



图像语义信息可视化 交互研究

HUMAN COMPUTER INTERACTION OF
IMAGE SEMANTICS VIA VISUALIZATION

陆泉 著



國家圖書館出版社
National Library of China Publishing House

国家社科基金后期资助项目研究成果

图像语义信息可视化交互研究

陆 泉 著



國家圖書館出版社
National Library of China Publishing House

图书在版编目(CIP)数据

图像语义信息可视化交互研究/陆泉著. --北京:国家图书馆出版社,2016.7

ISBN 978-7-5013-5825-0

I. ①图… II. ①陆… III. ①图像分析—语义分析—研究
IV. ①TP391.41

中国版本图书馆 CIP 数据核字(2016)第 108336 号

书 名 图像语义信息可视化交互研究
著 者 陆 泉 著
责任编辑 金丽萍

出 版 国家图书馆出版社(100034 北京市西城区文津街7号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真),66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京华艺斋古籍印务有限责任公司

版 次 2015年7月第1版 2015年7月第1次印刷

开 本 710×1000(毫米) 1/16

印 张 33.5

字 数 580千字

书 号 ISBN 978-7-5013-5825-0

定 价 120.00元

国家社科基金后期资助项目

出版说明

后期资助项目是国家社科基金设立的一类重要项目,旨在鼓励广大社科研究者潜心治学,支持基础研究,多出优秀成果。它是经过严格评审,从接近完成的科研成果中遴选立项的。为扩大后期资助项目的影响、更好地推动学术发展、促进成果转化,全国哲学社会科学规划办公室按照“统一设计、统一标识、统一版式、形成系列”的总体要求,组织出版国家社科基金后期资助项目成果。

全国哲学社会科学规划办公室

前 言

本书是2013年国家社科基金后期资助项目(第三批)“图像语义信息可视化交互研究”(批准号:13FTQ006)的主要成果。

随着互联网的发展,图像信息在网络应用中的地位日趋重要。就现有研究与应用而言,图像语义信息仍然是组织、管理与检索图像信息资源的主要途径。但是,由于图像及其语义信息的复杂特性,图像与图像语义中普遍存在语义鸿沟(Semantic Gap)问题,使计算机系统与图像用户均难以通过语义信息对图像进行准确描述,不利于图像信息资源的有效管理与利用。

目前,虽然已有许多图像语义与图像用户方面的研究与应用,但是现有研究往往只关注图像、图像语义或图像用户之一,很少从系统科学的角度,将图像、图像语义与图像用户作为相互联系、互相作用的整体来进行研究,这不利于从系统角度对图像、图像语义与图像用户建立全面深刻的认识。基于此,本书系统地梳理了现有的图像语义及图像用户行为研究,深入分析了图像语义鸿沟问题,从图像、图像语义与图像用户的整体系统视角,以图像语义信息可视化交互为三者结合的典型,开展图像语义信息及其用户行为的理论与实验研究。

本书的基础是对现有图像语义及图像用户行为研究进行调研、分析与述评,重点分析图像语义鸿沟问题,为全书从图像、图像语义与图像用户的整体系统视角开展研究明确了研究环境与主攻方向;核心是围绕图像语义信息可视化交互的一系列基本理论问题,以图像语义标注为典型用户行为进行实验研究:首先介绍研究方法、研究平台与研究数据集,然后对标注影响因素、标注耗时影响因素、用户行为与心理等子问题进行研究,并对基本标注方法进行了比较研究,得到一系列新结论,回答图像语义信息可视化交互中的一系列基本理论问题;扩展研究探索了图像语义与图像用户行为研究前沿问题,从图像、图像语义、图像用户与计算机系统等多个不同主体作用关系的视角,研究了语义辅助对图像用户行为的影响、图像用户行为与图像底层特征的相关性等热点问题。

