的制定,该标准为各类多媒体数据提供了一种与描述内容相关的标准化描述,大大促进了用户对各类多媒体数据的快速查询和访问^[4]。20世纪90年代,基于内容的图像检索技术(Content-based Image Retrieval, CBIR)应运而生,该方法对图像的视觉内容(颜色、纹理、形状等)进行分析并检索图像,其特点是不需要人为干预和解释图像包含的客观视觉特性,而是让计算机自动提取和存储图像特征^[5]。图1-3是图像内容的层次模型,CBIR技术利用第2层的低层视觉特征进行检索,特征提取是CBIR系统的基础,在很大程度上决定CBIR系统的成败^[6]。当前主要的技术有基于颜色特征的图像检索、基于纹理特征的图像检索、基于形状特征的图像检索和基于空间特征的图像检索等。与TBIR技术相比,CBIR技术在图像研究领域取得了重大突破,它从图像自身的内容出发,更能符合用户的实际需求,因此是目前比较流行的图像检索技术,其关键技术主要集中在研究合理的特征提取方法和相似性度量标准上。

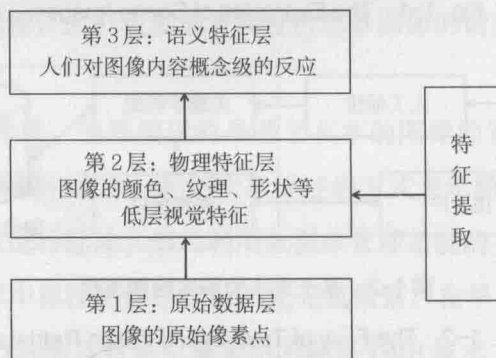


图 1-3 图像内容的层次模型

Fig. 1-3 The Hierarchical Model of Image Content

CBIR技术在检索时直接比较图像的低层视觉特征的相似度,并未考虑图像的高层语义,而图像的低层视觉特征与高层语义间存在巨大的“语义鸿沟”(Semantic Gap)(图1-3中第2层和第3层之间的差别),且人们在通常情况下是根据情感语义去观察和理解图像的,从而使得CBIR检索方法在实际应用中无法很好地满足用户需求。要架起“语义鸿沟”的桥梁,必须提取图像的高层语义特征。图像的语义检索主要包括对象语义、空间关系语义、场景语义、行为语义、情感语义及更高层的语义的检索。图像情感语义检索的目的就是使计算机检索图像的能力,达到人对图像的真实理解的水平,对图像进行情感语义分析并研究基于情感语义的检索技术目前是数字图像理解领域的一个研究热点,因此针对这一课题的深入研究具有很高的理论价值和广泛的应用前景。

1.1.2 研究意义

场景图像的情感语义分析是图像高层语义理解、模式识别和计算机视觉领域的重要研究内容,在处理许多实际问题中,如图像标注、分类、检索、人脸识别、室外监控、军事侦察等,都需要对场景图像先分析人的情感行为,提取其情感语义特征,然后再通过特征相似度计算等解决实际问题。场景图像情感语义分析的最终目的是使计算机能够表述人们观察场景图像时引起的情感反应。图像的语义从低到高可分为场景语义、行为语义和情感语义三层,其中,场景语义是指图像中包含的场景(图1-1中的标注结果);行为语义是指图像中包含的物体的行为及所做的活动(如一场联欢晚会等);情感语义是指图像给人们带来的主观感受(如愉悦、生气等),它属于图像语义中的最高层语义^[7]。图像情感语义特征的提取以图像低层视觉特征为基础,首先,通过相关的图像处理技术提取图像的颜色、纹理、形状和轮廓低层特征,其次,寻找图像低层特征与高层情感语义的

析方法, 目前仍是研究的热点和难点。

多图像与我们想要的结果差距很大。

检索的基础, 因此有着很强的理论研究价值和广阔的应用前景。



(a)



(b)

图 1-4 百度搜索引擎检索结果示例

Fig. 1-4 The Retrieval Results of Baidu Search Engine

1.2 国内外研究现状

自 20 世纪 90 年代 CBIR 技术诞生以来,国内外研究学者对图像低层和高层特征提取和图像检索技术进行了大量研究和实验,力图找到描述图像的有效特征和高效的图像检索方法。本节将分别介绍图像的情感语义分析和检索技术的研究现状。

1.2.1 图像的情感语义分析

情感问题的研究一直被视为心理学领域的研究内容,随着计算机技术的快速发展,国内外研究学者逐渐将情感问题的研究引入计算机学科中,形成了“情感计算”研究方向。早在 20 世纪 80 年代,美国麻省理工学院的媒体实验室就从计算机对人的情绪及感觉的感知开始了有关情感计算的研究^[9]。日本从 20 世纪 90 年代中期进行“感性工程”研究,旨在让计算机处理感性信息,实现“以人为本”^[10]。我国于 2003 年 12 月在北京成功举办

了“第一届中国情感计算及智能交互学术会议”，正式拉开了我国研究在该领域的序幕^[11]。2005年10月，“首届国际情感计算及智能交互学术会议”在北京开幕，围绕人机情感交互展开研讨，进一步推动了情感计算、智能交互这一前沿领域科学研究的发展^[12]。对于图像的高层情感语义，其研究在国内外如火如荼，但因情感的抽象性和主观性，图像的情感语义分析有着自身的特点。近年来研究的重难点主要有：①图像情感语义的表示方法及情感建模；②图像情感语义特征的提取。

1.2.1.1 图像情感语义的表示方法及情感建模

心理学家使用“维量”分析方法研究情感问题^[13]，施洛伯格提出了情绪的三个维量：愉快—不愉快，注意—拒绝，睡眠—紧张；奥斯古德提出了针对演员的表情的三个维量：愉快—不愉快，强度和控制；弗里达提出了六个维量：愉快—不愉快，激活，注意—拒绝，社会评价，惊奇，简单—复杂；后来普拉奇克使用一个倒立的锥体描述了情绪的复合维模式，刻画了八种情绪（喜悦—悲伤，赞同—反感，预期的一出乎预料的，恼怒—恐惧）和三种维量（强度、相似性和极性）。维量思想在情绪分析中起着重要作用，但其难点是分析维的含义和名称至今没有一个统一的标准。

在计算机领域，研究学者主要研究图像的视觉特征与情感理解的关系及情感的建模。Yuichi Kobayashi等^[14]通过实验证明了颜色和方向多分辨率的对比对人的主观感知的重要性。毛峡等^[15]通过分析图像的情感特征建立了一个二维波动数学模型，提出了一种图像波动分析方法，对图像给出了和谐感评价，实验证明了符合“ $1/f$ ”波动规律的图像能给人以和谐与美的感觉。王上飞等^[16]等从心理学的“维量”思想出发，使用语义量化技术和因子分析方法建立了情感空间。Yoshida等^[17]定义了图像的三种情感感受：舒适、杂乱、单调。Sung-Bae Cho等^[18]对图像讨论了高兴、沮丧和凉爽三