文章编号:1001-506X(2016)04-0936-13 网址:www. sys-ele. com

图像场景语义分类研究进展综述

顾广华[1],韩晰瑛[1],陈春霞[1],赵 耀[2]
(1. 燕山大学信息科学与工程学院,河北 秦皇岛 066004;
2. 北京交通大学信息科学研究所,北京 100044)

摘 要:场景语义分类是图像理解领域中一个重要的研究方向,涉及到信号处理、模式识别、计算机视觉和认知科学等多学科交叉。场景分类任务中,图像内容描述和分类判决是两大关键问题。图像内容描述力图得到关于场景图像最具判别意义的表示,而分类判决则对训练样本集的图像内容描述学习、训练,并建模得到某类场景图像区别于其他场景类图像的计算模型。目前,很多场景分类方法针对图像内容描述和图像分类进行了深入的研究,对室外人造场景、室外自然场景和室内场景图像进行分类,取得了较好的分类效果。然而,场景图像自身内容上的变化和差异,既会造成同一场景类内对象的差异性,同时也造成不同场景类之间图像的视觉相似性,特别是对于不同的室内场景类。因此,场景语义分类任务十分困难,是计算机视觉和认知心理学领域中一个颇具挑战性的难题。室外图像场景分类研究相对成熟,而室内图像场景分类研究却进展缓慢。本文综述了图像场景语义分类的研究进展,并分析了场景分类算法的性能,指出场景语义分类研究中存在的问题。

关键词:场景语义分类;特征提取;图像描述;主题模型;分类器设计
中图分类号:TP 391.4 **文献标志码**:A **DOI**:10.3969/j.issn.1001-506X.2016.04.31

Survey on semantic scene classification research

GU Guang-hua[1], HAN Xi-ying[1], CHEN Chun-xia[1], ZHAO Yao[2]
(*1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China;*
2. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Semantic scene classification is an important research in the field of image understanding. It is a multidisciplinary subject involving signal processing, pattern recognition, computer vision and cognitive sciences. The image representations and the classification decisions are the two key issues in the scene classification tasks. Image representations try to get the most discriminative descriptions of scene images. Classification decisions try to get a certain computational model different from other categories by learning and training the image samples. Currently, many scene classification methods have been proposed for the researches on the image representations and image classifications. These proposed methods perform the image classification for outdoor manmade scene images, outdoor natural scene images and indoor scene images, and achieve better classification results. However, the changes and differences of the content in the scene image itself can cause both the object differences within the class and the visual similarity between classes, especially for different indoor scene categories. So, it is a challenging problem for the semantic scene classification in the fields of computer vision and cognitive psychology. The research on the scene classification for outdoor images is relatively mature, but the indoor is on the opposite. This paper reviews the researches on the semantic scene classifications, analyzes the performance of the classification algorithms for scene images, and points out the problems of the semantic classifications.

Keywords: semantic scene classification; feature extraction; image representation; topic model; classifier design

收稿日期:2015-06-08; **修回日期**:2015-10-19; **网络优先出版日期**:2015-12-23。
网络优先出版地址:http://www.cnki.net/kcms/detail/11.2422.TN.20151223.1019.010.html
基金项目:国家自然科学基金(61303128);河北省自然科学基金(F2013203220);河北省高等学校科学研究青年基金(Q2012047)资助课题

0　引　言

图像包含的语义信息十分丰富。人眼不但可以观察到图像包含的全局空间结构信息和图像中局部的目标及其相对位置信息，而且还能感知图像蕴含的高层语义信息，比如图像的场景语义、行为语义和情感语义等[1]，即“千言不如一画”。人们在感知图像时更关心图像语义层面上的内容。近年来，网络图像数据爆炸性增长，互联网图像数据资源之间的信息互访越来越频繁。让计算机来管理图像资源，尤其是按照人类感知高层图像语义的方式对图像数据进行语义分类的问题迫在眉睫。图像场景理解不仅可以获取图像的整体信息，还可以感知到目标出现的上下景信息，为图像语义分类提供了一条思路。2006 年在麻省理工学院首次召开的场景理解研讨会上，场景语义分类被明确认定为图像语义分类中的一个关键课题。

很多研究机构都开展图像场景分类方向的研究。国外有斯坦福大学视觉实验室、普林斯顿大学视觉实验室、麻省理工学院视觉实验室、卡内基梅隆大学视觉与自主系统实验室和加利福尼亚大学伯克利分校机器视觉实验室等。网页 http://www.cs.cmu.edu/~cil/v-groups.html 上列出了国际上知名的科研机构关于计算机视觉实验室的链接。国内有中国科学院自动化研究所、清华大学、上海交通大学、国防科技大学和合肥工业大学等也都涉足图像场景分类领域的研究。很多 Top 期刊（如 IEEE Transaction on Pattern Analysis and Machine Intelligence（TPAMI）、International Journal of Computer Vision（IJCV）、IEEE Transaction on Image Processing（TIP）等）、顶级国际会议（IEEE Conference on Computer Vision and Pattern Recognition（CVPR）、IEEE International Conference on Computer Vision（ICCV）、European Conference of Computer Vision（ECCV）等）都经常报道场景图像分类方向的研究成果。

图像语义理解具有层次化结构[2]，包括低层、中层与高层，见图 1，分别对应图像处理层（低级视觉特征）、图像分析层（中间语义特征）和图像认知层（高级抽样语义）。从低层到高层之间形成了自底向上的数据驱动，从高层到低层之间形成了自顶向下的知识驱动[3]，其中，中层的存在是为了减小低层和高层之间的语义鸿沟[4]。图像场景语义理解必须建立低层视觉特征和高层场景语义之间的映射关系，它隶属于图像语义理解领域的范畴。

本文对近年来场景分类的发展进行了梳理和总结，介绍场景分类的主流方法。通常研究者更多关注的是分类模型和算法，然而事实上数据也是视觉识别研究中重要的因素之一。在数据充足的条件下，甚至利用最简单的模型算法也会得到很好的分类效果。文献[5]在文献[9]的 13 类场景数据集基础上加入 Industrial 和 Store 两类场景构成 15 类场景数据集（15-category），共包含 4 485 幅图像。文献[6]为室内场景分类识别提供了一个多样化的包含 67 个室内场景类的数据集（67-category），共包含 15 620 幅图像。文献[7]构建了一个覆盖场景、位置、人物变化较大的数据库（SUN397），包含室外自然场景、室外人造场景和室内场景共 397 个类别，总计 108 754 幅图像。Places 数据集[8]包含 700 万幅标记的图像，共 476 个类别，是目前场景和地点数目最大的图像数据集，可以为大数据计算提供足够的训练数据。文献[10]构造了层次化、多样化的大尺度数据集 ImageNet，是目前应用较多的数据集，可以应用于目标识别、图像分类和自动目标聚类等。通过观察实验发现，场景类别越多，类间差异越小，场景分类任务就会越加困难。

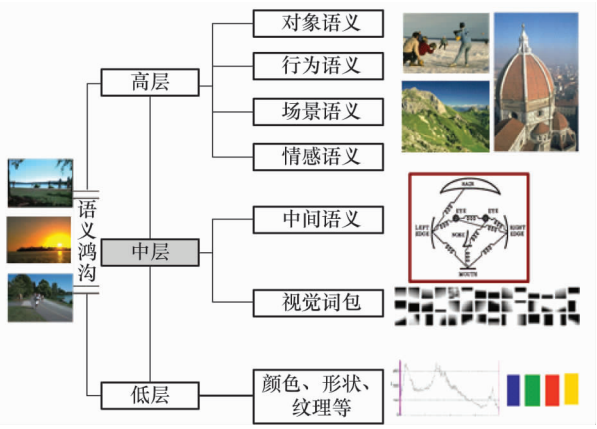


图 1　图像理解层次化表示

场景分类研究在许多方面都有着广泛的应用，比如应用于基于内容的图像检索（content-based image retrieval，CBIR），从数据库中检索图像，理解场景内容，有效地组建和扩展图像数据库[11]；应用于视频检索，把视频看作一系列图像的集合，按关键帧检索[12]；应用于数字照相洗印加工，对数字化的图片通过颜色平衡来校正偏色，增强图像对比度[13]；帮助机器人快速识别所在场景，使机器人针对不同场景做出自动快速的响应[14]；除此之外，场景分类还可以应用于医学图像浏览、卫星遥感图像分析、天气场景分类等方面。场景分类研究对于图像语义分类、反向图片搜索、自主机器人导航等，都具有重要的理论价值和广泛的应用前景。

场景语义分类作为模式识别、机器视觉与认知科学等学科交叉研究，已经发展成为一个重要的研究热点[15]。由于图像自身内容上的变化和差异，尤其是室内场景图像，会造成图像场景类内对象的差异性（比如凳子有靠背和无靠背）和图像场景类间的视觉相似性（比如阅览室场景和书店场景），图像场景语义理解与分类颇具挑战性。

本文首先介绍了场景语义分类的两大主要关键问题（图像内容描述和分类判决）以及场景分类的主要步骤；然后依据图像场景语义理解的层次化结构，对现有的场景语义分类算法进行了较为详细的描述；最后，总结了各种场景分类方法的性能比较，指出场景语义分类研究中存在的问题，并展望场景语义分类的前景。

