

# 一种基于视觉的手指屏幕交互方法

徐一华, 李善青, 贾云得

(北京理工大学计算机科学技术学院, 北京 100081)

**摘 要:** 本文提出了一种新的基于视觉技术的手指屏幕交互方法, 利用手指在屏幕上的移动和点击等动作完成与计算机的交互. 该方法使用两个摄像头从不同视角实时拍摄计算机屏幕, 结合人手肤色和轮廓信息对手指进行实时跟踪, 并根据两幅图像中的指尖位置和摄像头参数来判断指尖是否接触屏幕. 实验结果表明, 该方法可以完成鲁棒的手指跟踪和准确的点击检测, 对光照条件没有严格的限制, 实现了自然友好的人机交互操作, 可以运用于智能交互、数字娱乐、穿戴计算等多种应用场合.

**关键词:** 手指跟踪; 手指屏幕交互; 人机交互

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112 (2007) 11-2236-05

## A Vision-Based Method for Finger-Screen Interaction

XU Yi-hua, LI Shan-qing, JIA Yun-de

(School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** This paper presents a novel vision-based method for finger-screen interaction which achieves natural interactions through moving and clicking a single finger on a computer screen. The method uses two cameras at different viewpoints to capture video images covering the computer screen in real time. We integrate two cues of skin color and contour information to track finger's motions in the framework of particle filter. The clicking actions on the screen surface are detected with the obtained fingertip positions and the geometry parameters of cameras. Experimental results demonstrate that the interactive method can realize robust tracking of finger motion, against complex background and varying illumination conditions, as well as the instant detection of finger clicking actions. This method achieves robust and natural human-computer interaction, and can be widely used in many applications, such as intelligent interaction, digital entertainment and wearable computing environment.

**Key words:** finger tracking; finger-screen interaction; human-computer interaction

## 1 引言

键盘、鼠标和操纵杆等是传统的人机交互设备, 但在使用自然性和友好性等方面有一定的局限性. 人们习惯用双手与外界进行大部分交互操作, 因此利用人手的运动和姿势与计算机进行交互是很自然的. 基于视觉的手势交互技术利用摄像头等设备获取人手运动的图像序列, 通过视觉算法提取人手静态或动态特征, 识别人手的姿态或动作, 从而理解人的意图实现人机交互. 该技术在智能交互、计算机游戏和穿戴计算等领域有广泛的应用前景.

基于视觉的手势交互领域存在许多困难和挑战, 如复杂背景和光照变化的影响, 人手快速运动产生的严重变形, 手指之间的相互遮挡等, 这些都使得鲁棒精确的图像分割和特征提取成为难点<sup>[1]</sup>. 为降低视觉处理的复

杂性, 人们使用带有特殊标记的手套等来辅助获取手的位置、方向等信息<sup>[2]</sup>. 这些辅助设备可以在很大程度上提高算法的鲁棒性, 但也给使用者带来了不方便. 人们利用手部的自然信息如肤色、形状等进行特征提取, 开发出许多基于手势的交互系统. 如在 Hall<sup>[3]</sup>, Kjeldsen<sup>[4]</sup>, Hardenberg<sup>[5]</sup>, Takao<sup>[6]</sup>等人的工作中, 通过投影仪将图像投射到墙面或桌面, 利用摄像头捕捉人手在投影平面上的运动和姿势, 从而实现了手指绘图或点击按钮等交互操作. Zhang 等<sup>[7]</sup>设计的 Visual Panel 系统使用单个摄像头实现了虚拟鼠标和键盘. 它使用任意的矩形面板作为交互界面, 通过手指在纸板上的移动和停留分别实现光标定位和鼠标单击的功能. Malik 等<sup>[8]</sup>设计的 Visual Touchpad 系统使用两个摄像头获取现场视频, 通过人手在一块矩形纸板上的动作完成与计算机的交互. 它利用射影几何原理判断指尖点是否接触纸板表面, 可以模拟

