

基于卷积神经网络的手势识别

杨红玲, 宣士斌, 莫愿斌

(广西民族大学 信息科学与工程学院, 广西 南宁 530006)

摘要: 在手势识别的过程中, 手势变化的多样性和手势本身的复杂性会对手势识别的精确性和可靠性带来更大的影响。为了能够在实现高准确率手势识别的同时降低识别速度, 提出了一种基于深度卷积神经网络的准确的手势识别方法。该方法首先运用边缘检测算法和细化算法提取手势区域的边缘轮廓特征和手势骨架特征, 然后采用特征融合的方法获取手势融合特征, 最后通过对比几种常见机器学习算法(支持向量机、决策树、随机森林和卷积神经网络)在手势识别中的时间效率和准确精度, 选取最优的手势识别模型。实验结果表明, 在不同数据集下, 通过实验数据对比, 基于深度神经网络的手势识别虽然在平均时间开销上相对较高, 但在识别准确率上却提升了2%, 可以达到98.57%。

关键词: 机器学习; 卷积神经网络; 手势识别; 准确率

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2018)07-0011-04

doi: 10.3969/j.issn.1673-629X.2018.07.003

Hand Gesture Recognition Based on Convolutional Neural Network

YANG Hong-ling, XUAN Shi-bin, MO Yuan-bin

(School of Information Science and Engineering, Guangxi University for Nationalities, Nanning 530006, China)

Abstract: In the process of hand gesture recognition, the diversity of gesture changes and the complexity of gesture itself have greater impact on the accuracy and reliability of hand gesture recognition. In order to reduce hand gesture recognition speed while achieving high accuracy of gesture recognition, we propose an accurate gesture recognition method based on deep convolution neural network. Firstly, the edge detection algorithm and the refinement algorithm are used to extract the edge contour feature and gesture skeleton features of the gesture region. Then the feature fusion method is used to obtain the gesture fusion feature. Finally by comparing several common machine learning algorithms like support vector machines, decision tree, random forests and convolutional neural networks on the time efficiency and accuracy of gesture recognition, the optimal gesture recognition model is selected. The experiment shows that under the different data sets, the average time cost of the gesture recognition based on the deep neural network is higher than other algorithms, but the rate of recognition accuracy is improved by 2% and can reach 98.57%.

Key words: machine learning; convolutional neural networks; gesture recognition; accuracy

0 引言

近年来随着科学技术的高速发展, 人机交互的方式得到了很大改变, 各种新型的人机交互方式不断出现, 鼠标键盘的交互方式变为触摸屏与语音, 交互形式变得多样化、人性化。而更为高效的交互形式是让机器能够理解人的肢体语言, 在各类肢体语言中手势最为常见, 可将它作为一种简单、自由的人机交互手段。

基于手势进行人机交互时, 一个很重要的过程是对手势进行识别。手势识别时, 首先提取手势的特征,

然后对所提取的特征根据有效的识别方法进行手势识别。常见的手势识别方式有很多, 例如基于神经网络的识别方法具有较强的识别分类能力, 但是如果采用的神经网络层数较浅, 很容易出现过拟合现象^[1-2]; 基于几何特征的识别方法通过提取手势结构、边缘、轮廓等特征进行手势识别, 具有良好的稳定性, 但是不能在提升样本量的同时提升识别率^[3-5]; 基于隐马尔可夫模型的识别方法虽然具有描述手势时空变化的能力, 但识别速度却不尽如人意^[6]。随着机器学习和深度学

收稿日期: 2017-08-07

修回日期: 2017-12-27

网络出版时间: 2018-03-07

基金项目: 国家自然科学基金(21466008); 广西自然科学基金(2015GXNSFAA139311); 广西民族大学研究生科研创新项目(gxun-chxzs2017112)

作者简介: 杨红玲(1991-), 女, 硕士研究生, 研究方向为图像处理与识别; 宣士斌, 教授, 博士, 研究方向为图像处理与识别; 莫愿斌, 教授, 博士, 研究方向为智能信息处理与应用。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180307.1422.040.html>

习在计算机视觉的迅速发展,基于机器学习和深度学习的方法得到了更多的关注。其中基于深度卷积神经网络具有局部连接、权值共享、深度层次化结果、自动特征提取等特点,给手势识别^[7-8]带来了新的思路。

因此针对手势变化的复杂性,通过对比支持向量机、决策树、随机森林和邻近算法在手势识别中的特点和存在的问题,提出了一种基于深度卷积神经网络的手势识别方法。该方法提取手势的骨架与边缘相融合的特征图,将特征图作为深度卷积神经网络的输入,通过学习获取分类手势时的分类模型,实现手势识别。

