

视觉跟踪技术综述¹⁾

侯志强^{1,2} 韩崇昭¹

¹⁾(西安交通大学电子与信息工程学院综合自动化研究所 西安 710049)

²⁾(空军工程大学电讯工程学院指挥自动化教研室 西安 710077)

(E-mail: hou-zhq@sohu.com, czhan@mail.xjtu.edu.cn)

摘 要 视觉跟踪问题是当前计算机视觉领域中的热点问题, 本文对这一问题进行了详细的介绍. 首先, 对视觉跟踪技术在视频监视、图像压缩和三维重构等三个主要方面的应用进行了论述. 其次, 详细阐述了该技术的研究现状, 介绍了其中的一些常用方法, 为清楚说明这些方法, 先对视觉跟踪问题进行了分类, 然后介绍了处理视觉跟踪问题的两种思路即自底向上和自顶向下的思路, 最后将具体的视觉跟踪方法分为四类进行了介绍, 这四类分别是基于区域的跟踪、基于特征的跟踪、基于变形模板的跟踪和基于模型的跟踪. 最后, 从控制论角度给出视觉跟踪算法所面临的难点, 即算法要满足鲁棒性、准确性和快速性要求时所遇到的困难, 并对视觉跟踪问题的研究前景进行了展望.

关键词 计算机视觉, 视觉跟踪, 图像序列, 监视系统, MPEG-4, 三维重构
中图分类号 TP391

A Survey of Visual Tracking

HOU Zhi-Qiang^{1,2} HAN Chong-Zhao¹

¹⁾(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049)

²⁾(School of Telecommunication Engineering, Air Force Engineering University, Xi'an 710077)

(E-mail: hou-zhq@sohu.com, czhan@mail.xjtu.edu.cn)

Abstract This paper introduces in detail the research of visual tracking which is a hot spot currently in the domain of computer vision. Firstly, the applications of visual tracking in three areas including visual surveillance, image compression, and 3-D reconstruction are discussed. Secondly, the state of the art about visual tracking is introduced, especially the main approaches of visual tracking are shown. In order to explain these methods clearly, the problems of visual tracking are classified. Then two ways to research the visual tracking problem are presented. And the visual tracking algorithms are classified into four classes: the area-based methods, feature-based methods, deformable-template-based methods and model-based methods. Finally, from the point of view of control theory, the difficulties of visual tracking are discussed that the algorithms should have robustness accuracy and be fast. Meanwhile, some future directions of visual tracking are also addressed shortly.

Key words Computer vision, visual tracking, image sequence, surveillance system, MPEG-4, 3D reconstruction

1) 国家 973 项目 (2001CB309403) 和国家自然科学基金 (60574033) 资助

Supported by National Key Fundamental Research (973) Program (2001CB309403) and National Natural Science Foundation of P. R. China(60574033)

收稿日期 2005-7-12 收修改稿日期 2006-2-25

Received July 12, 2005; in revised form February 25, 2006

1 引言

随着计算机技术的不断发展, 计算能力得到了极大的提高, 使得利用计算机实现人类的视觉功能成为目前计算机领域中最热门的课题之一. 视觉跟踪 (Visual tracking) 问题是计算机视觉领域中的一个重要问题. 所谓视觉跟踪, 就是指对图像序列中的运动目标进行检测, 提取、识别和跟踪, 获得运动目标的运动参数, 如位置、速度、加速度等, 以及运动轨迹, 从而进行进一步处理与分析, 实现对运动目标的行为理解, 以完成更高一级的任务. 近年来, 国际上对视觉跟踪问题正处于研究高潮中. 许多重要国际期刊 (IEEE Transactions on Pattern Analysis and Machine Intelligence; IEEE Transactions on Image Processing; IEEE Transactions on Medical Imaging; IEEE Transactions on Circuits and Systems for Video Technology; IEEE Transactions on Vehicular Technology; International Journal of Computer Vision; Computer Vision and Image Understanding; Image and Vision Computing; Pattern Recognition; Pattern Recognition Letters; Machine Vision and Application; Real-time Imaging; *etc.*) 以及重要国际会议 (ICCV; CVPR; ECCV; ICIP; ICPR; IWVS; *etc.*) 发表了大量有关视觉跟踪方面的论文. 其中一些重要国际期刊还就视觉跟踪问题发行专刊进行了讨论 (IEEE Transactions on Pattern Analysis and Machine Intelligence, **22**(8), 2000; Proceedings of the IEEE, **89**(10), 2001; International Journal of Computer Vision, **37**, June 2000; Real-time Imaging, **7**(5), 2001).

视觉跟踪问题引起广泛关注是由于它能够应用于民用和军事的许多领域, 主要包括三个方面: 视频监视, 图像压缩和三维重构. 本文在随后一节中对视觉跟踪技术在这三方面的应用进行详细介绍. 本文的其它部分是这样安排的, 第 3 小节将对常用的一些视觉跟踪算法进行分类介绍; 在第 4 小节中, 对目前视觉跟踪技术中的难点问题以及未来的发展趋势作简要说明; 结束语在第 5 小节给出.

2 视觉跟踪技术的应用

2.1 视频监视

视频监视系统应用的范围非常广, 最常见的是对于民宅、停车场、公共场合、银行等的监视, 以防止偷盗、破坏行为的发生, 保障社会的安全. 实时监控系统 W⁴ 用于对人的行为的监视^[1], 用于民宅、停车场、银行等场合, 能够对携带物品的可疑人员进行有效的监视. DETER (Detection of events for threat evaluation and Recognition) 系统^[2] 用于对停车场进行监控, 以防止车辆被盗, 该系统在提取出运动目标之后, 能够对运动目标的行为进行跟踪, 通过跟踪轨迹对运动目标行为模式进行判定, 有效地提供异常行为报警, 从而确保停车场的安全. 1997 年, 美国国防部高级研究项目署 (Defense advanced research projects agency, DARPA) 设立了以卡内基梅隆大学为首, 麻省理工学院等高校参与的视觉监控项目 VSAM (Visual surveillance and monitoring)^[3], 该系统能够用于普通民用场景以及战场的实时监控, 卡内基梅隆大学在该系统建成之后, 建立了一个校园监控系统^[4]. PASSWORDS (Parallel and advanced surveillance system with operator assistance for revealing dangerous situations) 系统^[5] 是一个智能视频图像分析系统, 用于实现视频监视以满足不同用户的安全要求. Pfinder (Person finder) 系统^[6] 用于实现对室内人员行为的实时监控与判定.

在交通系统中, 视觉跟踪研究也具有非常广泛的应用, 主要包括交通流量控制、车辆异常行为检测、行人行为判定、智能车辆等很多方面. Coifman 等人^[7] 建立了一个基于视频图像处理系统 (Video image processing system, VIPS) 的交通监控系统, 用于交通流量监

控以及不同车型的统计, 这一方面的研究工作还有很多^[8~10], 研究要解决的重点问题是如何准确的实现对车辆的分割和跟踪. 道路上车辆异常行为检测在交通事故的预防和事后处理中都具有重要意义, 其核心技术是如何分割和跟踪车辆并对所跟踪车辆的运动轨迹进行判定, Tai 等人^[11] 研究了一个用于交通事故检测的视频监视系统, 能够自动检测运动车辆并对其运动轨迹进行判定. VISATRAM(Vision system for automatic TRAffic monitoring) 系统^[12] 能够对各个车道的车辆行为进行监控, 保证交通通畅. Haag 和 Nagel^[13] 专门对机动车辆跟踪问题进行了研究. 行人行为判定和跟踪对于车辆的安全驾驶以及道路交通管理有着重要帮助, Pai 等人^[14] 专门研究了十字路口的行人检测与跟踪以保证驾驶员在十字路口能够安全驾驶, Masoud 和 Papanikolopoulos^[15] 通过对道路行人的跟踪以实现对行人的记数, 为道路交通管理提供信息. 智能车辆是目前计算机视觉研究中一个热点^[16~18], 实现自动车辆驾驶是这一研究的最终目的, 其基本思路是利用安装在车辆上的摄像机实现对道路的检测、对前方车辆和行人的检测与跟踪, 以保证车辆的安全行驶, 同时, 该技术可以很容易的转移到机器人技术上^[19].

