

一种快速鲁棒的动态人手跟踪方法

吕金刚¹, 吴斌¹, 刘恒¹, 周翔宇¹, 邵延华²

(1.西南科技大学 信息工程学院, 四川 绵阳 621010;

2.重庆大学 光电工程学院 光电技术及系统教育部重点实验室, 重庆 400030)

摘要: 介绍了一种基于肤色与手形的动态人手跟踪方法。根据检测过程中所用到的人手基本特征, 提出了一种基于色度-饱和度自适应阈值肤色分割与成对几何直方图(PGH)匹配的人手图像检测方法。为了克服跟踪过程中背景色干扰的问题, 实现了基于 Kalman 滤波器的预测跟踪。实验结果表明, 该动态人手跟踪方法计算简便、实时性好、跟踪精度高, 可应用在大多数复杂场景。

关键词: 肤色概率直方图; 成对几何直方图; 卡尔曼滤波器

中图分类号: TP399

文献标识码: A

文章编号: 1674-7720(2012)01-0033-04

Fast and robust approach to tracking moving hand

Lv Jin'gang¹, Wu Bin¹, Liu Heng¹, Zhou Xiangyu¹, Shao Yanhua²

(1.School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China;

2.Key Laboratory of Optoelectronic Technology and Systems of the Education Ministry, College of Opto-Electronic Engineering, Chongqing University, Chongqing 400030, China)

Abstract: This paper presents an approach of moving hand tracking based on skin-color and hand shape. It proposes a method to detect moving hand based on chromatic adaptive thresholding and PGH matching. In order to solve the problem of the disturbances from background color, it achieves a predictive tracking method based on Kalman filter. The results show that the method is simple and accurate, and can be used in most complex scenes.

Key words: skin probability distribution histogram; pair-wise geometrical histogram; Kalman filter

随着计算机性能的不提高和机器视觉领域研究的不断深入, 人机交互技术日益成为机器视觉领域的研究热点。基于机器视觉的手势识别作为其重要的研究内容之一, 在许多领域有着重要的应用。最近几年, 尤其在智能家居以及 3D 体感游戏领域, 手势识别得到了广泛的应用, 因此受到越来越广泛的关注。

根据识别特征的不同, 手势识别分为基于标记的手势识别和基于裸手的手势识别系统, 而裸手手势识别又分为静态手势识别和动态手势识别^[1]。目前, 基于裸手的动态手势识别是手势识别中最重要的领域^[2], 相对于静态手势识别而言, 动态手势识别有着更灵活、更广泛的应用。但是人手是一个多关节非刚性物体, 在运动过程中会呈现出多种多样的手势形态, 在跟踪过程中还会伴有手势遮挡和肤色干扰等情况。所以, 如何对手进行有效的视觉检测与跟踪是实现动态手势识别和理解

的关键, 也是机器视觉机和人机交互领域中一个富有挑战的重要课题。

1 相关工作

按照运动目标跟踪策略的不同, 对于动态人手跟踪可以分为基于模型和基于表象两种方法。

基于表象的方法又叫非模型跟踪方法, 它可细分为基于区域的跟踪、基于活动轮廓^[3]以及基于其他图像特征的跟踪方法^[4]。该方法的优点是实时性较好, 但由于所跟踪特征很容易受遮挡或光线条件变化等外界因素的干扰, 难以保证鲁棒性。基于模型的方法要对跟踪目标建立特征模型, 通过比较观测和预测数据来修正模型中的参数。该类方法性能的实现依赖于初始化过程, 如果初始化合理, 即可达到精确的估计。同时该类方法还面临着高维空间中搜索的难题, 如采用粒子滤波器跟踪人手运动, 当状态空间维数过高时, 需要大量的粒子来

