

面向目标检测的稀疏表示方法研究进展

高仕博¹,程咏梅²,肖利平^{1,3},韦海萍^{1,3}

(1.北京航天自动控制研究所,北京 100854;2.西北工业大学自动化学院,陕西西安 710072;
3.宇航智能控制技术国家级重点实验室,北京 100854)

摘 要: 目标检测作为图像理解的一个基础而重要的课题深受国内外学者的重视,在军事和民用中具有广泛应用.应用背景的多样性和复杂性使得传统目标检测算法难以克服复杂背景、噪声干扰、光照变化以及非刚体形变、遮挡、弱特征、尺度、视角和姿态变化等因素的影响.近些年来发展起来的稀疏表示方法为图像处理及目标检测研究提供了新的思路,本文概述了稀疏表示基本概念和理论研究进展,综述了稀疏表示方法在目标特征学习、目标分类器和滤波器设计以及多源信息融合目标检测等目标检测领域中的国内外重要研究进展,并展望了稀疏表示方法在目标检测领域的发展方向.

关键词: 目标检测;图像处理;稀疏表示;特征

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112 (2015)02-0320-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2015.02.018

Recent Advances of Sparse Representation for Object Detection

GAO Shi-bo¹, CHENG Yong-mei², XIAO Li-ping^{1,3}, WEI Hai-ping^{1,3}

(1. Beijing Aerospace Automatic Control Institute, Beijing 100854, China;

2. College of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China;

3. National Key Laboratory of Science and Technology on Aerospace Intelligent Control, Beijing 100854, China)

Abstract: Object detection is a basic and important subject in image understanding, which has attracted much attention from domestic and foreign scholars. Object detection has been widely used in military and civilian. The diversity and complexity of applications makes the traditional detection technique be affected by many factors such as complex background, noise, illumination variations, non-rigid deformation, occlusion, feeble features, scale, visual angle attitude and, etc. Recently, the developing method of sparse representation provides a novel research approach for image processing and objects detection. This paper overviews the basic concept of sparse representation and its recent progress in the theoretical study. The domestic and foreign research advances of sparse representation in object detection are summarized, especially in object feature learning, classifier and filter designing, multisource fusion detection. Meanwhile, some future directions of sparse representation in object detection are also addressed.

Key words: object detection; image processing; sparse representation; features

1 引言

随着成像传感器技术的发展,人类扩展了获取图像信息的广度和深度,加深了人类对客观世界的认识,能观察到人眼能感知到和感知不到的物体,根据所用传感器的不同,常见的图像有彩色图像、红外图像、高光谱图像、合成孔径雷达图像及核磁共振图像等.对于获取的大量图像信息,人们期望借助计算机实现智能化处理,达到对场景的自动分析和理解.目标检测的任务是从获取的图像中分割出感兴趣的区域,作为图像理解的一个

基础而重要的问题深受国内外学者的重视,在目标侦查、精确制导、智能监控、视觉导航、人机交互、空间遥感及医疗辅助诊断等方面具有广泛应用.

在实际应用中,利用计算机图像处理技术完成对自然界庞杂物体的自动检测是一项挑战性任务,复杂背景、噪声干扰、光照变化以及非刚体形变、遮挡、弱特征、尺度、视角和姿态变化等因素严重影响目标的检测性能,传统的目标检测方法并不能达到理想的检测效果,为此人们结合新理论提出了多种方法以期能在实际应用中提高目标的检测性能.其中稀疏表示方法以符合人

类视觉系统的描述方式能有效捕捉图像的结构特征,可实现图像的有效表示,近几年的重要会议和期刊,如 ICCV、CVPR、ECCV、NIPS、PAMI 和 IJCV 等都对稀疏表示做了很多报道,基于稀疏表示的方法在许多领域均取得了当前最好的结果,例如图像恢复、特征提取及识别等,展现了稀疏表示方法的优势,理论上为目标检测提供了一种独特和非常有效的处理方法.实际应用中目标由于成像机理的不同,目标多种多样,如何将稀疏表示与应用背景相结合进而克服传统目标检测算法的不足是一个值得思考的问题,本文在简要介绍稀疏表示基本概念和理论的基础上,对稀疏表示在目标检测中的研究思路进行梳理,总结取得的重要进展,展望未来发展方向.

2 稀疏表示基本概念和理论研究进展

2.1 稀疏表示基本概念

大量研究表明视觉皮层复杂刺激的表达采用的是稀疏编码原则^[1-4],以稀疏编码为基础的稀疏表示方法能较好刻画人类视觉系统对图像的认知特性,已引起人们极大的兴趣和关注,在机器学习和图像处理领域得到了广泛应用,是当前国内外的研究热点之一.图像稀疏表示研究主要沿着两条线展开^[5]:单一基方法和多基方法.前者主要是多尺度几何分析理论^[6],认为图像具有非平稳性和非高斯性,用线性算法很难处理,应建立适合处理边缘及纹理各层面几何结构的图像模型,以脊波(Ridgelet)^[7]、曲波(Curvelet)^[8]等变换为代表的多尺度几何分析方法成为图像稀疏表示的有效途径;后者以 Mallat 和 Zhang 提出的过完备字典分解理论为基础^[9],根据信号本身的特点自适应选取能够稀疏表示信号的冗余基.鉴于后者良好的图像描述能力,人们最近以此为基础做了大量有关图像稀疏表示的工作.稀疏表示的基础是稀疏编码,所谓稀疏编码就是在过完备基向量中选择少数基向量的线性组合来表示样本数据,即:

$$\mathbf{x} = \sum_{k=1}^K a_k \mathbf{d}_k = \mathbf{D}\mathbf{a}, \text{ s.t. } \|\mathbf{a}\|_0 \leq s \quad (1)$$

其中 $\mathbf{x} \in R^{m \times 1}$ 为样本向量, $\mathbf{D} \in R^{m \times K}$ 称为字典 ($m < K$, 即过完备), \mathbf{d}_k 表示字典 \mathbf{D} 中的基向量,称为原子,通常满足 $\|\mathbf{d}_k\|_F = 1$,式中 $\|\cdot\|_F$ 为 F-范数,对 \mathbf{d}_k 进行归一化约束是防止 \mathbf{d}_k 无限大或无限小造成对应系数 a_k 的不唯一性, $\mathbf{a} = [a_1, \dots, a_K]^T \in R^{K \times 1}$ 为稀疏系数向量,稀疏性是对 \mathbf{a} 而言,使得 \mathbf{a} 中的非零元素个数少,也就是 \mathbf{a} 的 l_0 范数 $\|\mathbf{a}\|_0$ 尽量小(不超过 s),同时还能较好表示 \mathbf{x} .如果去掉稀疏性约束条件 $\|\mathbf{a}\|_0 < s$,上式是欠定线性方程, \mathbf{a} 具有多组解,Donoho 等^[10]证明如果矩阵 \mathbf{D} 满足如下条件,则上述 l_0 范数优化问题具有唯一解:

$$\|\mathbf{a}\|_0 \leq \text{spark}(\mathbf{D})/2 \quad (2)$$

其中 $\text{spark}(\mathbf{D})$ 定义为矩阵 \mathbf{D} 的最小线性相关的列数.但求解 l_0 范数是 NP 难的,Tao 等^[11]证明在受限正交条件(Restricted Isometry Property, RIP)下, l_0 范数优化问题与 l_1 范数优化问题具有相同的解,即:

$$\arg \min_{\mathbf{a}} \|\mathbf{a}\|_1 = \arg \min_{\mathbf{a}} \|\mathbf{a}\|_0, \text{ s.t. } \mathbf{x} = \mathbf{D}\mathbf{a} \quad (3)$$

其中 RIP 条件是:如果对所有 s 稀疏的向量 \mathbf{a} ,满足 $(1 - \delta_s) \|\mathbf{a}\|_2^2 \leq \|\mathbf{D}\mathbf{a}\|_2^2 \leq (1 + \delta_s) \|\mathbf{a}\|_2^2$ 的最小常数 δ_s 小于 1,则称矩阵 \mathbf{D} 满足 RIP 性质,RIP 条件是对于矩阵 \mathbf{D} 列向量正交性的一种衡量, l_1 范数定义为 $\|\mathbf{a}\|_1 = \sum_{k=1}^K |a_k|$.此时 l_1 范数优化是一个凸优化问题,有唯一解,其唯一解即为 l_0 范数优化问题的唯一解.考虑到数据中存在的噪声,松弛等式约束限制条件,上述优化问题可表示为:

$$\arg \min_{\mathbf{a}} \|\mathbf{a}\|_1, \text{ s.t. } \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_F^2 \leq \epsilon \quad (4)$$

