

智能视觉监控技术研究进展

王素玉 沈兰荪

(北京工业大学信号与信息处理研究室, 北京 100022)

摘要 新一代智能视觉监控技术的研究是一个极具挑战性的前沿课题,它旨在赋予监控系统观察分析场景内容的的能力,实现监控的自动化和智能化,因而具有巨大的应用潜力。视觉监控系统的智能化分析过程由运动目标检测、分类、跟踪和视频内容分析等几个基本环节组成,其中视频内容分析又包括异常检测、人的身份识别以及视频内容理解描述等。本文在总结以上有关关键技术研究进展的基础上,进一步提出将超分辨率复原技术引入视觉监控领域,介绍了超分辨率复原的主要算法及其在智能视觉监控中的应用。

关键词 智能视觉监控 目标检测 目标跟踪 步态识别 行为理解和描述

中图分类号: TP277 TP391 **文献标识码**: A **文章编号**: 1006-8961(2007)09-1505-10

Intelligent Visual Surveillance Technology: A Survey

WANG Su-yu, SHEN Lan-sun

(Signal & Information Processing Laboratory, Beijing University of Technology, Beijing 100022)

Abstract The research on new generation of Intelligent Visual Surveillance is a new arising font field with many challenges. It aims to endow surveillance systems with the ability of analyzing scene contents, make surveillance tasks fulfilled automatically and intelligently, which has a large potential of application. There are four main issues in the process of intelligent analysis of typical visual surveillance system, detection, classification, tracking of objects and analysis of video contents. According to the real applications, video contents analysis can be abnormal detection, person recognition, understanding and description of video contents etc. In this paper, in addition to make a survey on the research progresses of the key technologies of visual surveillance, the concept of super resolution restoration is further introduced to this field to enhance the quality of the surveillance sequence. The main algorithms of super resolution restoration and its applications in visual surveillance are discussed in detail.

Keywords intelligent visual surveillance, object detection, object tracking, gait recognition, video content understanding and description

1 引言

自9.11事件以后,如何对国家重要安全部门和敏感的公共场合进行全天候、自动、实时的监控,已成为世界各国高度重视的一个全球性问题。在这样的背景下,视频监控技术得到了广泛的应用,即通过分散设置的摄像机记录监控场景,并集中显示,使得

监控人员可以实时了解发生的事件,以及时做出反应。但是,有关研究已经表明^[1],仅依靠视频监控中的人眼检测,即使是专业操作人员也难以构成真正有效的安全系统。因为只要专注于监控屏幕超过20min,绝大多数工作人员的注意力都会降低至无法接受的水平。

显然,传统的基于人工操作的视频监控不能适应实际应用的需要,开发新一代的监控技术——智

基金项目: 国家自然科学基金项目(90304001, 60472036); 北京市自然科学基金项目(4052007); 武器装备预研基金项目(51434050105QT0101); 北京市教委基金项目(KM200410005022); 北京市科技新星计划基金项目(2005B08)

收稿日期: 2006-04-29; **改回日期**: 2006-06-07

第一作者简介: 王素玉(1976~),女。现为北京工业大学信号与信息处理研究室博士研究生。主要研究方向为图像/视频信号处理。
E-mail: auyuwang@emails.bjut.edu.cn

能视觉监控技术的研究,近年来已受到广泛的关注。即赋予监控系统观察分析场景内容的能力,使其更加智能化,能够在几乎不需要人为干预的情况下,对摄像机拍摄的视视频序列进行自动分析,并及时做出反应。

新一代的智能视觉监控技术涉及图像处理、图像分析、机器视觉、模式识别、人工智能等众多研究领域,是一个跨学科的综合问题,也是一个极具挑战性的前沿课题。图 1 给出了一个典型的视觉监控系统的基本工作流程。摄像机获取视频序列后,首先通过图像恢复或超分辨率复原技术提高图像质量,然后对场景中的目标进行检测、分类和跟踪,进而实现视频内容的分析理解,包括场景中的异常检测、人的身份识别以及视频内容的理解描述等。

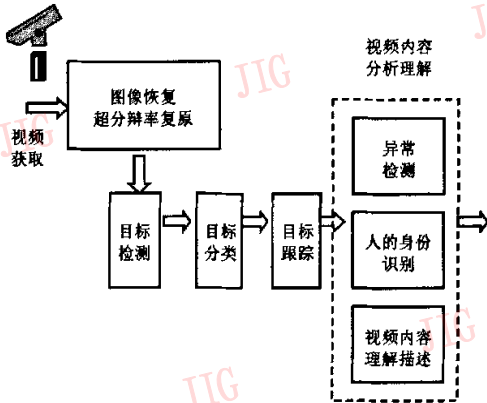


图 1 视觉监控系统工作流程

Fig. 1 Process flow of visual surveillance system

2 目标检测

目标的实时检测处于视觉监控技术的最底层,是各种后续处理如目标分类、行为分析理解的基础。人们希望设计能适用于各种监控环境,不受光照、天气等外界因素影响的目标检测(object detection)算法。但这种算法难免复杂度大,现有一些算法大多是针对某一类问题提出的,主要包括背景减法、相邻帧差法和光流法等^[2]。

2.1 背景减法

背景减法(background subtraction)是目前运动目标检测的主流方法^[3],其基本思想是将当前每一帧图像与事先存储或实时获取的背景图像相减,计

算出与背景偏离超过一定阈值 T 的区域作为运动区域。该算法实现简单,相减结果直接给出目标的位置、大小、形状等信息,能够提供关于运动目标区域的完整描述,特别是对于摄像机静止的情况,背景减法是实现运动目标实时检测和提取的首选方法。

背景减法实现的关键是背景模型的获取和更新。背景获取算法通常要求在场景中不存在运动目标的情况下获取背景图像,更新过程使背景能够适应场景的各种变化和干扰,如外界光线的改变,背景中对象的扰动和固定对象的移动,阴影的影响等。

