



同濟大學

TONGJI UNIVERSITY

博士学位论文

智能视频监控系统中的目标检测与跟踪
算法研究

姓名：李大威

学号：0810080012

所在院系：电子与信息工程学院

学科门类：工学

学科专业：控制理论与控制工程

指导教师：徐立鸿 教授

二〇一二年十二月



同濟大學
TONGJI UNIVERSITY

A dissertation submitted to

Tongji University in conformity with the requirements for
the degree of Doctor of Philosophy

**On the Object Detection and Tracking
Algorithms for Intelligent Video Surveillance
Systems**

Candidate: Dawei Li

Student Number: 0810080012

School/Department: School of Electronics and
Information Engineering

Discipline: Engineering

Major: Control Theory and Control Engineering

Supervisor: Prof. Lihong Xu

December, 2012

智能视频监控系统中的目标检测与跟踪算法研究

李大威

同济大学

学位论文版权使用授权书

本人完全了解同济大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版；学校有权保存学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或者部分的阅览服务；学校有权按有关规定向国家有关部门或者机构送交论文的复印件和电子版；在不以赢利为目的的前提下，学校可以适当复制论文的部分或全部内容用于学术活动。

学位论文作者签名：

年 月 日

同济大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

年 月 日

摘要

随着现代科学技术的发展和人们对自身安全关注度的提高，视频监控技术在社会生活和生产中得到了广泛的普及和应用。然而，传统的视频监控技术存在着智能化和自动化程度低下、需要人员长期值守、对特定监控内容检索困难等问题。智能视频监控作为一个涉及图像处理、计算机视觉、模式识别等学科的新兴技术领域，其研究意义和重要性与日俱增，已经得到了学术界，产业界和相关部门的高度重视。运动目标检测与跟踪不仅是智能视频系统中的核心组成部分，也是后续高级视频处理技术的基础。本文以固定摄像机的视频监控系统为对象，围绕着目标检测与跟踪这两大关键技术展开研究和应用，提出了两种目标检测算法和一种多目标跟踪算法。本文的贡献总结如下：

(1) 提出了一种基于 $srgb$ 色彩空间的核密度估计目标检测算法。这种方法能够在一定程度上抵抗光照变化的扰动，并利用了计算出的核密度估计值和马尔科夫随机场在贝叶斯框架下进行精细迭代决策改进检测结果。

(2) 设计了一种对光照变化具有鲁棒性的背景-前景目标检测算法 $spkmeansEM$ ，它具有两个主要优点：第一， $spkmeansEM$ 不仅能够抵消阴影对检测过程的干扰，还能排除高亮区域的不良影响。第二，不同于大部分其他算法， $spkmeansEM$ 不需要设计额外的抗光照环节。通过建立在混合高斯模型 GMM 上的迭代贝叶斯决策，我们消除了大部分散布的检测噪声并增强了前景轮廓。

(3) 在使用 $spkmeansEM$ 对 GMM 进行参数估计的研究中，我们发现了寻找多元高斯混合模型最大似然估计的一般规律，进而得到了理论上的突破口。我们首先验证了 Titterington 型和充分统计量型的两种在线 EM 算法可以对多元高斯混合模型进行在线参数更新，随后推导了各自的参数更新公式。这两种在线 EM 算法不仅能够用于智能监控系统的背景-前景检测任务，还能在其他领域发挥重要作用。

(4) 提出了可以应对目标遮挡的智能多假设目标跟踪算法。我们首先细化并完善了多假设跟踪的理论支持和模型，然后使用基于 K 均值聚类的机器学习方法解决目标在遮挡中产生的分裂和合并问题。通过与卡尔曼滤波器的结合，得到了令人满意的多目标跟踪结果。

关键词：智能视频监控系统，核密度估计，背景建模，期望最大算法，目标检测，目标跟踪，马尔科夫随机场，多假设跟踪

ABSTRACT

With the development of the modern science and technology and the awakening concern on security issues from people, video surveillance systems have become popular and found their applications in both our daily life and the manufacturing. However, the traditional video surveillance systems have many disadvantages such as the lack of intelligence and automation, the need for long-term monitoring and assistance from people, and the difficulty in content retrieving. Therefore, the importance of intelligent video surveillance technology, which is a novel research field concerning subjects as Image Processing, Pattern Recognition and Computer Vision, is now recognized by not only the academic world, but also by the whole industry and the administrative department. Moving object detection and tracking are the two core parts of the intelligent video surveillance technology, and they usually act as a fundamental step for higher level visual tasks. In this thesis, the object detection and tracking algorithms for video surveillance systems with fixed cameras are studied. We present two detection algorithms and a multiple target tracking algorithm, the contributions are summarized as follows:

(1) A Kernel Density Estimation algorithm in srgb color space is proposed for object detection. In the following noise reduction and binary mask enhancement step, the estimated density value and the Markov Random Field are utilized within the framework of Bayes' Theorem to improve the binary detection mask iteratively.