视频监视系统除了在上述民用方面应用之外, 在军事上也有极其重要的应用价值. 与民用技术相比, 在军事领域对于视频监视系统的要求要高得多, 主要由于战场环境远比普通的民用环境复杂和苛刻. 例如, 对隐藏的狙击手 (Sniper) 进行检测和跟踪就非常困难, 在这种情况下, 普通的民用技术由于算法所采用假设的局限性往往不能使用, LOTS(Lehigh omnidirectional tracking system) 系统^[20] 针对这一问题进行了详细的讨论. 从飞行器上对空中或地面目标进行跟踪和识别在现代战争中具有重要的意义, 是实施精确打击的必要手段, 但是在飞行状态下对运动或静止的目标实施跟踪与识别非常困难, AVS(Aerial video surveillance) 系统^[21,22] 在这一方面进行了大量的研究. 对战场环境中的敌方运动目标进行快速准确的搜索和跟踪, 是视频监视系统在军事上的一个重要任务, 但是恶劣的战场环境要求视频监视系统具有很强的适应性, 能够对快速变化的运动目标实施稳定可靠地跟踪, Tsao 和 Wen 对此问题进行了讨论^[23].

2.2 图像压缩

在新一代的视频图像压缩编码标准 MPEG-4 中, 提出了一个不同于以前的视频编码的新概念视频对象平面 (Video object planes, VOP)^[24], 以实现基于图像内容 (Content-based) 的压缩. 在压缩编码过程中, 视频中的场景被分割成若干个 VOP, 不同的 VOP 根据其特点采用不同的编码策略, 例如, 场景中的背景在很多情况下基本上保持不变或只有很少部分的变化, 则可以在连续的多帧图像中使用相同的编码或只有很少改动的编码来表示背景, 从而大大提高了压缩效率. 这种方案具有很高的压缩比, 同时还具有良好的视觉效果, 便于进行基于内容的交互和查询. 但是, 这一压缩编码的技术难点在于如何在任意场景下实现对运动目标的自动提取与分割^[25], 而这些问题也正是视频跟踪所要解决的关键问题.

与用于视频监视的图像跟踪技术不同的是, 在图像压缩编码中, 要求对运动目标的分割和提取的精度非常高, 能够尽量完整准确地将运动目标从背景中提取出来, 在这方面大量的文献对此进行研究^[26~31]. 此外, 一些重要国际期刊还刊发了有关 MPEG-4 压缩编码的的专刊 (IEEE Transactions on Circuits and Systems for Video Technology, 7(1), 1997; Signal Processing, 66(2), 1998; Signal Processing: Image Communication, 15(4-5), Jan. 2000; etc.).

2.3 三维重构

三维重构在医学上有着重要的实用价值. 在获得人体器官切片的医学图像之后, 通过对人体器官的三维重构能够为医生的准确诊断提供极其重要和直观的帮助. 三维重构还能在目标识别问题上提供一个很好的解决方案. 如果能够通过一组图像序列将其中的 2D 运

动目标重构成 3D 运动目标, 则无论该运动目标的姿态如何, 由于已知其三维信息, 总将该运动目标识别出来, 从而实施有效的跟踪. 这里要说明的是, 由于运动目标姿态的变化而导致跟踪过程中目标的丢失是目前视觉跟踪问题中的一个难点, 而三维重构技术能在克服这一难点方面提供一个潜在的解决办法. 三维重构与视觉跟踪要解决的问题都是在图像序列中找出相同或相似的感兴趣目标, 不同的是三维重构需要将该目标重构为三维形式, 而视觉跟踪却要得到该目标的相关运动参数. 有关三维重构的文献可以参见 [32~36].

视觉跟踪技术除了在以上所讲的三个大的方面应用之外, 其相关技术还可以应用到很多方面. 例如, 在视频监视系统中对人的行为的识别研究可以应用于人-机感知接口 (Human-computer perceptual interface) 中, 其主要解决的问题是基于计算机视觉的人的手势和表情识别^[37~39]. 对于人脸的检测、识别与跟踪 (Facial detection, recognition and tracking) 还能在远程电视会议中大大降低图像传输的比特率, 并使画面及时锁定在讲话人的身上^[27,40], 有关这方面的研究还可以参考重要国际期刊专刊 (IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 1997). 而上述应用中所涉及到的对于人运动的视觉跟踪是视觉跟踪问题中又一个研究的热点问题, 有大量文献对这一问题进行了研究, 其中代表性的综述性参考文献有 [41~44]. 在视觉跟踪技术中, 还可以利用图像序列中感兴趣目标的相似性实现对目标的识别与跟踪, 这一技术能够被利用来进行图像检索和图像查询 (Image Retrieval)^[45~47], 有关图像查询的文献在重要国际期刊上有专刊 (Computer Vision and Image Understanding, 94(1~3), 2004) 进行讨论. 此外, 视觉跟踪技术也是虚拟实现 (Virtual reality) 中的核心技术之一^[48~50].

综上所述, 开展视觉跟踪研究非常有意义, 其研究成果具有广阔的应用前景.

3 视觉跟踪技术的研究现状

视觉跟踪问题是随着计算机技术的发展而逐渐成为研究热点的. 二十世纪 80 年代以前, 由于计算机技术的限制, 对图像的处理与分析主要以静态图像为主^[51], 在动态图像序列的分析中, 对运动目标的跟踪带有很强的静态图像分析的特点^[52]. 二十世纪 80 年代初光流法 (Optical flow)^[53,54] 被提出之后, 动态图像序列分析进入了一个研究的高潮. 其中对光流法的研究热潮从其产生一直持续到了二十世纪 90 年代中期, 有关光流法的综述性文章可以参见文献 [55], 直到最近, 还有文章对 L-K 方法进行了回顾^[56]. 但是, 光流法所需的运算量即使在现阶段对计算机而言也是太大, 很难满足实时性要求, 同时由于采用假设的局限性使得光流法对噪声特别敏感, 很容易产生错误的结果, 这些缺点造成光流法与实际使用之间还存在着很大的距离. 从二十世纪 80 年代之后, 还出现了其他众多的视觉跟踪算法, 下边将介绍其中常用的一些方法. 为了清楚地对这些方法进行说明, 首先对视觉跟踪问题进行了分类, 然后介绍了处理视觉跟踪问题的两种思路, 最后对具体的视觉跟踪方法进行了分类介绍.

3.1 视觉跟踪问题分类

按不同标准可将视觉跟踪问题分成很多类型^[41,43,57]. 下边给出一些常见的分类.