估计高维空间中的概率分布。因此这种方法计算复杂度高、计算量较大,很难保证实时性。

针对以上问题,提出了一种快速鲁棒的人手检测与跟踪算法,在背景相对复杂、存在类肤色干扰的情况下,该方法也能够有效地对动态人手进行实时跟踪。具体框图如图 1 所示。

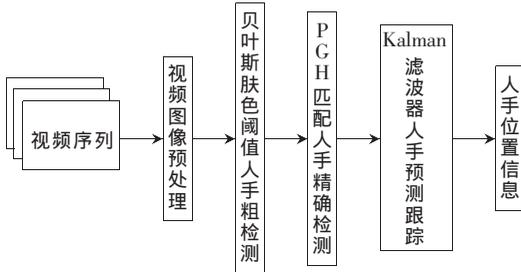


图 1 快速鲁棒的人手检测与跟踪方法框图

2 基于贝叶斯肤色模型的人手粗检测

色彩信息是物体重要的特征信息。利用肤色信息可以快速确定人手所在的候选区域,以缩小检测范围。HSV 色彩空间是从 RGB 空间转化而来的,比较接近人类的色彩感知习惯,适合建立肤色模型^[5],采用 HSV 空间二维彩色直方图,可以去除部分光照的影响,适合对颜色特征稳定的跟踪。

基于直方图的贝叶斯判决方法 SPM (Skin Probability Map) 是目前广泛采用的一种肤色检测方法,它利用贝叶斯准则将肤色直方图和非肤色直方图归一化,以概率来判断像素是否属于肤色像素,如式(1)所示。

$$P_{(skin|x)} = \frac{P_{(x|skin)}P_{skin}}{P_{(x|skin)}P_{skin} + P_{(x|nonskin)}P_{nonskin}} \quad (1)$$

其中, $P_{(x|skin)}$ 和 $P_{(x|nonskin)}$ 分别表示颜色值为 x 的像素属于肤色和非肤色像素的先验概率, P_{skin} 和 $P_{nonskin}$ 分别表示肤色区域与非肤色区域的概率。当后验概率大于某固定阈值时,则该像素为肤色像素。该方法的缺点在于,采用阈值固定分割不能很好地适应所有图像内容。

张明吉等人^[6]在 SPM 的基础上对肤色概率分布直方图 SPDH (Skin Probability Distribution Histogram) 进行了改进,用肤色概率的分布概率作为 SPDH 的横坐标,横坐标右面的值就是要找的阈值点,如式(2)所示。

$$P_x = \frac{P_{(skin|x)}}{\max_x \{P_{(skin|x)} | x \in I\}} \quad (2)$$

SPDH 的纵坐标表示具有相同归一化肤色概率的像素数,如式(3)所示。

$$spdh_x = \frac{\sum \delta(p_x - p_y)}{\sigma} \quad (y \in I) \quad (3)$$

其中, σ 是为了避免肤色和非肤色区域的重叠,它是所有这些像素的颜色均方差。

洪红等人^[7]提出了一种色度 SPDH 的算法,并提出了式(2)的改进公式,如式(4)所示。

$$P_x = \frac{\text{num}(P_{skin|x})}{\sum \text{num}(P_{skin|x})} \quad (P_{skin|x} > 0.5) \quad (4)$$

由式(4)可知,后验概率较大的点有可能是最佳阈值点。因此,可以设计一个贝叶斯分类器,输入每个候选阈值点,输出值表示候选点是否为最优阈值点。本文根据 SPDH 的 4 条线索^[6],提出最优阈值贝叶斯分类器:

$$P_{(optimal|x)} = \frac{P_{(x|thr)}P_{thr}}{P_{(x|thr)}P_{thr} + P_{(x|nonthr)}P_{nonthr}} \quad (5)$$

其中, $P_{(x|thr)}$ 和 $P_{(x|nonthr)}$ 分别表示候选阈值 x 是阈值和不是阈值的先验概率, P_{thr} 和 $P_{(nonthr)}$ 分别表示阈值概率和非阈值概率, $P_{(optimal|x)}$ 表示候选阈值 x 为最优阈值的后验概率。

最后,可以得到经过肤色分割的二值化图像,部分结果如图 2 所示,其中图 2(a)、图 2(b)、图 2(c)为原图像,图 2(d)、图 2(e)、图 2(f)为对应的二值化图像。

