



信息科学技术专著丛书

# 智能视频数据 处理与挖掘

梁美玉 著

INTELLIGENT VIDEO DATA PROCESSING AND MINING



北京邮电大学出版社  
[www.buptpress.com](http://www.buptpress.com)



信息科学技术专著丛书

# 智能视频数据处理与挖掘

梁美玉 著



北京邮电大学出版社  
[www. buptpress. com](http://www.buptpress.com)

## 内 容 简 介

本书主要研究了智能视频数据处理与挖掘的相关技术及应用,以提高视频数据中异常事件检测与识别的智能化、精准化水平和鲁棒性、实时性等性能为目的,从而实现对异常突发事件的及时预测和预警,保障公共安全。本书首先研究了视频数据的去噪技术,重点研究了基于残差卷积神经网络的视频去噪算法,然后研究了视频数据的超分辨率重建技术,包括基于深度学习、半耦合字典学习和时空非局部相似性特征的视频超分辨率重建算法;其次研究了视频的显著性时空特征提取算法,在此基础上研究了视频异常事件检测与识别技术,重点研究了基于稀疏组合学习的视频异常事件检测算法,以及基于时空感知深度网络的视频异常事件识别算法;最后构建了视频数据去噪和超分辨率重建系统,并以面向智慧旅游领域的旅游景区视频数据为例,将本书提出的相关模型和算法应用于旅游景区视频数据的智能挖掘,构建了旅游景区视频异常事件检测与识别系统,及时自动监测旅游突发事件。此外,本书相关技术还可应用于智慧城市、智能交通、智慧校园、智慧医疗等领域,切实加强保障公共安全,以实现智慧安防。

本书体系结构完整,注重理论联系实际,可作为计算机应用、人工智能、大数据、电子信息等相关专业的工程技术人员、科研人员、研究生和高年级本科生的参考用书。

### 图书在版编目(CIP)数据

智能视频数据处理与挖掘 / 梁美玉著. -- 北京:北京邮电大学出版社, 2022. 4

ISBN 978-7-5635-6444-6

I. ①智… II. ①梁… III. ①视频系统—数据处理②视频系统—数据采集 IV. ①TP274

中国版本图书馆CIP数据核字(2021)第156879号

策划编辑:姚 顺 刘纳新 责任编辑:王晓丹 左佳灵 封面设计:七星博纳

出版发行:北京邮电大学出版社

社 址:北京市海淀区西土城路10号

邮政编码:100876

发行部:电话:010-62282185 传真:010-62283578

E-mail: publish@bupt.edu.cn

经 销:各地新华书店

印 刷:唐山玺诚印务有限公司

开 本:787 mm×1 092 mm 1/16

印 张:12.75

字 数:240千字

版 次:2022年4月第1版

印 次:2022年4月第1次印刷

ISBN 978-7-5635-6444-6

定 价:48.00元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

# 前 言

通过视频数据智能处理和挖掘技术，实现视频数据增强处理及异常事件检测与识别，对于异常突发事件的及时预测和预警，公共安全保障工作的智能化、自动化、精准化，公共安全的保障，智慧安防目标的实现，具有重要的研究价值和应用前景。然而，监控视频大数据中往往包含背景混杂、目标间以及目标与背景间互相遮挡、视角变化、光照变化等复杂的运动场景，并且视频监控数据容易受噪声、运动或光学模糊、下采样等多种降质因素的影响，这些因素都给视频中异常事件的检测与识别带来了新的挑战，因此需要进一步研究对复杂场景具有较好鲁棒性的目标行为特征描述以及高效的异常事件检测与识别方法，提升其在复杂场景下的检测与识别性能，为异常突发事件的预测、预警以及应急决策提供有力的技术支撑，进而实现智慧安防。