1) 摄像机的数目: 单摄像机 VS 多摄像机. 根据所使用摄像机数目的多少, 可以将视觉跟踪问题分为单摄像机 (Monocular camera) 视觉跟踪和多摄像机视觉跟踪 (Multiple cameras). 目前, 绝大多数研究都属于单摄像机的视觉跟踪问题^[1,6,7,8,14,15,58], 其中对于图像压缩技术而言, 纯粹只考虑单摄像机获得的图像序列. 但是, 在实际的视频监视系统中, 单摄像机所能监视的区域很有限, 因此, 在一些大的公共场合, 更多的是使用多摄像机以扩大监视的范围^[2,4,57,59]. 此外, 单摄像机不能获得运动目标的距离信息, 而双目或三目视觉系统却能够得到运动目标的距离信息^[60]. 在对运动目标跟踪过程中, 往往发生运动目标

被遮挡或暂时消失等情况从而丢失运动目标, 这一问题 (Occlusion) 是运动目标跟踪中又一个难点问题, 使用单摄像机解决这一问题相当困难, 而使用多摄像机能在很大程度上解决这一难题^[61]. 但多摄像机跟踪面临一个单摄像机跟踪所没有的难题, 那就是如何将不同摄像机所获得的不同视角的运动目标信息正确地对应起来, 由于这一问题的难度甚至比单纯跟踪问题的难度还要大, 因此有关多摄像机跟踪的文献并不是很多.

2) 摄像机是否运动: 摄像机静止 VS 摄像机运动. 根据摄像机是否运动可以将图像跟踪问题分为摄像机静止时的运动目标跟踪和摄像机运动时的运动目标跟踪. 对于大多数的视频监控监视系统而言, 都是在摄像机静止时对某个需要特殊关注的区域进行监视^[1,2,6,7,11,14,15,20], 此刻, 背景是静止的, 而作为前景的运动目标是移动的, 这种情况下通常使用背景差法 (Background subtraction) 进行变化检测 (Change detection) 能够取得相当不错的结果, 已经应用到了很多的实际系统中^[1,2,6,11,12,20]. 但是, 在很多情况下, 摄像机是运动的, 其运动的形式可以分为两种, 一种是摄像机的支架固定, 但摄像机可以偏转 (Pan), 俯仰 (Tilt) 以及缩放 (Zoom)^[62]; 另一种是摄像机装在某个移动的载体上的, 如移动的车辆^[17,63] 或飞行的飞机^[22,64] 上. 在这些情况下, 由于背景和前景都是运动的, 要准确检测并跟踪运动目标是一件困难的事情, 通常, 如果摄像机的运动方式被限定在垂直于光轴的平面上, 可以采用图像拼接 (Mosaic) 的方法将背景拼在一起, 然后按照摄像机静止时的跟踪方法进行处理^[65~67]; 如果摄像机是沿着光轴运动的, 光流法或许是一种可以考虑的方法^[63,68].

3) 场景中运动目标的数目: 单运动目标 VS 多运动目标. 根据视频场景中运动目标数目的多少, 视觉跟踪问题可以被分为单目标跟踪和多目标跟踪两类. 单目标跟踪并不是一个简单的问题, 首先从场景中完整准确地检测和提取目标就比较困难, 例如在 MPEG-4 编码中, 要将画面上一个运动的目标从复杂的背景中完整地提取出来就很不容易^[28,69]. 由于场景中噪声的影响, 往往不能准确地检测到真正的运动目标; 即使检测到了运动目标, 也很难将其完整地提取出来, 例如由于阴影或光照 (Illumination) 的影响等造成提取出来的运动目标不完整或中心有空洞^[20]. 而多目标跟踪比单目标跟踪要困难得多, 在多目标跟踪过程中, 必须考虑到多个目标在场景中会互相遮挡 (Occlusion), 合并 (Merge), 分离 (Split) 等情况, 这是多目标跟踪问题的难点, 有大量文献是关于这些问题的^[70~72].

4) 场景中运动目标的类型: 刚体 VS 非刚体. 根据视频场景中运动目标类型的不同, 可将视觉跟踪问题分为对刚体 (Rigid) 的跟踪和非刚体 (Non-rigid) 的跟踪. 所谓刚体是指具有刚性结构、不易变形的物体, 例如车辆、飞机等等这些人造物体, 这些物体的共同特点是结构比较规范, 能够用 3D 几何模型描述, 因此, 对这类运动目标的跟踪采用基于 3D 模型的跟踪方法比较常见^[9,73,74]. 而非刚体是指外形能够变化的物体, 如细胞、动物、人等. 对这类目标一种通常的方法是采用变形模板 (Deformable template) 进行跟踪^[75,76], 此外, 在对人的跟踪方面, 大量工作尝试建立人的模型进行跟踪^[34,41,77,78], 重要国际期刊专刊 (Computer Vision and Image Understanding, 81(3), 2001, Machine Vision and Application, 14(4), 2003) 对有关人的建模和跟踪问题进行了讨论.

5) 传感器的种类: 可见光图像 VS 红外图像. 根据所使用的传感器种类的不同, 可以将视觉跟踪问题分为由 CCD 摄像头获得的可见光图像的视觉跟踪和由红外传感器获得的红外图像的视觉跟踪. 这两种类型的图像跟踪分别在不同的情况下使用. 通常的自然环境条件下, 白天使用可见光图像进行目标跟踪, 而晚上则使用红外图像进行目标跟踪. 如果将二者所获得的图像进行融合处理 (Image fusion), 则一般可以得到比单一传感器信息更多的图像, 从而极大地提高检测和跟踪运动目标的能力.

3.2 处理视觉跟踪问题的两种思路

对于视觉跟踪问题的处理, 总体上讲有两种思路^[4,79,80], 一种称之为自底向上 (Bottom-

up) 的处理方法; 另一种称之为自顶向下 (Top-down) 的处理方法. 自底向上的处理方法又称之为数据驱动 (Data-driven) 的方法, 这种方法不依赖于先验知识, 直接从图像序列中获得目标的运动信息并进行跟踪. 自顶向下的处理方法又称之为模型驱动 (Model-driven) 的方法, 这种方法一般依赖于所构建的模型或先验知识, 在图像序列中进行匹配运算或求解后验概率, 采用匹配运算时, 如果相似距离最为接近则认为跟踪上运动目标; 求解后验概率时, 选择最大后验概率所对应的状态向量作为运动目标的当前状态. 下边对这两种思路做进一步的介绍.

1) 自底向上. 自底向上实现视觉跟踪的思路以 Divad Marr 的视觉计算理论最具代表性^[81]. 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. 为此, Marr 将视觉过程分为三个阶段: 早期、中期和后期. 早期视觉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; 中期视觉实现由输入图像和要素图获得的 2.5D 图, 即在以观察者为中心的坐标系里可见表面的发祥、深度及不连续轮廓等等; 后期视觉由输入图像、要素图以及 2.5D 图获得的物体得 3D 表示.

与此相似, 在自底向上的视觉跟踪过程中, 跟踪的目的是获得场景中运动目标的位置、速度、加速度、运动轨迹等信息, 这相当于视觉过程中的后期阶段, 为得到这些信息, 在获得图像序列并对图像序列进行预处理之后, 首先直接从图像序列中检测运动目标, 这是早期阶段; 检测到运动目标之后, 提取出运动目标并进行识别, 以判定是否跟踪, 这相当于中期阶段; 最后对该目标进行跟踪并获得运动目标的相关运动信息. 上述过程按处理顺序共分为四步, 第一步为图像预处理, 一般对所获得的序列图像进行消噪或增强, 提高图像质量以方便后续处理; 第二步为目标检测, 在摄像机不动情况下一般采用帧差法^[82]或背景差法^[83]进行变化检测以获取运动目标, 此时得到的运动目标是一个斑点 (Blob); 第三步为目标分类, 要完成两个任务, 首先提取出检测到的运动目标, 其次对该目标进行识别; 最后在目标跟踪阶段获得运动目标的相关运动信息.

这种方法的优点在于能快速检测并跟踪运动目标, 便于工程实现, 目前已有大量实际系统利用该思路解决视觉跟踪问题^[1, 2, 4, 6~12, 14, 15, 20, 70, 72]. 该方法在摄像机静止情况下效果非常好, 但如果摄像机是运动的, 则实现运动目标检测与跟踪就比较困难, 因此在很大程度上局限了该方法的使用范围.

