

MIANXIANG SANWEI

shoushi genzong he jiaohu yanjiu de kexue wenti yu yanjiu shexiang

面向三维手势跟踪和交互 研究的科学问题与研究设想

冯志全 编著



电子科技大学出版社

面向三维手势跟踪和交互研究的 科学问题与研究设想

冯志全 编著

电子科技大学出版社

图书在版编目(CIP)数据

面向三维手势跟踪和交互研究的科学问题与研究设想/

冯志全编著. — 成都: 电子科技大学出版社, 2014. 4

ISBN 978-7-5647-2289-0

I. ①面… II. ①冯… III. ①人一机系统—研究

IV. ①TB18

中国版本图书馆 CIP 数据核字(2014)第 059684 号

内 容 提 要

本书旨在揭示三维手势跟踪研究的科学问题以及系统化研究设想,强调学术性、前瞻性和创新性。本书还介绍了作者所在团队的最新研究成果。本书以项目申请书的形式,系统性地介绍作者所在团队面向三维手势跟踪和交互研究的构想和思路,涉及计算机图形图像处理、人机交互理论、认识理论、计算机视觉、射影几何学、预测估计方法学、软计算理论等众多交叉学科或研究领域。

本书可以作为信息科学技术领域高年级本科生或研究生教材,也可以供从事人机交互方向的科研和技术人员参考,其主要特点在于它可以作为国家自然科学基金申请书的参考资料。

面向三维手势跟踪和交互研究的科学问题与研究设想

冯志全 编著

出 版: 电子科技大学出版社(成都市一环路东一段 159 号电子信息产业大厦 邮编:610051)

策划编辑: 谢应成

责任编辑: 谢应成

主 页: www.uestcp.com.cn

电子邮箱: uestcp@uestcp.com.cn

发 行: 新华书店经销

印 刷: 成都市火炬印务有限公司

成品尺寸: 185mm×260mm 印张 12.25 字数 300 千字

版 次: 2014 年 4 月第一版

印 次: 2014 年 4 月第一次印刷

书 号: ISBN 978-7-5647-2289-0

定 价: 30.00 元

■ 版权所有 侵权必究 ■

◆ 本社发行部电话:028-83202463;本社邮购电话:028-83201495。

◆ 本书如有缺页、破损、装订错误,请寄回印刷厂调换。

前 言

从 20 世纪的桌面计算时代转变为新世纪的普适计算时代,“以人为中心的信息社会”与“和谐人机环境”等相关研究也应运而生,并成为当前信息领域新的发展方向和研究热点。为此,研究新一代的自然和谐的人机交互理论和技术,发展和应用以用户为中心的设计方法,从而保证普适计算时代的信息技术能够创造和谐的人机环境,使全体公民都能分享信息技术发展的成果是 21 世纪人类社会需要解决的重大课题。随着计算机交互设备和交互技术的不断发展,二维图形用户界面的局限性越来越明显地体现出来。(1) 从界面的信息表示能力来看,二维图形技术的一个重要的缺点是不能用一种自然的方法表示复杂的多维关系。(2) 从交互方式而言,在虚拟现实等环境下很难用传统的交互方式来进行自然、和谐的表达,因为用传统方法反而大大增加了用户的交互难度,同时也加重了交互任务的整合工作。(3) 用户界面的发展历经了批处理、命令行、图形界面三个阶段,现在的研究和开发重点已经放在了 Post-WIMP 界面上。因此,三维人机交互界面(Human-Computer Interface, HCI)的研究已经成为一个紧迫的研究课题。目前, HCI 技术正在从以计算机为中心逐步转移到以人为中心,具体表现为:(1) 交互方式逐渐演化为适应人类的行为习惯,而不是计算机程序,更加强调以人为本;(2) 使用多种媒体、多种模式进行交互;(3) 基于多通道和多媒体的自然、高效、智能化、无障碍的 HCI 将是新一代智能 HCI 的主要发展方向。

自然人机交互是利用人的日常技能进行的,强调无须特别训练或不需要训练。在基于人脸、头部、手臂、人手、人眼以及整个人体的输入方式中,由于在通信和操作中的灵巧性,人手是最有效、用途最多的输入工具。手势是一种自然、直观、易于学习的人机交互手段,以人手直接作为计算机的输入方式,人机间的通信将不再需要中间媒体,用户可以简单地定义一种适当的手势来对周围的机器进行控制;手势是人与人之间的一种非口头交流形式,它包括从用手指示方向和移动物体的简单动作到能够表达感情以及允许彼此交流的复杂手势。考虑到人们拥有做手势的大量经验知识,如果人们能够把这些技能从日常的经验中转换过来并用在人机交互方面就可以期盼直观的、操作简便的、功能强大的人机接口。

VBI (Vision-Based Interface) 是人机交互 (HCI) 中的一个核心课题,而基于计算机视觉的自然手 (Natural Hand or Naked Hand) 的 3D 跟踪又是 VBI 中的一个重要内容,利用人体姿势尤其是手势作为输入设备已经成为人机智能交互 HCII (Human Computer Intelligent Interaction) 或感知用户界面 PUI (Perceptual User Interface) 的重要组成部分。HCI 技术已经从以计算机为中心逐步转移到以人为中心。一方面,交互方式逐渐演化