一种典型的背景建模方法是用混合高斯模型描述背景图像像素值的分布^[4],目标检测过程中判断图像的当前像素值是否符合该分布,若是被判为前景点,否则为背景点。同时根据新获取的图像,对背景图像参数进行自适应更新。该方法能够可靠处理光照变化、背景混乱运动的干扰以及长时间的场景变化等。文献[5]在此基础上,对背景、静止目标和运动目标三者采取不同的更新策略,以减弱背景更新过程中运动目标对背景的影响。

另外,Karman 等^[6]和 Kilger^[7]提出基于卡尔曼滤波的自适应背景模型,以适应天气和光照的变化;视觉监控系统 W^4 ^[8]中,利用最小、最大强度值和最大时间差值为场景中的每个像素进行统计建模,并对模型进行周期性的更新。

2.2 相邻帧差法

相邻帧差法(temporal differencing)利用序列中连续两帧或几帧图像间的差异进行目标的检测和提取。由于相邻帧的时间间隔一般较短,因此算法对场景中的动态变化不太敏感,具有较强的自适应性。但该方法一般不能完全提取出所有相关的特征像素点,在运动实体内容容易产生空洞现象。累积图像差分法^[9]进一步计算一阶和二阶差分图像,利用时间序列图像的历史积累信息,能够适应低对比度有噪声的时间序列,判断复杂情况下目标运动的多种状态。

以上两种目标检测方法适用于背景固定不变的情形,对于背景发生运动的情况,需要提前对由摄像机引起的背景移动进行补偿^[10]。

2.3 光流法

光流法(optical flow)是通过研究图像序列的光流场实现运动目标检测的一种方法。光流场是空间运动物体被观测表面上的像素点运动产生的瞬时速度场,它包含了物体编码结构和动态行为的重要信息。光流法实际上是通过光流场近似计算图像序列

中不能直接得到的运动场,根据运动场的特征对目标进行检测。

光流场的计算始于1981年,Horn和Schunck^[11]在相邻图像间的时间间隔很小,图像灰度保持不变的前提下,推导出灰度图像光流场的基本约束方程

$$\frac{\partial E}{\partial x}u + \frac{\partial E}{\partial y}v + \frac{\partial E}{\partial t} = 0 \quad (1)$$

该方程作为仅有的约束,不能解出相互独立的两个速度分量 u 和 v ,因此,如何为光流场添加合适的约束条件并从中解出 u 、 v 的解,是光流场算法的核心问题。

Horn和Schunck^[11]的方法是依据同一运动物体引起的光流场应该是连续的、平滑的,将全局平滑约束作为光流场的附加约束。由于未能考虑到实际运动物体边缘处光流矢量的不平滑性,该方法对于边界点计算结果很差。针对这一问题, Lucas等人^[12]提出一种局部平滑约束条件,即假设在一个足够小的空间邻域上,各像素点的光流可以用其中一个像素点的光流来近似,采用加权最小二乘技术推导出另一个约束条件。以Nagel^[13]为代表提出了“有向平滑”的概念, Ote等人^[14]则假定光流在局部空间和时间轴是线性变化的,推导局部约束条件。这些以基于梯度的光流场基本等式为基础,围绕寻求附加约束条件展开的光流计算方法,被统称为基于梯度的方法。

另外,也有研究抛开基本等式,提出基于区域匹配的方法、基于能量的方法、基于相位的方法等。Barron等人^[15]对上述几种典型的光流场计算方法从多个方面进行了分析比较,认为基于梯度的局部最优化方法总体性能上是最优的。

光流法的最大优势在于,它不需要预先知道场景的任何信息即能检测出独立运动的目标,并可用于摄像机运动的情况。但多数光流法计算复杂耗时,除非有特殊的硬件支持,否则很难实现实时检测。

另外,近年来相关文献中也有一些新的目标检测方法提出,包括基于小波分解的方法^[16]、基于神经网络的方法^[17]、基于知觉组织模型的方法^[18]等。

3 目标分类

目标分类(object classification)的两个关键问题

是特征量的提取和分类器的构造。

用于目标分类的特征主要包括空间特征和时间特征。空间特征是单帧图像里的目标相关的一些特征,如目标的形状、大小、位置等。时间特征与跟踪过程相联系,如目标大小的变化、运动的速度等。Lipton等人^[19]采用分散度和面积信息等空间特征对行人和车辆进行区分。Cutler等人^[20]根据人体运动的周期性特点,利用运动目标随时间变化的自相关函数将行人与场景中的其他运动目标区别开来。

分类器的设计方面,常用的分类器有贝叶斯分类器,神经网络分类器^[21]以及新兴的支撑向量机(SVM)^[22]分类器等。其中,支撑向量机及各种改进算法,因其在小样本分类方面的优势,得到了广泛的关注。

4 目标跟踪

运动目标的跟踪过程,就是依据目标及其所处的环境,选择能够唯一表示目标的特征,并在后续帧中搜索与该特征最匹配的目标位置的过程。它不仅是目标运动分析和场景分析的主要数据来源,也为目标的检测和识别提供帮助。目前关于目标跟踪(object tracking)的算法有很多,各种算法的根本区别在于选择何种特征对目标进行描述以及采用何种搜索匹配算法。这里介绍几种常用的目标跟踪算法,包括基于特征的方法、基于3维模型的方法、基于活动轮廓模型的方法等。

4.1 基于特征的跟踪方法

基于特征的方法通过各帧间运动目标个体特征的匹配实现跟踪。用于匹配的个体特征有目标的周长、面积、质心等全局特征和线段、角点等局部特征,也可将多个特征组合成更高级的特征。Polana等人^[23]采用质心作为基本特征实现了对人的运动跟踪,即使在发生遮挡的情况下,只要质心的速度可以区分,仍能实现有效的跟踪。Daniel等人^[24]利用了目标的颜色特征,对场景中的多个目标进行区分和跟踪。在汽车等刚性物体的跟踪中,角点是被经常采用的一种重要特征。