2) 自顶向下. 自顶向下的跟踪思路以 Bar-Shalom 的目标跟踪与关联思想最具代表性^[84]. 这一思路将目标跟踪问题转换为在贝叶斯 (Bayesian) 理论框架下, 已知目标状态的先验概率, 在获得新的量测 (Measurement) 后不断求解目标状态的最大后验概率的过程. 也就是说, 在贝叶斯理论框架下, 将视觉跟踪问题看作是“最优猜测”或者是一种“推理”过程, 通常采用状态空间法 (State space approach) 实现视觉跟踪. 首先确定目标的状态向量 $\{\mathbf{X}_k\}_{k=0,1,2,\dots}$, 其中 k 表示离散时间序列, 状态是跟踪系统的输出, 对应的状态方程如下:

$$\mathbf{X}_{k+1} = F_k \mathbf{X}_k + \mathbf{V}_k \quad (1)$$

其次是目标的观测向量 $\{\mathbf{Z}_k\}_{k=1,2,\dots}$, 同样 k 表示离散时间序列, 观测是指从图像中提取的各种特征, 对应的观测方程如下:

$$\mathbf{Z} = H_k \mathbf{X}_k + \mathbf{W}_k \quad (2)$$

在 (1) 和 (2) 式中, $\{\mathbf{V}_k\}_{k=0,1,2,\dots}$ 和 $\{\mathbf{W}_k\}_{k=1,2,\dots}$ 表示噪声序列, 通常假定它们是独立同分布的 (Independent and identically distributed, 记为 i. i. d.). 在确定了状态和观测之

后, 跟踪的任务是: a) 根据 k 时刻前的观测来预测 $k+1$ 时刻的状态; b) 收到 $k+1$ 时刻的观测后对预测做出修正.

假设在 k 时刻, 有状态的先验概率 $p(\mathbf{X}_k|\mathbf{Z}_{1:k})$, 其中, $\mathbf{Z}_{1:k} = \{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k\}$ 表示一直到当前 k 时刻的所有观测量, 该先验概率对于 $k+1$ 时刻是先验, 而对于 k 时刻是后验. 则跟踪过程由两步组成.

$$\text{预测: } p(\mathbf{X}_{k+1}|\mathbf{Z}_{1:k}) \propto p(\mathbf{X}_{k+1}|\mathbf{X}_k)p(\mathbf{X}_k|\mathbf{Z}_{1:k}) \quad (3)$$

$$\text{更新: } p(\mathbf{X}_{k+1}|\mathbf{Z}_{1:k+1}) \propto p(\mathbf{Z}_{k+1}|\mathbf{X}_{k+1})p(\mathbf{X}_{k+1}|\mathbf{Z}_{1:k}) \quad (4)$$

这样就得到了 $k+1$ 时刻的后验概率 $p(\mathbf{X}_{k+1}|\mathbf{Z}_{1:k+1})$. 其中状态转移概率 $p(\mathbf{X}_{k+1}|\mathbf{X}_k)$ 是建模系统的动态特性. 当系统噪声为高斯分布 (Gaussian) 且 F_k 和 H_k 为线性关系 (Linear) 时, 可以利用卡尔曼滤波 (Kalman filter)^[84, P.56] 求解后验概率, 此时 $p(\mathbf{X}_k|\mathbf{Z}_{1:k})$ 和 $p(\mathbf{X}_{k+1}|\mathbf{Z}_{1:k+1})$ 的分布都必须是高斯分布. 当 F_k 和 H_k 为非线性关系 (Nonlinear) 时, 可以利用扩展卡尔曼滤波 (Extended kalman filter, EKF)^[84, P.106] 求解后验概率, 此时后验概率仍然满足高斯分布. 一种替代 EKF 的方法是 UKF (Unscented Kalman Filter, UKF)^[85, 86]. 如果状态空间是离散的而且是由有限个状态组成的, 则可以使用隐马尔可夫模型 (Hidden Markov Models, HMM)^[87] 进行跟踪.

但是, 经典卡尔曼滤波只能处理线性、高斯、单模态 (Unimodal) 的情况. 实际的视觉跟踪过程中, 后验概率的分布往往是非线性、非高斯、多模态 (Multimodal) 的, 为此, 一种称之为序贯蒙特卡罗的方法 (Sequential monte carlo methods, SMC)^[88, 89] 被提出用于解决这类情况. 该方法在刚提出之时并未受到很大关注, 直到 1996 年, 由 Isard 和 Blake 提出 Condensation 算法之后^[90] 才引起广泛注意. 目前这一类方法被称为粒子滤波 (Particle filter), 是当前视觉跟踪方法中的研究热点^[91~96], 重要国际期刊专刊 (IEEE Transactions on Signal Processing, 50(2), 2002) 对此进行了讨论. UKF 思想也可被粒子滤波方法所借用, 这种情况下粒子滤波被称为 UPF (Unscented particle filter)^[97].

自顶向下思路利用先验知识对跟踪问题建立模型, 然后利用实际图像序列验证模型的正确性, 这种方法具有坚实的数学理论基础, 有很多数学工具可以使用, 因此一直是理论界研究视觉跟踪问题的主流方法. 但是, 先验知识的表述是人工智能 (Artificial intelligence) 中的困难所在, 而且人的很多先验知识也很难用数学形式来表达. 此外, 通过模型匹配来实现跟踪往往比较费时, 很难在实际工程中实现实时跟踪. 自底向上跟踪和自顶向下跟踪有着各自的优点和缺点, 若将自底向上和自顶向下这两种思路结合起来使用, 就有可能克服各自弱点, 实现更为有效的跟踪, 文献 [80] 中对此进行了尝试.

3.3 视觉跟踪算法

从 20 世纪 80 年代起到目前为止这近 25 年里, 出现了众多的视觉跟踪算法. 1988 年, Aggarwal 和 Nanadhakumar 对运动图像分析算法进行了总结^[98], 将算法分为两类, 一类是基于光流法的分析, 另一类是基于特征点的分析. 此后在视觉跟踪领域中, 又出现了许多新的方法, 其中在文献 [7] 和 [44] 中, 将视觉跟踪方法分为四类, 分别是基于区域的跟踪、基于特征的跟踪、基于变形模板的跟踪和基于模型的跟踪, 这种分类方法概括了目前大多数视觉跟踪算法, 因此下面用这种分类方法对视觉跟踪算法进行介绍.

1) 基于区域的跟踪 (Region-based tracking). 基于区域的跟踪算法基本思想是: 首先得到包含目标的模板 (Template), 该模板通过图像分割获得或是预先人为确定, 模板通常为略大于目标的矩形, 也可为不规则形状; 然后在序列图像中, 运用相关算法跟踪目标, 对灰度图像可以采用基于纹理和特征的相关, 对彩色图像还可利用基于颜色的相关. 最常用的相关准则是平方和准则 SSD (the sum of squared differences), 如下式

$$C = \sum (i - j)^2 \quad (5)$$

式中, i 和 j 分别是图像 I 和 J 图像的像素. 该算法还可和多种预测算法结合使用, 如线性预测、二次曲线预测、卡尔曼预测等, 以估计每幅图像中目标的位置. 这种算法的优点在于当目标未被遮挡时, 跟踪精度非常高, 跟踪非常稳定. 但其缺点首先是费时, 当搜索区域较大时情况尤其严重; 其次, 算法要求目标变形不大, 且不能有太大遮挡, 否则相关精度下降会造成目标的丢失. 有关基于区域的跟踪算法可参见文献 [99,100]. 近年来, 对基于区域的跟踪方法关注较多的是如何处理模板变化时的情况, 这种变化是由运动目标姿态变化引起的, 如果能正确预测目标的姿态变化, 则可实现稳定跟踪, 研究这一问题的有关文献可参见 [79,101,102,94,103~106].

