

Fuza Huanjing Zhong

Yundong Mubiao Jiance Yu Genzong Yanjiu

# 复杂环境中 运动目标检测与跟踪研究

张 辰 著

中国矿业大学出版社

# 复杂环境中运动目标 检测与跟踪研究

张 辰 著

中国矿业大学出版社

## 内 容 提 要

本书针对目标检测与跟踪方面的关键问题,结合科研工作中的研究案例,介绍了有关计算机视觉领域中涉及的运动目标检测与跟踪算法及具体应用。

本书内容全面,理论与实例分析相结合,可作为信息科学、计算机科学与技术、自动化、仿真与虚拟现实专业技术人员和科研工作者的参考书。

### 图书在版编目(CIP)数据

复杂环境中运动目标检测与跟踪研究/张辰著.

徐州:中国矿业大学出版社,2014.8

ISBN 978 - 7 - 5646 - 2466 - 8

I. ①复… II. ①张… III. ①运动目标检测—研究②  
目标跟踪—研究 IV. ①TP72②TN953

中国版本图书馆 CIP 数据核字(2014)第 202791 号

书 名 复杂环境中运动目标检测与跟踪研究

著 者 张 辰

责任编辑 郭 玉 张 岩

出版发行 中国矿业大学出版社有限责任公司

(江苏省徐州市解放南路 邮编 221008)

营销热线 (0516)83885307 83884995

出版服务 (0516)83885767 83884920

网 址 <http://www.cumtp.com> E-mail: cumtpvip@cumtp.com

印 刷 江苏徐州新华印刷厂

开 本 850×1168 1/32 印张 5.5 字数 143 千字

版次/印次 2014 年 8 月第 1 版 2014 年 8 月第 1 次印刷

定 价 26.00 元

(图书出现印装质量问题,本社负责调换)

## 前 言

运动目标检测和跟踪是运动目标行为分析的基础,是计算机感知和理解周围环境变化的重要前提,一直是计算机视觉领域中活跃的研究课题。该研究需要涉及多个理论学科,包括:图像处理、模式识别、机器学习、概率论与数理统计等。当前对运动目标检测和跟踪问题的研究尚处于对特定问题设计特定方法的阶段。由于复杂场景中的运动目标检测和跟踪是智能视觉监控系统的重要组成部分,其结果直接影响到对运动目标行为的分析和理解,因此,研究复杂环境中运动目标的检测和跟踪具有现实意义和理论研究价值。运动目标检测与跟踪的目的是通过建立背景模型,从视频序列图像中提取到感兴趣的运动目标,然后对该目标进行跟踪,获得目标各项参数,诸如位置、速度和加速度等,进而建立起目标在视频序列中的对应关系,为后期的高层视觉处理提供可靠的数据分析,其结果直接影响到对视频的分析 and 理解。本书针对目标检测与跟踪方面的关键问题,结合科研工作中的研究案例,介绍了有关计算机视觉领域中涉及的运动目标检测与跟踪算法及具体应用。

全书共分为 5 章。第 1 章阐述了本书研究课题提出的背景和研究意义,分析了煤矿视频数据的特点,描述了目标检测和跟踪的国内外研究现状。第 2 章介绍了一种基于鲁棒模糊核聚类方法,然后利用它进行目标检测,提高了背景模型对场景变化的适应能力。第 3 章介绍了一种基于 ELM 的最大间隔聚类(EMMC)方

法,然后将其引入 BoW 模型中,使用极速最大间隔聚类优化  $K$ -均值,可得到较好的视觉单词划分结果,提高目标分类性能。第 4 章提出一种基于 ELM 和多特征融合的粒子滤波跟踪算法。该算法利用极速学习机重新估计粒子权值,改善粒子退化问题,同时采用多特征融合构建粒子滤波中的目标表示模型,可以有效解决目标跟踪中的遮挡问题,同时提高算法的执行效率。第 5 章提出结合协同训练分类器的粒子滤波跟踪算法。该算法中分类器和粒子滤波跟踪器同时独立运行,然后将结果进行融合,得到目标定位。当跟踪失败后,采用无重叠的分类器网格重新获取目标,同时启动跟踪器对目标进行跟踪。

