

国内图书分类号: TP391.41
国际图书分类号: 681.39

工学博士学位论文

基于序列蒙特卡洛滤波算法的
视觉目标跟踪

博士研究生: 王建宇
指导教师: 高文教授
申请学位: 工学博士
学科专业: 计算机应用技术
所在单位: 计算机科学与技术学院
答辩日期: 2006年2月
授予学位单位: 哈尔滨工业大学

国内图书分类号: TP391.41

国际图书分类号: 681.39

工学博士学位论文

基于序列蒙特卡洛滤波算法的
视觉目标跟踪

博士 研究生: 王建宇
导 师: 高文教授
申 请 学 位: 工学博士
学 科、专 业: 计算机应用技术
所 在 单 位: 计算机科学与技术学院
答 辩 日 期: 2006 年 2 月
授 予 学 位 单 位: 哈尔滨工业大学

摘要

基于图像序列的目标跟踪作为计算机视觉领域的一个核心问题,得到了广泛而深入的研究。视觉跟踪研究的主要目的是模仿生理视觉系统的运动感知功能,赋予机器辨识图像序列中物体运动及其相互关系的能力,为图像序列理解提供重要途径。视觉跟踪技术具有广阔的应用前景,如视频监控、视频分析、视频检索、基于视频的运动分析和合成、基于运动信息的身份识别等。经过四十多年尤其是近十年的不懈研究,视觉跟踪技术取得了长足的进步,但实践表明一般意义上的视觉跟踪技术还远未成熟,要开发出真正鲁棒、实用的视觉跟踪应用系统还需要更为鲁棒的核心算法并需要解决大量的算法实现问题。

本文在序列蒙特卡洛滤波算法的框架下,以人脸和人体跟踪为研究对象,针对其中涉及的关键问题进行了探讨,研究了开发鲁棒实用的视觉跟踪系统所需要的核心技术和关键问题解决方案,重点探讨了目标表现建模,复杂运动的建模和推断,融合低端模型和高端模型的运动描述方法等几个关键问题。具体的研究内容如下:

- 1) 提出了可区分性目标表现模型的自适应建模和更新算法。表现建模是视觉跟踪算法性能的决定性因素之一。实践表明:图像特征选择和基于图像特征的目标表现描述模型从根本上决定了算法的鲁棒性和计算复杂性。虽然这一问题得到了领域内学者的极大重视和不懈努力,其仍是阻碍视觉跟踪技术进入实际应用的最困难问题之一。本论文中提出了一种自适应目标表现建模和更新算法。该算法在动态建模过程中不仅考虑目标表现信息,同时对目标所处环境中的背景信息进行考察,从而可对目标/背景的差异信息进行有效建模,从根本上保证了模型具有从变化的背景中区分前景的能力。实验结果表明,相比于目前最具代表性的跟踪算法之一 Mean Shift,提出的算法在公开的测试序列上取得了更好的跟踪结果。
- 2) 提出了集成多运动模型的复杂运动建模和推断算法。由于计算复杂性的限制,视觉跟踪算法通常基于局部搜索的策略确定目标的运动状态。所以,根据目标运动规律确定其以较高概率出现的局部区域成为算法效率的关键因素之一。如何针对复杂运动描述目标运动规律,是很多现实跟踪问题的效率瓶颈所在。本论文提出了采用多运动模型对目标复杂运动

进行建模和估计的基本框架。在此基础上,针对具有多种运动模式和具有高维运动状态的两类常见的复杂运动模式,将多模型的估计框架融入序列蒙特卡罗滤波算法中,从而针对两类复杂运动问题提出了标准序列蒙特卡罗滤波算法的两个改进:基于多模型切换和基于多模型协同的序列蒙特卡罗滤波算法。在人脸跟踪和面部表情估计问题上分别验证了改进的算法。实验结果表明,相对于标准序列蒙特卡罗滤波算法,在计算复杂度降低的同时,改进的算法得到了更高的跟踪精度。

- 3) 提出了融合光流和特定模型的面部特征点跟踪算法。面部特征点跟踪是基于特征点的运动感知研究的典型应用,也是基于特征点运动感知任务中的困难问题。现有的面部特征点跟踪方法主要可以分为基于特定描述模型(以下简称模型)和基于光流的方法。本论文在序列蒙特卡罗滤波算法的框架下融合了基于光流和基于特定模型的方法来解决面部特征点跟踪问题,以克服单独采用一类方法的不足,从而达到鲁棒跟踪面部特征点的目的。在基于尺度空间理论改进 KLT 光流算法的基础上,以光流估计结果约束基于模型的可变特征点估计的起始搜索位置,大大加速了序列蒙特卡罗滤波算法的搜索过程。对于估计结果中存在的跟踪误差,进一步采用特征点运动轨迹的空间约束迭代求精跟踪结果。相比于广泛使用的 KLT 特征点跟踪算法,实验结果证实了提出算法的有效性。

本论文的三个主要创新点,分别对应视觉跟踪中的两个关键问题:目标外观的建模和目标运动的描述。其中创新点一提出了动态建模前景/背景差异的理念,使其不同于已有的大部分视觉跟踪算法。创新点二和三点分别从显式地采用特定模型描述目标运动和隐式地采用离散特征点描述目标运动方面进行了创新尝试。三种方法互为补充,并适合不同的应用情境。

关键词 视觉目标跟踪; 序列蒙特卡罗滤波算法; 在线特征选择; 运动建模; 基于光流的跟踪

Abstract

Image sequence based object tracking is a fundamental problem for computer vision research and has been widely studied. The main goal of visual tracking is to imitate the motion sensibility of physical visual system, empower the machine with the ability of perceiving the object motion and their relations in the scene and provide an important way for image sequence understanding. Visual tracking technique has many applications, such as video surveillance, video analysis, video indexing, video based motion analysis and synthesis, motion-based human identification. After more than 40 years' development, visual tracking technique has made great progress especially in the past ten years. However, practical experience has shown that visual tracking technologies are currently far from mature. A great number of challenges need to be solved before one can implement a robust visual tracking system for commercial applications.

Under the framework of sequential Monte Carlo filtering algorithm, this thesis try to get insights on some key issues in visual tracking with application scenarios on face and human tracking. Some important technologies and solutions are studied which are necessary for robust and practical tracking systems, especially concentrate on how to model the object appearance variation, how to model and estimate the object complex dynamics, how to combine low level and high level motion estimation methods to enhance tracker robust and efficiency. The main contributions of this thesis can be concluded as follows.

