

基于 WebCam 的人机交互方法与应用

李锦鹏

(广东工业大学计算机学院, 广州, 510006, email:adison_lee@hotmail.com)

摘要: 本文探讨借助视频采集和运动分析技术, 利用普通的显示器, 建立一个模拟触摸屏的人机交互方法。该方法可以让用户(人)不借助鼠标、键盘等输入设备, 通过视频采集设备捕获用户的行为, 实现与计算机的交互。作为应用实例, 实现了基于 WebCam 的黑白棋游戏, 效果良好。

关键词: 人机交互; 模拟触摸屏; 运动分析; WebCam; 黑白棋。

A Human-Computer Interface Approach based on WebCam and its Application

LI Jinpeng

(Faculty of Computer, Guangdong University of Technology, Guangzhou, 510006)

Abstract: The purpose of the present paper is to introduce an approach to setup a touchscreen-like human-computer interface based upon a common monitor and a video capture device, which is essential to video acquisition and motion analysis techniques. This approach offers individuals, who have difficulties in manipulating a mouse or keyboard, an alternative interface to implement a touchscreen-like system. In this paper, we will also provide an application, Reversi Game, to demonstrate the well behavior of this approach with a video-chat webcam.

Keywords: Human-Computer Interface; touch screen simulation; motion analysis; WebCam; Reversi.

1. 引言

计算机用户界面是计算机系统的重要组成部分, 它直接关系到整个计算机系统的可使用性和效率。现代人机用户界面的发展, 使得人们能够借助一些辅助设备与计算机进行对话, 以实现某些特定领域的功能和任务。比如, 在微机系统下, 我们通常以鼠标操作作为输入来控制图形界面(GUI)。鼠标是一种定点设备(pointing device), 定位精确。使用鼠标能够激活“一点是否在一个区域内”的判断。尽管鼠标已经获得广泛应用, 但是, 鼠标的灵活使用需要一个熟练的学习过程; 而在一些应用领域, 如服务、宣传、广告等面向公众的场合, 鼠标往往为用户带来了一定程度的不便, 影响了其在这类领域的应用。

触摸是人类的一种本能操作。手之所及, 便可引发感兴趣对象的相应反映。触摸屏技术就是在这种背景下产生的, 它不依赖于传统的键盘-鼠标界面, 实现了用户和计算机之间简单、直接的互动方式。借助触摸屏技术, 用户几乎无需经过任何培训

就能马上使用计算机。触摸屏还使计算机能够工作在键盘或鼠标显得过于笨重、易受损或不实用的地方, 从而扩大了计算机的应用范围。

但是, 触摸屏相对于普通的显示屏, 由于需要特殊的传感材料, 其生产成本较高^[1]。近年来, 视频采集设备的成本不断下降和性能不断提升, 为人机用户界面的研究和应用提供了新思路。我们可以借助视频采集和运动分析技术, 来模拟实现触摸屏。这样, 用户(人)可以不用(显式地)借助鼠标、键盘等输入设备, 通过视频采集设备捕获用户的行为。文献[2]中提出类似的模拟结构。他使用投影仪, 摄像机和激光笔, 实现了点击系统。我们的研究最大的区别在于使用人类最自然的“手”与计算机的进行交互; 一种新的人机界面, 感知用户界面(PUI)^[3]。

本文探讨建立一个模拟触摸屏的人机交互系统的方法。图1是这种系统的示意图。令计算机屏幕和人手处在摄像机的视野里, 计算机对视频输入进行运动分析, 对人手的动作进行分类, 从而触发相应的事件, 运行计算机的相应的程序, 更新显示输出。在实验中, 我们采用一个普通的WebCam(一般上

网聊天使用的摄像头，默认解析度：320x240）作为计算机“视觉”输入设备，显示器可以是LCD或CRT，WebCam的位置和角度与屏幕都是固定的。并假定光源环境比较稳定。在这些条件下，我们以黑白棋为例，实现了一个模拟的触摸屏。效果良好。

