

图像内容表示模型综述

张琳波¹ 肖柏华² 王枫¹ 石磊¹

(交通运输部科学研究院 北京 100029)¹

(中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190)²

摘要 图像内容表示问题已成为计算机视觉领域广受关注的焦点问题,针对目标形变、外物遮挡、尺度变化以及背景干扰等挑战,研究者提出了很多有效的解决思路,做了很多优秀的工作。现对近年来比较经典的基于局部特征的图像表示模型做一个回顾,按照基于码本的模型、部分-结构模型、基于轮廓片段的模型、生物认知相关的模型和语境相关模型的顺序总结了过去出现的经典工作及其优点和局限性。

关键词 图像分类,基于内容,表示模型,综述

中图分类号 TP391.3 文献标识码 A

Review on Image Content Representation Models

ZHANG Lin-bo¹ XIAO Bai-hua² WANG Feng¹ SHI Lei¹

(China Academy of Transportation Sciences, Beijing 100029, China)¹

(State Key Laboratory of Management and Control for Complex Systems, Institute of Automation,
Chinese Academy of Sciences, Beijing 100190, China)²

Abstract Content-based image representation has become one of the most popular problem in the field of computer vision. To deal with the challenge of object deformation, object occlusion, scale variability and background confusion, a lot of strategies have been proposed, and a large number of works have been produced. This paper presented a review on the classic works related to content-based image representation, which are in the order of codebook-based models, part-structure models, contour-fragment-based models, biological-cognition-related models and context related models. In addition, the advantages and limitations of each model were also provided.

Keywords Image classification, Content-based, Representation model, Review

1 前言

图像内容表示是图像分类系统的重要组成部分,在机器学习领域现有分类技术性能确定的情况下,图像内容表示方法的优劣决定着最终分类系统性能的高低,因为如果不同类别的图像在被映射到的模式空间中不具备可分性,再好的分类器也无法得到令人满意的图像分类性能。鉴于图像内容表示在图像内容分类系统中的重要地位,寻找一种有效的图像表示方法,成为图像内容分类研究领域的一个热点问题。直观上讲,图像内容的表示模式,亦称图像特征,应该包含图像尽可能多的有用信息,如,颜色、纹理、连通域、特殊面、边缘和角点的分布情况等。然而,目前广泛应用的图像表示或图像特征并非如此,一方面,处理此类信息需要花费大量的计算资源,导致整个系统效率低下,而且此类信息之间存在一定的冗余并非全部有助于分类;另一方面,这些多方位的信息可能对特定模式的图像内容以及图像中目标的位置比较敏感,从而限制了图像内容表示和分类系统的推广能力。因此,研究

判别性和推广性都比较强的图像表示方法,成为图像分类技术中的研究重点。按照底层特征组织方式,图像内容表示方法大致可以分为:基于全局特征的表示方法和基于局部特征的表示方法。

全局特征的提取对象通常是整幅图像,它将整幅图像看作一个整体并抽取图像中包括感兴趣区域和背景区域在内的全部信息,用以表示图像。全局特征具有抽取速度快、可以有效利用图像上下文信息等优点,因此在图像分类系统中曾得到广泛的应用^[1],产生了大量优秀的全局特征描述方法,如全局颜色直方图^[2]、基于滤波器的全局特征^[3-5]、HOG^[6]、LBP^[7]等。然而,全局特征只能描述整幅图像的内容,而不能对目标和背景区别对待。在这种情况下,全局特征对图像亮度变化、目标遮挡、目标位移甚至同一目标的变形都比较敏感,这严重限制了它在图像内容分类中的应用。基于全局特征的上述局限性,许多研究者开始研究基于局部特征的图像表示方法,以便在变化的图像中寻找局部不变的部分。而以前大量对整幅图像进行操作的全局特征也开始被用于描述局

到稿日期:2012-09-29 返修日期:2013-01-26 本文受国家自然科学基金项目(60802055,60933010,60835001)资助。

张琳波(1983-),男,博士,助理研究员,主要研究方向为模式识别、计算机视觉、图像内容理解、交通运输信息化, E-mail: zhanglinbo@163.com;
肖柏华(1974-),男,研究员,主要研究方向为模式识别、机器学习、计算机视觉。

部图像区域,并演化为局部特征。

由于基本符合认知心理学家研究人类视觉得出的结论以及方法本身对图像内容变化所表现出的优越性能,基于局部特征的图像内容表示分类方法在过去几年得到了快速发展,出现了大量的优秀成果。纵观这些成果,其区别主要表现在以下几个方面:局部特征的检测和描述方法,局部特征的组织方法以及分类技术的选择。本文不涉及局部特征的检测、描述以及分类技术的选择,而重点介绍几种常见的用于图像表示的局部特征组织模型。为叙述方便,本文将按照基于码本(字典)的图像表示模型、部分-结构模型、基于轮廓片段的模型、生物-认知相关的模型以及语境相关的模型进行描述。严格地讲,这并不是严格意义上的划分,如,很多基于轮廓片段的模型也使用了部分-结构模型的思想,且许多语境相关的模型采用的是全局特征表示方法。但是,将这些方法按照模型进行归类,可使研究者更方便地掌握高层次图像内容表示方法的研究概况。下面各部分将对每种模型中一些经典的工作进行介绍。

2 基于码本的模型

基于码本(又称词典)的图像表示模型中,研究最广泛的当属 bag-of-words (亦称“词袋”, bag-of-features 或 bag-of-keypoints)模型。bag-of-words 方法是文档分类领域主要的文档表示方法^[8]之一,它首先被 Zhu 等^[9]引入计算机视觉用于图像检索。接着, Csuska 等人将 Zhu 等人用到的“key-blocks”扩展到具备一定程度不变性的“keypoints”(也即 visual words)^[10]。Bag-of-words 图像表示方法的主要思想是:首先,利用局部特征检测子和描述子在所有图像上提取局部特征;然后,利用得到的局部特征生成固定数目的具有代表性的局部特征,称为“码字”,所有的码字作为一个整体称为码本(或字典);最后,将每幅图像中提取的局部特征映射到码本中的码字上,并在每幅图像中统计码本中所有码字出现的次数,将统计得到的直方图作为该图像的 bag-of-words 图像表示。图 1 给出了这个表示方法的一个直观理解。

