

基于图像识别的虚拟翻书系统*

李 炜^{1,2}, 李 志^{1,2}, 郭 星^{1,2}, 吴建国^{2,3}

(1.安徽大学 计算机科学与技术学院,安徽 合肥 230039;

2.安徽大学 计算智能与信号处理教育部重点实验室,安徽 合肥 230039;

3.安徽大学 软件学院,安徽 合肥 230039)

摘 要: 设计了一种基于图像识别的虚拟翻书系统,采用普通摄像头作为传感器捕获人的手势动作,并利用图像识别技术和计算机视觉技术实现虚拟翻书效果。该系统具有虚拟触摸功能,可以点击播放页面中的视频、语音等。

关键词: 虚拟翻书;图像识别;序列图像处理;计算机视觉

中图分类号: TP39

文献标识码: A

Virtual flip book system based on the image recognition

LI Wei^{1,2}, LI Zhi^{1,2}, GUO Xing^{1,2}, WU Jian Guo^{2,3}

(1.School of Computer Science and Technology, Anhui University, Hefei 230039, China;

2.Key Laboratory of Intelligent Computing & Signal Processing of Ministry of Education, Anhui University, Hefei 230039, China;

3.School of Software Engineering, Anhui University, Hefei 230039, China)

Abstract: A virtual flip book system based on image recognition is designed, it uses a ordinary camera as a sensor to capture people's gestures, and uses image recognition technology and computer vision technology to achieve virtual flip page. At the same time it has the virtual touch capabilities, it can be clicked to play video, voice and so on.

Key words: virtual flip book; image recognition; sequence of image processing; computer vision

虚拟翻书系统是一种虚拟电子书,又称感应翻书、互动翻书等,它如同一本打开的书籍,里面可以记载丰富的资料(包括动画、视频、图片)。参观者可以自左向右或者自右向左挥动手臂“翻阅”书籍,还可以选择章节,快速翻阅相关内容,如翻阅一本普通的杂志一般。新颖的模式,强烈的视觉冲击,让虚拟翻书成为越来越不可或缺的创新产品,具有广泛的应用前景。目前的虚拟翻书系统主要利用红外感应或光电感应的方式获取参观者的动作,实现前后翻书效果。但是其存在只能进行简单的前后翻页操作,如果要增加点击播放视频、目录选页等功能需要在电子书模型上设置按钮,这样既不美观也不便于操作,并存在红外感应器的磨损和对翻书动作范围的限制等问题。

本文提出的基于图像识别的虚拟翻书系统主要是通过摄像头采集人手图像,然后对序列图像进行分析,

* 基金项目: 国家自然科学基金项目(60773114); 科技部创新基金项目(09C26213401408)

识别出人手动作,发送指令给计算机,并利用计算机视觉技术显示电子书页面内容从而达到虚拟翻书效果。

1 系统概述

系统的主要构成如图1所示。

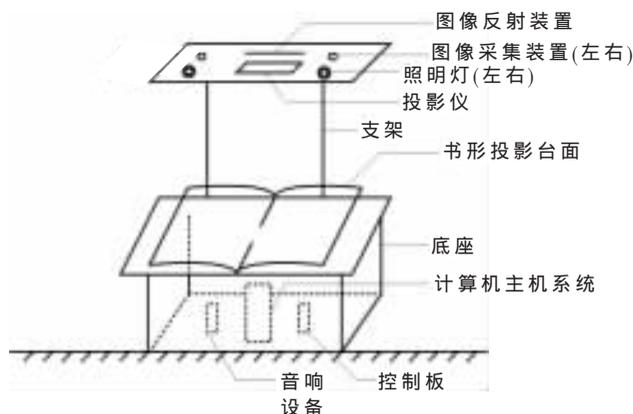


图1 系统主要构成

(1)书形投影台面:显示投影内容,包括文字、图像、声音、动画、视频等。

(2)底座:长方体框架结构,用于支撑书形投影台面,将计算机主机和音响设备放入其中。

(3)计算机主机系统:用于存储展示内容,并进行图像识别控制。

(4)音响设备:用于播放视频、语音等形式的内容。

(5)支架:2根长管,用于支撑照明灯、图像采集装置、投影仪及图像反射装置。

(6)图像采集装置(左右):可以是普通摄像头,用于采集人手运动图像。

(7)照明灯(左右):用于照亮书形投影台面。

(8)投影仪:用于把展示内容通过计算机投影到书形投影台面上。

(9)图像反射装置(镀膜高反射镜):将投影仪投影图像反射到书形投影台面上。

(10)控制板:控制整个系统供电、断电。

本系统中运用的算法主要包括:序列图像处理算法、仿真书边沿识别和系统标定、手指定位算法、计算机视觉算法等。其中仿真书边沿识别是获取算法所关注的兴趣区域,区域以外的部分不需要处理。系统标定的目的是获取摄像头的成像平面和计算机显示平面的映射关系。这样可计算出摄像头拍摄的图像中任一像素点在计算机屏幕平面的对应位置。手指定位算法实现人手在仿真书屏幕的准确定位,并根据虚拟电子书设置的参数实现翻页、点击等动作。计算机视觉算法是利用电影播放原理在电子书前后2页页面内容更换的过程中插入24帧更换过程画面,此时参观者就会看到翻页的动画效果。系统实现原理框图如图2所示。