本书试图突破现有图像语义与图像用户理论的局限性,从图像、图

像语义与图像用户的系统视角,研究图像语义信息可视化交互的基本理论。通过提出一系列新的研究问题,设计一些新的组合研究方法,得出了较系统化且富于理论意义的研究结论,其研究视角新颖,学术观点与理论体系具有基础性与原创性。本书对图像信息资源管理与利用以及图像用户研究有重要推动作用,对促进信息管理学科及其他相关学科的研究发展有重要启发意义。

衷心感谢武汉大学信息管理学院马费成教授、李纲教授、陆伟教授、查先进教授、唐晓波教授,华中师范大学夏立新教授、王伟军教授、李延辉教授、陈静副教授,威斯康星大学密尔沃基分校的 Jin Zhang 教授、Iris Xie 教授、穆祥明副教授,以及俄克拉荷马大学的 Kun Lu 博士等多位专家对本研究给予的大量无私的建议和帮助。此外,书中还引用了众多国内外同行的研究成果,谨此一并感谢。

有多位武汉大学及华中师范大学的本、硕阶段学生参与到本书研究与写作过程中。其中,韩阳、汪艾莉、郭怡婷、刘承谕及黄茜等参与了基础篇的研究与校对工作,刘高、韩阳、韩雪、赵琴、王宝、周思瑶及郑丁益等参与了核心篇的研究与校对工作,郭怡婷、金炜及吴孟澍等参与了扩展篇的研究与校对工作,来自上述两所大学的共 239 位同学参与了本书中的多个科学实验及数据处理,在此对他(她)们的参与和帮助一并表示感谢。

本书在实验研究中使用了国际情感图片系统(International Affective Picture System, IAPS)数据,在此对佛罗里达大学的 NIMH Center for Emotion and Attention(CSEA)授权本研究使用其数据集表示感谢;另外,研究中还利用了众多软件工具及数据服务,特别是来自 flicker、Google 及百度,在此也一并表示感谢。

借此也向本书出版过程中的所有相关人员表示感谢。

由于水平有限,书中错误或不妥之处在所难免,希望读者批评指正。

陆 泉

2015 年 6 月于珞珈山

目 录

前 言	(1)
第一章 图像语义信息	(1)
第一节 图像特征提取与分类研究	(1)
第二节 图像信息资源组织与检索研究分析	(9)
第三节 图像语义信息应用分析	(37)
本章小结	(44)
第二章 图像用户行为	(45)
第一节 图像用户行为研究	(45)
第二节 图像用户行为相关应用分析	(84)
本章小结	(89)
第三章 图像语义鸿沟问题	(90)
第一节 图像语义层次理论	(90)
第二节 图像语义鸿沟问题的研究	(95)
本章小结	(101)
第四章 图像用户行为研究方法与研究平台	(102)
第一节 一个国外用户行为研究的实验室案例	(102)
第二节 图像语义可视化交互方法	(108)
第三节 一个图像语义可视化交互标注研究平台	(118)
本章小结	(131)
第五章 图像语义标注用户行为数据集开发	(132)
第一节 数据集开发实验	(132)
第二节 数据集处理	(147)
本章小结	(179)

第六章 图像语义人机交互标注影响因素研究	(180)
第一节 引言	(180)
第二节 研究的理论与技术基础	(190)
第三节 用户图像标注影响因素的理论分析	(195)
第四节 用户图像标注影响因素模型的实验验证	(200)
第五节 实验结果及分析	(202)
第六节 用户图像标注影响因素模型修正与应用启示	(220)
本章小结	(223)
第七章 图像语义人机交互标注方法的比较研究	(225)
第一节 引言	(225)
第二节 相关理论及技术基础	(237)
第三节 图像情感语义社会化标注基本方法分类	(246)
第四节 图像情感语义强度社会化标注基本方法的比较指标	(253)
第五节 图像情感语义强度社会化标注基本方法的比较实验	(257)
第六节 实验数据统计及结果分析	(260)
第七节 图像语义的人机交互认知机理分析	(279)
本章小结	(282)
第八章 图像语义可视化交互标注耗时研究	(286)
第一节 引言	(286)
第二节 研究的理论与技术基础	(301)
第三节 影响用户图像情感标注耗时的因素模型构建	(322)
第四节 影响用户图像情感标注耗时的实验设计	(329)
第五节 实验结果分析	(335)
第六节 讨论与启示	(353)
本章小结	(358)
第九章 图像语义可视化交互的用户行为与心理研究	(360)
第一节 用户行为模式	(361)
第二节 用户标注心理	(385)