上式是一个标准的稀疏编码问题,根据应用目的的不同,其有多种变体,如:

$$\begin{cases} \arg \min_{\mathbf{a}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_F^2, \text{ s.t. } \|\mathbf{a}\|_1 \leq \epsilon \\ \arg \min_{\mathbf{a}} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_F^2 + \lambda \|\mathbf{a}\|_1 \end{cases} \quad (5)$$

其中 λ 表示非负稀疏性约束参数,式(5)的第二个式子就是著名的 Lasso 模型(Least Absolute Shrinkage and Selection Operator).除了求解 \mathbf{a} 的稀疏编码问题外,人们更感兴趣的是如何选择 \mathbf{D} 使得信号 \mathbf{x} 在 \mathbf{D} 上的表示是稀疏的和有效的,称为字典设计或学习问题,Olshausen 和 Field^[12]最早将稀疏性约束 l_1 范数作为最小二乘算法的规则项,从训练样本中学习到具有方向特性的原子,总体而言,稀疏编码和字典学习的基本定义可以表示如下:

$$\arg \min_{\mathbf{D}, \{\mathbf{a}_i\}} \sum_i \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_F^2 + \lambda \|\mathbf{a}_i\|_1, \text{ s.t. } \|\mathbf{d}_k\|_F^2 = 1 \quad (6)$$

通过上式学习到的过完备字典能更有效地揭示隐含在输入数据内部的结构与模式.上式对 \mathbf{D} 、 $\{\mathbf{a}_i\}$ 两个变量来说不是凸优化问题,但固定其中一项对另一项来说则是凸优化问题,一般采用交替优化的方法进行求解,即:初始化 \mathbf{D} ,然后固定 \mathbf{D} 求解 $\{\mathbf{a}_i\}$,再固定 $\{\mathbf{a}_i\}$ 更新 \mathbf{D} ,交替迭代完成优化.围绕稀疏编码和字典学习,人们提出了大量优化算法,关注的主要是计算速度和计算精度,常用的稀疏编码算法有正交匹配追踪算法(Orthogonal Matching Pursuit, OMP)^[13]、特征符号搜索算法(Feature-sign Search)^[14]等;字典学习方法有 K-SVD 算法^[15]、在线字典学习算法(SPAMS)^[16]等.

2.2 稀疏表示理论研究进展

鉴于稀疏描述在理论上和应用上的巨大优势,在上述稀疏编码概念的基础上,稀疏表示理论得到了快

速发展,目前受到广泛关注的主要有以下 2 个方面:

(1)结构稀疏性:上述稀疏编码方法只考虑了表示向量的整体稀疏性,由于稀疏系数通常存在某种结构属性,人们又引入解释性更强的结构稀疏性,近几年来常见的有组稀疏性(Group Sparsity)和联合稀疏性(Joint Sparsity).对于稀疏系数向量 $\mathbf{a} = [a_1, a_2, \dots, a_K] \in R^{K \times 1}$,其中的指标 $[1, 2, \dots, K]$ 可以划分为不重叠的 H 个组,即 $[1, 2, \dots, K] = \bigcup_{h=1}^H G_h, G_h \cap G_{l \neq h} = \emptyset$,如图 1 所示,其中每个节点表示 \mathbf{a} 中的一个元素,黑色点表示非零元素,灰色点表示零元素, \mathbf{a} 中的全部元素划分为 7 个组,组之间没有重叠的元素,这种表示形式能有效描述元素之间存在的相关性,Huang 等^[17]将组稀疏性定义为:

$$\arg \min_a \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_F^2 + \lambda \sum_{h=1}^H \|a_{G_h}\|_2 \quad (7)$$

上式中的第二项表示 \mathbf{a} 的 $l_{2,1}$ 范数,这里的稀疏性是指 \mathbf{a} 中非零组的数量,如果每组的元素数量都为 1,式(7)就是标准稀疏性,上式也称为组 Lasso 模型,在此基础上,又相继发展了重叠组稀疏性^[18]、块稀疏性^[19]等.组稀疏性优化算法有块坐标下降算法(Block-Coordinate Descent Algorithm)^[20]等.

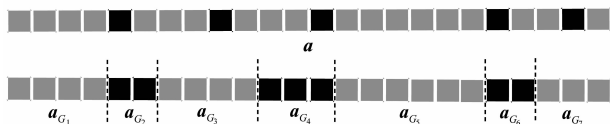


图1 标准稀疏向量(第一行)和组稀疏向量(第二行)

不同于组稀疏性,联合稀疏性则是对多个稀疏向量进行定义,设某一目标具有多个特征或属性 $\{\mathbf{x}_j\}$,每个特征或属性都有对应的字典 $\{\mathbf{D}_j\}$,该目标的多个特征或属性对各自字典的稀疏系数向量为 $\{\mathbf{a}_j\}$,联合稀疏性通常是指这些稀疏系数向量中的非零元素的位置保持一致(元素值大小可能不同),如图 2 所示,设稀疏系数向量组成的矩阵 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_J]$,联合稀疏性定义为^[21]:

$$\arg \min_A \frac{1}{2} \sum_j \|\mathbf{x}_j - \mathbf{D}_j \mathbf{a}_j\|_F^2 + \lambda \sum_i \|A^i\|_2 \quad (8)$$

其中 A^i 表示矩阵 \mathbf{A} 的第 i 行向量,联合稀疏性对多特征^[21]、多视角^[22]及多源图像^[23]的分类和识别具有重要意义.

将组稀疏性、联合稀疏性融入到字典学习模型中,相应的可以学习到结构化字典^[24];此外,在实际应用中,针对具体任务的特性,有多种结构化字典学习模型,比如:多字典^[25,26]、有监督字典^[27,28]、非负字典^[29]等,大量研究表明采用结构化字典学习方法能够有效提高任务处理的效果.

(2)矩阵稀疏性:稀疏编码问题是对一维向量进行

稀疏性约束,Candès 等^[30]基于图像的低秩特性把稀疏性约束从一维向量推广到二维矩阵,在此基础上提出了矩阵填充(Matrix Completion)的概念,低秩矩阵填充问题是利用不完整的采样数据 \mathbf{X} 来恢复低秩矩阵^[31],即:

$$\arg \min_L \text{rank}(\mathbf{L}), \text{ s.t. } \mathbf{L}_{mn} = \mathbf{X}_{mn}, (m, n) \in \Omega \quad (9)$$

其中 \mathbf{L} 是需要恢复的低秩矩阵, $\text{rank}(\cdot)$ 是矩阵秩函数,定义为矩阵非零奇异值的个数, Ω 是观测到的元素位置集合,矩阵 \mathbf{X} 位于 Ω 中的元素是已知的,其它元素缺失,上式就是要通过低秩约束恢复出其中未知的缺失元素.由于 $\text{rank}(\cdot)$ 是非凸函数,Candès 等^[30]证明矩阵秩优化问题在很多情况下可以等价于矩阵迹范数 $\|\cdot\|_*$ 优化问题,即:

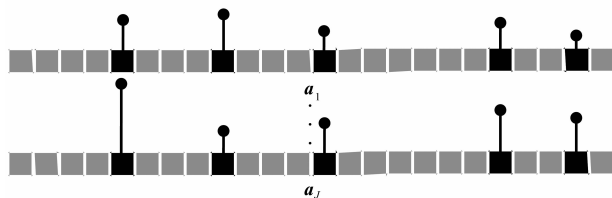


图2 联合稀疏性(线段长短表示非零元素值大小)

$$\arg \min_L \|\mathbf{L}\|_*, \text{ s.t. } \mathbf{L}_{mn} = \mathbf{X}_{mn}, (m, n) \in \Omega \quad (10)$$

其中矩阵迹范数 $\|\cdot\|_*$ 定义为矩阵奇异值之和,对比一维向量的 l_0 范数与 l_1 范数, $\text{rank}(\cdot)$ 相当于矩阵的 l_0 范数,而 $\|\cdot\|_*$ 相当于矩阵的 l_1 范数,优化上述问题常用方法有奇异值阈值算法(Singular Value Thresholding, SVT)^[32]、增广拉格朗日乘子算法(Augmented Lagrange Multiplier, AML)^[33]等.最近,矩阵填充思想在图像处理领域受到了大量关注,如:图像修复^[34]、分类^[35]等.在此基础上,Ma 等^[36]认为重构误差相对于总体样本而言是稀疏的,进一步给出下式:

$$\arg \min_{L,S} \text{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0, \text{ s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S} \quad (11)$$

其中 λ 是非负权重参数.上式就是著名的鲁棒主成分分析(RPCA)^[36,37],也称为矩阵的低秩和稀疏分解模型,是当前研究热点之一,在实际应用中,通常用矩阵迹范数 $\|\cdot\|_*$ 取代矩阵秩 $\text{rank}(\cdot)$,用 l_1 范数取代 l_0 范数,在优化过程中采用交替迭代方法交替更新低秩矩阵 \mathbf{L} 和稀疏矩阵 \mathbf{S} ,常用优化算法有 SVT 等.目前,鲁棒 PCA 方法得到了广泛应用,比如阴影去除^[37]、图像对齐^[38]、图像分割^[39]、运动目标检测^[40,41]以及显著性目标检测^[42,43]等.