2) 基于特征的跟踪 (Feature-based tracking). 基于特征的跟踪算法一般也采用相关算法. 与基于区域的跟踪算法的不同之处在于, 后者使用目标整体作为相关时的对象, 而前者使用目标的某个或某些局部特征作为相关时的对象. 这种算法的优点在于即使目标的某一部分被遮挡, 只要还有一部分特征可以被看到, 就可以完成跟踪任务, 另外, 这种方法与卡尔曼滤波器联合使用, 具有很好的跟踪效果. 这种算法的难点是: 对某个运动目标, 如何确定它的唯一特征集? 这也是一个模式识别问题. 若采用特征过多, 系统效率将降低, 且容易产生错误, 文献 [108,109] 对这一问题进行了讨论. 在特征提取时, 一般采用 Canny 算子获得目标的边缘特征^[110], 而采用 SUSAN 算子获得目标的角点信息^[111]. 有关基于特征的跟踪算法还可参见 [112,113].

3) 基于变形模板的跟踪 (Deformable-template-based tracking). 变形模板是纹理或边缘可以按一定限制条件变形的面板或曲线^[10]. 文献 [114,115] 中, 目标模板为一条手划 (Hand-Drawn) 的封闭曲线, 该曲线通过方向及方向的变形逐渐与图像中的真实目标相适应, 从而检索或跟踪复杂背景中的目标. 在视觉跟踪过程中, 更为常用的变形模板是由 Kass 在 1987 年提出的主动轮廓模型 (Active contour models, ACM)^[116], 又称为 Snake 模型, Kass 用于唇动的跟踪. Snake 模型非常适合可变形目标的跟踪, 如对运动细胞的跟踪^[75]. 这种模型与卡尔曼滤波相结合能够更好地进行跟踪^[117]. Vieren 等提出了一种基于 Snake 模型的多目标跟踪算法^[118], 思路比较新颖. 但 Snake 模型比较适合单目标的跟踪, 对于多目标的跟踪更多地是采用基于水平集 (Level Set) 方法的主动轮廓模型^[119~121], 有关的研究可以参见 [122~125].

4) 基于模型的跟踪 (Model-based tracking). 对人体进行跟踪时, 通常有三种形式的模型^[42], 即线图模型、2D 模型和 3D 模型. 但是在实际的视觉跟踪算法中, 更多的是采用运动目标的 3D 模型进行跟踪, 尤其对于刚体如汽车^[9] 的跟踪. 其基本思想是: 首先由先验知识获得目标的三维结构模型和运动模型, 然后根据实际的图像序列, 确定出目标的三维模型参数, 进而确定出目标的瞬时运动参数. 这种方法的优点是, 可以精确地分析目标的三维运动轨迹, 即使在运动目标姿态变化的情况下, 也能够可靠地跟踪. 但是其缺点在于, 运动分析的精度取决于几何模型的精度, 在现实生活中要获得所有运动目标的精确几何模型是非常困难的. 这就限制了基于模型的跟踪算法的使用, 同时, 基于 3D 模型的跟踪算法往往需要大量的运算时间, 很难实现实时的运动目标跟踪. 有关基于模型的跟踪算法可以参见 [73,74,126,127], 在最近的综述^[128] 中, 作者对基于模型的跟踪算法进行了详细的介绍.

4 视觉跟踪问题中的难点与展望

4.1 视觉跟踪问题中的难点

从控制的观点来看, 视觉跟踪问题所面临的主要难点可以归结为对视觉跟踪算法在三个方面的要求, 即对算法的鲁棒性、准确性和快速性^[43] 要求.

1) 鲁棒性 (Robustness). 所谓鲁棒性是指视觉跟踪算法能够在各种环境条件下实现对运动目标持续稳定的跟踪. 影响视觉跟踪算法鲁棒性的最主要原因在于^[101]:

- a) 被跟踪运动目标的姿态改变;
- b) 运动目标所处环境的光照变化;
- c) 部分遮挡引起的运动目标不规则变形和全部遮挡引起的运动目标的暂时消失.

对运动目标进行三维重构能够有效解决其姿态改变所引起的目标丢失, 基于 3D 模型的跟踪算法可在很大程度上解决这一问题, 但基于 3D 模型的跟踪算法所存在的局限性严重限制了它在实际系统中的应用. 当运动目标所处环境的光照发生改变时, 采用图像灰度信息或色彩信息作为跟踪基础的视觉跟踪算法一般都会失效, 而基于图像特征的方法往往不受光照改变的影响, 如利用运动目标的边缘信息能有效避免光照变化对运动目标的影响, 但在复杂环境中要将运动目标的边缘和周围其它目标边缘区分开来是非常困难的. 遮挡问题是视觉跟踪算法中又一难点问题, 利用单摄像机解决遮挡问题也一直是视觉跟踪领域中的热点^[15,102,129], 而利用多摄像机可以在很大程度上解决这一问题, 但正如前所述, 多摄像机的应用又会引入新的难题. 总之, 对视觉跟踪算法的鲁棒性要求是视觉跟踪问题研究的重点, 目前, 该问题还远远没有得到很好的解决.

2) 准确性 (Accuracy). 在视觉跟踪研究中, 准确性包括两个方面, 一是指对运动目标检测的准确性, 另一个是指对运动目标分割的准确性. 对运动目标检测准确性的目的是尽可能降低运动目标检测的虚警率和漏警率, 从而提高对真实运动目标的检测概率, 这一点在视频监视系统中要求比较高^[20]. 由于实际复杂环境中存在大量噪声, 很难实现各种环境条件下对运动目标地准确检测, 往往只能在虚警率和漏警率之间寻求一个折中 (Trade-off). 对运动目标分割的准确性主要应用于图像压缩编码中, 该问题实际上是图像分割 (Image segmentation) 问题, 至今已经出现了上千种各种类型的分割算法, 但由于尚无通用的分割理论, 目前并没有一种适合于所有图像的通用分割算法^[130]. 在视觉跟踪算法中, 对图像序列中运动目标的分割, 更多的是利用目标的运动信息实现对目标的提取^[131], 如利用光流法^[53,54]实现对图像序列中运动目标的分割. 而主动轮廓模型^[116,120]也是实现对运动目标提取的常用方法. 一旦能够实现对运动目标的准确检测和准确分割, 那么也就能够准确获得运动目标的相关运动信息.

3) 快速性 (Speed). 一个实用的视觉跟踪系统必须能够实现对运动目标的实时跟踪, 这就要求视觉跟踪算法必须具有快速性^[26,65,104,123]. 但是, 视觉跟踪算法处理的对象是包含巨大数据量的图像, 这些算法往往需要大量的运算时间, 很难达到实时处理的要求. 通常, 简单算法能够实现实时跟踪, 但跟踪精度却很差; 复杂算法具有很高的跟踪精度, 实时性却很差. 一种通用的减小视觉跟踪算法运算量的方法是利用金字塔分解或小波变换将图像分层处理; 另一种思路是设计专用硬件实现实时的视觉跟踪.