本章小结	(392)
第十章 语义辅助对图像用户行为的影响研究	(394)
第一节 研究问题	(394)
第二节 实验系统设计开发	(399)
第三节 实验用数据集的采集	(405)
第四节 数据分析	(409)
本章小结	(417)
第十一章 图像用户行为与图像底层特征的相关性	(419)
第一节 引言	(419)
第二节 相关研究述评	(423)
第三节 研究的理论和技术基础	(428)
第四节 研究设计与数据获取	(433)
第五节 研究结果分析	(446)
第六节 讨论与应用启示	(461)
本章小结	(468)
第十二章 总结与展望	(470)
参考文献	(472)
附录1 实验前问卷	(508)
附录2 实验后问卷	(509)
附录3 MATLAB 计算图像特征向量集源代码	(511)
3.1 能量、对比度、二维熵、相关性、同质性计算源代码	(511)
3.2 一维熵计算源代码	(512)
附录4 用户检索实验相关数据表格	(514)
4.1 用户检索实验中采纳的图片的底层特征数据表	(514)
4.2 检索实验中每个任务的点击选项卡次数及耗时表	(518)
4.3 检索实验中用户选择更改前后两张图的底层特征表	(522)
附录5 Flickr 图像集图像底层特征与标签数示例	(524)

第一章 图像语义信息

图像信息资源数量巨大、增长迅速,在人类社会生活中发挥着越来越大的作用。近些年来,人们对图像信息资源的需求日益增长,对图像信息资源管理水平要求也不断提高。相关的理论、技术与应用研究不断涌现,有必要对其进行系统梳理。

因此,本章将重点从图像及其语义的计算机处理理论与技术角度,首先,对图像语义信息研究现状进行综述,以系统梳理和归纳图像语义信息研究的主要领域与主要问题;其次,针对其中发展迅速的图像标注与图像检索两个热点领域,专门对其在海外的研究动态进行调查与梳理,以方便国内读者跟踪掌握国外相关研究进展;最后,选择三个典型应用领域,对图像语义信息应用进行了调研分析,方便读者了解图像语义信息领域理论与实践的结合情况。

第一节 图像特征提取与分类研究

图像与文本信息有着较大的差异,其高维的视角特征与复杂的语义内容难以被计算机有效理解,因此,在计算机中,如何提取与表示图像的特征,是通过计算机理解与管理图像的首要问题,在此基础上,对图像语义的相关研究主要从图像语义组织、图像检索与图像分类三方面展开,本节重点介绍图像特征提取与图像表示,以及图像分类基本理论方法研究,图像语义组织与图像检索研究将在下一节详细述评。

一、特征提取与图像表示

如何从原始图像中提取具有较强表示能力的图像特征是智能图像

处理的一个研究热点。图像特征提取包括基于全局的图像特征提取和基于局部的图像特征提取,而基于局部的图像特征提取已经逐渐成为当前研究趋势,基于局部的图像特征提取首先要求对图像进行分割。本部分将依据不同的图像特征从图像分割、颜色特征、纹理特征、形状特征、空间关系提取等方面对图像特征提取研究进行梳理。

(一) 图像分割

图像分割是一种重要的图像分析技术,它不仅得到人们广泛的重视和研究,也在实际中得到大量的应用。图像分割通常是提取图像特征并进行图像表示的第一步。图像分割是指依据图像特征的同质性将图像分成不同的区域,并使这些区域互不相交,且每个区域具有特定的一致性特点。图像分割的方法和种类有很多,有些分割运算可直接应用于任何图像,而另一些只能适用于特殊类别的图像。目前用于自动图像标注的图像分割方法主要包括基于网格的方法、基于聚类的方法、基于轮廓的方法、基于统计模型的方法、基于图表的方法等。