1 基于机器学习的手势识别

利用计算机代替人学习提高自身的处理问题的能力就是机器学习。随着计算机技术的高速发展,使用机器学习的领域逐渐扩大,基于机器学习的方法已经在语音、图像、文本、金融等领域取得了突破性进展。

文中通过对比常见的有监督学习算法,从中选择最优的学习算法进行手势识别。算法的输入为采集得到的原始手势图像,将原始图像通过滤波、去除噪声等预处理后进行骨架与边缘特征提取,获取这两种特征相融合的特征图,然后将融合后的特征作为输入,训练支持向量机、决策树、随机森林和卷积神经网络的手势识别模型,通过对比选取最优的分类模型作为输出来判断手势所代表的含义。具体算法流程如图 1 所示。

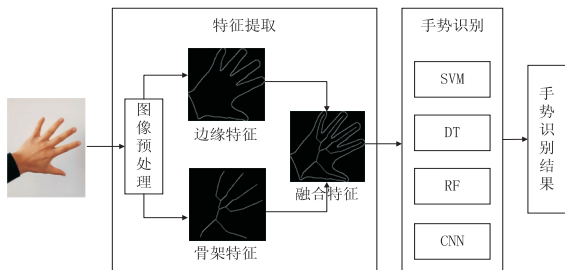


图 1 手势识别算法流程

1.1 特征提取

由于手势变化的复杂性,人们在进行手势识别时更加关注手势本身所代表的含义,而原始的手势图像中包含很多不必要的细节信息,从而增加了识别难度。为了增加手势识别的准确率,减少计算的复杂度,将手势的骨架特征与边缘特征相融合作为手势识别的输入,以减少不必要的细节信息对手势识别的干扰。

骨架作为手势的一种表示形式,能够保持手势体的几何、形状、拓扑信息,能够有效地描述手势。因此,骨架能够很好地描述手势所代表的物理含义,可以将手势骨架信息作为一类手势识别的特征描述,手势骨架提取结果如图 2(b)所示。

虽然单一的手势骨架特征能够很好地解释手势所代表的含义,但是在提取不准确或者一定的条件下,骨

架特征信息会有一定的缺失。对此,进一步利用形态学算子提取手势二值图像的边缘,获取具有更好解释效果的手势边缘图像,然后将其手势骨架图像相结合作为卷积神经网络的输入,获取更好的识别效果,融合结果如图 2(d)所示。

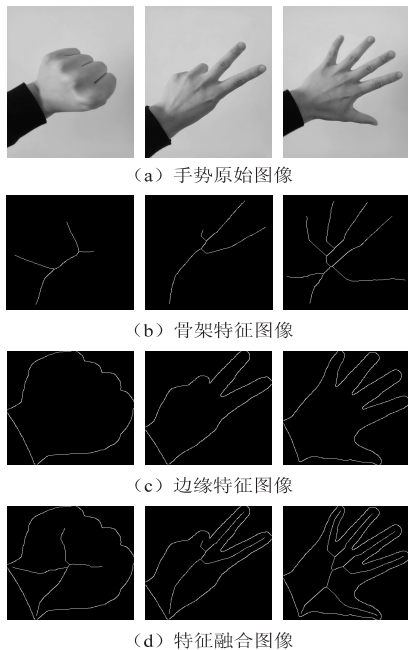


图 2 多特征融合效果图

1.2 手势识别

1.2.1 基于支持向量机的手势识别

支持向量机是建立在 VC 维理论和结构风险最小原理基础上的机器学习算法,能够很好地解决非线性以及高维度识别的问题。由于手势本身的复杂性,因此可以考虑将其引入到手势识别,将融合后的特征图像转化为支持向量机训练时所需要的一维特征向量并作为输入,训练获取分类模型,进行手势识别。

1.2.2 基于决策树的手势识别

决策树学习是以实例样本为基础的归纳学习算法,可以从一组无次序、无规则的事例样本中推理出决策树所表示形式的分类器和预测模型,从而实现了对未知数据样本的分类或预测。文中采用 ID3 学习算法生成决策树并进行剪枝,通过手势融合特征图像样本训练生成决策树模型进行手势预测。在利用决策树生成算法对手势进行识别时,由于独特的树形结构在预测时能减少识别时间,并且能够直接体现数据的特点,因此具有一定的可信度,但是对于图像数据来说,高维度的数据训练存在分类识别精度的问题。

1.2.3 基于随机森林的手势识别

在机器学习中,随机森林是一个包含多个决策树的分类器,其输出的类别由个别树输出的类别的众数而定。随机森林学习算法中每棵树的训练过程与决策树类似,只是无需对决策树进行剪枝。并且数据样本