为适应人类的行为习惯，而不是计算机程序；另一方面，基于多通道和多媒体的自然、高效、智能化、无障碍的 HCI 将是新一代智能 HCI 的主要发展方向。

基于计算机视觉的人体运动跟踪是目前国内外的一个研究热点，运动人手跟踪是人体跟踪中的一个重要组成部分。运动人体跟踪主要包括人脸跟踪与识别、面部表情跟踪与识别、头部运动跟踪、凝视跟踪、手势跟踪以及肢体跟踪与识别等等，并与 HCI 相融合，使得人机交互和多通道用户界面的研究具有应用价值大、研究内容丰富的特点。

基于自然人手的人机交互理论和应用已经成为国际上一个重要研究方向。在过去 20 多年中，国内外的研究者围绕关节式物体的视觉运动分析这一课题做出了大量的研究，覆盖了运动检测、三维建模、运动估计、跟踪与识别、行为理解与语义分析等内容，涉及图像处理与分析、计算机视觉、计算机图形学、人工智能、模式识别和统计数学等多个研究领域，由于其极具趣味性、挑战性以及广阔的应用前景，已经受到世界上越来越多国家、政府、企业界、大学和科研机构的关注。美国南加州大学制定了中长期研究计划，包括运动人手的实时跟踪与识别，以及在多模态人机交互中的应用；此外，牛津大学、剑桥大学、波士顿大学、大阪大学、瑞士皇家学院等众多大学和研究机构已经在自然人手跟踪领域投入大量人力和财力展开相关的理论研究和应用研究。尤其是随着多模态、多通道人机交互理论和技术的迅速发展，VR 技术（虚拟制造、协同虚拟装配）已经成为基于计算机视觉的自然人手跟踪研究的最重要的动力和需求源泉之一。

本书以申请书的形式揭示三维手势跟踪研究的科学问题以及系统化研究设想，强调学术性、前瞻性和创新性，本书还介绍了作者所在团队的最新研究成果。

由于研究水平有限，书中错误在所难免，欢迎读者批评指正。

冯志全

2012 年 12 月

（成都东山文化公司荣誉出品）

目 录

第一部分 科学问题与研究设想

第一章 基于计算机视觉的运动人手多鲁棒性和实时性 3D 跟踪理论与方法研究 (1)

第二章 面向实时性和鲁棒性的自然人手粒子滤波跟踪方法研究 (16)

第三章 面向人手三维跟踪的粒子滤波器中的实时性和鲁棒性问题研究 (40)

第四章 基于 Monte Carlo 方法的人手跟踪方法研究 (72)

第五章 虚拟样机人机交互界面设计中自然人手的跟踪方法研究 (83)

第六章 基于认知心理学的自然人手状态采样方法及其在运动跟踪中的应用研究 (91)

第七章 面向云计算的自然交互界面核心技术与应用示范 (105)

第二部分 研究进展

第一章 手势状态采样方法 (130)

第二章 手势跟踪理论与方法 (142)

第三章 手势跟踪与交互 (165)

第一部分 科学问题与研究设想

第一章

基于计算机视觉的运动人手多鲁棒性和实时性 3D 跟踪理论与方法研究

针对基于计算机视觉的运动人手多鲁棒性和实时性问题,研究对于光照条件和复杂背景均具有鲁棒性的肤色模型;研究人手 3D 模型与算法精度之间的关系;研究满足跟踪算法鲁棒性和实时性需要的自遮挡模型;研究满足跟踪算法鲁棒性和实时性需要的运动模型;集成性地、系统性地研究跟踪算法对于人手四模型的多鲁棒性和实时性,重点揭示在跟踪算法中速度和精度之间对立统一之形式;在满足对手进行实时性 3D 跟踪的条件下,确定粒子滤波器中采样粒子个数的上确界和下确界以及取得这种边界值的基本条件;反过来,在采样粒子数小于某给定值的前提下,进一步研究降低跟踪算法时间复杂度的理论和方法;研究跟踪算法,研究 Bayesian 滤波器和粒子滤波器中针对运动人手进行采样、获取粒子权重以及后验概率分布的新方法;重点研究降低算法时间复杂度的理论和方法;研究认知科学和神经生理学的成果尤其是达尔文自动机和运动神经内模理论;构建跟踪算法评价体系。

一、立项依据与研究内容

1. 项目的立项依据

人机交互和信息处理是公认的 21 世纪四项重大技术之一,“和谐人机交互理论和智能信息处理基础研究”是国家重点基础研究发展计划(973 计划)中“十五”后三年重点支持的方向之一,在《中华人民共和国国务院国家中长期科学和技术发展规划纲要(2006~2020 年)》中,把“智能感知技术”以及“虚拟现实技术 VR”作为重点发展领域及优先发展主题。

新型人机交互界面(HCI)是“智能感知技术”以及“虚拟现实技术 VR”中的核心内容之一。一方面,目前仍然以键盘和鼠标为主要的 HCI 工具,但在很多情况下,键盘、鼠标不能实现和谐的人机交流,或者不能确切地、直接地、自然地表达人们的意图,故传统的人机界面已经受到越来越多的批评^[1];另一方面,数字手套技术具有算法设计简单、精度高、速度快、无须摄像设备、数据采样结果不受光线等外界条件的影响、可以直接获

得手在空间的三维信息和手指的运动信息等优点,但存在影响操作者的沉浸感、价格昂贵、容易损坏等缺陷。所以,利用人体姿势^[2]尤其是运动人手的三维(3D)跟踪作为输入方式已经成为人机智能交互 HCII (Human Computer Intelligent Interaction) 或感知用户界面 PUI (Perceptual User Interface) 的重要组成部分,不仅在手语跟踪识别、虚拟现实、虚拟制造以及虚拟博物馆等领域有直接的应用,而且在 CAD/CAM、视觉监控、机器人控制以及远程会议等领域也具有重要的应用价值,基于自然手的人机交互界面已经成为自然、方便、和谐人机交互界面的基本要求^[3,4]。