由于自动图像分割的复杂性,许多研究者采用基于网格的方法将图像划分成不同区域,并从这些区域中进行特征提取^①。但是这一技术对图像语义构成的划分尚不理想,划分的单个区域往往包含多个不同物体,并且很难决定划分区域的大小。这一方法常用于特定领域的应用,如医学图像分类等。

基于聚类分析的图像分割方法是图像分割领域中一类应用相当广泛的算法。聚类分析是以相似性为基础,依据像素聚类为不同群集,从中提取图像的颜色和纹理特征。如 k-means 算法,它先将数据点集分为 k 个划分,每个划分作为一个聚类,然后从这 k 个初始划分开始,通过重复的控制策略,使某个准则最优化,而每个聚类由其质心来代表。这一算法收敛速度快,但其主要缺点是必须事先定义划分的数量,不合适的 k 的选择可能导致不准确的分割结果。

基于轮廓的图像分割方法是将均匀区域看作是被一闭合边缘所包围,而在这个边缘处的像素往往会发生剧烈的变化,基于轮廓的方法希望通过检测这种像素的骤变来确定目标边缘。与基于聚类的分割算法

^① Vailaya A, Figueiredo M A T, Jain A K, et al. Image classification for content-based indexing [J]. IEEE Transactions on Image Processing, 2001, 10(1): 117—130.

不同,基于轮廓的分割算法不需要预先假设聚类的数量。这一方法对边缘定位准确,运算速度较快,但是无法保证提取边缘的连续性。与此同时,如何平衡边缘检测的抗噪性以及检测精度使得检测结果达到最好也是个难点,过度的提高抗噪性可能会使检测结果出现位置偏差和漏检,而过度强调检测精度又会因为图像噪声导致的目标伪边缘而产生错误的轮廓结果。

基于统计模型的图像分割方法也被提出,Carson 等人于 2002 年提出 Blobworld^①,提供一种新的由原始像素数据到颜色、纹理区域一致性的图像表示方法,这一系统使用户能够看到图像内部表示及查询结果。

近几十年来,研究人员提出了许多著名的分割算法,其中 N-Cut 和 JSEG 是具有代表性的算法^②。到目前为止,没有一种自动图像分割方法可以取得较为满意的分割效果。这是由于目前的图像分割往往采用自底向上的分割方法,而实际上,图像分割问题的本身不仅仅是自底向上的图像处理问题,也会是自顶向下的对象理解问题,而对对象理解往往需要更为复杂的对象知识^③。

(二) 颜色特征提取

颜色是图像描述的一个重要特征,为了提取颜色特征,首先应该将图像映射到一个颜色空间或模型。当前使用的颜色空间主要包括 RGB、HSI、HSV^④、LUV、HMMD 等。在确定的颜色空间中提取颜色特征,包括颜色直方图、颜色矩、颜色相关图、颜色集等。

图像颜色特征提取的一个最有代表性的方法就是颜色直方图,其中 HSV 空间是直方图最常用的颜色空间。图像颜色统计直方图描述了不同颜色在一幅图像中所占的比例,这一方法在图像的移动和旋转下仍具有稳定性。但颜色直方图并不能详细表示图像的空间信息,不同的图像

-
- ① Carson C, Belongie S, Greenspan H, et al. Blobworld: image segmentation using expectation-maximization and its application to image querying[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(8): 1026—1038.
 - ② Deng Y, Manjunath B S, Shin H. Color Image Segmentation [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Fort Collins, CO: IEEE, 1999.
 - ③ Hare J S, Lewis P H. Saliency-based models of image content and their application to auto-annotation by semantic propagation[J]. Proceedings of Multimedia & the Semantic Web, 2005.
 - ④ Plataniotis K N, Venetsanopoulos A N. Color Image Processing and Applications [M]. Springer, Berlin, 2010: 340—344.