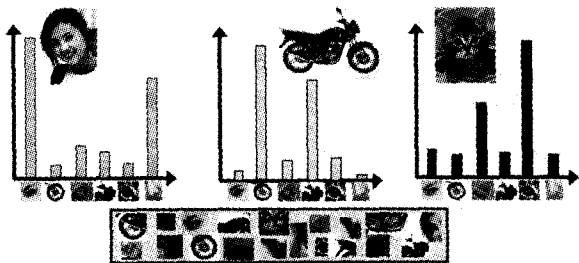


图 1 bag-of-words 表示方法示意图

bag-of-words 表示方法有许多优点:(1)受图像大小和复杂程度影响,从不同图像抽取的局部特征数量通常不同,bag-of-words 方法可以将具有不同数目局部特征的图像映射为具有相同长度的表示向量;(2)图像中的图像块是连续变化的,bag-of-words 模型的字典生成部分将这种连续的变化,通过聚类、高斯建模等方式映射到有限的、离散的码字上面,使得千变万化的图像变化体现在有限长度的向量中;(3)bag-of-words 方法放弃了局部特征之间的空间排列信息,使得 bag-of-words 方法对遮挡、目标位移以及类内变化都比较稳定。

大量研究表明,应用中间表示可以提高分类系统的性

能^[1,11]。基于这方面的考虑,部分研究者在基本 bag-of-words 表示方法的计算分析阶段,引入了文档分类中的话题(topic)模型。例如,pLSA(probabilistic Latent Semantic Analysis)^[12,13]和 LDA(Latent Dirichlet Allocation)^[14]等潜语义模型分别被 Sivic 等^[15]和 Li 等^[16]引入到计算机视觉领域。在低维话题图像表示的基础上,Rasiwasia 和 Vasconcelos^[17]引入了中间层空间(intermediate space)的概念,使得每幅图像即使在没有多个训练标签的情况下也可以与多个话题关联。Fritz 和 Schiele^[18]利用话题模型进行图像分解,然后在这个分解的基础上进行目标的检测和分类。此外,文献[19-21]中也应用了话题模型。上述潜语义模型的基本原理是:包含经常“共现”(co-occurring)的局部特征的不同图像在潜语义空间中的表示应该具有相似性。潜语义模型可以解决多义词和同义词问题——相同的局部特征可以在不同的图像语境中表示不同的内容,以及不同的局部特征可以在特定的图像语境中表示不同的内容^[20]。潜语义模型的另一个优点便是可以移除基本 bag-of-words 表示中的冗余信息,提供更有意义的图像表示。

还有一些研究者认为,空间几何信息在图像分类中起着重要的作用。局部特征之间的几何空间排列关系的全部丢失,严重限制了 bag-of-words 方法的性能。Lazebnik 等人^[22]采用将图像逐渐细分为子区域、然后在每个子区域内分别计算 bag-of-words 直方图向量的方法,来近似估计图像不同子区域局部特征之间的几何对应关系。这种加入位置信息的思路,提高了 bag-of-words 方法的表示性能,并被越来越多的研究者采用。类似地,文献[23]在基本 bag-of-words 的基础上提出了空间视觉词袋模型。针对上述问题,文献[24]从分析图像内容表示与文档分类中文档表示之间的关系出发,给出了短语包模型。

除了上述 bag-of-words 方法,稀疏编码理论及应用^[25]是基于码本模型的另一个研究热点。通过定义稀疏性约束来优化学习重构误差最小,稀疏编码可以得到类似于简单细胞响应特性的基函数。Wright^[26]假设一幅新的人脸图片可以由训练集中同类的训练图片稀疏线性表示,成功地将稀疏表示的思想运用到人脸识别当中。Yang^[27]对密集采样的局部区域 SIFT 特征进行稀疏编码,结合线性空间匹配金字塔策略,实现对图像的特征抽取,得到了非常好的分类性能。Chatfield 等人在文献[28]中对几种常用的局部特征编码方法进行了比较和评价。

3 部分-结构模型

上述 bag-of-words 模型的主要缺点之一即是局部特征之间相对空间几何关系的丢失。本节将介绍充分考虑局部图像块/局部特征之间几何关系的模型——部分-结构(part-structure)模型。“部分-结构”模型的概念,最早是由 Fischler 和 Elschlager^[29]提出的。这种模型的基本思想是:将图像中的每个局部区域进行独立描述,然后利用弹簧式的连接来建模这些部分之间可变的结构关系,如图 2 所示。表达这种思路的很自然的选择便是图模型,即让图节点(node)表示局部特征/图像块,然后利用图的边(edge)建模不同局部特征之间的相互位置关系^[29]。然而,这种图像表达方式的一个主要困难便是图模型的求解,从计算量考虑,研究者不得不在表达能力和计算复杂度之间取一个折中;此外,局部特征/图像块的选

择方式也是需要慎重考虑的问题。在部分-结构的思想框架下,研究者也做出了大量的工作。

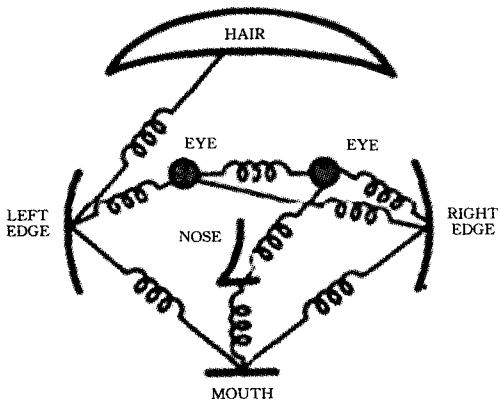


图2 部分-结构模型示意图^[29]

星座模型是部分-结构框架下的一种重要图像表示方法,该模型的研究最早可以追溯到 Burl 等人的文章^[30,31]。然后,该模型经过一系列改进^[32-35]日臻完善。在星座模型^[34]中,每个局部区域都有外观 A (即局部特征向量)、相对尺度 S 、相对位置 X 、启发 h (指明该局部区域是否被遮挡) 等描述。目标的形状使用局部区域块之间的相对位置关系来表示。局部外观 A 、相对尺度 S 、相对位置 X 、启发 h 都使用概率密度函数建模。在文献^[34]中采用高斯函数对上述描述进行建模,公式中参数采用最大似然估计 (maximum-likelihood, ML) 的方法从训练数据中获得。给定一幅测试图像,分类结果由预定的阈值与贝叶斯决策 R 的比较结果确定。随后, Li 等人给出了该模型一次性学习 (one-shot learning)^[36] 和增量 (incremental) 学习^[37] 的方法。