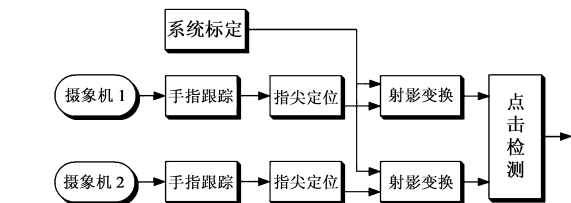
鼠标点击的效果。

本文提出了一种新的基于视觉的手指屏幕交互方法.该方法通过手指在计算机屏幕上的移动和点击来完成人机交互操作,是一种直观友好的交互方式.它使用两个摄像头从不同的角度实时拍摄视频图像,通过检测手指在显示器屏幕上的移动和点击,可以完成与计算机的鲁棒交互.这种新的交互方式有广泛的应用前景,如在手指绘图、触摸屏模拟、游戏驱动等方面都可以得到很好的应用。

2 系统概述

基于视觉的手指屏幕交互系统由一台普通 PC 和两个普通摄像头组成,两个摄像头从屏幕上方斜向下拍摄包含完整显示器屏幕的视频图像,如图 1 所示.这种系统配置可以有效解决手指遮挡的难题,以降低视觉处理的困难性,使该交互系统的鲁棒实现成为可能.系统在计算机屏幕上显示图形交互界面,通过指尖在屏幕上的移动和点击来完成与计算机的直接交互.实验部分给出了该系统在手指绘图、驱动游戏等方面的应用,结果表明该手指屏幕交互系统可以有效可靠的完成人机交互。

系统中的视觉处理算法主要包括系统标定、手指跟踪、指尖定位和点击检测等几部分,如图 2 所示.其中系统标定的目的是获得两个摄像头的成像平面与计算机屏幕平面之间的映射关系.利用该映射关系,可以计算出图像中任一像素点在屏幕平面中的对应位置.手指跟踪部分实现对手指的实时鲁棒跟踪,本文采用的粒子滤波器算法可以在复杂的背景下获得可靠的结果.指尖定位部分处理上述的跟踪结果,准确的计算指尖位置.根据获取的指尖位置和摄像机参数,点击检测部分判断指尖是否点击计算机屏幕.如果检测到点击动作,则完成预设的交互操作。



3 系统标定

系统标定的目的是分别确定两个摄像头的图像坐

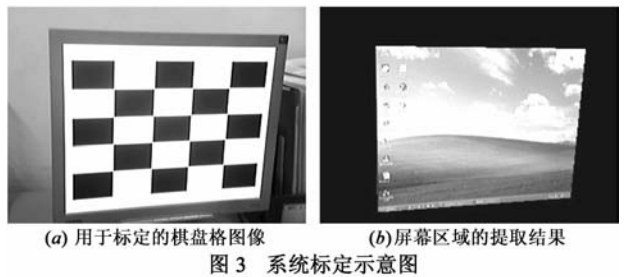
标系与计算机屏幕坐标系之间的映射关系.这里的屏幕坐标系是指显示器用于图像显示空间的坐标系.假设图像平面内任一点 $(X,Y)$ ,其在屏幕平面上的对应点为 $(x,y)$ ,其映射关系可以用射影变换公式表达如下:

$$[x,y,1]^T \equiv H[X,Y,1]^T \tag{1}$$

其中,3×3 矩阵  $H$  的独立参数只有八个,最少需要四组对应点即可确定该矩阵。

系统在计算机屏幕上显示棋盘格图像,如图 3(a),利用角点检测算法提取棋盘格的内角点作为对应点.这样可以获取多组对应点,从而得到精确的标定结果,如使用 5×5 的棋盘格可以得到十六组对应点.选取的对应点应适当远离屏幕边界,以适应屏幕边界被部分遮挡的情况。