(2) An illumination-robust background-foreground object detection method "spkmeansEM" is presented in this paper. It has two attractive features: First, spkmeansEM is not only resistant against the interference of the shadows, but also can exclude effects of highlights. Second, unlike most of other algorithms that can handle unstable illuminations, the proposed algorithm does not need to design an additional procedure to extract shadow regions or highlight areas. We eliminate most of dispersed noise and enhance the contours of foregrounds in the detection result via iterative bayesian decision making, with the information of Gaussian Mixture Model of the background.

(3) When using spkmeansEM to estimate the parameters of Gaussian Mixtures, we have discovered some general rules on finding maximum likelihood estimation of the Multivariate Normal Mixture (Gaussian Mixture Model), and this discovery

further became a theoretical breakthrough. We first prove that the Titterington's online EM algorithm and the other online EM variant using sufficient statistics are feasible for updating the parameters in Multivariate Normal Mixtures under streaming data. Then the equations for parameter updating are derived respectively. The two types of online EM variants can be used in many other fields where statistical techniques may apply.

(4) A multiple target tracking algorithm that can cope with occlusions is proposed. We first complete the theoretical support for this approach, and then use K-means algorithm to solve the merge and split problem of targets during occlusion. Finally, by applying Kalman filtering process, satisfying target tracking results can be obtained.

Key Words: intelligent video surveillance systems, Kernel Density Estimation, background modeling, EM algorithm, object detection, object tracking, Markov Random Field, Multiple Hypothesis Tracking

目录

| | |
|-------------------------------------|----|
| 第 1 章 引言..... | 1 |
| 1.1 概述..... | 1 |
| 1.2 智能视频监控的研究背景与现状..... | 3 |
| 1.3 智能视频监控的两项关键技术——目标检测和目标跟踪..... | 5 |
| 1.4 智能视频监控面临的难题和挑战..... | 6 |
| 1.5 本论文创新点和章节内容安排..... | 7 |
| 1.5.1 创新点..... | 7 |
| 1.5.2 章节安排..... | 7 |
| 第 2 章 视频目标检测技术概述..... | 9 |
| 2.1 目标（前景）检测方法..... | 9 |
| 2.2 目标检测算法的性能评价..... | 10 |
| 2.3 目标检测所采用的视频图像基准数据集与实验平台..... | 11 |
| 第 3 章 基于 SRGB 色彩空间的核密度估计目标检测算法..... | 12 |
| 3.1 核密度估计..... | 12 |
| 3.1.1 基本原理..... | 13 |
| 3.1.2 基于核密度估计的初步前景检测..... | 13 |
| 3.1.3 带宽选择..... | 15 |
| 3.2 在 SRGB 色彩空间的核密度估计目标检测..... | 17 |
| 3.3 迭代式的噪声去除和目标轮廓增强..... | 18 |
| 3.3.1 基于贝叶斯后验概率的精细决策..... | 18 |
| 3.3.2 马尔科夫随机场和模拟退火..... | 19 |
| 3.4 仿真实验..... | 21 |
| 3.4.1 检测效果..... | 21 |
| 3.4.2 复杂度分析和实时性..... | 25 |
| 3.5 本章小结..... | 26 |
| 第 4 章 一种对光照变化具有鲁棒性的背景-前景检测算法..... | 28 |
| 4.1 文献综述..... | 28 |
| 4.2 混合高斯模型与参数估计..... | 30 |
| 4.2.1 混合高斯模型简介..... | 30 |
| 4.2.2 用基本期望最大化算法来估计混合高斯模型的参数..... | 31 |
| 4.2.3 用在线 EM 算法对混合高斯模型参数进行估计..... | 32 |
| 4.2.4 光照变化时在 RGB 色彩空间中呈现的特性..... | 35 |
| 4.2.5 光谱反射理论与所观察到的线性特征..... | 36 |
| 4.2.6 球面 K 均值聚类算法 (Spkmeans)..... | 39 |