具体到虚拟考古数字博物馆的 HCI 研究,首先需要对用户的运动人手进行 3D 跟踪,生成并绘制亦步亦趋的 3D 虚拟手,在此基础上,用可视化虚拟手完成对虚拟展品的选择和操作。可见,对运动人手进行 3D 跟踪是一件基础性工作,多鲁棒性问题和实时性问题是运动人手 3D 跟踪理论和应用研究中的两个相互联系的核心问题,又是基于 3D 手势的 HCI 系统理论和应用研究中的两个最突出的矛盾。

(1) 多鲁棒性

所谓多鲁棒性是指人手肤色模型、人手 3D 模型、人手自遮挡模型以及人手运动模型(以下简称为四模型)中的每个模型在一定范围变化时,跟踪算法的精度仍然保持稳定之能力。

①跟踪算法对于肤色模型的鲁棒性

一般假设跟踪背景保持固定不动,从而影响鲁棒性的关键因素是光照条件以及背景的复杂性。为了解决鲁棒性问题,人们提出了 RGB 空间模型、标准化 RGB 空间模型、HSV 空间模型、YCbCr 空间模型以及 YIQ 空间模型等多种静态肤色模型,有的结合 RGB、HSV 和 Luminance/Chrominance 等实现背景分割^[5]。为了克服静态肤色模型适应性差等缺陷,人们又提出了动态直方图、高斯分布自适应、均值漂移和肤色轨迹白平衡等动态肤色分布模型^[6]。有的模型对光照条件具有较好的鲁棒性,有的对复杂背景具有较好的鲁棒性,但是,对于光照条件和复杂背景均具有鲁棒性的肤色模型还需要进一步研究。

②跟踪算法对于 3D 人手模型的鲁棒性

人手 3D 模型在跟踪算法设计中的主要作用有:(I) 定义所需要的可视化输出,即由每个关节角度所组成的状态向量;(II) 在状态与图像观测值之间定义映射,这使得配准(Registration)操作成为可能。目前广泛采用生理模型、带纹理的体模型、网格模型、骨架模型、轮廓模型和边界模型等基本模型,甚至采用最简单的纸板(Cardboard)模型^[7],但另一方面,由于模型的精确度对于跟踪精度影响很大,所以有的系统反而采用更加真实的模型,例如,在文献^[8]中,把骨架模型转化为 B 样条曲面,把关节点作为 B 样条曲面的控制点。

虽然人们认识到人手 3D 模型对算法精度影响很大,但对这种影响程度还缺乏定量的研究,例如,对跟踪算法精度影响最小的人手 3D 模型的研究,对人手 3D 模型与跟踪精度之间的独立性条件的研究尚不多见。

③跟踪算法对于自遮挡模型的鲁棒性

人手是一种典型的多链接体,具有多自由度且其运动过程具有随意性和复杂性等基本

特点,由此而产生的遮挡问题对于跟踪算法的精度以及实时性等系统的关键性能产生深刻影响。有的采用多相机解决遮挡问题,然而,算法的时间开销与相机的个数成正比例关系;有的根据局部遮挡不变量建立遮挡图(Occlusion Graph)^[9],但这种遮挡图与人手3D模型本身的精度关系极大。还有的通过建立可见性表(Visible Table),但限制在抓取过程中分析遮挡条件^[10]。

与此同时,这些研究忽略了建立人手自遮挡模型,这种模型应该与人手3D模型无关,且根据该模型可以正确地、最大限度地恢复被遮挡特征点。

④跟踪算法对于人手运动模型的鲁棒性

这里的运动模型包括:运动速度和加速度的改变方式、状态误差的分布方式以及状态观测与状态之间的关系。KF(Kalman Filter)、EKF(Extended Kalman Filter)和UKF(Unscented Kalman Filter)滤波算法具有速度快、精度高、易于并行处理等优点,但需要建立运动人手的运动模型,这显然不是一件容易的事情,所以一般采用1阶或2阶动态模型。也就是说,速度和加速度比较平稳地改变成为大多数滤波跟踪算法的一个基本假设——这种假设与人机交互的和谐性和自然性的基本目标是相悖的,因为任意性和突变性是两手运动的两个基本特点。任意性是指两手运动过程不遵循事先设定的固定运动模式;突变性是指两手运动之平衡状态被瞬间打破。不仅如此,几乎所有的基于模型(Model-Based)的人手跟踪算法的跟踪精度都与两手运动模型存在着密切的联系^[12]。

综合上面四种情形,可以看到,目前大部分人手3D跟踪算法往往只考虑对于个别特性的鲁棒性——事实上基本侧重于跟踪算法对于肤色模型的鲁棒性,同时,忽略了研究四模型对跟踪算法时间复杂度的影响方式。把跟踪算法对四模型的多鲁棒性和实时性集成性地、系统性地进行研究,既是基于运动人手的HCI理论和应用的基本要求,又是基于模型的人手3D跟踪算法研究的必然过程。

(2) 实时性

因为在基于模型的运动人手跟踪研究中固有的高维空间特性和状态模糊特性,使得跟踪速度和跟踪精度成为两个最突出的问题,一般采用的解决途径是:(Ⅰ)基于优化搜索范围技术;(Ⅱ)基于分析—合成(Analysis-by-Synthesis)技术;(Ⅲ)基于Monte Carlo技术。