1　场景语义分类

图像场景语义分类主要通过提取场景图像视觉特征，

进行特征映射,完成图像内容描述,设计分类器,最终完成图像场景分类识别,属于整体场景语义理解的范畴。文献[16]总结了一些常用的场景分类方法,对这些方法的优缺点进行了比较。场景分类主要包括两大关键问题:图像内容描述和分类判决。图像内容描述力图得到关于场景图像最具判别意义的表示,而分类判决则通过对训练样本集的图像描述学习训练,建模得到某场景类区别于其他场景类别的计算模型。图像内容描述包括:特征提取、视觉词典生成、图像特征映射、中间语义主题表示等。分类判决则包括分类器设计和分类。场景语义分类的主要步骤如图 2 所示。

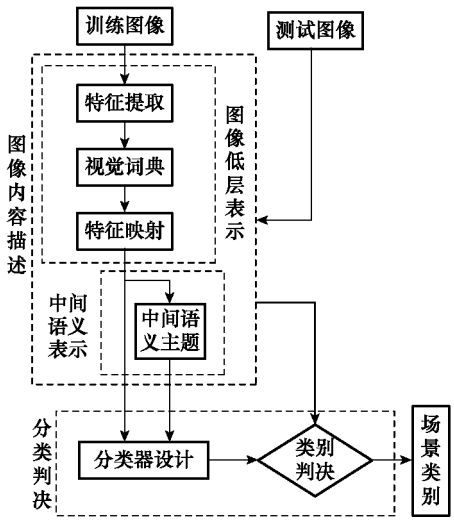


图 2 场景分类流程图

1.1 图像内容描述

1.1.1 特征提取

图像的视觉特征是对图像的特性描述。场景分类的第一步就是要提取图像显著视觉特征,包括局部特征和全局特征。文献[17]表明在缺乏明显局部信息的前提下,人类仍可利用空间结构信息完成对图像的场景认知。文献[18]定义了一组低维视觉感知词汇:自然度、开放度、伸展度、深度、粗糙度、复杂度、崎岖度和对称度,利用空间包络模型表示场景图像的全局空间结构信息。文献[19]提出了一种能够识别拓扑空间的全局视觉描述符 CENTRIST (CENsus TRansform hISTogram),它能描述场景识别任务(特别是室内图像场景)不同于其他视觉识别领域(如目标识别)中图像具有的属性,主要编码图像的整体结构信息,抑制细节结构信息,对图像的场景类别具有很强的概括性。文献[20]基于颜色、纹理、形状和小波提出一种用于目标和场景分类的图像描述符。首先提出 3 维局部二元模式(three dimensional local binary patterns,3D-LBP)描述符编码一幅彩色图像的颜色和纹理信息;然后整合 3D 的 LBP 颜色图和原始图像的方向直方图的小波变换,构造编码颜色、纹理、形状和局部信息的 H-descriptor;最后在几种颜色空间融合 H-descriptor 的主成分分析特征,形成 H-fusion 描述符。文献[21]基于监督共同信息提出重要度排序算法进行特征选择,获得的特征得到较高的识别率和较低的计算代价。利用全局特征进行场景分类,取得了较好的分类效果,但全局特征对于实际成像条件很敏感,鲁棒性和泛化能力差。

然而,局部特征却能有效抵抗各种仿射变换,具有某种不变性。文献[22]提出了一种具有里程碑式的局部特征描述子尺度不变特征变换(scale invariant feature transform,SIFT),具有良好的尺度不变性和旋转不变性。文献[23]在描述子 SIFT 的启发下提出了加速鲁棒特征描述子(speeded up robust features,SURF),构建积分图像,定位关键点,生成局部描述子。虽然 SURF 在尺度缩放和旋转不变性不及 SIFT,但是在模糊、光照变化等方面具有优越性,并且速度几倍快于 SIFT。文献[9]分别采用了 SIFT 特征和灰度像素矢量特征对图像进行描述,用于场景分类;文献[24]利用信息熵检测场景类图像的平坦度,据此对 SIFT 特征和灰度像素矢量特征加权形成融合图像描述;文献[25]获取图像块多方向的结构特征,形成对光照改变、几何扭曲、局部对比差异均具鲁棒性的局部描述符方向纹理曲线(oriented texture curves, OTC),但是局部特征明显缺乏空间信息。综上,全局特征和局部特征各有优缺点。在图像描述中,视觉特征提取应尽量做到既保持特征的不变性,又能融合特征的空间结构信息,实现二者的互补[26-28]。

1.1.2 视觉词典

视觉词典又称为视觉码本,把特征数据映射到各个码字上生成具有码本长度的特征向量[29]。构建视觉词典本质上是聚类问题,视觉码字对应聚类中心。所谓聚类,即把数据聚为若干簇,尽量使得簇内数据保持高相似性,而簇间数据保持低相似性,甚至无相似性。在图像场景分类任务中,视觉词典在图像视觉特征和场景语义之间承上启下。

视觉词典的设计是否有效,主要包括 3 个方面:分辨力、紧凑性和通用性。视觉词典的分辨力体现在视觉单词之间的相似度,相似度越低,表明分辨力越大;紧凑性体现在码本长度的选择,不同的码本长度对应不同的分类准确率,选择合适的视觉词典才能获得高识别率;通用性主要指如果加入了新的类别数据是否需要重新学习视觉词典。要做到这三方面的统一相当困难。文献[30]设计了通用视觉词典和类别视觉词典来竞争描述图像内容,其中通用视觉词典用来描述全部图像场景类,某一场景类图像依据通用视觉词典自适应学习可以获得该类的类别视觉词典。如果一幅图像属于某给定类,则类别视觉词典比通用视觉词典更适合描述该图像;反之,则通用视觉词典比类别视觉词典更适合描述该图像。文献[31]通过无监督学习获得所有场景公共字典的稀疏编码,再分解成一系列目标函数独立的多目标优化问题,有监督地学习各个类别字典。但传统的视觉词典易于出现缺乏明确含义或一词多义的情况。文献[32]为解决上述问题利用语义属性明确语义含义,并将语义属性融入视觉词包模型中,从而去除视觉单词的歧义。

1.1.3 特征映射

特征映射指依据视觉单词对视觉特征进行量化编码,生成视觉特征在视觉词典中的描述。特征映射主要有矢量

量化(vector quantization，VQ)、稀疏编码(sparse coding，SC)和局部约束线性编码(locality-constrained linear coding，LLC) 3 种方式。

VQ 方式的特征映射过程为遍历计算每一个数据特征向量与码字之间的欧式距离，选择距离最小的码字来表征该数据特征向量。虽然 VQ 方式简单方便，但其约束条件过于严苛，致使特征量化信息缺失。并且该方式忽略了码字之间的相互联系，最终造成特征映射编码结果比较粗糙。为了克服这种缺点，文献[33]采用了稀疏正则化方法，定义为矢量的 L1 范数，令矢量含有少量的非零元素，放宽了 VQ 方式中的约束条件。于是，视觉特征的编码过程就转化成为稀疏编码问题[33-34]。

SC 方式利用稀疏正则化方法来降低量化误差，以提高特征编码的独特性。文献[35]提出稀疏 Fisher 表示，在增加极少存储代价和计算代价的同时增加视觉码本的数目，从而提高分类正确率。针对编码稀疏性和编码局部性的关系，文献[36]在 SC 方式的基础上，提出局部坐标编码 (local coordinate coding，LCC) 方式。众所周知，特征编码的稀疏性不一定能保证其局部性，而特征编码的局部性一定能保证其稀疏性，所以 LCC 方式优于 SC 方式[36]。文献[37]以局部约束条件替换稀疏约束条件，得到 LLC 方式。

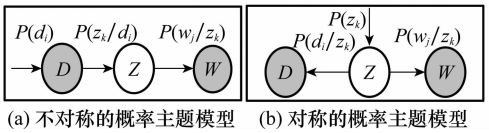
比较上述特征映射方式，LLC 方式计算效率高，并且具有更好的重构效果和局部平滑稀疏性。文献[38]对不同特征编码方式进行分类，总结当前算法的主要特点，在各个广泛使用的数据集上进行实验。文献[39]将高基数多类分类问题转化为按位解码问题，提出稀疏输出编码，有效地完成多类数据的分类问题。

1.1.4　主题模型

概率潜在语义分析(probabilistic latent semantic analysis，pLSA)模型[40]和潜在狄利克雷分配(latent dirichlet allocation，LDA)模型[41]是两种最具代表性的隐含语义主题模型。它们最初都是针对文本数据集的主题信息进行建模的概率生成模型，鉴于图像与文本有着很大的相似之处，研究者将 pLSA 和 LDA 引入到图像分类中，把高维的图像数据映射到低维的隐含语义层面是潜在主题模型的基本思路。

pLSA 模型如图 3 所示，主要通过参数估计描述实测数据的隐含变量，并计算其与实测变量的关系。图中 d 表示图像，z 表示隐含主题，w 表示视觉词汇。通过概率潜在语义空间的逐层推理，挖掘"d-z-w"之间的隐含概率关系。

pLSA 模型存在自身缺陷，其模型参量随着训练数据量的增加而线性增长，极易造成过拟合现象。它主要通过参数估计寻找训练数据样本的共现概率，不能给新数据分配概率。LDA 模型解决了 pLSA 模型的缺点。LDA 模型假设图像数据包含若干个线性的潜在主题，这样单幅图像表示为各个潜在主题的概率分布，而每个潜在主题表示为各个码字的概率分布。LDA 概率模型图如图 4 所示，其中 M 表示训练集中图像的个数，N 表示图像中视觉词汇的个数，α、β 和 θ 表示概率模型分布的参数。