近年来,稀疏表示理论得到了快速发展,上面讨论的稀疏表示方法广义上讲都是基于矩阵分解的概念,当前几种常见的矩阵分解形式如下表 1 所示.在应用方面,国内外学者结合具体任务开展了许多研究工作,并取得了显著的成果,后面将进一步讨论稀疏表示在目

标检测中的研究现状.

表 1 当前几种常见的矩阵分解形式

矩阵分解名称	表示形式 (矩阵 X 已知,其它矩阵未知)	主要应用领域
字典学习 (DL)	$X = DA$ 其中 A 稀疏约束	图像恢复、特征提取、识别
非负矩阵分解 (NMF)	$X = DA$ 其中 D 和 A 的元素非负	聚类、分类
稀疏 PCA (SPCA)	$X = DA$ 其中 D 稀疏约束	降维、特征选择
矩阵填充 (MC)	$P_{\Omega}X = P_{\Omega}L$ 其中 P_{Ω} 表示在已知子集 Ω 上的投影, L 低秩约束	图像修复
鲁棒 PCA(RPCA)	$X = L + S$ 其中 L 低秩约束, S 稀疏约束	图像对齐、图像恢复、检测、分类
噪声鲁棒 PCA	$X = L + S + N$ 其中 L 低秩约束, S 受稀疏约束, N 受 F 范数约束	复杂背景建模、检测

3 基于稀疏表示的目标检测研究进展

在实际应用中,目标多种多样,在目标检测过程中通常会面临复杂背景、噪声干扰、光照变化以及非刚体形变、遮挡、弱特征、尺度、视角和姿态变化等因素的影响,严重降低目标的检测性能.根据不同的目标检测任务国内外学者提出了许多研究思路和方法解决上述难题,如:模糊技术、机器学习、分形理论、数学形态学、遗传算法等检测算法,从实际应用的效果来看,还不能完全解决复杂场景下的目标检测问题,需要进一步研究有效的目标检测方法.一般来说,目标检测研究大体上可划分为两个部分:(1)特征提取,获取能够表示目标的有效特征;(2)分类器(或检测器)设计,构建分类器或者检测器判断符合某种规则的特征是否是目标.这其中又涉及到去噪、分割、降维、分类、跟踪、融合等技术,是一个很庞大的体系,各种技术的发展为目标检测提供了诸多研究思路.稀疏表示是近些年发展起来的一种有潜力的图像信息表示方法,稀疏表示是一种包容性强的数据描述方法,可以灵活设计字典等参数,而且稀疏性有利于突出目标特征,使其与背景更加线性可分,为目标检测研究提供了新的思路.人们在稀疏表示框架下针对具体的目标检测任务取得了一些成果,其成功应用的关键是建立合理的检测模型,下面着重从稀疏表示在目标特征学习、分类器和滤波器设计以及多源信息融合检测等目标检测领域中取得的重要研究进展进行讨论.

3.1 基于稀疏表示的目标特征学习

提取具有鲁棒性和可区分性的特征描述目标是目

标检测中的一个核心问题,是当前国内外目标检测研究的重点.近年来的一个研究热点是在尺度不变特征变换描述子(SIFT)^[44]等底层局部特征基础上建立起来的词袋特征(BOW)^[45,46],可以将目标的底层局部特征转化为能够描述目标整体的中层特征,对目标的尺度变化、旋转等具有较好的鲁棒性.构造过程通常是先对训练样本的局部特征描述子进行聚类,得到的聚类中心形成一个码本(Codebook),统计各聚类中心中描述子出现的频率,得到的频率直方图称为词袋,比如常用的纹理基元(Textons)就属于 BOW 的一种,Textons 所用的底层局部特征描述子是滤波器组响应向量^[47],主要用于描述纹理特征和检测纹理目标.构造 BOW 时,常用的聚类算法有硬分配算法(K 均值)、软分配算法^[48]等.实际应用中,需要根据不同的数据和任务学习更为鲁棒和可区分的 BOW 特征,而稀疏表示方法本身就是一种包容性强的数据描述方法,将其用于 BOW 特征学习的思路受到广泛关注. Yang 等^[49]采用稀疏编码方法扩展传统 K 均值聚类算法,用于构造 BOW 特征,将码本作为需要学习的字典,把提取的 SIFT 描述子作为训练样本去学习字典和稀疏系数向量,字典中的原子对应于聚类中心,稀疏系数向量中的非零元素指示原子(等价于聚类中心)是否被激活,元素值大小反映了描述子对原子的归属度,并考虑了平均 Pooling 和最大 Pooling 两种构造方法,其中平均 Pooling 方法是统计所有描述子对每个聚类中心归属的平均值,也就是对所有训练样本的稀疏系数向量进行平均获得 BOW 特征(硬分配中归属度为 1 或 0,平均 Pooling 等同于统计各聚类中心中特征描述子出现的频率),而最大 Pooling 方法则是统计所有描述子对每个聚类中心归属的最大绝对值,也就是求取所有稀疏系数向量每行的最大绝对值作为 BOW 特征,在很多情况下最大 Pooling 要比平均 Pooling 的目标检测识别效果更好^[50].在此基础上, Yang 等^[51]将稀疏系数向量的最大 Pooling 作为特征,以分类损失误差为准则对字典进行有监督的学习用于构造 BOW 特征,较无监督方法取得了更好的检测识别效果. Boureau 等^[52]则是将稀疏系数向量的平均 Pooling 作为特征,以分类损失误差为准则学习判别性字典进而构造 BOW 特征,也取得了比无监督方法好的效果. Wang 等^[53]考虑到图像中局部特征之间的位置关系,在构造字典的过程中进一步加入了空间局部性约束,使得相邻描述子对应的稀疏系数向量中的非零元素位置应尽量一致,也就是说相邻描述子应选择相似的原子,在该思路下去学习字典进而构造 BOW 特征,较 Yang 等^[49]所提方法效果更好. Lei 等^[54]在 SIFT 基础上通过构造视觉概念网络学习组字典,获得了更具判别性的 BOW. Naikal 等^[55]考虑到背景和无用的底层特征对识别

造成的不利影响,采用稀疏主成分分析方法 (SPCA) 选取具有判别性信息的底层特征,有利于后续的处理任务.但 SIFT、梯度方向直方图 (HOG) 等特征只能描述关键点、边缘等信息,对受复杂干扰的图像来说,这些特征不一定有效.一种更为理想的思路是从数据中学习解释性更强的目标特征用于检测,如:Ren 等^[56]基于稀疏表示方法学习的稀疏编码直方图特征 (HSC),在目标检测中明显优于 HOG 特征.当前常用的基于稀疏编码和字典学习的方法^[16],是对图像块进行建模,可以学习出类似 Gabor 滤波器的原子,仅能捕捉不同方向的边缘信息等,而且巨大的计算量使其只能处理较小的图像块.最近,Rigamonti 等^[57,58]、Kavukcuoglu 等^[59]将卷积方法与稀疏编码思想相结合相继提出了卷积稀疏编码模型,将稀疏编码模型中的字典看成是卷积滤波器组,对图像整体进行稀疏编码,代价函数表示为:

$$\arg \min_{\{f_m, \alpha_i\}} \sum_i \|I_i - \sum_{m=1}^M f_m * \alpha_i^m\|_F^2 + \lambda \sum_{m=1}^M \|\alpha_i^m\|_1, \text{ s.t. } \|f_m\|_F = 1 \quad (12)$$

其中 $I_i \in R^{L_1 \times L_2}$ 表示第 i 个图像, $f_m \in R^{(2H+1) \times (2H+1)}$ 表示第 m 个滤波器,每个滤波器 f_m 的 F 范数为 1, $\alpha_i^m \in R^{(L_1+2H) \times (L_2+2H)}$ 可以称为图像 I_i 的第 m 个隐含特征图 (Feature map), 对 α_i^m 的 l_1 范数约束 $\|\alpha_i^m\|_1 = \sum_{x,y} |(\alpha_i^m)_{x,y}|$ 可以使上式产生唯一解,并促使特征图是稀疏的, $\lambda \geq 0$ 是权重参数,通过随机梯度下降算法优化上式.相对于只能对较小图像块进行处理的稀疏编码算法,卷积稀疏编码模型更有优势,能有效处理图像整体,而且可以捕捉角点 (Corner)、中心-周围 (Center-surround) 等图像中固有的多种模态,在目标检测领域中具有广泛的应用前景,比如可以用于设计滤波器组响应向量进而构造 Texton.