①基于优化搜索范围技术

基于模型3D约束条件,参考被跟踪特征点的历史状态信息,缩小模型参数的搜索范围是降低计算代价的一种基本方法。

Edward Lin等人不仅对手的方向加以限制(大拇指始终在手的一侧),使得手的方向始终从图像的右边延伸出来,而且对手掌的方向和指缝大小亦加以限制,以降低搜索空间^[5]。在H. Zhou和T. Huang的研究中,限制手掌上必须至少有4个特征点存在可靠的跟踪结果,不然无法获取人手的全局运动^[11]。S. U. Lee and I. Cohen提出了一种基于3D链接手模(Articulated Model)与2D被检测手轮廓的匹配,在满足手指内约束(Intra-Finger Constraints)、手指间约束(Inter-Finger Constraints)以及整体运动约束(Global Motion Constraints)的条件下,可以对手进行有效的搜索和跟踪,计算时间开销比较小,

但它要求在运动过程中大拇指根、食指根和小拇指根必须可见^[12]。为了解决光照条件变化带来的问题, Amit Kale 等首先对手所产生的阴影进行跟踪, 使用多目相机的投影约束条件, 只需估计帧中投影的外极线, 从而缩小人手状态空间的搜索范围^[13]。

基于优化搜索范围技术在一定程度上优化了算法的处理时间, 但是, 它不仅与实时性跟踪目标差距甚远, 而且往往以牺牲跟踪算法对四模型的鲁棒性为代价。

②基于分析—合成技术

通过合成人手的 3D 模型进行手势分析, 然后调整模型参数, 对模型参数逐步求精, 直到 3D 模型和真实人手具有相同的可视化图像为止, 这是降低计算代价的另一种典型方法。

在 Nobutaka Shimada 等的研究中, 首先搜索可能的候选手势, 根据轮廓特征和运动特征, 用 EKF 滤波技术进行预测, 然后根据先验概率大小确定初步手势, 最后, 通过计算满足约束条件和图像观测的概率分布, 对初步手势进行求精。后来, 他们根据索引图像库, 找到粗略的 3D 手势模型; 为了进一步精化模型, 需要提取 3D 手势特征并建立它与图像特征之间的对应关系, 用滤波预测技术得到准确的手势^[14]。为了降低问题的复杂度, “分而治之”(Divide-and-Conquer) 的研究思想也被用到人手跟踪算法的研究中。S. Lu 等将手势跟踪问题分解为两个子问题: 跟踪手掌和跟踪手指。在自定义目标函数的基础上, 分别用 LMS (Least Median Squares) 和 Inverse Kinematics 对子问题求解^[15]。

基于分析—合成技术存在的主要问题是, 不容易定义驱动求精进程的目标函数; 并且, 在求精阶段, 又重新归属于高维空间搜索问题。

③基于 Monte Carlo 技术

从状态 (State) 观点出发, 建立系统的运动模型和观测模型, 用概率方法实现状态预测和刷新。概率方法是系统预测理论中的一种强有力的数学工具, 它包括 Bayesian 滤波 (BF)、Particle 滤波 (PF) 和 KF 等。

PF 作为 Bayesian 滤波的一种实现形式, 逐渐成为一种有效解决非线性、非高斯问题的主流数学工具之一。需要采样的粒子数目过大, 是把 PF 算法用于人手 3D 跟踪必须解决的一个关键问题。2000 年, J. Deutscher 等从减少采样数目入手, 提出一种基于模拟退火算法的 PF 算法, 与以前算法相比较, 效率可以提高 10 倍; 因为它可以找到全局优化值, 故定位精度也有所提高^[16]。2001 年, Y. Wu 等利用一组少数且特定的图像为基, 并以其建构一个低维的人手空间 (Manifold), 从而将高维的手势跟踪问题转换到低维空间处理, 采用 PF 估计手势, 但一般需要采样 100 个样本^[17]。2003 年, 在 Hanning Zhou 等的工作中^[18], 采用 Bayesian 框架, 提出了一种分层析因算法 (Hierarchical Factorization Method), 把手分解成链接的物体, 对它们进行分层跟踪, 然后得到手势形状。2004 年, Hang Zhou 等用 Bayesian 假设进行跟踪, 用 Poisson 分布建立观测模型。为了降低手势空间维数, John Y. 提出一种非线性空间的表示方法, 将鱼贯 Monte Carlo 技术与 Nelder—Mead 单纯形搜索法相结合^[19]。M. Bray 等把 Stochastic Meta—Descent optimization (SMD) 方法整合到 PF 中, 以求得较好的 particles, 其速度有所改善^[20]。同时, 他们进一步将 SMD 方法与 PF 滤波相结合, 以较少的样本跟踪高维链结构^[21]。尽管如此, 粒子

数目依然是影响跟踪算法时间开销之主要因素。

最近, 鉴于现有的状态估计方法很难利用表观信息改进跟踪性能, Wen-Yan Chang 等提出了面向表观的粒子滤波算法^[22], 它集成连续运动变换信息和表观信息, 从贝叶斯公式中推出一个概率扩散模型, 并用鱼贯 Monte Carlo 方法实现这种模型。实验表明, 这种算法比单纯用其中一种方法的效果要好。

从速度来看, Hanning Zhou 等人所提算法^[18]全局跟踪速度达到 30 帧/秒, 如果结合局部跟踪, 则速度降为 8 帧/秒 (Pentium3, 1.0GHz 处理器); Stenger 在其博士论文研究中^[23]用 UKF 滤波器进行跟踪, 取得了 3 帧/秒 (背景为黑色, Celeron CPU, 433MHz) 的跟踪速度。一般来说, PF 的速度比 UKF 的速度要低得多。