本书内容全面,理论与实例分析相结合,可作为信息科学、计算机科学与技术、自动化、仿真与虚拟现实专业的技术人员和科研工作者的参考书。在本书的写作过程中得到学院领导和同事的大力支持,在此对他们表示感谢。由于本书作者知识水平有限,书中不妥之处在所难免,敬请读者批评指正。

著 者

2014 年 6 月

# 目 录

|     |                          |    |
|-----|--------------------------|----|
| 1   | 绪论 .....                 | 1  |
| 1.1 | 研究背景及意义 .....            | 1  |
| 1.2 | 国内外研究现状 .....            | 5  |
| 1.3 | 主要研究内容及方法 .....          | 23 |
| 1.4 | 本章小结 .....               | 25 |
| 2   | 基于聚类的运动目标检测方法 .....      | 26 |
| 2.1 | 引言 .....                 | 26 |
| 2.2 | 鲁棒模糊核聚类算法 .....          | 28 |
| 2.3 | 基于鲁棒模糊核聚类的目标检测算法 .....   | 32 |
| 2.4 | 实验及结果分析 .....            | 36 |
| 2.5 | 本章小结 .....               | 50 |
| 3   | 基于改进 BoW 模型的目标分类方法 ..... | 52 |
| 3.1 | 引言 .....                 | 52 |
| 3.2 | 极速学习机模型 .....            | 54 |
| 3.3 | 极速最大间隔聚类 .....           | 58 |
| 3.4 | 利用极速最大间隔聚类构造视觉词典算法 ..... | 63 |
| 3.5 | 实验及结果分析 .....            | 69 |
| 3.6 | 本章小结 .....               | 81 |

|          |                               |     |
|----------|-------------------------------|-----|
| <b>4</b> | <b>基于 ELM 和多特征融合的粒子滤波跟踪方法</b> | 83  |
| 4.1      | 引言                            | 83  |
| 4.2      | 基于 ELM 的粒子滤波改进算法              | 86  |
| 4.3      | 基于多特征融合的目标表示模型                | 90  |
| 4.4      | 算法流程                          | 97  |
| 4.5      | 实验与结果分析                       | 98  |
| 4.6      | 本章小结                          | 114 |
| <b>5</b> | <b>结合协同训练分类器的粒子滤波跟踪方法</b>     | 116 |
| 5.1      | 引言                            | 116 |
| 5.2      | 系统流程                          | 118 |
| 5.3      | 基于协同训练分类器的目标检测                | 119 |
| 5.4      | 结合协同训练分类器的粒子滤波跟踪算法            | 127 |
| 5.5      | 实验与结果分析                       | 134 |
| 5.6      | 本章小结                          | 143 |
| <b>6</b> | <b>结论</b>                     | 145 |
| 6.1      | 主要研究成果                        | 145 |
| 6.2      | 进一步研究内容                       | 147 |
|          | <b>参考文献</b>                   | 148 |

# 1 绪 论

## 1.1 研究背景及意义

### 1.1.1 运动目标检测与跟踪

视觉是人类从外界获取信息的主要途径,而计算机视觉就是研究如何利用机器感知周围环境的变化,换句话说,就是利用计算机复制人类视觉的基本功能,例如运动感知和场景理解。计算机视觉的一个重要应用领域是智能视觉监控,它可以在没有人工帮助下利用计算机视觉方法实时获取、分析和处理视频数据,实现对监控场景中的目标检测、识别和跟踪,并根据结果执行更高层次的处理,例如对异常事件发出及时预警,从而使视频监控系统具有像人一样的智能,提高监控效率。智能视觉监控有着非常广阔的应用场景,如机场智能监控可以对停机坪场景进行目标跟踪和异常行为检测与报警;智能交通视频监控系统可以监控车辆速度、车辆逆行、车辆停驶、行人出现、交通流量等情况;城市安防监控系统可以对诸如停车场、火车站等公共区域及智能小区进行监控,对偷盗、破坏等异常行为提供实时报警,保证个人及公共财产和人身的安安全;军事视频监控可以利用无人机进行军事目标侦察与探测,跟踪与定位飞机、导弹等军事目标。