基于特征的算法简单,易于实时处理,且能够实现多个目标的跟踪。但由于是对3维平面内的特征进行分析,因而受观测视角的影响较大,跟踪识别率较低。

4.2 基于 3D 模型的跟踪

基于 3D 模型的跟踪方法是依据一定的投影规则,将预先建立的 3 维目标模型投影到 2 维平面,与平面图像中的目标进行匹配以实现跟踪。对于车辆等刚性目标,目标模型通常采用 3D 线框模型^[25]。Lou 等人^[26]进一步在此基础上提出地面约束 (ground-plane constraint, GPC) 条件,即车辆是限制在地面上运动的,以减少车辆的可能姿态数。基于模型的人体跟踪实际上是将序列中不同视角、不同姿态的人用一组随时间变化的姿态参数描述。需要预先建立人体的 3D 模型和运动模型,人体模型可采用骨架模型、椭圆模型和多边形模型等多种形式^[27]。运动模型包含关于人体运动的先验知识,对于运动参数的预测分析具有重要的意义。

基于模型的方法适用于刚性和非刚性目标,不易受观测视角的影响,具有较强的鲁棒性,但算法复杂度较高,实时性差。

4.3 基于主动轮廓 (snake) 模型的跟踪

主动轮廓 snake 模型^[28]自提出以来一直是图像分割和目标检测、跟踪领域的研究热点。snake 模型是在图像域内定义的可变形曲线,通过能量函数的最小化,调整 snake 的自然形状使其与对象轮廓一致。

snake 技术可以处理任意形状物体的任意形变。首先将分割得到的物体边界作为跟踪的初始模板,然后确定表征物体真实边界的目标函数,并通过降低目标函数值,使初始轮廓逐渐向物体的真实边界移动。跟踪过程实际上是解的寻优过程,其带来的计算量比较大,而且由于 snake 模型的盲目性,对于快速运动的物体或者形变较大的情况,跟踪效果不够理想。Park 等人^[29]提出一种带有方向性的 snake 模型,即根据边界曲线特征与图像特征,调节 snake 曲线的发展方向,防止错误的边缘定位。

目标跟踪过程中,直接对场景中的所有内容进行匹配计算,寻找最佳匹配位置,需要处理大量的冗余信息,造成算法复杂度的增加。采用运动估计方法对未来时刻目标的位置状态进行估计假设,缩小目标搜索范围具有重要意义。常用的运动估计方法有卡尔曼滤波法、扩展卡尔曼滤波法^[30]以及蒙特卡罗方法^[31]等。

5 视频内容分析

视频内容的分析是在完成运动目标的检测、分

类和跟踪后,对其内容的智能化分析和理解,并做出相应的判断。它处于智能监控技术的高级阶段,是真正实现监控智能化的关键。实际应用中,视频内容分析包括异常检测 (abnormal detection)、人的身份识别以及视频内容理解描述等几个方面。

5.1 异常检测

异常情况实际上是一个相对概念,并没有一般的定义形式。典型的异常情况主要有^[32,33]:

(1) 用户定义的异常情况,包括

① 场景目标是否运动及其运动特性。如博物馆中监控文物的移动;机场长时间未被移动的行李物品等。

② 异常目标。如机场跑道出现的除飞机以外的目标。

③ 异常行为。如停车场、车站等公共场合的可疑行为等。

(2) 非常规事件,即与特定场合的常规事件有显著区别的事件。如交通监控系统遵循“车辆走公路,行人走人行道”的规则,若检测到不符合该规则的事件均可被认为是非常规事件。

目前异常事件的检测方法主要可分为基于模型的方法和基于分类器的方法两种。

基于模型的方法强调对异常事件的检测而非分类,它对正常(异常)事件建立相应的统计模型,通过判断待检测事件是否符合该模型确定其是否为正常事件。该方法特别适合于其中一类样本不易获取的情况,只对易获取的一类样本建模,根据新的样本是否属于该类别确定检测结果。事件的建模是算法的关键,常用的模型有隐马尔可夫模型^[34]和混合高斯模型 (GMM) 等。

基于分类器的方法强调对场景中事件的分类,通过构建合理的分类器,实现对正常和异常事件的区分。如 Wu 等人^[35]采用支撑向量机分类器,实现了人群中奔跑、弯腰、招手等动作的有效检测。Tony 等人^[36]采用基于人工神经网络的方法,对停车场中人的可疑行为进行检测。

5.2 人的身份识别

人通常是监控场景的主体,人的身份识别 (person recognition) 对于场景内容的理解具有重要的意义。特别是在安全敏感场合 (如银行、机场等),非接触式远距离的身份识别技术备受关注,人脸和步态被认为是可供利用的两个重要特征^[37]。人脸识别技术涉及人脸的检测、跟踪、特征提取和识别等

4个方面,近年来得到了非常广泛的研究^[38,39]。本文着重讨论步态识别的相关技术及其研究进展。

步态识别(gait recognition)是生物识别技术的一个新兴领域,它旨在根据人们走路的姿势进行身份的认证。虹膜、指纹等生物特征通常要求近距离或接触性的感知(如指纹需要接触指纹扫描仪等),不能用于远距离的情况,而步态是可以远距离感知的,可以与监控系统更好的结合。

步态识别的基本过程如图2所示。首先进行步态的检测,将人体步态轮廓区域从背景图像中提取出来,然后选取一定的步态特征对其进行描述,并结合已存储的步态模式实现身份识别。步态的检测方法与运动目标的检测基本一致,关键在于步态的特征和识别。

