

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

王建宇

哈尔滨工业大学

2006年2月

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

Classified Index: TP391.41

U.D.C: 681.39

Dissertation for the Doctoral Degree in Engineering

SEQUENTIAL MONTE CARLO FILTERING BASED VISUAL TRACKING

Candidate:	Wang Jianyu
Supervisor:	Prof. Wen Gao
Academic Degree Applied for:	Doctor of Engineering
Speciality:	Computer Application
Affiliation:	School of Computer Science and Technology
Date of Defence:	February, 2006
Degree-Conferring-Institution:	Harbin Institute of Technology

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，重点探讨了目标表观建模，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观模型的自适应建模和更新算法。表观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观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观建模和更新算法。该算法在动态建模过程中不仅考虑目标表观信息，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，在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洛滤波算法中，从而针对两类复杂运动问题提出了标准序列蒙特卡洛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

蒙特卡洛滤波算法。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洛滤波算法, 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洛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形变特征点估计的起始搜索位置, 大大加速了序列蒙特卡洛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子空间约束来迭代求精跟踪结果。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: 目标表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。创新点二和三则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 视觉跟踪; 序列蒙特卡洛滤波算法; 在线特征选择; 运动模型; 光流跟踪

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 probelems in visul 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 exsiting visual tracking algorithms. The second and third ideas try to describe object dynamic explicitly from the motion model or impilcilty from discrete feature points' motions respectively. Therefore, the three proposed algorithms are somewhat complementary and can be choosen for different application scenarios.

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