图 1-1 描述了智能视觉监控系统的一般框架,从图中可以看出,目标检测与目标跟踪分别处于整个智能视觉监控系统的前期

和中期阶段,运动目标检测与跟踪目的是通过建立背景模型,从视频序列图像中提取感兴趣的运动目标,提取到目标后对该目标进行跟踪,获得目标各项参数,诸如位置、速度和加速度等,进而建立起目标在视频序列中的对应关系,为后期的高层视觉处理提供可靠的数据分析,其结果直接影响到对视频的分析 and 理解。因此,运动目标检测和跟踪作为智能视觉监控的关键技术,不仅直接决定了系统性能,同时也是计算机视觉中最重要和最具挑战性的研究课题之一。



图 1-1 智能视觉监控系统的一般框架

### 1.1.2 井下运动目标检测与跟踪

煤炭是我国现行使用的主要能源,煤矿安全是国家安全生产的重中之重。虽然目前煤矿已经装备了各类监控系统,并且这些系统也已经在煤矿的安全生产中发挥了重要作用(图 1-2 描述了某煤矿的视频监控的工作环境),但现阶段,煤矿监控系统仍主要以“人工监测”为主,即依赖机器观察环境同时用人的大脑进行判断的方式进行监控,从图 1-3 可以看出现在煤矿还是主要依靠工作人员观察监控视频,然后作出判断。但科学研究表明,当人眼连续监视屏幕超过 20 分钟就会出现视觉疲劳,导致大脑麻痹,不能实时处理各类突发事件。所以这种监控方式具有局限性,是造成煤矿安全事故频发的重要原因之一。因此,在煤矿视频监控系统中引入智能视觉技术,不仅要用机器取代人眼监控,还要将人脑用机器代替,将煤矿视频监控从事后取证改为基于事前预防和实时事件驱动的监控方式,使基于视频的报警/预警成为可能,这也是煤矿安全监控系统未来的发展趋势。

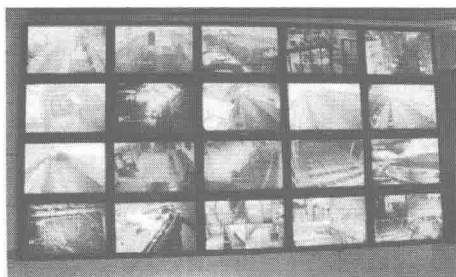


图 1-2 某煤矿的监控监测视频系统

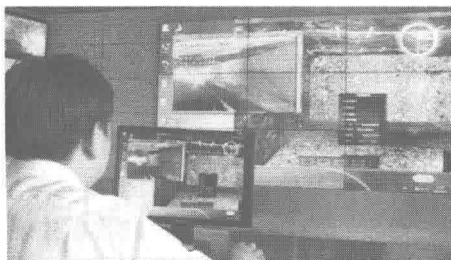


图 1-3 工作人员对煤矿监控视频进行观察

然而利用智能视觉技术实现对煤矿视频的监控还存在着许多问题：

### (1) 煤矿监控环境具有复杂性

与停车场、银行等公共场景不同,煤矿井下环境特殊,噪声大,需要 24 小时人工照明,亮度不均,目标边缘或区域特征不明显,几乎所有的图像都以黑、灰、白颜色为主(图 1-4),这导致目标和背景特征相似、分辨率低等。

### (2) 缺乏领域知识支持

目标检测与跟踪算法很少具有通用性,一般情况下是针对特定领域开发的。现有的机器智能视觉系统缺乏对煤矿现场知识和经验的有效利用,例如很少能够将这些领域知识表示和组织起来,



图 1-4 典型的井下图像

然后对它们进行处理融合并且能够逆向反馈到机器视觉系统,而这却是人类视觉常用的方法。

### (3) 目标外观变化

在目标检测跟踪过程中目标对象外观不可能一直不发生变化,这是由于目标自身有可能发生尺度、形状、角度旋转等变化;或者由于拍摄距离、角度、光照变化等导致的目标外观发生改变。要长时间检测跟踪目标,就要求检测跟踪算法适应这些变化。

### (4) 鲁棒性和实时性之间的平衡

煤矿监控数据量大,当处理复杂背景时,算法通常需要耗费大量时间。但对于井下环境而言,安全重于泰山,对实时性要求很高,既要求能实时检测物体,又要求能实时跟踪,快速分析判断是否发生异常信息,因而开发出同时具备快速性与准确性的目标检测与跟踪算法是至关重要的。这就需要鲁棒性和实时性进行折中考虑,使两者达到平衡。