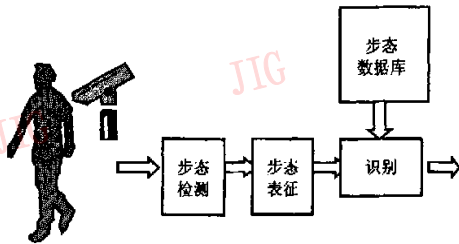


图2 步态识别系统示意图
Fig.2 Diagram of gait recognition

(1) 步态表征

步态表征又称为步态特征提取,就是采取一定的方法表示检测出的步态。其表征方法可分为基于模型和基于特征的两种。

基于模型的方法是将人用一定的模型表达,跟踪分析模型的参数进行步态的识别。例如 Amos 等人^[40]采用了如图3(a)所示的线性模型,其主要参数包括身高、头到髌骨的距离、髌骨到脚的最大距离以及两脚间的距离等静态参数。Lee 等人^[41]则将人的侧面二值化图像分为7个部分,用7个椭圆

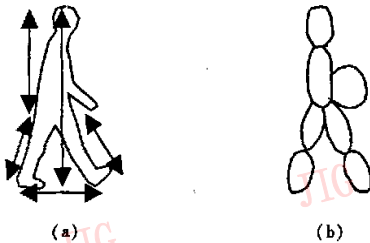


图3 基于模型的步态表征
Fig.3 Model-based gait representation

来表示,如图3(b)所示。基于模型的方法不受观测角度的影响,但是由于其依赖于对步态序列中特定参数的识别,对步态序列的质量要求比较高,另外参数的计算会受到行走过程中部位遮挡的影响。

基于特征的方法不需要假定人体模型,直接提取步态的特征数据用于识别,如侧面的轮廓、宽度等。文献[42]中利用行人的侧面轮廓特征,将2维轮廓形状转换为1维距离信号,然后通过特征空间变换(PCA)提取低维步态特征用于步态识别。

(2) 步态识别

由于步态数据通常是时变多变量数据,因此需要采用基于动态系统的识别方法,主要包括基于模板的方法和基于统计的方法^[43]。

基于模板的方法对每个人体样本建立一个或多个模板,识别过程实际上是特征数据与模板之间的匹配过程。为解决序列间运动快慢不一致情况下的匹配问题,通常采用动态时间规整(dynamic time warping, DTW)^[44]的方法,其基本原理如图4所示,将两个不同时间长度的运动特征模板 T_1 和 T_2 按照一定的曲线进行时间规整,使时间长度达到一致,然后再匹配。

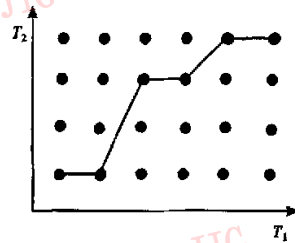


图4 动态时间规整 DTW
Fig.4 Dynamic time wrapping

基于统计的方法是动态系统模式识别中广泛采用的方法,其中最具有代表性的是隐马尔可夫模型(hidden Markov model, HMM)^[45]方法,其基本思想是用双重随机过程来描述一个模式:用一个马尔可夫链描述模式内部的状态序列,另一个随机过程描述状态与观测值之间的关系。每一个步态类别对应一个隐马尔可夫模型。隐马尔可夫模型的使用涉及训练和分类两个阶段,训练阶段制定 HMM 的隐藏状态数,并且优化相应的状态转换概率和输出概率;分类阶段计算该 HMM 产生所观察图像特征序列的概率。HMM 在学习能力和处理未分割的连续数据流方面具有显著的优越性,已被广泛用于步态识别中。

步态识别技术的研究目前仍处于探索性理论研究阶段,现有算法大多是在一定的限制条件下进行的,如假定人的行走方向与摄像机的镜头主轴方向垂直,利用人的侧面图像进行识别。研究步态识别中各种因素对识别结果的影响,减少约束条件,提高识别效率将具有重要的意义。

5.3 视频内容理解描述

视频内容的理解描述是指在对序列进行低级处理(目标检测、跟踪)的基础上,进一步对场景中的事件进行分析和识别,并用自然语言等加以描述。其中最典型的是对场景中的运动目标特别是人的行为的理解和描述。其实现过程首先是一个时变特征数据的分类过程:将未知测试序列的相关特征与一组预先存储的典型行为参考序列进行匹配,以确定其所属的类别;并在此基础上应用自然语言的概念,选择恰当的词语或短句对行为进行描述。

由于自然语言中,关于动作、事件和状态的描述有丰富的概念,现有的研究尚处于探索阶段,主要是针对某种特定应用场合中特定行为的描述。Remagnino 等^[46]采用两个贝叶斯网络对 3 维场景中的行人和车辆的活动状态进行文字性描述,其中一个网络用于描述车和行人的状态如静止、匀速或加速前进等,另一个网络用于描述两车或人之间的相互关系。Ram 等^[47]定义了一种事件本体论(ontology),通过简单事件合成的方法对物理世界中具有复杂时空关系的一般事件进行描述。

视频内容理解和描述本质上是一个人工智能问题,如何借助先进的视觉算法和人工智能等领域的成果,将现有的简单行为识别和语义理解推广到更为复杂场景下的自然语言描述,建立能够不断积累知识、丰富经验、具有自主学习能力的系统将是该领域的主要发展方向。

6 超分辨率复原技术及其在视觉监控中的应用

在各种监控应用中,人们总是希望获取的视频序列具有尽可能高的空间分辨率。采用超分辨率复原算法,可以在不改变原有系统硬件设备的前提下大大提高视频序列的质量。

6.1 超分辨率复原技术

图像获取过程中,受成像条件和成像方式的限制,成像系统并不能获取原始场景中的所有信息。成像过程往往会受到光学模糊、运动模糊、变形和噪声等诸多因素的影响,造成获取图像质量的下降。超分辨率复原技术采用信号处理的方法,在改善图像质量的同时,试图恢复成像系统截止频率之外的信息,是提高空间分辨率改善图像质量的有效方法。这一技术的最大优点在于成本低,现有的低分辨率成像系统仍可以使用。实际上,超分辨率复原技术和图像恢复技术在理论上有着密切的联系,可认为是第二代图像恢复技术。