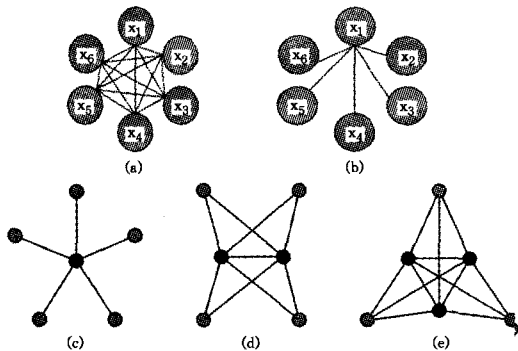


图3 依次为星座模型、HSM 模型、1-fan、2-fan 和 3-fan 模型直观示意图

然而,基本的星座模型^[34]也存在一些不足:(1)局部图像块特征的增加,将会导致运算量呈指数级增长,因此严重限制了该模型在表示每幅图像过程中使用的局部特征的数目;(2)每幅图像仅选择很少的局部特征,使得这个模型对类别的属性比较敏感;(3)模型需要学习大量参数,在样本量不足的情况下,很容易造成过拟合现象。鉴于这些问题, Fergus 等人^[35]提出了星座模型的改进版——异构星形模型 (heterogeneous star-shaped model, HSM)。该方法在局部区域(局部特征)中选择一个地标块 (landmark part), 然后对所有图像块与地标块的位置关系进行建模 (如图 3(c) 所示)。与基本的星座模型^[34]相比, HSM 可以显著减少在相同数量局部特征 (或图像块) 情况下整体的运算量。然而, 当被 HSM 模型选作地标块的部分被遮挡时, 这种方法就开始显得无能为力。后来, Crandall 等人提出了 k -fans 模型^[38], 即通过一个参数 k 来控制

制处理遮挡能力和计算量的关系,如图 3(c)–(e) 3 幅图片分别给出了当 $k=1, 2$ 和 3 时的示意图。从图 3 可以看出, 星座模型和 HSM 模型分别对应着 k -fans 模型中 $k=N$ 和 $k=1$ 的特殊情况。

星座模型不适合对铰链式目标 (例如伸展或收回上肢的人) 进行建模, 因为, 这类目标局部区域之间位置关系变化幅度很大, 很难利用高斯分布函数捕获这些部分之间的排列关系。作为 Fischler 和 Elschlager 工作^[29] 的直接发展, Felzenszwalb 等人提出了铰链结构模型 (pictorial structure model)^[39,40], 如图 4 所示。在这些工作中, Felzenszwalb 等人的贡献可以归为几点^[40]: (1) 为大量铰链结构模板在图像中寻找全局最优匹配提供了一种有效的算法, 该算法中的能量函数不需要任何初始化; (2) 提供了一种利用训练图像学习模型参数的框架; (3) 采用概率采样技术, 可以为目标在图像中的位置提供多个好的候选, 而不仅仅给出一个最优的。

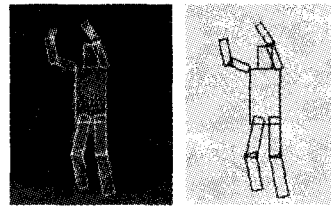


图4 铰链结构模型建模实例^[40]

在后来的工作^[41]中, Felzenszwalb 等人又提出了一种判别式训练的、多尺度的、可变形的模型, 用来进行目标检测, 如图 5 所示。在该方法中, 每个目标的表示模型包含一个全局根滤波器 F_{root} 和许多部分模块 M_i ; 每个部分模块 M_i 包含一个位置模块 L_i 和一个部分滤波器 F_i ; 位置模块 L_i 包含着部分模块 M_i 相对于全局根滤波器 F_{root} 所有可以停靠的位置及每个位置对应的变形代价 (deformation cost)。全局滤波器 F_{root} 和每个部分模块 M_i 都会对当前位置给出一个评分, 最终的结果取决于全局滤波器和各个部分模块得分情况的综合结果。该检测方法在 PASCAL VOC2008 目标检测竞赛中取得了第一名的好成绩。

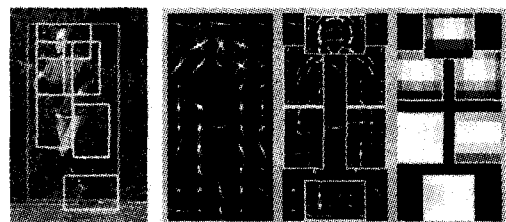


图5 目标建模示意图^[41]

此外, Felzenszwalb 等人还提出了利用多个三角形描述变化的形状, 然后利用图模型对这些三角形进行建模, 如图 6 所示, 并为上述图模型提供了一种有效的动态规划 (dynamic programming) 求解方法^[42]。后来他们还提出了利用层次化描述目标边缘来描述可变形状的弹性匹配框架^[43], 这个框架可以同时应用目标的全局和局部的几何属性。所有这些, 都或多或少地沿用了部分-结构模型的基本思想。

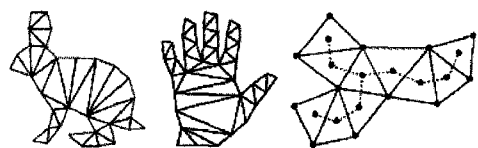


图6 三角形图像表示和图模型建模示意图^[42]

4 基于轮廓片段的模型

前面几节介绍的建模方法的局部特征,大部分采用局部特征检测子检测到局部图像块,然后描述这些局部图像块的纹理等信息的方式得到;但是,上述模型大都没有完全挖掘图像中包含的轮廓信息。人类可以利用轮廓信息识别上百类目标^[44,45]的事实表明,轮廓边缘信息对正确分类具有重要的作用,如图7所示。在这里,轮廓指的是目标的外部边缘(outline)和目标内检测到的边缘(edge)。刚开始,研究者采用模板的方式,如:刚性模板(rigid template)^[46]、可变模板(deformable template)^[47]以及类原型(class prototype)^[48]和相似视角组合^[49]来表示目标。但是,这种采用模板的方法是匹配整个目标的轮廓,泛化能力(generalization)不强。因此,研究者开始关注基于轮廓片段的图像表示及分类方法^[50,51]。

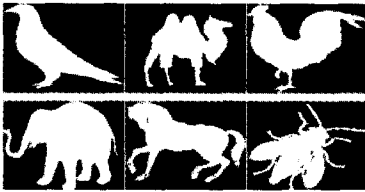


图7 目标轮廓样

Shotton 等人在文献^[52]中采用了两阶段分类的方法。首先,训练阶段利用边缘检测得到边缘局部特征训练码本。这里的码本与前面 bag-of-words 中码本的不同在于,这里的码字是轮廓边缘片段,且包含得到它的目标的中心位置信息;然后,分类阶段利用码本中的边缘片段拟合新图像中检测到的边缘,并对目标类别和目标中心位置进行投票,进而完成图像内容分类和目标检测,如图8所示。

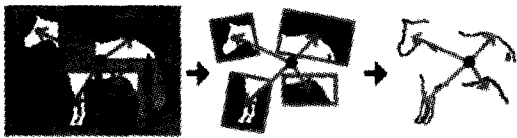


图8 Shotton 等人的基于轮廓的模型^[52]

Opelt 等人^[53]与 Shotton 等人^[52]一样,也采用了 boosting 的方式。但与 Shotton 随机选择边缘片段方法的不同之处在于,Opelt 等人^[53]采用的方式是随机地在轮廓上选择一个种子点,然后使种子点沿着边缘生长(grow),直到获得最优长度轮廓片段。这里的最优,指的是可以在校验集中得到最大的斜面匹配分数和最精确的预测目标的位置。此外,文献^[53]采用 $k > 1$ 个轮廓片段构建弱分类器,而文献^[52]采用了 $k = 1$ 。在文献^[54]中,Opelt 等人进一步改进了他们自己的方法,使其不仅可以进行多目标识别,而且还可以增量学习新类别(enabling incremental additions of new categories)。作为自己方法^[52]和 Opelt 等人^[53,54]方法的扩展,Shotton 在文献^[55]中使用了多尺度。