纵观上面三种基本研究技术, 前两种技术存在的共同问题是, 没有从理论上解决实时获取全局最优搜索解之方法, 后一种技术的主要弊端是计算量太大, 也就是说, 达到实时性 3D 跟踪的目标还是一个极具挑战性课题; 同时, 上面三种基本研究方法把重点放在跟踪算法本身的设计上, 对影响跟踪算法实时性的基本因素缺乏系统性分析, 尤其是对四模型与跟踪算法之间关系的研究还很少。

(3) 评价体系

基于可视化分析方法的自我评价方法是目前普遍采用的一种方法: 把跟踪得到的 3D 人手直接投影到相应的帧图像上, 通过计算帧图像特征和投影图像特征之间的重合度来评价跟踪的精度^[12]。另一种方法是把跟踪结果与实打实的数据 (Ground Truth Data) 作为比较的标准, 计算跟踪结果的精度。但是, 这种方法往往需要借助数字手套或特殊传感器才能完成, 有人用手工方法获取这种实打实的数据^[24], 存在的问题是精度低, 且极易受到数据采集者主观性的影响。最近, Cristina Manresa 等通过若干试用者的使用情况进行评价^[25]。

设计简单易行、可以对跟踪算法设计起积极推动作用的评价体系是一个亟待解决的问题。

综上所述, 多鲁棒性和实时性分别决定跟踪系统的精度和速度这两个关键性能, 对其进行融合性研究, 既是基于模型的人手 3D 跟踪算法研究的必然过程, 又是基于运动人手的 HCI 理论和应用研究中的基础性、关键性和挑战性课题, 且这种研究涉及计算机图形图像处理、人机交互、计算机视觉、射影几何学、预测估计方法学、软计算理论以及神经生理学等众多学科, 满足我国信息化建设和中长期科学和技术发展规划之需要; 结合已有的工作基础, 深入研究基于计算机视觉的运动人手 3D 跟踪中的多鲁棒性和实时性这两个突出问题和主要矛盾, 对于本领域其他关联问题的解决、推动基于自然手的 HCI 系统在智能感知技术领域以及 VR 领域的理论和应用研究都具有重要意义, 对于研究和实现自然、方便、和谐的虚拟博物馆人机交互界面系统更具有现实意义——因为本项目的基本应用背景正是虚拟博物馆的研究与设计。

参 考 文 献

- [1] Ali Erol George Bebis; Mircea Nicolescu; Richard D. Boyle; Xander Twombly. A Review on Vision—Based Full DOF Hand Motion Estimation. IEEE Computer Society Conference on Computer Vision And Pattern Recognition (CVPR'2005). San Diego, CA, USA, 2005, pp. 15-22.
- [2] 王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述 [J]. 计算机学报, 2002 (3): 226-237.
- [3] Guangqi Ye, Jason J. Corso, and Gregory D. Hager. Gesture Recognition Using 3D Appearance and Motion Features. In Proceedings of CVPRHCI (Vision and Pattern Recognition for Human Computer Interaction), Hong Kong, 2004, vol. 10, pp. 60-170.
- [4] 董士海. 人机交互的进展及面临的挑战 [J]. 计算机辅助设计与图形学学报, 2004 (1): 1-13.
- [5] Edward Lin, Andy Cassidy, Dan Hook, Avinash Baliga, Tsuhan Chen. Hand Tracking Using Spatial Gesture Modeling and Visual Feedback for a Virtual DJ System. Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02), NY, USA, 2002, pp. 97-202.
- [6] 陈锻生, 刘政凯. 肤色检测技术综述 [J]. 计算机学报, 2006 (2): 196-207.
- [7] Hang Zhou; Qiuqi Ruan. A Method of Hand Tracking Based on Articulated Model. The 6th International Conference on Signal Processing Proceedings. Beijing, China, 2002, vol. 2, pp. 961-964.
- [8] J. J. Kuch and T. S. Huang. Human Computer Interaction Via The Human Hand: A Hand Model. In Twenty—Eighty Asilomar Conference on Signal, Systems, and Computers, Pacific Grove, CA, USA, 1995, pp. 1252-1256.
- [9] J. M. Rehg and T. Kanade. Model—based Tracking of Self—occluding Articulated Objects. In International Conference on Computer Vision (ICCV), Boston, 1995, MA, pp. 612-617.
- [10] Lathuiliere, F.; Herve, J. — Y. Visual Tracking of Hand Posture with Occlusion Handling. Proceedings of 15th International Conference on Pattern Recognition (ICPR00), Quebec, Canada, 2000, vol. 3, pp. 1129-1133.
- [11] H. Zhou, T. Huang. Recovering Articulated Motion with A Hierarchical Factorization Method. in Proc. of 5th Intl. Workshop on Gesture and Sign Language based Human—Computer Interaction, Genova, Italy, 2003, pp. 140-151.
- [12] S. U. Lee and I. Cohen. 3D Hand Reconstruction From Monocular View. International Conference on Pattern Recognition [C]. Proceedings of the 17th International Conference on Pattern Recognition (ICPR' 04). Cambridge, UK, 2004, vol. 3, pp. 310-313.
- [13] Amit Kale, Kenneth Kwan, Christopher Jaynes. Epipolar Constrained User Pushbutton Selection In Projected Interfaces, Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Washington, DC, 2004, vol. 10, pp. 156-164.
- [14] Nobutaka Shimada, Kousuke Kimura, Yoshiaki Shirai, and Yoshinori Kuno. Hand Posture Estimation by Combining 2—D Appearance—based and 3D Model—based Approaches. Proceedings of 15th International Conference on Pattern Recognition, Barcelona, Spain, 2000, vol. 3, pp. 709-712.
- [15] S. Lu, D. Metaxas, D. Samaras, and J. Oliensis. Using Multiple Cues For Hand Tracking And Model Refinement. In Proc. Conf. Computer Vision and Pattern Recognition, Madison, WI, 2003, vol. 2, pp. 443-450.
- [16] J. Deutscher, A. Blake, and I. Reid. Articulated Body Motion Capture By Annealed Particle Filtering. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00), Hilton Head, USA

