

文章编号: 1004-4353(2013)03-0211-04

# 一种基于视频的手势识别算法

关键, 何良华\*

( 同济大学电子与信息工程学院 计算机科学与技术系, 上海 201804 )

**摘要:** 研究了一种基于视频的手势识别算法,该算法利用均值法去除噪声,根据物体颜色配合最大类间方差法对目标和背景进行区分,并使用 BP 神经网络进行手势的分类.实验表明,该算法对部分典型手势识别的准确率达到 74.7%,具有较高的实际应用价值.

**关键词:** 手势识别; 神经网络; 视频; 算法

**中图分类号:** TP391.4      **文献标识码:** A

## A gesture recognition algorithm based on video

GUAN Jian, HE Lianghua\*

( Department of computer science and technology, College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China )

**Abstract:** We study a kind of hand gesture recognition algorithm based on video. Average method is used in removing noise, the algorithm based on object color together with the OTSU is used in distinguishing target and background, and the BP neuron network is used in classifying hand gestures. Experiments show that the accuracy of the algorithm is 74.7%, so its application value is high.

**Key words:** gesture recognition; artificial neuron network; video; algorithm

手势是人机交互的方式之一,通过这种方式可以实现人与机器的自然沟通<sup>[1]</sup>.目前,主流的手势识别方法有基于数据手套的手势识别方法和基于视频的手势识别方法两种<sup>[2-3]</sup>.基于数据手套的手势识别方法是利用数据手套来确定手势的形状,并根据手势形状来分析该手势的种类;而基于视频的手势识别方法是利用摄像头来采集手势识别所需的手势信息并加以分析,从而实现手势识别的目的<sup>[4]</sup>.基于视频的手势识别方法与基于数据手套的手势识别方法相比虽然存在处理过程复杂、需要处理的数据量大等缺点,但是该方法具有设备简单、价格低廉、可应用范围广和不干扰用户等优点,因此拥有较高的实际应用价值<sup>[5]</sup>.

基于视频的手势识别是一个根据视频采集到

的信息对手势进行分类的问题,而人工神经网络恰好可以完成对数据进行分类的任务,这使得利用神经网络来实现基于视频的手势识别成为可能.本文以视频采集到的手势信息为依据,以人工神经网络理论为基础,实现了一种具有实用性的手势识别算法,其准确率达到 74.7%.

### 1 噪声处理与背景目标识别

摄像头在采集视频帧图片时会具有一定的噪声,而且视频帧图片当中还会存在着一些无用的对象,因此就需要通过图像处理来去除噪声、划分目标与背景.

**1.1 去除噪声**

在摄像头获取的图像中,其噪声主要是由周

围电磁波影响而产生的高斯噪声和椒盐噪声,这 2 种噪声的去除方法如下:

1) 高斯噪声的去除. 一种简单的方法是使用摄像头在短时间内取多个图像,并求出这些图像的平均值. 这是因为高斯噪声是随机产生的噪声,它随时间变化而变化,而且平均值为 0,因此只要求出多张图片的平均值就可以有效地去除高斯噪声. 此方法的优点是处理简单、执行速度快,其缺点是采集运动物体图像的效果不够理想,因为在不同时刻图像中运动物体的位置不同. 石美红等<sup>[6]</sup>提出了一种通过滤波来去除高斯噪声的方法.

2) 椒盐噪声的去除. 采取对邻近的多个像素相加取平均值的方法可以减弱椒盐噪声,其原因是:椒盐噪声是随机分布的,它聚集在一个较小区域内的概率较低,因此对像素取平均值可以使椒盐噪声减弱. 此方法的优点是处理简单、执行速度快,其缺点是对图像细节的保留效果不理想和对摄像头的像素要求较高等. 王晓凯等人<sup>[7]</sup>提出了一种使用中值滤波器来去除椒盐噪声的方法.