可能产生相同的颜色直方图,并且颜色直方图的维数通常很高。

Stricker 和 Orengo 提出了颜色矩的方法^①。颜色矩是一种简单而有效的颜色特征表示方法,有一阶矩(均值, mean)、二阶矩(方差, variance)和三阶矩(斜度, skewness)等,由于颜色信息主要分布于低阶矩中,所以用一阶矩,二阶矩和三阶矩足以表达图像的颜色分布。颜色矩已证明可有效地表示图像中的颜色分布,该方法的优点在于不需要颜色空间量化,特征向量维数低,但实验发现该方法的检索效率比较低,因而在实际应用中往往用来过滤图像以缩小检索范围。

为了能够在大规模图像数据集中进行快速的搜索,Smith 和 Chang 等人提出了颜色集的概念^②。首先将 RGB 颜色空间转化为视觉上的一致化空间,如 HSV,并将颜色空间量化成若干个 bin,然后运用颜色自动分割技术将图像分为若干个区域,每个区域用量化颜色空间的某个颜色分量来索引,从而将图像表达成一个二进制的颜色索引表。在图像匹配中,比较不同图像颜色集之间的距离和颜色区域的空间关系。因为,颜色集表达为二进制的特征向量,可以构造二分叉树来加快检索速度,对大规模的图像集合十分有利。

Huang 提出颜色相关图用于图像检索,并证明在基于内容的图像检索中颜色相关图特征要优于比颜色直方图^③。传统的颜色直方图只刻画了某一种颜色的像素数目占像素总数目的比例,只是一种全局的统计关系,而颜色相关图可以看作 3D 的颜色直方图,它还表达了颜色随距离变换的空间关系,也就是颜色相关图不仅包含图像颜色统计信息,同时包括颜色之间的空间关系。

(三) 纹理特征提取

纹理特征是图像的另一重要特征。纹理是图像所有物体表面所具有的特性,它包含了物体表面的结构特征及物体间的关系。由于纹理特

① Stricker M, Orengo M. similarity of color images[J]. Proc. SPIE Storage and Retrieval for Image and Video Databases, 1995, 2420: 381—392.

② Smith J R, Chang S F. Single color extraction and image query[C]//International Conference on Image Processing, Proceedings. Washington, DC: IEEE, 1995: 528—531.

③ Huang J, Kumar S R, Mitra M, et al. Image indexing using color correlograms[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Juan: IEEE, 1997: 762—768.

征具有很强的识别性,被广泛应用于图像检索及语义学习技术中。纹理特征的研究主要集中于图像处理及计算机视觉领域。纹理分析的方法有多种,如空间自相关法、共生矩阵法、Tamura 方法等。

Tamura^①于1978年提出了与人的视觉感受相关的6个纹理特征,分别是粗糙度、对比度、方向性、相似性、规则性和粗略度。Mallat^②于1989年首先将小波分析引入纹理分析中之后,随后基于小波的纹理分析方法如雨后春笋般涌现出来。随着小波理论不断发展,小波分析在纹理特征提取中的应用也在不断发展。

目前常用的纹理特征提取方法主要包括统计方法、结构方法、模型方法。统计方法是基于像元及其邻域的灰度属性研究纹理区域中的统计特性。实践证明,GLCM^③(灰度共生矩阵)在基于统计的纹理特征提取方法中是一枝独秀,具有较强的稳定性和适用性。这一方法是建立在估计图像的二阶组合条件概率密度的基础上的。GLCM是描述在某一方向上,相隔 n 个像元距离的一对像元分别具有的灰度层出现的概率。GLCM能够导出14种纹理特征。尽管这一方法对图像纹理特征提取具有较好的鉴别能力,但其对像素级的复杂纹理分类应用仍然受限,因此不断有研究者尝试对其进行改进。

(四)形状特征提取

形状是识别物体的重要特征,利用物体形状特征进行图像检索已经得到广泛应用。Zhang和Lu^④将物体的形状特征提取方法分为基于轮廓的形状特征提取方法和基于区域的形状特征提取方法。基于轮廓的形状特征提取方法仅通过图像轮廓边缘的特征来计算图像的形状特征,而基于区域的形状特征提取方法则是通过计算整个区域特征来提取图像形状特征。图像的轮廓特征只是图像区域特征的一部分,所以图像形状的细微改变会引起图像轮廓很大的变化。因此,研究较多地使用基于区