针对噪声干扰、运动或光学模糊、下采样等多种降质因素对视频数据中异常事件检测和识别精准度的影响，本书对视频数据的去噪和超分辨率重建处理技术进行了研究。本书提出了基于残差卷积神经网络的视频去噪算法，解决了传统视频去噪算法对复杂噪声泛化能力不强的问题；针对现有的视频超分辨率重建方法对视频的时空一致性保持能力不强以及对复杂运动场景鲁棒性不高的问题，提出了基于半耦合字典学习和时空非局部相似性的视频超分辨率重建算法；针对现有方法由于依赖有限规模的外部训练实例或者内部相似块实例不匹配而产生噪声、过平滑或者视觉瑕疵的问题，提出了基于深度学习和时空特征相似性的视频超分辨率重建算法；综合利用外部关联映射学习和内部时空非局部自相似性先验约束，通过两者的优势互补，构建了内外部联合约束的视频超分辨率重建机制；针对现有事件表示方法没有充分考虑视频的帧间时空相关性，难以适用于背景混杂、目标间相互遮挡等复杂运动场景的问题，提出了一种视频显著性时空特征提取方法；针对现有视频异常事件检测方法在复杂运动场景下鲁棒性和时效性不高，无法适用于实际应用中异常事件实时检测的问题，提出了一种基于稀疏组合学习的视频异常事件检测方法；针对现有视频异常事件识别方法大多只在空间域上学习视频特征，没有考虑视频时间域信息的问题，提出了一种基于时空感

知深度网络的视频异常事件识别方法。

本书共分 10 章：第 1 章为绪论；第 2 章介绍了视频数据智能处理与挖掘的相关技术；第 3 章介绍了基于残差卷积神经网络的视频去噪算法；第 4 章和 5 章分别介绍了基于半耦合字典学习和时空非局部相似性的视频超分辨率重建算法以及基于深度学习和时空特征相似性的视频超分辨率重建算法；第 6 章介绍了视频显著性时空特征提取算法；第 7 章介绍了基于稀疏组合学习的视频异常事件检测算法；第 8 章介绍了基于时空感知深度网络的视频异常事件识别算法；第 9 章综合以上各章方法，设计并实现了视频数据去噪和超分辨率重建系统；第 10 章介绍了本书相关技术在智慧旅游领域的应用，设计并实现了旅游景区视频异常事件检测与识别系统。

因作者水平有限，书中错误在所难免，请读者多批评指正。

作 者

# 目 录

第 1 章 绪论 .....	1
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	2
1.2.1 视频超分辨率重建方法的研究现状 .....	2
1.2.2 视频去噪方法的研究现状 .....	7
1.2.3 视频异常事件检测方法的研究现状 .....	9
1.2.4 视频异常事件识别方法的研究现状 .....	11
参考文献 .....	13
第 2 章 相关技术 .....	23
2.1 稀疏字典学习 .....	23
2.1.1 低秩矩阵的生成 .....	23
2.1.2 低秩矩阵的分解 .....	24
2.2 视频的时空相似性学习 .....	25
2.3 视频特征提取 .....	26
2.3.1 运动目标检测 .....	26
2.3.2 背景建模方法 .....	28
2.3.3 行为特征表示 .....	29
2.4 基于深度学习模型的视频超分辨率重建 .....	30
2.4.1 基于深度卷积神经网络的超分辨率重建算法概述 .....	30