(a) 不对称的概率主题模型　(b) 对称的概率主题模型

图 3　概率潜在语义分析模型

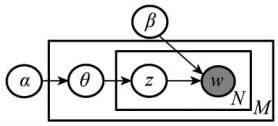


图 4　隐含狄利克雷分配图模型

1.2　分类

数据分类是数据挖掘和机器学习领域中一种重要的数据分析方法。进行数据分类时，通常把数据集按照一定比例随机地分成训练数据和测试数据，设计数据分类器对训练数据进行分析建模；然后把测试数据输入到分类模型中，输出数据分类结果。

数据分类模型分为生成式概率模型和判别式概率模型，两类各有侧重。生成式概率模型侧重于对数据类类内的概率分布建模，学习各类数据自身蕴含的概率分布。生成式概率模型结构清晰，易解释，比如高斯混合模型、Bayes 模型和隐马尔可夫模型等。判别式概率模型侧重于对数据类类间的概率分布建模，学习各类数据的分类边界，但缺点是无法有效利用数据类内的概率分布信息，比如支持向量机、逻辑回归、神经网络等分类模型。

生成式概率模型估计学习数据的联合概率分布，而判决式概率模型估计学习数据的条件概率分布，两类数据分类模型各有侧重，互有特点。当训练数据样本数量较大时，判别式概率模型性能好于生成式模型，但建模时间长；反之，生成式概率模型性能优于判别式模型，且建模时间短[42]。文献[43]表明由于这两种模型性质互补，混合两种模型的方法分类性能好于单一模型。

2　算法分类

如前所述，图像语义理解包含低层、中层和高层的层次化结构，对于场景语义分类来说，场景语义对应着场景语义理解层次化中的高层。据此，场景语义分类算法可以分为基于低层视觉特征和中层语义建模的两大类算法，分别体现出场景语义内容在图像理解低层和中层上的描述[17]。

2.1　基于低层视觉特征的场景分类

文献[44]指出，场景图像类别可以由低层视觉特征属性直接来描述，比如雪山场景布满了皑皑白雪，高楼场景具有垂直边缘信息，森林场景充满密集的纹理等。基于低层视觉特征的场景分类算法提取图像的低层视觉特征(比如颜色、形状和纹理等)，之后进行特征描述并设计分类器，对图像的场景语义信息进行推理。

该类方法代表性学者主要有 Vailaya 和 Szummer 等。Vailaya 等提取图像的空间颜色矩和边缘纹理方向一致性两个全局特征，训练两类贝叶斯分类器和支持向量机分类器来

完成场景分类[44]。Szummer 等将图像分割为若干子域,对各子块抽取特征点,采用大数投票思想对场景图像进行分类[45]。但是,图像内容比较复杂,为了弥补单一低级视觉特征的不足,Shen 等融合多种视觉特征进行场景分类,取得了优于单一特征的分类效果[46]。整体上来说,这种基于低层局部或全局特征及其特征融合的方法分类效果往往较差。最核心的问题是场景语义分类需要从低层特征向上推理能够得到图像的中间语义表示,而该类方法恰恰缺乏这种语义表示,对于训练图像样本集以外的图像数据无能为力,泛化性差。

2.2 基于中层语义建模的场景分类

一幅图像可以表示为一些特定目标的组合,比如天空、草地等,这些对象目标可以看作连接低层视觉特征和高层场景语义之间鸿沟的桥梁,相当于中间语义特征的角色。早期语义属性特征通常描述的都是图像的整体属性,从图像的全局结构信息和空间布局统计入手,定义一些感官属性概念来解释图像的场景语义。根据语义概念形式的不同,该类算法可以分为语义属性、语义对象和局部语义概念 3 种方法[16]。

2.2.1 语义属性

通过语义属性获取图像中层语义的代表性算法有文献[18]提出的 Gist 特征。Oliva 等定义了一组低维的视觉感知词汇,并利用空间包络模型来描述场景图像的空间结构信息[18]。文献[47]提出生物启发面向任务 Gist 模型(biologically inspired task oriented Gist model, BT-Gist),仿真生物 Gist 以整体场景为中心的空间布局属性和面向任务的分辨率测定属性。文献[48]利用视觉和语义信息构造语义含义图像层,估计整个层次概念的分布来表示图像。文献[49]计算两幅图像的语义相似性,用待测图像和 ImageNet[10] 数据库中最近邻的图像标签来表示待测图像。

文献[19]提出视觉描述符 CENTRIST (CENsus TRansform hISTogram),识别拓扑空间进行场景分类,主要编码图像的结构属性,抑制细节结构信息。为克服 CENTRIST 某些情况下无法正确获取局部图像结构的问题,文献[50]在 CENTRIST 基础上将局部结构分布和上下文信息组合,提出上下文平均统计变换(contextual mean census transform, CMCT),能够有效区分外观相似近邻不同的图像块。随后将全局 Gist 和描述局部形状的文献[50]组合生成 GistCMCT[51],对修正统计变换(modified census transform,MCT)采用 PCA 降维,建模 3 层空间信息生成 spatial MCT,结合两种特征的分类器输出获取最后类别,在不增加特征矢量维数的情况下,提高分类正确率[52]。统计变换只对灰度图像进行处理,因此 CENTRIST 特征忽略了颜色信息,而颜色信息在图像场景识别中起着重要作用。文献[53]将颜色信息融入 CENTRIST 框架提出颜色 CENTRIST,用亮度值的梯度和局部图像块像素间颜色变化来描述全局形状信息。多通道信息具有互补性,现有多通道描述符大多直接串联每个通道的矢量,实际上通道间信息并不独立。文献[54]提出 mCENTRIST,明确从多通道图像获取共同信息进行场景分类,将传统 RGB 色彩空间转换为对立色彩空间,并加入 Sobel 梯度图得到超对立色彩空间,来获取更多通道,从而得到更多互补信息提高分类正确率。上述基于语义属性的方法描述粗略,泛化能力弱,一旦图像内容复杂,则通常会得到较差的场景分类效果。

2.2.2 语义对象

依据人类视觉感知经验可知,包含相似目标的图像有相似的场景标签,在定义一个场景类别时,目标的重要程度不同。比如,一幅图书馆的图像一定含有书和书架,但不一定含有桌子等。这些先验知识为场景图像描述提供了思路。

用语义属性矢量描述图像中的目标进行图像分类,对每个语义属性训练分类器。文献[55]最先提出利用视觉语义属性的集合来识别熟悉的目标,提供边框注释来描述新图像中不熟悉的目标。文献[56]提出对其他类别训练语义属性分类器,即使待识别的目标类没有训练图像,也可以用语义属性的高层描述符来识别。文献[57]通过编码图像和区域层的视觉结构获取压缩的视觉属性,在保证分类正确率的前提下大幅度降低特征的维数。文献[58]提出基于超图的建模方法,考虑场景语义属性的高阶交互进行场景分类。文献[59]先将图像的场景分类问题转化为目标识别问题,用大量目标分类器的输出表示图像,应用于图像分类和图像检索任务。文献[60]从弱标记的视觉图像获取视觉概念,这种 ConceptLearner 用来识别图像层的概念,探测图像区域层精确的概念,然后用于特定的图像表示。

文献[6]定义图像原型,实现图像和场景标签之间的映射,不直接利用目标,但认为某些目标在定义场景类别时更为重要。文献[15]在文献[6]的基础上提出结合全局和局部信息的混合图像表述,分别获取基于内容的信息和基于目标的信息。文献[61]提出基于可变形部分模型(deformable part-based models,DPM),用固定根节点和可移动的感兴趣区域描述图像场景,对于具有稳定的全局结构场景和具有突出目标的场景图像来说,都有较好的分类效果。文献[62]利用潜在支持向量机框架,学习区域探测器集,获取场景关键特性,进行场景分类;将每个区域称为潜在金字塔区域(latent pyramidal region, LPR),用非线性约束编码空间金字塔进行表示。文献[63]用潜在支持向量机训练 DPM 模型,对大量含有对象位置注释的数据库学习部分滤波器;利用语义层次自动获取合并多个对象相似的部分,生成混合滤波器。这种表示含有丰富细化的信息,更具判别力。这些方法进一步用语义对象描述图像中的目标,为场景识别打下基础。

文献[64]提出一种高层图像特征表示目标库(object bank, OB),编码目标外观和图像的空间位置信息,获取未知图像的语义信息。从常用数据库如 ESP[65]、LabelMe[66]、ImageNet 等,选中目标探测器,在高层图像识别任务中进行有效的计算。文献[67]用 OB 提供的特征,考虑数据结构,对一些基准库进行场景分类,在提高分类正确率的同时,大大降低特征维数。文献[68]提出利用目标和场景的共生概率来建模场景的内容,在语义空间表示图像,每一维对应概念基于外观的后验概率,然而由于分类图像块时固