-
- ① Tamurah, Moris, Yamawakit. Texture features corresponding to visual perception [J]. IEEE Trans on System, Man and Cybernetics, 1978, 8(6): 460—473.
 - ② Mallat S G. A theory for multi resolution signal decomposition: the wave let representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674—693.
 - ③ Park S B, Lee J W, Kim S K. Content-based image classification using a neural network [J]. Pattern Recognition Letters, 2004, 25: 287—300.
 - ④ Zhang D, Lu G. Review of shape representation and description techniques [J]. Pattern Recognition, 2004, 37(1): 1—19.

域的图像形状特征提取方法。2005 年 Yang^① 等利用贝叶斯模型解决图像分类问题,并将其应用于基于区域的图像特征提取。

(五)空间关系提取

图像的空间关系是指物体在图像中的位置以及图像中物体之间的关系。空间关系表明图像中分割的多个目标之间存在着一定的空间位置关系和方向性的关系,如图像的邻接与连接关系、图像的包容和包含关系等。常用的图像空间特征提取方法有 2 种:根据图像中的对象或者颜色等其他特征对图像进行分割后提取特征;把图像分割成规则的子块,分别对图像的每个子块进行特征提取。运用空间关系特征描述图像内容能起到更完备的功效,但是一旦图像或目标发生反转、旋转等变化时,空间关系特征发生的变化就非常明显。当采用空间特征关系以提高检索准确率时,一般不单独使用,而是经常和其他特征提取方法综合使用。

图像的对象空间关系对图像数字图书馆有着重要意义。考虑到人们对空间关系认知的主观性,Wang 等提出的基于模糊 k-NN 分类器的元数据自动生成框架,可自动生成能够描述图像的对象空间关系的模糊语义元数据,用来表达图像中两两对象之间的空间关系,如上、下、左、右、近、远、内、外等^②。

图像的底层视觉特征,如颜色、纹理、形状等的提取方法已经得到大量的研究。目前存在着许许多多的图像底层特征表示方法,如何自动筛选这些特征才能更有效地达到图像分类和检索的目的,是一个较为困难的问题。现有研究在大多数情况下,是依据经验和大量的实验结果来确定哪些特征更适合于所要研究的问题。未来可以考虑采用通过机器学习的方法来自动选择对当前问题有效的特征。另外,语义表示涉及计算机学科、心理学、语言学等多门学科。由于语义之间关系复杂,且语义具有模糊性,因此有效表示语义是非常困难的。虽然已有的语义表示方法

① Yang C, Dong M, Fotouhi F. Image content annotation using Bayesian framework and complement components analysis[C]//International Conference on Image Processing, Geneva, Italy: IEEE, 2005.

② Wang Y H, Makedon F, Ford J, et al. Generating fuzzy semantic metadata describing spatial relations from images using the R-Histogram[C]//Proceedings of the Fourth ACM/IEEE Joint Conference on Digital Libraries: Global Reach and Diverse Impact, 2004: 202—211.

在某些方面证明是有效的,但仍缺乏一种通用的表示方法。因此,在未来一段时间内,建立一个通用的能够广泛认可的语义表示方法极具挑战性,计算机视觉与深度学习领域的发展可能会提供较为根本性的解决办法。

二、图像分类研究

图像分类问题涉及计算机视觉、情报学、机器学习等领域,是一个集中了机器学习、计算机视觉、信息组织和图像处理等多个研究领域的交叉研究方向。早期的图像分类多是依靠人工加注标签来实现,无法由计算机直接分析图像的内容,近些年来随着图像处理和机器学习理论的发展,基于内容的图像分类技术变得可行并逐渐成为研究的热点。图像分类研究的关键技术在于特征获取方法及分类研究方法两个方面。

