ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, No.10, October 2007, pp.2423−2433
DOI: 10.1360/jos182423

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种面向实时交互的变形手势跟踪方法*

王西颖+, 张习文, 戴国忠

(中国科学院 软件研究所 人机交互技术与智能信息处理实验室,北京 100080)

An Approach to Tracking Deformable Hand Gesture for Real-Time Interaction

WANG Xi-Ying+, ZHANG Xi-Wen, DAI Guo-Zhong

(Laboratory of Human-Computer Interaction and Intelligent Information Processing, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-10-62540451, E-mail: wangxy0823@sohu.com, http://iel.iscas.ac.cn/

Wang XY, Zhang XW, Dai GZ. An approach to tracking deformable hand gesture for real-time interaction. *Journal of Software*, 2007,18(10):2423−2433. http://www.jos.org.cn/1000-9825/18/2423.htm

Abstract: The tracking of deformable hand gesture is a very important task in vision-based HCI (human-computer interaction) research. A novel real-time tracking approach is proposed to capture the motion of deformable hand gesture with single camera. The proposed approach uses a set of 2D hand models in place of high-dimensional 3D model. It achieves auto-initialization by firstly using Bayesian classifier to do posture recognition, and then locating fingers and fingertips to fit image features to recognized posture. It solves the problem of interference among fingers during tracking successfully by the integration of *K*-means clustering and particle filter. Moreover, a state checking process is embedded into tracking method, and it realizes resumption from tracking failure and update of hand models automatically. Experimental results show that the proposed method can achieve continuous real-time tracking of deformable hand gesture with high precision, and thus it can meet the requirements from real-time vision-based human-computer interaction.

Key words: hand tracking; multifinger tracking; particle filter; auto-initialization; state checking

摘 要: 变形手势跟踪是基于视觉的人机交互研究中的一项重要内容.单摄像头条件下,提出一种新颖的变形手势实时跟踪方法.利用一组2D手势模型替代高维度的3D手模型.首先利用贝叶斯分类器对静态手势进行识别,然后对图像进行手指和指尖定位,通过将图像特征与识别结果进行匹配,实现了跟踪过程的自动初始化.提出将 *K*-means 聚类算法与粒子滤波相结合,用于解决多手指跟踪问题中手指互相干扰的问题.跟踪过程中进行跟踪状态检测,实现了自动恢复跟踪及手势模型更新.实验结果表明,该方法可以实现对变形手势快速、准确的连续跟踪,能够满足基于视觉的实时人机交互的要求.

关键词: 手势跟踪;多手指跟踪;粒子滤波;自动初始化;状态检验

中图法分类号: TP391 **文献标识码**: A

手势交互是人机交互领域近年来的一个研究热点[1−5].手势作为一种输入方式包含了丰富的信息,但由于

* Supported by the National Basic Research Program of China under Grant No.2002CB312103 (国家重点基础研究发展计划(973))
Received 2006-07-13; Accepted 2006-08-16

人手是多关节非刚性物体,手掌和手指的状态在运动中不断发生变化,要获得不断运动变化的手势信息并非易事.数据手套和传感器的方式(glove-based)是通过附着在人手上的特殊设备获得手的运动和形状信息,但同时也带来了对人手运动的约束,所以,这种方式并非是一种自然的交互方式.随着计算机视觉技术的发展,利用摄像头可以实现对手势信息的非接触性捕获,对于使用者来说,这是一种更加自然和符合人自身行为习惯的交互方式.手势的形态在交互过程中是不断发生变化的,对变形手势进行视觉跟踪是实现动态手势识别和理解的关键,也是计算机视觉和人机交互领域中一个富有挑战性的重要问题.其困难主要在于手势运动的自由度大(可高达 27 个自由度(degree of freedom,简称 DOF)[6]),跟踪特征向量的维度高,难以满足人机交互中要求实时反馈的要求,而且通过视觉技术获得 3D 模型的参数是一项非常困难和复杂的任务[7].此外,跟踪初始化问题一直依赖于人工标记[8,9],一旦跟踪失败,则无法实现跟踪的自动恢复,难以保证人机交互的连续性.

通常,对手或手指进行跟踪研究的方法大都假设手的形状在跟踪过程中保持不变[10−12]或者不考虑手指状态的变化[13,14],即便是针对多关节的变形手势,也往往是利用 3D 手模型进行跟踪[15,16],或采用两个或多个摄像头的方式[6],存在计算复杂、实时性差的问题.

上述的现有手势跟踪算法或者无法对变形手势进行有效跟踪,或者无法满足人机交互的实时性要求.本文在单摄像头条件下,针对变形手势提出一种无须人工干预的连续跟踪方法,使之满足人机交互中对实时性、准确性及连续性的要求.从仿生学的角度来看,生物视觉系统对目标的有效跟踪,通常都是建立在对目标对象理解的基础上,而不仅仅是依赖视觉信息.本文的跟踪方法就是建立在对跟踪目标识别和理解的基础上.首先,为了提高跟踪速度并便于进行手势理解,我们采用“分而治之”的策略,对高维度特征空间降维,利用一组低维度的 2D 模型替代高维度 3D 模型.然后通过图像分割和轮廓提取,将图像特征与识别后手势模型进行匹配以完成跟踪模型的自动初始化.跟踪过程中利用 Camshift 算法对整体手区域进行跟踪,将聚类算法与粒子滤波算法相结合,跟踪多个手指指尖.跟踪检测部分判断跟踪是否失败以及跟踪模板是否需要更新.一旦跟踪失败或需要进行模板更新,则利用前期跟踪过程积累的知识和当前图像特征,快速地自动恢复跟踪.