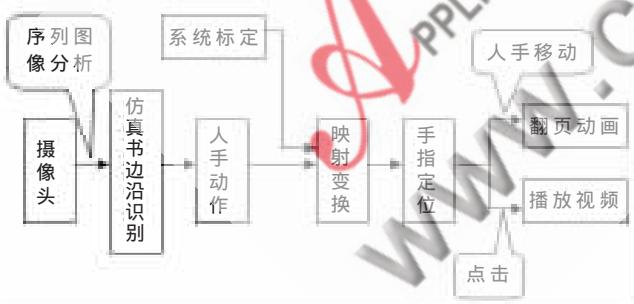


图2 系统实现原理框图

2 序列图像处理

利用 Windows 的 DirectShow 技术采集摄像头拍摄的序列图像,并对图像进行滤波处理去除噪声,得到图像细节。中值滤波是众多图像滤波器中既能很好地抑制噪声,又能很好保护图像细节的一个滤波器,本研究即采用此滤波器对得到的序列图像进行处理。普通的中值滤波对图像进行全局中值滤波处理时,可能导致图像边缘模糊,本算法采用判断图像上的该像素是否为边缘点的方法:(1)边缘判断以 (x,y) 为中心的 3×3 像素区域,

并选取8种代表性边缘,如图3所示。以中心画十字,得到4个邻近像素,设阈值为 T ,邻近像素灰度值与中心像素灰度值的差为 d ,当 $d < T$ 时, s 自动加1;当 $a < s < b$ 时,可确定该像素为边缘点(a, b 为常数, s 为邻近像素与中心像素灰度相似的个数);(2)该像素点是边缘点则不进行任何变换,直接将该像素值 $f(x,y)$ 输出;(3)不是边缘则进行中值滤波。

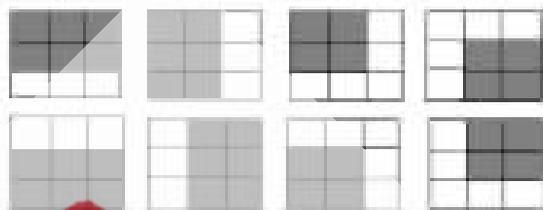
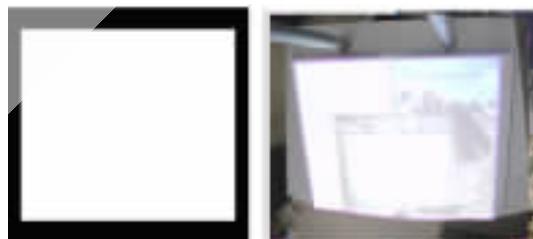


图3 3×3 像素区域的8种代表性边缘

3 仿真书边沿识别和系统标定

仿真书的边沿识别是为了找到算法关注的兴趣区域。本实验中的计算机显示器分辨率为 1024×768 。首先调整投影仪的投影区域,将其刚好完全覆盖电子书系统上的书模型;然后在屏幕上显示1个 1024×768 的白色像素图像,则在摄像头捕获的序列图像中书模型区域为白色,其他区域为黑色,如图4(a)所示。通过本文的序列图像处理算法对序列图像进行滤波处理,并对处理过的序列图像进行基于亮度的信息抽取,获取图像中像素点为白色的区域,从而可以找到图像中书模型的形状,其结果如图4(b)所示。边框线包围了整个仿真书的书面,该区域内即为本文算法所关注的兴趣区域,区域以外的部分不做处理。



(a) 1024×768 白色图像 (b) 屏幕区域的提取结果

图4 仿真书的边沿识别

系统标定是确定摄像头的图像坐标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仿真书模型上的坐标系。将计算机屏幕划分出 16×18 个点,并以这些点为中心画 3×3 的矩形框棋盘格,第1次在书模上这些点显示为白色,如图5(a)所示,其他显示黑色,通过基于亮点的信息抽取方法逐行扫描图像确定这些亮点在摄像头图像中的坐标;接着将这些点为中心的棋盘格小方块显示为黑色,其他像素点显示为白色,通过找如图5(b)所示的图像中的暗点校正之前所找到的这些点在摄像头中的位置,从而确定屏幕上指定的这 16×18

