

基于视觉的媒体播放器控制算法

吴家洲, 江泽涛, 李志杰

(南昌航空大学 信息工程学院, 江西 南昌 330063)

摘 要: 采用有限状态自动机思想, 设计了通过动态手势来控制媒体播放器的算法. 首先, 将序列图像从 RGB 颜色空间转变到 YCrCb, 用大津法得到手势目标区域并计算目标中心; 接着, 把目标中心位置和帧延时作为状态机的输入, 播放器的功能作为其状态, 通过输入信号和当前状态确认各种手势命令; 最后调用 windows 系统函数 `keybd_event()` 实现播放器的暂停/继续、下一首/上一首和音量调节等功能. 实验表明, 本算法能实时有效地控制播放器, 状态机的应用很好地解决了 Midas Touch 问题.

关键词: 有限状态机; 动态手势; 大津法; 人机交互; 4-连通

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)02-0180-04

Media Player Control Algorithm Based on Vision

WU Jia-zhou, JIANG Ze-tao, LI Zhi-jie

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: Control algorithm of media player was proposed by dynamic hand gesture, using thinking in finite state automation. Firstly, sequences image was converted from RGB color space to YCrCb, in which gesture target region and it's center were obtained by means of otsu. Secondly, center of target region and frame delay time were input signal of finite state automation, different functions of media player were it's state, different gesture commands were decided by input signal and current state. Finally, functions of media player such as pause/continue, next/previous and volume adjustment were realized by windows system function `keybd_event()`. Experimental results show that media player can be controlled real-timely and effectively using the algorithm, Midas Touch program is perfectly solvable.

Key words: finite state automation; dynamic hand gesture; otsu; human computer interaction; 4-connectivity

1 引言

人与计算机的交互(HCI), 是机器视觉领域研究的关键问题, 而采用手势实现对机器的控制, 是目前研究的热点. 对手势理解的研究, 主要集中在手势建模、分析和识别等方面, 主要差别在于手势模型是以 3D 还是基于表观的建模; 3D 模型包含丰富的信息, 但建模过程比较复杂, 且计算量大, 不能满足实时的要求; 而表观的手势模型, 信息量较少, 适用于交流性手势, 容易实现^[1-3]. 文献[2]中叙述了“Midas Touch”(点石成金问题): 系统本身往往无法有效区

分哪些是用户真正的交互动作, 哪些仅仅是用户随意的下意识动作, 从而正确感知用户的交互意图; 文献[4-5]提供了手势识别与追踪问题的有效算法, 但没有对手势控制做具体的研究.

本文给出了从手势分割到手势控制播放器的完整算法.

2 控制算法

2.1 原理和流程

手势控制是一种非接触操作, 通过普通网络摄像头获得序列图像, 将每帧图像从 RGB 转变到

YCrCb 颜色空间,并提取 Cb 分量;在 Cb 空间,用大津法将手势目标从背景中分割出来,采用 4-连通算法对图像遍历,滤除噪声点,按目标区域的两组最大最小坐标值构造包围手势的最小矩形,并把矩形中心当作手势中心;根据每帧图像手势中心点在图像上的位置和帧延时来确认消息,调用系统函数实现播放器控制,其流程框图如图 1 所示。