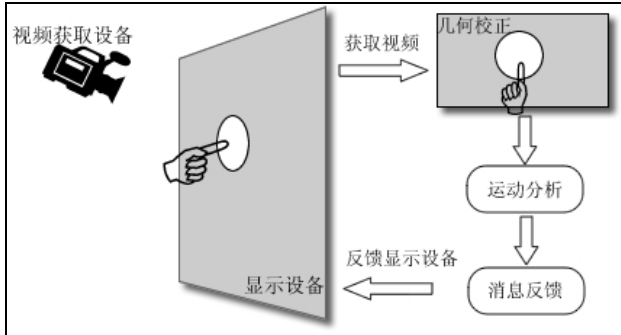


图1 基于视频输入的人机交互系统

本文安排如下。第2节介绍系统的初始化方法，其目的是进行几何校正，使得屏幕在摄像机镜头上的投影和屏幕坐标相匹配。第3节介绍运动分析方法，其目的是对运动对象(人手)的动作进行分类，以完成相应的操作。接下来，作为该模拟触摸屏的应用，我们在第4节实现黑白棋。最后，为结束本文，第5节给出结束语。

2. 初始化

系统初始位置是任意的，只要摄像头视野能够覆盖屏幕和人手即可。因此，在进行运动分析之前，必须进行几何校正，使得视频中的屏幕对象和屏幕的图形坐标合理地匹配起来。



图2 显示器图像

图2为显示器在WebCam中的成像，由于显示器四个角顶点到WebCam距离不一，造成显示器投影到WebCam的平面时，产生几何畸变，原本矩形的屏幕变成了一个凸的曲边四边形。由于凸起的程度不大，对于此系统影响较小，所以我们将其近似为一个四边形。

2.1 定位四个角顶点

在利用摄像头模拟触摸屏时，计算机屏幕发光和反光，以及环境光等因素对成像都有不同程度的

影响。为保证成像质量，为视频输入提供可靠的运动对象，提高触摸的精度，就必须保证拥有一个稳定的环境。由于普通WebCam的摄像质量的限制，色度作用不明显，因而我们重点考虑亮度。

初始化时，采用强对比度的亮度进行全屏作图，如采用黑白两色，分别制造黑白帧（看上去屏幕在闪烁），二者作差，即可检测到屏幕的位置。这种方式，可以有效降低对环境的依赖性。

先做一个全屏闪烁，检测出显示器的二维包围盒，即感兴趣区域(ROI)，从而缩小其在摄像头视野中的范围，避免角点检测时的噪声干扰。

接着进行角点检测，作为几何校正的基础。作黑白帧差之后，可以获得角的最大连续变化点集，设为 S ，在其中找出一离显示器包围盒中心最远的点作为角顶点 $D(x_d, y_d)$ ，即：

$$\begin{aligned} |x_d - x_c| \cdot |y_d - y_c| &= \max \\ x_c &= \text{ScreenWidth} / 2 \\ y_c &= \text{ScreenHeight} / 2 \\ D(x_d, y_d) &\in S \end{aligned}$$

其中， \max 意指达到最大值， ScreenHeight 和 ScreenWidth 分别表示显示器包围盒的长和宽。

2.2 几何校正

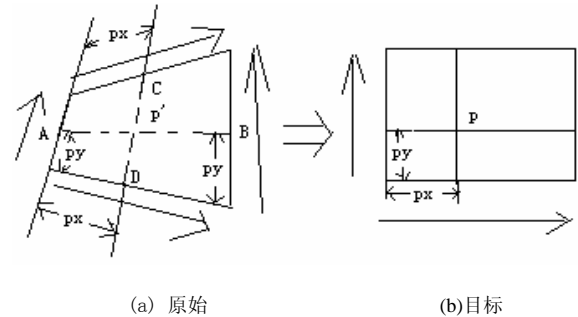


图3 几何校正

假设屏幕成像是由上面求得的角顶点构成的四边形，该四边形对应着一个矩形。几何校正的目的就是通过该四边形求出矩形。我们采用双线性模型来求解^[4]。如图3，在目标转换窗口中，任意找一点 p ，计算出这点在目标窗口比例坐标，所谓比例坐标指如图3.b中 p 点的 x, y 与其边长的比例。设比例为 px, py ，设边长为 TargetWidth 和 TargetHeight ：

$$\begin{cases} px = x / \text{TargetWidth} \\ py = y / \text{TargetHeight} \end{cases}$$

根据目标窗口的 px, py 以及之前检测的四个顶点坐标可以计算出对应原始窗口(图3.a)的 CD, AB 两直线方程，设其方程分别为：

$$\begin{cases} A_1 \cdot x + B_1 \cdot y = C_1 \\ A_2 \cdot x + B_2 \cdot y = C_2 \end{cases}$$