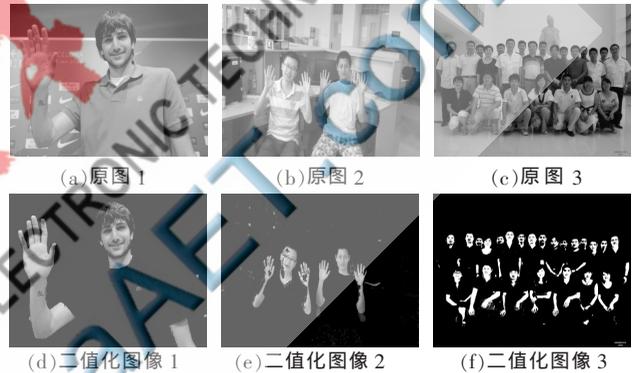


图 2 基于贝叶斯肤色阈值分割的部分图片

3 PGH 人手精确定位

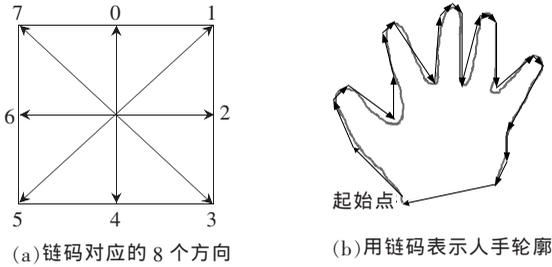
根据肤色分割结果,将视频图片进行二值化处理,肤色区域为白点(255),非肤色区域为黑点(0),缩小了检测范围。由于人体其他肤色部分以及复杂背景下类肤色的影响,以上方法所分割到的肤色区域可能包含人手以外的肤色部分,因此,需要去除这些“混淆”区域,以便精确地提取出人手区域。

3.1 人手轮廓特征描述

轮廓是点的序列,可以用来表示图像中间的曲线^[8]。二值图像轮廓可用于提取图像中的目标区域,以便对目标区域作进一步处理,如区域填充、计算轮廓长度及面积、特征提取和轮廓匹配等。轮廓可以采用链码和线段表两种不同方法来描述。在链码中,多边形被表示为一些列的位移,每一个位移都有 8 个方向,这 8 个方向使用整数 0~7 来描述,如图 3(a)所示。则图 3(b)中人手轮廓对应的链码可以表示为:7-7-1-3-0-7-1-3-0-1-3-4-1-2-3-5-4-1-3-5-4-5-6。

3.2 人手轮廓匹配

PGH 实际上是链码直方图的一个扩展和延伸^[9]。人手的外形是不规则形状,可以将其描述为许多线段组成



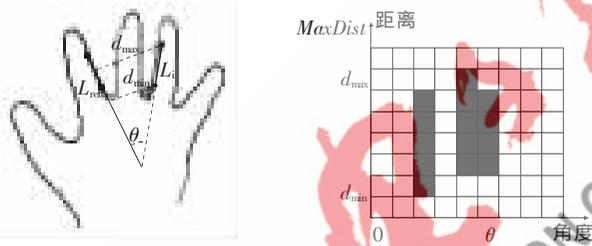
(a) 链码对应的 8 个方向

(b) 用链码表示人手轮廓

图 3 链码表的 8 个方向及用链码表示的人手轮廓

的多边形。由于多边形 PGH 基于多边形线段间距离与角度信息,因此这种编码方式具有很好的平移、旋转和尺度不变性,而且计算简单,只需要少量的手势模板即可达到很好的匹配效果^[9]。

本文将提取到的人手轮廓近似为多边形,这样就可以用一系列连续的边缘点来表示人手轮廓,定义其边缘点的集合为 $(x(t), y(t)) \in R^2$ 。将每一个轮廓边缘视为其方向上的基准线,那么它与其他轮廓边缘间的相对角 $\theta \in [0, \pi]$ 以及最大、最小垂直距离 (d_{\min} 和 d_{\max}) 可以计算出来,如图 4(a) 所示。定义边缘之间的相对角度为直方图的行,对应的相对最大距离与最小距离为直方图的列,这样就得到了当前人手多边形的 PGH,如图 4(b) 所示。