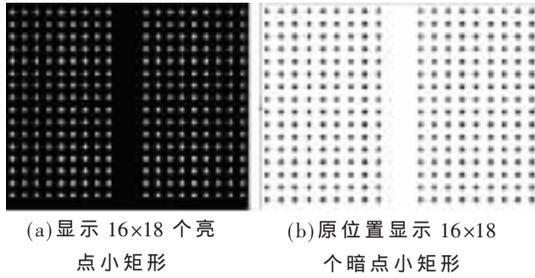


图5 系统标定

个点的屏幕坐标对应的摄像头坐标。

4 人手定位算法

在本系统中,利用图像采集装置监控人手的动作,故图像采集装置是静止的,而目标即人手是运动的。本系统采用基于帧间差分的方法来实现人手的定位。

帧间差分法就是将之前处理过的图像序列,通过逐个像素比较可直接求取前后2帧图像或者相邻几帧图像之间的差别。假设前后图像之间光线不变化,那么差分后图像不为零的像素表示该像素发生了移动;如果前后图像之间的光线有稍许变化,可以采用1个阈值来判断变化。帧间差分法就是利用图像序列中连续2帧或几帧图像的差异来进行运动物体的检测,即对图像序列中时间相邻的2幅或几幅图像求绝对差,然后用1个阈值来判断变化的区域。

假设第 k 帧和第 $k-1$ 帧图像的灰度分别为 $I_k(i,j)$ 和 $I_{k-1}(i,j)$,则其他的差分图像为:

$$D_k(i,j) = |I_k(i,j) - I_{k-1}(i,j)| \quad (1)$$

对式(1)得到的差分图像,通过1个阈值 T 来判断图像中的各个像素点是运动目标还是背景,从而提取出运动区域的图像 $M_k(i,j)$:

$$M_k(i,j) = \begin{cases} I_k(i,j), & D_k(i,j) \geq T \\ 0, & D_k(i,j) < T \end{cases} \quad (2)$$

为处理方便,本文将差分的结果二值化,这种二值化图像称为掩模,并对掩模图像进行连通性分析和数学形态滤波。当某一连通区域面积大于某一个给定的阈值时,则检测到目标,并认为该区域就是目标区域。在本算法中一般情况下将阈值 T 设置为30左右。

利用帧间差分法,通过前后2帧的图像分析人手运动方向,并利用系统标定将人手在摄像头中的坐标转化为屏幕坐标。设前1帧中人手的位置为 $P_1(x,y)$,后1帧中人手的位置为 $P_2(x,y)$ ($P_1(x,y)$ 、 $P_2(x,y)$ 均为屏幕坐标)。根据人手在前后2帧的位置比较可得出人手动作。

(1)前后翻页动作:根据 X 方向的位移大小来判断人手左右方向移动如下:

$$\begin{cases} P_{1,x} - P_{2,x} > 0, & \text{人手向左运动,电子书向后翻页} \\ P_{1,x} - P_{2,x} < 0, & \text{人手向右运动,电子书向前翻页} \end{cases} \quad (3)$$

(2)点击判断动作:设 $P_1, P_2 \dots P_n$ 为 n 幅序列图像中人手的位置,若点 $P_1, P_2 \dots P_n$ 的 X, Y 坐标相等,则认为

人手在此位置产生点击动作,此时电子书完成预先在此设置好的动作,如播放语音、视频等。

5 计算机视觉和图像合成算法

因人眼在某个视像消失后,仍可使该物像在视网膜上滞留0.1~0.4s左右。电影胶片以每秒24格画面匀速转动,一系列静态画面就会因视觉暂留作用而造成一种连续的视觉印象,产生逼真的动感。所以本系统根据此原理在前后2页内容更换的过程中插入24帧更换过程画面,则参观者就会看到翻页的动画效果。

图6是向后翻页过程图像示意图。数字1所在框为书的左页,数字2所在框为书的右页,数字3所在的三角形为翻上来的左页部分,而数字4则是翻上来的右页露出来的部分。 P 为书的右下角翻上来的位置。若整个翻页过程点 P 做弧形运动若干步,则由此会形成若干幅翻上来的画面,根据计算机视觉原理,如果在短时间内将这若干幅画面依照点 P 的移动顺序依次显示出来,则就会给人产生翻页的动画效果。