超分辨率复原的基本思想最初是由 Harris^[48]和 Goodman^[49]以单张图像复原的概念提出的。1984 年 Tsai 和 Huang^[50]开创了序列图像超分辨率复原的工作,利用多帧低分辨率图像中不同但相互补充的信息以及图像的先验信息,恢复高分辨率图像。

6.1.1 观测模型

建立一个联系原始高分辨率图像和实际获取的低分辨率图像的观测模型是实现图像超分辨率复原的关键,一个典型的观测模型^[51]如图 5 所示。

设理想的高分辨率图像为 X ,第 k 个低分辨率图像为 Y_k ,这样,变形、模糊、降采样作用在图像 X 后,就得到退化的低分辨率图像,假定存在加性噪声的影响,观测模型可以用下式表示为

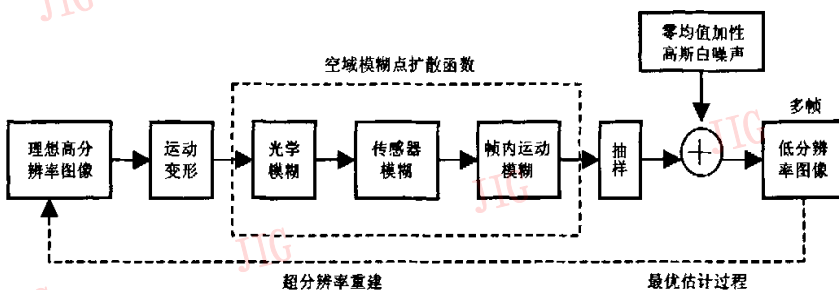


图 5 图像观测模型

Fig. 5 Observation model of image

$$Y_k = DB_k M_k X + n_k \quad 1 \leq k \leq p \quad (2)$$

其中, M_k 为几何失真引起的变形矩阵, B_k 为点扩散函数或衍射极限引起的模糊矩阵, D 为降采样矩阵, n_k 为噪声。

图像的超分辨率复原过程,就是根据获取的低分辨率图像序列 Y 恢复理想的高分辨率图像 X 的过程。它往往不能满足一个或多个 Hadamard 条件,因而是一个病态求逆问题。解决问题的有效方法是正则化方法,即用图像的先验知识(局部平滑、边缘保持、正数、能量有限等)来约束解空间。换言之,图像复原是一个最小化代价函数的优化问题。

6.1.2 超分辨率复原算法

Tsai 和 Huang 等人^[50]最初提出的方法是通过在频域消除频谱混叠来恢复图像的高频信息,提高空间分辨率,其理论基础是傅里叶变换的移位特性。该方法理论简单,适合并行处理,但观测模型局限于全局位移和线性空间不变,难以将空域的先验知识考虑进来,目前已不成为主流技术。

空域方法可以将复杂的运动模型与相应的插值、迭代、滤波和重采样过程放在一起处理,更符合图像退化的复杂过程,因而具有较好的复原效果。空域方法主要包括非均匀空域样本内插法^[52]、迭代反投影方法^[53]、凸集投影法(POCS)^[54]、统计复原方法——最大后验概率(MAP)估计器^[55]和最大似然估计器(ML)、混合 MAP/POCS^[56]方法以及滤波方法等。Borman^[57]认为,在各种空域算法中,MAP方法和 POCS 方法是最有前途的两种。

(1) 凸集投影法^[54]

超分辨率解空间中的解有多个限制条件,如正定、能量有界、数据可靠、平滑等,通常是超分辨率图像的一些比较理想的性质。凸集投影法(projection onto convex sets)将每个限制条件定义为向量空间中的凸集合。超分辨率复原问题的解空间就是这些凸的限制集的交空间。POCS 是指从成像空间中任意一点开始投影定位到凸集的交集上的过程。

凸集投影法的原理直观简单,具有灵活的空域观测模型,特别是具有强有力的利用先验知识的能力。但由于其解空间定义为所有凸集约束的交集,通常不是唯一的,因而对初始值具有较强的依赖性。投影过程需要的运算量较大,这在一定程度上限制了 POCS 的应用。

(2) 最大后验概率方法^[55]

最大后验概率(MAP)方法是一种典型的统计

复原方法,将超分辨率复原问题看成一个统计估计问题。该方法的含义是在已知低分辨率视频序列的前提下,使出现高分辨率图像的后验概率达到最大。计算中可通过先验概率项直接引入先验约束(例如采用 Markov 随机场模型作为图像的先验信息),具有很好的利用先验知识的能力。

最大后验概率框架能够提供方便、直观且符合实际情况的成像模型,复原过程中能够充分利用先验知识,提高图像复原的效果,并且可以实现运动估计和超分辨率复原的同时求解,已成为目前最灵活、最有前景的算法之一。

(3) 基于学习的超分辨率复原算法

前述的各种超分辨率复原算法大多数是基于“重建约束”的,即当高分辨率图像经过变形和下采样后,应该可以生成低分辨率的输入图像。因此这些方法也被统称为基于重建的方法。随着放大因子的增加,重建约束所能提供的有用信息越来越少,使得放大图像会产生细节信息丢失,边缘模糊,边缘锯齿明显等问题。

为此,人们提出基于学习的超分辨率复原方法,通过学习低分辨率图像与高分辨率图像之间的关系,指导图像的超分辨率复原。Freeman^[58]首次提出基于例子插值的概念,将一组高分辨率图像进行降采样和模糊处理后,分成多个小块,建立一组低分辨率图像块和高分辨率图像块之间一一对应的图像库,采用马尔可夫网络学习训练库中与低分辨率图像各区域相对应的高分辨率图像的精细节,然后利用学习得到的关系预测任意输入低分辨率图像的细节信息,对其进行超分辨率。

