

# 网络环境下智能监控综述<sup>①</sup>

杨 戈<sup>1,2</sup>, 尤晓旭<sup>1</sup>

<sup>1</sup>(北京师范大学珠海分校 信息技术学院, 珠海 519087)

<sup>2</sup>(北京大学深圳研究生院 深圳物联网智能感知技术工程实验室, 深圳 518055)

**摘 要:** 网络环境下的智能监控在各类技术背景的支持下迅速发展了起来, 广泛应用于生产制造、智能安检、图像检索、医疗影像分析等领域. 与此同时, 对其核心计算机视觉的研究更是处在高新技术研究领域的前沿. 现今计算机视觉技术面临着语义信息描述模糊、效率低下等诸多问题. 本文旨在研究网络环境下的智能监控技术的基础上, 回顾并分析其发展轨迹, 梳理各核心技术的内在联系并加以归纳总结, 展望其发展趋势, 并针对其热点研究之一视觉跟踪, 特别是多摄像头协同工作的情况, 进行了较为详细的介绍.

**关键词:** 计算机视觉; 视觉跟踪; 智能监控; 多摄像头

## Intelligent Monitoring under Network Environment

YANG Ge<sup>1,2</sup>, YOU Xiao-Xu<sup>1</sup>

<sup>1</sup>(Beijing Normal University (Zhuhai Campus), College of Information Technology, Zhuhai 519087, China)

<sup>2</sup>(Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen 518055, China)

**Abstract:** For the support of a variety of technologies, the intelligent monitoring under the network environment develops rapidly. It is widely used in manufacturing, smart security, image retrieval, medical image analysis and other fields. Meanwhile, the study of its key technology, computer vision is at the leading edge of the field of high-tech research. Currently, the computer vision technology faces many problems such as vague description of semantic information and inefficient. Based on the study of current technology of intelligent monitoring under the network environment, this article aims to review and analyze the pathway of its development, then discover and summarize the internal relations among the core technologies and try to predict the development trends of the intelligent monitoring under the network environment. Additionally, the detailed description of visual tracking, one of the hotspot studies especially the condition of multi-camera collaborative work is presenting in the article.

**Key words:** computer vision; visual tracking; intelligent monitoring; multi-camera

众所周知, 视觉对每个人都十分重要, 科学数据表明我们所接触到的信息几乎有 85% 是通过眼睛得到的, 可是随着科技的发展, 人类涉足的领域越来越多, 渐渐发现我们光靠自身的视觉, 已很难兼顾到各个方面的发展, 甚至在一些监控领域我们的人类视觉已无能为力, 因此智能监控技术开始一步步发展起来. 就

目前的发展趋势而言, 智能监控将会被越来越多的人所熟悉. 智能监控技术在得到推广的过程中, 其发展趋势逐渐向高清化以及网络化的方向发展<sup>[1]</sup>. 到如今网络环境下的智能监控系统已经成为各行各业计算机系统应用的重要组成部分, 其应用领域日渐广泛, 配套的环境设备也日益增多. 对其核心计算机视觉的研

① 基金项目: 国家自然科学基金(60875050, 61272364); 国家高技术研究发展计划((2006AA04Z247)); 广东省自然科学基金(9151806001000025); 深圳市科技计划及基础研究计划(JC201005270275A); 深圳市战略性新兴产业发展专项资金(JCYJ20120614144655154); 北京师范大学珠海分校科研创新团队(多媒体传输与计算机视觉研究团队, 201251006); 北京师范大学珠海分校教改项目(201329).

收稿时间: 2013-04-24; 收到修改稿时间: 2013-07-01

究更是处在高新技术研究领域的前沿。

在网络环境下的智能监控日渐深入社会各个层面的今天, 人们的生产生活越来越依赖这一技术的运用, 在网络环境强大的覆盖能力下, 其核心 CV<sup>1</sup> 技术的发展十分的迅速, CV 技术主要关注如何利用传感器提供的原始数据, 并基于不同的应用, 准确并完善地描述出数据中可能蕴含的可供计算机识别的视觉信息并建立相应的数学模型. 可是虽然 CV 技术取得了一定的成就, 但是和精密且灵活的人类视觉相比, 仍显得过于粗糙. 现今 CV 技术的发展已成为高新技术领域热议的话题之一, 本文旨在研究当今网络环境下的智能监控技术的基础上, 回顾并分析其发展轨迹, 梳理各核心技术的内在联系并加以归纳总结, 并展望其发展趋势. 希望能以此文为那些对网络环境下的智能监控技术不是十分了解的研究人员提供引导与帮助.

### 1 计算机视觉

计算机视觉就是用各种成象系统代替视觉器官作为输入敏感手段, 由计算机来代替大脑完成处理和解释. 计算机视觉的最终研究目标就是使计算机能象人那样通过视觉观察和理解世界, 具有自主适应环境的能力. 英国已故科学家戴维·马尔(David Marr)是计算机视觉发展史上的一位领军人物. 他提出了第一个较为完善的视觉系统框架<sup>[2]</sup>, 即计算视觉理论框架<sup>2</sup>. 随着 CV 技术的发展, 研究人员发现如图 1 所示的 Marr 视觉理论框架并不十分完整, 因此人们针对 Marr 视觉理论框架的不足, 集各类观点综合成了如图 2 所示的计算机视觉理论框架<sup>[3]</sup>.



图 1 Marr 视觉理论框架

在此框架中存在的不足之处为:

- ① 输入为被动的, 提供何种图象就处理何种图象, 无主动性和目的性可言, 这根本不同于生物视觉.
- ② 整个处理过程基本上自底向上, 没有信息反

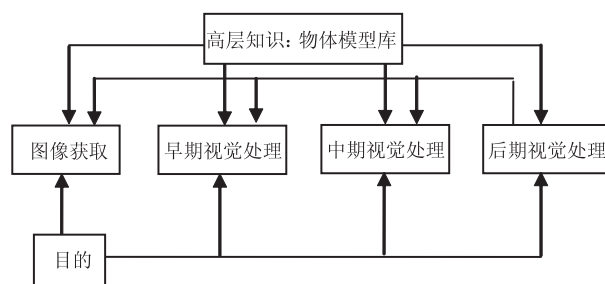


图 2 新的理论框架<sup>[3]</sup>

馈, 这种纯粹数据驱动的视觉机理往往会导致相应系统功能的失败.