2.4.2	深度卷积神经网络在超分辨率重建算法中的应用	32
2.5	视频异常事件检测	33
2.5.1	基于深度学习的异常事件检测方法	33
2.5.2	基于稀疏表示的异常事件检测方法	34
2.6	视频异常事件识别	35
	本章小结	36
	参考文献	37
<b>第3章</b>	<b>基于残差卷积神经网络的视频去噪</b>	<b>42</b>
3.1	基于残差卷积神经网络的视频去噪算法框架	43
3.2	基于残差卷积神经网络的视频去噪算法实现	43
3.2.1	基于低秩矩阵分解的稀疏字典去噪	43
3.2.2	残差图像预处理	45
3.2.3	卷积神经网络训练	47
3.2.4	基于残差卷积神经网络的视频去噪算法的实现步骤	48
3.3	实验结果及分析	49
3.3.1	实验一:不同噪声方差下对随机噪声的去噪效果对比实验	49
3.3.2	实验二:相同噪声方差下对不同噪声的去噪效果对比实验	55
	本章小结	59
	参考文献	60
<b>第4章</b>	<b>基于半耦合字典学习和时空非局部相似性的视频超分辨率重建</b>	<b>62</b>
4.1	引言	62
4.2	视频超分辨率重建观测模型	64
4.3	基于非局部相似性的超分辨率重建	65
4.4	基于半耦合字典学习和时空非局部相似性的视频超分辨率重建算法	67
4.4.1	CNLSR 算法研究动机	67
4.4.2	CNLSR 算法框架	67

4.4.3	CNLSR 算法数学模型 .....	69
4.4.4	CNLSR 算法描述 .....	69
4.4.5	CNLSR 算法步骤 .....	76
4.5	实验结果与分析 .....	76
4.5.1	实验数据集 .....	76
4.5.2	客观评价指标 .....	77
4.5.3	实验结果与分析 .....	78
	本章小结 .....	93
	参考文献 .....	94
<b>第 5 章</b>	<b>基于深度学习和时空特征自相似性的视频超分辨率重建 .....</b>	<b>98</b>
5.1	引言 .....	98
5.2	基于深度学习和时空特征相似性的视频超分辨率重建算法 .....	101
5.2.1	DLSS-VSR 算法研究动机 .....	101
5.2.2	DLSS-VSR 算法框架 .....	101
5.2.3	DLSS-VSR 算法数学模型 .....	102
5.2.4	DLSS-VSR 算法描述 .....	103
5.2.5	DLSS-VSR 算法步骤 .....	112
5.3	实验结果与分析 .....	113
5.3.1	实验数据集 .....	113
5.3.2	客观评价指标 .....	114
5.3.3	实验结果与分析 .....	114
	本章小结 .....	127
	参考文献 .....	128
<b>第 6 章</b>	<b>视频显著性时空特征提取 .....</b>	<b>131</b>
6.1	视频显著性时空特征提取的算法框架 .....	131
6.2	视频显著性时空特征提取的算法实现 .....	132

6.2.1	视频背景建模和前景检测 .....	132
6.2.2	视频显著性时空特征提取 .....	134
6.2.3	视频显著性时空特征降维 .....	135
6.3	视频显著性时空特征提取实验结果 .....	136
6.3.1	视频背景建模和前景检测结果 .....	136
6.3.2	视频显著性时空特征提取结果 .....	137
	本章小结 .....	138
	参考文献 .....	139
<b>第7章</b>	<b>基于稀疏组合学习的视频异常事件检测 .....</b>	<b>140</b>
7.1	基于稀疏组合学习的视频异常事件检测的算法框架 .....	140
7.2	基于稀疏组合学习的视频异常事件检测的算法实现 .....	141
7.2.1	基于稀疏组合学习的异常事件检测的主要思想 .....	141
7.2.2	基于稀疏组合学习的异常事件检测的训练过程 .....	142
7.2.3	基于稀疏组合学习的异常事件检测的测试过程 .....	144
7.3	实验结果与分析 .....	145
7.3.1	数据集 .....	145
7.3.2	评价指标 .....	145
7.3.3	ScenicSpot 景区数据集的实验结果与分析 .....	146
7.3.4	Avenue 标准数据集的实验结果与分析 .....	148
7.3.5	UCSD Ped1 数据集的实验结果与分析 .....	152
	本章小结 .....	155
	参考文献 .....	156
<b>第8章</b>	<b>基于时空感知深度网络的视频异常事件识别 .....</b>	<b>157</b>
8.1	基于时空感知深度网络的视频异常事件识别的算法框架 .....	157
8.2	基于时空感知深度网络的视频异常事件识别的算法实现 .....	158
8.2.1	时空感知深度网络 .....	158