1.2 背景目标识别

在摄像头获取的视频帧图片中,会存在对手势识别没有用的背景,为了去除背景对识别结果的干扰,需要把背景与目标区分开. 本文使用物体颜色配合最大类间方差法来区分目标和背景,具体方法如下:首先将采集到的图像由 RGB 色彩空间转换为 YCbCr 色彩空间,并去除其中受光照影响较大的 Y 分量;然后根据目标物体的颜色构造一个函数,使颜色接近目标物体的函数值大,颜色远离目标物体的函数值小,并使用这个函数将彩色图片转换为一个灰度图片,在该图片中目标物体灰度值较高,背景灰度值较低;最后采用最大类间方差法<sup>[8]</sup>即可区分出目标与背景.

但是在使用以上方法划分目标和背景过程中,容易将背景当中颜色与目标物体相近的区域误判为目标物体,对此可以采取以下两种方法纠正:一种方法是戴上与背景颜色不同的有色手套;另一种方法是事先采集一张背景的图片,然后使用这张图片来减弱背景的影响. 其中后一种方法要求背景不变,因此不适合在摄像头移动的情况下使用,而且在采集背景图片时,目标物也不应出现在画面当中.

2 手势图像归一化

在通过图像处理得到的目标与背景的二值图像中提取出相应的数据来训练神经网络. 数据选择的好坏能较大地影响神经网络训练和识别的效果,因此在提取数据时应掌握如下原则:不同类别提取到的数据差别尽可能大,而同一类别提取到的数据差别尽可能小.

在手势识别过程中,手势只与手指的动作有关,而与手的大小、位置和角度无关,因此在选择的数据当中应去除手的大小、位置和角度对判别结果的影响. 去除手的位置对判别结果影响的方法是:计算手势整体重心的坐标,然后通过坐标变换使其成为所用坐标的原点,由于手势边缘各点的坐标对于手势识别最为重要,所以需要将手势边缘各点的坐标作为神经网络的输入数据. 该方法抗噪声能力较差,因此采用如下的方法来提高算法的抗噪声能力和识别准确度:首先把手的图像以手势重心为中心按照角度等分成  $n$  个区域,然后再分别计算每个区域的目标物重心,并根据每个区域目标物重心与手势整体重心间的位置关系得出输入神经网络的数据.

消除手势角度对结果影响的方法是:把各区域目标物重心与手势整体重心的距离作为研究对象,并把该距离最大的区域定为 0 号区域,其余的区域按逆时针方向依次定为 1 号、2 号、3 号等区域. 只要手势相同,无论手势角度如何变化,各区域的该距离均不发生变化,这样就消除了手势角度对结果的影响. 图 1 就是  $n=4$  时手势整体重心与区域重心的一种可能分布.

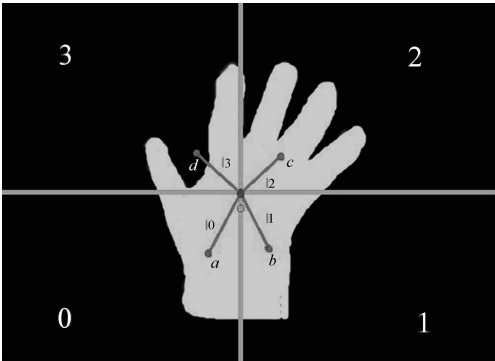


图 1  $n=4$  时手势整体重心( $o$ )与区域重心( $a, b, c, d$ )的一种分布

消除手势在图像中的大小对结果影响的方法是:计算手势整体重心到各区域重心的距离与该距离最远值的比值,这样只要手势相同,无论手势在图像中大小如何变化,该组比值始终不变,由此可判定是否是同一手势.为了使神经网络训练的效果更加明显,可以将该比值的平方作为输入神经网络的数据,即输入神经网络的数据由公式(1)得到:

$$d_i = (\frac{l_i}{l_0})^2, i = 0, 1, 2, \dots, n - 1, \tag{1}$$

式中  $d_i$  表示第  $i$  个数据,  $l_i$  表示第  $i$  个区域的重心与手势整体重心的距离,  $l_0$  表示第 0 个区域的重心与手势整体重心的距离,且  $l_0$  是所有  $l_i$  中最大的.