③ 处理目的局限于场景中物体的三维形状及空间位姿关系.

这一框架充分体现了生物视觉的特点:

①从输入来说, 它嵌入了主动性、选择性机制; ②从目的来说本框架在原有基础之上集成了“目的性视觉”和“定性视觉”, 即视觉系统的功能取决于应用目的, 这一方法论指导无论对视觉系统的开发还是应用探索都极为重要; ③高层知识的利用, 目对计算机视觉不适定视觉过程(表现为病态数学问题)的求解. 需要用各种形式的知识加以限定, 以期得到合理的结果. ④控制策略, 在 D Marr 视觉理论中, 处理流程总是自下而上, 各阶段之间不存在反馈, 因而基本不存在控制问题; 而在本框架中, 仿照生物机制在前后处理之间引入了交互作用.

计算机视觉经过了漫长的发展, 其中几何模型以及利用几何模型进行三维推理、多传感(算子)信息的集成和融合、动态视觉, 定性视觉、目的性视觉等的运用与推理都是计算机视觉中很有潜力的发展方向. 这在文[5]中有较为详细的猜想.

随着目前互联网络技术的不断发展, 在网络环境下的智能监控中, 用于识别真实世界中较为复杂的图像内容的技术是发展的趋势之<sup>[4]</sup>. 另一个值得重视的趋势是智能监控技术与计算机图像技术、社交媒体技术等其它计算机技术的融合. 现在手机上越来越普遍运用的人脸识别技术, 随着网络环境下的智能监控技术的发展, 人脸识别技术的准确度可以通过发现图像中人物的社会关系、周边环境来得以提高, 这是基于

1 计算机视觉的简称.

2 D.Marr 认为:人们生活于其间的世界是三维的.投射于人的视网膜上的图像则是二维的.他将整个视觉过程所要完成的任务分成如图 1 所示的三个阶段.

对象在社交网络上分享的照片一般能反映固定的社会人际关系这一事实，通过统计确定特定人物出现在特点场景中的概率大小，从而进行人身份信息的识别<sup>[15]</sup>。这一方法也能应用在刑侦领域，特别是当目标人物无法清晰识别而周围的其他相关人物或相关场景较易判定的情况下。

### 1.1 计算机视觉相关技术

随着 CV 技术的发展，CV 领域中诞生了许多较为重要的技术，如图 3 所示<sup>[4]</sup>。



图 3 计算机视觉核心技术发展时间线<sup>[4]</sup>

计算机视觉系统的具体实现过程中，摄像机标定、图像配准，视觉跟踪和视觉测量是关键的技术环节，是计算机视觉系统研究内容中的关键问题，其中视觉跟踪是现今研究的热点问题。在这种背景下，下

文将对视觉跟踪这一当今重点研究课题做深入的理论探索和工程实践研究。根据国内外学者和研究机构针对这些问题取得得很多有价值的研究成果，表 1 将对这些关键技术研究成果的现状进行分析、归纳。

表 1 计算机视觉关键技术一览表

	摄像机标定技术	图像配准技术	视觉跟踪技术	视觉测量技术
原理	指在已知标定物三维几何信息的前提下，针对具体的摄像机成像模型，建立标定物的二维图像与三维世界之间的几何关系，然后根据几何约束采取一定的求解方法，求解成像模型参数的过程。	由于采集的时间、传感器或者视角不同，同一场景的多幅图像之间会有一些的差异，对这些图像匹配的过程就是图像配准。	指自动跟踪像素在图像序列中的运动，同时辅助于目标识别、提取等手段，获得运动目标的位置、速度、加速度等信息。	以图像为信息检测和分析的对象，综合运用计算机视觉和图像处理等技术，测量目标的各种参数。
特点	具有多元化，成本低，方法相对简单，稳定性、精度高等特点	应用领域不同，具有的特点也不同。在军事目标识别领域，其特点是自动化，效果精确；在图像实时性要求高的领域，其特点是运算量小，配准快速；在图像易受噪声和干扰信息的应用领域，其特点是具有较好的抗噪性和抗干扰性。	有鲁棒性、准确性、快速性等特点(下文有详细介绍)	具有精度高，自动化程度高，成本低，非接触等特点

常用方法简介	<p><b>直接线性转换法</b></p> <p>该类型方法将摄像机的物理参数转换成过渡参数,以过渡参数为未知量构建线性方程,最后从方程的闭合解中分解出摄像机的物理参数.</p>	<p><b>互相关法</b></p> <p>互相关法通常使用迭代优化法计算图像的几何变换参数,以使图像的相似性测度最大化.</p>	<p><b>点跟踪法</b></p> <p>在图像序列中检测到目标点后,运用点跟踪法确定相邻图像帧中多个目标点之间的对应关系.</p>	<p><b>双目视觉测量法</b></p> <p>双目视觉测量是由摄影测量理论发展而来的,根据同一物点在两台立体摄像机中的成像视差测距.</p>
	<p><b>非线性优化法</b></p> <p>本类方法采用迭代优化方法搜索使优化准则的函数值最小的摄像机参数.由于使用大范围非线性优化计算摄像机参数,算法所选用的成像模型越接近真实模型,估计结果越准确.</p>	<p><b>频域法</b></p> <p>基于频域法的配准技术具有抗噪性强、速度快,且对灰度变化不敏感的优点,一直是图像配准研究的重点.</p>	<p><b>区域跟踪法</b></p> <p>基于区域的跟踪方法使用图像区域信息跟踪目标,区域信息包括光流、颜色、灰度、纹理等等.</p>	<p><b>单目视觉测量法</b></p> <p>单目视觉测量法使用单目摄像机采集图像,避开了双目视觉中立体匹配的问题.通过调整摄像机焦距长度,可以根据透视成像公式求得被测点的距离.</p>
	<p><b>两步法</b></p> <p>该类标定方法是直接线性变换法和直接非线性优化法的折衷,首先使用直接线性变换法求解大部分参数,计算直接简单,然后只用非线性优化法迭代求解少部分参数,迭代求解的参数数目很少、速度快.</p>	<p><b>针对非刚体模型的配准方法</b></p> <p>对如透视模型、多项式模型、弹性模型等模型的研究.</p>	<p><b>轮廓跟踪法</b></p> <p>早期轮廓跟踪方法局限于图像边缘特征的使用,当目标边缘不够清晰时难以得到好的效果,为了摆脱这种限制,目前的研究方向侧重于边缘信息与区域信息的结合.</p>	

相关科学	计算机科学和工程、信号处理、物理学、应用数学和统计学,神经生理学和认知科学等
应用领域	制造业、检验、文档分析、医疗诊断、交通和军事等领域中各种智能/自主系统等领域.