本文第 1 节介绍国内外的相关工作及本文方法的梗概.第 2 节详细介绍本文提出的跟踪方法,分别描述其 3 个组成部分:静态手势识别部分、图像与模型的匹配部分和手势跟踪部分.第 3 节是实验和结果分析.第 4 节是本文方法的总结.

1 相关工作

多关节非刚体运动的跟踪,国内外的研究人员已经提出了一些不同的解决方案.总的来说可分为基于模型(model-base)和非模型的两种方法[17−19].非模型的方法主要是基于表观(appearance-based)的方法,它又可分为基于区域的跟踪[20]、基于活动轮廓或变形模板的跟踪[21]以及基于其他图像特征的跟踪方法.实际上,完全基于图像表观特征的非模型跟踪方法很难保证跟踪的鲁棒性,因为它们所跟踪的特征,如兴趣点、轮廓线、2D 区域等,很容易受遮挡或光线条件变化等因素的影响.利用变形模板(deformable template,或 snake 模型)是近年来进行变形目标跟踪的常用办法,但同样容易受到周围复杂环境的影响,使模型收敛于非目标边缘,并且轮廓通常无法收敛到深度凹陷的区域.GVF Snake 模型[22]通过计算梯度矢量流解决深陷问题,但它需要事先求解一个偏微分方程组,极大地增加了计算量,同时在初始化时还存在一个“临界点”问题.基于模型的跟踪算法大多采用固定的跟踪对象模型,为了解决多关节对象产生的外观形变问题,研究者往往需要对所有的关节参数建立 3D 模型[15,16],在跟踪过程中,将 3D 模型投影到图像空间,并将投影与图像特征进行比对.由于模型的特征矢量维度很高,造成计算量大,而且在跟踪的过程中,由于错误的不断累积很容易导致跟踪失败.

对于跟踪初始化,以往的跟踪方法中往往是通过手工标出或假定初始位置及其参数符合平均概率分布[8,9].Cheng[23]的方法是搜索整个运动模型空间,通过计算代价函数的方法实现步态跟踪的初始化,计算量很大.Sminchisescu 和 Triggs[24]采用层次化的 3 个步骤来得到初始位置,但如何将 3D 人体模型与 2D 图像进行对应仍是一个难题.文献[25]的方法则是通过视频序列的前若干帧得到初始模型.

目前常用的多目标跟踪方法为多目标数据关联[26],典型的数据关联方法有最近邻方法、概率数据关联滤波

(probability data association filter,简称 PDA)、联合概率数据关联滤波(joint PDA)等.但数据关联本身是 NP-hard 问题,计算代价高而且主要用于单跟踪器同时跟踪多个目标的情况[26].此外,数据关联未能利用手的结构特征,即手掌与手指、手指与手指之间的关系特征.在 Letessier 和 Berard[27]的多手指的跟踪方法中,将当前帧中检测的手指位置与上一帧中距离其最近的手指相关联,这种方法虽然简单,但当手指运动幅度较大时,将会导致关联错误.Kenji Oka 和 Yoichi Sato 的方法[12]为了计算简单,默认手指间的顺序关系在跟踪过程中不会发生变化.

总的来说,在人机交互系统中,变形手势的跟踪算法应满足下面 3 个要求:

(1) 实时性好,避免高维度特征矢量的计算和复杂的搜索过程.

(2) 足够的鲁棒性.不受跟踪对象旋转、平移和比例改变以及摄像头视角改变的影响.

(3) 跟踪的连续性和自动初始化,能够在跟踪失败后自动恢复跟踪,尽量减少人为干预.

本文提出的跟踪方法是结合了基于模型方法与基于表观方法的特点,首先将高维度的多关节手势模型分解为若干低维度的 2D 模型,跟踪系统中包含一个检测机制,按一定规则触发识别过程,根据识别结果更新所采用的手势模型.通过识别,系统可以对当前的手势状态有一个基本的理解,并通过模型与图像特征的匹配自动建立 2D 跟踪模型,例如建立一个拇指和一个食指的 2D 模型.由于跟踪模型是建立在理解的基础上,在很大程度上将有助于解决光线变化以及视角改变等问题.针对变形手势跟踪,本文提出的连续跟踪方法主要包括 3 个部分:

(1) 静态手势识别部分.识别部分实现了对当前帧中手势姿态的理解.

(2) 手势图像与模型的匹配部分.利用第 1 部分的识别结果,包含手势的图像需要与 2D 模型进行匹配,得到跟踪所需要的特征矢量和初始参数.

(3) 跟踪部分.先进行手区域的粗定位,再确定手指指尖位置的变化,以及利用手的结构特征定位手掌位置.与传统的基于模型的跟踪过程不同,它还包括一个检测模块用以检测当前的 2D 手模型是否需要更新.一旦检测到模型需要更新或发生跟踪失败,则识别模块将被触发.3 个模块构成了整个系统的迭代结构,如图 1 所示.

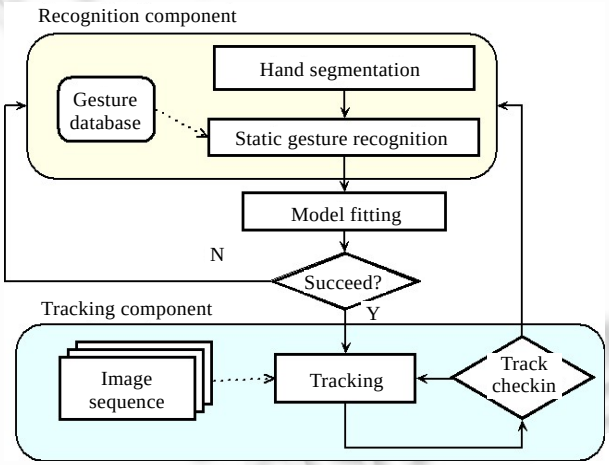


Fig.1 The tracking framework of deformable hand gesture

图 1 变形手势跟踪方法框架

2 手势跟踪框架

2.1 静态手势识别

通过静态手势的识别,使系统能够对被跟踪对象有一个基本的理解,为实现自动跟踪初始化与跟踪的自动恢复奠定了基础.首先,手部区域需要从场景中分割出来.我们采用了一种基于模糊集和模糊运算的方法进行手的区域和轮廓提取.具体分割算法参见文献[28].静态手势的识别是基于轮廓特征的识别,对手势轮廓按顺时针方向进行轮廓追踪,得到完整的轮廓边缘.图 2 中,左图为分割出的手部区域,右图为对应的手势轮廓.