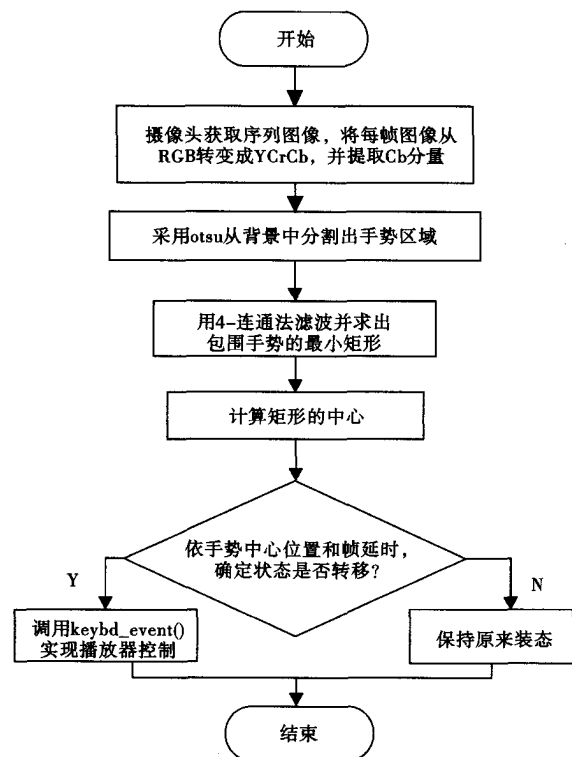


图1 手势控制算法流程框图

2.2 手势分割

手势分割的目的是为了获得手势的中心位置,作为有限状态机的输入。RGB 颜色空间转变成 YCrCb 颜色空间,Cr 分量表示红色分量和亮度的差值,Cb 分量表示蓝色分量与亮度的差值,它是一种线性变换,如式(1)所示。大津法是根据最小二乘原理推导出的最大类间方差法,是一种很好的自动选取阈值的方法,其基本思想是:选取的最佳阈值应是用该阈值分割得到的两类间具有最好的分离性,即类间特性方差最大或类内特性方差最小^[6]。

$$Y = 0.229R + 0.587G + 0.114B$$

$$Cr = -0.147R - 0.289G + 0.436G$$

$$Cb = 0.615R - 0.515G - 0.100B \quad (1)$$

大津法得到手势的大致区域,由于光照不均匀、背景中可能出现与肤色相近的物体、以及其他不可预见的原因,可能会出现如图 4~6 所示的孤立或者

小片的噪声点,会严重影响手势中心计算的正确性;采用 4-连通算法来解决这个问题;对整副图像进行遍历,计算所有的 4-连通区域,最大的连通区域即为手势区域,再计算包括这个区域的最小矩形和它的中心位置,把矩形的中心当作手的区域中心。

2.3 有限状态机算法

有限状态机是一种概念性建模工具,能根据当前状态来响应一个外部事件,是程序语言设计的一种处理方法,软件上也称为 FMM—有限消息机;状态机采取的操作不仅取决于接收到的事件,还取决于当前状态和各个事件的相对发生顺序,也就是说,一个事件的响应不仅取决于事件本身,还取决于其内部(当前)状态,同时还会决定并更新它的当前状态,任何逻辑都可建模成一系列事件和状态的组合,其状态机编程模型如图 2 所示。

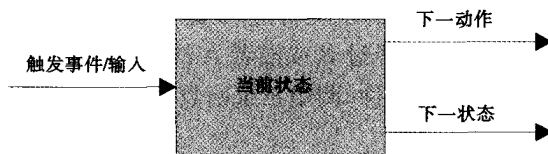


图2 状态机编程模型

根据图 2 模型把手势中心位置和帧延时作为状态机的触发事件或输入,播放器的暂停/继续、下一首/上一首和音量调节等控制功能作为状态机的状态,状态机根据触发事件和当前状态决定下一动作和下一个状态,最后通过调用系统函数来实现播放器功能控制。

采用自动机编程思想,将经过预处理后的视频帧划分成不同功能区,如图 3 所示;区域 1~6 为手势输入区,前面计算的手势中心作为输入信号,A 线和 B 线只在上一首/下一首控制操作时有效,图中中部序号为 0 的区域只在音量调节时起作用,其他时候不接收输入;图下部序号为 0 的区域为功能扩展区,输入信号在该区域无效。

控制算法描述:

播放控制包含三个过程:进入新状态、确认状态改变和发送指令改变当前状态;根据手势中心在 1~6 号区域的位置和帧延时情况,确认播放器是否进入新状态、是否确认状态改变,一旦状态改变确认,就调用系统函数来实现。