### 1.2 核心问题

计算机视觉技术复杂性的本质在于相对声音等物理信号的描述,视觉信号充满了非常丰富的信息,描述起来也更加困难.比如,很多图像中蕴含了大量简单(如颜色、形状、纹理、几何特征等)及复杂(如场景、字符、物体分布、人物面部特征、人体姿势等)信息并具有较强的动态范围和主观性,如何攻克图像信息提取过程中的各种难题一直是当今计算机图像学研究的热点问题.

在科学家们还未完全破译生物视觉系统的奥秘的前提下,大多数 CV 问题只能采用“逆向推导机制”——依据已知或假设的关联将视觉系统的输入(数字图像)和输出(语义描述)对应起来,通过图片猜测真实世界物体具有的形状,照明度以及颜色分布.因此基于概率论和数理统计的数学模型是最适合解决这类逆推问题(inverse problem)的工具,这也是目前 CV 领域普遍采用各种统计模型和机器学习算法的本质原因<sup>[4]</sup>.

David Marr 于 1982 年发表的著作讨论了丰富的视觉原理问题<sup>[5]</sup>.其中,计算机视觉系统的开发问题被归纳为 3 个要素.该阐述在 31 年之后的今天看来仍不过时,研究者们仍可以从其描述中找到解决各种 CV 相关问题的基本思路.

#### (1) 数学理论

考虑数学计算层面的目标及可以引人的合理约束条件.

#### (2) 描述和算法

重点解决计算机视觉中的输入和输出的数据格式问题,并设计合理的算法实现其系统功能.

#### (3) 硬件的合理使用

使用符合算法要求的硬件并考虑该硬件对所需要的算法和描述的反作用.当今流行的图形处理器(Graphics Processing Unit, GPU)加速技术及多核处理使该问题变得更加具有现实意义<sup>[4]</sup>.

## 2 单摄像头视觉跟踪算法

上文针对视觉跟踪的原理及应用做了简要叙述,



接下来将介绍一些单摄像头基础的跟踪算法, 来为本文的重点第 3 节多摄像头协同跟踪做理论铺垫。

## 2.1 理想的视觉跟踪算法特性及思路

理想的视觉跟踪算法应具有以下几种特性<sup>[8]</sup>:

### (1) 快捷性:

视觉跟踪算法应该能够有效地跟踪运动目标, 并且对场景的突然变化做出快速反应, 这是视觉跟踪算法所希望达到的最根本目标;

### (2) 鲁棒性:

鲁棒性意味着可用性. 因为被跟踪对象从 3D 投影到 2D 时会造成信息损失、图像中的噪声干扰、物体的复杂运动、物体的非刚性<sup>3</sup>或关节的本质、部分和全部遮挡造成的信息暂时消失、物体姿态的复杂性、场景的光照变化等, 而希望的是在各种环境条件下都能实现对运动目标持续稳定的跟踪;

### (3) 透明性:

视觉跟踪算法对客户应是透明的, 客户得到的结果仅仅是快速的响应和良好的可用性;

### (4) 高效性:

视觉跟踪算法带来的运算开销越小越好;

### (5) 稳定性:

视觉跟踪算法不应给后续的运动识别带来不稳定因素;

### (6) 简单性:

视觉跟踪算法越容易实现则越容易被普遍接受, 一个理想的视觉跟踪配置应简单易行。

除了以上特性之外, 在处理视觉跟踪问题时总体上有两种思路, 一种是自底向上的处理方法; 一种是自顶向下的处理方法. 当中自底向上的处理方法又称之为数据驱动的方法。

自底向上的方法一般是通过分析图像的内容来重建目标状态. 这种方法在计算量上是有效的, 但它的健壮性很大程度上依赖于对图像的分析能力, 在自底向上的跟踪思路中, 以 Divad Marr 的视觉计算理论最具代表性. Marr 认为视觉过程的主要任务是从 2D 图像中定量地恢复出图像所反映的场景中 3D 物体的形状和空间位置. 基于这个原因, Marr 将视觉过程分为了三

个阶段: 早期、中期和后期. 早期阶段由输入图像获得要素图, 包括图像中强度变化剧烈处的位置、几何分布和组织结构; 中期阶段实现根据输入图像和要素图获得的 2.5 维图; 后期阶段将根据输入图像、要素图以及 2.5D 图而获得的物体用 3D 描述<sup>[6]</sup>.

自顶向下方法产生和估计一系列基于目标模型的状态假设, 通过估计和校正这些图像观测的假设来实现跟踪<sup>[8]</sup>, 在自顶向下的跟踪思路中, 以 Bar-Shalom 的目标跟踪与关联思想最具代表性, 它是将目标跟踪问题转换为贝叶斯理论<sup>4</sup>框架下, 把视觉跟踪问题抽象为一种推理问题; 利用状态空间法建立目标状态观测模型及转换模型, 然后根据目标状态的先验概率及最新的观测值信息, 递归地估计系统状态的后验概率密度<sup>[10]</sup>. 早期, 本思路最先考虑的是使用卡尔曼滤波器(Kalman filter, KF), 但是其只能处理线性系统和高斯噪声的跟踪问题. 而后又考虑使用无极卡尔曼滤波算法(unscented Kalman filter, UKF), 因为它能处理非线性系统. 可是事实上, 在实际的视觉跟踪过程中, 受各种干扰因素影响, 目标的后验概率分布呈现非曲线、多模态的特点, 这是之前的方法解决不了的, 因此有了粒子滤波算法(Particle filter, PF, 也称为 Sequential Monte Carlo 方法)的诞生。

## 2.2 四类基本的跟踪算法

### (1) 基于区域的跟踪算法

基于区域的跟踪算法的基本思想是: 首先得到包含有目标的模板, 该模板是通过图像分割或者是人为预先确定的, 模板的形状一般为略大于模板的矩形, 当然也可为不规则的形状; 然后在序列图像中, 运用相关的算法跟踪目标, 对灰度图像可以采用基于纹理和特征的相关, 对于彩色的图像还可以利用基于颜色的相关<sup>[6]</sup>. 通常最常用的相关准则是平方和准则 SSD(the sum of squared differences). 此算法的优点在于, 当目标没有被遮挡时, 跟踪的精度非常高, 且十分稳定. 但是它的缺点在于费时, 当搜索范围较大时尤为突出, 而且如果目标变形了或者被较大地遮挡, 它的跟踪精度就会下降, 从而可能造成目标丢失. 具体相关研究可参见文献[19, 20].