- 1) An algorithm for online modeling and adapting discriminative object appearance model is proposed. How to model the appearance of the object is one of the key factors determining the performance of a visual tracking system. Practical experience shows that feature selection strategy and how to model features fundamentally determine the robustness and computational complexity of a tracker. This problem has been extensively studied by many researchers. However, it is still one of the biggest difficulties to prevent the tracking technique into practical applications. An adaptive appearance model with online updating process is proposed. The algorithm considers both object appearance and its relevant background when constructing object

model. The constructed model encodes the difference between the object and background dynamically. Therefore, during the tracking process, the discriminability of the updated model is guaranteed basically. Compared with one of the state-of-the-art tracking algorithms, Mean Shift, experimental results show that our algorithm performs better on publicly available test sequences.

- 2) An algorithm for modeling and estimating complex object dynamics by integrating multiple models is proposed. Due to the limitation of computational resource, tracking algorithms are almost based on local search methods to find object motion state. Therefore, predicting the object future positions according to its motion trajectory is one of the key factors to determine the algorithm efficacy. How to model these complex dynamics is the bottleneck of these kinds of tracking tasks. This thesis proposes a framework to model and estimate complex motion by incorporating multiple motion models. Based on the proposed framework and aim to solving two kinds of complex motions, two new variations of sequential Monte Carlo filter are proposed, termed as multi-model switching sequential Monte Carlo filter and multi-model cooperation sequential Monte Carlo filter respectively, by combining the proposed framework with the standard sequential Monte Carlo filter. Experimental results show that the proposed algorithms perform better than standard sequential Monte Carlo filter and simultaneously lower the computational burden.
- 3) A facial feature tracking algorithm by combining optical flow and specific description model is proposed. Facial feature tracking is one of the classical applications of local feature based motion perception. Tracking facial feature is also a challenging problem. Existing facial feature tracking algorithms can be categorized into two kinds: specific description model based and optical flow based methods. The thesis proposes to combine the optical flow based and specific model based methods under the sequential Monte Carlo filter to solve the facial feature tracking problem. The classical KLT feature tracker is improved with the scale space theory. Based on the fine initial conditions constrained by improved KLT tracker, mouth description model is employ for those deformable features and the searching process of sequential Monte

Carlo filtering algorithm can be accelerated significantly. Considering remained tracking errors, subspace constraint on motion trajectories of all features is furtherly adopted to iteratively refine tracking results. Compared with original KLT tracker, experimental results confirm the effectiveness of the proposed method on facial feature tracking task.

Three proposed novel ideas in the thesis are try to solve two basic problems in visual tracking research: object appearance modeling and object dynamics description. The first idea proposes to dynamically modeling the difference between foreground/background, which make it different from most existing visual tracking algorithms. The second and third ideas try to describe object dynamic explicitly from the motion model or implicitly from discrete feature points' motions respectively. Therefore, the three proposed algorithms are somewhat complementary and can be chosen for different application scenarios.

Keywords Visual object tracking, sequential Monte Carlo filtering algorithm, online feature selection, motion modeling, optical flow based tracking

目录

摘要	I
Abstract	III
第1章 绪论	1
1.1 研究背景	1
1.2 跟踪研究中面临的主要问题	2
1.3 视觉目标跟踪研究概述	5
1.3.1 目标表现的建模和提取	5
1.3.2 数据关联技术	13
1.3.3 滤波框架	14
1.4 本工作的主要技术路线和目标	15
1.5 本论文的主要贡献	16
1.6 本文组织及各章间关系	18
第2章 序列蒙特卡洛滤波算法	20
2.1 引言	20
2.2 跟踪问题的定义和贝叶斯时序滤波框架	20
2.3 序列蒙特卡洛滤波算法	23
2.3.1 核心思想	23
2.3.2 算法的退化问题	26
2.3.3 选择合适的提议分布	26
2.3.4 粒子重采样技术	28
2.3.5 弥补非最优提议分布缺陷的技术	30
2.4 小结	30
第3章 可区分性目标模型的动态构建	31
3.1 引言	31
3.2 问题的提出和相关工作	31
3.3 方法概述	33
3.4 基于目标/背景差异信息的特征选择	34
3.4.1 特征集合	34
3.4.2 目标建模	35

3.5 跟踪过程中目标模型的动态更新	37
3.5.1 维护目标/背景差异性的模型更新	38
3.5.2 维护目标描述一致性的更新	39
3.6 采用动态建模的目标跟踪	40
3.7 将模型更新融入序列蒙特卡洛滤波算法	41
3.7.1 背景粒子的存在	41
3.7.2 融入自适应目标模型更新的跟踪算法	42
3.8 实验结果	44
3.8.1 人体跟踪实验	44
3.8.2 汽车跟踪实验	55
3.9 小结	59
第4章 基于多运动模型的复杂运动建模和推断	60
4.1 引言	60
4.2 问题的提出和相关工作	60
4.2.1 具有多种运动模式的复杂运动	60
4.2.2 具有高维状态空间的复杂运动	61
4.3 多模型运动估计框架	62
4.3.1 定义	63
4.3.2 模型交互过程	64
4.4 多模型切换序列蒙特卡洛滤波算法	65
4.5 实验部分	66
4.5.1 基于可控视频序列的算法验证	66
4.5.2 基于公共测试序列的算法验证	71
4.6 基于多模型协同的序列蒙特卡洛滤波算法	75
4.7 头部运动估计	76
4.7.1 概述	76
4.7.2 头部运动的表示	77
4.7.3 运动模型	80
4.7.4 评估粒子权重	81
4.7.5 算法的定量性能评测	81
4.7.6 算法性能的定性分析	85
4.8 小结	87

第5章 融合光流和模型的面部特征点跟踪算法	88
5.1 引言	88
5.2 算法的理论层面分析	89
5.3 结合光流和模型的面部特征点跟踪	90
5.3.1 KLT光流跟踪算法	90
5.3.2 基于尺度空间理论的特征点尺度自动选择	92
5.3.3 多尺度Harris特征点选择算法	92
5.3.4 尺度空间理论	95
5.3.5 嘴部特征点跟踪	97
5.4 采用子空间约束的跟踪结果求精过程	101
5.5 实验部分	105
5.5.1 对KLT增强算法的实验验证	105
5.5.2 子空间约束特征点跟踪算法的实验验证	108
5.6 小结	115
结论	116
参考文献	118
攻读学位期间发表的学术论文	127
哈尔滨工业大学博士学位论文原创性声明	129
哈尔滨工业大学博士学位论文授权使用书	129
致谢	130
个人简历	132

