

基于双目视觉的低复杂度高精度手势提取算法

吕俊宏¹, 任屹奕¹, 谢 翔¹, 李国林²

(1 清华大学 微电子学研究所, 北京 100084; 2 清华大学 电子工程系, 北京 100084)

摘 要: 为满足实时人机交互的需求, 解决在不同光照、不同背景下, 与肤色无关的手势提取问题, 提出了一种低复杂度、高精度、适合硬件实现的手势提取算法. 该算法对双目图像基于逐行扫描的机制提取手势. 首先检测边缘, 由边缘对图像形成分割, 获得手势轮廓, 随后通过匹配提取出手势, 最后消除背景噪声的影响, 获得准确的手势. 实验采用两个相同型号的网络摄像头在室内光照条件下采集图像. 结果表明, 该方法在复杂背景中能够快速准确地提取出各种手势, 相比传统的基于肤色的手势提取方法干扰更少.

关键词: 手势提取; 低复杂度; 高精度; 双目视觉

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-7180(2014)01-0120-04

A Low Complexity and High Accuracy Algorithm of Gesture Extraction Based on Binocular Stereo Vision

LV Jun-hong¹, REN Yi-yi¹, XIE Xiang¹, LI Guo-lin²

(1 Institute of Microelectronics, Tsinghua University, Beijing 100084, China;

2 Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: A gesture extraction algorithm with low complexity and high accuracy is presented in this paper. This algorithm meets the needs of real-time human-computer interaction, solves the problem of gesture extraction which has nothing to do with skin color in different light conditions and different backgrounds. And it is suitable for hardware implementation. The algorithm extracts the gesture based on the row-by-row basis in binocular images. Firstly, edges in both images are detected and images are segmented by the edges so that the gesture contour can be obtained. Secondly, the gesture is extracted by a matching algorithm. Finally, background noise are eliminated and accurate gesture comes out. In this paper, the binocular images are captured by two same web cameras in the indoor lighting conditions. The experimental results show that the algorithm is able to extract various gestures quickly and accurately in complex background, and is more robust compared to traditional gesture extraction algorithm based on skin color.

Key words: gesture extraction; low complexity; high accuracy; binocular stereo vision

1 引言

手势作为人类表达思想和情绪的重要方式, 在人机交互领域的扮演着重要角色. 手势识别技术在这种背景下应运而生, 多种基于视觉的手势识别方法已经被研究人员提出^[1-2]. 在诸多方法中, 精准的手势识别都需要以正确的手势分割作为基础.

现有的手势分割提取方法往往利用了人的肤色

作为特征, 采用肤色阈值分割的方法来得到手势^[3-4], 然而这种方法受到背景以及环境光照的影响较大, 使用者肤色的不同也会给这种方法检测手势带来困难. 另有研究人员提出利用深度图像信息来精确地获取手势^[5-6], 这种方法适用于复杂背景, 多种光照环境, 但深度图像的采集使用了 Kinect, 无疑增加了系统的复杂度, 不利于系统的便携性.

本文主要致力于研究利用双目视觉的特点进行

手势提取,同时考虑系统的实时性,提出一种基于图像逐行扫描机制、适合硬件实现的手势提取算法。

2 基于逐行扫描的手势提取算法

2.1 算法总体框架

该算法同时对经过校正的双目手势图像进行逐行扫描.首先利用 RGB 分量的变化趋势检测边缘,由边缘信息将不同物体进行分割,获得手势轮廓;随后将分割出的前景物体(手势)进行分段匹配;再利用图像形态学中的腐蚀膨胀操作消除背景噪声的影响,提取出精确的手势。

2.2 基于边缘检测的图像分割算法

图像中不同颜色的物体交叠处存在人眼可以明确判断的边缘,检测出边缘即可将不同的物体分割开.传统的边缘检测算法大多基于灰度图像进行处理,将灰度值不连续处视为边缘.然而在不同的光照环境下,颜色有明显区别的物体边缘可能存在着灰度值缓慢变化甚至不变的现象,如图1中方框所示的区域.图2所示为图1方框中一行像素点的 RGB 分量值以及其灰度值,从图中不难看出,手势与背景交界处 RGB 三个分量的变化趋势有明显差异,而在同一个位置,灰度值几乎没有变化.在这种情况下,一般的基于灰度图像的边缘检测算子(例如 Sobel 算子)就难以得到准确的边缘,如图3所示。