如上所述, SIFT、HOG 等特征只能描述关键点、边缘信息等,难以整合原始图像信息描述对象、场景等,会造成特征信息提取的不完备;除了算法本身的不足外, SIFT 等还需要巧妙的构造思路,并且在不同场景下对不同具体数据所起的作用不同,适应性较差,需要较多的人工干预,增加了特征的不确定性,难以满足实际应用中目标检测方法的鲁棒性要求,在检测性能上难以保证满意的检测效果.最近在人工神经网络基础上发展的深度学习 (Deep Learning) 方法可以直接从数据中无监督地自动学习层次式的特征表示,是机器学习研究中的一个新领域和热点,希望通过模仿人脑的多层抽象机制来实现对数据 (图像等) 的抽象表达和解释,将特征学习和分类器整合到了一个学习框架中,减少人工设计特征的工作,为目标检测方法的研究提供了理论和应用上的可能性.深度学习的实质是通过构建

具有很多隐层的机器学习模型和海量的训练数据,学习更有用的特征,从而最终提升分类或预测的准确性^[60].深度学习中的几种常用模型包括: Hinton 等^[61]提出的自动编码器 (AutoEncoder)、Bengio 等^[62]提出的限制波尔兹曼机 (Restricted Boltzmann Machine)、Hinton 等^[63]提出的深度信念网络 (Deep Belief Networks) 等.最近几年,稀疏表示越来越多的融入到深度学习发展中, Lecun 等^[64]提出的卷积神经网络模型,就是通过稀疏编码得到的稀疏系数向量训练非线性前向反馈预测器,完成无监督特征学习,是一个多层结构学习算法,并初步应用于行人检测问题中^[65];在此基础上, Zeiler 等^[66]借助定义的稀疏潜在特征图提出的解卷积网络深度学习模型,也是当前一个重要的深度学习模型.卷积神经网络和去卷积网络深度学习模型是基于最小化预处理数据要求而产生的,将图像的局部作为分层结构的最底层输入,通过不同网络层次传递信息,在每一层能够获取对平移、缩放和旋转不变的观测数据的显著特征.对于具有潜在复杂结构规则的图像,深度学习算法先从原始图像去学习得到它的一个低层次表示,然后通过线性或者非线性运算组合这些低层次表示形成更加抽象的高层次表示 (属性类别或特征),重复这个过程,可以提取数据中潜在 (隐含) 分布的特征表示,通过深度学习获取的本质特征非常适合解决旋转、几何形变及尺度变化的目标检测问题.当前多伦多大学、斯坦福大学、纽约大学以及 Google、微软、百度等机构正积极开展这方面的研究,目前深度学习在大规模图像搜索、物体识别、语音识别等领域中取得了不错的效果.但深度学习方法的一个重要不足^[67]是深度学习中需要较多的经验、人工尝试及技巧,如:网络层数、每层节点数、节点的激发函数等, Ngiam 等^[68]提出了一个值得参考的思想:稀疏滤波方法,需要学习的只有特征数目一个参数,该方法的核心思想是认为每个样本特征是稀疏的、样本间的特征是稀疏的以及特征的分布是均匀的等 3 个描述特征分布的准则,设计了一个简单的 l_2 范数稀疏约束的特征学习代价函数,避免了对数据分布的显式建模,通过优化特征分布的稀疏性得到好的特征表达,能有效处理高维输入,并可扩展为多层结构.

另外,将稀疏表示方法和核方法、多线性描述等方法相结合,能学习出复杂样本数据的目标内在特征,在目标特征学习中具有重要应用,限于篇幅不再展开介绍.

3.2 基于稀疏表示的目标分类器和检测器设计

目标检测时需要分类器或检测器区分目标和背景,是目标检测过程中的重要一步,实际应用中噪声、遮挡等干扰的存在使得诸如支持向量机 (SVM) 等传统分类器或检测器面临着极大的困难.随着稀疏表示研

究的深入,人们发现稀疏表示方法可以作为一种良好的分类器,用于对目标进行检测识别,基于稀疏表示构造的分类器通常采用重构误差作为分类准则,主要有 2 种思路:

(1)对每类样本单独训练一个字典,将测试样本在相同稀疏性约束下对各类别的字典分别进行稀疏编码,根据测试样本对各类别字典的重构误差大小进行分类,Mairal 等^[25]以局部图像块为样本,在字典学习过程中加入了判别性重构误差约束项训练出一组判别性字典,使得每类样本对该类别字典的重构误差较小,对其它类别字典的重构误差较大,将测试样本对学习到的判别性类别字典分别进行稀疏编码,根据测试样本对各类别字典的重构误差大小进行判断,较好实现了边缘、纹理和场景中物体(自行车)等目标的分类和检测。

(2)Wright 等^[69]提出的稀疏表示分类器(SRC),其思想是把所有类别的训练样本直接组合成为一个字典,将测试样本在该字典上进行稀疏编码获得稀疏系数向量,分别只保留稀疏系数向量中对应某一类别训练样本的元素,其余设为 0,得到测试样本对该类别的稀疏系数向量,再与原字典进行重构,计算该类别重构项与测试样本的重构误差,依此计算所有类别对测试样本的重构误差,比较重构误差大小进行分类, SRC 对光照和表情变化及噪声、遮挡等情况下的人脸检测识别具有很好的鲁棒性。

从当前研究现状来看,稀疏表示分类器是一个研究热点,稀疏表示分类器的潜在含义是:同类样本处于同一线性子空间中,每类训练样本的数量要足够多,能够张成该类样本的子空间,使得测试样本能在该类训练样本上进行线性表示,这样测试样本在所有训练样本所构成的字典上的表示是稀疏的,对应的稀疏系数向量也包含了样本的类别信息。因此,稀疏表示分类器要求样本图像已经对齐,而且同类训练样本的数量充分多,能张成该类样本子空间。Wagner 等^[70]引入二维参数变换实现图像对齐,提出了一种稀疏表示鲁棒对齐算法(RASR),通过优化对齐误差稀疏性函数求解其中的变换参数,实现稀疏表示分类器中对样本图像对齐的要求。Yang 等^[71]考虑到对齐图像的低秩特性,在对齐误差稀疏性函数中引入低秩约束对变换参数进行由粗到细的估计,极大改进了 RASR 算法的实时性。He 等^[72]采用最大相关熵准则(Maximum Correntropy Criterion)赋予人脸中各像素以不同的权重,对其中遮挡或损毁的区域赋予小权重,降低这部分区域在检测识别过程中的作用,更利于处理遮挡等复杂条件下的人脸检测识别,提高了稀疏表示分类器的鲁棒性。为满足 SRC 方法中训练样本能张成该类样本子空间的要求,需要同类

训练样本的数量充分多,训练样本数量的增加,造成训练样本字典增大,增加了测试样本在字典上进行稀疏编码的计算量,人们期望从某类别训练样本中学习新的反映该类的子字典,使得该类子字典能够表示该类别样本子空间,并且子字典中的原子数量要远小于训练样本数,这样联合子字典构成新字典的原子数量会远少于训练样本数量,更便于对测试样本进行稀疏编码。Yang 等^[73]通过训练样本学习类别子字典代替原始训练样本字典对 SRC 改进,其中利用了稀疏系数向量的 Fisher 准则变体增加稀疏系数向量的判别性,使得稀疏系数向量具有较小的类内方差和较大的类间方差,并将类内方差与类间方差的比值简化为类内方差与类间方差的差值构造代价函数,进而学习到具有判别性的类别子字典,提高了人脸检测识别的效果,该方法采用的判别性指标并不是纯粹的 Fisher 准则,对学习判别性字典的利弊需要做进一步考虑。人们以 SRC 分类为基础在目标检测识别方面也做了很多工作,Yuan 等^[74]在 SRC 中引入联合稀疏性融合颜色、纹理等多特征通过分类对目标进行检测,考虑到测试样本的每种特征可以表示为训练样本该种特征的稀疏线性组合,而且不同特征的稀疏系数向量具有共同的稀疏性表示形式,也就是这些稀疏系数向量中的非零元素的位置保持一致(元素值大小可能不同),即联合稀疏性,在联合稀疏性约束下,可以得到具有某种结构特性的稀疏系数向量,再结合稀疏表示分类器对目标通过分类进行检测,并进一步扩展到核空间下,提高了花朵等目标的检测性能, SRC 其它方面的应用在下一节会有进一步讨论。