8.2.2	时空金字塔池化 .....	159
8.2.3	基于时空感知深度网络的异常事件识别 .....	160
8.3	实验结果与分析 .....	161
8.3.1	数据集 .....	161
8.3.2	评价指标 .....	162
8.3.3	旅游景区数据集的实验结果与分析 .....	162
8.3.4	Avenue 数据集的实验结果与分析 .....	164
	本章小结 .....	166
	参考文献 .....	167
<b>第 9 章</b>	<b>视频数据去噪和超分辨率重建系统 .....</b>	<b>168</b>
9.1	引言 .....	168
9.2	系统总体设计 .....	168
9.3	系统详细设计 .....	170
9.3.1	视频预处理模块详细设计 .....	170
9.3.2	视频超分辨率重建模块详细设计 .....	171
9.3.3	性能评价模块详细设计 .....	171
9.4	系统主要功能模块的实现 .....	172
9.4.1	视频预处理模块的实现 .....	172
9.4.2	视频超分辨率重建模块的实现 .....	173
9.4.3	性能评价模块的实现 .....	176
	本章小结 .....	178
<b>第 10 章</b>	<b>旅游景区视频异常事件检测与识别系统 .....</b>	<b>179</b>
10.1	系统需求分析 .....	179
10.2	系统总体设计 .....	180
10.3	系统功能模块的详细设计 .....	182
10.3.1	视频特征提取模块 .....	182

10.3.2	异常事件检测模块 .....	182
10.3.3	异常事件识别模块 .....	183
10.4	系统的实现与测试 .....	184
10.4.1	视频特征提取模块功能的实现 .....	184
10.4.2	异常事件检测模块功能的实现 .....	186
10.4.3	异常事件识别模块功能的实现 .....	186
10.4.4	系统测试 .....	190
	本章小结 .....	191

# 第 1 章

## 绪 论

---

### 1.1 研究背景与意义

近年来,伴随着人口密度的日益增长,各种社会安全问题也日益突显,拥堵、踩踏、打架斗殴等异常突发事件频发,这给公共安全管理带来了极大的威胁和挑战。如上海外滩陈毅广场拥挤踩踏事件、华山游客被捅事件等,这些事件的发生均给公共安全敲响了警钟,引起了社会各界和相关监管部门的极大重视。近年来,视频监控系统被广泛部署在校园、景区、机场、火车站、高速公路、停车场、商场、办公室等公共场所,以加强保障公共安全。通过对这些公共场所的视频全天候、全方位的监控,利用视频智能处理和挖掘技术对公共场所进行实时监控和异常事件检测与识别,对于异常突发事件的及时预测和预警,以及公共安全的保障,具有重要的研究价值和广阔的应用前景。

随处可见的视频监控设备使相关部门可以监控并记录相应区域的现场情况,但是这需要耗费大量的人力物力,而且没有完全发挥出监控的作用。目前传统视频监控系统有如下三个弊端。(1)在需要对监控视频实时监测的场景下,传统视频监控系统只能用人眼进行监测,这是一项非常耗费人力的任务。在长时间查看监控视频的情况下,人眼极易容易产生视觉疲劳,多种人为因素的干扰直接导致监测到有效信息的可靠性和效率都很低。(2)很多监控系统处于“只记录不判断”的模式。在异常事件发生后,相关部门需要对监控视频进行调查取证,这种方式虽然能针对性地调取监控视频进行查证,但存在很大的信息滞后性,无法实时地对监控视频中的异常信息做出响应,也就无法从根源上降低异常事件的发生所带来的损失和影响。(3)监控系统的安装范围越来越广,传