在上面的许多文献中,已经提到了切割的概念^[50,51],因为轮廓中用到的边缘映射(edge map)也正是切割算法的研究热点。严格来讲,边缘检测、目标检测与分类以及目标切割往往是耦合在一起的^[55];精确地切割有助于目标的精确定位和分类,而正确的分类和定位可以为切割提供好的初始条件。究竟是先切割后分类还是先分类后切割,是计算机视觉经典

的“鸡与蛋”的问题,这里不再赘述。下面,本文将介绍几种将切割融入图像分类的算法,使读者对图像切割在图像分类中的应用有所了解。Leibe 等人的隐式模型^[56,50]将识别过程和切割过程融合到了一个框架下面,该隐式模型并没有采用显式的框架来定义目标的出现形式,而是定义了目标的不同局部区域相互位置的可取范围。图像切割方法大致可以分为自上而下(top-down)和自下而上(bottom-up)两种形式。自上而下的方法是利用从图像样例中学习到的先验知识(形状、颜色和纹理信息等)来指导切割;自下而上的方法是先将图像切割为小的区域,然后确定区域的类别标签。鉴于自上而下的分割方法无法很好地拟合变化多端的目标形状和纹理,而自下而上的方法容易把相同目标的区域分为不同的部分等问题,Borenstein 等人^[57]将多尺度多层次的自下而上图像切割与自上而下的学习得到的目标形状的先验知识融合到了一个全局的代价函数(global cost function)中。与之类似的是,Kumar 等人^[58]将自上而下的图示结构^[40]和自下而上的马尔科夫随机场(Markov Random Fields, MRFs)^[59]整合到一起进行图像切割与分类。

为了凸显边缘、轮廓片段在图像分类(或者同时进行图像切割和分类)中的作用,上述方法很多只应用了边缘轮廓信息。其实,将其它的信息(例如,外观、颜色纹理等)和边缘轮廓片段信息进行融合来提高分类性能,也是研究的热点问题^[60]。在文献^[61]中,Shotton 等提出了一种图像分类的判别式模型,其有效地融入了外观、轮廓形状和上下语境信息。

5 生物认知相关的模型

图像内容分类(识别)的长远目标就是能够建立可达到甚至超过人类视觉系统的目标定位、分类识别系统。然而截至目前,人类等灵长类的视觉系统在几乎所有的度量上都超过最好的机器视觉系统^[62]。在人类等灵长类视觉系统的启发和挑战下建立分类系统也是非常具有吸引力的研究方向。虽然生理学方面的研究至今不能为人类视觉优于机器视觉性能给出一个令人满意的解释^[63],研究者在视觉皮层研究方面还是做了很多有价值的工作^[63-66]。

神经生理学和心理学中关于注意力(attention)的研究开始于20世纪80年代。由于对注意力中“自上而下”的意愿的作用研究相对较少^[67,68],大量关于注意力的研究集中于对“自下而上”的基于显著性的研究^[69]。大量的实例和严格的量化分析(rigorous quantitative analysis)表明,注意力的研究可以将图像内容分类问题分解为计算要求不高的、快速的局部视觉分析问题,进而显著提高复杂场景中的目标分类和识别性能^[64]。人类大脑皮层中基于注意力的表示中有一个反馈模块^[70,71],其可以增强大脑皮层在感兴趣区域的表示能力,因此开发一个描述大脑皮层中注意力分配的计算模型具有重要的潜在应用价值^[69]。在人类视觉皮层的图像内容分类与注意力分配等方面,研究者做出了大量的工作^[72-74],只是这些工作大部分停留在理论阶段或者仅仅在几幅特定图像上进行实验,而没有在国际通用的标准大数据集上与其它方法进行对比。

除此之外,还有部分研究者更关注大脑视觉皮层(visual cortex)的结构。他们关注的焦点是视觉场景在大脑皮层是如何分层次处理的^[75],进而建立人工系统来模拟这种行

为^[76,77]。视觉皮层中后向传递(back projection)模块的作用目前还存在争议,但是视觉皮层中基本信息处理采用前馈的方式却得到了实验的证明^[78,79]。虽然上面的实验结果没有排除区域内的局部反馈,但却说明核心层前馈结构是解释大脑皮层快速识别理论的一个合理的起点。在借鉴与灵长类视觉皮层的腹侧流(ventral stream)在目标识别过程前100~200毫秒对应的前馈路径相关的HMAX理论^[75,77]的基础上,Serre等人提供了一种分层的系统^[62,63],即通过交替使用模板匹配和最大池(maximum pooling)操作,建立一个逐步复杂和稳定的表示方法,如图9所示。计算神经学的Olshausen和Field发现^[80],利用自然图像的统计信息增加稀疏约束对学习一个好的生物认知模型很关键。在这个结论的影响下,Mutch和Lowe^[81]对Serre等人^[62,63]的工作进行了改进,增加了一些从生理学研究得到的特性,特别是稀疏表示和局部中间层特征,进一步改进了分类性能。此外,Bileschi等人HMAX模型^[77]的基础上,提出了多尺度的HMAX描述子^[82]。

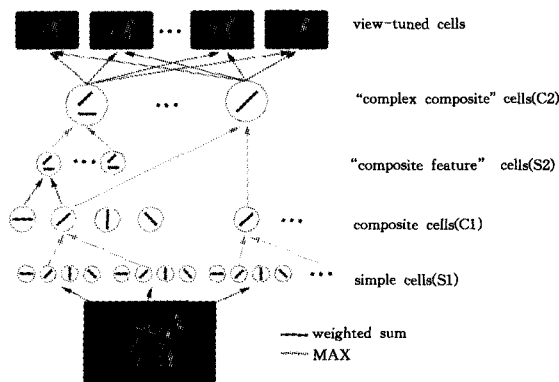


图9 Riesenhuber等人提供的HMAX模型示意图^[75,77]

生理学研究发现在灵长类视觉系统的大脑皮层中,反馈连接(feedback connection)与前馈连接(feedforward connection)所占的比例大约为10:1^[83]。这说明,一个完整的模拟系统应该包含反馈的模块。Hochstein和Ahissar在反向层次理论(Reverse Hierarchy Theory, RHT)^[84]中提出,在人类的视觉系统中,信息首先采用自动的“自下而上”的处理方式,然后知觉采用相反的方向给出反馈。这个结论为在现有模型的基础上增加反馈模块来提高分类性能的策略,提供了生理学上的支持。