3 非刚性就是物体会在力等外界因素下发生形态的变化。

4 贝叶斯决策是在信息不完全的情况下, 对部分未知的状态用主观概率估计, 然后用贝叶斯公式对发生的概率进行修正, 最后再利用期望值和修正概率做出最优决策。

(2) 基于模型的跟踪算法

基于模型的跟踪算法的基本思想是: 首先由先验知识获得目标的三维结构模型以及运动模型, 然后再根据实际的图像序列, 来确定目标的三维模型的各项参数, 从而确定出目标的瞬时运动参数<sup>[6]</sup>. 这种方法的优点在于可以精确地分析目标的三维运动轨迹, 就算在运动目标的姿态变化情况下仍然能够可靠地跟踪. 而其缺点在于运动分析的精度太过依赖几何模型的精度, 而生活中所以运动目标的精确几何模型的获得是十分困难的, 而其这种算法的实时性比较差. 具体相关研究可参见文献[21, 22].

(3) 基于特征的跟踪算法

基于特征的跟踪算法的基本思想是: 一般采用相关算法, 以目标的某个或某些局部特征(例如: 颜色、材质、纹理等)作为相关时的对象. 此算法的优点在于, 就算目标的某一部分被遮住, 只要还有一部分的特征可以被观察到, 就可以完成跟踪任务, 这里与基于区域的跟踪算法有很大的区别. 而此算法的难点在于, 如何确定对某个运动目标的唯一特征集, 这和提高算法的效率有很大关系. 根据文献[11, 12]中的讨论, 得出在特征提取时, 一般采用 Canny 算子获得目标的边缘特征, 采用 SUSAN 算子来获得目标的角点信息. 相关研究可参见文献[23, 24].

(4) 基于主动轮廓的跟踪算法

基于主动轮廓的跟踪算法的基本思想是: 在曲线内部控制力和图像力相互作用下使曲线运动到物体轮廓来完成对物体轮廓的搜索和提取<sup>[13]</sup>. 基于主动轮廓的跟踪一般利用的不是整体对象的空间和运动信息, 而是依靠视频对象边界的信息来实现跟踪. 它再现了目标的轮廓, 并在后续的视频帧中动态地更新目标的轮廓<sup>[8]</sup>. 主动轮廓模型(Snake)是 Kass 等人在 1987 年提出的. 相关研究可参见文献[25, 26].

2.3 其他跟踪算法

视觉跟踪算法还有很多种, 下表将其中一些常用的跟踪算法及优缺点进行了一个小的汇总, 如表 3<sup>[9]</sup>所示, 原理可参阅文献[27~29]等.

表 3 常用的视觉跟踪算法<sup>[9]</sup>

	CM+NCC	CM Track-er	NCC Trac-ker	FF Trac-ker	OF、KLT	LOC Trac-ker	LOC +BG	BD Track-er
--	--------	-------------	--------------	-------------	--------	--------------	---------	-------------

优点、适合场景	1. 跟踪器间切换, 可以处理不同的情形; 2. 和识别过程紧密结合.	可以很好地跟踪运动的肤色物体, 包括快速运动的变形物.	适合在固定区域内跟踪低速运动的物体, 更新模板可处理多尺度问题.	低速物体.	KLT >OF	适合情形和 NCC 差不多.	实验分析和 LOC 相比基本没有改进.	性能最低的一种.
缺陷与不足		在跟踪慢速物体准确.	不能跟踪快速物体.	不适合跟踪高速物体, 并且容易受同色物体干扰.	都不适合跟踪快速物体.	性能要不 NCC 低.		

LOC:local orientation correlation; FF:locks of features tracking; OF:optical flow tracking using templates on a regular grid; KLT:local feature tracking, KLT-tracker; BD:boosted detection.

3 多摄像头协同跟踪

在上文介绍的网络环境下智能监控的相关背景及应用技术的基础上, 接下来将介绍其核心研究——多摄像头的协同跟踪. 监控系统可分为单摄像头监控系统<sup>[17]</sup>和多摄像头监控系统. 在网络环境下的智能监控中, 仅仅靠一个摄像头监控去看到所有方位角度的场

景几乎是不可能的,因此就需要有多个可以转动的摄像头一起协同跟踪.当考虑多摄像头协同跟踪时,一般是从提高跟踪算法可靠性方面入手,如利用跟踪目标的三维信息解决目标的遮挡问题等.

在多摄像头监控系统中,首要是要解决两个关键问题:

① 必须解决单摄像头系统<sup>[31]</sup>中的一些跟踪问题,如运动目标的检测、跟踪与分类等;

② 必须处理多摄像头系统中出现的一些新的问题.下文就将对对其进行简要分析.

多摄像头协同跟踪遭遇到的挑战:

- ① 多个摄像机信息的融合;
- ② 针对目标和任务,摄像机的任务分配;
- ③ 场景的可视化.

多摄像头监控系统在一定程度上也解决了很多在单摄像头监控系统中遇到的难题,比如监控场景的目标的遮挡问题,摄像头的视野的局限问题,无法对目标进行三维跟踪等.多摄像头协同跟踪系统能全方位、大面积、长时间的对目标进行跟踪.因此多摄像头协同跟踪对网络环境下的智能监控具有十分重要的意义.

### 3.1 系统结构

多摄像头协同跟踪监控系统的分布式结构如图 4 所示<sup>[18]</sup>,每个客户端都带有一个摄像头,每个摄像头内进行目标检测、跟踪、分类的工作,然后由服务器和各个摄像头保持通信状态,融合来自多个摄像头的跟踪数据,调度摄像头进行物体跟踪.