上述讨论仅仅只是从控制角度简要地给出了视觉跟踪技术中的几个主要难点问题. 视觉跟踪技术所面临的更为一般的难题是: 如何从图像序列中实现有效的图像分割和图像对应. 这里的图像分割是指将被跟踪目标从图像或图像序列中提取出来, 而图像对应是指将被跟踪目标在图像序列中始终能够准确的对应起来, 无论该目标是否变形、被遮挡等等. 图像分割问题目前尚未得到有效解决, 而图像对应问题则是与模式识别紧密相连的人工智能难题. 更多的有关视觉跟踪技术所面临的困难在综述文献 [37,41~44,98,128] 中都有所论述, 其中所涉及的每一个难点都值得进行深入研究.

4.2 展望

实现一个具有鲁棒性、准确性和快速性的视觉跟踪系统是当前视觉跟踪技术努力的方向. 但视觉跟踪技术在这几个方面中每前进一步都是非常困难的, 因为该技术的发展与

人的感知特性的研究紧密联系在一起. 由于目前对人的感知特性没有一个主流的理论, 其数学模型更是难以建立. 同时, 在计算机视觉中大多数问题是不适定的, 这就更增加了视觉跟踪技术发展的难度. 但是, 视觉跟踪技术具有广阔的应用前景, 并不仅仅局限于前边所介绍的几个方面. 随着计算机技术的不断发展, 视觉跟踪技术作为计算机视觉领域中的核心技术之一, 在一些关键技术中具有越来越重要的作用, 例如机器人技术、智能武器系统、虚拟现实技术等. 同时近几十年来, 数学理论方面取得了巨大的进步, 使计算机视觉技术在发展过程中可以很好的利用这些数学工具处理视觉跟踪问题. 因此, 尽管视觉跟踪技术在前进的道路上还存在着许多困难, 但是其所具有的强大的生命力必将使这一技术得到强有力的发展, 并对人们未来的生活产生深远的影响.

5 结束语

视觉跟踪问题是当前计算机视觉领域中的热点问题, 本文对这一问题进行了详细的介绍. 首先, 对该技术应用的主要领域, 即在视频监控、图像压缩和三维重构等三个方面的应用进行了论述. 其次, 详细阐述了该技术的研究现状, 介绍了其中的一些常用方法, 为了清楚地对这些方法进行说明, 先对视觉跟踪问题进行了分类, 然后介绍了处理视觉跟踪问题的两种思路即自底向上和自顶向下的思路, 最后将具体的视觉跟踪方法分为四类进行了介绍, 这四类分别是基于区域的跟踪、基于特征的跟踪、基于变形模板的跟踪和基于模型的跟踪. 最后, 从控制论的角度给出视觉跟踪算法所面临的难点问题, 即算法要满足鲁棒性、准确性和快速性要求时所遇到的困难, 同时还简要的对视觉跟踪问题的研究前景进行了展望.

本文着重选取了近十年来国内外在视觉跟踪技术方面的研究成果, 从上千篇参考文献中选出一百多篇重要参考文献. 由于期刊文献具有更好的理论完整性, 因此, 在文献选择上更多地关注了期刊论文. 通过本文对视觉跟踪技术的介绍, 希望能对相关领域的研究人员和工程技术人员提供有益帮助.

References

- 1 Haritaoglu I, Harwood D, Davis L. W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 809~830
- 2 Pavlidis I, Morellas V, Tsiamyrtzis P, Harp S. Urban surveillance system: From the laboratory to the commercial world. *Proceedings of the IEEE*, 2001, **89**(10): 1478~1497
- 3 Collins R, Lipton A, Kanade T, Fujiyoshi H, Duggins D, Tsin Y, Tolliver D, Enomoto N, Hasegawa O, Burt P, Wixson L. A system for video surveillance and monitoring, VSAM final report, Carnegie Mellon University: Technical Report CMU-RI-TR-00-12, 2000
- 4 Collins R, Lipton A, Fujiyoshi H, Kanade T. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE*, 2001, **89**(10): 1456~1477
- 5 Bogaert M, Chleq N, Cornez P, Regazzoni C S, Teschioni A, Thonnat M. The PASSWORDS project. In: *Proceedings of IEEE International Conference on Image Processing*. Switzerland: Lansanne, 1996, **3**: 675~678
- 6 Wren C R, Azarbayejani A, Darrell T, Pentland A P. Pfnder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 780~785
- 7 Coifman B, Beymer D, McLauchlan P, Malik J. A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C*, 1998, **6**(4): 271~288
- 8 Magee D. Tracking multiple vehicles using foreground, background and motion models. *Image and Vision Computing*, 2004, **22**(2): 143~155
- 9 Koller D, Daniilidis K, Nagel H. Model-based object tracking in monocular image sequences of road traffic scene. *International Journal of Computer Vision*, 1993, **10**(3): 257~281
- 10 Dubuisson-Jolly M, Lakshmanan S, Jain A K. Vehicle segmentation and classification using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(3): 293~308

- 11 Tai J, Tsang S, Lin C, Song K. Real-time image tracking for automatic traffic monitoring and enforcement application. *Image and Vision Computing*, 2004, **22**(6): 485~501
- 12 Zhu Z, Xu G, Yang B, Shi D, Lin X. VISATRAM: A real-time vision system for automatic traffic monitoring. *Image and Vision Computing*, 2000, **18**(10): 781~794
- 13 Haag M, Nagel H. Tracking of complex driving maneuvers in traffic image sequences. *Image and Vision Computing*, 1998, **16**(8): 517~527
- 14 Pai C, Tyan H, Liang Y, Liao H M, Chen S. Pedestrian detection and tracking at crossroads. *Pattern Recognition*, 2004, **37**(5): 1025~1034
- 15 Masoud O, Papanikolopoulos N P. A novel method for tracking and counting pedestrians in real-time using a single camera. *IEEE Transactions on Vehicular Technology*, 2001, **50**(5): 1267~1278
- 16 Foresti G L, Murino V, Regazzoni C. Vehicle recognition and tracking from road image sequences. *IEEE Transactions on Vehicular Technology*, 1999, **48**(1): 301~318
- 17 Betke M, Haritaoglu E, Davis L S. Real-time multiple vehicle detection and tracking from a moving vehicle. *Machine Vision and Applications*, 2000, **12**(2): 69~83
- 18 Ferryman J M, Maybank S J, Worrall A D. Visual surveillance for moving vehicle. *International Journal of Computer vision*, 2000, **37**(2): 187~197
- 19 Desouza G. N, Kak A. C. Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(2): 237~267
- 20 Boulton T E, Micheals R J, Gao X, Eckmann M. Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. *Proceedings of the IEEE*, 2001, **89**(10): 1382~1402
- 21 Kumar R, Sawhney H, Samarasekera S, Hsu S, Tao H, Guo Y, Hanna K, Pope A, Wildes R, Hirvonen D, Hansen M, Burt P. Aerial video surveillance and exploitation. *Proceedings of the IEEE*, 2001, **89**(10): 1518~1539
- 22 Tao H, Sawhney H S, Kumar R. Object tracking with Bayesian estimation of dynamic layer representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(1): 75~89
- 23 Tsao T, Wen Z. Image-based target tracking through rapid sensor orientation change, *SPIE Optical Engineering*, 2002, **41**(3): 697~703
- 24 Sikora T. The MPEG-4 video standard verification model. *IEEE Transactions on Circuits and Systems for Video Technology*, 1997, **7**(1): 19~31
- 25 Castagno R, Ebrahimi T, Kunt M. Video segmentation based on multiple features for interactive multimedia applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998, **8**(5): 562~571
- 26 Kim B, Park R. A fast automation VOP generation using boundary block segmentation. *Real-time Imaging*, 2004, **10**(2): 117~125
- 27 Valente S, Dugelay J. A visual analysis/synthesis feedback loop for accurate face tracking. *Signal Processing: Image Communication*, 2001, **16**(6): 585~608
- 28 Meier T, Ngan K N. Automatic segmentation of moving objects for video object plane generation. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998, **8**(5): 525~538
- 29 Wang D. Unsupervised video segmentation based on watersheds and temporal tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998, **8**(5): 539~546
- 30 Mech R, Wollborn M. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera. *Signal Processing*, 1998, **66**(2): 203~217
- 31 Zhu Z J, Jiang G Y, Yu M, Wang R D, Wu X W. New algorithm for extracting and tracking moving object in object-based video coding. *Acta Electronica Sinica*, 2003, **31**(9): 1426~1428
- 32 Cohen L D, Cohen I. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(11): 1131~1147
- 33 Jaynes C, Riseman E, Hanson A. Recognition and reconstruction of buildings from multiple aerial images. *Computer Vision and Image Understanding*, 2003, **90**(1): 68~98
- 34 Remondino F. 3-D reconstruction of static human body shape from image sequence. *Computer Vision and Image Understanding*, 2004, **93**(1): 65~85
- 35 Park S, Subbarao M. Automatic 3D model reconstruction based on novel pose estimation and integration techniques. *Image and Vision Computing*, 2004, **22**(8): 623~635
- 36 Zheng J Y. Acquiring a complete 3D model from specular motion under the illumination of circular-shape light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 913~920