对一部分研究者来说,将现有的不同特征融合在同一个系统当中是提高系统性能最有效的方式之一。然而,特征的融合不仅仅是不同特征的简单累加,而应该根据这些特征的不同特点进行相应的处理。Riesenhuber指出^[66],在大脑皮层中,相邻的神经细胞对相似的刺激产生响应,顶下小叶(inferior parietal lobule, IPL)区域是将不同信息融合在一起的关键部位。因此,利用IPL区域的研究成果指导多特征融合系统的设计也是一种重要的思路。

6 语境相关模型

图像中的上下语境(context)信息,对图像分类的结果起着重要作用^[85,86,88],来自心理学的研究成果也为这些信息的重要作用提供了心理学上的证据。Biederman等人发现,人类对异常上下语境目标(如天空中出现汽车)的识别要花费比

正常情况(天空中出现飞机)更多的时间^[87]。此外,Cox等^[88]和Bar等^[89]也为人类利用语境帮助识别提供了功能性磁共振成像(functional MRI)的证据。上下语境所指的范围比较广泛,它可以是感兴趣目标周围其它目标的性质^[90,91]、相对尺度或位置^[92];也可以指场景中所有底层特征作为一个整体的统计特性^[86,93]。Wolf和Bileschi在文献^[94]中给出了上下语境的一个定义:与检测或分类任务相关但却不是感兴趣目标的直接物理外观的信息。一般来讲,上下语境相关的方法大致可以分为两类,一类对目标或者分割的局部片段之间的关系建模,另一类直接将图像整体作为输入。

作为第一类的实例,Kumar等人^[95]首先对图像进行切割,然后对切割后图像块之间的关系进行建模。类似地,Parikh等人^[96]在切割好并分类的图像块上应用条件随机场(Conditional Random Field, CRF)建立这些图像块之间目标与目标的关系。Torralba等人^[86]应用boosting和图形化网络(graphical networks)来学习容易“共现”(co-occurring)的目标之间的语义及位置关系。Tu^[97]提出了自动语境(auto-context)算法,该算法利用局部图像块学习分类器,然后利用分类器的判别结果作为上下语境连同开始时的局部图像块再次训练分类器,直到与训练图像的真实标签相同。He等人^[98]提供了隐话题模型(latent topic model)的新形式,该形式可以灵活地对高阶的联合上下语境进行建模。上述利用切割后的图像块建模目标之间关系的上下语境模型的性能,往往依赖于图像切割的质量以及对切割后图像块进行分类的分类器的性能,当图像数据集中仅含图像类别标签而不包含切割信息时,这种方法的劣势很难发挥^[94]。

对于将整幅图像作为输入的上下语境模型,则不存在对图像切割等标注的过分依赖。Murphy等人^[90]利用上下语境作为场景的gist,它影响着场景中目标存在与否以及存在位置的先验。在文献^[99]中,Torralba等人提出一个简单的框架来对语境信息和目标属性之间的关系进行建模。Mei等人^[100]也提出了一种新的图像标注算法,同时利用图像标注信息和通过图像标注信息学习得到的语义距离,达到语义一致的图像标注结果。

一般说来,在目标外观信息相对不足的情况下,如分辨率低、图像噪声大、目标在图像中尺度小等,上下语境信息对正确分类作用明显。而对目标在图像中的外观信息比较丰富的情况,上下文语境的作用相对有限。为诠释对这个结论,Torralba等人^[86]给出了一个简单的例子:给定一幅电脑图片,对显示器进行识别,鼠标等小构件提供的上下语境信息的作用是有限的;而要识别鼠标时,显示器和主机的位置等上下语境信息就非常重要。

7 模型比较

对基于码本的模型来说,当使用的局部特征具有局部变形鲁棒特性的时候,无需对目标的各方面显式建模即可处理目标的各种移位和变形。该模型所提取的局部特征既包括目标上的局部特征也包括背景上的局部特征,当分类对象是整幅图像时,该模型的方法几乎可以达到最好的效果;然而,当识别任务包含目标定位需求,尤其是每幅图像中包含多类别目标的时候,该模型必须借助于其他模型的辅助。此外,特征之间空间排列关系信息的全部或部分丢失,某种程度上成为

该模型性能进一步大幅提升的瓶颈。

与基于码本的模型相比,部分-结构模型最显著的特点是它既能建模局部区域,又能建立这些局部区域之间的空间位置排列关系。空间位置信息的建模,在很大程度上解决了传统基于码本的模型不能进行目标定位等问题。然而,部分-结构模型也存在自己的局限性,例如,该模型大部分方法将局部图像块的数量控制在一个相对低的水平,以保障计算复杂度的可行;部分方法还需要提供半监督学习的保障;该类模型大部分方法要求待识别目标在图像中比较显著,以保证在目标上能够提取足够的局部图像块。

基于轮廓片段的模型与基于码本的模型、部分-结构模型等并没有明显的界线,但该模型的方法由于强调轮廓片段和边缘在目标识别中的作用而常常被研究者单独划分出来与基于码本的模型和部分-结构模型等相提并论。轮廓片段采用基于码本的方式建模,即可认为是基于码本的模型;如果在描述这些片段的过程中同时考虑了轮廓片段和边缘的位置关系,即可归于部分-结构模型的范畴。与传统的基于码本的模型、部分-结构模型等不同,基于轮廓片段的模型更关注充分利用轮廓片段的特点对上述模型进行相应的调整。

生物认知相关的模型更强调在人类等灵长类视觉系统的启发和挑战下建立分类系统,通过进一步研究灵长类的视觉系统改进性能。虽然该模型与上述其它模型基于不同的理念,但是两类模型之间存在着相互作用。其它模型需要经常借鉴生物认知领域的相关知识,为所取得成果寻求生物认知方面的解释,以期为自身性能的进一步提高获取生物认知领域的启示。生物认知相关模型也关注其他分类模型,以得到自身方法的计算机建模思路。

语境相关模型强调不同视觉元素之间的相互作用,利用周围其它元素的识别结果改进整体识别性能。总体来说,大部分语境相关模型并不独立存在,而是借助其他模型的识别结果、“共现”元素的统计特性等信息实现,并进一步对上述信息的改进提供反馈。因此,语境相关模型一般只提供一种不同视觉元素之间相互作用的思路,通常与其他模型方法混合使用。

结束语 图像内容表示是计算机视觉和模式识别领域的重要问题,图像内容表示性能直接决定着图像分类结果的准确度。本文回顾了近几年基于局部特征的图像内容表示的发展,重点介绍了几种经典的局部特征组织模型,并对各种图像表示模型的优点和局限性进行了描述。

虽然图像内容表示工作在过去几年已经取得了很大发展,但是受图像亮度变化、尺度变化、目标遮挡、目标位移甚至同一目标各种变形等因素的影响,该领域的研究成果与实际应用的需求之间还存在着一定差距。因此,图像内容表示模型算法研究还存在很大的发展空间。主要的研究方向包含如下几个方面:

(1)新算法模型的研究

图像内容表示是一个跨学科且极具挑战性的研究课题,研究人员在充分借鉴现有算法模型的基础上,需要积极借鉴其他领域的研究成果,将机器学习、神经生物学以及认知心理学的相关研究成果引入到图像内容表示模型算法的设计中,研究更适合图像内容表示的新模型。

(2)多特征融合

将现有的不同特征进行融合,是提高现有图像表示模型表示性能的一个有效途径。然而,特征的融合不仅仅是不同特征的简单累加,而应该根据这些特征的不同特点,进行相应的处理。Riesenhuber^[66]指出,在大脑皮层中,相邻的神经细胞对相似的刺激产生响应,顶下小叶(inferior parietal lobule, IPL)区域是将不同信息融合在一起的关键部位。因此,利用IPL区域的研究成果指导多特征融合系统的设计,是未来图像内容表示模型设计的重要思路。此外,抛开神经生物方面的研究,充分利用统计学、机器学习等领域的新结论,并利用充足的实验进行验证,也是实现已有异源特征融合的重要手段。

(3)反馈模块的加入

生理学研究表明,灵长类视觉系统的大脑皮层中,反馈连接(feedback connection)与前馈连接(feedforward connection)所占的比例大约为10:1^[83]。反向层次理论(Reverse Hierarchy Theory, RHT)^[84]也认为,在人类的视觉系统中,信息首先采用自动的“自下而上”的处理方式,然后知觉采用相反的方向给出反馈。这说明,一个完整的模拟系统应该包含反馈的模块,而现有的大部分图像表示模型都没有或者仅仅利用实验结果对模型参数进行细微调整,未充分研究反馈在图像内容表示和分类中的作用及其在模型中的表现形式。研究在现有图像表示模型乃至整个分类系统中引入反馈模块,也是未来提高图像内容表示和图像内容分类性能的重要方向。

参考文献

- [1] 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
- [2] Swain M J, Ballard D H. Color indexing [J]. *International Journal of Computer Vision*, 1991, 7(1): 11-32
- [3] Freeman W T, Adelson E H. The design and use of steerable filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(9): 891-906
- [4] Li J, Allinson N. A comprehensive review of current local features for computer vision [J]. *Neurocomputing*, 2008, 71(10-12): 1771-1787
- [5] Scharalitzky F, Zisserman A. Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?" [C]// *Proceedings of the 7th European Conference on Computer Vision*. London, UK, 2002: 414-431
- [6] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// *Conference on Computer Vision and Pattern Recognition*. June 2005: 886-893
- [7] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987
- [8] Joachims T. Text categorization with support vector machines: learning with many relevant features [C]// *Proceedings of 10th European Conference on Machine Learning*. No. 1398, Chemnitz, DE, 1998: 137-142
- [9] Zhu L, Rao A B, Zhang A D. Theory of keyblock-based image retrieval [J]. *ACM Transactions on Information Systems*, 2002, 20: 224-257

- [10] Csurka G, Dance C R, Fan L X, et al. Visual categorization with bags of keypoints[C]// Workshop on Statistical Learning in Computer Vision. ECCV, 2004; 1-22
- [11] Vogel J, Schiele B. A semantic typicality measure for natural scene categorization[C]// Pattern Recognition Symposium. DAGM, 2004
- [12] Hofmann T. Probabilistic latent semantic analysis[C]// Proceedings of Uncertainty in Artificial Intelligence. UAI, Stockholm, 1999
- [13] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. *Machine Learning*, 2001, 42(1/2): 177-196
- [14] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. *Journal of Machine Learning Research*, 2003, 3(4/5): 993-1022
- [15] Sivic J, Russell B C, Efros A A, et al. Discovering objects and their location in images[C]// International Conference on Computer Vision. volume 1, 2005; 370-377
- [16] Li F F, Perona P. A bayesian hierarchical model for learning natural scene categories[C]// Conference on Computer Vision and Pattern Recognition, vol. 2, June 2005; 524-531
- [17] Rasiwasia N, Vasconcelos N. Scene classification with low-dimensional semantic spaces and weak supervision[C]// Conference on Computer Vision and Pattern Recognition, 2008; 1-6
- [18] Fritz M, Schiele B. Decomposition, discovery and detection of visual categories using topic models[C]// Conference on Computer Vision and Pattern Recognition, 2008; 1-8
- [19] Bosch A, Zisserman A, Munoz X. Scene classification via pls[C]// Proceedings of the European Conference on Computer Vision, 2006
- [20] Quelhas P, Monay F, Odobez J M, et al. Modeling scenes with local descriptors and latent aspects[C]// Tenth International Conference on Computer Vision, 2005; 883-890
- [21] Liu J G, Shah M. Scene modeling using co-clustering[C]// 11th International Conference on Computer Vision, 2007; 1-7
- [22] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial-pyramid matching for recognizing natural scene categories[C]// Conference on Computer Vision and Pattern Recognition, vol 2, 2006; 2169-2178
- [23] 王宇新, 郭禾, 何昌钦, 等. 用于图像场景分类的空间视觉词袋模型[J]. *计算机科学*, 2011, 38(8): 265-268
- [24] 张琳波, 王春恒, 肖柏华, 等. 基于 Bag-of-phrases 的图像表示方法[J]. *自动化学报*, 2012, 38(1): 46-54
- [25] Olshausen B A, Field D J. Sparse coding of sensory inputs[J]. *Current opinion in neurobiology*, 2004, 14(4): 481-487
- [26] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227
- [27] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]// Conference on Computer Vision and Pattern Recognition, 2009; 1794-1801
- [28] Chatfield K, Lempitsky V, Vedaldi A, et al. The devil is in the details; an evaluation of recent feature encoding methods[C]// British Machine Vision Conference, 2011
- [29] Fischler M A, Elschlager R A. The representation and matching of pictorial structures[J]. *IEEE Transactions on Computers*, 1973, C-22(1): 67-92
- [30] Burl M C, Perona P. Recognition of planar object classes[C]// Conference on Computer Vision and Pattern Recognition, 1996; 223-230
- [31] Burl M C, Weber M, Perona P. A probabilistic approach to object recognition using local photometry and global geometry[C]// Proceedings of European Conference Computer Vision, 1998
- [32] Leung T K, Burl M C, Perona P. Probabilistic affine invariants for recognition[C]// Conference on Computer Vision and Pattern Recognition, 1998; 678-684
- [33] Weber M, Welling M, Perona P. Unsupervised learning of models for recognition[C]// Proceedings of European Conference Computer Vision Dublin, Ireland, 2000
- [34] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning[C]// Conference on Computer Vision and Pattern Recognition, vol 2, 2003; 264-271
- [35] Fergus R, Perona P, Zisserman A. A sparse object category model for efficient learning and exhaustive recognition[C]// Conference on Computer Vision and Pattern Recognition, vol 1, 2005; 380-387
- [36] Li F F, Fergus R, Perona P. A bayesian approach to unsupervised one-shot learning of object categories[C]// International Conference on Computer Vision, vol. 2, 2003; 1134-1141
- [37] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples; An incremental bayesian approach tested on 101 object categories[C]// Conference on Computer Vision and Pattern Recognition Workshop, 2004; 178
- [38] Crandall D, Felzenszwalb P, Huttenlocher D. Spatial priors for part-based recognition using statistical models[C]// Conference on Computer Vision and Pattern Recognition, vol. 1, 2005; 10-17
- [39] Felzenszwalb P F, Huttenlocher D P. Efficient matching of pictorial structures[C]// Conference on Computer Vision and Pattern Recognition, vol. 2, 2000; 66-73
- [40] Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition[J]. *International Journal of Computer Vision*, 2005, 61(1): 55-79
- [41] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]// Conference on Computer Vision and Pattern Recognition, 2008; 1-8
- [42] Felzenszwalb P F. Representation and detection of deformable shapes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(2): 208-220
- [43] Felzenszwalb P F, Schwartz J D. Hierarchical matching of deformable shapes[C]// Conference on Computer Vision and Pattern Recognition, 2007; 1-8
- [44] Biederman I. Surface versus edge-based determinants of visual recognition[J]. *Cognitive Psychology*, 1988, 20(1): 38-64
- [45] De W J, Wagemans J. Contour-based object identification and segmentation; stimuli, norms and data, and software tools[J]. *Behavior Research Methods, Instruments, and Computers*, 2004, 36(4): 604-624
- [46] Huttenlocher D P, Rucklidge W J. A multi-resolution technique for comparing images using the hausdor distance[C]// Conference on Computer Vision and Pattern Recognition, 1993; 705-706
- [47] Yuille A L, Cohen D S, Hallinan P W. Feature extraction from faces using deformable templates[C]// IEEE Computer Society