(一) 分类研究中的特征获取方法

特征区域求取是图像分类任务的关键环节之一。大致包括显著特征获取、随机特征获取和基于区域分割特征获取。由于特征提取方法在第一章有所介绍,这里将简单介绍随机特征提取的相关研究。

Maree 等采用随机窗的方法,在随机窗的位置、大小和角度上都采用随机的方式获取,并对获取后的区域进行尺寸归一化,使特征具备一定的旋转不变性和尺度不变性^①。但是完全由随机方式获取特征区域,特征区域的辨别性不强,而且由于每次获取的区域不一致,试验效果的稳定性不高。另外,J. Winn 等采用单个像素步进的方式进行 Dense 特征区域获取,在每一个像素点处,利用其周围像素求取 17 维的滤波向量^②。但由于该方法针对每个像素进行特征区域划定来求取特征向量,使得特征的分辨性不强,不利于分类训练。

(二) 分类方法研究

国内外的许多学者从不同的角度,对图像分类及其相关任务做了大

① Maree R, Geurts P, Piater J, et al. Random Subwindows for Robust Image Classification[C]// IEEE conference on computer vision and pattern recognition. San Diego, USA: IEEE, 2005, 1: 20—25.

② Winn J, Criminisi A, Minka T. Object categorization by Learned Universal Visual Dictionary [C]// International Conference on Computer Vision. Beijing, China, 2005, 2: 1800—1807.

量的研究。在分类算法方面学者们也进行了大量研究,有学习分类和无学习分类研究方法。基于学习的图像分类方法是图像分类研究的重点。

国外方面,Li^①等应用 LDA (Latent Dirichlet Allocation) 模型解决了自然场景图像的分类问题。支持向量机 (Support Vector Machines, 简称 SVM) 应用的典型流程是首先提取出图像的局部特征,并形成特征码,然后将每幅图像的局部特征所形成的特征单词的直方图来作为特征,最后通过 SVM 进行训练得到模型^②。斯坦福大学的 Li Fei-Fei 利用贝叶斯理论和潜在狄利克雷分析理论 (Latent Dirichlet Analysis) 对图像分类进行了深入研究,构造了经典的 Caltech 101 目标分类数据集^③。Yang 利用稀疏编码技术进行视觉词汇学习和特征编码,利用线性 SVM 分类,与非线性的直方图交叉核 SVM 相比,在不降低分类准确率的同时提高了分类器的训练和识别速度^④。

国内方面,陈海林提出了双金字塔匹配思想,即同时在图像特征空间和空域建立一个统一的多分辨率框架,构造了一个具有线性复杂度的 Mercer 核,提高了分类性能^⑤。董立岩等采用朴素贝叶斯对分割后的医疗图像中的尿液残渣颗粒进行分类取得了很好的效果^⑥。赵英等采用贝叶斯网络进行遥感图像的检索,而贝叶斯网络充分利用了变量之间的独立性和条件独立性关系,大大减小了为定义联合概率分布所需要指定的概率数目,同时也避免了朴素贝叶斯分类器要求所有变量都是独立的不足^⑦。

-
- ① Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003 (3): 993—1022.
 - ② Zhang J, Marszalek M, Lazebnik S, etc. Local features and kernels for classification of texture and object categories: A comprehensive Study [J]. International Journal of Computer Vision, 2007, 73 (2): 213—238.
 - ③ Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental Bayesian Approach tested on 101 object categories [J]. Computer Vision & Image Understanding, 2004, 106 (1): 178.
 - ④ Yang J, Yu K, Gong Y, et al. linear spatial pyramid matching using sparse coding for image classification [C] // IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 1794—1801.
 - ⑤ 陈海林, 吴秀清. 基于双空间金字塔匹配核的图像目标分类 [J]. 中国科学技术大学学报, 2010 (3): 313—320.
 - ⑥ 董立岩, 苑森森, 刘光远等. 基于贝叶斯分类器的图像分类 [J]. 吉林大学学报, 2007, 45 (2): 249—253.
 - ⑦ 赵英, 刘佳佳. 基于贝叶斯定理的遥感图像检索 [J]. 现代图书情报技术, 2006 (5): 36—39.