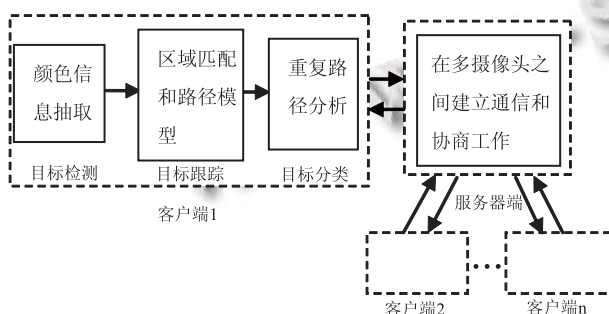


图 4 多摄像头监控系统结构

多摄像头协同跟踪系统主要由三个部分组成<sup>[30]</sup>:

第一部分:主系统部分,主要负责对场景的监控,对场景中的运动物体进行检测,并与服务器、子系统进行信息交互,控制子系统跟踪的开始与结束;

第二部分:子系统部分(可多个),主要负责接收主系统的命令,对监控场景内检测到的运动物体进行实时的跟踪和显示,同时把跟踪结果发送给服务器并显示给 PC 客户端;

第三部分:协同部分即服务器和 PC 客户端部分,服务器部分主要负责与底层网络数据传输部分的通信,接收主系统和子系统发送的相关摄像头的信息并转发给客户端的 IE,并接受用户在客户端的输入命令,发送给相应的摄像头,从而实现对特定摄像头的实时远程交互控制.

当然既然是在网络环境下协同跟踪那么网络协议必不可少,在网络实时监控中,主要的网络实时视频传输协议为 RTP/RTCP 协议.

### 3.2 网络拓扑结构

多摄像头协同跟踪的一个关键性问题的在于,采用何种网络拓扑结构才能合理组织摄像机去最大程度的满足通信的需求.接下来将简介常用的三种拓扑结构:中心式结构、分布式结构,以及层次结构<sup>[30]</sup>.

(1) 中心式结构:有一个中央处理单元,用来对所有摄像头得到的信息进行融合、检测和预测等处理.这种结构与单摄像头系统类似,具有较高的跟踪精度和较高的系统耦合度,但系统的通信开销相对较大且中央处理单元所承担的计算任务也比较大,这样就容易遇到计算处理时的瓶颈,而且系统性能很大程度上依赖于中央处理单元的性能,降低了系统的可靠性.

(2) 分布式结构:系统由多个相对独立的摄像头跟踪系统组成,每个子系统的局部跟踪结果送给其它子系统,参与其它子系统的跟踪.分布式系统的可靠性较高,且不会形成通信和计算瓶颈,但是系统的复杂性大大增加,同时导致成本开销也大大提高.

(3) 层次结构:每一个摄像头都是独立的子跟踪系统,可以维持本地的局部跟踪,并将局部跟踪的结果送到中央处理单元,由中央处理单元建立全局的目标信息和各个摄像头之间的关联,它是中心结构和分布式结构的一种折中方案.它不仅减少了中央处理单元的计算量,同时也降低了系统的通信开销.

在实际运用中,常常在这三种常用的拓扑结构的基础上,根据环境的实际情况衍生出更加符合实际情况的网络系统拓扑结构.

### 3.3 关键算法

#### 3.3.1 数据融合算法

在了解了多摄像头协同跟踪的框架基础上,可以发现一个多摄像头监控系统不能只是简单地收集单摄像头检测、跟踪所收集到的目标信息,还必须使所收集到的所有的观测结果连贯起来,形成连贯的动态场景描述.因此,有必要将同一目标在不同摄像头内所观测的结果融合成一个完整的整体目标描述.实现多摄像头的的数据融合需要利用目标外形变化估计和创建路径模型方法.具体方法在文献[18]有详细描述.数据融合的具体算法如下:

- ① 在单摄像头内检测和跟踪运动目标,抽取它们的轨迹特征,建立路径模型;
- ② 估计目标穿过多个摄像头的外形变化,计算新目标与数据库中存在的测试目标的距离  $D$ ;
- ③ 路径模型分析,比较新目标与测试目标的空间特征、速度特征和曲度特征;
- ④ 判断,如果距离  $D$  小于某一非零常量,或者新目标的路径特征满足空间和速度特征的限制,则认为新目标与测试目标相同,以同一标号标记此目标,否则将它置以新标记.

##### 3.3.1.1 目标外形变化估计<sup>[18]</sup>

对目标从一个摄像头到另一个摄像头的外形变化进行模型化,目标外形用颜色直方图来描述.以此来表示目标在多个摄像头之间运动时一般的颜色变化,使用这个线索建立多个摄像头之间的通信,识别在不同摄像头内运动的同一目标.对于有  $m$  级的直方图  $p$  有(1)、(2)式两种定义关系,对于有  $m$  级的直方图  $q$  有(3)、(4)式两种定义关系,其中  $\hat{p}, \hat{q}$  分别是  $p$  和  $q$  的最大似然估计.

$$p = \{\hat{p}_u\}_{u=1 \dots m} \quad (1); \quad \sum_{u=1}^m \hat{p}_u = 1 \quad (2);$$

$$q = \{\hat{q}_u\}_{u=1 \dots m} \quad (3); \quad \sum_{u=1}^m \hat{q}_u = 1 \quad (4);$$

计算它们之间距离  $d$  的定义如式(5)

$$d(p, q) = \sqrt{1 - \sum_{u=1}^m \sqrt{\hat{p}_u \hat{q}_u}} \quad (5)$$

通过分析  $d(p, q)$  值可知两个颜色直方图的不匹配

程度.一般情况下,场景中的不同摄像头内会有光照亮度变化可以使用亮度变换函数(BTF)来解决这个问题.考虑摄像头对  $C_i$  和  $C_j$  假定  $f_{ij}$  是摄像头  $C_i$  到摄像头  $C_j$  的 BTF, 则亮度变换函数如式(6).

$$B_j = f_{ij}(B_i) \quad (6)$$

其中,  $B_i$  和  $B_j$  分别表示摄像头  $C_i$  和  $C_j$  内图像的亮度值.