得到上述映射关系,即可以计算出屏幕四个顶点在图像中的位置.定屏幕的左下顶点为原点,并由当前显示器分辨率得到其它三个顶点的坐标,然后根据公式(1)可以反推出它们在图像中的坐标,依次连接四个顶点可得到图像中屏幕的形状.四边形所围的区域是算法所关注的兴趣区域,区域以外的部分不需要处理,其结果如图 3(b)所示。



4 手指跟踪与指尖定位

手指跟踪实现对手指的实时鲁棒跟踪,为后续的处理提供手指的位置、形状等信息.由于实验背景及手指的运动都比较复杂,本文采用 Isard 等人<sup>[9]</sup>给出的使用重要采样的粒子滤波器来实现对单个手指的跟踪.因为单个手指的轮廓形状比较简单且相对固定,该方法可以在保证实时性的基础上提供准确的跟踪结果.手部检测部分根据肤色信息可以给出指尖的粗略位置,该信息用于指导粒子的重要采样,可以有效的提高采样的准确性.由于同时使用重要采样和标准的粒子滤波器采样,只有两部分粒子采样同时失败才会导致跟踪错误.这种方法提高了跟踪的鲁棒性,同时解决了手指大幅度运动造成的目标丢失问题。

4.1 人手检测

人手检测的作用在于获取图像中的人手区域,进而计算指尖的粗略位置.在实际的应用中,某些场合的

背景比较复杂,为手部的准确检测增加了很多困难.本文在 HSV 空间使用高斯分布来描述肤色的色调 H 及饱和度 S 的分布规律.该方法将输入的彩色图像转换为灰度图像,越接近肤色的像素其灰度值越大,如图 4(b)所示.这样,设置一个阈值即可将人手区域分割出来.由于背景中可能存在颜色接近肤色的物体,检测结果中会存在多个较小的噪声区域,利用连通区检测技术选择图像中最大的肤色区域为人手区域.

根据手部检测的结果,本文采用 Takao 等<sup>[6]</sup>提出的计算模型来确定指尖的粗略位置.如图 4(c)所示,人手区域的中心点与该区域下边界中心点的连线为主轴,将沿主轴方向距中心点最近的边缘点定为粗略指尖点.上述得到的灰度图像和粗略指尖位置在后续粒子滤波跟踪算法中将分别用于提取手指的边缘轮廓和指导粒子的重要采样.

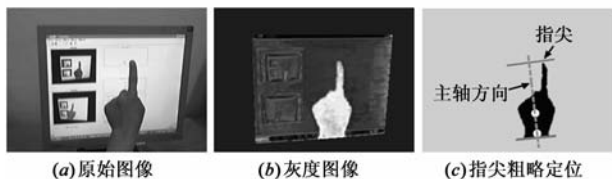


图 4 手部检测结果

## 4.2 粒子滤波跟踪

粒子滤波算法主要包括随机采样,运动预测,置信度计算,参数估计等几部分.随机采样指从上一次结果的样本集中随机选取粒子产生新的样本集;接下来进行预测,由随机动态模型产生粒子的状态向量;然后根据观测模型计算粒子的置信度;最后计算样本集中所有粒子的数学期望得到对参数向量的估计.采用粒子滤波器可以对手指进行实时鲁棒的跟踪,下面讨论手指跟踪中使用的具体模型.

手指轮廓模型使用 7 个控制点的 B 样条曲线来表示,如图 5(a)所示.由于手指运动过程中轮廓的形状相对固定,手指轮廓跟踪过程即为确定如下状态向量的过程:

$$\mathbf{x} = [P_x, P_y, \alpha_x(\cos\theta - 1), \alpha_y(\cos\theta - 1), -\alpha_x\sin\theta, \alpha_y\sin\theta]^T$$

其中  $P_x, P_y$  分别表示手指模板的水平和垂直位移,  $\alpha_x, \alpha_y$  表示模板在水平和垂直方向的尺度变化量,  $\theta$  表示模板旋转的角度.