Contents

Chinese Abstract	I
Abstract	III
Chapter 1 Introduction	1
1.1 Research Background	1
1.2 Main Challenges	2
1.3 Previous Works	5
1.3.1 Object Representation and Measurement Extraction	5
1.3.2 Data Association Techniques	13
1.3.3 Filtering Framework	14
1.4 Main Purpose and Strategy	15
1.5 Main Contributions	16
1.6 Organization of the Dissertation	18
Chapter 2 Sequential Monte Carlo Filter	20
2.1 Introduction	20
2.2 Definition of Visual Tracking and Bayesian Temporal Filter	20
2.3 Sequential Monte Carlo Filtering algorithm	23
2.3.1 Basic Idea	23
2.3.2 Sampling Degeneracy Problem	26
2.3.3 Choosing Appropriate Proposal Distribution	26
2.3.4 Re-sampling Techniques	28
2.3.5 Compensating Non-Optimal Proposal Distribution	30
2.4 Conclusion	30
Chapter 3 Dynamic object modeling for Visual Tracking	31
3.1 Introduction	31
3.2 Problems and Related Works	31
3.3 Method Outline	33
3.4 Selecting Features Based on the Difference of Object/Background	34
3.4.1 The Feature Set	34
3.4.2 Object Appearance Modeling	35

3.5 Updating the Appearance Model during Tracking.....	37
3.5.1 Updating Process to Keep Model Discriminative.....	38
3.5.2 Updating Process to Keep Up Model with Object Variations.....	39
3.6 Object Tracking with Online Model Updating.....	40
3.7 Embedding Model Adapting Process into Sequential Monte Carlo Filter..	41
3.7.1 Existing of "Background" Particles.....	41
3.7.2 Sequential Monte Carlo Filter with Online Model Updating.....	42
3.8 Experimental Results.....	44
3.8.1 People Tracking.....	44
3.8.2 Car Tracking.....	55
3.9 Conclusion.....	59
Chapter 4 Multi-model Based motion modeling and estimation	60
4.1 Introduction.....	60
4.2 Problems and Related Works.....	60
4.2.1 Complex Dynamics Consists of Multiple Motion Modes.....	60
4.2.2 Complex Dynamics with High Dimensional States.....	61
4.3 Framework of Multi-Model Motion Estimation.....	62
4.3.1 Notations.....	63
4.3.2 Model Interacting Process.....	64
4.4 Multi-Model Switching Sequential Monte Carlo Filtering algorithm.....	65
4.5 Experiments.....	66
4.5.1 Experiments on Self-Recorded Video Sequences.....	66
4.5.2 Experiments on Public Available Video Sequences.....	71
4.6 Multi-Model Cooperating Sequential Monte Carlo Filtering algorithm...	75
4.7 Head Motion Estimation.....	76
4.7.1 Introduction.....	76
4.7.2 Head Motion Representation.....	77
4.7.3 Motion Models.....	80
4.7.4 Evaluating Particle Weights.....	81
4.7.5 Quantitative Evaluation of Proposed Algorithm.....	81
4.7.6 Qualitative Evaluation of Proposed Algorithm.....	85
4.8 Conclusion.....	87

Chapter 5 Feature Tracking by Combining Optical Flow and Specific Motion Model	88
5.1 Introduction.....	88
5.2 Theoretical Analysis of Proposed Algorithm.....	89
5.3 Facial Feature Tracking Combining Optical Flow and Model.....	90
5.3.1 KLT Optical Flow Tracker.....	90
5.3.2 Selecting Interest Point Scale by Scale Space Theory.....	92
5.3.3 Multi-Scale Harris Interest Point Detector.....	92
5.3.4 Scale Space Theory.....	95
5.3.5 Tracking Mouth Features.....	97
5.4 Refining Tracking Results using Subspace Constraints.....	101
5.5 Experiments.....	105
5.5.1 Evaluating Sacle Space Theory Enhanced KLT Tracker.....	105
5.5.2 Evaluating Proposed Tracker with Subspace Constraints.....	108
5.6 Conclusion.....	115
Conclusion	116
References	118
Papers published in the period of Ph.D. education	127
Statements of Copyright	129
Letter of Authorization	129
Acknowledgement	130
Resume	132

图表目录

图 1-1 视角、尺度、遮挡等因素变化而引起的目标表现变化	3
图 1-2 光照对目标表现的影响示例, 上行图片在室内环境下采集, 下行图片在室外环境下采集(NIST-NSWC-USF数据库)	4
图 2-1 标准序列蒙特卡洛滤波算法流程	25
图 2-2 基于重采样的标准序列蒙特卡洛滤波算法	29
图 3-1 嵌入在线特征选择过程的目标跟踪方法流程图	34
图 3-2 当前跟踪系统中所采用的Haar特征	35
图 3-3 弱分类器及其与前景背景特征值分布的关系	36
图 3-4 表现采样和弱分类器训练示例	38
图 3-5 图像标注信息的示例	45
图 3-6 MS算法在人体序列A上的部分跟踪结果示例	46
图 3-7 MSR算法在人体序列A上的部分跟踪结果示例	47
图 3-8 提出的算法在人体序列A上的部分跟踪结果示例	47
图 3-9 MS跟踪算法在人体序列A上的平均跟踪误差	48
图 3-10 MSR跟踪算法在人体序列A上的平均跟踪误差	48
图 3-11 提出的算法在人体序列A上的平均跟踪误差	49
图 3-12 MS算法在人体序列B上的部分跟踪结果示例	50
图 3-13 MSR算法在人体序列B上的部分跟踪结果示例	51
图 3-14 提出的算法在人体序列B上的部分跟踪结果示例	52
图 3-15 提出的算法在人体序列B上跟踪性能的定量分析	52
图 3-16 MS跟踪算法在人体序列B上的平均跟踪误差	53
图 3-17 MSR跟踪算法在人体序列B上的平均跟踪误差	54
图 3-18 提出的算法在人体序列B上的平均跟踪误差	54
图 3-19 MS算法在汽车序列上的部分跟踪结果示例	55
图 3-20 MSR算法在汽车序列上的部分跟踪结果示例	56
图 3-21 提出算法在汽车序列上的部分跟踪结果示例	56
图 3-22 MS算法在汽车序列上的跟踪误差曲线	57
图 3-23 MSR算法在汽车序列上的跟踪误差曲线	57
图 3-24 提出的算法在汽车序列上的跟踪误差曲线	58
图 3-25 MSR跟踪算法的背景采样策略。从前景和前景周围的背景分别	