(a) 两条线段间的相对角与垂直距离

(b) 成对几何直方图

图 4 两条线段间的相对角与垂直距离以及成对几何直方图

同样,利用相同的方式可以获得手势模板的 PGH。将两个直方图归一化,假设其分布分别为 $p_c(\theta)$ 和 $q(\theta)$,其中 c 表示候选区域中心。应用 Bhattacharyya 系数计算两个分布间的相似度 $\rho(q, p_c)$,然后计算 Bhattacharyya 距离,得出一个描述其相似性的数据 $p_{\text{match}} (0 \leq p_{\text{match}} \leq 1)$:

$$p_{\text{match}} = d_{\text{Bhattacharyya}} = \sqrt{1 - \rho(q, p_c)} \quad (6)$$

p_{match} 的值越小,手形的匹配度越高,即当前手形与模板手形的相似度越大。通过多次实验确定,PGH 匹配能够准确去除人手混淆区域,精确定位人手所在区域。

4 Kalman 动态人手跟踪

复杂背景下的人手跟踪往往需要一些具有预测功能的跟踪算法。当人手肤色区域与“混淆”区域重叠面积较大时,普通的跟踪方法很难准确地定位人手位置。Kalman 滤波器是一个对动态系统的状态序列进行线性最小方差估计的算法,具有计算量小、可实时计算的特点,而且能够较为准确地预测目标的位置和速度。

如果短时间内将人手运动看作是一种近似恒定运动,即将将人手的运动状态在该时间序列内看作时

不变系统,则可以设计一个 Kalman 跟踪器对人手进行预测跟踪。Kalman 系统状态模型如式(7)和式(8)所示。

状态方程:

$$X(k+1) = F(X(k), u(k)) + \omega_k \quad (7)$$

观测方程:

$$Z(k+1) = H(X(k+1)) + v_k \quad (8)$$

其中, F 和 H 分别为状态转移矩阵和观测矩阵。 ω_k 和 v_k 分别为系统噪声和观测噪声,分别服从方差为 R 和 Q 的正态分布。 $u(k)$ 是 k 时刻对系统的控制量,本文中 $u(k)$ 可以忽略不计。用 X 表示人手状态向量, (x, y) 和 (\dot{x}, \dot{y}) 分别表示人手特征点在图像中的坐标位置和运动速度,则 $X = (x, y, \dot{x}, \dot{y})^T$ 。 $Z = (x, y)^T$ 表示观测向量,即人手的位置信息。

设计 Kalman 滤波方程对系统参数进行估计,如式(9)和式(10)所示。

预测:

$$X(k+1) = F(X(k)) + \omega_k \quad (9)$$

$$P(k+1|k) = P(k|k) + Q_k$$

更新:

$$Kg(k+1) = P(k+1|k) \cdot F^T(k+1) \cdot [F(k+1) \cdot P(k+1|k) \cdot F^T(k+1) + R(k+1)]^{-1}$$

$$X(k+1|k+1) = X(k+1|k) + Kg(k+1)[X(k+1) - F(X(k+1|k))] \quad (10)$$

$$P(k+1|k+1) = [I - Kg(k+1)F(k+1)] \cdot P(k+1|k)$$

其中, T 为相邻帧之间的采样周期, ω_k 和 v_k 的协方差矩阵 $Q = I_1$, $R = I_2$, I_1 为四维单位矩阵, I_2 为二维单位矩阵。

5 实验结果分析

在训练 SPM 时,采用 6000 幅样本图片,其中 4500 幅通过网络摄像机采集,1500 幅来自于因特网。训练样本 5000 幅,包含肤色图片 2000 幅,非肤色图片 3000 幅。测试样本集 1000 幅,包含肤色图片 400 幅,非肤色图片 600 幅。

5.1 实验结果

对该方法的检测与跟踪性能做了测试,利用普通 USB 摄像头制作了两段视频序列。运行环境为 2.86 GHz 奔腾 E6300 双核处理器,2 GB 内存,Windows7 操作系统,VS2008+Opencv2.1 开发平台,视频窗口显示大小为 320 pix×240 pix,灰色矩形框为跟踪到的手势区域,图 5 给出了部分实验结果截图。



(a) 第 60 帧 (b) 第 82 帧 (c) 第 97 帧 (d) 第 145 帧 (e) 第 182 帧 (f) 第 218 帧

(a) 简单背景下跟踪实验结果



(a) 第 48 帧 (b) 第 92 帧 (c) 第 133 帧 (d) 第 185 帧 (e) 第 211 帧 (f) 第 245 帧

(b) 复杂背景下跟踪实验结果

图 5 实验结果

5.2 对比实验分析

为了验证本文方法的快速性与鲁棒性,在背景环境相对复杂的场景下,分别与 Camshift+Kalman 方法和 Camshift+Particle filter 方法、基于 Haar 特征的 Adaboost 方法和 Chamfer Matching 方法进行了对比性实验测试,实验运行环境同上。测试结果如表 1 所示。

