

复杂背景下基于傅立叶描述子的手势识别

刘寅, 滕晓龙, 刘重庆

(上海交通大学图像所, 上海 200030)

摘要:人的手势是人们日常生活中最广泛使用的一种交流方式。由于在人机交互界面和虚拟现实环境中的应用, 手势识别的研究受到了越来越广泛的关注。但是目前基于单目视觉的手势识别技术中, 手势分割要求背景简单或者要求识别者戴着笨重的数据手套。而该文结合了运动信息和基于 KL 变换的肤色模型, 在复杂背景下进行手势分割, 与传统的基于 RGB 肤色模型的手势分割相比, 在复杂背景环境下得到了很好的分割效果。在对分割的手势区域进行预处理后, 该文使用了一种归一化的傅立叶描述子进行手势的特征提取, 相比传统的傅立叶描述子更加准确, 最后采用了传统的三层 BP 网络作为模式识别器, 手势训练集和测试集的认识率分别达到了 95.9% 和 95%。

关键词:肤色模型; 手势分割; 手势识别; 归一化的傅立叶描述子

中图分类号: TP391.41 **文献标识码:** B

Hand Gesture Recognition Based on Fourier Descriptors with Complex Backgrounds

LIU Yin, TENG Xiao - long, LIU Chong - qing

(Image Institute, Shanghai Jiaotong University, Shanghai 200030, China)

ABSTRACT: Hand gesture is one of the most popular communication methods in everyday life. Hand gesture recognition research has gained a lot of attentions because of its applications for interactive human - machine interface and virtual environments. But currently, in the vision - based hand gesture recognition, almost all the technologies on hand gesture segmentation are based on simple background or on gloves in special colors. However, this paper presents a method that segments the hand gestures with complex backgrounds through the combination of motion and skin color based on KL Transformation, in contrast with traditional hand gesture segmentation based on RGB color model in some environments. After the pretreatment to hand gesture region, we use a normalized Fourier descriptor, which is more accurate than the traditional Fourier descriptor, to select the hand gesture features and use the traditional 3 levels BP network to perform hand gesture recognition. Finally, the average recognition rate is 95.9% on the training set and 95% on the testing set.

KEYWORDS: Skin color model; Hand gesture segmentation; Hand gesture recognition; Normalized Fourier descriptor

1 引言

人与计算机的交互活动越来越成为人们日常生活的一个重要的组成部分。特别是最近几年来, 随着计算机技术的迅速发展, 研究符合人际交流习惯的新颖人机交互手段取得了可喜的进展。这些研究包括人脸识别, 手势识别, 姿态识别等等。由于手势是人们日常生活中最常使用的一种交流方式, 因此手势识别的研究近些年得到了很大的发展。包括基

于数据手套的识别和基于视觉的手势识别技术。由于基于数据手套的识别需要使用者戴上笨重的识别设备, 因此并不实际。因此我们主要研究基于计算机视觉的手势识别技术。

在复杂背景下基于计算机视觉的手势分割, 由于受背景和环境光照变化的影响, 成功地手势分割面临着诸多的困难。不少研究人员采用对手势图像加上种种限制的方法, 例如采用特殊颜色背景或者识别者带着特殊颜色的手套来简化分割的难度。但是人为的增加了诸多限制, 不利于自然的人机交互。

本文使用了多种可用信息, 包括边界及运动信息, 并且

收稿日期: 2004 - 09 - 14

使用了 KL 坐标系下的肤色信息,并对初始的结果进行了一些处理,得到了比较令人满意的分割效果。分割得到的手势区域的特征提取,利用传统的傅立叶描述子具有运算量大,相位信息丢失而造成辨别能力下降等缺点,本文采用了一种归一化的傅立叶描述子,克服了上述传统描述子的缺点。然后选用了传统的三层 BP 网络作为模式识别器,最终得到识别效果。