该方法的依据是:很多看似不同的图像在高频细节上有很多相似的地方,利用训练图像学习得到的高频恢复信息,可以用于其他图像的超分辨率。

6.2 超分辨率复原技术在视觉监控中的应用

对于视觉监控系统,需要在给定低分辨率视频序列的前提下,快速估计出相应的高分辨率视频序列,用于视频内容的进一步分析处理。显然,开发适用于视觉监控系统的高效超分辨率复原算法具有重要意义。

将超分辨率复原技术引入视觉监控系统中可采用以下两种方案(图6)。

方案1 对获取的视频序列直接进行超分辨率复原,提高整体质量。为保证较高的复原速度,可采用滤波的方法,如自适应滤波、维纳滤波和卡尔曼滤

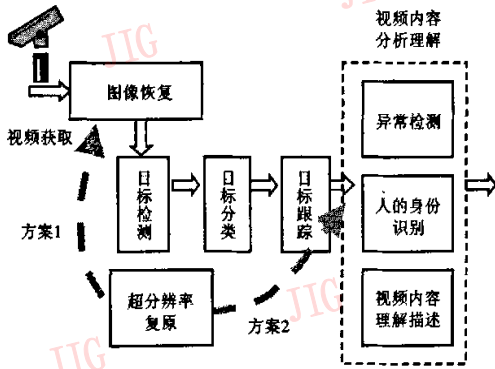


图6 超分辨率复原在视觉监控系统中的应用
Fig.6 Applications of super resolution restoration in visual surveillance

波等。自适应滤波算法^[59]等价于一种时间轴上不加任何先验约束的线性最小均方差估计器,根据低分辨率视频序列的当前帧和前一帧高分辨率图像估计快速估计出当前高分辨率图像帧。运动估计只在两个相邻帧间进行,运算速度很快。维纳滤波方法^[60]是采用加权近邻法将运动补偿后的低分辨率图像帧快速映射到一个均匀的高分辨率网格上,然后采用维纳滤波器完成去模糊和降噪过程。这些方法虽然能够实现较快速的超分辨率复原,但复原效果较差。

方案2 在完成序列中运动目标的检测、分类和跟踪后,对运动目标进行超分辨率复原。由于视觉监控系统通常背景相对固定,人们更为关心的是场景中的目标及其运动情况。因此采取基于目标的超分辨率复原,不仅可以有效提高感兴趣目标的空间分辨率,便于进一步分析处理,也节省了大量的运算量,允许采用较为复杂的复原算法。若能针对目标的特点,加入更为具体的先验约束,将有望进一步提高算法的性能。

7 展望

智能视觉监控技术是一个极具挑战性的研究领域,许多新的研究问题又相继提出,主要有:

(1) 超分辨率复原技术的利用。采用有效的超分辨率复原技术,提高监控视频的质量,对于提高后期视频分析算法的性能,降低复杂度具有重要意义。将其用于视频序列的存储,平时存储低分辨率序列,

需要时采用超分辨率技术恢复高质量图像,也是一种经济有效方法。

(2) 分布式视觉监控系统。采用多个摄像机从不同角度对同一场景进行监控,被认为是提高系统性能,特别是解决目标遮挡问题的有效方法。这将涉及到多源视频序列的融合问题。如何实现多角度信息的合理互补,有效提高监控系统的性能将是分布式监控系统需要解决的一个重要问题。

(3) 多波段成像监控系统。受多光谱遥感的启发,对监控内容在可见光、红外等多个波段同时成像,不仅可以使监控设备少受外界条件的影响,而且有利于各成像波段间信息的互补,实现对场景内容更为全面的分析。

(4) 步态识别与其他生物特征识别的结合。目前关于步态识别的研究尚处于起步阶段,其识别性能还有待进一步提高。将步态识别技术与其他生物特征如人脸甚至语音信息相结合,是提高识别效果的有效方法。

8 结语

智能视觉监控技术作为一个富有挑战性的新的研究方向,具有重要的理论研究意义和实用价值。目前,相关的研究尚处于探索研究阶段,要真正实现监控技术的智能化还需要大量的工作。随着相关技术的日益成熟,真正智能化的监控系统必将在未来的生产生活中发挥重要的作用。

参考文献(References)

- 1 Green Mary W. The Application and Effective Use of Security Technologies in U. S. Schools: A Guide for Schools and Law Enforcement Agencies[R]. NCJ 178625, Washington DC, USA: National Institute of Justice, U. S. Department of Justice, 1999.
- 2 Shen Lan-sun, Zhuo Li. Wavelet Coding and Network Video Transmission[M]. Beijing: Science Press, 2005. [沈兰荪, 卓力. 小波编码与网络视频传输[M]. 北京: 科学出版社, 2005.]
- 3 Hu Wei-ming, Tan Tie-niu, Steve Maybank. A survey on visual surveillance of object motion and behaviors[J]. IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, 2004, 34(3): 334-352.
- 4 Stauffer Chris, Grimson W E L. Adaptive background mixture models for real-time tracking[A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Vol. 2[C], Colorado, USA, 1999: 1063-1069.
- 5 Liu Ya, Ai Hai-zhou, Xu Guang-you. Moving object detection and