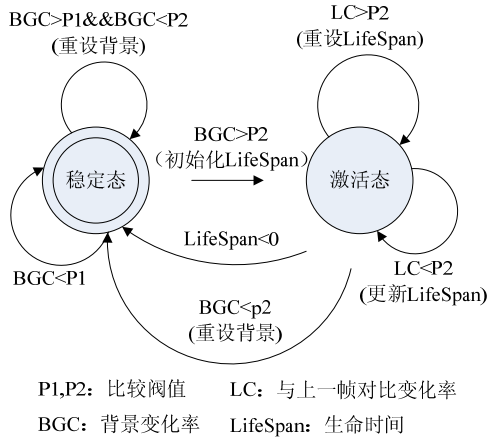
$$D = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}, D_1 = \begin{vmatrix} C_1 & B_1 \\ C_2 & B_2 \end{vmatrix}, D_2 = \begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix}$$

其交点坐标为： $x = D_1 / D$, $y = D_2 / D$ 。

计算出目标窗口每一点在原始坐标对应的坐标，保存为一张索引表，便于下次使用，无需重新计算所有映射关系。

3. 运动分析

3.1 网格检测



把校正后的窗口分解为一个适当大小（本系统划分为 64×64 ）的网格，所谓适当大小指网格宽度应至少为手指宽度的二分之一，但也不能太小，小到与像素同一个数量级。分解为网格，有两个好处：

- 1 降低计算量，不需要从像素数量级进行计算。
- 2 便于动态适应背景，网格可以局部反映一个区域内的运动情况，调整背景很方便。

系统的网格初始化在稳定态，当有运动物体在表面经过的时候，网格就会自动地触发到激活态，它是检测系统的最基本组成部分。

图 4 表示单个网格的状态图。其术语解释如下：

$P1, P2$: 两个触发的阈值 $0 < P1 < P2 < 1.0$ ，本文采用 $P1=0.3$, $P2=0.9$

BGC （“背景变化率”）：当前帧与背景对比的变化率。

LC （“与上帧对比变化率”）：当前帧与前一帧对比的变化率。

“重设背景”：指重设网格的背景，不是整个系统的背景重设。

$LifeSpan$ （“生存时间”）：此参数为了让系统从

“错误”中恢复到正常状态，例如环境光发生了剧烈变化，整个系统的所有网格被触发到激活态，“生存时间”就可以让系统经过一段时间，恢复到稳定态，本文采用 5 秒恢复时间。

重设 $LifeSpan$ ：把 $LifeSpan$ 重新恢复到初始值。

更新 $LifeSpan$ ：根据系统逝去的时间，更新 $LifeSpan = LifeSpan - \text{逝去的时间}$ 。

3.2 稳定态事件分析

$BGC < P1$ ： $P1$ 是一个比较低的阈值，如果过小于 $P1$ 的变化率，证明此网格可能只是受非常小的光变化影响或者没有变化，为了提高效率，暂不需要重设网格的背景。

$BGC > P1 \& \& BGC < P2$ ：网格有变化，但是变化还没有足够能力触发网格，因为手指有足够大小把整个网格遮掩，所以认为是噪声干扰，重设网格的背景，但不激活。

$BGC > P2$ ：网格被接近完全覆盖了，触发其网格。

3.3 激活态事件分析

$LC > P2$ ：非常大的变化率表明物体还在移动，所以重设其生命时间。

$LC < P2$ ：表明运动物体静止了，所以更新“生存时间”。

$LifeSpan < 0$ ：物体停留在此网格时间太长，可能是出错，所以恢复稳定态。

$BGC < P2$ ：标明移动物体离开了，所以恢复到稳定态。

3.4 网格整体分析

前节阐述了如何构建网格的更新触发，本节对网格进行整体分析。

为了便于阐明问题，我们不妨约定“手”必须从显示器边框进入系统。换句话说：中间突然出现的运动物体都不是“手”。

据此，我们对所有网格进行整体分析，找出所有激活态的区域^[5]（所谓区域就是按照 4 或 8 邻域规则得到一个连通激活网格的集合）。把靠近显示器边框（以下简称“边框”）且最大的区域保留并称为目标区域，其余的非靠近边框的区域和靠近边框但非最大区域中的网格强制设置为稳定态。系统做一个额外的约定，在进行点击的过程中，显示器的边框部分网格不能有变化（即使系统限定了边框部分不能有变化，但是由于手对屏幕附近环境的光产生影响及 CRT 逐行扫描也有可能激活边框附近的网格，但网格数量一般来说比“手”产生的少，所以靠近边框但非最大区域中的网格也要设置为稳定态）。