表 1 对比实验测试结果

方法	性能		
	帧率/ (frame/s)	误检率 $frame_{\text{err}}/frame_{\text{sum}}$	漏检率 $frame_{\text{miss}}/frame_{\text{sum}}$
SPDH+PGH+Kalman	31	0.14	0.09
Camshift+Kalman	37	0.26	0.18
Haar+Adaboost	9	0.11	0.07
Chamfer Matching	2	0.05	0.03
Camshift+Particle filter	23	0.18	0.12

其中, $frame$ 为帧数, $frame_{\text{err}}$ 为视频序列中检测错误帧数, $frame_{\text{miss}}$ 为视频序列中检测遗漏帧数, $frame_{\text{sum}}$ 为视频序列总帧数。

经过对比可以看出, Camshift+Kalman 实时性最好, 但误检率最高; Haar+Adaboost 方法漏检率较低, 但是跟踪速度相对较慢, 误检率较高; Chamfer Matching 方法误检率与漏检率都较低, 但是跟踪速度最慢, 很难达到实时跟踪的要求; 本文所提出的 SPDH+PGH+Kalman 方法, 误检率和漏检率与 Haar+Adaboost 方法相当, 跟踪速度接近 Camshift+Kalman 方法, 满足实时性跟踪, 具有较强的鲁棒性, 且能够适用于大多数复杂场景。

本文提出了一种基于肤色与形状的手势检测与跟踪方法, 该方法简单可靠, 具有较强的快速性与鲁棒性。在跟踪过程中, 手部运动尽量保持恒速运动, 跟踪效果比较好。本文提出的方法尚有需要改进之处, 当背景光照较强烈时, 容易出现漏检情况。因为强光照的干扰, 视频中部分肤色区域的色调“污染”严重, 所以影响了自适

应阈值肤色分割的效果。

参考文献

- [1] 杨波, 宋晓娜, 冯志全, 等. 复杂背景下基于空间分布特征的手势识别算法[J]. 计算机辅助设计与图形学学报, 2010, 22(10): 1841-1848.
 - [2] 滕达, 杨寿保, 刘岩, 等. 基于“大小手”的徒手手势实时识别[J]. 计算机应用, 2006, 26(9): 2041-2043.
 - [3] 王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述[J]. 计算机学报, 2002, 25(3): 225-237.
 - [4] HOCH M, LITWINOWICZ P. A practical solution for tracking edges in image sequences with snakes [J]. The Visual Computer, 1996, 12(2): 75-83.
 - [5] 赵怀勋, 郑敏, 李志强. 一种自适应光照的肤色识别算法[J]. 微电子学与计算机, 2006, 24[3]: 158-160.
 - [6] Zhang Mingji, Wang Weiqiang, Zheng Qingfang, et al. Skin-color detection based on adaptive thresholds [C]. Proceedings of the Third International Conference on Image and Graphics, 2004: 250-253.
 - [7] 洪红, 朱秀昌. 一种新的人手图像跟踪算法[J]. 计算机技术与应用, 2010, 20(3): 88-91.
 - [8] 于仕琪, 刘瑞祯. 学习 OpenCV [M]. 北京: 清华大学出版社, 2009.
 - [9] 刘玉进, 蔡勇, 武汇岳. 一种肤色干扰下的变形手势跟踪方法[J]. 计算机工程与应用, 2009, 45(35): 164-167.
- (收稿日期: 2011-10-19)

作者简介:

吕金刚, 男, 1986 年生, 硕士研究生, 主要研究方向: 多运动目标跟踪与手势识别。

吴斌, 男, 1965 年生, 博士, 教授, 博士生导师, 主要研究方向: 智能控制、图像处理与模式识别、人机交互与虚拟现实技术等。

刘恒, 男, 1976 年生, 博士, 副研究员, 硕士生导师, 主要研究方向: 模式识别与计算机视觉。