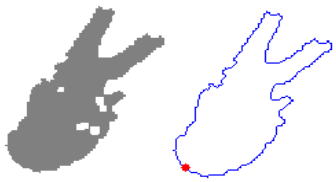


Fig.2 Extraction of hand region and its contour

图 2 手势区域与边缘的提取

接下来是提取手势轮廓的形状特征.傅里叶描述子(Fourier descriptor,简称 FD)是一种经典的形状描述方法[29].它首先对物体轮廓上的所有边缘像素(x_i,y_i)进行坐标变换,得到集合$\{p_i(x_i,y_i)|p_i=x_i,y_ij,i=0\ldots K,j^2=-1\}$,然后对其进行 Fourier 变换,用得到的系数 $FD_u(u=0,1,2,\ldots,K-1)$描述形状特征.按公式(1)可以构建具备平移、旋转与缩放不变性的轮廓特征向量 X:

$$X=\{|FD_2|/|FD_1|,|FD_3|/|FD_1|,\ldots,|FD_n|/|FD_1|\} \quad (1)$$

在本文的跟踪方法中,静态手势的确定是将连续手势状态中相似的手势类型归并为一类.这通常是参考具体应用的需要,并且考虑到遮挡对跟踪造成的影响,将能够连续跟踪的手势状态归为一类手势类型.本文按照手指的状态,将手指个数和手指类型(食指、中指等)相同的手势归为同一类手势.

静态手势分类采用的是贝叶斯分类器.贝叶斯分类器是基于贝叶斯决策理论的分类器,有着成熟而完善的数学基础,它通过选择最小化条件风险的分类方式来使预期的损失最小化,即为了最小化总风险,对所有的 $i=1,2,\ldots,m$,计算条件风险: $R(\alpha_i|X)=\sum_{j=1}^{c}\lambda(\alpha_i|\omega_j)p(\omega_j|X)$,其中,$\lambda(\alpha_i|\omega_j)$为风险函数,描述了类别状态为 ω_j 采取行动 α_i 的风险.使总风险最小的类别划分 ω_j 就是 X 所属的类.$p(\omega_j|X)$通常被认为符合高斯分布模型,这种模型运算简单,而且现实世界中的很多事件都与高斯分布有极大的相似性.高斯型贝叶斯分类器的模型参数(均值矢量与协方差矩阵)是通过对训练样本的学习得到的.

2.2 跟踪的自动初始化

跟踪的初始化是确定运动目标的最初状态,初始化效果的优劣直接影响到跟踪的成败.以前的跟踪方法往往是默认在跟踪的第 1 帧已经完成了初始化,或通过手工标定的方式确定第 1 帧中目标对象的状态[30].本文提出一种将图像特征与识别结果进行匹配,从而实现跟踪自动初始化的方法.识别的结果被转换为跟踪所使用的 2D 模型,我们采用了与文献[31]中类似的简化 2D 手模型,如图 3 所示.

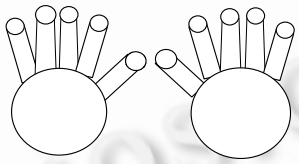


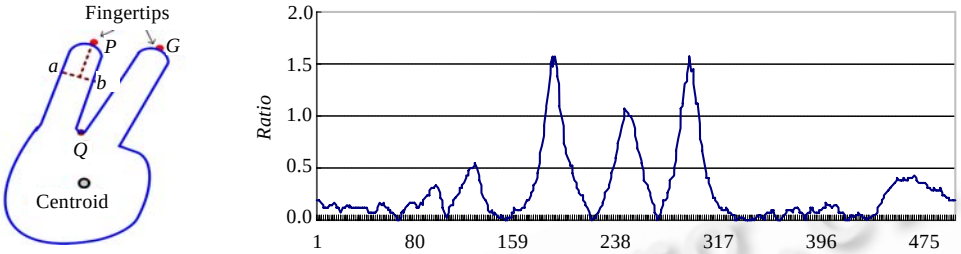
Fig.3 2D hand model

图 3 2D 手模型

识别的结果指出了当前手势的基本状态以及手指的个数、手指的类型等信息,通过对已识别的手势图像进行手指和指尖位置搜索,将图像与模板进行匹配.文献[10]中给出的手指指尖位置查找方法存在明显不足:需要预先设定的参数过多,而且当两个手指距离较近时会干扰指尖位置的查找.Kenji Oka 和 Yoichi Sato 的指尖搜索方法[12]是首先在一个较大的搜索窗口内扫描确定 20 个候选指尖位置,然后再对匹配度最大的候选位置周围的候选进行抑制,同时按一定规则去除位于手指中间的部分候选.该方法由于需要对搜索区域进行多次逐像素的扫描,造成计算量较大,而且除去手指中部候选位置的方法的鲁棒性也较差.本文提出一种根据手势轮廓的曲率进行指尖位置搜索的方法,通过对手势轮廓按轮廓点顺序进行定长扫描,将满足公式(2)的点设为指尖候选点:

$$ratio=D_p/D_{ab}\geq\varepsilon \quad (2)$$

其中,D_{ab}为扫描轮廓起始点 a 与终点 b 连线 AB 的长度,D_p为扫描轮廓中点 p 到 AB 的垂直距离,参照图 4(a),ε为比值 *ratio* 的最小阈值,比值 *ratio* 大于ε的情况下,p 点被设置为指尖候选点.对于图 2 中得到的轮廓,从其右图上的起始点(实心圆点)开始按顺时针方向进行扫描,计算各个轮廓点处的 *ratio* 值.如图 4(b)曲线所示,X 轴为轮廓点索引值,Y 轴为各轮廓点处 *ratio* 值,曲线中可以明显看到 3 个波峰(设ε=1),分别对应图 4(a)示意图中的 P,G 和 Q 这 3 个候选点.候选点中又有轮廓凹顶点(Q 点)与凸顶点(P 点、G 点)的区分,参考两个端点连线 AB 的中心点,考察它的某矩形邻域范围内是否为肤色区域,若为肤色区域,则判定为指尖点,否则为非指尖点.



(a) The map of fingertip searching
(a) 指尖搜索示意图

(b) The *ratio* curve of the contour in Fig.2
(b) 图 2 轮廓的 *ratio* 值曲线

Fig.4 Location of fingertips

图 4 指尖定位

整个手指的分割是在找到指尖位置的基础上,分别从 a 点出发向前和从 b 点出发向后进行等步长搜索,设新的端点分别为 a'和 b'点,如果仍然满足公式(2),则继续向前、向后搜索,直到不满足公式条件为止.最终的 a'和 b'点则标识出手指的根部.

如果经过搜索匹配的手势 2D 模型与识别结果一致,即手指个数一致、手指指尖位置符合手势的形状特征,则开始进行跟踪,否则继续对下一帧进行识别和匹配,直至它们的结果一致.通过这种先识别后匹配然后再跟踪的方式进行跟踪初始化,保证了初始化的正确和效率,为下一步的跟踪奠定了良好的基础.

2.3 手部跟踪

对于手部跟踪,我们设计了 CAMshift(continuously adaptive mean shift)[32]与粒子滤波[33]相结合的跟踪算法,它综合利用了手势图像的颜色、区域和轮廓特征.CAMshift 是 Meanshift 算法的推广,是一种有效的统计迭代算法,它使目标点能够“漂移”到密度函数的局部极大值点.CAMshift 跟踪算法是基于颜色概率模型的跟踪方法,在建立被跟踪目标的颜色直方图模型后,可将视频图像转化为颜色概率分布图,每一帧图像中搜索窗口的位置和尺寸将会被更新,使其能够定位跟踪目标的中心和大小.本文中,CAMshift 算法被用于手势位置的粗定位,即确定当前手势区域的外包矩形 R_h.

粒子滤波也称为条件概率密度传播算法(conditional density propagation,简称 CONDENSATION),它是通过非参数化的序列蒙特卡罗方法(sequential Monte Carlo method)实现递推的贝叶斯滤波,用加权样本(粒子)的形式而不是函数的形式对先验和后验概率进行描述.在测量的基础上,通过调节各粒子的权重大小和样本的位置来近似实际概率分布,并以样本的期望作为系统的估计值.与卡尔曼滤波(Kalman filter)相比,粒子滤波能够处理非线性、非高斯分布的情况,而卡尔曼滤波只能解决线性、高斯的估计问题[34].目前,粒子滤波已被应用到目标跟踪及导航、参数估计和目标辨识等领域.变形手势的跟踪问题属于非线性、非高斯问题,所以,利用粒子滤波作为指尖位置跟踪器.在手区域外包矩形 R_h 范围内,基于粒子滤波的多指尖跟踪算法被用来确定多个手指指尖的位置,最后根据手掌的形状特征以及手掌与指尖的位置关系确定手掌位置.

指尖部位的跟踪特征采用了轮廓特征,为了降低计算量,我们利用半圆弧对指尖轮廓进行拟合,而不是经常被使用的 B 样条曲线[33].跟踪的特征向量为 $f=\{x,y,r,\alpha\}$,其中,x,y 为半圆弧所对应圆的中心点坐标,r 为圆弧半径,

α为半圆弧的直径与 X 轴正向的夹角,$-\pi/2<\alpha\leq\pi/2$.

在指尖跟踪过程中,为了计算各个粒子估计状态的置信度,我们从估计位置的半圆圆心按一定角度β进行径向方向扫描,如图 5 所示,p 点为某径向方向与估计位置半圆弧的交点,q 点则为图像中该径向方向灰度梯度变化大于某阈值的点,粒子置信度计算公式如下:

$$p(x=s_i\mid Z)=\exp-\left(\delta\times\sum_{n=1}^{N}Dis(p_n,q_n)\right) \tag{3}$$

其中,$N=\lfloor 180/\beta\rfloor$,$\delta$为常系数.图 5 右图中,指尖部位下方圆弧的点为估计位置圆弧上的 p 点,上方圆弧的点为对应的 q 点.