建立直方图, 然后将二者的相似性度量作为可区分性判据	58
图 4-1 多模型运动估计框架的流程	63
图 4-2 MSMCF算法的流程	67
图 4-3 采用MSMCF算法在测试序列上获得的部分跟踪结果	69
图 4-4 MSMCF, SMCF的跟踪结果和目标真实状态之间的比较	69
图 4-5 MSMCF, SMCF的 D_{σ} 变化和真实状态间的关系	70
图 4-6 MSMCF算法和SMCF算法的 C_e 指标变化	70
图 4-7 目标的运动模式可用NCHTMM运动模型解释的概率	71
图 4-8 SMCF算法(上行)和MSMCF算法(下行)在“seq_fast.tar.gz”序列上跟踪结果的部分比较	72
图 4-9 在序列“seq_fast.tar.gz”上的MSMCF和SMCF算法的跟踪误差比较	73
图 4-10 在“seq_jw.tar.gz”序列上SMCF算法的部分跟踪结果	73
图 4-11 在“seq_jw.tar.gz”序列上MSMCF算法的部分跟踪结果	74
图 4-12 在“seq_jw.tar.gz”序列上MSMCF和SMCF算法的跟踪误差比较	74
图 4-13 MCMCF算法的主要流程	78
图 4-14 MPEG-4 标准中定义的三维头部模型的示例	79
图 4-15 标准序列蒙特卡洛滤波算法产生的无用假设示例	80
图 4-16 头部跟踪问题的多模型粒子滤波推断过程	81
图 4-17 原始图像帧和采用估计结果的合成帧之间的比较	82
图 4-18 头部俯仰的估计值和真实值的比较	83
图 4-19 头部深度旋转的估计值和真实值的比较	83
图 4-20 头部平面旋转的估计值和真实值的比较	84
图 4-21 跟踪过程中粒子数在粒子模型上的分布情况	85
图 4-22 原始视频帧与根据估计结果合成帧之间的对比	86
图 4-23 原始视频帧与根据估计结果合成帧之间的对比	87
图 5-1 等权窗和高斯窗	91
图 5-2 特征点纹理对其跟踪性能的影响	93
图 5-3 图像区域的自回归矩阵的物理意义	95
图 5-4 尺度响应函数示例	96

图 5-5 基于尺度空间理论和多尺度Harris算法的特征点选择流程	97
图 5-6 嘴部特征点的定义及其运动描述模型	98
图 5-7 融合帧间运动估计的面部特征点跟踪算法流程	100
图 5-8 基于尺度空间和多尺度Harris特征点检测算法的结果	105
图 5-9 KLT特征点跟踪算法(下行)和基于尺度空间理论增强的KLT特征点跟踪算法(上行)部分跟踪结果的比较	106
图 5-10 在“wczhang.avi”序列上KLT算法的跟踪误差	107
图 5-11 在“wczhang.avi”序列上基于尺度空间理论增强的KLT算法的跟踪误差	107
图 5-12 特征点可跟踪性度量在自适应特征尺度选择和经验设定之间的差异对比	108
图 5-13 (a) 初始选定的人脸特征点集合 (b) 可以被提出的算法完整跟踪的特征点集合; (c) 可以被KLT算法完整跟踪的特征点集合	110
图 5-14 在“bcao.avi”序列上KLT算法的跟踪误差	110
图 5-15 在“bcao.avi”序列上提出的算法的跟踪误差	111
图 5-16 迭代使用子空间约束后平均误差递减的曲线	111
图 5-17 提出算法的部分跟踪结果图例	112
图 5-18 提出的跟踪算法的部分跟踪结果(上行)和KLT跟踪算法的部分跟踪结果(下行)的比较	113
图 5-19 在“foreman.mpeg”序列上KLT算法的跟踪误差	114
图 5-20 在“foreman.mpeg”序列上提出的算法的跟踪误差	114
表 1-1 本论文主要章节之间的关系	19
表 3-1 MS, MSR和提出的算法在平均跟踪误差上的比较	49
表 3-2 在序列B上MS, MSR和提出的算法平均跟踪误差的比较	55
表 3-3 在汽车跟踪序列上MS, MSR和提出的算法的平均跟踪误差的比较	59
表 5-1 图像区域自回归矩阵和图像灰度值分布之间的关系	95
表 5-2 KLT跟踪算法和提出的跟踪算法的跟踪性能定量分析	109

List of Figures and Tables

Fig.1-1 Examples of observation variations due to viewpoint, scale, occlusion changed	3
Fig.1-2 Illustration the impact of illumination variations on target appearance, top row and bottom row are recorded in indoor environment and outdoor environment respectively (from database NIST-NSWC-USF). 4	
Fig.2-1 Flowchart of standard sequential Monte Carlo filtering algorithm 25	
Fig.2-2 Flowchart of standard sequential Monte Carlo filtering algorithm with particle resampling	29
Fig.3-1 Flowchart of object tracking method with online feature selection embedding	34
Fig.3-2 Haar Features used in current system implementation	35
Fig.3-3 Weak classifier and its relation to distribution of foreground/background feature values	36
Fig.3-4 Illustration of observation sampling and classifier training	38
Fig.3-5 One example of image annotation information	45
Fig.3-6 Some tracking results on people sequence A by the MS tracker	46
Fig.3-7 Some tracking results on people sequence A by the MSR tracker	47
Fig.3-8 Some results on people sequence A by the proposed tracker	47
Fig.3-9 Mean tracking errors on sequence A by the MS tracker	48
Fig.3-10 Mean tracking errors on sequence A by the MSR tracker	48
Fig.3-11 Mean tracking errors on sequence A by the proposed tracker	49
Fig.3-12 Some tracking results on people sequence B by the MS tracker	50
Fig.3-13 Some tracking results on people sequence B by the MSR tracker	51
Fig.3-14 Some tracking results on people sequence B by the proposed tracker	52
Fig.3-15 Quantitatively analyzing the performance of the proposed tracker on people sequence B	52
Fig.3-16 Mean tracking errors on sequence B by the MS tracker	53
Fig.3-17 Mean tracking errors on sequence B by the MSR tracker	54
Fig.3-18 Mean tracking errors on sequence B by the proposed tracker	54