##### 3.3.1.2 创建路径模型<sup>[18]</sup>

设模型化路径为  $P(r, s, c)$ , 其中  $r$  为选取运动目标路径的空间特征,  $s$  为速度特征,  $c$  为曲度特征, 用这三个参数来区别空间上不相似、空间上相似但速度不同、曲度不同的轨迹.

###### (1) 空间特征 $r$

路径的空间特征表现在两个方面:

- 第一, 测试的轨迹中有 90% 的点在路径的包围内;
- 第二, 平均路径轨迹和测试路径轨迹之间的 Hausdorff 距离<sup>5</sup> 应小于路径包围边界之间的 Hausdorff 距离. 满足这两个条件则认为路径在空间上相似, 否则, 标记为异常.

###### (2) 速度特征 $s$

对于任一条轨迹  $P_i(x_i, y_i, t_i)$  来说,  $x_i$  和  $y_i$  是轨迹  $P_i$  在时刻  $t_i$  的  $x$  坐标和  $y$  坐标, 则计算速度的定义如式(7)

$$v'_i = \left( \frac{x_{i+1} - x_i}{t_{i+1} - t_i}, \frac{y_{i+1} - y_i}{t_{i+1} - t_i} \right), i = 0, 1, \dots, N-1 \quad (7)$$

抽取测试轨迹的速度, 使用高斯分布模型化路径模型中轨迹的速度, 使用 Mahalanobis 距离<sup>6</sup> 判断测试的速度是否异常, 判断方法如式(8)

$$d(v'_i, m_p) = \sqrt{(v'_i - m_p)^T (\Sigma)^{-1} (v'_i - m_p)} \quad (8)$$

其中,  $v'_i$  是测试轨迹的速度,  $m_p$  是均值,  $\Sigma$  是路径轨迹分布的协矩阵,  $d$  是其马氏距离.

###### (3) 曲度特征 $c$

使用速度和加速度计算轨迹的曲度, 曲度的定义如式(9)

$$k = \frac{\sqrt{y''(t)^2 + x''(t)^2 + (x'(t)y''(t) - x''(t)y'(t))^2}}{(\sqrt{x'(t)^2 + y'(t)^2} + 1)^3} \quad (9)$$

5 Hausdorff 距离是描述两组点集之间相似程度的一种量度.

6 Mahalanobis 距离表示数据的协方差距离, 是一种有效的计算两个未知样本集的相似度的方法.



其中,  $x'$  和  $y'$  是速度  $v'$  的  $x$  部分和  $y$  部分. 对路径模型中的轨迹, 测量曲度  $k$  的均值和方差, 符合高斯分布. 使用 Mahalanobis 距离方法比较测试路径的曲度和路径的曲度分布, 根据比较结果检测出不规则运动.

### 3.3.2 多摄像头协同算法

在监控系统中, 每一个运动目标都可能被多个摄像头同时跟踪, 如何去选择最适合的摄像头去跟踪运动目标, 在系统资源紧张时是十分重要的. 在参考文献[18]中引用了一个权函数来描述多摄像头中的任务分配方案, 以达到监控系统协同多个摄像头无缝跟踪监控范围内的运动目标. 做法如下<sup>[18]</sup>:

(1) 确定权函数中的三个变量:

① 可见性  $V_{ij}$ , 表示摄像头  $C_i$  能否看到目标  $O_j$ , 值用 0 或 1 表示;

② 优先级  $P_j$ , 表示用户分配给系统任务  $T_j$  的优先级, 依据多摄像头数据融合的结果进行动态分配, 值越大表示优先级越高;

③ 距离  $D_{ij}$ , 表示摄像头  $C_i$  与目标  $O_j$  在物理空间上的距离;

(2) 权函数的定义如式(10)

$$W_{ij} = V_{ij}(k_1 P_j + k_2 D_{ij}) \quad (10)$$

$k_1$  和  $k_2$  是  $P_j$  和  $D_{ij}$  的调谐常量,  $W_{ij}$  代表摄像头  $C_i$  与目标  $O_j$  的权值, 它越大, 摄像头  $C_i$  跟踪目标  $O_j$  的概率就越大, 值为零时, 则摄像头  $C_i$  不对目标进行跟踪;

(3) 确定分配方案:

对每一个任务选择一个有最大非零权值的摄像头来执行它;

有更高优先级的任务在任何时候到来时, 执行高优先级的任务;

分配完所有任务后, 没有任务的摄像头执行最基本的单摄像头跟踪任务.

(4) 采用的自动分配方式:

优先级高的任务有高的权值;

目标选择在某一区域有更好视角的摄像头, 如果最近的摄像头被遮挡, 就选择比它稍远的那个;

在目标离开一个摄像头的监控区域进入另一个摄像头的视野的过程中, 前一个摄像头的跟踪权值将减小, 后一个的将增加, 直到后一个自动执行这个监控任务.

## 3.4 相关问题研究现状

### 3.4.1 跟踪方法研究现状

上文 2.2 节中已对四类跟踪方法的基本思想进行

了简要介绍, 下文将简要介绍多摄像头协同跟踪中这四类基本方法的研究现状<sup>[32]</sup>.

#### 3.4.1.1 基于模型的跟踪

Boyd 等人于 1999 年提出了使用网络单位图构建基于网络的活动模型<sup>[38]</sup>.

Ellis 等人于 2002 年提出了基于目标运动轨迹的空间和频率分布的活动模型, 不过此活动模型适合单个摄像头场景和视野域重叠的多个摄像头场景, 但是不适合非重叠的多摄像头系统, 即这些方法没有考虑摄像头之间有非重叠区域或者多摄像头系统中有遮挡区域的情况<sup>[39]</sup>.

Makris 和 Junejo 等继续在 2004 年提出了无监督学习的活动模型和多特征路径模型<sup>[42,43]</sup>, 其中文献[42]中仅利用来自摄像头的空间和时间信息, 不受摄像头特征和方向的限制, 在非重叠的摄像头区域(即盲区)建立活动路径的内连网, 确定各摄像头间的空间关系, 不过由此算法产生的活动路径内连网仍包含一些冗余链.

Jain 等提出了利用校准的摄像头和环境模型来获得人的 3D 位置<sup>[44]</sup>.