统的视频监控系统很难从海量的视频数据中智能地提取出有效信息,导致资源利用率低下。因此,利用视频智能处理和挖掘技术,构建能够自动检测、自动识别、自动定位、自动报警的智能监控系统是发展的必然趋势。

通过智能视频数据处理和挖掘技术,对监控视频中的异常事件模式信息进行提取和分析,可以从海量的视频数据中精准地筛选出有效的数据,自动检测出异常事件的发生,定位出异常事件发生的位置,并识别出异常事件的类型。这样的智能视频监控系统实现了实时的监控,能在检测出异常事件时及时发出报警,提醒相关部门及时做出响应。近年来,视频中的异常事件检测与识别技术受到国内外学术界和产业界的广泛关注,许多有效的异常事件检测和识别算法都相继被提出,为智能监控系统的建设作出巨大的贡献。目前,随着人工智能和大数据技术的日益发展,视频监控系统已在智能交通、智慧旅游、智慧城市、智慧校园、智慧医疗、智能家居等领域取得了显著成效。虽然目前面向视频的异常事件检测与识别方法在简单的场景下效果较好,然而在复杂的运动场景下往往存在着检测率低、误检率较高的问题,背景混杂、目标间以及目标与背景间互相遮挡、视角变化、光照变化等复杂运动场景给面向视频的异常事件检测与识别带来了新的研究挑战,因此需要进一步研究对复杂场景具有较好鲁棒性的目标行为特征描述以及高效异常事件检测与识别模型和算法,提升其在复杂场景下的检测与识别性能,为异常突发事件的预测、预警以及应急决策提供有力的技术支撑,保障公共安全,最终实现智慧安防的目标。

此外,视频监控设备所拍摄的视频数据易受噪声干扰、光照变化、运动或光学模糊、下采样等多种降质因素的影响,视觉分辨率、细节清晰度和对比度较低,严重影响监控视频中异常事件检测和识别的精准度。视频超分辨率重建和去噪技术能够通过推断低分辨率视频序列中丢失的高频细节信息,重建出高质量的细节更为清晰而丰富的高分辨率视频序列。因此,研究高效且鲁棒性强的视频超分辨率重建和去噪算法对于提升视频异常事件检测和识别的精准度,具有重要的研究意义。

## 1.2 国内外研究现状

### 1.2.1 视频超分辨率重建方法的研究现状

超分辨率重建(SR, super-resolution)<sup>[1,2]</sup>是指通过融合多帧相互间存在全局或局部

位移、信息互补的低分辨率图像序列或视频序列,以获取视觉分辨率更高、质量更佳、细节信息更为丰富的图像、图像序列或视频的过程。超分辨率重建的概念最早由 Tsai 和 Huang<sup>[3]</sup>在题为“Multi-frame image restoration and registration”的文章中提出,并从此引起了国内外学术界和工业界的广泛关注和深入研究,特别是近年来,它一直是热门的研究方向。目前超分辨率重建研究主要分为频率域方法和空间域方法两大类。

频率域方法通过在频域内消除频谱混叠来提升图像分辨率。Tsai 和 Huang 最早提出的基于傅里叶变换的卫星图像超分辨率重建方法就是一种频率域方法,采用频率域逼近的策略对全局平移的多幅低分辨率图像进行超分辨率重建。然而,该方法的退化模型没有考虑模糊和噪声等因素,随后 Kim 等<sup>[4]</sup>对该方法进行了改进,针对模糊、噪声干扰和全局平移等问题提出了图像序列的频率域递推重构算法。虽然频率域方法机理简单,但由于没有充分利用图像的先验信息,且仅适用于全局平移运动模式,因而该类算法具有一定的局限性。