| | |
|---|-----|
| 4.3 一种自组织 (SELF-ORGANIZING) 的机制..... | 42 |
| 4.4 基于统计方法的初步前景检测..... | 43 |
| 4.5 噪声抑制和前景轮廓增强..... | 45 |
| 4.6 实验与仿真..... | 47 |
| 4.6.1 背景建模与前景检测..... | 47 |
| 4.6.2 初步前景检测的最优参数调整..... | 56 |
| 4.6.3 迭代式去噪与前景增强实验..... | 58 |
| 4.6.4 计算量与实用性..... | 60 |
| 4.7 算法小结和展望..... | 61 |
| 第5章 最大似然估计、多元高斯混合模型及在线 EM 算法..... | 63 |
| 5.1 多元高斯混合模型的最大似然估计..... | 63 |
| 5.1.1 一些常见的最大似然估计方法..... | 64 |
| 5.1.2 用基本 EM 算法估计多元高斯混合模型..... | 66 |
| 5.2 两种在线 EM 算法..... | 67 |
| 5.2.1 用 Titterington 的在线 EM 算法更新多元高斯混合模型..... | 68 |
| 5.2.2 Titterington 型在线 EM 算法与梯度上升 EM 算法的关系..... | 69 |
| 5.2.3 用基于充分统计量的在线 EM 算法更新多元高斯混合模型..... | 70 |
| 5.3 实验仿真..... | 72 |
| 第6章 多假设目标跟踪技术..... | 80 |
| 6.1 目标跟踪技术概述和研究背景..... | 80 |
| 6.2 多假设跟踪技术概述..... | 81 |
| 6.2.1 概念准备..... | 81 |
| 6.2.2 基于多假设的多目标跟踪技术..... | 83 |
| 6.3 目标测量和跟踪器的关联——一种图论方法..... | 84 |
| 6.3.1 代价矩阵 (Cost Matrix)..... | 84 |
| 6.3.2 相关的图论基础..... | 86 |
| 6.3.3 通过求最小权值边覆盖求最优跟踪假设..... | 87 |
| 6.4 次优跟踪假设的获取..... | 90 |
| 6.5 应对目标分裂和合并..... | 92 |
| 6.6 用卡尔曼滤波器刷新跟踪器状态..... | 94 |
| 6.6.1 适用于跟踪器的卡尔曼滤波运动模型..... | 94 |
| 6.6.2 卡尔曼滤波器对跟踪器状态的预测和更新..... | 97 |
| 6.7 实验结果..... | 97 |
| 第7章 结束语..... | 99 |
| 7.1 本文工作小结..... | 99 |
| 7.2 未来工作展望..... | 100 |
| 致谢..... | 101 |
| 参考文献..... | 102 |

| | |
|--|-----|
| 附录 A SCORE 函数的条件期望..... | 108 |
| 附录 B 定理 5.1 的证明..... | 108 |
| 附录 C 定理 5.3 的证明..... | 109 |
| 附录 D 定理 5.4 的证明..... | 109 |
| D.1 推导 TITTERINGTON 型在线 EM 算法对 ω_j 的更新公式..... | 109 |
| D.2 推导 TITTERINGTON 型在线 EM 算法对 μ_j 的更新公式..... | 111 |
| D.3 推导 TITTERINGTON 型在线 EM 算法对 c_j 的更新公式..... | 112 |
| 附录 E 定理 5.5 的证明..... | 113 |
| 附录 F 定理 6.2 的证明..... | 115 |
| 个人简历、在读期间发表的学术论文与研究成果..... | 118 |

第1章 引言

1.1 概述

随着现代科学技术的发展和人们对自身安全关注度的提高,视频监控技术得到了广泛的普及和应用,在社会生活和生产上发挥着重大作用。那么,什么是视频监控?顾名思义,视频,就是通过摄像机采集视频图像的数据;而监控则是利用其传回的数据对环境进行监视和控制并且能够对特定的事件和行为作出响应,视频监控系统是一种具有防范和报警能力的综合系统。