综上所述,当前对运动目标检测和跟踪问题的研究尚处于对特定问题设计特定方法的阶段。由于复杂场景中的运动目标检测和跟踪是智能视觉监控系统的重要组成部分,其结果直接影响到对运动目标行为的分析和理解,因此,成为国内外众多学者研究的热点。本书针对煤矿复杂的工作环境下安全监控系统中存在的问

题,在借鉴已有研究成果的基础上,对煤矿井下环境中运动目标的检测及跟踪中的关键问题进行研究,并将这些研究成果应用到煤矿安全监控实际问题中去。因此,研究煤矿井下环境的运动目标检测和跟踪具有现实意义和研究价值。

## 1.2 国内外研究现状

运动目标检测与跟踪首先都需要考虑怎样对运动目标进行有效描述,即采用什么样的目标特征描述对象使其在特征空间上能较好地目标从背景中分离出来。两者的区别在于,目标跟踪由于完全标注的训练样本较少和实时性要求较高,因此在选用特征或者描述时往往对计算效率有更高的要求,比如要求具有较低的维数以及要求计算尽量高效等。本节首先介绍视觉特征的研究现状,然后介绍目标检测和目标跟踪的研究现状,最后给出目标检测和跟踪的关系。

### 1.2.1 视觉特征研究现状

从原始图像中提取具有较强表示能力的图像特征是运动目标检测与跟踪的关键,视觉特征可以严重影响检测与跟踪的效果。一般情况下,良好的特征应具有四个特点:区别性、可靠性、独立性和小数目。在计算机视觉领域,根据图像特征应用范围不同,可以将特征分为不同类型,例如,依据特征在图像上的表现形式可分为点特征、线特征和面特征;依据特征提取的方法可分为颜色特征、纹理特征、形状特征等;根据特征提取区域范围不同可以分为全局视觉特征和局部视觉特征。

#### (1) 基于全局特征的视觉表示

一个全局视觉表示反映了运动目标外观的全部统计特征。一般情况下,全局视觉特征表示主要应用于需要实现简单、快速的目标检测与跟踪中。光照变化、阴影、变形等导致的外观变化,常采

取将多种类型的可视化信息(如位置、形状、质地、几何结构等)融合到外观模型中。但是,全局特征无法从背景区分前景,并且有可能把这两部分的信息混合在一起。它主要包括以下几个方面:

### ① 原始像素

原始像素在计算机视觉中被认为是最基本的特征,由于其具有简单性和高效率,原始像素被广泛应用于可视对象跟踪。图像可以直接利用像素的原始颜色值或强度值表示待识别的对象区域。在文献中,通常将原始像素以两种形式表示:基于矢量和基于矩阵。基于矢量表示方法直接拼合图像区域得到一个高维向量,而且往往受到小样本大小问题的影响。为了减轻小样本大小问题,基于矩阵表示方法通常采用二维矩阵或高阶张量来减少所描述对象的维度属性。原始像素由于不用考虑特征提取,因此针对可视化对象跟踪是简单有效的。然而,仅靠原始像素信息是不够的,为了能够得到更好的检测与跟踪,研究人员试图将其他视觉特征(例如,形状或纹理)融合到原始像素表示中。

现在,为了改进手工设计的视觉特征(如 SIFT、HOG、SURF 等)计算困难、耗费时间和专业性强,尤其进行特定视觉应用时,还需要对这些视觉特征花时间调整以适应需求的缺陷,可以将原始像素作为无监督特征学习的输入,自动学习得到合适的视觉特征。

### ② 光流

运动是视觉感知的重要特征,而光流代表一个密集领域中各像素的图像区域内的位移矢量,并且通常是用来捕获运动对象的时空信息。典型的光流可以分成两大类:恒定亮度约束光流和非恒定亮度约束光流。光流法通过光照变化感知物体运动,可以利用光流法计算动态目标的运动特征。