相对于频率域方法,空间域方法<sup>[5]</sup>具有较强的灵活性。空间域的超分辨率重建方法主要有基于插值的超分辨率重建方法、基于学习机制的超分辨率重建方法、基于运动估计的超分辨率重建方法以及基于时空相似性的超分辨率重建方法等。基于插值的超分辨率重建方法首先采用一个基函数或者插值核将低分辨率图像像素点映射入高分辨率图像栅格中,然后进行去模糊和去噪处理,获得最终的重建图像。基于插值的方法简单快速,往往能使平滑区域取得较好的重建效果,但是由于该方法没有充分利用多帧图像间的互补冗余信息为低分辨率图像提供新的图像信息,丢失的细节信息也不能被恢复,而且经常引入明显的锯齿效应,所以插值重建后的图像视觉效果往往不够理想。下面将重点对其他几种比较热门的方法进行综述。

### 1. 基于学习机制的超分辨率重建方法研究现状

基于学习机制的超分辨率重建方法是近些年的研究热点。该方法通过利用低分辨率(LR)和高分辨率(HR)块对作为训练数据集来学习 LR 和 HR 块间的关联映射关系,或者利用一个 LR-HR 过完备字典对作为先验知识,并以此来预测 LR 图像或视频帧中丢失的高频细节信息,获取 HR 图像或视频帧。基于学习的超分辨率方法主要包括基于稀疏编码的方法<sup>[6]</sup>、基于邻域嵌入的方法、基于回归的方法以及基于深度学习的方法等。Dong 等<sup>[7,8]</sup>通过综合学习局部和非局部稀疏性约束,提出了一种更为精确的集中化稀疏表示模型,并将其应用于超分辨率重建领域,获取了较好的效果。Yang 等<sup>[9]</sup>提出了一种基于学习的超分辨率重建框架,利用图像的稀疏表示和支持向量回归模型进行自学习,

实现重建。Zhou 等<sup>[10]</sup>通过字典学习提出了一种基于稀疏表示的超分辨率重建方法。近年来随着压缩感知理论<sup>[11]</sup>的兴起,在学习类的方法中还衍生出了一种基于压缩感知的超分辨率重建方法<sup>[12,13]</sup>。这类方法首先需要学习和训练样本图像集合,然后利用该样本集合重建图像。

Gao 等<sup>[14]</sup>提出了一种基于稀疏邻域嵌入的超分辨率重建方法。Chang 等<sup>[15]</sup>提出了基于流行学习的超分辨率重建算法,分别将高分辨率和低分辨率图像视为高维空间流行和低维空间流行,根据流行邻域在低维和高维空间的一致性,由低维空间结构重建高维空间结构。受局部线性嵌入(LLE)的启发,Chang 等<sup>[16]</sup>首次提出了一种基于邻域嵌入的超分辨率重建方法,通过学习从 LR 图像块流行到 HR 图像块流行的局部几何关系之间的映射,来实现 HR 图像块的重建。随后大量相关方法相继被提出,并具有较好的性能。Gao 等<sup>[17]</sup>基于稀疏邻域嵌入对该方法进行了扩展,通过自适应选取每个 LR 块的  $K$  近邻邻域(KNN)来利用 HOG (histograms of oriented gradients) 特征描述局部结构信息。Timofte 等<sup>[18]</sup>提出了一种新颖的基于邻域嵌入回归的快速超分辨率方法,该方法通过计算与字典元素的相关性而非欧几里得距离来选取最近邻域。然而当处理大规模的训练块时,该方法搜索最近邻的时间开销较大,并且需要大量内存空间。此外,随着超分辨率倍数的递增,LR 图像块和相应的 HR 图像块之间的相关性会变得十分模糊<sup>[19]</sup>。