2 复杂背景下的手势分割

2.1 手势分割系统流程

对视频图像序列进行肤色区域检测,运动分析以及边缘检测,然后对它们的结果进行与运算,并进行预处理(去除噪声点),即可分割出手势区域。见图 1。

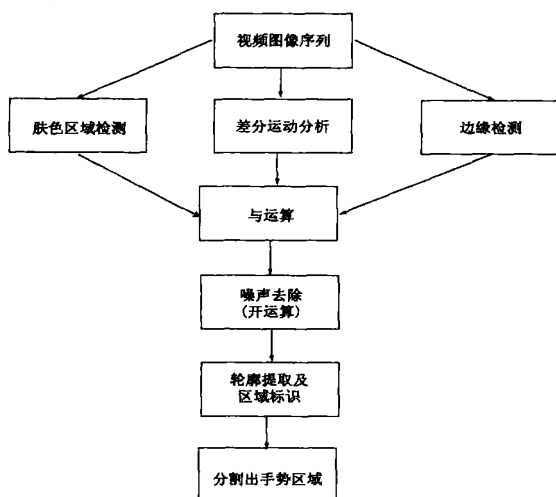


图 1 手势分割系统流程图

2.2 基于 KL 坐标系的肤色模型的区域检测^[2]

常用的彩色表示方法是基于三基色的 RGB 颜色空间,但是如果在 RGB 颜色空间使用约束条件 $R > G > B$ 进行人肤色区域的检测的话,那么检测的结果将会包含很多和肤色相近的区域,例如黄色,桔黄色等等。因此,在复杂背景情况下,特别是存在近似肤色的区域的话,将无法分割出手势区域。

所以,我们将 KL 变换应用到肤色定位上,建立了 KL 肤色坐标系。

通过 KL 变化,得到其变换矩阵为:

$$\begin{pmatrix} K_1 \\ K_2 \\ K_3 \end{pmatrix} = \begin{pmatrix} 0.666 & 0.547 & 0.507 \\ -0.709 & 0.255 & 0.657 \\ 0.230 & -0.797 & 0.558 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

KL 肤色过滤器的阈值为:

$$110.2 < K_1 < 376.3, -61.3 < K_2 < 32.9, -18.8 < K_3 < 19.5$$

在本系统中,由于受环境光照的影响,取上述阈值的检测效果并不理想。

因此,在本系统的实验环境中,我们将 KL 肤色坐标系的阈值调整为:

$$145.2 < K_1 < 376.3, -61.3 < K_2 < 32.9, -4.5 < K_3 < 19.5$$

检测效果如图 2 和图 3 所示。



图 2 取原始阈值的检测结果

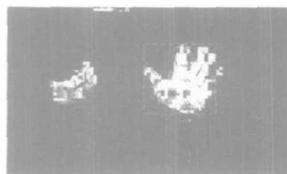


图 3 取调整后的阈值的检测结果

2.3 差分运动分析^[3]

假设摄像机位置固定并且光照恒定,那么在不同时刻上获取的图像之间的简单相减使得运动检测成为可能。差分图像(difference image)是一个二值图像,其中非零值代表了具有运动的图像区域,也就是连续图像 f_1 和 f_2 之间具有较大灰度差值的区域:

$$d(i, j) = \begin{cases} 0, & \text{当 } |f_1(i, j) - f_2(i, j)| < \mu \text{ 时} \\ 1, & \text{其它} \end{cases}$$

其中 $\mu = w\mu$, 这里 w 为权重指数,在不同的环境光照情况下, w 的取值的大小会影响检测的结果,因此在本系统中我们可以根据环境的亮度对 w 进行手动调整; μ 是整幅图像的平均亮度。见图 4 和图 5。



图 4 w 的取值不恰当时的检测结果

2.4 边缘检测

用常用的边缘检测算子(例如 Sobel 算子)对每一帧的原始图像进行处理,并对得到的结果取阈值,即 $e(i, j) = \begin{cases} 0, & \text{当 } e(i, j) < 125 \text{ 时} \\ 1, & \text{其它} \end{cases}$,从而得到边缘检测的结果。

2.5 噪音的去除

为了提升手势分割的精确度和分割出手势形状的精



图 5 调整了 w 值后的检测结果

确度,我们对图像进行了基于数学形态学的预处理,即先对一幅图像进行腐蚀,然后再进行膨胀,得到的不是原始图像。结果图像要比原始图像更简单,可以去除大量的噪声,并简化形状。见图 6 和图 7。