除了稀疏表示分类器在目标检测识别方面的成功应用外,另一个值得关注的是基于稀疏表示设计滤波器,在目标局部部件检测模型中得到了应用。在实际应用中,遮挡、类内变化及视角变化等因素会导致目标的整体特征发生变化,影响目标的检测效果,解决该问题的一个重要方法是基于目标局部部件的目标检测方法,该方法将整个目标分解成多个局部部件,通过提取各部件的特征构造局部组件检测器,检测是否存在相应部件并结合部件间的几何关系推断场景中是否存在感兴趣的待检测目标,能有效解决目标局部遮挡和柔性目标形变较大的检测问题^[75]。比较有代表性的方法是 Felzenszwalb 等^[76]把目标整体表示成部件的混和模型,将未标记部件作为隐含变量,由已标记样本的各部件训练潜在 SVM 分类器,用于预测隐含变量的类别;其后, Girshick 等^[77]提出了视觉语法模型(Visual Grammars),其中采用可变的多层结构描述目标,扩展了形变部件检测模型。由于在目标部件检测模型中需要将大量的滤波器与图像特征金字塔进行卷积运算,巨大的

计算量使得采用该模型对多类目标进行检测时无法满足实时性要求.最近, Song 等^[78]给出了一种 Sparselet 模型用于多类目标检测, 考虑到部件滤波器之间存在的冗余性结构, 采用稀疏编码方法先对大量部件滤波器学习一个滤波器公共字典, 这样每个部件滤波器都可以由该公共字典的稀疏线性组合来表示, 将图像特征金字塔和滤波器公共字典进行卷积, 再结合部件滤波器在滤波器字典上的稀疏系数向量来重构出图像特征金字塔与相应部件滤波器卷积的结果, 这种处理方式只需要图像特征金字塔和滤波器公共字典进行一次卷积, 降低了卷积运算所消耗的时间, 在保持检测性能的同时, 极大提高了多类目标检测的实时性, 并在此基础上结合稀疏约束的结构 SVM 进一步给出了判别性 Sparselet 模型^[79], 对多类目标检测更为有效.

除此之外, 采用稀疏约束度量重构时的残留误差矩阵也常用于目标检测模型中, 尤其是显著性目标检测. Shen 等^[43]先提取图像的颜色、边缘及纹理等底层特征构造图像联合特征, 以联合特征为基础采用均值漂移算法对图像进行过分割, 将每个过分割区域的代表性特征组成的特征矩阵分解为描述图像背景的低秩矩阵和描述显著性目标的稀疏残留误差矩阵, 其中引入线性特征变换对特征矩阵进行变换确保其能分解为低秩和稀疏两部分, 并设计了一种线性特征变换学习方法, 同时融合更高层特征的先验知识改进检测性能, 通过低秩和稀疏矩阵整合底层和更高层的特征完成显著性目标检测. Lang 等^[42]对图像的颜色、局部能量、局部对比度等特征采用低秩描述的方法进行显著性目标检测, 由于目标不同于背景块, 背景块之间一般存在较强的相关性, 背景块可由其自身描述, 相应的系数矩阵应具有低秩特性, 残留的误差矩阵表示了图像的显著性区域, 具有稀疏特性, 对于多特征来说, 误差矩阵的稀疏性应相同, 也就是非零列及其所在位置一致, 利用矩阵低秩和联合稀疏分解进行图像的显著性区域检测, 取得了很好的效果. 除了在显著性目标检测方面应用外, Guo 等^[80]采用低秩和稀疏的思想实现目标的联合检测, 将各种待检测目标样本和全部目标训练样本放到一起, 提取所有目标样本的 BOW(SIFT)、Gabor 及局部二值模式(LBP)等特征, 对于每种特征, 每个目标样本都可用其它所有样本进行线性重构, 联合所有样本的系数向量可以获得一个系数矩阵, 对全部特征来说, 该系数矩阵理论上应该相同, 是一个公共的系数矩阵, 其度量了不同待检测框的互相关性, 应是低秩矩阵, 这样样本特征可以分解为 2 部分: 特征与公共低秩系数矩阵的乘积、残留误差矩阵, 残留误差包括噪声、野点等, 采用稀疏性度量, 对分解得到的公共低秩系数矩阵采用谱聚类(Normalized Cuts)进行聚类, 相同的目标会聚为

一类, 根据聚类结果实现目标联合检测. 结合多特征, 利用重构过程中残留误差矩阵的稀疏性约束进行目标检测是一个值得关注的方法.

从上述讨论中可以看出, 采用稀疏表示方法直接构造合理的目标分类器和检测器模型, 能有效解决目标检测中面临的噪声、遮挡等干扰, 是一种有潜力的目标检测方法.

3.3 基于稀疏表示的多源信息融合目标检测

仅使用单源信息进行目标检测, 从场景中获取的信息有限, 面临一些难以克服的困难, 比如在光照变化下的可见光运动目标检测等, 单纯依赖设计新的和复杂的算法很难提高目标检测性能. 一种可行的方法是融合诸如多帧、多传感器、多光谱、多视角等多源信息进行检测, 以弥补单源信息的不足, 有助于获得更准确的检测结果. 但如何利用多帧图像、多传感器图像、多光谱图像以及多视角图像等提供的冗余性、互补性与合作性信息实现定量化、智能化和动态化处理, 以获得被感知目标的一致性和正确解释或描述, 是一个积极而富有挑战的任务, 目前已提出了一些基于稀疏表示的融合检测方法.

3.3.1 基于稀疏表示的多帧图像融合目标检测

对于多帧图像, 人们感兴趣的是利用多帧图像信息检测其中的运动目标, 常用的方法是混合高斯背景建模算法^[81]及其一系列衍生算法, 混合高斯背景建模算法需要复杂的统计知识和先验假设, 难以对非平稳等复杂背景进行建模, 目标检测效果并不理想. 最近, 基于低秩和稀疏分解思想的运动目标检测算法以其良好的检测性能广泛用于运动目标检测问题, 该算法具有清晰的物理概念和坚实的理论保证, 不需要复杂的背景建模过程, 而且便于根据实际问题融合多种模型, 成为国内外学者研究的热点之一. Zhou 等^[40]提出了一种基于低秩和稀疏分解的运动目标检测模型, 考虑到不同帧的背景之间存在很强的线性相关性, 将所有帧的背景图像向量化后组成的背景项矩阵建模为低秩矩阵; 而前景运动目标项相对背景来说数量较少, 采用稀疏性约束; 并通过图剪切方法加入空间约束, 保持目标区域的完整性; 另外, 引入二维参数变换补偿由于视觉传感器震动引起的图像晃动, 联合上述约束项建立了一种运动目标检测模型, 通过交替优化方法估计其中各变量, 可将多帧图像分解为低秩背景项和稀疏运动目标项, 在非平稳背景下取得了不错的检测效果. Or-eifej 等^[41]采用矩阵低秩和稀疏分解方法研究了大气湍流干扰下的红外运动目标检测问题, 将图像序列分解为稀疏约束的运动目标项、低秩约束的背景项和 F 范数约束的湍流干扰项(稠密误差矩阵), 其中考虑到湍流干扰项在运动域和强度域呈现的类似高斯分布特

性,构造目标可信图用于约束目标检测模型,在湍流干扰复杂背景下能较好检测出红外运动目标,由于没有考虑目标的空间一致性约束,目标轮廓还不够完整. Li 等^[82]没有对图像矩阵进行向量化处理,而是直接对多帧图像采用张量进行处理,其中采用张量迹范数度量背景的低秩性,较向量化处理方式能更好描述背景的空间结构相关性,对张量模型通过多模展开进行求解,对漂浮在水面的运动目标实现了较好检测,但模型的优化求解精度需进一步完善. 当前基于低秩和稀疏分解的目标检测方法能较处理好非平稳背景下的运动目标检测问题,但大都只考虑了目标的稀疏性和背景的低秩性,忽略了目标和背景固有特征的使用,而且是离线处理方式,难以在线处理新输入数据的目标检测问题,同时在实际应用中还面临着模型参数的选取问题.