图1 方框内区域颜色变化明显而亮度变化不明显

因此本文提出一种基于图像 RGB 分量变化趋势的边缘检测算法.该算法将颜色的不连续定义为物体的边缘,颜色的不连续在图像中反映为 RGB 分量的变化趋势不同.数据的变化趋势可以由导数方便的表示,在数字图像中可以借助空域差分近似空域微分,由梯度 ∇I_c 代替一阶导数来表示:

$$\nabla I_c(x) = I_c(x+1) - I_c(x) \quad (1)$$

$c \in \{R, G, B\}$

式中, $I_c(x)$ 表示图像某一行中横坐标为 x 的像素点的 RGB 分量值。

随后利用 RGB 分量两两之间上述变化趋势的差值 D ,即可反映出图像颜色的不连续性.三个分量

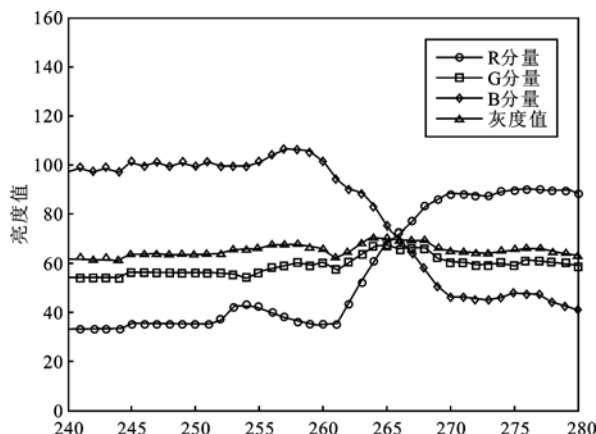


图2 亮度缓慢变化处 RGB 分量变化趋势不同



图3 利用 Sobel 算子检测到的边缘

的变化趋势两两相减后取绝对值,共有三个差值,取其中最大的差值作为该像素点处的颜色不连续性:

$$D(x) = \max_{c1, c2 \in \{R, G, B\}, c1 \neq c2} (|\nabla I_{c1}(x) - \nabla I_{c2}(x)|) \quad (2)$$

不连续性 D 大于预先设定的阈值的像素点即被认定为边缘点。

然而由阈值判断边缘存在一定的局限性,阈值的选取需要通过大量的图像数据来确定,而光照等环境条件的变化又极易导致该阈值失效,同一幅图像中不同局部的边缘对应的阈值可能也会有较大差异.因此,本文将阈值判断法改进为检测局部极大值的方法,即对上述不连续性 D 进行高斯滤波,再选取其局部极大值点对应的像素为最终的边缘点。

该算法有效地解决了基于灰度值的边缘检测算法对于颜色不同而灰度值变化不明显的边缘不敏感的问题,对图1所示图像的整体边缘检测效果如图4所示.该算法的另一优势是能够在逐行扫描的机制下检测出图像边缘,便于硬件实现.逐行扫描带来的效果是对于垂直边缘有较强的响应,而对于水平边缘检测效果一般.后采用的匹配算法亦是逐行进行,故水平边缘的缺失对于最终手势提取效果的影响可以忽略。

2.3 分段匹配的手势提取算法

用户采用手势与带双摄像头的系统进行人机交



图4 利用图像 RGB 分量变化趋势检测到的边缘
互时,手的位置位于摄像头拍摄到的所有场景物体



图5 双目图像中手势的分段方式一致

该算法从左到右对右视点中的每一个分段在左视点的对应行中寻找匹配段. 由于手势距离摄像头较近,在左右视点中的视差较大,故搜索范围可以限制在较大的视差范围内,降低了算法的计算量.

匹配准则分别考虑了分段长度与分段颜色. 其一是两段的长度差必须小于阈值. 造成长度差异的原因主要是两个摄像头的视角不同,拍摄到手势的不同侧面. 该阈值的选取与分段的长度有关,一般设为进行比较的两个分段中较短分段长度的 20%:

$$|\text{length}^L - \text{length}^R| < \min(\text{length}^L, \text{length}^R) \cdot 20\% \quad (3)$$

其二是两段居中的三个像素点的颜色差异不能超过阈值. 造成颜色差异的因素有两个摄像头成像元件的不完全相同以及外部环境光照的影响. 通过对不同光照条件下的图像进行分析,发现每个像素的 RGB 三个通道的亮度值差异均不应超过 25,即:

$$\sum_{x=M-1}^{M+1} \sum_{c \in \{R, G, B\}} (I_c^L(x) - I_c^R(x)) < T = 25 \times 3 \times 3 \quad (4)$$

式中, M 为分段中居中像素点的横坐标.

3 实验结果与分析

本文的实验使用两个相同型号的罗技网络摄像头在室内光照、复杂背景条件下拍摄手势图像,图像分辨率为 640×480 ,手势距离摄像头 30 cm 到 80 cm 之间,模拟近距离的手势识别应用. 获得的手势

的前方,手势在左右视点中的视差大于其他物体. 同时手势区域在场景中属于前景,不受任何遮挡,两个摄像头拍摄到的手势基本一致. 利用这一特性,将场景中的前景物体进行匹配,即得到手势.

本文提出一种逐行对图像进行分段匹配的手势提取算法. 前文所述基于边缘检测的图像分割算法将一行图像中的场景分割为若干段,如图 5 所示,且左右视点图像的对应行中手势的分段方式一致,故左右视点中的手势分段之间具有一一对应的关系.

提取结果如图 6 所示. 由实验结果图像可以看出,本文的算法在复杂的背景下可以准确地分割提取出不同的手势,且不受与肤色相近的背景干扰.