### ③ 直方图

由于直方图能够很好地捕获运动对象区域的视觉特征分布,很受视觉对象检测与跟踪欢迎。在一般情况下,它有两个分支:单

特征直方图和多特征直方图。单特征直方图通过构建一个直方图用来捕捉对象区域内的分布信息。例如,Bradski 使用颜色直方图表示 HSV 色彩空间中的目标对象,然后把它应用到连续自适应均值漂移跟踪框架中去。单特征直方图表示虽然能够高效统计目标对象区域的视觉特征的分布信息,但是由于其缺少被跟踪对象的空间结构信息,往往受到被跟踪对象的相似背景颜色干扰。为了捕捉更多空间信息,空间颜色直方图表示被引入可视对象跟踪。通常情况下,空间信息编码在一个共同的空间颜色特征的空间,或通过划片策略将图像的空间信息编码到外观模型。但是,在上述颜色—空间直方图表示中没有考虑到对象外观的形状或纹理信息,为了解决这个问题,形状或纹理信息(例如,渐变或边缘)表示被建议加入到直方图中,并构成形状—纹理直方图表示。

#### ④ 协方差矩阵表示

协方差矩阵是一类基于图像二阶统计特性的特征描述,在目标检测与跟踪中取得了良好的应用场景。Porikli 等提出使用区域内所有特征点的协方差矩阵来表示目标区域的相关信息,实现了多特征融合,同时能有效适应旋转、尺度缩放以及亮度变化。根据黎曼流形度量,协方差矩阵表示可分为基于仿射不变距离和基于 Log-Euclidean 距离。基于仿射不变黎曼测量指的是目标模型和候选目标模型的协方差矩阵之间的距离计算公式为:

$$\rho(C_1, C_2) = \sqrt{\sum_{j=1}^d \ln^2 \lambda_j(C_1, C_2)} \quad (1-1)$$

式中:  $d$  为协方差矩阵维度;  $\{\lambda_j(C_1, C_2)\}_{j=1}^d$  是指两个协方差矩阵  $C_1$  和  $C_2$  的广义特征值,其定义如下:  $\lambda_j C_1 x_j = C_2 x_j, j = \{1, \dots, d\}$ 。

基于 Log-Euclidean 的黎曼测量定义了两个协方差矩阵在欧氏向量空间上的计算公式:

$$d(C_1, C_2) = \|\log C_i - \log C_j\| \quad (1-2)$$

使用的协方差矩阵表示的优点如下:可以捕捉对象外观内在

的自相关属性;提供了一种有效融合从不同模型中获取的不同的图像特征;由于其低维,计算效率较高;可以用于比较不同尺寸的区域或形状;很容易实现;对光照变化、遮挡、形状变形等情况具有鲁棒性。使用协方差矩阵表示的缺点为:由于使用像素统计,很容易受到噪声干扰,同时失去很多有用的信息,如纹理、形状和位置。

#### ⑤ 基于小波变化的视觉表示

基于小波滤波表示是利用小波变换对目标区域在不同的尺度或方向进行滤波,提取到边缘、角点、形状等有用信息。比如,He等人利用二维 Gabor 小波变换(GWT)进行可视化表示。具体而言,就是利用 GWT 系数高的特征点去表示目标对象。此外,Li等人提出了一种跟踪算法,该算法是基于生物启发的三层(SBI)架构,优点是可以将丰富的空间频率信息编码成一个统一的 SBI 的特征向量。

#### ⑥ 活动轮廓表示

活动轮廓表示通常被设计用来解决非刚体目标跟踪,可以精确地分割具有复杂形状的物体。活动轮廓表示采用符号距离地图隐含边界信息的对象进行编码。典型地,主动轮廓表示被定义为一个符号距离映射  $\Phi$ :

$$\Phi(x,y) = \begin{cases} 0, & (x,y) \in C, \\ d(x,y,C), & (x,y) \in R_{\text{out}} \\ -d(x,y,C), & (x,y) \in R_{\text{in}} \end{cases} \quad (1-3)$$

在式中,  $R_{\text{in}}$  和  $R_{\text{out}}$  分别表示轮廓  $C$  的内部区域和外部区域,  $d(x,y,C)$  表示点  $(x,y)$  到轮廓  $C$  的距离。此外,活动轮廓表示与能量函数紧密关联,能量函数主要包括三个方面:内部能量,外部能量和形状能量。内部能量项描述目标轮廓上的内部约束;外部能量项衡量图像数据属于前景对象的可能性;形状能量项主要描述形状特征对目标轮廓的约束。活动轮廓的主要思想就是定义一条由控制点构成的初始能量曲线,将其初始化在待分割轮廓的周围,在