- Conference on Computer Vision and Pattern Recognition, 1989; 104-109
- [48] Tomaso J, Poggio M. Model-based matching by linear combinations of prototypes [R]. MIT Series/Report no. AIM-1583, CBCL-139. December 1996
- [49] Ullman S. Three-dimensional object recognition based on the combination of views[J]. *Cognition*, 1998, 67(1/2): 21- 44
- [50] Leibe B, Leonardis A, Schiele B. Combined object categorization and segmentation with an implicit shape model [C] // ECCV workshop on statistical learning in computer vision, 2004; 17-32
- [51] Leibe B, Leonardis A, Schiele B. Robust object detection with interleaved categorization and segmentation [J]. *International Journal of Computer Vision*, 2008, 77(1): 259-289
- [52] Shotton J, Blake A, Cipolla R. Contour-based learning for object detection [C] // Tenth IEEE International Conference on Computer Vision, vol. 1, 2005; 503-510
- [53] Opelt A, Zisserman A. A boundary-fragment-model for object detection [C] // European Conference on Computer Vision, 2006; 575- 588
- [54] Opelt A, Pinz A, Zisserman A. Incremental learning of object detectors using a visual shape alphabet [C] // Conference on Computer Vision and Pattern Recognition, vol. 1, 2006; 3-10
- [55] Shotton J, Blake A, Cipolla R. Multiscale categorical object recognition using contour fragments [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(7): 1270 -1281
- [56] Leibe B, Schiele B. Interleaved object categorization and segmentation [C] // Proceedings of The British Machine Vision Conference, 2003; 759-768
- [57] Borenstein E, Ullman S. Combined top-down and bottom-up segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(12): 2109-2125
- [58] Kumar M P, Ton P H S, Zisserman A. Obj cut [C] // IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005; 18 -25
- [59] Kindermann R, Snell J L. Markov Random Fields and Their Applications. Contemporary Mathematics [M]. American Mathematical Society, Providence, Rhode Island, 1980
- [60] He X M, Zemel R S, Carreira-Perpinan M A. Multiscale conditional random fields for image labeling [C] // Proceedings of Conference on Computer Vision and Pattern Recognition, 2004
- [61] Shotton J, Winn J, Rother C, et al. Texton boost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context [J]. *International Journal of Computer Vision*, 2009, 81; 2-23
- [62] Serre T, Wolf L, Bileschi S. Robust object recognition with cortex-like mechanisms [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(3): 411-426
- [63] Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex [C] // Conference on Computer Vision and Pattern Recognition, vol. 2, 2005; 994-1000
- [64] Rutishauser U, Walther D, Koch C, et al. Is bottom-up attention useful for object recognition? [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2004
- [65] Gaussier P, Cocquerez J-P. Neural networks for complex scene recognition; simulation of a visual system with several cortical areas [C] // International Joint Conference on Neural Networks, 1992; 233-259
- [66] Riesenhuber M. Appearance isn't everything: News on object representation in cortex [J]. *Neuron*, 2007, 55(3): 341-344
- [67] Hopfinger J B, Buonocore M H, Mangun G R. The neural mechanisms of top-down attentional control [J]. *Nature Neuroscience*, 2000, 3(3): 284-291
- [68] Kincaid J M, Ollinger J M, McAvoy M P, et al. Voluntary orienting is dissociated from target detection in human posterior parietal cortex [J]. *Nature Neuroscience*, 2000, 3(3): 292-297
- [69] Itti L, Koch C. Computational modelling of visual attention [J]. *Nature reviews. Neuroscience*, 2001, 2(3): 194-203
- [70] Treisman A M, Gelade G. A feature-integration theory of attention [J]. *Cognitive Psychology*, 1980, 12(1): 97-136
- [71] Reynolds J H, Desimone R. The role of neural mechanisms of attention in solving the binding problem [J]. *Neuron*, 1999, 24(1): 19-29
- [72] Guigon E, Grandguillaume P, Otto I, et al. Neural network models of cortical functions based on the computational properties of the cerebral cortex [J]. *Journal of Physiology-Paris*, 1994, 88 (5): 291-308
- [73] Schill K, Umkehrer E, Beinlich S. Scene analysis with saccadic eye movements; top-down and bottom-up modeling [J]. *Journal of Electronic Imaging*, 2001, 10; 152-160
- [74] Deco G, Zihl J. A neurodynamical model of visual attention: Feedback enhancement of spatial resolution in a hierarchical system [J]. *Journal of Computational Neuroscience*, 2001, 10(3): 231-253
- [75] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [J]. *The Journal of physiology*, 1962, 160; 106-154
- [76] Gabor D. Theory of communication [J]. *Journal of the Institution of Electrical Engineers*, 1946, 93; 429-457
- [77] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex [J]. *Nature Neuroscience*, 1999, 2(11): 1019-1025
- [78] Perrett D I, Hietanen J K, Oram M W, et al. Organization and functions of cells responsive to faces in the temporal cortex [J]. *Philosophical Transactions: Biological Sciences*, 1992, 335 (1273); 23-30
- [79] Hung C P, Kreiman G, Poggio T. Fast readout of object identity from macaque inferior temporal cortex [J]. *Science*, 2005, 310 (5749); 863-866
- [80] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. *Nature*, 1996, 381(6583): 607-609
- [81] Mutch J, Lowe D. Object class recognition and localization using sparse features with limited receptive fields [J]. *International Journal of Computer Vision*, 2008, 80(1): 45-57
- [82] Poggio T B, Stanley M. A multi-scale generalization of the hog and hmax image descriptors for object detection [R]. MIT-CSAIL-TR-2008-019, CBCL-271. Massachusetts Institute of Technology, Apr. 2008
- [83] Bullier J. Integrated model of visual processing [J]. *Brain research. Brain research reviews*, 2001, 36(2/3): 96-107
- [84] Hochstein S, Ahissar M, Neurobiology D O. View from the top: Hierarchies and reverse hierarchies in the visual system [J]. *Neuron*, 2002, 36; 791-804
- [85] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(4): 509-522