3.3.2 基于稀疏表示的多传感器图像融合目标检测

在实际应用中,不同传感器提供的数据及其特征很难归一化到相同的尺度和空间下,融合这些不均等的复杂数据及其特征是一个困难的任务,需要对不同传感器数据及其特征做各种变换和组合,变换和组合方式的不同对检测结果影响很大,因此,人们很少在数据层和特征层上进行融合检测,通常更愿意在决策层上融合多传感器信息进行目标检测,但决策层上的融合方式很难利用目标的原始互补和合作信息,影响目标检测性能. 而稀疏表示方法能通过稀疏系数向量中的非零元素建立不同传感器数据及其特征之间的关系,将不同类型数据及其特征的融合问题转化到稀疏系数向量这个新的均等空间下进行处理,便于多传感器数据及其特征的融合检测. Shekhar 等^[83]结合稀疏表示分类器和联合稀疏性对多模生物数据进行检测识别,先对样本数据提取 Gabor 特征,考虑到每种模态测试样本特征可以表示为该模态训练样本特征的稀疏线性组合,而且不同模态数据的稀疏系数向量具有共同的稀疏性表示形式,即联合稀疏性,再结合稀疏表示分类器对目标进行检测识别,并进一步扩展到核空间下,能在特征层上融合指纹、虹膜等多模数据对人进行检测识别,对噪声和遮挡等干扰具有较好鲁棒性. Nguyen 等^[23]也采用相似的思想融合声波传感器、地震传感器、红外传感器和超声波传感器等数据,区分边境线上人类和动物的脚步. 在实际应用中,一种常见的多传感器组合方式是融合可见光和红外图像进行目标检测^[84],既能克服运动阴影、光照变化、夜间环境等可见光图像的不足,又能处理对温度变化敏感、目标轮廓和边缘模糊等红外图像的局限. 目前融合检测算法大多是决策层融合,先对可见光和红外图像分别进行目标检测,得到初步检测结果,再使用先验融合策略融合两者的检测结果,获得最终检测结果. 这其中单独从可见光或红

外图像中提取目标,在目标提取过程中没有充分利用可见光与红外图像的互补信息,而且需要人为选择合适的融合策略完成检测任务. Gao 等^[85]基于低秩和稀疏分解方法提出了一种可见光与红外图像相融合的运动目标检测模型,考虑到真实的运动目标应该在对应帧的可见光与红外图像中占据相同的位置,并且可见光图像中的运动非目标项和红外图像中的运动非目标项在相同位置同时发生的可能性很小,以此为基础通过建立低秩联合背景项、稀疏公共真实目标项以及运动非目标项构造运动目标检测模型直接在数据层上对可见光与红外图像信息进行融合检测,在目标检测过程中不但利用了可见光与红外图像的互补信息,而且不需要较多的先验融合规则,在运动阴影、光照变化和夜间环境等复杂场景下达到较好的目标检测效果,但该模型同样面临着参数的合理选取问题.

3.3.3 基于稀疏表示的多光谱图像融合目标检测

利用目标和背景在不同光谱上的区分信息,能有效解决弱目标检测问题. 传统方法很多都是基于统计假设检验,统计假设模型的准确与否对检测结果影响很大,而且光照和大气变化等干扰因素对检测结果影响严重. Chen 等^[86]基于稀疏表示分类器提出了一种高光谱目标检测算法,将待检测像元的光谱特征向量对目标和背景所组成的联合字典进行稀疏编码,再通过稀疏系数向量分别在目标字典和背景字典上对原始像元光谱特征向量进行重建,根据重建误差大小判断待检测像元属于目标还是背景;鉴于相邻像元具有较强的空间相关性,应归属于相同的类别,相邻像元的光谱差异主要来自于传感器本身噪声和大气变化,采用 4 邻域向量拉普拉斯算子引入空间约束,通过联合邻域像元的稀疏系数向量组成更大的稀疏系数向量,以此为基础结合稀疏表示分类器计算待检测像元对目标字典和背景字典的重建误差大小,实现对目标的检测,这种处理方式有效融合了像元的光谱和空间信息,显著改进了目标检测效果;其中需要结合物理模型和 MODTRAN 大气模型仿真生成不同大气情况下的目标训练样本用于学习目标字典,从样本中随机截取数据作为背景训练样本学习背景字典,在字典学习过程中采用的是 K-SVD 算法,学习到的是重构性字典,将来可以考虑学习判别性字典以提高检测效果. Chen 等^[87]还考虑了邻域联合稀疏模型,即:相邻像元具有相同的稀疏性结构,在稀疏表示分类器基础上引入联合稀疏性,采用同时正交匹配追踪算法(SOMP)进行求解优化,并扩展到核空间中^[88],也具有较好的检测结果. Huang 等^[89]进一步给出了权重 l_1 范数优化方法避免稀疏编码优化过程中出现的局部最优问题,改进了高光谱目标检测效果. 上述基于稀疏表示的高光谱目标检测方法不需要

对目标和背景像元进行统计分布特性假设,而且对光照和大气变化等干扰更为鲁棒.

稀疏表示方法除了在上述融合多帧、多传感器、多光谱等图像信息进行目标检测外,在多视角图像融合目标检测方面也有一些应用,例如 Zhang 等^[90]结合稀疏表示分类器和联合稀疏性对多视角 SAR 目标的检测,也取得了不错的效果.

4 总结与展望

目标检测是一个很庞大的体系,稀疏表示为目标检测提供了一种新的研究思路,已经初步渗入到目标检测领域的多个方面,从报道的结果看,很多取得了目前最好结果,显示出了稀疏表示方法的优势,但其在目标检测领域中的应用还远没有充分发掘,针对具体目标检测问题的相关研究仍然很少.另外,由于目标检测问题的多样性和复杂性,以及稀疏表示理论研究的不完善,还有不少问题需要深入研究,将来的研究方向主要有:

(1)构建鲁棒的目标检测模型:稀疏表示是一种处理方式,要达到良好的目标检测性能,需要在深入分析目标成像机理的基础上,构建鲁棒的目标检测模型.目前大部分研究集中在稀疏表示的传统模型上,没有充分考虑目标的固有特性和内在结构,探索结构化的稀疏表示目标检测模型是将来的重要研究方向,如组稀疏性、联合稀疏性及结构化字典等.在实际应用中,基于稀疏表示的检测模型需要对一些关键参数进行选取,如字典中原子的个数、稀疏性约束的强弱等,目前大都是凭经验设置和调整,如何结合目标检测模型从理论上对参数做出较优选取需要进一步研究,从贝叶斯和信息论角度上进行探讨是一个注意的方向;另外,在目标检测中,通常需要对目标进行去噪、分割等处理,如何将去噪、分割等思想有机融合到目标检测模型中需要进一步的考虑.

(2)算法实时性,大规模数据及在线处理:基于稀疏表示的目标检测算法虽然能有效改善目标检测的性能,但是其付出的计算时间代价是巨大的,是稀疏表示在实际应用中的一个瓶颈,需要进一步研究如何提高优化算法的计算速度和效率,尤其是需要研究面向大规模数据的快速有效优化算法,使得稀疏表示方法进一步工程化.同时,当前基于稀疏表示的方法大都采用的是离线处理方式,在实际应用中需要考虑稀疏表示的在线处理方式,能够根据实时数据实时检测目标.

(3)稀疏表示度量空间:目前基于稀疏表示的算法大都是基于欧式距离,数据本身的复杂模式使得欧式空间难以描述数据的内在关系,根据数据本身的特点,将稀疏表示进一步扩展到流形空间或核空间中是今后

的一个研究方向.

(4)多线性、多尺度稀疏表示研究:目前稀疏表示算法大都是将图像或图像块向量化后进行处理,这种处理方式破坏了图像或图像块各像素之间的空间结构关系,同时易生成高维向量导致“维数灾难”,一种可行的解决方式是建立二阶甚至高阶张量的稀疏表示框架,对多图像块、多时相图像、多波段图像和多模图像等进行描述以期能建立更优的稀疏表示模型;另外,将稀疏表示与多尺度相结合,构建多尺度稀疏表示模型,能更好捕捉目标的有用特征,有利于改善目标检测的性能和鲁棒性,是一个值得研究的方向.