3 手势判别神经网络的构造

3.1 手势判别神经网络的种类

本研究采用的手势判别神经网络是 BP 神经网络,它是一种按误差逆传播算法训练的多层前馈网络,由输入层、隐层和输出层节点组成,其中输入层和输出层只有 1 个,而隐层可以有多个.神经网络的各层中都有若干个节点,每一个节点表示一个神经元,且前层与后层节点之间由带有权值的边相连. BP 神经网络进行训练时,训练样例数据所对应的信号通过隐层从输入层传向输出层,并将得到的结果与期望结果进行比较得到误差;之后,将误差沿连接通路反向传播,并采用梯度下降法调节各节点的权值,使误差逐渐减小,最终达到条件所允许的程度,满足训练的要求.训练完成后,即可以使用该神经网络对数据进行判别和分类.

3.2 手势判别神经网络的构造

手势判别神经网络由多个神经元组成,各神经元的输入、输出关系如公式(2)和(3)所示:

$$O_i = f(net_i) = \frac{1}{1 + e^{-net_i}}, \tag{2}$$

$$net_i = \frac{x_1w_{1i} + x_2w_{2i} + \dots + x_mw_{mi}}{m}. \tag{3}$$

公式(2)和(3)表达的是某一层第  $i$  个神经元的输入输出情况,其中  $f$  是该 BP 神经网络的激活函数,  $x_j$  是该神经元第  $j$  个输入,  $w_{ji}$  是该神经元第

$j$  个输入的权值,  $m$  是该神经元的输入个数,  $O_i$  是该神经元的输出.训练时权值的变化方式采用带有冲量的 BP 算法.

本文分别对隐层数为 1、2、3、4,隐层中的神经元个数为 8、16、32、64、128、256、512 的神经网络进行了测试,发现当隐层数为 2,每个隐层中神经元个数为 128 时算法的综合效果最为理想.

4 系统的实现及测试

利用 VC++ 2008、OpenCV2.45 编写了一个基于视频的神经网络手势识别程序,其界面如图 2 所示.图 2 中左上的大图是摄像头的实时图像(粗略处理所得),左下的小图是添加样例和判断时程序截取的图片(精细处理所得),右下的数字表示经判断分析后所确定的手势类别,右侧是该程序的各种功能按钮.测试时首先添加样例,点击添加样例按钮后程序会向 traindata.data 添加由手势图片生成的手势数据.每一组数据由  $n+1$  个数组成,其中第 1 个整数是该数据的类别,后面  $n$  个有理数是  $d_0, d_1, \dots, d_{n-1}$ ,如图 3 所示.之后,点击训练按钮训练神经网络,程序会读取 traindata.data 中的各组样例数据,开始训练神经网络,训练完成后会弹出训练成功的提示.

上述工作完成后即可进行手势识别.点击判断按钮程序就会截取摄像头采集到的手势图像,并判断该手势属于哪一个类别.为了能有效识别各种手势,需添加多种训练样例,例如:包袱(0 类)、锤(1 类)、剪子(2 类)等,相应的手势识别结果如图 2、图 4 和图 5 所示.

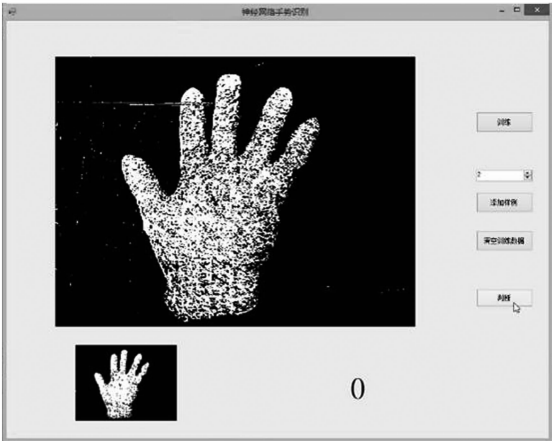


图 2 实验结果示例 1(包袱)