文献[20]提出了基于人工神经网络和分类的视频超分辨率算法,利用人工神经网络来学习 LR 和 HR 视频帧间的复杂时空关系,并基于训练得到的网络参数来预测 LR 帧的目标 HR 估计。文献[21]提出了基于局部数据驱动的 3D 核回归(3DKR)方法来估计目标 HR 图像。文献[22]提出了贝叶斯自适应的视频超分辨率重建方法,通过采用最大后验概率估计来迭代估计目标 HR 估计,该方法在精确的光流运动估计下往往可获得较好的 HR 细节重建效果。Yang 等<sup>[23]</sup>利用  $k$ -means 算法<sup>[24]</sup>将图像空间划分为若干子空间,并分别在划分后的子空间内构建 LR-HR 关联映射函数来实现超分辨率重建。

基于深度学习的超分辨率方法是近年发展起来的热门研究方向。文献[25],[26]提出了用于超分辨率重建的深度卷积神经网络模型(SRCNN),通过构建深度网络结构来学习 LR 和 HR 图像块间的端到端关联映射,该方法与传统的稀疏编码方法相比,能挖掘出更多的细节信息用于超分辨率处理。Wang 等<sup>[27]</sup>提出了一种基于深度网络和稀疏先验的超分辨率重建方法,该方法综合深度学习和稀疏编码的优点取得了较好的超分辨率性能,并且利用级联网络增强了算法对任意的超分辨率倍数的适应能力。文献[28]提出了一种新的基于深度学习的快速超分辨率重建算法,首先生成高分辨率候选块集,获取局部高分辨率结构,然后利用深度卷积神经网络来选择最优的候选块用于生成最终的高

分辨率块。该方法不依赖精确的运动估计,能够适用于复杂的局部运动。

基于学习的超分辨率重建方法需要构建一个足够大的训练样本数据集,而且在训练图像的内容与待重建图像的内容相似时往往能取得更好的重建效果。基于学习的超分辨率方法能够重建出新的目标细节信息,但往往由于不相关实例的匹配或者特征映射过程中出现的错误而在重建结果中产生视觉瑕疵。基于实例学习的超分辨率方法有助于从外部实例中产生合理有用的细节信息,但是容易产生额外的干扰细节瑕疵,并且难以抑制锯齿现象的产生。基于重建的超分辨率方法有助于保护锋利的边缘,但是容易模糊微小的细节信息,导致视觉效果不自然。综合基于实例学习和基于重建的超分辨率方法的优势,Zhang 等<sup>[29]</sup>提出了由粗到精的超分辨率统一框架,其优势体现在:(1)避免因基于学习的超分辨率方法产生的无干扰细节瑕疵;(2)有效地恢复出基于重建的超分辨率方法中平滑掉和丢失的高频细节信息。基于同种思想,Yu 等<sup>[30]</sup>提出了一种用于单帧图像超分辨率的统一学习框架,综合利用基于学习和基于重建的超分辨率方法的互补特性,利用非局部均值滤波增强边缘并且有效抑制瑕疵,进一步提升超分辨率性能。

## 2. 基于运动估计的超分辨率重建方法研究现状

依赖精确运动估计的超分辨率重建思路先采用块匹配、光流等运动估计方法进行帧间的亚像素运动估计<sup>[31, 32]</sup>,并基于运动估计所获取的运动向量实现时空配准,然后再通过帧间信息融合来获取重建后的高分辨率图像序列。该类方法目前的研究主要集中在图像序列的配准以及超分辨率重建算法两个方面。图像序列的配准是确保重建效果的一个关键因素,其精度直接影响重建的质量。许多学者针对这一问题展开研究,试图提升图像序列间配准算法的精度。图像序列间的配准方法主要有基于像素区域的方法<sup>[33]</sup>和基于特征匹配的方法<sup>[34, 35]</sup>。