(5)目标鲁棒特征学习:良好的目标特征对目标检测性能的提高具有重要意义,直接设计不变性和可区分性的目标特征是一个困难的任务,通过稀疏表示如何从数据中有监督和无监督的学习较好的目标底层特征、中层特征及高层特征需要进一步的研究;如何将深度学习概念与稀疏表示方法深入的结合,提取更为本质的目标特征改进目标的检测识别效果,是将来的一个研究重点.

本文着重讨论了近年来稀疏表示在目标检测领域中的研究进展,希望能对从事该方面研究的人员提供有益帮助.

参考文献

- [1] Vinje W E, Gallant J L. Sparse coding and decorrelation in primary visual cortex during natural vision[J]. Science, 2000, 287 (5456): 1273 - 1276.
- [2] Nirenberg S, Carcieri S, Jacobs A, et al. Retinal ganglion cells act largely as independent encoders[J]. Nature, 2001, 411 (6838): 698 - 701.
- [3] Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 411 - 426.
- [4] 赵松年, 姚力, 金真, 等. 视像整体特征在人类初级视觉层上的稀疏表象: 脑功能成像的证据[J]. 科学通报, 2008, 53 (11): 1296 - 1304.
- [5] 杨静宇, 金忠, 杨健. 模式特征抽取研究进展[A]. 2009 中国自动化大会暨两化融合高峰论坛[C]. 杭州: 2009.
- [6] 焦李成, 谭山. 图像的多尺度几何分析: 回顾和展望[J]. 电子学报, 2003, 31(12): 1975 - 1981.
- [7] Jiao L C, Tan S. Development and prospect of image multiscale geometric analysis[J]. Acta Electronica Sinica, 2003, 31(12): 1975 - 1981. (in Chinese)
- [8] Candes E J. Ridgelets: Theory and applications[D]. Stanford: Stanford University, 1998.
- [9] Starck J-L, Candès E J, Donoho D L. The curvelet transform for image denoising[J]. IEEE Transactions on Image Process-

- ing, 2002, 11(6): 670 – 684.
- [9] Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397 – 3415.
 - [10] Donoho D L, Elad M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization[J]. Proceedings of the National Academy of Sciences, 2003, 100(5): 2197 – 2202.
 - [11] Candes E J, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(12): 4203 – 4215.
 - [12] Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: A strategy employed by V1? [J]. Vision Research, 1997, 37(23): 3311 – 3326.
 - [13] Pati Y C, Rezaifar R, Krishnaprasad P. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition[A]. Asilomar Conference on Signals, Systems and Computers[C]. Pacific Grove, CA: IEEE Press, 1993. 40 – 44.
 - [14] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[A]. Advances in neural information processing systems (NIPS)[C]. MIT Press, 2007. 801 – 808.
 - [15] Aharon M, Elad M, Bruckstein A. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311 – 4322.
 - [16] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding[J]. The Journal of Machine Learning Research, 2010, 11: 19 – 60.
 - [17] Huang J, Zhang T, Metaxas D. Learning with structured sparsity[J]. The Journal of Machine Learning Research, 2011, 12: 3371 – 3412.
 - [18] Jacob L, Obozinski G, Vert J-P. Group lasso with overlap and graph lasso[A]. International Conference on Machine Learning (ICML)[C]. New York, NY: ACM, 2009. 433 – 440.
 - [19] Zhang Z, Rao B. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation[J]. IEEE Transaction on Signal Processing, 2013, 61(8): 2009-2015
 - [20] Qin Z, Scheinberg K, Goldfarb D. Efficient block-coordinate descent algorithms for the group lasso[J]. Mathematical Programming Computation, 2013, 5(2): 143 – 169.
 - [21] Yuan X-T, Yan S. Visual classification with multi-task joint sparse representation[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. San Francisco, CA: IEEE Press, 2010. 3493 – 3500.
 - [22] Zhang H, Nasrabadi N M, Zhang Y, et al. Multi-observation visual recognition via joint dynamic sparse representation[A]. IEEE International Conference on Computer Vision (ICCV) [C]. Barcelona: IEEE Press, 2011. 595 – 602.
 - [23] Nguyen N H, Nasrabadi N M, Tran T D. Robust multi-sensor classification via joint sparse representation[A]. International Conference on Information Fusion [C]. Chicago, IL: IEEE Press, 2011. 1 – 8.
 - [24] Szabó Z, Póczos B, Lorincz A. Online group-structured dictionary learning[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Providence, RI: IEEE Press, 2011. 2865 – 2872.
 - [25] Mairal J, Bach F, Ponce J, et al. Discriminative learned dictionaries for local image analysis[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Anchorage, AK: IEEE Press, 2008. 1 – 8.
 - [26] Smith L N, Elad M. Improving dictionary learning: Multiple dictionary updates and coefficient reuse[J]. IEEE Signal Processing Letters, 2013, 20(1): 79 – 82.
 - [27] Mairal J, Bach F, Ponce J, et al. Supervised dictionary learning [A]. Advances in neural information processing systems (NIPS)[C]. Hyatt Regency, Vancouver; 2008. 1033 – 1040.
 - [28] Mairal J, Bach F, Ponce J. Task-driven dictionary learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 791 – 804.
 - [29] Dikmen O, Févotte C. Nonnegative dictionary learning in the exponential noise model for adaptive music signal representation[A]. Advances in neural Information processing systems (NIPS)[C]. Lake Tahoe, NV; 2012. 2267 – 2275.
 - [30] Candès E J, Recht B. Exact matrix completion via convex optimization [J]. Foundations of Computational mathematics, 2009, 9(6): 717 – 772.
 - [31] 林杰, 石光明, 董伟生. 基于信息自由度采样的信号重构方法研究进展[J]. 电子学报, 2012, 40(8): 1640 – 1649.
LIN J, SHI G M, DONG W S. Research Advances in Reconstruction Methods Based on Information Degree-of-Freedom Sampling[J]. Acta Electronica Sinica, 2012, 40(8): 1640 – 1649. (in Chinese)
 - [32] Cai J-F, Candès E J, Shen Z. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956 – 1982.
 - [33] Lin Z, Chen M, Ma Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices[A/OL]. <http://arxiv.org/abs/1009.5055>, 2010-09-26.
 - [34] Liu J, Musialski P, Wonka P, et al. Tensor completion for estimating missing values in visual data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 208 – 220.
 - [35] Cabral R S, De La Torre F, Costeira J P, et al. Matrix completion for multi-label image classification[A]. Advances in neural information processing systems (NIPS)[C]. Sierra Nevada, Spain; 2011. 190 – 198.
 - [36] Wright J, Ganesh A, Rao S, et al. Robust principal component