有的模糊性，这种表示会存在一些上下文噪声，通过建模语义空间中每个概念的分布来保证鲁棒性。

场景（尤其是室内场景）涉及复杂语义模式下许多对象的相互作用。文献[69]利用学习 kLog 核（kLog 是一种基于核学习的逻辑和关系语言）定义语义目标的高阶空间关系，自动探测语义对象及其空间关系，用逻辑表示或超图描述场景。文献[70]利用对象的高阶成分提出对象组的概念，这些对象组类似于场景特定的空间结构，可以为场景分类提供细节信息和其他对象的位置。文献[71]提出三层（像素层，对象层，场景层）生成层次模型进行场景理解，在训练过程实现自动明确对象注释，用概率链结构表示对象共现情况获取上下文信息，用简单整体策略解决局部优化问题。

上述算法的思路是将图像的场景分类问题先转化为目标的识别问题[72-73]，再利用目标或多个目标或包含多个目标的图像块来描述场景，但实际上目标识别问题仍是一个较难处理的任务。

2.2.3　局部语义概念

为了避免目标检测与识别过程，可以将图像规则划分，提取各子块的局部图像描述子，建立局部描述子和局部语义概念之间的对应，利用局部语义概念的概率分布完成图像场景语义分类。这种基于局部语义概念方式的算法可以分为两种：基于语义主题建模的场景分类算法和基于视觉词包模型的场景分类算法。

（1）基于语义主题建模的场景分类算法

基于码字对场景语义内容进行概率分布建模，采用概率分布模型来学习图像中的隐含语义主题，然后根据隐含语义主题的概率分布来进行场景语义分类。基于语义主题建模的场景分类，符合人类视觉的认知过程。

语义主题建模分为生成式概率模型和判别式概率模型。生成式概率模型按照场景类别在特征空间中的联合概率分布进行场景建模，代表性学者主要有 Bosch 和 Li 等。Bosch 等利用 pLSA 模型[40]挖掘视觉单词的隐含语义信息[74]。由于场景每类含有不同的空间层次，可以在局部语义内容的基础上对空间信息进行加权融合。文献[75]基于 pLSA 加权空间信息提出室内家居场景识别模型，考虑整幅图像的空间信息，以及在家居场景分类中不同的空间位置比例，融合各层每个网格模型和分类器的分类结果。将概率主题模型应用于场景分类能够获取自然图像的潜在语义成分，在训练过程中不需要监督，人类视觉和心理学研究者设计高层监督主题模型进行场景理解。Li 提出了两种 LDA 模型[41]的变形模型，将图像局部区域建模为不同的语义主题[9]。文献[76]提出主题监督 LDA（topic supervised LDA，ts-LDA），用指定主题代替自动获取主题的方式，解决 LDA 监督扩充用通用图像规律而不是感兴趣的语义规律来分类图像的问题，提出特定类 LDA（class specific simplex LDA，css-LDA）为每类产生一个主题，用最具判别的形式，解决主题能力受限的问题。文献[77]基于特征分类设计出分类码本，通过调整分类码本来控制各场景类别特征的贡献。文献[28]通过上下景特征设计类别码本，训练改进的生成模型，完成场景分类任务。融合上下景特征也越来越多地被用来完成场景语义分类任务[26-27,78]。文献[7]采用 4 种几何上下文构造特定几何直方图，再将全部直方图级联进行图像表述。文献[31]利用语义属性的预测，在决策层将词包直方图和语义描述符进行组合，并且采用语义信息融入视觉词典的方法，大幅提高词包模型应用于场景分类的正确率。为了增加空间结构信息，文献[5]提出空间金字塔匹配核思想，设定空间金字塔层数，进行图像子域分割并提取子块特征，进行特征映射量化编码后构建空间金字塔加权矢量形成图像描述。文献[33,37，79－80]为了融入全局空间信息也都采用了空间金字塔匹配思想进行场景分类。

判别式概率模型根据场景类别在特征空间中的条件概率分布进行建模，其任务核心是设计核函数。文献[81]证明对于直方图形式的数据，基于直方图交叉核（histogram intersection kernel，HIK）的支持向量机比径向基函数核更有效。

生成式概率模型和判别式概率模型各有所长，特点互补。计算量和模型复杂度之间的矛盾是生成式概率模型最大的问题，这种矛盾对于判别式概率模型则不是问题。判别式概率模型对于不同场景类建模时没有考虑各场景类之间的联系，属于独立建模。文献[82]将两种概率模型结合起来完成场景语义分类任务，场景识别效果好于单一概率模型。总之，基于主题模型的研究还存在很多问题，需要进一步考虑不同区域间的内容及空间关系，还有很大的研究改进空间。

（2）基于视觉词包模型的场景分类

预先定义视觉码本，由视觉码字出现的概率分布来描述图像内容，然后根据码字的概率分布进行场景语义分类。其中图像特征的提取、视觉词典的学习、特征的映射方式、是否增加空间上下文信息等，都会对分类结果产生影响。

考虑不同视角下多特征场景在低维空间的构造，获取图像特征组，利用特征的互补特性，学习归一化低维子空间，有效地融合特征，解决多视角维数降低问题。文献[83]在多视图特征成对约束中加入用户标注，最初由用户提供待测图像，通过搜索引擎检索图像集，由用户标记检索到的图像与待测图像的相关性，用交替优化处理整合不同视角信息和用户标记信息，为场景分类提供多视角维数降低方法。文献[84]提出稠密采样局部描述符，用核密度估计获取描述符的概率密度函数；对概率密度函数梯度进行方向编码，生成方向码本；用方向码本聚合生成概率密度函数方向梯度直方图；将 BoF（bag-of-feature）离散表示变为概率密度函数的连续表示。文献[85]构造潜变量重构词包模型（reconfigurable BoW，RBoW）用于场景分类，在重构模式下将一个场景看作是一些区域模型的集合。Redi 等提出描述符聚合边缘估计（marginal estimation for descriptor aggregation，MEDA）[86]，Multi-MEDA[87]，C-BoW（Copula BoW）[88]和关联边缘签名（copula and marginal signature，COMS）[89]进行场景分类，这 4 种方法均完成局部图像描述符的聚合。MEDA 不考虑局部图像描述符间的关系，Multi-MEDA 利用

平移不变核，串联局部图像描述符，从 MEDA 矢量获取边缘近似，建模多元模型。C-BoW 融入 Copula 理论，基于局部特征边缘分布的相关性，为矢量量化构造一种二次高效的码本。而 COMS 是基于 Copula 的 MEDA 扩充，利用 Copula，允许图像局部描述符的有效多元分析。上述方法均应用于场景识别和图像检索等任务，与语义特征结合，实现很好的效果。文献[90]还提出全局图像描述符 Saliency Moments，在场景全局表示中嵌入局部解析信息，如场景显著区域或对象。

字典学习是机器学习一个重要任务，现有字典学习包括生成（无监督）和判别（监督）方法。文献[91－92]采用多示例学习进行字典学习，把监督问题转化为弱监督问题进行处理。文献[91]对每一幅图像获取显著块和非显著块作为示例，为每个类训练单概念分类器和多聚类视觉概念，通过空间金字塔图像表述进行场景分类。而文献[92]通过最大化边缘多示例字典学习（max-margin multiple instance dictionary learning，MMDL）明确最大化不同聚类中心，生成学习分类器，利用已学习分类器（G-codes）进行图像分类。文献[93]利用层次类的相关性进行字典的学习，为层次类结构中的每个节点学习类模型集合的字典，不同层次的字典能够获取不同尺度的视觉属性。文献[94]提出两步构造语义词典的方法，首先用传统词包模型方法对局部特征进行矢量量化获取视觉单词，然后将中层特征融入低维语义空间，再通过 k 均值聚类生成语义词典。文献[95]和文献[96]分别对训练样本集中各个区域的视觉特征进行聚类，生成视觉词典，然后依据视觉单词按照矢量量化编码方式，生成图像区域在视觉词典中的编码描述。语义词典比传统的词典更具识别力，可以将语义信息融入词典中。

字典学习之后，需要将图像局部描述子与字典进行映射，对不同映射方法的改进，可以提高图像表述力，进而提高场景分类正确率。然而稀疏编码会把相似局部特征量化到同一个视觉单词中，空间稀疏编码（sparse spatial coding，SSC）将稀疏编码字典学习、空间约束编码和在线分类方法组合应用于分类[97]。文献[98]提出线性距离编码，可以获得比传统编码更具判别性的信息，同时缓解对图像空间结构合并的依赖。

传统特征词包表示具有无序性，图像特征描述时需要融入特征空间分配来获取场景中固有的高阶结构。文献[99]提出局部成对码本（local pairwise codebook，LPC）方法，在联合特征空间将空间位置上接近的描述符看作一个数据点，再进行码本的构造。文献[100]将码本结构看作特征空间，基于最近邻度量函数进行室内场景识别。文献[101]用方向金字塔匹配（orientational pyramid matching，OPM）代替空间金字塔匹配，正是由于区分室内场景时目标的 3 维方向因素非常重要，故采用 3 维方向信息形成金字塔产生合并区域，与空间金字塔匹配进行互补能得到非常好的分类效果。文献[102]提出语义流形空间金字塔匹配（spatial pyramid matching semantic manifold，SPMSM），用于场景识别。SPMSM 基于语义概率单纯形图像表示，使用粗糙空间信息编码，建立语义单纯形和黎曼流形的联系，进行语义空间流形结构的相似性测量。利用目标作为特征进行场景描述，能有效地提高场景分类正确率。但是现有图像分割通常将图像直接分割为规则的矩形，会导致同一目标被分割到不同的图像子块，或将不同目标过多地分到同一图像块中，导致识别率低下。文献[103]采用基于超像素网格的分割方法，使同一图像子块中的对象具有相对完整性，同时子块图像特征具有一致性，将上下景信息融入到空间金字塔描述，场景内容描述既有局部梯度信息、局部结构信息，又有全局空间信息。文献[104]在传统空间划分方式上加入自底向上的显著驱动获取视觉结构，对图像获取的显著图划分出显著性和非显著性区域（各占图像 50％且不重叠），获取图像的相对语义，是对传统空间描述的补充。