3.5 确定点击位置

在刚才的分析基础上,可以获得一个目标区域。就如图 5 的情况,分析区域的靠边情况,可以得出点击物从屏幕的左边进入。

如果不考虑手指弯曲的情况,指尖的位置应该是离进入边框最大的点(如果是笔直的棒状物效果会更好,因为手掌有时会造成视角上遮掩)。

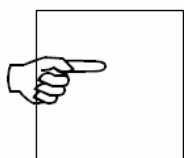


图 5: 一般情况

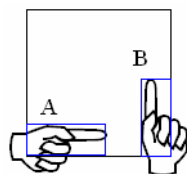


图 6: 角位置

3.6 角位置判定

如果手会刚好在角的位置,如图 6,可以利用手的包围盒,判定其指向。手 A 的宽大于高,所以方向是从左向右。同理 B 的方向是从底向上。

3.7 点击事件

按照上述方法就可以确定点击位置。在一段时间内,如果点击位置稳定在一个范围(可以由用户设置阈值)内,就触发一个点击事件,点击位置就是这个范围的均值。本文设置了 1 秒的稳定时间,3 个网格的稳定范围。

4. 黑白棋的实现

现以黑白棋为例,实现上文所描述的触摸屏模拟系统。

按照黑白棋的行棋规则,参与下棋的双方的行为主要有如下两个特征:

1: 同一时间只有一个人下棋,即只有一只手在屏幕上。

2: 每个人下完棋后,手离开屏幕,接着下一个人开始下棋,之间有一段空白时间。

根据第一个特征,用上一小节的方法,我们可以实现点击操作。点击完毕,系统可能会接着出现误点:原因是手指点击处背景会发生变换,这样就会使指尖附近的网格都被激活。对这种情况,我们可以利用第二个特征,第一个人下完棋后手会离开屏幕(由于手离开屏幕,且系统设定了屏幕的边框部分不能有变化,边框部分的激活网格就变为稳定态,这样指尖附近的激活网格变为非目标区域),接着下一个人开始下棋,之间有段空白时间。就在第一个人下完棋后,系统可以设置一个空白时间进行

等待(本文中,采用 2 秒时间间隔),不立刻进入下一个点的判定,而等待下一个人的行棋。依照上面两个特征,我们就可以利用这种虚拟的触摸屏,进行黑白棋游戏。

5. 结束语

本文研究了建立一个模拟触摸屏的人机交互系统的要素和方法。仔细地就初始化方法和运动分析方法进行了深入的分析。并以黑白棋威力,实现了模拟的触摸屏系统。这里所探讨的方法可以用来构建一些实用系统,如可用作服务业、广告业的宣传和介绍。当然,由于 WebCam 成像能力较差,不可能实现高精度的模拟。在应用中,为了减少误差,系统需要符合一些环境要求:

1: 显示器表面不能产生强烈反光。

2: 系统需要合适的环境光,不能让摄像头产生过分曝光,或者曝光不足。

3: WebCam 视角不能太偏,否则容易造成视角误差和手指的遮掩。

毕竟 WebCam 精度相对较低,误差是必然的,所以及时反馈必要的信息,能帮助用户能够自我调整点击位置^[6]。在黑白棋的实现中,系统提供一个红点来标志点击位置,增加了用户点击期望位置的可靠性。

参 考 文 献

- [1] 杨邦朝,张治安. 触摸屏产业及材料市场现状[J]. 新材料产业, 2003, (1): 31-32.
- [2] 刘芳,林学闯,史元春. 基于激光笔的远程人机交互技术[J]. 中国图象图形学报. 2003, 8(11).
- [3] 江 激,周曼丽. 一种新的人机界面—PUIs[J]. 计算机自动测量与控制. 2001. 9(3).
- [4] Mark Gordon. Anti Aliased Image Transformation [EB/OL]. <http://www.codeproject.com/useritems/Aaform.asp>.2005.11.
- [5] 王耀南,李树涛,毛建旭. 计算机图像处理与识别技术[M]. 北京:高等教育出版社, 2001.6:158-159.
- [6] 李乐山. 人机界面设计[M]. 北京:科学出版社,2004.