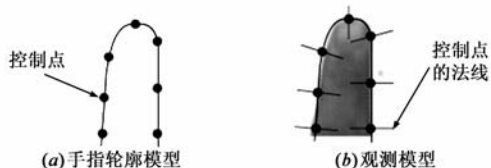


图 5 手指轮廓及观测模型

观测的具体步骤是:(1)对手部检测获取的如图 4

(b)的灰度图像进行 Canny 滤波;(2)沿预测的样条曲线控制点的法线方向寻找边缘点,如图 5(b)所示,选择距离控制点最近的边缘点作为手指轮廓上的点;(3)根据确定的边缘点与控制点之间的最短距离,计算第  $n$  个粒子的置信度:

$$\pi_t^{(n)} = \exp\left(-\sum_{i=1}^N \frac{d_i^2}{2\sigma^2}\right) \quad (2)$$

其中  $\sigma$  为密度函数的方差,  $N$  为控制点的个数.

本文采用匀速直线模型对手指的运动进行预测,根据  $t$  和  $t-1$  时刻的位置可以计算出  $t+1$  时刻手指的位置.该运动模型可以在一定程度上提高跟踪的鲁棒性.

重要采样是借助重要函数来指导粒子到最有可能的目标区域进行采样的方法,其目的是提高采样的准确性.本文采用手部检测获取的指尖粗略位置来指导粒子的采样.为了避免因指尖检测错误导致的目标丢失,本文同时使用重要采样和标准的粒子滤波器采样.即一部分粒子到检测出的指尖区域进行重要采样,另一部分则到由上一帧信息预测出的区域进行标准的粒子滤波器采样.目前,我们人为设置两部分的粒子数目的比例;将来,系统可以根据两部分粒子的跟踪结果进行自适应调整采样的比例.该方法只有两部分粒子同时采样失败才会导致跟踪错误,不但提高了采样的准确性,而且解决了手指突然大幅度运动造成的目标丢失问题,实现了对手指的鲁棒跟踪.

由粒子滤波器的实现过程知,参数估计使用全部粒子的加权平均值,当置信度较小的粒子数目很多时结果会不准确.本文选取样本集中置信度大于给定阈值的粒子进行参数估计.实验表明该方法可以有效提高计算结果的精度.由于在跟踪的过程中可能出现目标丢失等错误,有效的错误恢复算法是非常必要的.本文通过设定一个重新启动区域来解决上述问题.当目标丢失时,将手指移动到该区域附近即可重新启动跟踪的过程.手指跟踪的结果如图 6 所示.



图 6 手指跟踪和指尖定位

手指跟踪部分获得的是手指轮廓的 B 样条曲线,计算指尖所在控制点的坐标就可以获得指尖的位置.手指样条曲线共使用 7 个控制点,第 4 控制点定义在指尖的位置.由于上述粒子滤波器可以提供精确的跟踪结果,指尖定位是准确而有效的,如图 6 所示(黑色曲线为跟踪的手指轮廓,红色十字表示检测出的指尖位

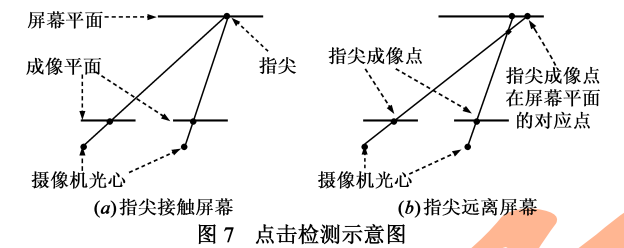


置). 后续的点击检测部分需要利用该信息判断指尖是否接触屏幕表面.

5 点击检测

点击检测部分用于检测指尖是否接触屏幕表面. 本文通过双摄像头获取指尖的三维信息, 利用指尖点击计算机屏幕来模拟鼠标点击事件, 其特点是操作自然方便, 具有很好的鲁棒性.