获取图像局部特征时，可以稠密划分网格，也可以提取图像关键点处特征，增加多视角特征等方式，以期提高词包表示的性能。由无监督聚类算法生成的视觉词典，必然会存在同义问题和多义问题，从而无法准确表示视觉单词与相应局部语义之间的对应关系。还有，视觉词典容量过大或过小，都会导致分类性能下降，从而影响场景分类性能。编码中特征量化不可避免地会丢失信息，合并同时会依赖于图像的空间层次，因此不同映射方式的改进，在传统词包表示中增加空间结构等方式均会对场景分类性能产生影响。

2.3 深度模型

获取图像局部语义、全局语义或语义目标的方法通常采用无监督学习的方式获取图像特征，而深度学习通常采用监督学习方式获取图像特征，通过卷积滤波器层次反向传播的训练获取大尺度识别任务数据集的深度模型，进行图像中层表示，在各种识别任务中达到目前最好的效果。早在 1989 年 Lecun 就利用卷积神经网络（convolutional neural networks，CNN）进行手写字符识别，取得较好成果[105]。2012 年，文献[106]将 CNN 应用于 ImageNet 图像分类，随后文献[107]将 CNN 应用于目标识别，取得较好成果。文献[108]提出迁移学习策略，在某个数据集上用监督方式预训练 CNN 特征，实现在其他数据集上的各种分类识别任务。一些现有的 CNN 特征获取工具包括 OverFeat[109]，DeCAF[110]，Caffe[111]等。文献[112]提出简单有效的框架多尺度无序合并（MOP-CNN），获取多尺度局部块的 CNN，在每个尺度层采用有序向量局部描述符聚合，串联得到结果描述符。CNN 可以获取丰富的中层表示，但是它的几何不变性比较差，文献[113]用多尺度金字塔合并，来提高 CNN 激活对几何不变鲁棒性的判别力。文献[114－116]通过实验验证 CNN 不同方面的特性，通过迁移学习的策略完成各个数据集上的分类和识别问题。

文献[8]指出由于 ImageNet 数据集是目标类的数据集，在 ImageNet 上训练的特征，应用于场景类数据集上效果不如应用于目标类数据集的效果好。他们构建的包含 700 万幅图像，60 倍于 SUN 数据集的 Places 数据集用于场景类图像的训练和识别，是目前数目最大，种类最齐全的场景类。实验表明，对 Places 训练的 CNN 特征应用于场景类的识别

效果很好，而对 ImageNet 训练的 CNN 特征应用于目标识别效果很好。由于图像由目标组成，对 Places 训练的 CNN 进行场景识别的同时可以进行目标的定位[117]。与定位目标的区域 CNN(R-CNN)[107]类似，对复杂情况下的纹理识别也可以采用 CNN。文献[118]对 CNN 滤波器组进行 Fisher 矢量合并，得到 D-CNN 描述符，可以应用于目标和场景识别、纹理识别和分割。CNN 的缺陷是需要大量的训练数据和精细的参数优化，但如本文表 1和表 2 所示，CNN 的分类识别率很高。

表 1　各种算法在 15-category 和 67-category 数据集上的分类性能比较

文献	15-category	67-category	文献	15-category	67-category
Quattoni[6]	—	～26.50	Nakayama[128]	86.10	45.50
Gu[103]	87.13%	—	Juneja[121]	—	46.10
Han[47]	80.20	25.70	Li[91]	83.40	46.40
Redi[90]	—	28.00	Zhou[129]	84.20	46.50
Zhu[130]	—	～28.00	Cakir[100]	82.24	47.01
Pandey[61]	—	30.40	Wu[124]	—	47.15
Wang[131]	80.43	33.70	Bo[132]	—	47.60
Chu[53]	—	36.09	Bu[133]	89.38	48.3
Wu[19]	83.88	36.88	Garg[35]	89.60	49.87
Parizi[85]	78.60	37.93	Wang[92]	86.35	50.15
Singh[119]	—	38.10	Fornoni[104]	84.39	50.54
Morioka[99]	83.40	39.63	Sun[120]	86.00	51.40
Zhu[134]	81.80	39.70	Han[135]	89.13	57.72
Zheng[63]	84.70	39.80	Kobayashi[84]	85.63	58.91
Niu[15]	—	40.19	Doersch[122]	—	64.03
Gazolli[52]	86.30	42.42	Azizpour[115]	—	66.3
Li[64]	—	42.90	Gong[112]	—	68.88
Zhou[136]	85.20	42.90	Razavian[108]	—	69.0
Mesnil[67]	86.44	44.00	Koskela[116]	92.1	70.1
Kwitt[102]	82.30	44.00	Zhou[8]	91.59	70.8
Oliveira[97]	—	44.35	Cimpoi[118]	—	80.0
Wu[54]	—	44.60	Yoo[113]	—	80.78
Sadeghi[62]	85.81	44.84			

注：表中符号“～”表示“约等于”。

表 2　各种算法在 SUN397 数据集上的分类性能比较

文献	SUN 397	文献	SUN 397
Xiao[7]	38.00	Zhang[21]	42.72
Kwitt[102]	24.3	Xie[101]	45.91
Margolin[25]	34.56	Azizpour[115]	49.9
Su[32]	35.60	Gong[112]	51.98
Agrawal[114]	40.40	Zhou[8]	53.86
Donahue[110]	40.94	Koskela[116]	54.70

通过弱监督学习，最小化判决块对场景分类也能取得很好的结果。文献[119]对大量图像块数据库用判决聚类获取判决块、交替聚类和训练判决块的迭代处理。文献[120]从有类别标签的图像集学习判决部分探测器，通过组稀疏归一化在最大边缘框架下联合选择和优化判决部分探测器进行建模，用梯度近似算法解决对应的优化问题。文献[121]利用增量学习部分的方式解决判决部分自动学习问题。文献[122]利用 mean-shift 算法和判决方式获取视觉一致的块。文献[123]提出判别共享特征学习(discriminative and shareable feature learning，DSFL)，得到灵活数目的共享滤波器来表示不同场景类间的通用模式，而与卷积神经网络训练的特征互补，分类效果非常好。

2.4　分类器的简单讨论

大多数学者将研究重点集中于如何有效地表示图像场景内容，分类器通常采用现有的有监督机器学习算法，如支持向量机分类器、k 近邻分类器、贝叶斯分类器、boosting 等。随着图像数据海量的增长，图像数据集数目的扩大，对分类算法的改进主要集中于寻求计算速度和计算精度之间的折中，尽可能地在保证低内存消耗和低计算代价的前提下，得到尽可能高的正确率。空间金字塔匹配核[5]、直方图相交核[81]都可以作为支持向量机分类器的核函数，用于提高场景分类的正确率。文献[124]提出幂平均支持向量机分类器(power mean SVM，PmSVM)，解决线性分类器精度较差，非线性分类器分类时间较长的问题。文献[125]提出牛顿-拉富生框架，在低计算代价和低内存需求下完成分类任务。文献[126]用树或有向无环图结构提出层次分类器，实现大尺度数据集的分类。

通常分类器的选取，对场景图像分类的结果会产生重要的影响，以后有待进一步研究新的分类机理与方法。

3　结　论

图像场景语义分类研究实验所用场景数据集主要为文献[5]提供的 15 类场景数据集(15-category)，文献[6]提供的 67 类室内场景数据集(67-category)和文献[7]提供的 397 类室内室外场景数据集(SUN397)。最近由文献[8]构建了目前场景和地点数目最多且图像最多样化的 Places 场景图像数据集。

表 1 为各种算法在 15 类场景数据集和 67 类室内场景数据集中分类识别率的比较。表 2 为各种场景分类算法在 SUN397 数据集中分类识别率的比较。由于 SUN397 数据集和 Places 数据集是近年来才提出的,且图像数目繁多,种类多样化,对这两个数据集进行实验的算法远不如对 15 类场景数据集和 67 类室内场景数据集的算法多。但随着设备存储计算能力的增强,大尺度数据集的出现和研究会成为主流趋势。观察表 1 和表 2 可知,无论是 15 类场景、67 类室内场景,还是 397 类场景,图像分类正确率都逐渐增加。早期 15 类场景平均分类正确率为 70%多,现在可以达到 90%以上[116],接近人类分类的效果。67 类室内场景由于场景类别数目相对增多,且室内场景相对室外场景更加复杂,最初分类正确率只有 26.5%左右[6],通过弱监督学习获取特征,平均分类正确率可以达到 60%多[122],随着深度特征的获取,甚至可以达到 80.78%的正确率[113]。文献[7]将 GIST、HOG、稠密 SIFT、LBP、自相似描述符(self-similarity descriptors, SSIM)等特征组合用于 397 类场景的描述,得到平均分类识别率为 38%。虽然有些学者对于传统方法进行改进,分类识别率得到了一定的改善,但是总的来说,分类识别率仍低于 46%。而利用 CNN 获取深度特征的方法,则具有明显的优势,文献[112]方法分类识别率为 51.98%,文献[116]方法分类识别率则高达 54.70%。综合表 1 和表 2 可见,CNN 方法的场景分类识别率明显优于传统方法,是未来研究的重点方向。