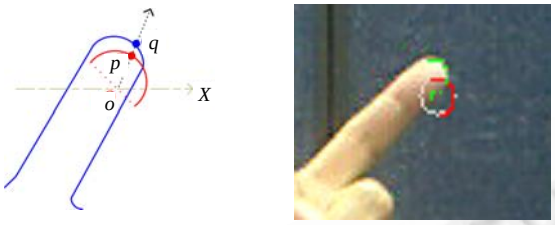


Fig.5 The map of shape estimation of fingertip and its practical effect

图 5 指尖轮廓估计示意图和实际效果图

为了解决多手指在粒子传播过程中互相干扰的问题,我们提出了基于聚类的多目标跟踪算法,这是一种基于聚类的多个粒子滤波器的跟踪方法,粒子滤波中粒子状态的传播模型 $X_{t+1}=AX_t+BW_t$,传播后的粒子通过观测图像计算估计值的后验概率密度,作为其置信度 $\pi_t^{(n)}=p(z_t\mid x_t=s_t^{(n)})$,再通过置信度得到粒子的权重.由于图像中手指与手指的距离很近,往往会影响后验概率密度的计算,即观测特征 Z_t 不能保证是当前对象的特征.我们采用 K-means 聚类辅助多目标粒子跟踪的方法,首先对全体粒子单独进行随机传播,传播过程符合二阶马尔可夫随机过程的特征.对所有粒子计算其置信度 $p(z|x=s_i)$,对于置信度大于某阈值的粒子,我们称其为强粒子(strong particles).对所有强粒子,利用识别阶段给出的手指个数作为聚类的类别个数,采用 K-means 迭代算法对其进行聚类.根据聚类结果为所有强粒子标注类别标签 C_i,如果 C_i 与强粒子所属的手指对象 f_i 不相符合,则认定其为可疑粒子.

对聚类结果进行的标注是标识聚类与手指对象的对应关系,标注是根据当前手势类型和上阶段各个手指的相对位置关系进行的.由于被跟踪的目标是经过识别的手势对象,而且在跟踪过程中其结构不会发生大的改变,即不会出现跟踪目标消失或出现新的跟踪目标的情况,我们根据手势结构特征的约束关系可以很容易地确定聚类与手指对象的对应关系.

最后,粒子权重的计算公式为 $\pi_t=\mu\times p(x_t=s_t|z_t)$,其中,$\mu$为信任系数,可疑粒子取值为 0.2,其他粒子为 1.粒子聚类示意图如图 6 所示,上方椭圆内的绿色点、中间椭圆内的红色点和下方椭圆内的浅蓝色点标识出 3 个指尖跟踪器所对应的强粒子簇.图中箭头所指向的被聚类到中间椭圆所标识的 *ClusterB* 中的两个点即为可疑粒子.通过聚类,可疑粒子被筛选出来,并且在粒子重取样的过程中,这些可疑粒子被复制的概率将大为降低,它们对跟踪过程所产生的干扰也将大为降低.

基于聚类的多目标粒子跟踪算法流程如下:

Step 1. 按二阶马尔可夫过程对全体粒子进行传播.根据 Random Walk 假设对下一时刻的粒子特征进行估计:

$$P(f_{t+1}|f_t)\propto\exp[-(f_t-f_{t-1}-1)^2/2].$$

Step 2. 根据观测图像,计算各个粒子的置信度 p_i.选出置信度大于阈值ε的粒子组成强粒子集 $S_p=\{s_i|p(s_i)>\varepsilon, i\leq n'\}$,$n'$为全体粒子的总数.

Step 3. 对 S_p 进行 K-means 聚类,根据聚类结果对其进行类别标注.对于类别标注与所属对象不符的强粒子,

设为可疑粒子.

Step 4. 计算权重$\pi_t=\mu\times p(x_t=s_t|z_t)$,可疑粒子的信任系数$\mu$为 0.2,其他粒子为 1.

Step 5. 重采样,更新粒子簇.

经过手势的整体定位和指尖位置跟踪后,手掌位置的确定方法是:根据人手的比例特征,我们取手掌最小外包圆的直径为指尖最小外包圆平均直径的 3~4 倍,手掌外包圆的圆心在手的质心 2D 邻域内依次扫描,设圆心在(x_p,y_p)点处时,手掌外包圆内的肤色像素比例最大,则手掌中心位置被确定在(x_p,y_p)点.

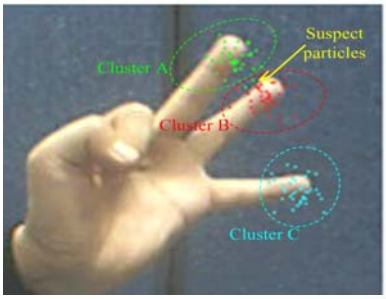


Fig.6 Particle clustering

图 6 粒子聚类

2.4 跟踪过程检测

由于光线变化和环境干扰等因素,跟踪失败是难以完全避免的,通过跟踪检测判断跟踪系统是否失败是很有必要的一项工作.跟踪过程的检测在以往文献中并不多见,特别是当跟踪失败后如何进行后续处理,则更少有人提及.本文的跟踪框架中加入了一个检测部分,检测部分的作用是判定是否触发新的识别过程来完成跟踪的再次初始化.有两种检测结果会触发新的识别过程:

(1) 跟踪失败.被跟踪的对象与模板无法正确匹配,或跟踪的特征消失.导致跟踪失败的原因大致有光线变化的原因、手关节运动产生了自遮挡或其他外部对象产生的遮挡等等.

(2) 出现新的可能特征区域.由于手关节的运动,新的手指出现在跟踪视野中.例如,由一个手指的手势变为两个手指的手势.

以上这些都会触发识别过程,对当前手势再次进行识别,从而重新对跟踪进行初始化.再次初始化涉及到指尖跟踪失败的检测与手势模板变化的检测.

对于跟踪失败的检测,本文采用的是统计跟踪粒子的置信度的方法.跟踪对象的全体粒子的平均置信度若低于最低置信度阈值δ,则认为跟踪失败,即满足公式(4)时识别过程被触发.

$$\sum_{i=1}^{n} p(x=s_i \mid z)/n<\delta \tag{4}$$

手势跟踪模板变化的检测是基于手及其手指的结构特征,根据对手和手指结构特征的先验知识以及当前被跟踪手指的位置,对手势区域中其他手指可能出现的方位进行检测,一旦检测到手指个数与被跟踪模板手指个数不符,识别过程将被触发.手和手指的结构特征如图 7 所示,箭头所指为当前手的朝向,弧形虚线部分是以手的质心为中心的半圆形检测条带,灰色阴影部分是检测条带检测到的手指区域,也就是手指可能出现的区域.