- analysis: Exact recovery of corrupted low-rank matrices via convex optimization[A]. Advances in neural information processing systems (NIPS) [C]. Whistler, BC: 2009. 2080 – 2088.
- [37] Candès E J, Li X, Ma Y, et al. Robust principal component analysis? [J]. Journal of the ACM, 2011, 58(3): 11.
- [38] Peng Y, Ganesh A, Wright J, et al. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2233 – 2246.
- [39] Wang X, Wan W, Liu G. Multi-task low-rank and sparse matrix recovery for human motion segmentation[A]. IEEE International Conference on Image Processing (ICIP) [C]. Orlando, FL: IEEE Press, 2012. 897 – 900.
- [40] Zhou X, Yang C, Yu W. Moving object detection by detecting contiguous outliers in the low-rank representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 597 – 610.
- [41] Oreifej O, Li X, Shah M. Simultaneous video stabilization and moving object detection in turbulence[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(2): 450 – 462.
- [42] Lang C, Liu G, Yu J, et al. Saliency detection by multitask sparsity pursuit[J]. IEEE Transactions on Image Processing, 2012, 21(3): 1327 – 1338.
- [43] Shen X, Wu Y. A unified approach to salient object detection via low rank matrix recovery[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Providence, RI: IEEE Press, 2012. 853 – 860.
- [44] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [45] Avila S, Thome N, Cord M, et al. Pooling in image representation: The visual codeword point of view[J]. Computer Vision and Image Understanding, 2013, 117(5): 453 – 465.
- [46] Koniusz P, Yan F, Mikolajczyk K. Comparison of mid-level feature coding approaches and pooling strategies in visual concept detection[J]. Computer Vision and Image Understanding, 2013, 117(5): 479 – 492.
- [47] Leung T, Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons[J]. International Journal of Computer Vision, 2001, 43(1): 29 – 44.
- [48] Van Gemert J C, Veenman C J, Smeulders A W, et al. Visual word ambiguity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1271 – 1283.
- [49] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Miami, FL: IEEE Press, 2009. 1794 – 1801.
- [50] Boureau Y-L, Ponce J, Lecun Y. A theoretical analysis of feature pooling in visual recognition[A]. International Conference on Machine Learning (ICML) [C]. Haifa, Israel: Omnipress, 2010. 111 – 118.
- [51] Yang J, Yu K, Huang T. Supervised translation-invariant sparse coding[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. San Francisco, CA: 2010. 3517 – 3524.
- [52] Boureau Y-L, Bach F, Lecun Y, et al. Learning mid-level features for recognition[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. San Francisco, CA: 2010. 2559 – 2566.
- [53] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. San Francisco, CA: 2010. 3360 – 3367.
- [54] Lei H, Mei K, Zheng N, et al. Learning group-based dictionaries for discriminative image representation[J]. Pattern Recognition, 2014, 47(2): 899 – 913.
- [55] Naikal N, Yang A Y, Sastry S S. Informative feature selection for object recognition via sparse PCA[A]. IEEE International Conference on Computer Vision (ICCV) [C]. Barcelona: IEEE Press, 2011. 818 – 825.
- [56] Ren X, Ramanan D. Histograms of sparse codes for object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Portland, OR: IEEE Press, 2013. 3246 – 3253.
- [57] Rigamonti R, Brown M A, Lepetit V. Are sparse representations really relevant for image classification? [A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Providence, RI: IEEE Press, 2011. 1545 – 1552.
- [58] Rigamonti R, Sironi A, Lepetit V, et al. Learning separable filters[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Portland, OR: IEEE Press, 2013. 2754 – 2761.
- [59] Kavukcuoglu K, Sermanet P, Boureau Y-L, et al. Learning convolutional feature hierarchies for visual recognition[A]. Advances in neural information processing systems (NIPS) [C]. Vancouver, BC: 2010. 1090 – 1098.
- [60] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展, 2013, 50(9): 1799 – 1804.
YU K, JIA L, CHEN Y Q, et al. Deep Learning: Yesterday, Today, and Tomorrow[J]. Journal of Computer Research and Development, 2013, 50(9): 1799 – 1804. (in Chinese)
- [61] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504 – 507.
- [62] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends® in Machine Learning, 2009, 2(1): 1 – 127.

- [63] Hinton G E, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527 – 1554.
- [64] Jarrett K, Kavukcuoglu K, Ranzato M A, et al. What is the best multi-stage architecture for object recognition? [A]. *International Conference on Computer Vision (ICCV)*[C]. Kyoto: IEEE Press, 2009. 2146 – 2153.
- [65] Sermanet P, Kavukcuoglu K, Chintala S, et al. Pedestrian detection with unsupervised multi-stage feature learning [A]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*[C]. Portland, OR: IEEE Press, 2013. 3626 – 3633.
- [66] Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[A]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [C]. San Francisco, CA: IEEE Press, 2010. 2528 – 2535.
- [67] 张长水. 机器学习面临的挑战[J]. *中国科学: 信息科学*, 2013, 43(12): 1612 – 1623.
ZHANG C. Challenges in machine learning [J]. *SCIENCE CHINA Information Sciences*, 2013, 43(12): 1612-1623. (in Chinese)
- [68] Ngiam J, Koh P W, Chen Z, et al. Sparse Filtering[A]. *Advances in Neural Information Processing Systems (NIPS)* [C]. Sierra Nevada, Spain; 2011. 1125 – 1133.
- [69] 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.
- [70] Wagner A, Wright J, Ganesh A, et al. Toward a practical face recognition system: Robust alignment and illumination by sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(2): 372 – 386.
- [71] Yang M, Zhang L, Zhang D. Efficient misalignment-robust representation for real-time face recognition [A]. *European Conference on Computer Vision (ECCV)* [C]. Florence, Italy: Springer Berlin Heidelberg, 2012. 850 – 863.
- [72] He R, Zheng W-S, Hu B-G. Maximum correntropy criterion for robust face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1561 – 1576.
- [73] Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation[A]. *IEEE International Conference on Computer Vision (ICCV)* [C]. Barcelona: IEEE Press, 2011. 543 – 550.
- [74] Yuan X, Liu X, Yan S. Visual Classification With Multitask Joint Sparse Representation[J]. *IEEE Transactions on Image Processing*, 2012, 21(10): 4349 – 4360.
- [75] 潘泓, 金立左, 夏思宇, 等. 基于多层次互补特征的通用目标检测模型[J]. *电子与信息学报*, 2012, 34(7): 1531 – 1537.
PAN H, JIN L Z, XIA S Y, et al. A hierarchical and complementary feature-based model for generic object detection[J]. *Journal of Electronics & Information Technology*, 2012, 34(7): 1531 – 1537. (in Chinese)
- [76] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627 – 1645.
- [77] Girshick R B, Felzenszwalb P F, Mcallester D. Object detection with grammar models[A]. *Advances in neural Information processing systems (NIPS)* [C]. Sierra Nevada, Spain; 2011. 442 – 450.
- [78] Song H O, Zickler S, Althoff T, et al. Sparselet models for efficient multiclass object detection [A]. *European Conference on Computer Vision (ECCV)* [C]. Florence, Italy: Springer Berlin Heidelberg, 2012. 802 – 815.
- [79] Song H O, Darrell T, Girshick R B. Discriminatively activated sparselets[A]. *International Conference on Machine Learning (ICML)*[C]. Atlanta, GA; 2013: 196 – 204.
- [80] Guo X, Liu D, Jou B, et al. Robust Object Co-detection[A]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*[C]. Portland, OR: IEEE Press, 2013. 3206 – 3213.
- [81] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[A]. *International Conference on Pattern Recognition (ICPR)* [C]. Cambridge, UK: IEEE Press, 2004. 28 – 31.
- [82] Li Y, Yan J, Zhou Y, et al. Optimum subspace learning and error correction for tensors [A]. *European Conference on Computer Vision (ECCV)* [C]. Crete, Greece: Springer Berlin Heidelberg, 2010. 790 – 803.
- [83] Shekhar S, Patel V, Nasrabadi N, et al. Joint sparse representation for robust multimodal biometrics recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(1): 113 – 126.
- [84] Davis J W, Sharma V. Background-subtraction using contour-based fusion of thermal and visible imagery[J]. *Computer Vision and Image Understanding*, 2007, 106(2): 162 – 182.
- [85] Gao S, Cheng Y, Zhao Y. Method of visual and infrared fusion for moving object detection[J]. *Optics Letters*, 2013, 38(11): 1981 – 1983.
- [86] Chen Y, Nasrabadi N M, Tran T D. Sparse representation for target detection in hyperspectral imagery[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(3): 629 – 640.
- [87] Chen Y, Nasrabadi N M, Tran T D. Simultaneous joint sparsity model for target detection in hyperspectral imagery [J]. *IEEE Geoscience and Remote Sensing Letters*, 2011, 8(4): 676 – 680.
- [88] Chen Y, Nasrabadi N M, Tran T D. Kernel sparse representation for hyperspectral target detection [A]. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* [C].

Munich: IEEE Press, 2012. 7484 – 7487.

- [89] Huang Z, Shi Z, Qin Z. Convex relaxation based sparse algorithm for hyperspectral target detection[J]. Optik-International Journal for Light and Electron Optics, 2013, 124(24): 6594 –

6598.

- [90] Zhang H, Nasrabadi N M, Huang T S, et al. Joint sparse representation based automatic target recognition in SAR images [A]. Proc. of SPIE [C]. Orlando, FL: SPIE Press, 2011. 805112 – 1.

作者简介



高仕博 男, 1982 年出生, 河北邢台人. 2006 年和 2013 年在西北工业大学自动化学院分别获得工学学士和工学博士学位, 现在北京航天自动控制研究所工作, 主要研究方向为目标检测、图像处理、机器学习等.

E-mail: gaohbob@gmail.com



程咏梅 女, 1960 年出生, 陕西西安人. 教授、博士生导师. 1997 年和 2001 年在西北工业大学分别获工学硕士和工学博士学位, 现为西北工业大学自动化学院教授, 主要从事信息融合、图像处理、证据推理等方面的研究工作.