在超分辨率重建算法方面,目前已有的研究方法主要包括最大后验概率(MAP)法、凸集投影法(POCS)、迭代反向投影法(IBP)、极大似然估计法(ML)、非均匀插值法、正则化法以及如上几种方法的组合等。其中 MAP 法和 POCS 法相对被采用得较多。MAP 法是一种基于随机概率论的方法,采用目标高分辨率图像的先验概率在贝叶斯框架下实现超分辨率重建。Liu 等<sup>[36]</sup>提出了一种基于贝叶斯框架的自适应视频超分辨率重建方法,将高分辨率图像重建、光流估计、噪声估计和模糊核估计集成在一个统一的框架下。Chen 等<sup>[37]</sup>提出了一种基于广义高斯马尔科夫随机场(GGMRF)的超分辨率重建方法,获得了较好的重建质量和边缘保持效果。POCS 方法<sup>[38]</sup>是一种基于集合论的方法,这类方法原理简单且容易引入先验知识,但也存在解不唯一以及收敛速度较慢的问题。

### 3. 基于时空相似性的超分辨率重建方法研究现状

基于运动估计的超分辨率重建算法对帧间运动估计和配准的精度十分敏感,往往依赖精确的亚像素运动估计,因而不适用于一些复杂的摄像机/场景运动模式。为解决这一问题,近几年兴起和发展了一种基于时空相似性的超分辨率重建方法,是一种模糊配准机制下的超分辨率重建思想,为视频的超分辨率重建问题提供了一种崭新的思路,且目前已被验证是基于多帧的视频超分辨率重建算法中的一种优秀算法。基于这一新思想的超分辨率重建方法最早始于 Protter<sup>[39]</sup>提出的一种基于 3D 非局部均值滤波的超分辨率重建方法。非局部均值滤波(NLM)最早是在单帧图像去噪领域被提出的<sup>[40]</sup>。随后 A. Buades<sup>[41]</sup>将 NLM 扩展到 3D 域,并在此基础上提出了一种基于 3D NLM 的视频序列去噪方法。受这一思想的启发,Protter 将 3D NLM 拓展至超分辨率重建领域,这是一种基于概率运动估计的模糊配准机制下的超分辨率重建策略,图像配准和重建的过程通过非局部邻域像素的相似性匹配和相似像素的加权平均同步完成。这种基于非局部相似性的超分辨率重建机制充分利用相邻帧的高分辨率信息对目标帧图像进行修正,利用图像非局部相似性引入图像先验知识和相邻帧的细节信息,在无须进行精确的帧间运动估计的情况下,就可以将相邻帧信息运用到目标帧超分辨率重建中,并取得较好的效果。然而,当低分辨率图像序列中存在或缺失一些对象,或者存在不同角度旋转时,帧间信息的关联性就很微弱,因此此时该机制无法充分利用低分辨率图像之间的相似信息实现有效的重建。为解决这一问题,Gao 等<sup>[42]</sup>提出了一种基于 Zernike 矩的超分辨率重建方法,通过引入具有较好旋转、平移和尺度不变特性的 Zernike 矩特征来充分利用图像序列中的细节信息,从而获取高质量的重建效果。然而,基于 Zernike 矩的模糊配准机制的时间复杂度是相当高的,其时间代价主要表现在相似性权重的计算过程中,尤其是随着参与超分辨率重建的低分辨率视频帧数目和大小的增多,以及重建视频帧超分辨率倍数的增大,这种时间代价的累积是十分严重的。

基于模糊配准机制下的重建策略催生了超分辨率重建领域新的研究流派,成为国内外一大研究焦点。近年来,学习图像结构已被验证对于构建超分辨率的重建约束具有重要作用,涌现出了一系列的相关算法,如非局部相似性和局部结构规律性的互补特性<sup>[43,44]</sup>,以及稀疏性和结构规律性的综合<sup>[45]</sup>。Takeda 等<sup>[46]</sup>提出了一种局部自适应 3D 迭代核回归机制,通过挖掘图像序列内的时空邻域关系来实现超分辨率重建,从而有效避免了精确的亚像素运动估计。Zhang 等<sup>[47]</sup>提出了一种非局部核回归(NL-KR)超分辨率重建方法,充分利用了图像局部结构信息和非局部相似性进行帧间信息融合重建,后