Fig.3-19 Some tracking results by the MS tracker on car sequence	55
Fig.3-20 Some tracking results by the MSR tracker on car sequence.....	56
Fig.3-21 Some tracking results by the proposed tracker on car sequence ..	56
Fig.3-22 Tracking error curve on car sequence by the MS traker.....	57
Fig.3-23 Tracking error curve on car sequence by the MSR traker	57
Fig.3-24 Tracking error curve on car sequence by the proposed traker	58
Fig.3-25 Background sampling strategy used in MSR tracker. Constructing color histogram from foreground and its relevant background respectively and treat likelihood ratio between histograms as the discriminative measurement.....	58
Fig.4-1 flowchart of multi-model motion estimation framework	63
Fig.4-2 Flowchart of MSMCF algorithm.....	67
Fig.4-3 Some tracking results obtained by MSMCF on test sequence.....	69
Fig.4-4 Comparison of tracking results obtained from MSMCF, SMCF and target true states.....	69
Fig.4-5 Trajectories of D_{opt} of MSMCF, SMCF and their relations to object true states.....	70
Fig.4-6 Trajectories of C_{opt} of MSMCF and SMCF.....	70
Fig.4-7 Probability that the target resides on NCHTMM	71
Fig.4-8 Comparison of tracking results obtained from SMCF (top row) and MSMCF (bottom row) on video "seq_fast.tar.gz"	72
Fig.4-9 Comparing tracing errors between MSMCF and SMCF on video "seq_fast.tar.gz".....	73
Fig.4-10 Some tracking results obtained by SMCF on video "seq_jw.tar.gz"	73
Fig.4-11 Some tracking results obtained by MSMCF on video "seq_jw.tar.gz"	74
Fig.4-12 Comparing tracking errors between MSMCF and SMCF on video "seq_jw.tar.gz".....	74
Fig.4-13 Flowchart of MCMCF algorithm	78
Fig.4-14 Illustrating the 3D face model defined in MPEG-4.....	79
Fig.4-15 Useless hypotheses generated by standard sequential Monte Carlo	

filter	80
Fig.4-16 The inference structure for head motion estimation.....	81
Fig.4-17 Comparing original and synthesized frames using estimations....	82
Fig.4-18 Comparing estimations and ground truth of head tilt.....	83
Fig.4-19 Comparing estimations and the ground truth of head yaw	83
Fig.4-20 Comparing estimations and the ground truth of head roll.....	84
Fig.4-21 Distribution of particles during tracking in one sub-model	85
Fig.4-22 Comparing original frames with synthesized frames using estimated FAP values.....	86
Fig.4-23 Comparing original frames with synthesized frames using estimated FAP values.....	87
Fig.5-1 Gaussian kernel and uniform kernel	91
Fig.5-2 Characteristic texture and its relation to trackability	93
Fig.5-3 The vivid explanation of auto-correlation matrix computed on image region	95
Fig.5-4 Illustration of scale response function	96
Fig.5-5 Flowchart of interest point selection based on scale space theory and multi-scale Harris algorithm	97
Fig.5-6 Mouth feature points definition and its motion description model	98
Fig.5-7 Flowchart of facial feature tracking algorithm with inter-frame motion estimation	100
Fig.5-8 Feature selection results based on scale space theory and multi-scale Harris detector	105
Fig.5-9 Comparing results from scale-space theory enhanced KLT tracker (top row) and original KLT tracker (bottom row)	106
Fig.5-10 Tracking errors on sequence "wczhang.avi" by KLT tracker.....	107
Fig.5-11 Tracking errors on sequence "wczhang.avi" by scale space enhanced KLT tracker.....	107
Fig.5-12 Comparing features trackability with adaptive scale and fixed scale	108
Fig.5-13 (a)Initial feature set (b)Tracked feature by proposed tracker (c)Tracked features by KLT tracker	110
Fig.5-14 Tracking errors on sequence "bcao.avi" by KLT tracker.....	110

Fig.5-15 Tracking errors on sequence "bcao.avi" by proposed tracker	111
Fig.5-16 Error reduction curve by applying subspace constraints iteratively	111
Fig.5-17 Some tracking results by proposed tracker	112
Fig.5-18 Tracking errors on sequence "foreman.mpeg" by the proposed tracker.....	113
Fig.5-19 Tracking errors on sequence "foreman.mpeg" by the KLT tracker	114
Fig.5-20 Tracking errors on sequence "foreman.mpeg" by the proposed tracker.....	114
Table 1-1 Relations among main chapters.....	19
Table3-1 Comparing mean tracking errors among MS, MSR and proposed algorithms.....	49
Table3-2 Comparing mean tracking errors on sequence B among MS, MSR and proposed algorithms	55
Table 3-3 Comparing mean tracking errors on car sequence among MS, MSR and proposed algorithms	59
Table 5-1 Relations between auto-correlation matrix of image region and distribution of image pixel values.....	95
Table5-2 Quantitative comparison of proposed method and KLT feture tracker.....	109

第1章 绪论

1.1 研究背景

作为视觉的一项基本功能,人类视觉系统可以敏锐地感知所处场景中的物体运动。除非在很困难的情况下,如辨识乒乓球球的运动轨迹,对生理视觉系统而言,运动感知已经成为一种自然而然的无意识视觉行为。然而对机器来说,迄今为止,赋予其感知运动的能力仍然是一项极大的挑战和尚未解决的问题,并作为视觉跟踪领域的主要研究目标而具有广泛的商业应用前景。

随着硬件成本的降低和制造工艺的改进,摄像机等成像设备的性价比快速提高,并迅速普及到生产生活中的方方面面,在安全保障等方面发挥着越来越重要的作用。但是,当前的视频监控系统只能简单地记录事情发生的映像。自“911”事件以来,恐怖主义的猖獗使国家安全危机凸显,赋予视觉监控系统能够理解所拍摄场景的能力成为日益迫切的需求。在2000年,美国 DARPA 启动了 HumanID 项目,其目标是试图通过视觉方法远距离地辨认人的身份信息,而步态等运动信息在其中扮演了重要角色。除视觉监控之外,视觉跟踪技术还可以在基于视频的运动分析等领域发挥重要作用,如可以采用跟踪技术分析运动员的技术动作与理想动作之间的差别。另外运动分析技术还可以用来指导真实感运动的合成技术,从而为虚拟现实技术的广泛应用提供技术支持。

视觉跟踪技术还经常作为很多系统的重要组成部分。在视频分析领域,系统输入是一组在时间轴上具有关联的图像集合。如果计算机能够将图像集合分解为事件的集合,从而得到视频的高层描述和语义特征,将对视频检索等领域产生重要影响。跟踪技术可以通过寻找在时间轴上相同物体的对应,不仅将图像集合分解为以物体为中心的对象集合,同时能够得到物体的运动信息,从而为场景语义的提取提供了重要支持。近些年提出的视频压缩领域的国际标准 MPEG-4 和 MPEG-7 将基于对象的编码和描述列为视频处理的研究方向。而采用跟踪技术的运动分层技术可以对这一策略提供有效支持。另外,在机器人自动导航、智能人机交互、视频分割、军事目标定位等领域,都迫切需要成熟稳定的视觉跟踪技术。

除了上述应用方面的需求之外,研究视觉跟踪技术对理解人类视觉的机

制和探讨人工智能的实现手段也具有重要的指导意义。

1.2 跟踪研究中面临的主要问题

基于视觉的目标跟踪可以定义为根据时间轴连续的图像集合推断目标运动状态的问题^[1]。目标运动状态是问题相关的,可以是目标的2D位置、3D位置、大小、运动方向以及更为复杂的运动姿态等,而目标表现则指包含了目标图像投影的可见光图像。