能量函数极小值的约束下,经过不断地演化曲线最终收敛得到图像轮廓。

## (2) 基于局部特征的视觉表示

相比之下,局部特征的可视化表示能够捕捉到局部结构对象外观。局部特征主要针对某些点或由某些点组成的区域或者单一目标的特征进行详细有效的提取,使其具有更高的可区分度,利用局部视觉特征可以获得精确的目标局部外观描述。因此,当全局外观由于光照变化,形状变形,旋转和部分遮挡等原因引起变化时,局部外观要比全局外观更具有鲁棒性。但是,要对局部特征作精确描述,当识别的局部区域较大时会导致局部特征数量庞大,进行目标匹配时计算量大且耗时。一般情况下,局部特征的可视化表示方法可以大致分为基于本地模板、基于分割、基于仿射不变性特征和基于显著性特征等。

### ① 基于本地模板

一般情况下,基于本地模板的可视化表示是使用一组局部模板表示目标区域。相对于基于全局模板可视化表示,它们能够有效应对部分遮挡和灵活的关节模型。例如,林提出将分层模板形状模型应用于人体检测与分割,他将人体分解成一组第三方模板,并用这些模板组成一棵树用来表示形状模型。通过分级匹配测试图像的部分模板,可以产生一套可靠的检测假说,然后放入一个全局优化框架得到人的最终定位。

### ② 基于分割

通常情况下,基于分割的可视化表示是将图像分割信息合并到对象跟踪的过程中,这能得到相对鲁棒的跟踪结果。另一种替代方法是基于超像素分割。所谓超级像素就是将多个像素组成具有感知意义的原子区域。例如,王等人利用超像素分割构造一个基于本地模板的可视化表示,具体而言,一个对象的周围的区域分割分为以下几种超级像素的,每一个超级像素都对应于一个本地

模板。通过建立一个基于均值漂移聚类的本地模板字典,对象状态可以通过候选样本与超级像素在字典中的模板被预测出来。

### ③ 基于仿射不变性特征

#### a. 尺度不变特征变换(Scale-Invariant feature transform)

D. G. Lowe 提出的 SIFT 算法用于在图像中检测和描述关键点,关键点是指那些当图像缩放和图像的旋转时不变的特征,同时相对于图像的噪声、光照、细微变化等具有鲁棒性。这些关键点可以定位不同图像之间的转换关系。SIFT 将图像特征描述为一个三元组:尺度、方向、纹理描述,其中纹理描述 =  $f(\text{尺度}, \text{方向})$ 。计算 SIFT 步骤大致可以分为以下四步:首先在尺度空间进行极值检测,利用高斯微分函数检测到潜在的对于尺度和旋转不变的兴趣点;接着在兴趣点位置上,确定关键点的位置和尺度;然后,给每个关键点分配基于图像局部的梯度方向;最后,用一个特征向量来表达在每个关键点的领域内测量图像局部的梯度。

#### b. 加速稳健特征(Speeded Up Robust Features)

Surf 算法是对 Sift 算法在算法执行效率方面的一种改进,是 SIFT 的一个变种。它也有类似 SIFT 的属性,但它的运算速度要快得多。这是由于 SURF 采用变化尺度的窗口滤波器建立尺度空间,同时 SURF 特征描述算子的维数要小于 SIFT 算子维数,并且 SURF 算子的计算过程不存在插值。在适当的条件下, SURF 利用积分图像 haar 求导可以实时完成两幅图像中物体的匹配。Wei 等人使用基于 SURF 的可视化表示提出了一个跟踪算法,该算法通过判断局部 SURF 特征全局对象运动的相容性跟踪算法,针对外观变化和背景杂波具有很强的鲁棒性。但是, SURF 算子在计算时可能产生信息的不对称现象。

#### c. 最大稳定极值区域(Maximally Stable Extremal Region)

MSER 被业界认为是处理图像之间的仿射形变最好的算法之一,它是一种基于图像局部对比度的区域特征检测子,利用不同