识别过程一旦被触发,则对当前的手势图像进行再识别.再识别的过程是将当前手势图像的图像特征与上一阶段的手势类型相结合,基于最大后验概率(maximum a posteriori,简称 MAP)的方法(参见公式(5))得到当前手势的分类结果:

$$\omega_t=\text{argmax}_i(p(Z_t|C_t=\omega_i,C_{t-1})p(C_t=\omega_i)) \tag{5}$$

其中,条件概率$p(Z_t|C_t=\omega_i,C_{t-1})$由下面的公式(6)得到:

$$p(Z_t|C_t=\omega_i,C_{t-1})=p(C_t=\omega_i|C_{t-1})p(Z_t|C_t=\omega_i) \tag{6}$$

转移概率$p(C_t=\omega_i|C_{t-1})$可以通过经验设定,或者通过样本学习的途径,利用 Baum-Welch 算法[35]得到各个转

移概率的最大似然估计.

新的识别过程的识别结果将与当前图像帧中手势特征进行匹配,按照第 2.2 节中的方法完成跟踪模板的初始化,从而开始新一轮的跟踪.

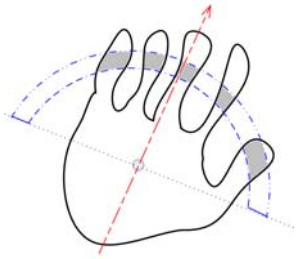


Fig.7 The structural template of gesture

图 7 手势结构模板

3 实验结果及分析

我们用 VC++实现了本文提出的变形手势跟踪方法,为了验证算法的有效性,通过普通 USB 摄像头,采集了一组实验室光照条件下,不带任何特殊标记的手势单目视频录像,并在一台 CPU 为 Pentium IV 1.6GHz,内存为 256M 的普通 PC 机上进行了实验.可识别的静态手势(posture)状态被分为 6 种类型,分别为 *A*:握拳、*B*:单个食指手势、*C*:食指+中指手势、*D*:中指+无名指+小指手势、*E*:食指+中指+无名指+小指手势、*F*:全部手指伸开手势.实验中参数设置如下:CAMshift 算法的最大迭代次数为 10;每个粒子跟踪器的粒子个数设为 40 个;*K*-means 聚类中误差阈值设为 0.1,最大迭代次数为 30;手指检测的比例阈值设为 1.0.变形手势的整个跟踪过程无人工干预.

表 1 给出了 6 种手势的识别率与匹配成功率.匹配成功率是指在正确识别出手势类型的情况下,按照第 2.2 节中给出的匹配方法将图像特征与手势类型进行匹配的成功比率.出现匹配失败的主要原因是公式(2)中的比例阈值目前还未实现自适应选取,在手指长度比较短的情况下未能作出正确判断.图 8 给出了其中两个视频手势片断跟踪过程的部分截图,每张截图的右上角标出了当前的手势类型.

通过实验我们对本文的跟踪框架进行了性能测试,结果如下:

(1) 跟踪的实时性.在上述机器配置下,对分辨率为 320×240 的手势视频进行跟踪.对握拳手势进行跟踪时跟踪速度最快,为 26 帧/s;对 5 根手指完全伸开的手势进行跟踪时跟踪速度最慢,为 9 帧/s;平均跟踪速度为 18 帧/s.可以满足手势交互对实时性的要求.

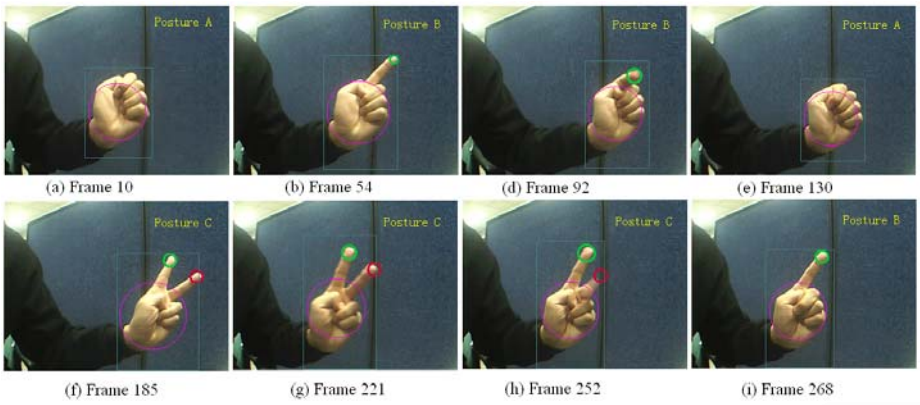
(2) 跟踪的准确性.通过目测,指尖、手掌以及整个手势的跟踪准确度满足人机交互的需要(如图 8 所示).

(3) 跟踪的连续性.连续性主要取决于跟踪的准确度与跟踪状态的检测能力.实验使用的测试视频片断平均长度为 60s,每段视频中手势变换类型不少于 4 种,结果显示准确跟踪时间长度占视频总时间的 95%以上.实验结果也验证了本方法中跟踪检测能力的可靠性.通过跟踪检测,可以及时发现跟踪失败以及检测到手形发生的改变,从而触发再识别过程,使跟踪的连续性得以保持.