对目标跟踪来说,首先需要面对的问题是能够在图像中将目标的投影完整而清晰地分割出来。但是,从目标的表现推断目标的状态,首先就是一个具有病态解的问题。当现实世界中的三维物体投影到二维图像平面时,深度信息丢失,而深度信息是将目标从纷繁复杂的背景图像中分割出来的重要依据之一。另外,相对于理论研究中的成像模型,现实世界中的投影矩阵异常复杂,不仅受着诸多随机因素的影响,同时在求解的过程中必须考虑噪声和误差的作用。例如,对图像中的一个像素,其成像过程要受到物体本身的材质、外界的光源、摄像机的视角、透明度和遮挡关系等诸多因素的影响,并且像素值与诸多影响因素的关联关系并非都可采用线性函数表达。虽然外界的变化因素是繁杂的,但反映到图像上,却仅仅是像素值的简单变化,从而掩盖了造成变化的众多原因。对视觉跟踪算法来说,其输入往往是以万为数量级的像素在时间轴上的变化,算法需要从如此高维的输入中选择信息进行推理,同时需要抵抗噪声和不确定性因素的干扰。

视觉跟踪问题的研究经过了几十年的积累和诸多学者的不懈努力之后,取得了不俗的成果并出现了一些进入某些特定应用场合的商业系统,但总体来说,对大规模应用和达到工业标准,仍面临重重困难:

- 1) 应用场合中具有复杂的背景变化:复杂背景是引起跟踪失效的最重要因素之一。背景本身是不断变化的因素。另外,出于实时计算的考虑,要求刻画目标表现所采用的图像特征及其模型必须简单,从而有可能造成背景在所选择的图像特征上与目标相似,在根本上使算法无法区分目标和背景,造成跟踪的失效。例如,选择颜色直方图作为目标表现刻画方法,如果背景中有区域和目标在颜色分布上相近时,则在算法被背景“吸引”从而丢失目标的危险。需要指出,所谓“复杂背景”是与所选择的图像特征和应用场合紧密相关的。
- 2) 目标具有复杂的运动模式:基于计算性和鲁棒性的考虑,跟踪算法主要

采用局部搜索的策略,即只关心目标以较大概率出现的区域。但当目标具有复杂的运动模式时,如目标的运动速度或者方向突然发生改变,未对特殊情况进行建模的简单预测机制容易失效,造成算法在目标并不存在的区域中进行目标的搜索,从而造成目标丢失。

- 3) 场景中物体对目标的遮挡和目标的自遮挡(如图 1-1(c)、(e)、(f)所示):目标遮挡情况更多出现在多目标跟踪的问题中。遮挡是造成目标的图像表现突然变化的重要原因之一,并且使图像表现变化具有突然性和不连续性等非线性特征,从而容易引起既有目标表现模型的失效。
- 4) 光照变化:如图 1-2 所示,相同的目标状态在不同的光照条件下具有差异明显的目标表现。尤其当目标处于室外环境中,复杂的光照变化如阴影等会引起目标图像表现的剧烈变化。光照影响是很多实际应用中阻碍系统性能提升的瓶颈因素问题。
- 5) 视点变化和目标的非刚性形变(如图 1-1(b)、(d)所示):该两因素会造成目标图像表现的非线性变化,而非线性变化的建模本身就是困难的问题。



图 1-1 视角、尺度、遮挡等因素变化而引起的目标表现变化示例

Fig.1-1 Examples of observation variations due to viewpoint, scale, occlusion changed

上面的诸多困难因素中,在根本上可以归结为两类问题,首先是如何建立和维护正确的目标表现模型的问题。前面描述的困难 1、3、4 和 5 中,都需要鲁棒准确的目标表现建模和更新过程。而现实世界中的跟踪问题,由于

原书缺页

要处理的是时变信号的分类问题：即将时变的前景(目标)从时变的背景中正确地、连续地分离出来。

如何排除种种不确定和变化因素的影响，在“未来”的场景映像中推断出目标在时间轴上的对应关系，最主要的倚重因素就是正确地建立目标表现的刻画模型，从而对目标和背景的分类提供根本的保证。表现的提取过程则主要指在目标表现模型的基础上找到相邻帧间目标表现的对应，提取方法与所采用的目标模型是密切相关的。

本小节按照目标表现的建模及其提取策略的不同对前人的工作进行分类总结。

1.3.1.1 基于颜色分布的方法

图像区域颜色分布具有对旋转、小幅度仿射变换等变化和图像噪声较鲁棒的优点，并且计算简单，符合实时处理的计算要求。但是，颜色分布在对目标表现的表面颜色进行统计的同时，丢失了其颜色分布的几何特性，无法根据图像区域的颜色直方图重建图像区域，从而使完全不同的图像区域可能具有相同的颜色分布，所以，其对图像区域的刻画能力是偏弱的，只能编码目标相关的部分图像信息。

但颜色分布仍然是视觉跟踪中应用最广泛和最成功的目标刻画方法之一^[10-12]。通过将人的头部用椭圆近似，Birchfield提出了采用直方图结合椭圆周梯度进行头部跟踪的方法^[13]。作者认为，颜色直方图刻画了椭圆内部的纹理信息，椭圆周梯度则表示了椭圆轮廓的信息，根据集合论的观点，二者在目标刻画上具有互补性，即当一种表示方法失效时，另一种方法往往能起到矫正的作用。该方法计算简单，满足实时跟踪的要求。所采用的颜色空间为RGB空间的变形，三个颜色分量分别为色度分量G-R和B-G，光照强度分量 $(R+G+B)/3$ 。

相对于RGB颜色空间，Bradski认为HSV色彩空间中的H分量(色调分量)能够更好地区分人的肤色和其它自然色^[14]。算法将在H分量上肤色隶属度的分布用直方图概率密度估计的形式求取，然后对整幅图像中每个像素的肤色隶属度进行计算，最后采用Camshif算法找到一个与人脸形状相似的具有最大概率可能的图像区域作为人脸跟踪结果。

为了弥补直方图的刻画能力不足的弱点，Comaniciu采用了空间加权直方图。Enpanolkov核函数使靠近图像区域中心的像素具有进行较大的权重从而部分编码了像素颜色分布的空间信息，使加权直方图相对于标准直方图具有了更好的目标定位精度^[15]。该算法根据直方图是密度分布估计的一

种，采用Bhattacharyya距离度量模型和图像区域直方图之间的相似度，并引入mean-shift梯度上升算法来迭代匹配模型到正确的图像区域。不同于局部穷举搜索的方式，基于梯度上升的mean-shift方法降低了搜索的计算复杂性。为了满足mean-shift算法的应用条件，算法要求相邻帧间的目标表现具有重叠的区域，该缺陷可以采用基于滤波的预测框架加以弥补。