图 6 未进行开运算的分割结果

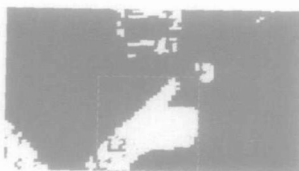


图 7 进行开运算后的分割结果

2.6 基于上述各种信息集合的手势分割

我们对上述的肤色检测图像、运动差分结果以及边缘检测结果进行逻辑与运算,并去除噪声后,即可进行手势区域的分割。通常情况下,由于手势区域在整幅图像中最大,因此我们首先对结果(二值化的图像)进行边缘的轮廓线的搜索,并且计算各个轮廓线包含区域的面积大小,即可得到最大的区域,即手势区域,如图 8。



图 8 最终的分割结果(蓝色框内)

3 基于傅立叶描述子的特征提取及识别

至此,我们已经成功分割出了手势区域的二值图像,下面就需要对分割结果进行特征提取并识别。在计算机中物体的形状有多种表示方式,如:边界曲线上各点的坐标序列,边界上的点到物体重心的距离以及边界上的点随弧长变化的切线方向和曲率等。基于不同的形状表示方式,许多种形

状识别方法已经被提出,如:基于边界特征点、不变矩、傅立叶描述子和自回归模型等方法。

在本文中我们使用傅立叶描述子对分割的手势进行特征提取及形状识别。

3.1 传统的傅立叶描述子^[4]

传统的傅立叶描述子的基本思想是:取分割的手势边界上的某一点为起始点,以逆时针方向沿着边界曲线,得到一个函数 $z(s)$,其中 s 为到起始点的沿着边界的曲线距离。将 $z(s)$ 展开为傅立叶级数,则其系数便是 $z(s)$ 的傅立叶描述子。而对于数字图像数据,边界坐标是离散的,因此其傅立叶描述子可以从 $z(s)$ 的离散傅立叶变换中计算出来。

假设手势图像的边界由坐标对 (x_m, y_m) 表示, $(x_1, y_1) = (x_L, y_L)$ 。则:

$$a_n = \frac{1}{L-1} \sum_{m=1}^{L-1} x_m e^{-i[2\pi/(L-1)]nm}$$

$$b_n = \frac{1}{L-1} \sum_{m=1}^{L-1} y_m e^{-i[2\pi/(L-1)]nm}$$

系数 a_n 和 b_n 不是不变量,但是经过变换,

$$r_n = (\sqrt{a_n^2 + b_n^2})^{1/2}$$

则 r_n 是平移和旋转不变量。同时为了达到缩放不变性,使用描述子 $w_n = r_n / r_1$ 即可将手势区域的形状信息提取出来。

但是对于传统的傅立叶描述子,由于需要对连续的手势边界进行离散化,因此采样的误差造成了傅立叶描述子精度的降低;与此同时,由于采样的间隔很小,因此很大的采样点数目也导致了极大的傅立叶变换运算量。而且传统的傅立叶描述子丢失了傅立叶变换系数中的相位信息,因此对于物体形状的辨别能力有所下降。

3.2 归一化的傅立叶描述子^[1]

为了提高傅立叶变换的运算速度并且降低边界曲线离散化造成的误差,通过边界曲线的多边形近似方法^[1],对近似多边形的各条直线段进行连续傅立叶变换来代替原来的离散傅立叶变换。

边界多边形的顶点 P_n 坐标为 x_n, y_n ; l_n 表示多边形顶点 P_n 到起点 P_0 的边长; L 表示多边形的边长;则边界多边形的傅立叶变换如下:

$$P(l) = \sum_{n=1}^{N-1} P_n(l) = x(l) + jy(l)$$

$$z(k) = \frac{1}{L} \int_0^L P(l) e^{-j\frac{2\pi}{L}kl} dl = \frac{1}{L} \sum_{n=0}^{N-1} \int_{l_n}^{l_{n+1}} P_n(l) e^{-j\frac{2\pi}{L}kl} dl$$

傅立叶描述子与形状的尺度、方向和曲线的起始点 P_0 位置有关。为了识别具有旋转、平移和尺度不变性的形状,需要对傅立叶描述子进行归一化。对于起始点位置平移 a 长度,物体放大 λ 倍,旋转角度 θ , 平移位置 (x_0, y_0) 后的物体边界形状的傅立叶描述子变为:

$$z(0) = re^j z(0) + F(x_0 + jy_0), k = 0$$

$$z(k) = re^j e^{j\frac{2\pi}{L}ka} z(k), k = 1, 2, \dots, N-1$$

然后定义归一化形状,其主方向 = 0为水平方向,选择边界起始点的位置和形状大小恰好使其傅立叶变换系数 $z(1)$ 的相位为零,模为 1,即 $z(1) = 1$ 。