任何一个状态改变在确认过程中如果被终止,则进入初始状态。

(1) 1~3 号区域(包括 A 线和 B 线)控制上一首/下一首调节:当中心点进入区域 1 或 3 并经过 A

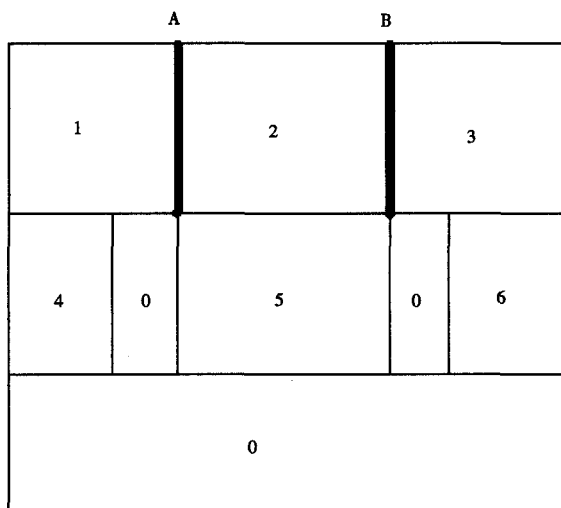


图3 视频帧区域划分

线或者 B 线,表示系统将进入上一首/下一首的状态,中心点继续经过 B 线或者 A 线,确认状态改变,触发状态改变事件,调用系统函数实现上一首/下一首的转换;

(2) 第二行的 4、0、5、0 和 6 是音量调节作用区:中心点在区域 4 或 6 表示系统将进入音量调节状态,在 4 或 6 区停留 10 帧后确认为音量控制手势,手在 4 至 6 号区域间移动时,向媒体播放器发送音量调节控制指令,目标中心离开中间区域,音量调节终止;

(3) 5 号区为暂停/继续播放手势触发区:中心点停留在 5 号区表示系统将进入暂停/继续的状态,中心点在此停留 10 帧后确认为暂停/继续播放控制手势,并同时向媒体播放器发送暂停/继续播放命令,但不会连续触发事件,只有当手势离开 5 号区域并再次进入该区域,才能再次触发其状态的变化。

3 实验及结果分析

根据上面描述的算法,以酷狗播放器为例,实现对其暂停/继续,上一首/下一首和音量调节等功能。

播放器对应快捷键如下:

暂停/继续播放: F5

下一首: Ctrl+Right

上一首: Ctrl+Left

减小音量: Ctrl+Down

增大音量: Ctrl+Up

快捷键对于播放器的控制采用的是消息传递机制,在程序中通过 windows 提供的系统函数 `keybd_event()` 来实现,其函数原型如下:

```
void keybd_event(
    BYTE bVk, // virtual-key code
    BYTE bScan, // hardware scan code
    DWORD dwFlags, // function options
    ULONG_PTR dwExtraInfo // additional key-
    stroke data
);
```

该函数发送键盘消息到系统键盘队列,当前活动窗体接收到消息后发生响应,从而实现控制。

在本例,发送 Ctrl+Left 消息可以使用下面函数:

```
keybd_event(VK_CONTROL, EYEVENTF_
EXTENDEDKEY, 0);
keybd_event(VK_LEFT, 0, KEYEVENTF_
EXTENDEDKEY, 0);
keybd_event(VK_LEFT, 0, KEYEVENTF_
EXTENDEDKEY | KEYEVENTF_
KEYUP, 0);
keybd_event(VK_CONTROL, 0, KEYEVENTF_
EXTENDEDKEY | KEYEVENTF_
KEYUP, 0);
```

图 4~6 是用手势控制酷狗媒体播放器的对比图片,图左半部分是计算机处理的图像区域,右半部分是与左半部分相对应时刻播放器的播放界面;左半部分中白色区域为手的区域,白色区域中的符号“×”是手的计算中心。