- 37 Pavlovic V I, Sharma R, Huang T S. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 677~695
- 38 Lien C, Huang C. Model-based articulated hand motion tracking for gesture recognition. *Image and Vision Computing*, 1998, **16**(2): 121~134
- 39 Choi H I, Rhee P K. Head gesture recognition using HMMs. *Expert System with Applications*, 1999, **17**(3): 213~221
- 40 Viola P, Jones M. J. Robust real-time face detection. *International Journal of Computer Vision*, 2004, **57**(2): 137~154
- 41 Gavrilu D M. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 1999, **73**(1): 82~98
- 42 Aggarwal J K, Cai Q. Human motion analysis: A review. *Computer Vision and Image Understanding*, 1999, **73**(3): 428~440
- 43 Moeslund T B, Granum E. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 2001, **81**(3): 231~268
- 44 Wang L, Hu W M, Tan T N. A survey of visual analysis of human motion. *Chinese Journal of Computers*, 2002, **25**(3): 225~237
- 45 Tieu K, Viola P. Boosting image retrieval. *International Journal of Computer Vision*, 2004, **56**(1/2/3): 17~36
- 46 Fablet R, Bouthemy P. Non-parametric motion activity analysis for statistical retrieval with partial query. *Journal of Mathematical Imaging and Vision*, 2001, **14**(3): 257~270
- 47 Chen Y, Wang J Z. A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(9): 1252~1267
- 48 Stricker D, Kettenbach T. Real-time and markerless vision-based tracking for outdoor augmented reality application. *IEEE and ACM International Symposium on Augmented Reality*, 2001, 189~190
- 49 You S, Neumann U, Azuma R. Hybrid inertial and vision tracking for augmented reality registration. *Proceedings of IEEE Virtual Reality*, 1999, 260~267
- 50 Kanbara M, Yokoya N, Takemura H. A stereo vision-based augmented reality system with marker and natural feature tracking. In: *Proceedings of IEEE the 7th International Conference Virtual System and Multimedia, VSM'01*. Washington, DC, USA: IEEE Press, 2001. 455~462
- 51 Rosenfeld A. From image analysis to computer vision: An annotated bibliography. *Computer Vision and Image Understanding*, 2001, **84**(2): 298~324
- 52 Ballard D H, Brown C M. *Computer Vision*, Prentice-Hall, Inc, 1982
- 53 Horn B K, Schunck B G. Determine Optical Flow. *Artificial Intelligence*, 1981, **17**(1-3): 185~204
- 54 Lucas B D, Kanade T. An iterative image-registration technique with an application to stereo vision. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. Vancouver, BC, Canada: William Kaufmann, 1981, 674~679
- 55 Beauchemin S S, Barron J L. The computation of optical flow. *ACM Computing Surveys*, 1995, **27**(3): 433~467
- 56 Baker S, Matthews I. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 2004, **56**(3): 221~255
- 57 Matsuyama T, Ukita N. Real-time multitarget tracking by a cooperative distributed vision system. *Proceedings of the IEEE*, 2002, **90**(7): 1136~1150
- 58 Wachter S, Nagel H. Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, 1999, **74**(3): 174~192
- 59 Zhang W J, Qi F H, Jiang Z J. A practiceal method for detecting and tracking objects in image sequences in a real-time video surveillance and monitoring system. *Chinese Journal of Shanghai Jiaotong University*, 2002, **36**(12): 1837~1840
- 60 Ma S D, Zhang Z Y. *Computer Vision: Calculation Theory and Algorithm Foundation*, Beijing: Science Press, 1998
- 61 Dockstader S L, Tekalp A M. Multiple camera tracking of interacting and occluded human motion. *Proceedings of the IEEE*, 2001, **89**(10): 1441~1455
- 62 Ye Y, Tsotsos J K, Harley E, Bennet K. Tracking a person with pre-recorded image database and a pan, tilt, and zoom camera. *Machine Vision and Application*, 2000, **12**(1): 32~43
- 63 Smith S M, Brady J M. ASSET-2: Real-time motion segmentation and shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(8): 814~820

- 64 Pless R, Brodsky T, Aloimonos Y. Detecting independent motion: The statistics of temporal continuity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 768~773
- 65 Hsieh J. Fast stitching algorithm for moving object detection and mosaic construction. *Image and Vision Computing*, 2004, **22**(4): 291~306
- 66 Hsu C, Tsan Y. Mosaics of video sequences with moving objects. *Signal Processing: Image Communication*, 2004, **19**(1): 81~98
- 67 Wang D S, Li Z M. The technology of real-time detection and tracking for spatial video motion object. *Chinese Journal of Signal Processing*, 2005, **21**(2): 195~198
- 68 Yang Y, Zhang T. Moving target tracking based on feature-optical-flow. *Chinese Journal of Astronautics*, 2000, **21**(2): 8~15
- 69 Soares L D, Pereira F. Spatial shape error concealment for object-based image and video coding. *IEEE Transactions on Image Processing*, 2004, **13**(4): 586~599
- 70 Zhao T, Nevatia R. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9): 1208~1221
- 71 Yu T, Wu Y. Collaborative tracking of multiple targets. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'04*, Washington, DC, USA: IEEE Press, 2004. **1**: 834~841
- 72 McKenna S, Jabri S, Duric Z, Rosenfeld A, Wechsler H. Tracking groups of people. *Computer Vision and Image Understanding*, 2000, **80**(1): 42~56
- 73 Jurie F. Tracking objects with a recognition algorithm. *Pattern Recognition Letters*, 1998, **19**(3-4): 331~340
- 74 Drummond T, Cipolla R. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 932~946
- 75 Leymarie F, Levine M D. Tracking deformable objects in the plane using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(6): 617~624
- 76 Jain A K, Zhong Y, Dubuisson-Jolly M. Deformable template models: A review. *Signal Processing*, 1998, **71**(2): 109~129
- 77 Delamarre Q, Faugeras O. 3D articulated models and multiview tracking with physical forces. *Computer Vision and Image Understanding*, 2001, **81**(3): 328~357
- 78 Plankers R, Fua P. Tracking and modeling people in video sequences. *Computer Vision and Image Understanding*, 2001, **81**(3): 285~302
- 79 Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564~577
- 80 Oliver N M, Rosario B, Pentland A P. A Bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 831~843
- 81 Marr D. *Vision: a computational investigation into the human representation and processing of visual information*, Freeman W.H. and Company, San Francisco, 1982
- 82 Wang S, Ai H. Z, He K. Z. Difference-image-based multiple motion targets detection and tracking. *Chinese Journal of Image and Graphics*, 1999, **4A**(6): 470~475
- 83 Hou Z Q, Han C Z. A background reconstruction algorithm based on pixel intensity classification. *Chinese Journal of Software*, 2005, **16**(9): 1568~1576
- 84 Bar-Shalom Y, Fortmann T. *Tracking and Data Association*, Academic Press, 1988
- 85 Julier S, Fleet D, El-Maraghi T. A new extension of the Kalman filter to nonlinear system. *Proceedings of SPIE*, 1997, **3068**: 182~193
- 86 Li P, Zhang T, Ma B. Unscented Kalman filter for visual curve tracking. *Image and Vision Computing*, 2004, **22**(2): 157~164
- 87 Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989, **77**(2): 257~285
- 88 Salmond D, Gordon N, Smith A. A novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings on Radar, Sonar and Navigation*, 1993, **140**(2): 107~113
- 89 Arulampalam M, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 2002, **50**(2): 174~188
- 90 Isard M, Blake A. CONDENSATION-Conditional density propagation for visual tracking. *International Journal of Computer Vision*, 1998, **29**(1): 5~28
- 91 Chen Y, Rui Y. Real-time speaker tracking using particle filter sensor fusion. *Proceedings of the IEEE*, 2004, **92**(3): 485~494