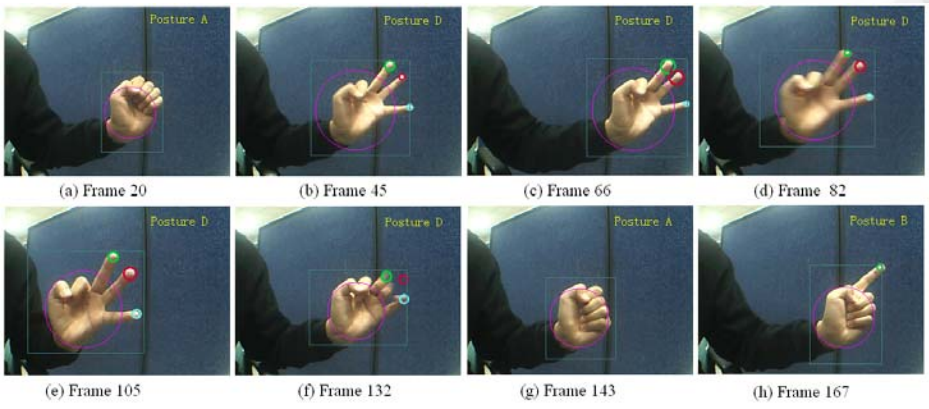
Table 1 Recognition and matching ratios of static gestures

表 1 静态手势识别与匹配率

Posture type	Recognition ratio (%)	Matching ratio (%)
Posture *A*	100	100
Posture *B*	100	100
Posture *C*	99	100
Posture *D*	93.5	100
Posture *E*	92	100
Posture *F*	98	98.5



(a)



(b)

Fig.8 Tracking results of two video clips of deformable hand gesture

图 8 两组变形手势跟踪结果

表 2 给出了本文方法与其他方法的比较结果.与其他手势跟踪方法相比,本方法实现了自动恢复跟踪和手势模型的更新.在实时性方面明显好于 Ivan Laptev[36]的多状态手势跟踪方法(10 帧/s).Yoichi Sato[11]方法的单手平均跟踪速度 25~30 帧/s,略好于本文方法,但它只能对一种固定形状的手形进行跟踪,无法解决手势变形问题.Stenger[15]的基于 3D 模型的多关节手势跟踪方法虽然能对手势变形进行有效跟踪,但在单摄像头条件下跟踪速度仅为 3 帧/s,无法满足实时性要求.

Table 2 The comparison of tracking methods

表 2 跟踪算法比较

	Our method	Ivan Laptev method	Yoichi Sato method	B.Stenger method
Average tracking rate (f/s)	18	10	>=25	3
Deformable gesture tracking	Yes	No	No	No
Tracking auto-resumption	Yes	No	No	No
Tracking auto-initialization	Yes	No	No	No

4 总 结

本文针对人机交互领域基于视频手势的实时交互任务提出一种快速、连续的变形手势跟踪方法.它结合了基于模型与基于表观方法的特点,是建立在对目标对象-手势的理解基础上,通过识别静态手势实现了自动跟踪初始化和跟踪失败后的自动恢复.与传统的基于模板的跟踪方法不同,它能够动态地更新跟踪模板,从而适应多关节手势对象不断变化的外观轮廓.通过将复杂的高维度特征向量分解为多个 2D 跟踪模板,跟踪计算量大为减

小,实验表明可以满足交互的实时性要求.该跟踪方法还实现了跟踪的自动恢复,保证了交互的连续性.

在下一步工作中,我们将进一步利用手势的先验知识,提高跟踪的鲁棒性和对复杂背景环境的抗干扰能力.

References:

[1] Wu Y, Huang TS. Human hand modeling, analysis and animation in the context of HCI. In: Proc. of the Int'l Conf. on Image Processing. 1999. 6−10. http://www.informatik.uni-trier.de/~ley/db/conf/icip/index.html

[2] Zhu YX, Ren HB, Xu GY, Lin XY. Toward real-time human-computer interaction with continuous dynamic hand gestures. In: Proc. of the IEEE Int'l Conf. on Automatic Face and Gesture Recognition. 2000. 544−551. http://www.informatik.uni-trier.de/~ley/db/conf/fgr/index.html

[3] Nielsen M, Störring M, Moeslund TB, Granum E. A procedure for developing intuitive and ergonomic gesture interface for HCI. In: Proc. of the GW 2003. LNAI 2915, 2003. 409−420. http://www.informatik.uni-trier.de/~ley/db/conf/gw/gw2003.html

[4] Mo ZY, Lewis JP, Neumann U. SmartCanvas: A gesture-driven intelligent drawing desk system. In: Proc. of the IUI 2005. 2005. 239−243. http://www.informatik.uni-trier.de/~ley/db/conf/iui/iui2005.html

[5] Malik S, Laszlo J. Visual touchpad: A two-handed gestural input device. In: Proc. of the ACM ICMI 2004. 2004. 289−296. http://www.informatik.uni-trier.de/~ley/db/conf/icmi/icmi2004.html

[6] Rehg JM, Kanade T. Visual tracking of high DOF articulated structures: An application to human hand tracking. In: Proc. of the 3rd European Conf. on Computer Vision. 1994. 35−46. http://www.informatik.uni-trier.de/~ley/db/conf/eccv/eccv1994-2.html

[7] Ren HB, Zhu YX, Xu GY. Vision-Based recognition of hand gestures: A survey. Chinese Journal of Electronics, 2000,28(2): 118−122 (in Chinese with English abstract).

[8] Wachter S, Nagel HH. Tracking of persons in monocular image sequence. Computer Vision and Image Understanding, 1999,74(3): 174−192.

[9] Sidenbladh H, Black M, Fleet D. Stochastic tracking of 3D human figures using 2D image motion. In: Proc. of the 6th European Conf. on Computer Vision. 2000. 702−718. http://www.informatik.uni-trier.de/~ley/db/conf/eccv/eccv2000-2.html

[10] Hardenberg CV, Berard F. Bare-Hand human-computer interaction. In: Proc. of the ACM workshop on Perceptual User Interface. 2001. 15−17. http://conferences.cs.ucsb.edu/PUI/

[11] Sato Y, Kobayashi Y, Koike H. Fast tracking of hands and fingertips in infrared images for augmented desk interface. In: Proc. of the IEEE Int'l Conf. on Automatic Face and Gesture Recognition. 2000. 462−468. http://www.informatik.uni-trier.de/~ley/db/conf/fgr/index.html

[12] Oka K, Sato Y. Real-Time fingertip tracking and gesture recognition. Proc. of IEEE Computer Graphics and Applications, 2002, 22(6):64−71.