该手势提取算法产生误差的原因主要有左右两个视点中检测到的边缘不完全相同,导致两个视点中分割出的段长度区别较大,无法正确匹配,造成手势中某一行信息的丢失. 边缘不完全相同的问题是由于环境光照影响导致左右视点中手势图像的颜色不完全相同,检测到的边缘亦产生了区别. 由于在整个手势区域中只有少量几行出现了上述问题,利用数学形态学中的膨胀操作即可圆满解决. 另外还有少量背景图像带来的干扰,这些干扰则可通过腐蚀操作来消除.

4 结束语

本文提出了一种对双目图像基于逐行扫描机制提取手势的算法,适合硬件实现. 实验结果证明,该算法在复杂背景下准确地提取出了多种代表不同含义的手势,精确度能够满足手势识别应用的需求. 该算法提取近距离手势的能力可以被应用在穿戴式智能设备(例如 Google Glass 等)中,其适合于硬件实现的特点使得设备的小型化便携性要求能够得到满足.

本文进一步的工作将研究更加多样化的手势提取,同时提出同样适合硬件实现的手势识别方法,并在硬件平台上验证其性能.

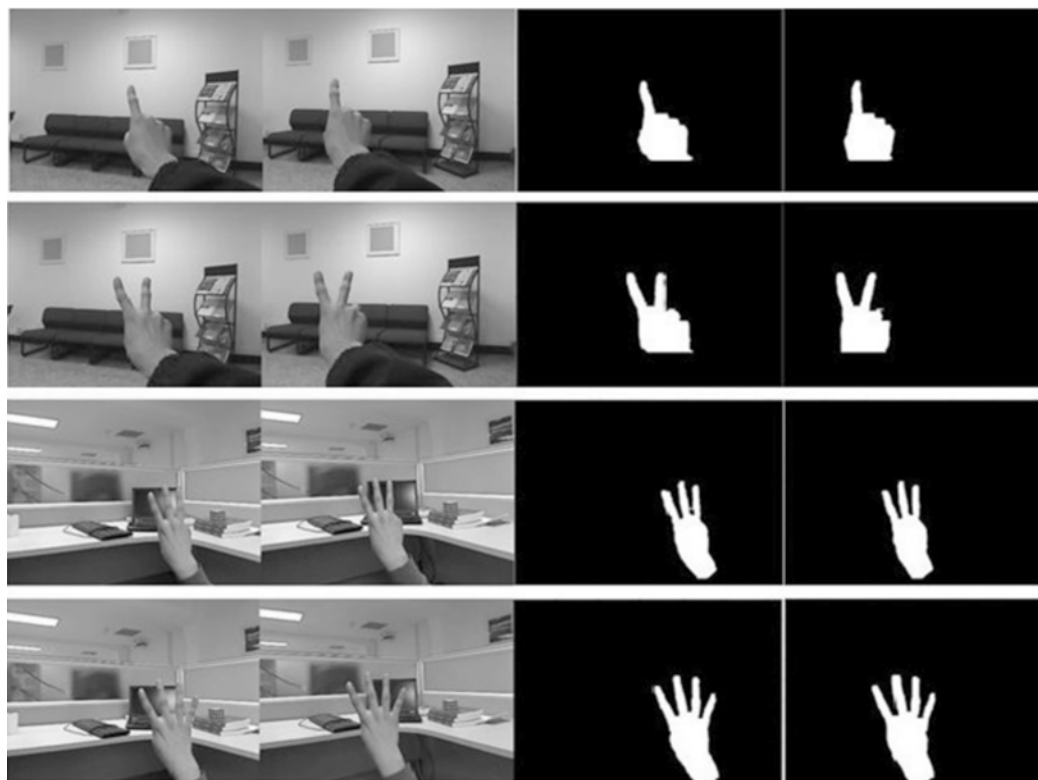


图6 使用本文算法提取出的手势图像

参考文献:

- [1] Erol A, Bebis G, Nicolescu M, et al. Vision-based hand pose estimation: a review[J]. Computer Vision and Image Understanding, 2007(108):52-73.
- [2] 任海兵,祝远新,徐光佑,等. 基于视觉手势识别的研究-综述[J]. 电子学报,2000,28(2):118-121.
- [3] Wachs J P, Kolsch M, Stern H, et al. Vision-based hand-gesture applications[J]. Communications of the acm,2011,54(2):60-71.
- [4] 雷海龙,杨胜,刘红新. 融合图像轮廓矩和傅立叶描叙子的手势识别[J]. 微电子学与计算机, 2011(4): 103-105.
- [5] Suarez J, Murphy R R. Hand gesture recognition with depth images: A Review[C]//IEEE RO-MAN. Paris:IEEE,2012:411-417.
- [6] 曹维清,李瑞峰,赵立军. 基于深度图像技术的手势识别方法[J]. 计算机工程,2012,38(8):16-21.

作者简介:

吕俊宏 男,(1987—),硕士研究生. 研究方向为 3D 图像处理.

任屹奕 男,(1987—),博士研究生. 研究方向为集成电路设计、图像处理.

谢翔 男,(1971—),副研究员. 研究方向包括集成电路设计、图像处理、无线通信、生物医学电子学等.

李国林 男,(1970—),副教授. 研究方向包括互连分析以及 RFIC 设计.