- 92 Wu Y, Huang T. Robust visual tracking by integrating multiple cues based on co-inference learning. *International Journal of Computer Vision*, 2004, **58**(1): 55~71
- 93 Perez P, Vermaak J, Blake A. Data fusion for visual tracking with particles. *Proceedings of the IEEE*, 2004, **92**(3): 495~513
- 94 Zhou S K, Chellappa R, Moghaddan B. Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Transactions on Image Processing*, 2004, **13**(11): 1491~1506
- 95 Hue C, Cader J, Perez P. Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 2002, **38**(3): 791~812
- 96 Spengler M, Schiele B. Towards robust multi-cue integration for visual tracking. *Machine Vision and Applications*, 2003, **14**(1): 50~58
- 97 Merwe R, Doucet A, Freitas N, Wan E. The unscented particle filter, technical report CUED/F-INFENG/TR380, Engineering Department, Cambridge University Aug. 2000
- 98 Aggarwal J K, Nandhakumar N. On the computation of motion from sequences of images - A review. *Proceedings of the IEEE*, 1988, **76**(8): 917~935
- 99 Meyer F, Bouthemy P. Region-based tracking using affine motion models in long image sequences. *CVGIP: Image understanding*, 1994, **60**(2): 119~140
- 100 Bascle B, Deriche R. Region tracking through image sequences. In: *Proceedings of IEEE International Conference of Computer Vision*, 1995, 302~307
- 101 Hager G, Belhumeur P. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(10): 1025~1039
- 102 Jepson A, Fleet D, El-Maraghi T. Robust online appearance models for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(10): 1296~1311
- 103 Matthews I, Ishikawa T, Baker S. The template update problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(6): 810~815
- 104 Nguyen H, Smeulders A. Fast occluded object tracking by a robust appearance filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8): 1099~1104
- 105 Avidan S. Support vector tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8): 1064~1072
- 106 Blac M, Jepson A. Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation. In: *Proceedings of European Conference of Computer Vision*. Cambridge, UK: Springer Verlag, 1996, 329~342
- 107 Krishnan S, Raviv D. 2D feature tracking algorithm for motion analysis. *Pattern Recognition*, 1995, **28**(8): 1103~1126
- 108 Kaneko T, Hori O. Feature selection for reliable tracking using template matching. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'03*. Madison, Wisconsin, USA: IEEE Press, 2003. **1**: 796~802
- 109 Mitra P, Murthy C, Pal S. Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(3): 301~312
- 110 Canny J F. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, **8**(6): 679~698
- 111 Smith S, Brady J. SUSUN - A new approach to low level image processing. *International Journal of Computer Vision*, 1997, **23**(1): 45~78
- 112 Tissainayagam P, Suter D. Object tracking in image sequences using point feature. *Pattern Recognition*, 2005, **38**(1): 105~113
- 113 Nickels K, Hutchinson S. Estimating uncertainty in SSD-based feature tracking. *Image and Vision Computing*, 2002, **20**(1): 47~58
- 114 Jain A K, Zhong Y. Lakshmanan, Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(3): 267~278
- 115 Yu Z, Jain A K, Dubuisson-Jolly M. Object tracking using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(5): 544~549
- 116 Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *International Journal of Computer Vision*, 1988, **1**(4): 321~331
- 117 Terzopoulos D, Szeliski R. Tracking with Kalman Snakes, in *Active Vision*, Blake A. and Yuille A, Eds. Cambridge, MA: MIT Press, 1992. 3~20

- 118 Vieren C, Cabestaing F, Postaire J. Catching moving objects with snakes for motion tracking. *Pattern Recognition Letters*, 1995, **16**(7): 679~685
- 119 Osher S, Sethian J A. Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulation. *Journal of Computational Physics*, 1988, **79**(1): 12~49
- 120 Malladi R, Sethian J A, Vemuri B C. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(2): 158~179
- 121 Caselles V, Kimmel R, Sapiro G. Geodesic active contours. *International Journal of Computer Vision*, 1997, **22**(1): 61~79
- 122 Paragios N, Deriche R. Geodesic active contours and level set for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(3): 266~280
- 123 Dgoldenberg R, Kimmel R, Rivlin E, Rudzsky M. Fast geodesic active contours. *IEEE Transactions on Image Processing*, 2001, **10**(10): 1467~1475
- 124 Mansouri A. Region tracking via level set PDEs without motion computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 947~961
- 125 Freedman D, Zhang T. Active Contours for tracking distributions. *IEEE Transactions on Image Processing*, 2004, **13**(4): 518~526
- 126 Gerard P, Gagalowicz A. Three dimensional model-based tracking using texture learning and matching. *Pattern Recognition Letters*, 2000, **21**(13-14): 1095~1102
- 127 Jung S K, Wohn K Y. A model-based 3-D tracking of rigid objects from a sequence of multiple perspective views. *Pattern Recognition Letters*, 1998, **19**(5-6): 499~512
- 128 Hu W M, Tan T N, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 2004, **34**(3): 334~352
- 129 Gentile C, Camps O, Sznaiar M. Segmentation for robust tracking in the presence of severe occlusion. *IEEE Transactions on Image Processing*, 2004, **13**(2): 166~178
- 130 Zhang Y J. Image Engineering: Image Processing and Analysis, Beijing: Tsinghua University Press, 1999
- 131 Gao W, Chen X L. Computer Vision: Algorithm and System Theory. Beijing: Tsinghua University Press, 1999

侯志强 2005年毕业于西安交通大学,获博士学位,目前为空军工程大学电讯工程学院副教授.研究方向为图像处理,视觉跟踪和信息融合.

(**HOU Zhi-Qiang** Received his Ph.D. degree from Xi'an Jiaotong University in 2005. He is currently an associate professor, his research interests include image processing, visual tracking, and information fusion.)

韩崇昭 西安交通大学电信学院教授,博士生导师.研究领域是随机控制与自适应控制、工业过程控制与优化、非线性频谱分析,决策理论与决策支持系统等,近年来主要从事信息融合理论研究.

(**HAN Chong-Zhao** Professor in School of Electronic and Information Engineering at Xi'an Jiaotong University. His research interests include stochastic control and adaptive control, industrial process control and optimization, nonlinear spectral analysis, decision support theory and systems, and information fusion.)