视频监控系统可以说是当今发展最为迅猛的系统之一,其原因是多方面的:首先,视频监控系统的制造技术已趋于成熟,生产成本越来越低廉,这使得大规模的应用成为可能。其次,得益于计算机与网络技术的进步,视频监控系统的运行速度不断提升,其功能也不断增强,能够胜任在各式各样的监控环境下的工作。第三,由于当今世界形势的动荡和恐怖主义的威胁,许多国家都在安防领域提出了更高的要求。视频监控作为安防系统的重要组成部分,引起了国家和个人高度的重视。事实上,视频监控技术已经走入寻常百姓家,我们的身边遍布着以摄像机为前端的视频监控系统,它们在社会生活和生产的方方面面发挥着重要的作用。

视频监控系统的发展主要经历了三个阶段:第一代的模拟视频监控系统,第二代的部分数字化的监控系统和第三代的分布式数字视频监控系统。

第一代视频监控系统(1960-1980)以基于模拟信号的图像采集和传输为基础。摄像机所采集的视频信号通过同轴电缆传输到监控室,然后在监视器上显示给执勤人员。这个系统的主要特点包括以下几个方面:(1)视频信息使用录像带这种介质进行存储,容量有限也不方便长期保存,在面对大量的视频资料时无法方便的检索和查询所需要的内容;(2)由于同轴电缆的固有缺陷,模拟视频监控系统的信号传输距离受到限制;(3)模拟视频监控系统不能自动的触发报警,监控场景的突发情况和异常的判断全靠人工完成,管理维护复杂,价格昂贵[1]。

第二代视频监控系统(1980-2000)主要依赖于混合模拟和数字或者全数字的视频传输和处理方法。采用 Motion JPEG、H. 263 和 MPEG 等多媒体数字压缩技术将视频数字化,节省了传输带宽和存储空间。初步的视频分析算法开始出现在这一代监控系统中,可以根据一些特定的事件和情况触发自动报警。该系统主要有如下几种特点:(1)相比模拟信号,数字视频信息的抗干扰和抗信号衰减的能

力有了明显增强,而且可以进行加密传输;(2)经过数字化压缩后的视频信息可以在硬盘等可靠的存储介质上进行长时间的保存,所需要的存储容量大大减少,同时也便于检索;(3)初步的视频分析技术可以进行联动报警,减轻执勤人员的负担[2]。

第三代视频监控系统(2000-),也就是我们常说的智能视频监控,利用低价位高性能的计算机和先进的通信网络来传输和处理各种监控信息。网络摄像机和嵌入式的智能前端能够自动的分析监控环境,把视频与分析信息传输给监控中心,实现自动监控。该系统的主要特点有:(1)能够自适应的根据带宽情况进行视频压缩率的调整;(2)理论上可以进行无限扩展,组成复杂的监控网络;(3)由于信号的传输采用先进可靠的网络,所以可以实现异地实时监控;(4)智能的图像处理和视频分析技术的运用使得这一代视频监控系统可以自动监控和报警,无需人员值守。

传统的第一代和第二代视频监控系统虽然解决了视频的采集,传输,显示和存储问题,但这些系统无不需要人员长时间的值守来判断是否在场景中有异常情况发生。对监控人员而言,监视任务枯燥,劳动强度大,同时对注意力的要求也很高。有研究表明,即便是一个经验丰富的监控人员,当监视屏幕达到20分钟后,注意力也将迅速下降到不能满足监控要求的水平。在目前广泛应用于银行,商场等场所的视频监控系统中,由于摄像机布控点太多,人员往往无法及时掌握所有场景的状况。如果发生了异常和突发情况,只能去检索录像记录来进行事后分析。此时损失已经造成,事后的调查无异于亡羊补牢。我们希望能够对异常和突发情况进行预警,把安全隐患消灭在萌芽状态。智能视频监控系统的诞生,为解决上述的问题带来了希望。

智能视频监控(技术)是在以先进的网络摄像机、可靠的网络传输和高性能的视频服务器为主要结构的监控系统中,利用计算机视觉、图像处理、模式识别和人工智能等技术,使用计算机对视频序列中的兴趣目标进行检测、分类与跟踪,以及对监视场景中的行为和事件进行理解和推测的技术。与传统技术相比,智能视频监控不仅能代替人类的“眼睛”,还能成为我们的“大脑”,自动根据各种场景情况进行决策和操作,把监控人员从枯燥繁重的监视任务中解放出来。当前,视频监控系统的智能化已成为一个备受学术界、产业界和管理部门所瞩目的新兴研究领域,特别是美国9·11事件后,世界各国政府对本土安全政策的关注和投入都达到了新的高度。国家机关、敏感军事设施、机场、电力设施、供水设施、公共交通枢纽还有人员密集的楼宇都可能成为恐怖分子袭击的潜在目标。因此,智能视频监控技术以其强大的视频处理和分析能力以及突发事件预警能力成为应对恐怖主义和民用安防问题的有力工具。除了安全防范以外,智能视频监控技