与上述方法采用单一色彩空间计算直方图的方法不同，自适应地选择合适的色彩空间的方法首先在[16]的工作中得到了体现。在该工作中，作者注意到没有任何色彩空间在所有人脸跟踪任务中总是优于其它色彩空间的。在算法实现中，预定义了多个候选色彩空间，并以对被跟踪目标和背景的分度作为选择色彩空间的标准，从而通过对背景的评估，自适应地从多个色彩空间中选择一个构建面部直方图来保证跟踪算法的鲁棒性。

自适应地直方图特征选择思想在[17]的工作中被用来进行一般图像区域的跟踪。在算法中，作者通过对RGB颜色空间中的三个色彩分量进行加权 $w_1R+w_2G+w_3B$ 得到不同包含直方图的特征空间，其中，

$\{w_i | i=1,2,3\} \in [1,5]$ ，共有49种不同的特征空间可供选择。与[16]中的自适应选择标准类似，算法定义了似然比来度量特定直方图对前景和背景的区分能力，作为特征空间选择的标准。

国内方面，刘明宝博士在[18]中提出了一种在复杂背景中实时跟踪人脸自由运动的方法。在人脸直方图的表达上，将图像的RGB空间变换到色度空间，然后利用最大能量坐标和矩表征色度空间的直方图分布。该方法中的运动模型则结合了运动检测与运动预测，以减少搜索区域，提高检测速度。

在[19]中，姚鸿勋教授分析得出了关于色度与彩色坐标系的关联关系，建立了肤色和唇色色系的坐标变换映射，从而得到了对象姿态、背景鲁棒的人脸面部定位与跟踪的方法。通过进一步地在线学习，还可以去除光照条件变化的影响及摄像设备参数变化的影响。

1.3.1.2 基于模板的方法

模板可以有效编码目标表现的信息，而且通常仅依靠较少的初始图像数据量就可以进行构造^[20, 21]。但是由于模板通常带有全局性的信息，并以像素为基本单位，难以有效地得到更新来反映目标表现的实时变化，所以基于模板的跟踪方法更适合于短时跟踪任务，而在长时序的跟踪任务中，基于模板的方法容易产生目标“漂移”现象^[22]。

Frey 在^[23]中对传统的模板方法做了改进。对目标区域的每一像素,算法根据一段时间内的目标表观求取其变化范围。该模板对噪声具有更好的鲁棒性,但是增加了在训练集合上对图像像素的灰度变化进行预学习的过程。在随后的工作中,Jojic 针对场景中存在多物体的情况,建立了层次性的基于模板的目标模型。由于该生成式模型中包含了叠加性高斯噪声和阿尔法通道,具有处理多目标深度信息和透明性遮挡等情况的能力^[24]。模型的学习采用了效率优于最大后验概率方法的迭代算法。在迭代过程中,在不同运动层间传递关于灰度的概率分布信息,在层内水平地传递多目标的标识信息。但在算法中,运动层的个数和目标物体的数目都是假设预知,从而限制了算法的实用性。

Rucklidge 在^[25]中提出了一种有效率地定位灰度模式的方法。算法基于块匹配策略,采用平方差之和(Sum of Squared Differences, SSD)作为图像匹配度准则。算法将特定模式匹配到图像区域的函数变换空间划分为若干子空间。在每个子空间中,考虑所有可能的变换形式,对变换后的特定模式和图像区域之间的差值根据经验设定阈值,从而忽略那些最佳匹配值亦大于阈值的子空间来降低计算量。对最佳匹配值低于阈值的子空间则进行进一步划分,并采用深度优先的搜索策略进行求解。

Olson 等人认为广泛使用的模板匹配准则 SSD 对所有像素等同看待,所以对噪声等外部影响敏感,并且该准则不随目标表观和表观的偏离情况呈线性关系^[26]。算法将模板匹配过程纳入到基于极大似然估计的概率求解框架中。为达到亚像素的匹配精度和处理遮挡等诸多非确定因素,算法中采用高斯分布来拟合模型表观间相似性分布的峰值区域,并将高斯分布的峰值作为匹配的位置,高斯分布的方差则给出了估计结果的信度。在相同条件下的实验表明,该算法优于基于 SSD 匹配准则的方法。

Morris 分析了运动范围约束对跟踪有关节运动特性的运动物体的作用,认为当运动与摄像机的视角同方向时,会造成约束矩阵的奇异^[27]。如果矩阵奇异,算法的收敛性变差甚至丢失被跟踪目标。作者针对该问题提出了二维尺度棱镜模型(2D Scaled Prismatic Model, SPM),相对于三维运动模型,该模型更少发生矩阵奇异问题,并且不需要预先学习特定的三维运动规律。算法还可以对发生矩阵奇异的情况进行判断从而在发生奇异时通过借助三维模型进行跟踪。

在面部图像分析领域,高文教授在国内研究领域做出了早期的工作^[28]。该方法建立了基于部件分解组合的人脸图像模型。通过对部件的分

析,采用分类树建立表情模型的向量表示,从而可以根据能量优化原理,利用模板匹配方法提取目标特征,得到人脸表情的表征向量从而实现表情的识别。

苗军博士提出了一种基于重心模板的实时人脸检测方法^[28]。并将该方法应用于人脸检测和跟踪任务中,该方法首先将图像马赛克化,通过计算马赛克图像中显著边缘的重心,利用规则的方法确定人脸的位置。该方法能够检测水平和深度旋转的人脸。

1.3.1.3 基于轮廓的方法

轮廓是目标刻画中广泛使用的图像特征,轮廓点通常包含目标的重要信息。基于轮廓的方法一般求解策略是通过迭代方法趋近于问题相关的能量函数的局部最优解。基于轮廓的方法通常对搜索的起始位置比较敏感,所以初始位置的选择是基于轮廓方法性能的重要倚赖。在视觉目标跟踪中,算法通常将前帧的搜索结果作为后帧的起始搜索位置。

Snake 算法是视觉跟踪中最著名的轮廓模型之一,其采用样条曲线连成的封闭轮廓来逼近目标物体的形状^[29]。算法中的能量函数包含三个子项:控制轮廓光滑程度的内力,控制初始位移的外力和控制边缘、特征点吸引程度的图像力。通过梯度下降来迭代最小化上面三个能量项的加权和,最终收敛到能量一维分布中的某局部极小作为算法的输出,其中三个能量项的权重往往需要根据具体应用问题进行调整。