场景图像包括室外场景图像和室内场景图像,如 15 类场景数据集、SUN397 场景数据集、Places 数据集均包含室内图像和室外图像。事实上,在常见的场景数据集中应用各种的场景分类算法,大多数算法对于室外类场景图像识别率较高,对室内类场景图像测试时分类识别性能普遍较差。文献[127]将 15 类场景数据集划分为 10 类室外场景和 5 类室内场景两个子数据集分别进行分类实验。在相同的实验条件下,10 类室外场景数据集的平均分类正确率高达 94.20%,而 5 类室内场景数据集的平均分类正确率仅为 78.20%。由此可见,室外场景分类识别率要优于室内场景类。为了进一步证明该结论,针对 67 类室内场景数据集,选取文献[6]方法中场景分类识别率最高的前 15 类室内场景,按照文献[5]的方法进行场景分类实验,得到的场景分类识别率仅为 47.30%(而对于 15-category 场景数据集,Lazebnik[5]方法的分类识别率为 81.40%)。

通常来说,室外场景分类识别率优于室内场景类。室外场景类图像往往具有比较明显的全局空间结构信息,能由全局属性很好的描述图像内容;相反,只有少数室内类场景(如走廊)图像才具有全局空间特征,大多数室内类场景图像更适合通过局部的目标对象来表征(比如卧室场景中的床)。并且相比于室外场景,室内场景通常包含复杂的结构和种类繁多的人工制品,更容易造成较大的场景类内差异和类间相似;室内场景更容易受到光照变化、视角变化和尺度变化的影响;室内场景中更容易出现遮挡问题[6]。因此,对室内场景图像分类问题进行研究是非常有必要的。

在图像描述方面,本文将场景分类划分为基于低层特征和基于中层语义建模的场景分类算法。基于低层特征的方法很难识别训练集之外的图像数据,泛化性较差。基于中层语义建模的方法能够降低低层和高层图像处理之间的语义鸿沟,将场景模型与人类感知相匹配。然而场景图像的高层语义含义并不唯一,对图像内容复杂的场景要想获得理想的分类效果十分困难。利用语义目标的频率进行场景的描述能够处理内容复杂的场景识别,然而目标识别和探测本身就是一个较难的任务。为避免目标的分割和探测,可以获取关键点局部描述符的中间属性进行图像的语义描述。根据局部语义概念表示方式不同,进行视觉词包建模或者语义主题建模。视觉词包建模中视觉单词容易出现同义问题和多义问题,并且视觉码本的维数会对分类性能产生较大影响。基于主题模型的方法通常不考虑不同区域间内容的关联和空间关系,仍存在较大改进空间。近几年深度学习成为一种研究趋势,能大幅提高场景分类识别率,但仍需大量的训练样本和精细的参数优化,需要较长的运算时间和较大的内存代价。

目前大部分学者将场景分类研究重点集中在图像内容描述上,而对分类器进行改进融合以期提高室内场景分类性能的方法仍然较少。但是分类器的选取,对场景图像分类的结果会产生重要的影响,尤其是现在随着深度特征的出现,将深度特征与新的分类机理与方法组合,能够更好地识别场景,完成场景分类任务。

由于图像自身内容的复杂性,尤其是室内场景图像,场景分类任务十分困难,场景语义分类是图像理解、计算机视觉和认知科学领域中一个具有挑战性的课题。目前,场景分类的方法并不局限于单一的类型,可以将局部目标和全局结构组合以期提高分类正确率。室外场景分类研究相对比较成熟,而室内场景分类研究进展却相对缓慢,现在越来越多的学者致力于室内场景分类的研究。室内场景分类研究将是该领域下一步的重点研究内容。

参考文献:

[1] Liu S Y. Research on perception-oriented image scene and emotion categorization[D]. Beijing: Beijing Jiaotong University, 2011.(刘硕研. 面向感知的图像场景及情感分类算法研究[D]. 北京:北京交通大学, 2011.)

[2] Gao J, Xie Z. *Theroy and method of image understanding*[M]. Beijing:Science Press, 2009: 399-430.(高隽, 谢昭. 图像理解理论与方法[M]. 北京:科学出版社,2009: 399-430.)

[3] Datta R, Joshi D, Li J, et al. Image retrieval: ideas, influences, and trends of the new age[J]. *ACM Computing Surveys*, 2008, 40(2): 1-60.

[4] Gu G H. Image semantic representation based scene classification research[D]. Beijing: Beijing Jiaotong University, 2013.(顾广华. 面向图像语义描述的场景分类研究[D]. 北京:北京交通大学, 2013.)

[5] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern*

Recognition, 2006: 2169 - 2178.

[6] Quattoni A, Torralba A. Recognizing indoor scenes[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 413 - 420.

[7] Xiao J, Hays J, Ehinger K, et al. Sun database: large-scale scene recognition from abbey to zoo[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010: 3485 - 3492.

[8] Zhou B, Lapedriza A, Xiao J, et al. Learning deep features for scene recognition using places database[C]// *Proc. of the Advances in Neural Information Processing Systems*, 2014: 2 - 7.

[9] Li F F, Perona P. A Bayesian hierarchical model for learning natural scene categories[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005: 524 - 531.

[10] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 1 - 9.

[11] Wang J Z, Li J, Wiederhold G. SIMPLIcity: semantics-sensitive integrated matching for picture libraries[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001, 23(9): 947 - 963.

[12] Zhang H J, Wu J, Zhong D, et al. An integrated system for content-based video retrieval and browsing[J]. *Pattern Recognition*, 1997, 30(4): 643 - 658.

[13] Boutell M R, Brown C B, Luo J. Review of the state of the art in semantic scene classification[R]. Rochester, NY: University of Rochester, 2002.

[14] Caro L, Correa J, Espinace P, et al. Indoor mobile robotics at Grima, PUC[J]. *Journal of Intelligent & Robotic Systems*, 2012, 66 (1/2): 151 - 165.

[15] Niu Z B. Research on key techniques of image representation in recognition [D]. Shanghai: Shanghai Jiaotong University, 2011.(牛志彬. 图像识别中图像表达的关键技术研究[D]. 上海:上海交通大学, 2011.)

[16] Bosch A, Munoz X, Marti R. A review: which is the best way to organize/classify images by content? [J]. *Image and Vision Computing*, 2007, 25(6): 778 - 791.

[17] Tang Y J. Research on semantic topic model based image scene classification[D]. Beijing: Beijing Jiaotong University, 2010.(唐颖军. 基于语义主题模型的图像场景分类研究[D]. 北京交通大学, 2010.)

[18] Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. *International Journal of Computer Vision*, 2001, 42(3): 145 - 175.

[19] Wu J X, Rehg J M. CENTRIST: a visual descriptor for scene categorization[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1489 - 1501.

[20] Banerji S, Sinha A, Liu C. New image descriptors based on color, texture, shape, and wavelets for object and scene image classification[J]. *Neurocomputing*, 2013, 117: 173 - 185.

[21] Zhang Y, Wu J, Cai J F, Compact representation for image classification: to choose or to compress? [C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014:1 - 8.

[22] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91 - 110.

[23] Bay H, Tuytelaars T, Van Gool L. SURF: speeded up robust features[C]// *Proc. of the European Conference of Computer Vision*, 2006: 404 - 417.

[24] Gu G H, Zhao Y, Zhu Z F. Integrated image representation based natural scene classification[J]. *Expert Systems with Applications*, 2011, 38(9): 11273 - 11279.

[25] Margolin R, Zelnik L, Tal A. OTC: a novel local descriptor for scene classification[C]// *Proc. of the European Conference of Computer Vision*, 2014: 377 - 391.

[26] Qin J, Yung N H C. Feature fusion within local region using localized maximum-margin learning for scene categorization[J]. *Pattern Recognition*, 2012, 45(4): 1671 - 1683.

[27] Jiang Y, Chen J, Wang R S. Fusing local and global information for scene classification[J]. *Optical Engineering*, 2010, 49 (4): 047001 - 047010.

[28] Qin J, Yung N H C. Scene categorization via contextual visual words[J]. *Pattern Recognition*, 2010, 43(5): 1874 - 1888.

[29] Wu J X, Rehg J M. Beyond the Euclidean distance: creating effective visual codebooks using the histogram intersection kernel[C]// *Proc. of the IEEE International Conference on Computer Vision*, 2009: 630 - 637.

[30] Perronnin F. Universal and adapted vocabularies for generic visual categorization[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008, 30(7): 1243 - 1256.

[31] Duan F, Zhang Y J. Scene categorization via supervised subspace modeling and sparse representation[J]. *Journal of Image and Graphics*, 2012,17(11): 1409 - 1417.(段菲, 章毓晋. 有监督子空间建模和稀疏表示的场景分类[J]. 中国图象图形学报, 2012,17(11): 1409 - 1417.)

[32] Su Y, Jurie F. Improving image classification using semantic attributes [J]. *International Journal of Computer Vision*, 2012, 100(1): 59 - 77.