Javed 等利用摄像头的 FOV 边界建立多个摄像头之间的通信, 通过划定多摄像头之间的 FOV 界线来建立运动目标之间的对应关系<sup>[45]</sup>.

Lim 等使用场景的预定视角创建场景模型<sup>[46]</sup>.

#### 3.4.1.2 基于特征

Huang 等于 1997 年描述了跟踪穿过高速公路上两个摄像头的车辆的方法<sup>[47]</sup>.

J.Orwell 等提出了在视觉监控中使用目标颜色进行跟踪的系统<sup>[48]</sup>.

Javed 等提出了在非重叠多摄像头系统中实现目标跟踪的外形模型, 此模型是在基于特征的跟踪算法的基础上提出的<sup>[49]</sup>.

#### 3.4.1.3 基于区域

在参考文献[50]中, 描述了在多摄像头系统中使用区域匹配检测和跟踪多人的算法. 该算法首先分割每一个图像, 然后比较每一对图像的区域, 找到分割结果的中心, 通过匹配中心点找到场景中可能对人通信的三维点(3D, three-dimension), 最后将 3D 点投影到 2D 平面, 使用拒绝框架方案对人的 2D 位置给出鲁棒性估计.

#### 3.4.1.4 基于活动轮廓

在文献[51]中, 描述了在不同天气条件下的室外

多摄像头视频监控系统. 文中定义了新的可信措施——外形比 AR(Appearance Ratio)来自动评价每个时刻传感器的性能, 通过比较它们的 AR, 系统能够选择最合适的摄像头来执行特定的任务.

### 3.4.2 协同算法研究现状

在多摄像头协同监控系统中, 针对如何选择最佳视角, 如何排除环境变化干扰等问题, 国内外已做了一些相关研究, 提出了用以解决的相关算法.

Snidaro 等描述了一个室外有天气变化的多摄像头视频监控系统. 系统中定义了一个新的可信措施——外形比(AR), 通过比较传感器的 ARs 选择最合适的摄像头来执行特定的任务<sup>[33]</sup>.

Nummiaro 等描述了一种通过彩色颗粒滤波方式从不同摄像头中自动选择最佳视角的方法<sup>[34]</sup>.

Lim 等设计了一个可变的覆盖大范围的视觉监控系统. 系统利用基于图像的信息控制摄像头, 使用场景的预定视角创建场景模型<sup>[35]</sup>.

Henriksson 等提出了一种基于多摄像头的反馈控制的动态资源分配方法, 通过最小化位置估计误差的协方差来选择一个合适的摄像头子集<sup>[36]</sup>.

Nguyen 等提出了分布式监控系统中多摄像头的协同算法. 该算法根据人与摄像头之间的距离以及目标的遮挡情况来分配目标给摄像头, 引入了服务质量(Qos, Quality of Service)函数评价监控系统性能<sup>[37]</sup>.

## 4 总结

本文以网络环境下的智能监控的核心技术为切入点, 介绍了计算机视觉的发展简史、核心问题、原理和关键技术. 尤其是针对多摄像头协同跟踪技术, 在剖析视觉跟踪技术的组成部分、基本思路、主要应用和常用算法的基础上, 理解多摄像头协同跟踪的相关理论知识. 从它的研究现状、相关算法、核心问题以及未来的展望等方面进行了分析, 并依据各个技术的发展趋势和社会的需求, 对网络环境下的智能监控的未来发展进行了预测与展望.

通过对已有的网络环境下智能监控的核心技术的发展的简要回顾和分析, 结合近年出现的研究热点, 在现有的科研人员的推测的基础上, 大胆提出几个作者认为的今后网络环境下的智能监控领域将有可能的发展方向.

从 2009 年末微软公司所推出的 xbox360 kinect 体

感游戏在全世界引起了巨大反响, 到近期微软新推出体感系列 Xbox One, 参考文献[16]中的理论, 可以推知 kinect 体感技术作为计算机视觉中较为新奇的应用将来很有可能成为网络环境下的智能监控的一个研究热点. 朱松纯教授曾提出过一个有趣的观点, 计算机和人最大区别在于人能够做梦而计算机却不能, 所以研究计算机视觉首先就要解决计算机做梦的问题, 让计算机也能够根据各种零散的信息来生成特定的图像和图形. 作者认为这也是网络环境下的智能监控的一个很有潜力的发展方向. 如果这个观点能实现, 基于网络强大的覆盖能力, 网络环境下的智能监控系统必将能形成一个大型的, 让人能“梦想成真”的虚拟世界. 总之, 网络环境下的智能监控的发展是势不可挡的, 或许曾经人们天马行空的想法很多都能随着它的发展而得以实现.

## 参考文献

- 1 林海平. 智能监控开启新时代. 大科技. 2013, 2: 301-302.
- 2 Marr D. Vision - A Computational Investigation into the Human Representation and Processing of Visual Information. New York. WH Freeman.
- 3 项海兵. 计算机视觉发展中存在的问题. 新浪潮, 1996, (4): 29-30.
- 4 许志杰, 王晶, 刘颖, 范九伦. 计算机视觉核心技术现状与展望. 西安邮电学院学报, 2012, 17(6): 1-8.
- 5 陈丹. 计算机视觉技术的发展及应用. Computer Knowledge and Technology, 2008, 12(4): 2449-2450.
- 6 侯志强, 韩崇昭. 视觉跟踪技术综述. 自动化学报, 2006, (4): 603-617.
- 7 张进. 视觉跟踪技术发展和难点问题的分析. 信息技术与信息化, 2008, (6): 63-64.
- 8 杨戈, 刘宏. 视觉跟踪算法综述. 智能系统学报. 2010, 5(2): 95-105.
- 9 VisualTrackingReview. <http://www.cnblogs.com/CVArt/archive/2011/07/03/2096683.html>. 2011-7-3.
- 10 王法胜, 郭权. 视觉跟踪中的粒子滤波算法研究进展. 山西大学学报(自然科学版), 2011, 34(4): 528-533.
- 11 Itra PM, Murthy C, Pal S. Unsupervised feature selection using feature similarity. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(3): 301-312.
- 12 Kaneko T, Hori O. Feature selection for reliable tracking