用 TDMA 小了 1 到 3 个数量级。随着网络业务量不断增大, MCPS 协议延时几乎没有变化, 维持在 3.44 ms。但 TDMA 协议在业务量到达一定程度时, 延时突然激增, 网络性能明显下降。

图 8 中随着网络的业务量增加, MCPS 协议网络的丢包率略有增加, 满足 MCPS 协议网络中丢包率较小的设计要求。对于 TDMA 协议, 当网络较空闲时, 丢包率为 0; 当网络业务量达到一定程度, TDMA 网络的丢包率急剧增加, 甚至达到了 72.56%, 网络已无法继续工作。

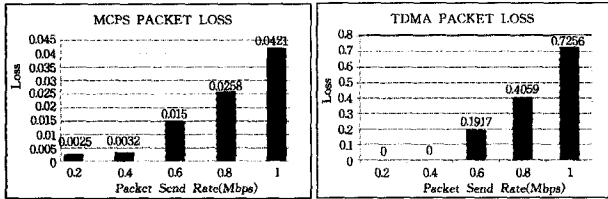


图 8 MCPS(上)vs. TDMA 协议丢包率比较

从图 9 可看出, TDMA 网络能够承受的网络业务量远远小于 MCPS 网络的承载业务量。这也是导致 TDMA 协议网络在网络业务量较大时, 端到端延时和丢包率急剧上升的原因。

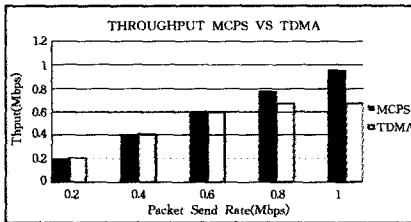


图 9 MCPS(左)vs. TDMA 协议吞吐量比较

结束语 本文通过 MCPS 协议的理论设计、协议实现以及协议仿真验证, 验证了 MCPS 协议的性能。无论是在不同网络规模下, 还是在同等网络规模不同业务量配置的场景下, 基于 MCPS 协议的数据链网络在业务端到端时延、网络的丢包率以及网络的承载业务量等方面远远优于基于 TDMA 的数据链网络, 能够满足新型战术数据链设计要求。

参考文献

- [1] 童志鹏, 刘兴. 综合电子信息系统(第二版)[M]. 北京: 国防工业出版社, 2008
- [2] 孙义明, 杨丽萍. 信息化战争中的战术数据链[M]. 北京: 北京邮电大学出版社, 2005
- [3] 陈志辉, 李大双. 对美军下一代数据链 TTNT 技术的分析与探讨[J]. 通信技术, 2011, 5: 76-79
- [4] 俞胜男. 无线动态 TDMA 网络的时隙分配技术研究[D]. 西安: 西安电子科技大学, 2009
- [5] 王宝康, 陈强. 一种改进的 Ad Hoc 网络中动态 TDMA 时隙分配方法[J]. 电子世界, 2011(14)
- [6] Hadid N, Guitton A. Michel Mission, Adaptive Slotted CSMA/CA Algorithm for The Traffic Accumulated During The Inactive Period[J]. PE-WASUM 09 Proceedings of the 6th ACM Symposium on Performance Evaluation of Wireless Ad hoc, Sensor, and Ubiquitous Networks. 2009; 143-146
- [7] Wang Xin. Mobile Ad-Hoc Network; Applications [M]. Tech, 2011-01
- [8] QualNet 5. 0 Programmer's Guide[M]. Scalable Network Technologies, Inc

(上接第 8 页)

- [86] Torralba A, Murphy K P, Freeman W T. Contextual models for object detection using boosted random fields[C]// Advances in Neural Information Processing Systems, NIPS. 2005: 1401-1408
- [87] Biederman I, Mezzanotte R J, Rabinowitz J C. Scene perception: Detecting and judging objects undergoing relational violations [J]. Cognitive Psychology, 1982, 14(2): 143-177
- [88] Cox D, Meyers E, Sinha P. Contextually evoked objectspecific responses in human visual cortex[J]. Science, 2004, 304(5667): 115-117
- [89] Boshyan J, Fenske M, Gronau N, et al. The contribution of context to visual object recognition[J]. Journal of Vision, 2005, 5(8)
- [90] Murphy K, Torralba A, Freeman W T. Using the Forest to See the Trees; A Graphical Model For Recognizing Scenes and Objects[M]// Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, eds. Advances in Neural Information Processing Systems. Cambridge, MA; MIT Press, 2004
- [91] Carbonetto P, de Freitas N, Barnard K. A statistical model for general contextual object recognition[C]// Proceedings of the European Conference on Computer Vision. 2004
- [92] Singhal A, Luo J, Zhu W Y. Probabilistic spatial context models for scene content understanding[C]// IEEE Conference on Com-

- puter Vision and Pattern Recognition. vol. 1, 2003: 235-241
- [93] Torralba A. Contextual influences on saliency[Z]. Neurobiology of attention. 2005: 586-592
- [94] Wolf L, Bileschi S. A critical view of context[J]. International Journal of Computer Vision, 2006, 69(2): 251-261
- [95] Kumar S, Hebert M. Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification [C]// Proceedings of the 9th International Conference on Computer Vision. Washington, DC, USA, 2003: 1150
- [96] From appearance to context-based recognition; Dense labeling in small images[DB]. June 2008
- [97] Tu Z W. Auto-context and its application to high-level vision tasks[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8
- [98] He X M, Zemel R S. Latent topic random fields; Learning using a taxonomy of labels[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8
- [99] Torralba A. Contextual priming for object detection[J]. International Journal of Computer Vision, 2003, 53(2): 169-191
- [100] Mei T, Wang Y, Hua X S, et al. Coherent image annotation by learning semantic distance[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8