根据系统标定获取的参数, 由公式(1)可以计算出左右两幅图像中的指尖点在屏幕坐标系内的对应点位置. 由射影几何知, 如果指尖恰好位于屏幕表面, 屏幕坐标系下两个对应点是重合的, 如图 7(a)所示; 而当指尖位于屏幕前方一定距离时两个对应点不重合, 如图 7(b)所示. 于是判断指尖是否接触屏幕的问题转化为判断两对应点是否重合. 设置一个距离阈值, 当两对应点间的距离小于等于给定阈值时即认为指尖接触屏幕.



6 实验结果

本文实验所使用计算机的主频为 2.0GHz, 图像采集设备为普通的网络摄像头. 系统运行时, 两个摄像头近似同步地采集分辨率为  $320 \times 240$  像素的图像序列, 两个跟踪器分别跟踪两个角度下的单手指运动, 从而实现手指运动和点击操作的实时检测. 实验系统的视觉处理部分速度可以达到 15 帧/秒. 我们拍摄用手指在屏幕上写字母 ABC 的 400 帧图像序列用于验证系统的跟踪精度. 计算图像序列中跟踪值的误差结果为:  $X$ 、 $Y$  坐标的算术平均误差分别为 0.65 和 1.76 像素, 标准误差分别为 0.95 和 2.19 像素. 图 8 是系统跟踪获取的指尖位置与人工标定真实值的对比结果 (前 60 帧). 实验结果表明, 系统可以对手指进行鲁棒的跟踪, 完成比较精确的指尖定位, 能迅速响应指尖的点击操作, 从而有效地实现人机交互的功能. 系统抗干扰能力很强, 在环境光照发生很大变化时依然能够鲁棒运行. 基于该交互方法, 本文设计了一些简单的应用场景, 重载了操作系统中鼠标的驱动程序, 可以通过手指来驱动所有使用鼠标的应用程序. 如图 9 展示的是手指绘图的功能, 指尖接触屏幕并不断移动就可以绘制出其运动轨迹曲线. 图 10 是手指驱动一个拼图游戏的实验结果, 通过指尖移动来模拟鼠标移动, 指尖点击屏幕来模拟鼠标的

单击.