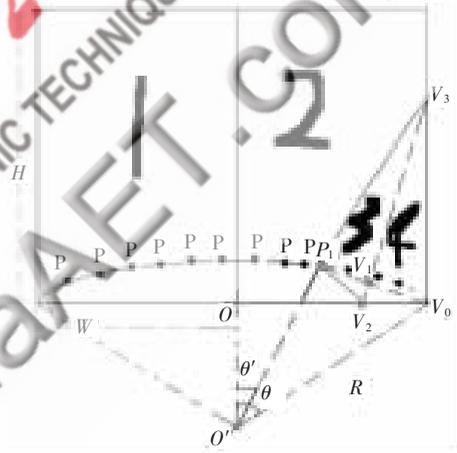


图6 动画产生示意图

利用计算机视觉算法显示翻页过程中的图像,需要对翻页中产生的图像进行实时的图像合成处理,该算法如下:

(1)假定翻页过程动画数为 N ,每次翻动相同的角度 θ_0 ,每页电子书页面的宽度为 W ,高度为 H ,以屏幕左上角为坐标原点 $(0,0)$ 。翻页的圆心角为 θ ,如图6所示,则可得 P 点运动轨迹所形成的圆弧半径 $R = W/\sin\theta$ 。图中, $OO' = W/\tan\theta$,每次翻页角度 $\theta_0 = 2\theta/N$,则图中的 $\theta' = \theta - i \times \theta_0$ 。由此得到每次翻页动画中的 $P_i(x,y)$ ($i=1,2,3 \dots N$)的坐标:

$$\begin{cases} P_i x = R \times \sin\theta' - W \\ P_i y = W - (R \cos\theta' - OO') \end{cases} \quad (4)$$

(2)将坐标系转换成以图6中所示的电子书底边中点 O 点为坐标原点,则 V_0 点坐标为 $(W,0)$, P_i 的坐标为 $(P_i x - W, H - P_i y)$, V_0 与 P_i 的中点 V_i 的坐标为 $(\frac{P_i x - V_0 x}{2}, \frac{P_i y - V_0 y}{2})$ 。 $V_2 V_3$ 的直线方程为:

$$y=k_1x+b_1 \quad (5)$$

式中, k_1 表示斜率, b_1 表示截距。由 P_i 、 V_0 可求得直线斜率 $k_1 = \frac{P_i \times x - V_0 \times x}{P_i \times y - V_0 \times y}$, 以斜率 k_1 、点 V_1 的坐标代入直线方程 $y=k_1x+b_1$ 求出截距 b_1 。因为 V_2 纵坐标也为 0, V_3 横坐标为 0, 利用 V_2V_3 直线方程可求出 V_2 、 V_3 对应的坐标。再利用两点确定一条直线, 可求出直线 P_iV_2 、 P_iV_3 的直线方程。

(3) 利用第 2 步求出的直线 V_2V_3 、 P_iV_2 、 P_iV_3 , 翻页过程中 P 点每向前移动 1 次, 判断每一个像素点属于图 6 所示的 1、2、3、4 四部分的所属部分并进行 1 次图像合成, 从而得到若干幅合成的图像, 同时依次显示到电子书模型上, 通过计算机视觉原理使人产生翻页的动画效果。图 7 是人手在书模型上挥动手臂电子书向后翻页过程中的 1 幅动画。



图 7 仿真书向后翻页过程动画

本文介绍了一种新的基于图像识别算法的虚拟翻书系统的设计, 实验所使用的计算机主频为 2.0 GHz,

图像采集设备为普通的网络摄像头, 投影设备为日立投影机。本系统运行时, 摄像头近似同步地采集分辨率为 640×480 像素的图像序列, 视觉翻书处理部分以每秒 24 帧的速度显示图像。本系统的特点在于利用图像采集装置(摄像头)获取图像数据来确定人的手势动作从而完成虚拟翻页效果, 同时准确定位人手位置, 具有虚拟触摸功能, 实现点击播放视频等。系统样机已成功运用于第四届中国中部博览会, 取得了良好的展示效果。但存在光线太强, 人手定位不理想等问题, 有待进一步改进。

参考文献

- [1] HALL D, GAL C L. Magicboard: a contribution to an intelligent office environment[J]. Robotics and Autonomous Systems, 1999, 35(3-4):211-220.
- [2] ZHANG Z, WU Y. 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.
- [3] 徐一华, 李善青, 贾云得. 一种基于视觉的手指屏幕交互方法[J]. 电子学报, 2007, 35(11):2236-2240.
- [4] 党韧. 一种多点触摸屏: 中国, 101271372A [P]. 2008-09-24.
- [5] 何斌, 马天予, 王运坚, 等. Visual C++ 数字图像处理[M]. 北京: 人民邮电出版社, 2001.

(收稿日期: 2009-10-12)

作者简介:

李炜, 女, 1969 年生, 博士, 副教授, 主要研究方向: 图像处理, 计算机视觉, 中文信息处理等。

李志, 男, 1985 年生, 硕士, 主要研究方向: 图像处理。

郭星, 男, 1983 年生, 硕士, 主要研究方向: 智能 EDA, 图像处理。