- , 2000, vol. 2, pp. 1144-1149.
- [17] Y. Wu, J. Y. Lin, and T. S. Huang. Capturing Natural Hand Articulation. Proc. IEEE Int. Conf. Computer Vision, Vancouver, Canada, 2001, vol. 2, pp. 426-432.
- [18] Hanning Zhou and Thomas S. Huang. A Bayesian Framework For Real-time 3D Hand Tracking In Highly Cluttered Background. in Proc. of 10th Intl. Conf. on Human-Computer Interaction, Crete, Greece, June 22-27, 2003, vol. 2, pp. 188-190.
- [19] John Y. Lin, Ying Wu, Thomas S. Huang. 3D Model-Based Hand Tracking Using Stochastic Direct Search Method. Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04), Seoul, Korea, 2004, pp. 693-698.
- [20] M. Bray, E. Koller-Meier, P. Muller, L. Van Gool, and N. N. Schraudolph. 3D Hand Tracking By Rapid Stochastic Gradient Descent Using A Skinning Model. In 1st European Conference on Visual Media Production (CVMP), London, 2004, pp. 231-237.
- [21] M. Bray, E. Koller-Meier, and L. Van Gool, Smart Particle Filtering For 3D Hand Tracking, Proc. IEEE Int. Conf. Automatic. Face & Gesture Recognition, Seoul, Korea, 2004, pp. 675-680.
- [22] Wen-Yan Chang, Chu-Song Chen, and Yi-Ping Hung. Appearance-Guided Particle Filtering for Articulated Hand Tracking. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2005, San Diego, CA, USA, 2005, vol. 1, pp. 235-242.
- [23] Bjorn Dietmar Rafael Stenger. Model-Based Hand Tracking Using A Hierarchical Bayesian Filter. Ph. D. Thesis. March 2004, University of Cambridge, St. John's College.
- [24] Eric Foxlin and Michael Harrington. Weartrack: A Self-referenced Head And Hand Tracker For Wearable Computers And Portable VR. In International Symposium on Wearable Computing, Atlanta, GA, 2000, pp. 155-162.
- [25] Cristina Manresa, Javier Varona, Ramon Mas and Francisco J. Perales. Hand Tracking and Gesture Recognition For Human-Computer Interaction. Electronic Letters on Computer Vision and Image Analysis, 2005, vol. 5 (3): pp. 96-104.

2. 项目的研究内容、研究目标以及拟解决的关键问题

研究目标

提出基于计算机视觉的运动人手的多鲁棒性和实时性 3D 跟踪的新理论和新方法是本项目的主要研究目标。多鲁棒性强调系统精度的总体优化, 即跟踪系统对于人手肤色模型、3D 模型、自遮挡模型以及运动模型这四模型均具有鲁棒性。实时性则强调系统总体的时间复杂度要控制在用户可以容忍的响应范围内。

研究内容

(1) 研究四模型。研究对于光照条件和复杂背景均具有鲁棒性和实时性的肤色模型, 包括静态模型和动态模型, 研究肤色反射模型和皮肤的波谱特性, 在此基础上进一步研究皮肤的物理检测新方法; 研究人手 3D 模型与算法精度之间的关系, 尤其是探讨二者独立性条件, 探讨或者降低算法精度对手模型的依赖程度、或者降低人手 3D 模型对算法精度的影响程度之方法; 研究满足跟踪算法鲁棒性和实时性需要的自遮挡模型, 根据这种模型易于最大限度地恢复被遮挡特征点; 研究跟踪算法可以正确恢复被遮挡特征点的充分必要条件; 研究满足跟踪算法鲁棒性和实时性需要的运动模型, 使得该模型既能描述状态的

连续变化,又能描述状态的突变现象。

(2) 集成性地、系统性地研究跟踪算法对于人手四模型的多鲁棒性和实时性,重点揭示在跟踪算法中速度和精度之间对立统一之形式。

(3) 在满足对手进行实时性 3D 跟踪的条件下,确定粒子滤波器中采样粒子个数的上确界和下确界以及取得这种边界值的基本条件;反过来,在采样粒子数小于某给定值(例如,上面所述的下确界)的前提下,进一步研究降低跟踪算法时间复杂度的理论和方法。

(4) 研究跟踪算法,研究 Bayesian 滤波器和粒子滤波器中针对运动人手进行采样、获取粒子权重以及后验概率分布的新方法;重点研究降低算法时间复杂度的理论和方法。

(5) 研究认知科学和神经生理学的成果尤其是达尔文自动机和运动神经内模理论,从神经生理学研究人手运动的精确运动模型和预测模型,仿真神经系统控制人手运动的过程和方式。

(6) 一方面,重点研究实打实数据(Ground Truth Data)新的获取方法;另一方面,研究获取实打实数据的替代方法——容许对三维数据按比例进行缩放。最后,构建跟踪算法评价标准。