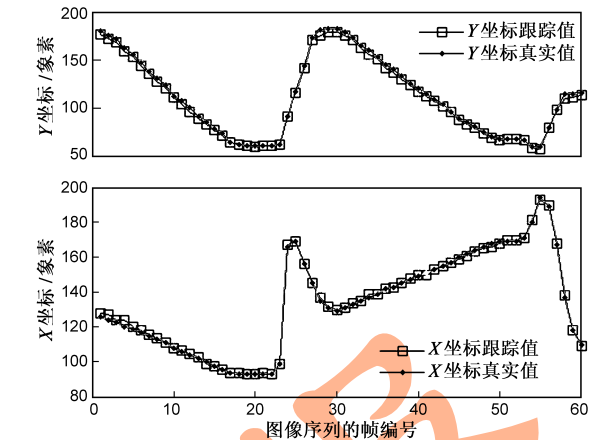


图 8 前 60 帧图像序列的跟踪结果

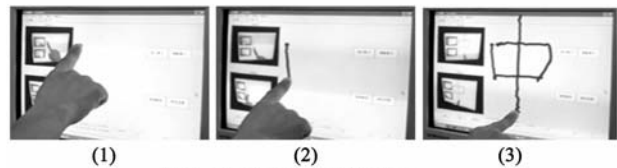


图 9 手指绘图功能的演示

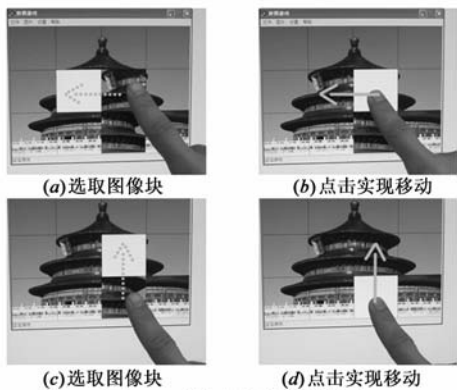


图 10 手指驱动的拼图游戏

7 总结与讨论

本文介绍了一种新的基于视觉技术的手指屏幕交互方法, 通过手指在计算机屏幕上的移动和点击很好的完成人机交互操作. 该方法使用两个普通摄像头实现了自然、有效的人机交互操作. 通过对摄像头空间位置的适当配置, 本文有效地解决了手指的互遮挡等问题, 使得实时鲁棒的手指跟踪和点击检测成为可能. 该系统可以鲁棒的工作, 对环境光照等没有严格要求, 可以应用于人机交互、数字娱乐等多种应用场合.

目前, 本系统实现的功能是指尖的定位和点击检测. 后续可以进行更多的研究工作, 如可以增加复杂手势识别的功能, 由手势指令和点击操作构建更强大、更人性化的人机交互系统. 另外三维信息是很有用的数据, 可以考虑用来提高交互操作的准确性和鲁棒性.

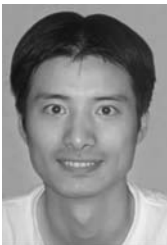
## 参考文献:

- [1] Pavlovic VI, Sharma R, Huang TS. Visual interpretation of hand gestures for humancomputer interaction: A review [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1997, 19(7): 677 – 695.
- [2] Dorfmueller-Ulhaas K, Schmalstieg D. Finger tracking for interaction in augmented environments [A]. IEEE and ACM International Symposium on Augmented Reality [C]. New York: IEEE Computer Society, 2001. 55 – 64.
- [3] Hall D, Gal CL, Martin J, et al. Magicboard: A contribution to an intelligent office environment [J]. Robotics and Autonomous Systems, 1999, 35(3 – 4): 211 – 220.
- [4] Kjeldsen R, Pinhanez C, Pingali G, et al. Interacting with steerable projected displays [A]. Proc Conf Automatic Face and Gesture Recognition [C]. Washington: IEEE Computer Society, 2002. 402 – 407.
- [5] Hardenberg CV, Berard F. Bare-hand human-computer interaction [A]. Proc ACM Workshop on Perceptive User Interfaces [C]. New York: ACM Press, 2001. 155 – 162.
- [6] Takao N, Shi J, Baker S. Telegraffiti: A camera-projector based remote sketching system with hand-based user interface and automatic session summarization [J]. International Journal of Computer Vision. 2003, 53(2): 115 – 133.
- [7] Zhang Z, Wu Y, Shan Y, et al. Visual panel: Virtual mouse, keyboard, and 3D controller with an ordinary piece of paper [A]. Proc. ACM Workshop on Perceptive User Interfaces [C]. New York: ACM Press, 2001. 219 – 226.
- [8] Malik S, Laszlo J. Visual touchpad: A two-handed gestural input device [A]. Proc International Conference on Multimodal

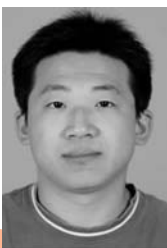
Interfaces [C]. New York: ACM Press, 2004. 289 – 296.

- [9] Isard M. Visual Motion Analysis by Probabilistic Propagation of Conditional Density [D]. Oxford: Oxford University, 1998.

## 作者简介:



**徐一华** 1998 年、2001 年、2007 年分别获得北京理工大学计算机专业学士、硕士、博士学位; 2001 ~ 2004 年担任微软研究院助理研究员研究方向为计算机视觉、图像视频处理、媒体计算、嵌入式计算等. E-mail: yihuaxu@bit.edu.cn



**李善青** 2004 年获北京理工大学计算机专业学士学位, 现为北京理工大学计算机专业博士研究生. 研究方向为计算机视觉、智能人机交互等. E-mail: shanqingli@bit.edu.cn



**贾云得** 北京理工大学计算机学院教授, 博士生导师. 主要研究领域: 计算机视觉、媒体计算、智能人机交互等. E-mail: jiaiyunde@bit.edu.cn