- tracking based on background subtraction [J]. *Information and Control*, 2002, 31(4): 315 ~ 319. [刘亚, 艾海舟, 徐光祐. 一种基于背景模型的运动目标检测与跟踪算法[J]. *信息与控制*, 2002, 31(4): 315 ~ 319.]
- 6 Karmann Klaus-Peter, Brandt Achim Von. Moving object recognition using an adaptive background memory[A]. In: Cappellini V. *Time-Varying Image Processing and Moving Object Recognition* [C], 2, Elsevier, Amsterdam, The Netherlands, 1990; 289 ~ 296.
- 7 Kilger M. A shadow handler in a video-based real-time traffic monitoring system [A]. In: *Proceedings of IEEE Workshop on Applications of Computer Vision* [C], California, USA, 1992; 11 ~ 18.
- 8 Haritaoglu Ismail, Harwood David, Davis Larry S. W⁴: who? when? where? what? a real time system for detecting and tracking people [A]. In: *Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition* [C], Nara, Japan, 1998; 222 ~ 227.
- 9 Jain R, Nagel H H. On the analysis of accumulative difference of picture from image sequences of real world scenes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979, PAMI-1(1): 206 ~ 214.
- 10 Wang Dong-zheng, Li Zai-ming. The technology of real-time detection and tracking for spatial video motion object [J]. *Signal Processing*, 2005, 21(2): 194 ~ 198. [王东升, 李在铭. 空域视频场景监视中运动对象的实时检测与跟踪技术[J]. *信号处理*, 2005, 21(2): 194 ~ 198.]
- 11 Horn Berthold K P, Schunck Brain G. Determining optical flow[J]. *Artificial Intelligence*, 1981, 17(1-3): 185 ~ 203.
- 12 Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision [A]. In: *Proceedings of DARPA Image Understanding Workshop* [C], Pittsburgh, Pennsylvania, USA, 1981; 121 ~ 130.
- 13 Nagel Hans Hellmut. Constrains for the estimation of displacement vector fields from image sequences [A]. In: *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-83)* [C], Karlsruhe, Gemany, 1983; 945 ~ 951.
- 14 Otte M, Nagel H H. Optical flow estimation: Advances and comparisons [A]. In: Eklundh J. *Proceedings of European Conference on Computer Vision*, volume 800 of LNCS-Series [C], Stockholm, Sweden, 1994; 51 ~ 60.
- 15 Barron J L, Fleet D J, Beauchemin S S, *et al.* Performance of optical flow techniques [A]. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [C], Illinois, USA, 1992; 236 ~ 242.
- 16 Stephane Mallat, Wen Liang Hwang. Singularity detection and processing with wavelets [J]. *IEEE Transactions on Information Theory*, 1992, 38(2): 617 ~ 643.
- 17 Jasdish Chandra Patra, Ferdinan Widjaja, *et al.* A fast neural network-based dection and tracking of dim moving targets in FLIR imagery [A]. *Proceedings of International Joint Conference on Neural Networks* [C], Montreal, Canada, 2005; 3144 ~ 3149.
- 18 Stilla U, Michaelsen E, Soergel U, *et al.* Perceptual grouping of regular structures for automatic detection of man-made objects [A]. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03)* [C], Toulouse, France, 2003, 6: 3525 ~ 3527.
- 19 Lipton A J, Fujiyoshi H, Patil R S. Moving target classification and tracking from real-time video [A]. In: *Proceedings of Applications of Computer Vision* [C], Princeton, New Jersey, USA, 1998; 8 ~ 14.
- 20 Cutler R, Davis L S. Robust real-time periodic motion detection, analysis and applications [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 20(8): 781 ~ 796.
- 21 Collins Robert T, Lipton Alan J, Kanade Takeo. A System for Video Surveillance and Monitoring [R]. Technical report CMU-RI-TR-00-12, Pittsburgh, USA; Robotics Institute, Carnegie Mellon University, 2000.
- 22 Fang Shuai, Wang Dong-zhu, Chi Jian-nan, *et al.* The algorithm of small targets classification in video surveillance system [J]. *Information and Control*, 2005, 34(2): 201 ~ 204, 222. [方帅, 王东署, 迟建男等. 视频监控系统中运动目标分类算法[J]. *信息与控制*, 2005, 34(2): 201 ~ 204, 222.]
- 23 Ramprasad Polana, Randal Nelson. Low level recognition of human motion. Motion of Non-Rigid and Articulated Objects [A]. In: *Proceedings of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects* [C], Texas, USA, 1994; 77 ~ 82.
- 24 Wojtaszek Daniel, Laganiere Robert. Tracking and recognition people in color using the earth mover's distance [A]. In: *Proceedings of IEEE International Workshop on Haptic Virtual Environments and Their Applications* [C], Ottawa, Canada, 2002; 91 ~ 96.
- 25 Warren F Gardner, Daryl T Lawton. Interactive model-based vehicle tracking [J]. *IEEE Transactions on Pattern Analysis and Intelligence*, 1996, 18(11): 1115 ~ 1121.
- 26 Lou Jian-guang, Tan Tie-niu, Hu Wei-ming, *et al.* 3-D model-based vehicle tracking [J]. *IEEE Transactions on Image Processing*, 2005, 14(10): 1561 ~ 1569.
- 27 Ralf Plankers, Pascal Fua. Tracking and modeling people in video sequences [J]. *Computer Vision and Image Understanding*, 2001, 81: 285 ~ 302.
- 28 Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models [J]. *International Transactions on IT*, 1992, 38(2): 617 ~ 643.
- 29 Hyunwook Park, Schopflin Todd, Yognin Kim. Active contour model with gradient directional information: directional snake [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2001, 11(2): 252 ~ 256.
- 30 Bozic S M, Ling Yun-dan. *Digital and Kalman Filtering* [M]. Beijing: Science Press, 1984. [鲍齐克, 凌云旦. *数字滤波和卡尔曼滤波* [M]. 北京: 科学出版社, 1984.]
- 31 Gordon N J, Salmond D J, Smith A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation [J]. *IEE Proceedings of Radar & Signal Processing*, 1993, 140(2): 107 ~ 113.
- 32 Song Lei, Huang Xiang-lin, Shen Lan-sun. Rapid detection &