[33] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 1794 - 1801.

[34] Lee H, Battle A, Raina R. Efficient sparse coding algorithms[C]// *Proc. of the Advances in Neural Information Processing Systems*, 2007: 801 - 808.

[35] Garg V, Chandra S, Jawahar C. Sparse discriminative Fisher vectors in visual classification[C]// *Proc. of the 8th Indian Conference on Computer Vision, Graphics and Image Processing*, 2012: 1 - 8.

[36] Yu K, Zhang T. Improved local coordinate coding using local tangents[C]// *Proc. of the 27th International Conference on Machine Learning*, 2010: 1215 - 1222.

[37] Wang J, Yang J, Yu K. Locality-constrained linear coding for image classification [C] // *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010: 3360 - 3367.

[38] Huang Y, Wu Z, Wang L, et al. Feature coding in image classification: a comprehensive study[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2014, 36(3): 493 - 506.

[39] Zhao B, Xing E P. Sparse output coding for large-scale visual

recognition[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 3350 - 3357.

[40] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. *Machine Learning*, 2001, 42(1): 177 - 196.

[41] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. *Journal of Machine Learning Research*, 2003, 3(4/5): 993 - 1022.

[42] Zeng P. Research on key techniques of image classification for semantic extraction[D]. Changsha: National University of Defense Technology, 2009. (曾璞. 面向语义提取的图像分类关键技术研究[D]. 长沙:国防科技大学, 2009.)

[43] Liu K, Zhao J. Domain adaptation in NLP based on hybrid generative and discriminative model[C]// *Proc. of the IEEE Chinese Conference on Pattern Recognition*, 2008: 7 - 12. (刘康, 赵军. 基于"产生/判别"混合模型的分类器领域适应性问题研究[C]// 2008 年全国模式识别学术会议, 2008: 7 - 12.)

[44] Vailaya A, Figueiredo M, Jain A, et al. Image classification for content-based indexing[J]. *IEEE Trans. on Image Processing*, 2001, 10(1): 117 - 130.

[45] Szummer M, Picard R W. Indoor-outdoor image classification[C]// *Proc. of the IEEE Workshop on Content-based Access of Image and Video Databases*, 1998: 42 - 51.

[46] Shen J, Sheperd J, Ngu A. Semantic-sensitive classification for large image libraries[C]// *Proc. of the IEEE Conference on Multimedia Modeling*, 2005: 340 - 345.

[47] Han Y N, Liu G Z. Biologically inspired task oriented gist model for scene classification[J]. *Computer Vision and Image Understanding*, 2013, 117(1): 76 - 95.

[48] Li L J, Wang C, Lim W, et al. Building and using a semantic visual image hierarchy[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010: 3336 - 3343.

[49] Ferrari T D V. Visual and semantic similarity in ImageNet[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011: 1777 - 1784.

[50] Gazolli K, Salles E. A contextual image descriptor for scene classification[C]// *Proc. of the Information Assurance and Security Trends in Innovative Computing*, 2012: 66 - 71.

[51] Gazolli K, Salles E. Combining holistic descriptors for scene classification[C]// *Proc. of the International Conference on Computer Vision Theory and Applications*, 2013: 315 - 320.

[52] de Souza Gazolli K A, Salles E O T. Using holistic features for scene classification by combining classifiers[J]. *Journal of WSCG*, 2013, 21(1): 41 - 48.

[53] Chu W T, Chen C H, Hsu H N. Color CENTRIST: embedding color information in scene categorization[J]. *Journal of Visual Communication and Image Representation*, 2014, 25(5): 840 - 854.

[54] Xiao Y, Wu J, Yuan J. mCENTRIST: a multi-channel feature generation mechanism for scene categorization[J]. *IEEE Trans. on Image Processing*, 2014, 23(2): 823 - 836.

[55] Farhadi A, Endres I, Hoiem D, et al. Describing objects by their attributes[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 1778 - 1785.

[56] Nickisch H, Lampert C N, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 951 - 958.

[57] Su Y, Jurie F. Learning compact visual attributes for large-scale image classification[C]// *Proc. of the European Conference of Computer Vision*, 2012: 51 - 60.

[58] Choi S W, Lee C H, Park I K. Scene classification via hypergraph-based semantic attributes subnetworks identification[C]// *Proc. of the European Conference on Computer Vision*, 2014: 361 - 376.

[59] Torresani L, Szummer M, Fitzgibbon A. Efficient object category recognition using classemes[C]// *Proc. of the European Conference on Computer Vision*, 2010: 776 - 789.

[60] Zhou B, Jagadeesh V, Piramuthu R. Concept learner: discovering visual concepts from weakly labeled image collections[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1492 - 1500.

[61] Pandey M, Lazebnik S. Scene recognition and weakly supervised object localization with deformable part-based models[C]// *Proc. of the IEEE International Conference on Computer Vision*, 2011: 1307 - 1314.

[62] Sadeghi F, Tappen M F. Latent pyramidal regions for recognizing scenes[C]// *Proc. of the European Conference of Computer Vision*, 2012: 228 - 241.

[63] Zheng Y, Jiang Y G, Xue X. Learning hybrid part filters for scene recognition[C]// *Proc. of the European Conference of Computer Vision*, 2012: 172 - 185.

[64] Li J L, Su H, L F F, et al. Object bank: an object-level image representation for high-level visual recognition[J]. *International Journal of Computer Vision*, 2014, 107(1): 20 - 39.

[65] Ahn L. Games with a purpose[J]. *Computer*, 2006, 39(6): 92 - 94.

[66] Russell B C, Torralba A, Murphy K P. LabelMe: a database and web-based tool for image annotation[J]. *International Journal of Computer Vision*, 2008, 77(1): 157 - 173.

[67] Mesnil G, Rifai S, Bordes A, et al. Unsupervised and transfer learning under uncertainty: from object detections to scene categorization[C]// *Proc. of the International Conference on Pattern Recognition Applications and Methods*, 2013: 345 - 354.

[68] Rosiwasia N, Vasconcelos N. Holistic context models for visual recognition[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2012, 34(5): 902 - 917.

[69] Antanas L, Hoffmann M, Frasconi P. A relational kernel-based approach to scene classification[C]// *Proc. of the IEEE Workshop on Applications of Computer Vision*, 2013: 133 - 139.

[70] Li C, Parikh D, Chen T. Automatic discovery of groups of objects for scene understanding[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 2735 - 2742.

[71] Li X, Guo Y. An object co-occurrence assisted hierarchical model for scene understanding[C]// *Proc. of the British Machine Vision Conference*, 2012: 1 - 11.

[72] Luo J, Savakisa A E, Singhal A. A Bayesian network-based framework for semantic image understanding[J]. *Pattern Recognition*, 2005, 38(6): 919 - 934.

[73] Fan J, Gao Y, Luo H. Statistical modeling and conceptualization of natural images[J]. *Pattern Recognition*, 2005, 38(6): 865 - 885.

[74] Bosch A, Zisserman A, Munoz X. Scene classification via PL-

SA[C]// *Proc. of the European Conference on Computer Vision*, 2006: 517 - 530.

[75] Wang Z, Wang R, Ma X. Indoor scene recognition based on the weighting spatial information fusion[C]// *Proc. of the 2nd International Conference on Intelligent System Design and Engineering Application*, 2012: 1040 - 1044.

[76] Rosiwasia N, Vasconcelos N. Latent dirichlet allocation models for image classification[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013,35(11):2665 - 2679.

[77] Yang X, Xu D, Feng S H. Scene categorization with classified codebook model[J]. *IEICE Trans. on Information and Systems*, 2011, 94*D*(6): 1349 - 1352.

[78] Cheng H H, Wang R S. Natural scene classification based on spatial context of local semantics[J]. *Journal of Circuits and Systems*, 2010, 15(6):39 - 46.(程环环, 王润生. 融合空间上下文的自然场景语义建模[J]. 电路与系统学报, 2010, 15(6): 39 - 46.)

[79] Sun X, Fu K, Wang H Q. Spatial semantic objects-based hybrid learning method for automatic complicated scene classification[J]. *Journal of Electronics and Information Technology*, 2011, 33(2): 347 - 354.(孙显, 付琨, 王宏琦. 基于空间语义对象混合学习的复杂图像场景自动分类方法研究[J].电子与信息学报, 2011, 33(2): 347 - 354.)

[80] Meng X L, Wang Z Z, Wu L Z. Building global image features for scene recognition[J]. *Pattern Recognition*, 2012, 45(1): 373 - 380.

[81] Wu J X. A fast dual method for HIK SVM learning[C]// *Proc. of the European Conference on Computer Vision*, 2010: 552 - 565.

[82] Bosch A, Zisserman A, Muñoz X. Scene classification using a hybrid generative/discriminative approach[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008, 30(4): 712 - 727.

[83] Yu J, Tao D, Rui Y, et al. Pairwise constraints based multiview features fusion for scene classification[J]. *Pattern Recognition*, 2013, 46(2): 483 - 496.

[84] Kobayashi T. BFO meets HOG: feature extraction based on histograms of oriented pdf gradients for image classification[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 747 - 754.

[85] Parizi S, Oberlin J, Felzenszwalb P. Recon-figurable models for scene recognition[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 2775 - 2782.