Leymarie 采用 Snake 算法进行细胞轮廓的跟踪^[30]。细胞轮廓的初始位置在参考帧中由人工确定,并采用前帧中的估计结果作为后帧中的初始搜索位置。当相邻帧尖的细胞位移和形变较小,能观察到算法具有较好的跟踪性能。与原始 Snake 算法中的能量项不同,作者采用轮廓上的分段能量平均最小值作为优化标准,并认为相对于求取^[29]中定义的目标函数最小,新的能量函数在避免搜索中的振荡和定位到最优解的频率上都有提高。

Paragios 根据水平集(Level Set)理论提出了测地主动轮廓模型(Geodesic active contour)^[31]。在该算法中,根据水平集理论递归地改变预定义轮廓的位置和形状来拟合目标的轮廓,达到同时跟踪和检测目标的目的。

上述的基于轮廓的目标跟踪方法主要采用经验定义的能量函数,并试图利用通用的能量函数解决不同的跟踪问题。在^[32, 33]的工作中,学习的机制被引入到主动轮廓模型中,虽然得到模型的过程需要繁琐的学习训练,并且学到的模型是应用相关的,但这种针对特定问题的模型表现出了比通用模型优秀的定位和跟踪特性。

主动形状模型(Active Shape Model, ASM)和主动表面模型(Active Appearance Model, AAM)都是基于点分布模型(Point Distribution Model, PDM)的算法[32-35]。在 PDM 中,特定类别物体(如人脸、人手)的轮廓形状通过若干关键的特征点进行定义,这些特征点的坐标串接构成描述目标的原始形状向量。对训练集中的所有形状向量进行对齐操作后,对他们进行 PCA 分析建模,保留的主成分形成最终的形状模型,形状模型的参数个数反映了形状的主要可变化模式^[34]。ASM 搜索首先通过局部纹理模型匹配得到各个特征点的更佳的位置,经过相似变换对齐后,通过统计形状模型对其进行约束,而后再进行局部纹理模型的匹配,形成一个迭代过程,以期形状模型最终匹配到输入的形状模式上去。在 ASM 中,仅使用了特征点局部纹理特征作为启发式信息,没有使用全局的纹理约束,实践中发现 ASM 很容易陷入局部极小。

而在 AAM 中,则采用了形状和纹理二者融合的统一约束,即所谓的统计表现模型。AAM 搜索借鉴了基于合成的分析技术(Analysis-By-Synthesis, ABS)的思想,通过模型参数的优化调整使得模型能够不断逼近实际输入模式。模型参数的更新则放弃了 ASM 中的局部纹理搜索过程,仅使用一个线性预测模型根据当前模型和输入模式之间的差别来预测和更新模型参数。AAM 尽管利用了全局纹理,但却抛弃了局部纹理匹配过程,因此会在一定程度上降低关键特征点配准的精度,而且其线性预测模型也有较大的局限性,在初始位置偏离目标位置过大时,则很难收敛到正确位置。

除了在轮廓模型本身的改进,Isard 将主动轮廓模型融入到 CONDENSATION(Conditional Density Propagation)算法中^[3]。在算法实现中,轮廓形状的后验分布采用一组离散的粒子及其相应的权重表示。这种非参数的离散表示方法可以求解任何形式的后验分布。而前述算法中主要采用基于梯度下降的策略,相当于力图求解目标轮廓形状的后验概率分布的期望。当后验分布呈尖峰状分布的时候,期望具有较好的解性质,否则,很容易造成被跟踪目标的丢失。而在[3]中,采用粒子模拟后验概率分布的方法不仅解决了由于暂时的误差所造成的误差传递问题,同时也在一定程度上解决了主动轮廓模型对其搜索位置要求较苛刻的问题。

1.3.1.4 基于子空间的方法

从本质上说,跟踪问题处理的是非静态信号,前景目标和背景都在随着时间的变化和目标的运动而改变。虽然现在有无数的方法能在短时间段内和可控环境中很好地跟踪目标,但是当目标表现本身发生剧烈变化或者环境

(比如光照)发生不可预期改变时,往往会造成跟踪中的“漂移现象”,造成跟踪任务的失败。在基于子空间建模的方法中,由于其能够充分地利用历史数据或者训练集得到目标物体表现变化的子空间,从而提供了完整描述前景目标变化的能力,减少因为目标的表现变化所产生“漂移现象”的可能。

Black 提出了基于视角的子空间表示方法并将两不同视角的子空间匹配过程归结为求解优化问题^[7]。该工作的主要贡献有两点:首先,对匹配中具有较大误差的特征项进行降权,提高了误差范数的鲁棒性,从而避免了等权使用特征系数时对噪声和奇异点敏感的缺点,比如发生部分遮挡等情况。其次,算法中泛化了光流计算中的灰度恒定假设,提出了子空间一致性假设。该假设认为,目标的视角图像,如果找到正确的参数变换来调整图像,可以用相同的特征基组进行重建,并且可以保证重建图像和原图像具有相同的图像灰度分布。在该假设的基础上,通过求取目标视角图像和子空间重构图像之间的形变参数,做到视角无关的目标跟踪。在其后续工作中^[34],作者进一步采用混合模型构造子空间来建模目标表现的变化。

Torre 在[37]的基础上,将基于子空间的表示方法应用到人体跟踪任务中^[39]。根据当前需处理的帧数据和训练集中图像的相似性,选择训练集中的相似度高的一些子集来构造所跟踪人体的特定子空间。算法中采用肤色进行面部区域的分割,并采用了卡尔曼滤波算法来估计仿射运动模型的参数。

Hager 和 Belhumeur 提出了基于参数化运动模型的模板匹配方法。通过梯度下降策略来求解运动模型的参数从而将固定的模板匹配到正确的图像位置^[7]。由于是应用在视觉跟踪中,为了避免每帧都要计算雅可比矩阵,该工作将矩阵分为由模板的亮度梯度乘以其变换的空间微分的常量项和针对运动模型的参数进行微分的可变量,从而大大简化了计算复杂度。该方法在模板的基础上利用子空间方法处理光照问题,通过对涵盖各种光照情况的训练图像的学习得到光照描述的子空间,算法中部分遮挡问题也采用 M 估计算子(M-estimator)进行了简单地处理。

Jepson 等人提出了一种融合稳定部分,噪声部分和帧间变化部分的目标模型^[40]。该模型采用小波基响应作为图像特征,采用期望最大化(expectation-maximization)算法估计混合模型中三部分的权重并且采用稳定部分进行仿射运动的估计。在该模型能够处理目标表现变化和光照变化的同时,作者指出在背景变化平缓的情况下,该模型也可能同时估计了背景的稳定部分,从而有可能造成被跟踪中常见的“漂移”问题。