- analysis of anomalies in video surveillance [J]. *Chinese High Technology Letters*, 2004, 14(10): 1~6. [宋磊, 黄祥林, 沈兰荪. 视频监控中的快速异常检测与分析[J]. *高技术通讯*, 2004, 14(10): 1~6.]
- 33 Arun Hampapur, Lisa Brown, Jonathan Connell, et al. Smart surveillance: applications, technologies and implications [A]. In: *Proceedings of Fourth International Conference on Information [C]*, Singapore, 2003, 2C1.2: 1133~1138.
- 34 Nair Vinod, Clark James J. Automated visual surveillance using hidden markov models [A]. In: *Proceedings of the 15th Vision Interface Conference [C]*, Calgary, Canada, 2002: 88~93.
- 35 WU Xin-yu, OU Yong-sheng, Qian Hui-huan, et al. A detection system for human abnormal behavior [A]. In: *Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2005) [C]*. Edmonton, Canada, 2005: 1204~1208.
- 36 Tony Jan, Massimo Piccardi, Thomas Hintz. Neural network classifiers for automated video surveillance [A]. In: *Proceedings of 2003 IEEE XIII Workshop on Neural Network for Signal Processing [C]*, Toulouse, France, 2003: 729~738.
- 37 Steffens J, Elagin E, Newen H. Person spotter-fast and robust system for human detection, tracking and recognition [A]. In: *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C]*, Nara Japan, 1998: 516~521.
- 38 Liu Dang-hui, Shen Lan-sun, Kin-Man Lam. Face recognition: a survey [J]. *Journal of Circuits and Systems*, 2004, 9(1): 85~94. [刘党辉, 沈兰荪, Lam Kin-Man. 人脸识别研究进展[J]. *电路与系统学报*, 2004, 9(1): 85~94.]
- 39 Liu Dang-hui, Kin-Man Lam, Shen Lan-sun. Illumination invariant face recognition [J]. *Pattern Recognition*, 2005, 38(10): 1705~1716.
- 40 Johnson Amos Y, Bobick Aaron F. A multi-view Method for gait recognition using static body parameters [A]. In: *Proceedings of 3rd International Conference on Audio- and Video-based Biometric Person Authentication [C]*, Halmstad, Sweden, 2001: 301~311.
- 41 Lee L, Grimson W E L. Gait analysis for recognition and classification [A]. In: *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02) [C]*, Washington DC, USA, 2002: 148~155.
- 42 Wang Liang, Hu Wei-ming, Tan Tie-niu. Gait-based human identification [J]. *Chinese Journal of Computers*, 2003, 26(3): 353~360. [王亮, 胡卫明, 谭铁牛. 基于步态的身份识别[J]. *计算机学报*, 2003, 26(3): 353~360.]
- 43 Boulgouris N V, Hatzinakos D, Plataniotis K N. Gait recognition: a challenging signal processing [J]. *Technology for Identification*, 2005, 22(6): 78~90.
- 44 Boulgouris Nikolaos V, Plataniotis Konstantinos N, Dimitrios Hatzinakos. Gait recognition using dynamic time warping [A]. In: *Proceedings of 2004 IEEE 6th Workshop on Multimedia Signal Processing [C]*, Sienna, Italy, 2004: 263~266.
- 45 Kale Amit, Sundaresan Aravind, Rajagopalan A N, et al. Identification of humans using gait [J]. *IEEE Transactions on Processing*, 2004, 13(9): 1163~1173.
- 46 Remagnino P, Tan T, Baker K. Agent orientated Annotation in Model Based Visual Surveillance [A]. In: *Proceedings of IEEE International Conference on Computer Vision [C]*, Bombay, India, 1998: 857~862.
- 47 Nevatia Ram, Zhao Tao, Hongeng Sombon. Hierarchical language-based representation of events in video streams [A]. In: *Proceedings of the Second IEEE Workshop on Event Mining [C]*, Madison, Wisconsin, USA, 2003: 1~8.
- 48 Harris J L. Diffraction and resolving power [J]. *Journal of the Optical Society of America*, 1964, 54(7): 931~936.
- 49 Goodman J W. *Introduction to Fourier Optics [M]*. New York: Mc Graw-Hill, 1968.
- 50 Tsai R Y, Huang T S. Multipleframe image restoration and registration [A]. In: *Advances in Computer Vision and Image Processing V. 1 [M]*. Greenwich, USA: JAI Press Inc., 1984: 317~339.
- 51 Zhang Xin-ming. Research on Image Fusion Technique—Super resolution Restoration [D]. Beijing: Beijing University of Technology, 2002. [张新明. 图像融合技术的初步研究——超分辨率复原[D]. 北京: 北京工业大学, 2002.]
- 52 Komatsu T, Igarashi T, Aizawa K, et al. Very high resolution imaging scheme with multiple different aperture cameras [J]. *Signal Processing Image Communication*, 1993, 5(5-6): 511~526.
- 53 Irani M, Peleg S. Improving resolution by image registration [J]. *CVGIP: Graph. Models and Image Processing*, 1991, 53(3): 231~239.
- 54 Ozkan M K, Tekalp A M, Sezan M I. POCS-based restoration of space-varying blurred images [J]. *IEEE Transactions on Image Processing*, 1994, 3(4): 450~454.
- 55 Schultz R R, Stevenson R L. Extraction of high-resolution frames from video sequences [J]. *IEEE Transactions on Image Processing*, 1996, 5(6): 996~1011.
- 56 Elad M, Feuer A. Restoration of a single super resolution image from several blurred, noisy, and undersampled measured images [J]. *IEEE Transactions on Image Processing*, 1997, 6(12): 1646~1658.
- 57 Borman Sean, Robert Stevenson. Spatial Resolution Enhancement of Low-resolution Image Sequences: A Comprehensive Review with Directions for Future Research [R]. Indiana, USA: University of Notre Dame, 1998.
- 58 Freeman William T, Pasztor Egon C. Learning low-level vision [J]. *International Journal of Computer Vision*, 2000, 40(1): 25~27.
- 59 Elad M, Feuer A. Superresolution restoration of an image sequence: adaptive filtering approach [J]. *IEEE Transactions on Image Processing*, 1999, 8(3): 387~395.
- 60 Alam M S, Bognar J C, Hardie R C, et al. Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames [J]. *IEEE Transactions on Instrumentation and Measurement*, 2000, 49(5): 915~923.