拟解决的关键问题:

(1) 研究和设计四模型。四模型是基于模型的人手跟踪算法中的四个基本要素,不仅影响跟踪结果的多鲁棒性,而且影响跟踪结果的实时性,还影响到跟踪算法本身的设计与实现。

(2) 研究跟踪算法,重点研究降低跟踪算法时间复杂度的理论和方法。跟踪算法的研究与设计是本项目的核心内容,降低跟踪算法时间复杂度是实现实时性跟踪目标的关键途径。

(3) 研究多鲁棒性、实时性、四模型和跟踪算法之间融合以及进一步优化这种融合的途径和方式。只有将这几个方面有机地融合起来、协调和统一起来,才能最终实现本项目的研究目标。

(4) 研究准确、快速获取被跟踪特征点状态观测值的理论和方法。状态观测值是人手跟踪算法的基础性输入数据,直接影响跟踪算法的成败或性能。

3. 拟采取的研究方案及可行性分析

研究方法

首先,进行理论调研与研究;调研分析国内外同类研究的最新成果。

高度模型化、层次化和系统化是本项目的一个基本研究方法。从系统的角度,剖析影响跟踪算法多鲁棒性和实时性的关键因素,然后,用系统化理论和方法,根据功能性原则建立所有因素的层次化结构,再优化这种结构及其关系。图 1-1 中进一步诠释这种基本思想。系统由动态数据结构、算法和性能三个层次构成,通过选择合适的滤波器(例如粒子滤波器),可以把图 1-1 中的边 e 去掉,从而可以进一步优化层次关系。所谓动态数据结构是指数据结构具有一定的可形变特征。我们更强调动态数据结构对跟踪算法实时性的影响,尽管同一个算法可以设计不同的数据结构。

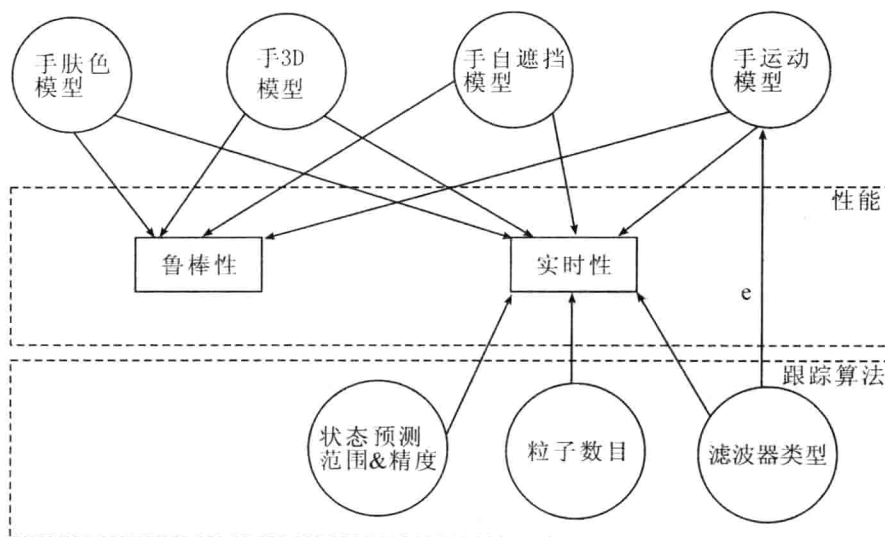


图 1-1 影响跟踪算法多鲁棒性和实时性的关键因素

其次，融合多种信息，重新组合信息可能使问题发生质的变化，在这种变化过程中，找到解决问题的突破口；采用交叉式和立体式研究策略，融合机器学习、人工神经网络等软计算方法、Monte Carlo 技术等多种研究手段，结合多种研究工具尤其是跨学科理论（例如运动神经内模理论），从新的角度找到问题的研究点或突破口。

综合分析相关领域的主要研究理论和方法，分析它们的时间复杂度、跟踪精度、鲁棒性以及适用条件，重点关注主流研究理论和方法，探索、引进先进的数学工具和先进的研究理论。

把信息可视化理论和方法引入本系统的研究将推动本项目的研究。

技术路线

(1) 四模型的设计

研究颜色现象的生物学本质，在颜色空间、色度测量以及颜色恒常性等理论的基础上，建立肤色模型。重点研究新的颜色空间和颜色向量的量化，研究相似性度量尤其是颜色分类器的设计。离开肤色建模方法和分类器原理与实现，探讨所谓的肤色检测最佳颜色空间就没有太大意义。结合空间信息的肤色检测，将肤色特征与空间分布有关的其他信息，如梯度、形状或纹理特征等相结合。

研究生理模型、带纹理的体模型、网格模型、骨架模型、轮廓模型和边界模型等几种基本模型中的不变量，重点研究骨架模型，研究上述几种模型对算法鲁棒性和实时性的影响方式，进而建立标量化的可形变 3D 骨架模型。

对被跟踪特征点进行分类，剖析各类特征点在自遮挡问题中的作用，分别建立每类特征点的数学模型，研究不同类特征点数学模型之间的关系，进一步建立人手自遮挡模型。

一方面，以现有的认知科学和神经生理学研究成果，特别是以达尔文自动机和运动神经内模理论为基础，建立具有神经生理学和认知科学依据的人手运动模型，通过计算机模拟或仿真，理解人手运动系统运动控制的形成和发展，在此基础上，设计人手运动模型及