[86] Redi M, Merialdo B. Marginal-based visual alphabets for local image descriptors aggregation[C]// *Proc. of the 19th ACM International Conference on Multimedia*, 2011: 1429 - 1432.

[87] Redi M, Merialdo B. Exploring two spaces with one feature: kernelized multidimensional modeling of visual alphabets[C]// *Proc. of the 2nd ACM International Conference on Multimedia Retrieval*, 2012: 20 - 27.

[88] Redi M, Merialdo B. Fitting Gaussian copula for efficient visual codebooks generation[C]// *Proc. of the 10th International Workshop on Content-Based Multimedia Indexing*, 2012: 1 - 6.

[89] Redi M, Merialdo B. Direct modeling of image keypoints distribution through copula-based image signatures[C]// *Proc. of the 3rd ACM International Conference on Multimedia Retrieval*, 2013: 183 - 190.

[90] Redi M, Merialdo B. Saliency moments for image categorization[C]// *Proc. of the 1st ACM International Conference on Multimedia Retrieval*, 2011: 1 - 8.

[91] Li Q, Wu J, Tu Z. Harvesting mid-level visual concepts from large-scale internet images[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 851 - 857.

[92] Wang X, Wang B, Bai X, et al. Max-margin multiple-instance dictionary learning[C]// *Proc. of the 30th International Conference on Machine Learning*, 2013: 846 - 854.

[93] Shen L, Wang S, Sun G, et al. Multi-level discriminative dictionary learning towards hierarchical visual categorization[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 383 - 390.

[94] Liu J, Yang Y, Shah M. Learning semantic visual vocabularies using diffusion distance[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 461 - 468.

[95] Csurka G, Dance C, Fan L. Visual categorization with bags of keypoints[C]// *Proc. of the European Conference on Computer Vision*, 2004: 1 - 16.

[96] Jurie F, Triggs B. Creating efficient codebooks for visual recognition[C]// *Proc. of the IEEE International Conference on Computer Vision*, 2005: 604 - 610.

[97] Oliveira G, Nascimento E, Vieira A. Sparse spatial coding: a novel approach for efficient and accurate object recognition[C]// *Proc. of the IEEE International Conference on Robotics and Automation*, 2012: 2592 - 2598.

[98] Wang Z L, Feng J S, Shui C Y, et al. Linear distance coding for image classification[J]. *IEEE Trans. on Image Processing*, 2013, 22(2): 537 - 548.

[99] Morioka N, Satoh S. Building compact local pairwise codebook with joint feature space clustering[C]// *Proc. of the European Conference on Computer Vision*, 2010: 692 - 705.

[100] Cakir F, Güdükbay U, Ulusoy Ö. Nearest-neighbor based metric functions for indoor scene recognition[J]. *Computer Vision and Image Understanding*, 2011,115(11): 1483 - 1492.

[101] Xie L, Wang J, Guo B, et al. Orientational pyramid matching for recognizing indoor scenes[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 1 - 8.

[102] Kwitt R, Vasconcelos N, Rasiwasia N. Scene recognition on the semantic manifold[C]// *Proc. of the European Conference of Computer Vision*, 2012: 359 - 372.

[103] Gu G H, Zhao Y, Zhu Z F. Spatial distribution descriptor based keypoints matching algorithm[J]. *Optical Engineering*, 2011, 50(9): 1 - 9.

[104] Fornoni M, Caputo B. Indoor scene recognition using task and saliency-driven feature pooling[C]// *Proc. of the British Machine Vision Conference*, 2012: 1 - 12.

[105] LeCun Y, Boser B, Denker J, et al. Back-propagation applied to handwritten zip code recognition[J]. *Neural Computation*, 1989, 1(4): 541 - 551.

[106] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// *Proc. of the Advances in Neural Information Processing Systems*, 2012: 1097 - 1105.

[107] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//

Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 1 - 20.

[108] Razavian A S, Azizpour H, Sullivan J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 512 - 519.

[109] Sermanet P, Eigen D, Zhang X, et al. OverFeat: integrated recognition, localiza-tion and detection using convolutional networks[C]// *Proc. of the International Conference on Learning Representations*, 2014: 1 - 16.

[110] Donahue J, Jia Y, Vinyals O. DeCAF: a deep convolutional activation feature for generic visual recognition[C]// *Proc. of the International Conference on Machine Learning*, 2014: 1 - 10.

[111] Jia Y Q. Caffe: an open source convolutional architecture for fast feature embedding[EB/OL]. [2015 - 05 - 30]. http://caffe.berkeleyvision.org/, 2013.

[112] Gong Y C, Wang L W, Guo R Q, et al. Multi-scale orderless pooling of deep convolutional activation features[C]// *Proc. of the European Conference on Computer Vision*, 2014: 392 - 407.

[113] Yoo D, Park S, Lee J, et al. Fisher kernel for deep neural activations[J]. *The Computing Research Repository*, 2014: 1 - 11.

[114] Agrawal P, Girshick R, Malik J. Analyzing the performance of multilayer neural networks for object recognition[C]// *Proc. of the European Conference of Computer Vision*, 2014: 329 - 344.

[115] Azizpour H, Razavian A S, Sullivan J, et al. From generic to specific deep representations for visual recognition[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 36 - 45.

[116] Koskela M, Laaksonen J. Convolutional network features for scene recognition[C]// *Proc. of the ACM International Conference on Multimedia*, 2014: 1169 - 1172.

[117] Zhou B, Khosla A, Lapedrza A, et al. Object detectors emerge in deep scene CNNs[C]// *Proc. of the International Conference on Learning Representations*, 2015: 1 - 12.

[118] Cimpoi M, Maji S, Vedaldi A. Deep filter banks for texture recognition and segmentation[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3828 - 3836.

[119] Singh S, Gupta A, Efros A. Unsupervised discovery of mid-level discriminative patches[C]// *Proc. of the European Conference of Computer Vision*, 2012: 73 - 86.

[120] Sun J, Ponce J. Learning discriminative part detectors for image classification and cosegmentation[C]// *Proc. of the IEEE International Conference on Computer Vision*, 2013: 3400 - 3407.

[121] Juneja M, Vedaldi A, Jawahar C. Blocks that shout: distinctive parts for scene classification[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 923 - 930.

[122] Doersch C, Gupta A, Efros A. Mid-level visual element discovery as discriminative mode seeking[C]// *Proc. of the Advances in Neural Information Processing Systems*, 2013: 1 - 9.

[123] Zuo Z, Wang G, Zhao L, et al. Learning discriminative and shareable features for scene classification[C]// *Proc. of the European Conference of Computer Vision*, 2014: 552 - 568.

[124] Wu J. Power mean SVM for large scale visual classification[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 2344 - 2351.

[125] Ali W, Nock R, Nielsen F, et al. Fast newton nearest neighbors boosting for image classification[C]// *Proc. of the IEEE International Workshop on Machine Learning for Signal Processing*, 2013: 1 - 6.

[126] Gao T, Koller D. Discriminative learning of relaxed hierarchy for large-scale visual recognition[C]// *Proc. of the IEEE International Conference on Computer Vision*, 2011: 2072 - 2079.

[127] Gu G H, Li F C, Zhao Y, et al. Scene classification based on spatial pyramid representation by superpixel lattices and contextual visual features[J]. *Optical Engineering*, 2012, 51(1): 1 - 9.

[128] Nakayama H, HaradaT, KuniyoshiY. Global Gaussian approach for scene categorization using information geometry[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010: 2336 - 2343.

[129] Zhou L, Zhou Z, Hu D. Scene classification using a multi-resolution bag-of-features model[J]. *Pattern Recognition*, 2013, 46(1): 424 - 433.

[130] Zhu J, Li L J, Li F F, et al. Large margin learning of upstream scene understanding models[C]// *Proc. of the Advances in Neural Information Processing Systems*, 2010: 2586 - 2594.

[131] Wang L, Li Y, Jia J, et al. Learning sparse covariance patterns for natural scenes[C]// *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 2768 - 2774.

[132] Bo L, Ren X, Fox D. Unsupervised feature learning for RGB-D based object recognition[J]. *Experimental Robotics*, 2013, 88: 387 - 402.

[133] Bu S, Liu Z, Han J, et al. Superpixel segmentation based structural scene recognition[C]// *Proc. of the 21st ACM International Conference on Multimedia*, 2013: 681 - 684.

[134] Zhu J, Wu T, Zhu S C, et al. A reconfigurable tangram model for scene representation and categorization[J]. *IEEE Trans. on Image Processing*, 2014: 1 - 13.

[135] Han Y, Liu G. Efficient learning of sample-specific discriminative features for scene classification[J]. *Signal Processing Letters*, 2011, 18(11): 683 - 686.

[136] Zhou L, Zhou Z, Hu D. Scene classification using multi-resolution low-level feature combination[J]. *Neurocomputing*, 2013, 122: 284 - 297.

作者简介:

顾广华(1979 -),男,副教授,博士,主要研究方向为图像信号处理、模式识别。

E-mail: guguanghua@ysu.edu.cn

韩晰瑛(1990 -),女,硕士研究生,主要研究方向为图像场景分类。

E-mail: hxy_0510@126.com

陈春霞(1990 -),女,硕士研究生,主要研究方向为图像场景分类。

E-mail: 15033502352@163.com

赵 耀(1967 -),男,教授,博士研究生导师,博士,主要研究方向为跨媒体技术。

E-mail: yzhao@bjtu.edu.cn