定义归一化的傅立叶描述子 $d(k)$ 如下:

$$d(k) = \frac{z(k)}{e^{-\frac{2}{L}(k-1)a} z(1)}, \text{其中 } e^{\frac{2}{L}a} = \frac{z(1)e^{-j}}{z(1)}, k = 1, 2, \dots, N - 1$$

重新定义的归一化傅立叶描述子 $d(k)$ 计算简便,同时保留了傅立叶变换系数的模与相位特性。 $d(k)$ 值恒等于归一化形状的傅立叶变换系数 $z(k)$ 与 $z(1)$ 之比,与形状的大小、旋转方向、平移和边界起点位置无关。由于新的归一化傅立叶描述子 $d(k)$ 比传统的傅立叶描述子多包含了相位信

息,因此避免了由于归一化傅立叶描述子丢失相位信息引起的形状混淆,可以更加准确地识别物体的形状。

$$\text{取 } dis\ tan\ ce = \sum_{k=2}^M d_i(k) - d_j(k)^2$$

利用傅立叶描述子便可以计算任意两个手势的相似程度,可以将分割手势边界的傅立叶描述子同模板库中手势的傅立叶描述子进行比较。通常来说,利用描述子的低频分量足以计算差异,本实验系统中 M 的值根据实际情况取为 20。

3.3 实验结果 (见表 1和表 2)

本实验系统中,预先定义了 10种手势 (对应于人用手势表示从 0到 9的含义),每种手势取 40个样本,20个作为训练集,20个作为测试集。本实验使选用的是三层 BP网络。

表 1 训练集的手势识别正确率

手势	0	1	2	3	4	5	6	7	8	9	平均
正确率	94%	100%	100%	100%	98%	98%	93%	90%	93 %	93%	95.9%

表 2 测试集的手势识别正确率

手势	0	1	2	3	4	5	6	7	8	9	平均
正确率	94%	100%	100%	97%	97%	97%	93%	90%	91%	91%	95%

4 结论

本文利用多种信息对手势进行分割,得到了令人满意的结果;在对分割手势的特征提取及识别方面,使用了一种归一化的傅立叶描述子,实验结果表明这些算法是有效的,令人满意的。

但是,在对手势分割过程中,如果出现了运动的人脸区域或者其它大面积的肤色区域,则最终的分割效果无法保证,因此如何解决这个问题是增强手势分割效果的面临的一个难点。而且本文中使用的傅立叶描述子对于出现遮挡现象的手势的识别也无能为力。因此解决手势识别系统中的遮挡问题是面临的一个需要解决的问题。因此,进一步的工作是解决上述问题。

参考文献:

[1] 王涛,刘文印,孙家广,张宏江. 傅立叶描述子识别物体的形状[J]. 计算机研究与发展, 2002, 39(12).

[2] 杨盈昀,谢婷婷,施美楠. 基于肤色的人脸检测算法研究[J].

北京广播学院学报, 2002, 9(4): 11 - 20.

[3] Feng - Sheng Chen, Chih - Ming Fu, Chung - Lin Huang Hand Gesture Recognition Using a Real - time Tracking Method and Hidden Markov Models[J]. Image and Vision Computing 2003, 21: 745 - 758.

[4] Milan Sonka, Vaclav Hlavac, Roger Boyle Image Processing, Analysis, and Machine Vision[M]. 人民邮电出版社, 2003.



[作者简介]

刘 寅 (1979. 8. 1 -),男 (汉族),山东滕州人,上海交通大学模式识别与智能系统 2002级硕士生,主要研究方向:图像处理,手势识别和人机交互技术。

滕晓龙 (1975),男 (汉族),上海人,上海交通大学模式识别与智能系统 2002级博士生,主要研究方向:图像处理和模式识别。

刘重庆 (1939 -),男 (汉族),上海人,上海交通大学模式识别与智能系统教授,博导,模式识别与计算机视觉研究室主任,主要研究方向:模式识别和计算机视觉。