- using template matching. Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR03. Madison, Wisconsin, USA. IEEE Press. 2003.1. 796–802.
- 13 岑峰,戚飞虎,曾文珺.基于边缘吸引力场正则化的短程线主动轮廓模型.电子学报,2003,31(1):82–84.
- 14 戴维.A.福赛斯,简.泊斯.计算机视觉——一种现代方法.电子工业出版社.2004.
- 15 刘露.人脸识别技术新发展.百科知识,2013,(2):22.
- 16 刘鑫,许华荣,胡占义.基于 GPU 和 Kinect 的快速物体重建.自动化学报,2012,38(8):1288–1297.
- 17 谢鹏程.基于单摄像头的运动目标跟踪定位技术研究.计算机光盘软件与应用,2012,18:82–84.
- 18 陈伟宏,肖卫初.监控系统中的多摄像头协同算法.重庆工学院学报,2008,22(4):117–123.
- 19 Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564–577.
- 20 Hager G, Belhumeur P. Efficient region tracking with parametric models of geometry and illumination. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998, 20(10): 1025–1039.
- 21 Jurie F. Tracking objects with a recognition algorithm. Pattern Recognition Letters, 1998, 19(3-4): 331–340.
- 22 Drummond T, Cipolla R. Real-time visual tracking of complex structures. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(7): 932–946.
- 23 Tissainayagam P, Suter D. Object tracking in image sequence using point feature. Pattern Recognition, 2005, 38(1): 105–113.
- 24 Nickels K, Hutchinson S. Estimating uncertainty in SSD-based feature tracking. Image and Vision Computing, 2002, 20(1): 47–58
- 25 Dgoldenberg R, Kimmel R, Rivlin E, Rudzsky M. Fast geodesic active contours. IEEE Trans. on Image Processing, 2001, 10(10): 1467–1475.
- 26 Mansouri A. Region tracking via level set PDEs without motion computation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(7): 947–961.
- 27 Stenger B, Woodley T, Cipolla R. A vision-based remote control. Studies in Computational Intelligence, 2010, 285: 233–262.
- 28 Lim J, Ross D, Lin RS, Yang MH. Incremental learning for visual tracking. Advances in neural information processing systems, 2004, 17: 793–800.
- 29 Kulaczewski MB, Siegel HJ. SIMD and mixed-mode implementations of a visualtracking algorithm. Parallel. Processing Symposium, 1998. IPPS/SPDP 1998.
- 30 章建祥.多摄像头协同跟踪系统关键技术研究[硕士学位论文].杭州:浙江工商大学.2011.
- 31 Intille S, Davis J, Bobick A. Real-time closed-world tracking. IEEE Conference on Computer Vision and Pattern Recognition. San Juan, Puerto Rico.1997. 697–703.
- 32 陈伟宏.基于多摄像头的目标跟踪算法研究[硕士学位论文].长沙:湖南大学.2005.
- 33 Snidaro L, Niu R, Varshney PK. Automatic camera selection and fusion for outdoor surveillance under changing weather conditions. IEEE Conference on Advanced Video and Signal Based Surveillance. Miami Beach, Florida, USA. 2003.364–369.
- 34 Nummiaro K, Koller-Meier E, Svoboda T. Color-based object tracking in multi-camera environments. 25th Pattern Recognition Symposium. Magdeburg. 2003. 591–599.
- 35 Lim SN, Davis LS, Elgammal A. A scalable image-based multi-camera visual surveillance system. IEEE Conference on Advanced Video and Signal Based Surveillance. Miami, Florida, USA. 2003. 205–212.
- 36 Henriksson D, Olsson T. Maximizing the use of computer resources in multi-camera feedback control. Proc. of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium(RTAS). Toronto. 2004.360–367.
- 37 Nguyen NT, Venkatsh S, West G. Multiple camera coordination in a surveillance system. Acta Automatica Sinica, 2003, 29(3): 408–422.
- 38 Boyd JE, Meloche J, Vardi Y. Statistical tracking in video traffic-surveillance. Proc. ICCV99. Corfu, Greece. 1999. 163–168.
- 39 Ellis T. Multi-camera video surveillance. Proc. of 36th Annual 2002 International Carnahan Conference. 2002. 228–233.
- 40 Makris D, Ellis TJ. Path detection in video surveillance. Image and Vision Computing, 2002, 20(12): 895–903.

- 41 Makris D, Ellis T. Automatic learning of an activity- based semantic scene model. IEEE International Conference on Advanced Video and Signal Based Surveillance. Miami, FL, USA. 2003. 183–188.
- 42 Makris D, Ellis T, Black J. Bridging the gaps between cameras. Proc. the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington. 2004, 2. 205–210.
- 43 Junejo I, Javed O, Shah M. Multi feature path modeling for video surveillance. 17th conference of the International Conference on Pattern Recognition (ICPR). Cambridge. 2004. 716–719.
- 44 Jain R, Wakimoto K. Multiple perspective interactive video. Proc. of IEEE International Conference on Multimedia Computing and Systems. 1995. 202–211.
- 45 Javed O, Rasheed Z, Alata O. KnightM: A real time surveillance system for multiple overlapping and non-overlapping cameras. Invited paper in IEEE conference on Multimedia and Expo, Special Session on Multi-Camera Surveillance Systems. Baltimore. 2003.
- 46 Lim SN, Davis LS, Elgammal A. A scalable image-based multi-camera visual surveillance system. IEEE Conference on Advanced Video and Signal Based Surveillance. Miami, Florida, US. 2003. 205–212.
- 47 Huang T, Russell S. Object identification in a bayesian context. Proc. of IJCAI. Nagoya, Japan. 1997. 1276–1283.
- 48 Orwell J, Remagnino P, Jones GA. Multi-camera colour tracking. IEEE International Workshop on Visual Surveillance. Fort Collins, Colorado. 1999. 14–21.
- 49 Javed O, Shafique K, Shah M. Appearance modeling for tracking in multiple non-overlapping cameras. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). San Diego. 2005, II. 26–33.
- 50 Mittal A, Davis L. Unified multi-camera detection and tracking using region-matching. IEEE Workshop on Multi-Object Tracking. Vancouver, Canada. 2001. 3–10.
- 51 Snidaro L, Niu R, Varshney PK. Automatic camera selection and fusion for outdoor surveillance under changing weather conditions. IEEE Conference on Advanced Video and Signal Based Surveillance. Miami Beach, Florida. 2003. 364–369.