其运动预测方式；另一方面，如果选择对手运动模型具有天然鲁棒性的滤波器，则可以大大简化甚至无须人手运动模型的设计。

(2) 跟踪算法设计尤其是降低跟踪算法时间复杂度的理论和方法

基于贝叶斯原理的 Particle 滤波技术通过非参数化的蒙特卡罗模拟方法来实现递推贝叶斯滤波，适用于任何能用状态空间模型以及传统的卡尔曼滤波表示的非线性系统，精度可以逼近最优估计，所以将其作为设计跟踪算法的首选数学工具，同时，融合 Kalman 类滤波、分析—合成技术等其他跟踪方法。

一方面，在限制跟踪算法的时间复杂度的条件下，确定采样粒子数的上确界和下确界，同时，进一步研究取得这种边界值的基本条件；反过来，在采样粒子数小于某给定值（例如，上面所述的下确界）的前提下，进一步研究降低跟踪算法时间复杂度的理论和方法。

另一方面，融合上一帧的跟踪结果信息、人手 3D 模型、人手 3D 模型的约束条件以及各类特征点的数学模型等多线索到 PF 算法中，去除无效样本点。

同时，把人手的运动过程分为连续形变和非连续形变过程，再分别研究其采样的特点和规律，研究减少采样粒子数的途径和方法。

(3) 四模型、多鲁棒性、实时性和跟踪算法之间融合的途径与方式

一方面，对手四模型分别进行分析和研究，确定每个模型满足实时性跟踪的解空间，然后把这四个解空间求交集运算，得到满足实时性的动态数据结构（称为实时数据结构）；用类似的方法得到满足多鲁棒性需要的动态数据结构（称为鲁棒数据结构）；然后，或者通过实时数据结构和鲁棒数据结构抽象出上位动态数据结构，或者通过寻找实时数据结构和鲁棒数据结构的平衡点得到上位动态数据结构；或者通过寻找实时数据结构和鲁棒数据结构的交点得到上位动态数据结构。显然，这种上位动态数据结构均满足跟踪算法的多鲁棒性和实时性之性能要求。

另一方面，研究四模型与实时性之间的独立性条件，研究把四模型与实时性之间的关联关系降到最低的途径；同时，研究四模型与鲁棒性之间的独立性条件，研究把四模型与鲁棒性之间的关联关系降到最低的途径。

(4) 获取观测

一方面，把关节作为被跟踪特征点，结合多尺度（Multi-Scale）理论和距离地图（Distance Map）研究准确、快速地从帧图像中提取关节状态观测值的方法。

另一方面，采用多相机硬件平台也有助于准确获取观测值。另一种折中的方法是单目体系中加入一个镜面，利用镜面对称成像理论，重点研究遮挡存在条件下获取观测值的基本方法。

同时，用不变量理论研究确定观测值的必要条件，研究由已知量推测未知量测的产生式规则。

另外，用层次化结构表示帧图像信息（例如，可以分为低层信息和高层信息等），研究各层的表示模型、提取方法以及通信方式。

可行性研究

本研究采用的技术路线是可行的。国内外相同或相近领域的研究成果和理论,尤其是基于 Bayesian 滤波器和 Particle 滤波器的人手跟踪理论和实践为我们奠定了良好的研究基础,使得本项目具有较高的研究起点和广阔的研究平台。

融合上一帧的跟踪结果信息、手势 3D 模型、3D 人手模型的约束条件以及各类特征点的数学模型等多线索,必然影响粒子的采样范围和采样方式,由此,可以研究粒子数目变化范围界限的存在性,采用线性的或非线性的方法,探索满足实时性条件下采样粒子数的上确界和下确界。其他的一些技术,比如,其他几个模型的设计以及算法评价体系,我们已经进行部分预研,对本项目的研究积累了一定的理论基础和实践经验。

另外,本课题组已经拥有本项目所需要的设备,具有很好的科研环境,而且课题组成员年富力强,具有很好的协作精神和团队精神,大部分成员曾参与山东省自然科学基金的研究,具有丰富的科研经验,有能力、有信心按时完成本项目。

4. 本项目的特色与创新之处

(1) 提出多鲁棒性的基本概念及其相关理论,集成性地、系统性地研究跟踪算法对于人手四模型的多鲁棒性和实时性,深刻揭示在跟踪算法中速度和精度之间对立统一之形式。

(2) 以系统化、层次化和模型化为基本研究方法;把获取满足多鲁棒性和实时性需要的人手四模型的解空间及其边界条件作为解决本项目关键问题的突破口。

(3) 融合历史帧跟踪结果信息、人手 3D 模型、模型的约束条件以及各类特征点的数学模型等多线索,在限制跟踪算法的时间复杂度的条件下,确定采样粒子数的上确界和下确界,同时,进一步确定取得这种边界值的基本条件;反过来,在采样粒子数小于某给定值的前提下,进一步研究降低跟踪算法时间复杂度的理论和方法。

(4) 限制系统瓶颈,追求整体优化,研究多鲁棒性、实时性、四模型和跟踪算法之间融合的途径和方式,把它们有机地统一于本项目的总体研究目标之中。

5. 年度研究计划及预期研究结果

(1) 年度研究计划

2007.1~2007.3

调研分析,检索相关文献,提出系统总体设计方案,并请专家组论证方案的可行性。

2007.4~2007.10

研究和设计四模型。

2007.11~2008.7

研究跟踪算法,重点研究降低跟踪算法时间复杂度的理论和方法。

2008.8~2009.7

研究鲁棒性、实时性、四模型和跟踪算法之间融合以及进一步优化这种融合的途径和方式。

2009.8~2009.10