术还在许多领域有着用武之地。在军事国防方面，移动智能视频监控分析系统不仅用来识别战场上的敌军伪装色，还可以加载红外夜视设备，在夜间进行侦查和防范。在智能交通系统中，智能视频监控能够实时的监控车流密度以实现自动交通调度，甚至可以完成违章车辆查处、车辆跟踪和车牌识别等任务。在现代医学当中，智能视频分析技术不仅可以分析 CT 和 MRI 的医学图像，还可以辅助医生进行远程手术。智能视频监控在工业生产中也发挥着重要的作用，许多零件的加工和质量检测离不开自动视频分析技术。

1.2 智能视频监控的研究背景与现状

在智能视频监控领域，国内已有许多高等院校和研究机构正在进行相关的研究并取得了一些成果，如中国科学院自动化研究所、上海交通大学、清华大学、北京邮电大学、北京航空航天大学、国防科技大学和华中科技大学等。其中，中国科学院自动化研究所模式识别国家重点实验室在人体运动分析、交通行为事件分析、交通场景视频监控和智能轮椅视觉导航等领域取得了许多科研成果。为了促进国内智能视觉监控的发展，中国科学院自动化研究所在 2002 年和 2003 年分别举办了第一届和第二届全国智能视觉监控会议，北京图像图形学学会在 2006 年举办了中国首届城市公共视觉听觉科技研讨会。《模式识别与人工智能》、《计算机学报》、《自动化学报》、《中国图像图形学报》、《软件学报》和《控制与决策》等一些杂志也时常发表智能视频监控和图像处理的一些研究成果。由于智能视频监控系统在国家反恐和犯罪威慑上显现出与日俱增的重要作用，我国近几年开始重视对这个领域的投入和产业扶植并且提出了“平安城市”的口号。

在国外，智能视觉监控方面的工作开展得更早。在 1996 年至 1999 年间，美国国防高级研究项目署 (DARPA) 设立了以卡内基梅隆大学为首、联合美国十几所高等院校和研究机构参加的视频监控重大项目 VSAM (Video Surveillance and Monitoring)。主要目标是利用视频理解、网络通信、多种传感器融合等技术实现对未来城市、战场的自动监控。DARPA 随后在 2000 年又资助了重大项目 HID (Human Identification at a Distance) 计划，研究开发多模式的监控技术以实现远距离情况下人的检测、分类和识别，以增强国防、民用场合对恐怖袭击的保护能力。欧盟 EULTR (European Union Long Term Research) 资助比利时 Katholieke 大学工程系、法国国家计算机科学和控制研究院 (INRIA) 等欧洲著名的大学和研究机构联合警察、法庭等司法机关开发基于图像处理的监控系统，目标是利用监控系统和图像视频理解技术，使司法机关能从现有的监控系统的

录像资料中获得更多有用的犯罪证据。包括伦敦大学在内的六所科研院所共同参与了 ADVISOR 项目,其主要内容是通过多个摄像机对地铁站点行人的检测与跟踪监控,估计出人群密度和运动量,并分析人群的行为,从而检测危险或者犯罪行为,并在限定的时间内向监控员作出告警[3, 4]。由英国爱丁堡大学负责的 BEHAVE 项目,通过图像分析手段对视频进行处理,检测和提取监控场景中的危险、异常行为,并对感兴趣的目标进行分析和识别[5, 6]。美国 ISS 公司的 AUTOSCOPE 2004 视频检测系统是一种分布式智能视频监控系统,该系统用于检测在铁路线上的无人值守候车室和铁路沿线重要路段出现的遗弃物品。这套系统在北美,欧洲和亚洲数以千记运行的交通智能管理监控系统中得到了实践的验证[3, 7]。