[13] Kolsch M, Turk M. Fast 2D hand tracking with flocks of features and multi-cue integration. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2004. 158−164. http://www.informatik.uni-trier.de/~ley/db/conf/cvpr/cvpr2004-2.html

[14] Zieren J, Unger N, Akyol S. Hands tracking from frontal view for vision-based gesture recognition. In: Proc. of the DAGM 2002. LNCS 2449, 2002. 531−539. http://www.informatik.uni-trier.de/~ley/db/conf/dagm/dagm2002.html

[15] Stenger B, Mendonca PR, Cipolla R. Model based 3D tracking of an articulated hand. Proc. of IEEE Conf. of Computer Vision and Pattern Recognition, 2001,2:310−315.

[16] Bray M, Koller-Meier E, Van Gool L. Smart particle filtering for 3D hand tracking. In: Proc. of the 6th IEEE Conf. on Automatic Face and Gesture Recogntion. 2004. 675−680. http://www.informatik.uni-trier.de/~ley/db/conf/fgr/index.html

[17] Aggarwal JK, Cai Q. Human motion analysis: A review. Computer Vision and Image Understandin, 1999,73(3):428−440.

[18] Jun J, Wang L, Singh S. Video analysis of human dynamics—A survey. Real-Ttime Imaging, 2003,9(5):321−346.

[19] Wang L, Hu WM, Tan TN. A survey of visual analysis of human motion. Chinese Journal of Computers, 2002,25(3):225−237 (in Chinese with English abstract).

[20] Wren C, Azarbayejani A, Darrell T, Pentland A. Pfinder: Real-Time tracking of the human body. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997,19(7):780−785.

[21] Hoch M, Litwinowicz P. A practical solution for tracking edges in image sequences with snakes. The Visual Computer, 1996,12(2): 75−83.

[22] Xu CY, Prince JL. Snake, shapes, and gradient vector flow. IEEE Trans. on Image Processing, 1998,7(3):359−369.

[23] Cheng JC, Moura J. Capture and representation of human walking in live video sequences. IEEE Trans. on Multimedia, 1999,1(2): 144−156.

[24] Sminchisescu C, Triggs B. A robust multiple hypothesis approach to monocular human motion tracking. Technical Report, RR-4208, INRIA, 2001.

[25] Ning HZ, Tan TN, Wang L, Hu WM. People tracking based on motion model and motion constrains with automatic initialization. Pattern Recognition, 2004,37(7):1423−1440.

[26] Jaward MH, Mihaylova L. A data association algorithm for multiple objects tracking in video sequence. In: Proc. of the 2006 IEE Seminar on Target Tracking: Algorithms and Applications. Birmingham, 2006. http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=10860

[27] Letessier J, Berard F. Visual tracking of bare fingers for interactive surfaces. In: Proc. of the 17th ACM Symp. on User Interface Software and Technology. 2004. 119−122. http://www.informatik.uni-trier.de/~ley/db/conf/uist/index.html

[28] Zhu JY, Wang XY, Wang WX, Dai GZ. Hand gesture recognition based on structure analysis. Chinese Journal of Computers, 2006,29(12):2130−2137 (in Chinese with English abstract).

[29] Zhang DS, Lu GJ. Review of shape representation and description techniques. Pattern Recognition, 2004,37(1):1−19.

[30] Krahnstoever N, Yeasin M, Sharma R. Automatic acquisition and initialization of articulated models. Machine Vision and Applications, 2003,14(4):218−228.

[31] Ju SX, Black MJ, Yacoob Y. Cardboard people: A parameterized model of articulated image motion. In: Proc. of the 2nd Int'l Conf. on Automatic Face and Gesture Recognition. 1996. 38−44. http://www.informatik.uni-trier.de/~ley/db/conf/fgr/index.html

[32] Bradski GR. Real time face and object tracking as a component of a perceptual user interface. In: Proc. of the IEEE 4th Workshop on Application of Computer Vision. 1998. 214−219. http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=5940

[33] Isard M, Blake A. Contour tracking by stochastic propagation of conditional density. In: Proc. of the European Conf. of Computer Vision. 1996. 343−356. http://www.informatik.uni-trier.de/~ley/db/conf/eccv/eccv1996-1.html

[34] Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. on Signal Processing, 2002,50(2):174−188.

[35] Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. of the IEEE, 1989,77(2): 257−286.

[36] Laptev I, Lindeberg T. Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features. In: Scale-Space 2001. LNCS 2106, 2001. 63−74. http://www.informatik.uni-trier.de/~ley/db/conf/scalespace/scalespace2001.html

附中文参考文献:

[7] 任海兵,祝远新,徐光佑.基于视觉手势识别的研究——综述.电子学报,2000,28(2):118−122.

[19] 王亮,胡卫明,谭铁牛.人运动的视觉分析综述.计算机学报,2002,25(3):225−237.

[28] 朱继玉,王西颖,王威信,戴国忠.基于结构分析的手势识别.计算机学报,2006,29(12):2130−2137.

王西颖(1974－),男,江苏睢宁人,博士,主要研究领域为计算机视觉,视频分析,人机交互.

戴国忠(1944－),男,研究员,博士生导师,CCF 高级会员,主要研究领域为人机交互,计算机图形学.

张习文(1971－),男,博士,副研究员,主要研究领域为模式识别,图像理解.