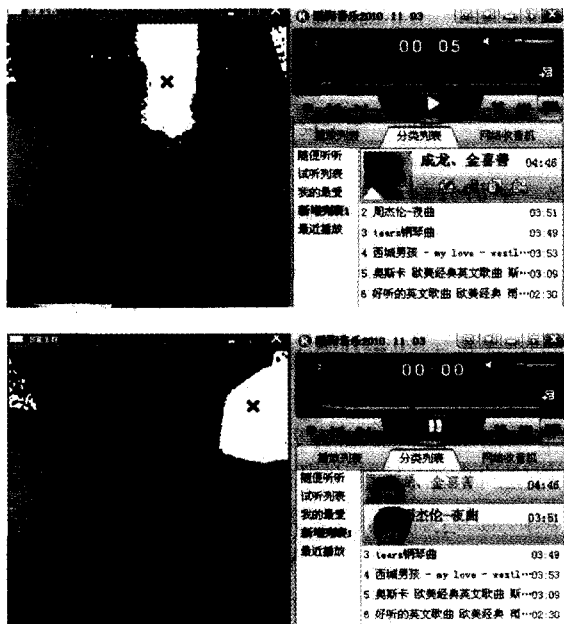


图4 上一首/下一首调节对比图

从图 4 可以看出:上图是手势从 1 号区经过 A 线到达 2 号区,播放器正常播放当前曲目,下图是手

势从2号区经过B线到达3号区,可以明显地看到曲目从下一首到上一首翻转的瞬间过程,事件响应完毕后,曲目实现转换。



图5 音量调节对比图



图6 暂停/播放调节对比图

由图5可知:上图是手势进入4号区的瞬间,10个序列帧延时过后触发音量调节事件,下图播放器是手势依次经过左边4、0、5、0到达右边6区,音量减小。

图6说明:上图是手势进入5号区的瞬间,播放器正常播放,下图显示出手在该区域停留10帧后播放器转换为停止状态;可以看出手形略有不同,但这并不影响手势的控制作用。

4 结束语

(1) 采用有限状态机思想,设计了通过动态手势来控制媒体播放器算法,算法简单高效,鲁棒性好,对于解决人机交互中的“Midas Touch”问题,提供了新的思路和解决方法;

(2) 手在运动过程中可以任意改变其姿态,不需要进行手势识别,既减少了由于模板匹配所消耗的时间,同时降低了操作要求;

(3) 采用图像处理中的4-连通法对图像进行遍历,求解手的目标区域,排除了视频帧中与肤色相近像素的干扰。

参考文献:

- [1] Vladimir I Pavlovic, Rajeev Sharma and Thomas S. Visual interpretation of hand gestures for human-computer interaction: A review [J], IEEE Transactions on Pattern Analysis and Machine, 1997, 19(7): 677-694.
- [2] 武汇岳, 张凤军, 刘玉进, 等. 基于视觉的手势界面关键技术研究 [J], 计算机学报, 2009, 32(10): 2030-2041.
- [3] 任海兵, 祝远新, 徐光佑, 基于视觉手势识别的研究——综述 [J], 电子学报, 2000, 28(2): 118-121.
- [4] Lars Bretzner, Ivan Laptev, Tony Lindeberg, Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering [C] // Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002: 63-74.
- [5] Cristina Manresa, Javier Varona, Ramon Mas et al. Hand tracking and gesture recognition for human-computer interaction [J]. Electronic Letters on Computer Vision and Image Analysis, 2005, 5(3): 96-104.
- [6] Otsu N. A threshold selection method from gray2level histograms [J]. IEEE Trans 1979, SMC, 1979, 9(1): 62-66.

作者简介:

吴家洲 (1973—), 硕士, 讲师. 研究方向为图像处理模式识别和计算机视觉。

江泽涛 博士, 教授, 博士生导师. 研究方向为图像处理模式识别和计算机视觉。