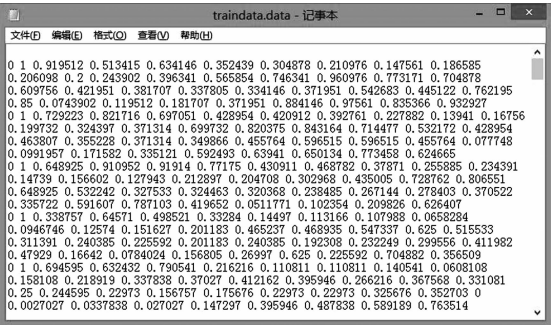


图 3 traindata.data 中的各组数据

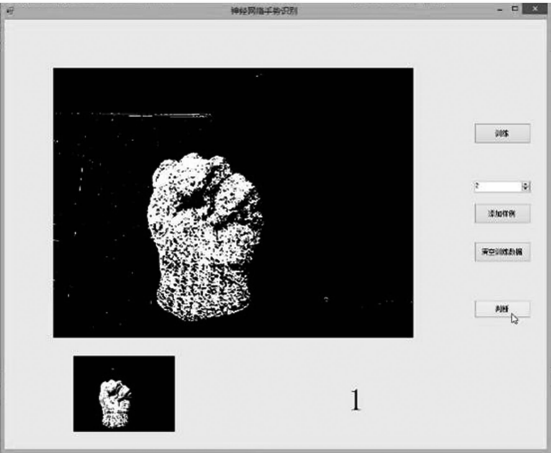


图 4 实验结果示例 2(锤)



图 5 实验结果示例 3(剪子)

为了评估本文算法的效果,使用 30 组数据(其中剪子、包袱、锤各 10 组)对神经网络进行训练,然后对另外 150 组数据(其中剪子、包袱、锤各 50 组)进行识别,其结果如表 1 所示。

表 1 手势识别实验结果

分类	测试次数	正确次数	正确率/%
剪子	50	43	86.0
包袱	50	32	64.0
锤	50	37	74.0
总计	150	112	74.7

5 结束语

以摄像头作为信息采集设备,以 BP 神经网络作为手势判别工具,实现了一种具有实用性的手势识别算法,为人机交互提供了一种可靠的途径。该算法不受手势图像旋转、缩放和平移的影响,具有可识别手势种类多、抗噪声能力较强和手势识别的准确度较高等特点,其不足是处理过程较为复杂、处理数据量过大等,这有待于今后进一步完善。

参考文献:

[1] 张凯. 基于立体视觉的自然手势识别[D]. 北京: 北京大学, 2005.

[2] 何阳清, 葛元, 王林泉. 应用几何矩和边缘检测的手势识别方法[J]. 计算机工程, 2005, 31(15): 165-174.

[3] 任海兵, 徐光祜, 林学闯. 基于特征线条的手势识别[J]. 软件学报, 2002, 13(5): 987-993.

[4] Javier Molina, Marcos Escudero-Vinolo, Alessandro Signoriello, et al. Real-time user independent hand gesture recognition from time-of-flight camera video using static and dynamic models[J]. Machine Vision and Applications, 2013, 24(1): 187-204.

[5] 江立, 阮秋琦. 基于神经网络的手势识别技术研究[J]. 北京交通大学学报, 2006, 30(5): 32-36.

[6] 石美红, 毛江辉, 梁颖, 等. 一种强高斯噪声的图像滤波方法[J]. 计算机应用, 2007, 27(7): 1637-1640.

[7] 王晓凯, 李锋. 改进的自适应中值滤波算法[J]. 计算机工程与应用, 2010, 46(3): 175-176.

[8] Otsu N. A threshold selection method from gray level histograms[J]. IEEE Trans Systems Man & Cybernetics, 1979, 9(1): 62-66.