国际上的许多视频信息和图像处理的权威杂志如 International Journal of Computer Vision (IJCV), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on Image Processing (TIP), Pattern Recognition (PR), Pattern Recognition Letters (PRL), Computer Vision and Image Understanding (CVIU), Image and Vision Computing (IVC), International Journal of Mathematical Imaging and Vision (JMIV) 和重要年度学术会议 International Conference on Computer Vision (ICCV), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), European Conference on Computer Vision (ECCV), IEEE International Workshop on Visual Surveillance (IWVS) 等都将智能视觉监控系统作为主要关注内容之一。近几年,智能视觉监控的相关研究成果更是在学界引起广泛的兴趣,如 EURASIP Journal on Image and Video Processing 于 2008 年出版了专刊“Human-Centric Applications of Distributed Camera Networks”, Springer 的期刊 Signal, Image and Video Processing 在 2007 年出版了“Multi-Sensor object detection and tracking”专刊, International Journal of Computer Vision 在 2007 年出版了专刊“Modeling and Representations of Large-Scale 3D Scenes”和于 2008 年出版的专刊“Evaluation of Articulated Human Motion and Pose”中均涉及到智能视频监控的内容。

目前,智能视频监控的研究主要集中在目标检测与分类、目标跟踪、目标匹配、目标识别、行为识别以及视频内容理解等方面。其中,目标检测与目标跟踪是最为重要的两项技术,它们是后续的各种高层次的视频处理和理解的基础,也是视频监控系统的自动化、智能化和能够胜任实时任务的关键。本论文将主要围绕着目标检测和目标跟踪这两项技术展开讨论。

1.3 智能视频监控的两项关键技术——目标检测和目标跟踪

那么，什么是目标检测？视频序列的目标检测（也常被称为前景检测）是视频图像处理中的第一个重要问题，也是后续处理过程的基础，它所要完成的任务就是用算法检测出监控场景中我们所感兴趣的人或者交通工具等目标区域并且用一个气泡（blob）或者检测框在视频中标识其位置。一个标准的目标检测结果如图 1.1 所示，所有显著的行人被绿色的方框所标识。



图 1.1 一些场景中的目标检测结果

什么又是目标跟踪呢？目标跟踪一般建立在目标检测的基础之上，是智能视频监控系统的第二个重要环节。其主要任务是对进入监控场景中的运动目标进行身份标识，在相邻时刻对同一个目标观测位置之间建立对应关系[1]。在有多个目标出现的场景中，目标跟踪技术通常用不同的数字标号或者是不同的颜色来区别不同的目标气泡与检测框。一个多目标跟踪应用如图 1.2 所示，图中在街道上行走的行人在一段视频中被不同颜色的目标框所标记，该实例来自 [8]。



图 1.2 一些目标跟踪示例[8]

1.4 智能视频监控面临的难题和挑战

尽管在智能视频监控领域已经有了大量的研究成果，但是在实际应用中仍然存在许多难题亟待解决，诸如：

(1) 实时性。一个智能视频监控系统往往要求对场景中的突发和异常情况作出快速而及时的响应。同时，视频系统的图像的传输速率一般是每秒三十帧左右，虽然不要求对每一图像帧都进行处理，但监控系统的视频分析算法应该能以较快的处理速度与最新的图像帧保持同步。在像素级的目标检测和跟踪算法中，虽然 CPU 能够在很短的时间内处理每个像素，但是随着视频分辨率的提高，处理速度将会越来越慢。一幅分辨率为 704×576 的图像帧中像素已经超过了四十万，因此一个优秀的智能视频监控系统不仅要有性能强大的硬件，还需要在算法的设计和优化上下功夫。

(2) 适应性。正如 1.1 节中所述，智能视频监控的应用领域多种多样，监控环境千差万别。目前还不存在一种智能视频监控系统能够同时胜任在室内，室外，白天，夜间各种场景下的工作。因此如何增强智能视频监控系统的目标检测、跟踪、识别、理解等关键技术对纷繁芜杂环境的适应性，是一个重大课题。

(3) 鲁棒性。以晴朗的户外这一单独的监控环境为例，云层对太阳光照的遮挡会产生环境光线的剧烈变化，另外，目标对直射光照的遮挡会产生投射阴影，这些亮度和颜色改变都会对目标检测产生不良影响。在多运动目标交互运动的场景中时常发生几个目标相互遮挡的情况，这对目标跟踪技术提出了重大的挑战。为了应对这些实际干扰。智能视频监控系统应该具有一定的鲁棒性。