和特征选择是一个随机过程,每棵树的具体构造如下:

- (1)用 N 表示训练样本的个数, M 表示图像转化为一维数据结构作为训练的特征;
- (2)从 N 个训练样本中采用又放回的抽样方式,取样 N 次,形成一组训练集;
- (3)对抽取的样本,随机选择 m 个特征 ($m \ll M$),计算其最佳的分割方式,训练生成一棵决策树;
- (4)选取 20 个数据集进行训练,每棵树都会完整地成长而不会剪枝。

利用随机森林算法对手势进行识别时,由于结果需要根据多棵树输出的众数而定,因此相对决策树分类来说,精确度会有一定的提升,但是由于多棵树的预测,时间将会增加。

1.2.4 基于卷积神经网络的手势识别

随着深度学习的快速发展,卷积神经网络已经在语音识别^[9]、手写字体识别^[10]、车牌识别^[11]、人脸识别^[12]等领域得到了广泛的应用,其高效的识别精度和速度对手势识别也具有一定的促进作用。因此可以采用基于深度学习的方法来进行手势识别。

卷积神经网络(CNN)具有三个最基本的特征^[13]:局部连接、权值共享和下采样。通过局部连接和权值共享减少训练参数,通过下采样提升模型的鲁棒性,减少训练参数。因此根据卷积神经网络的特征,其一般包含两个特殊的网络神经元层:卷积层和下采样层。由于文中的分类任务较为简单,因此基于 AlexNet 的网络结构进行精简,具体的网络结构如图 3 所示。

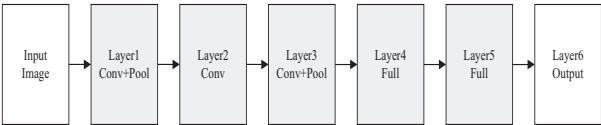


图 3 手势识别的卷积神经网络结构

该结构共有 6 层,Input Layer 为输入层,具体为 64×64 的手势特征融合图像,Layer1–Layer3 为卷积层,Layer4–Layer5 为全连接层,Layer6 Output 为输出层,输出层神经元有 3 个,分别代表手势类别:石头、剪刀、布。卷积核和各偏置等参数的初始值均随机产生,输入样本后通过前向传播和反向传播算法对网络进行训练来更新参数。

卷积滤波实质就是用卷积核在图像矩阵中滑动遍历,卷积核与图像上相对位置的元素作乘积,将所得结果相加得到一个结果值,最后通过激活函数获得卷积结果。当卷积核滑动遍历整张图像,结束特征提取,获取一个新的图像特征矩阵(feature map)。同时卷积核滑动的步幅也和最后获取的特征矩阵存在以下关系:

$$a_{i,j} = f(\sum_{m=0}^M \sum_{n=0}^N w_{m,n} x_{i+m,j+n} + w_b) \tag{1}$$
$$f(x) = \frac{1}{\max(1, x)} \tag{2}$$

$$W_2 = (W_1 - F + 2P)/S + 1 \tag{3}$$

$$H_2 = (H_1 - F + 2P)/S + 1 \tag{4}$$

式 1 为卷积计算,式 2 为激活函数,式 3 和式 4 为卷积变化。其中, $x_{i,j}$ 为图像的第 i 行第 j 列元素, $w_{m,n}$ 为卷积核中第 m 行第 n 列权重, w_b 为卷积核的偏置项; f 为激活函数,即 relu 函数; W_2 为卷积后 feature map 的宽度, W_1 为卷积前图像的宽度, F 为 filter 的宽度, P 为 Zero Padding 数量,Zero Padding 是指在原始图像周围补几圈 0,如果值是 1,那么就补 1 圈 0, S 为步幅; H_2 为卷积后 Feature Map 的高度, H_1 为卷积前图像的宽度。

卷积滤波后再通过下采样图像特征矩阵进行降维,减少计算量,同时避免特征过多导致出现过拟合,增强网络结构对位移的鲁棒性。具体的卷积和下采样计算如下所示:

$$b = P \begin{pmatrix} a_{i,j} & \cdots & a_{i+p,j} \\ a_{i,j+1} & \cdots & a_{i+p,j+1} \\ a_{i,j+q} & \cdots & a_{i+p,j+q} \end{pmatrix} \tag{5}$$

其中, $a_{i,j}$ 为卷积后的第 i 行第 j 列元素; P 为下采样函数,一般为 MaxPoling 或 MeanPoling,文中采用 MaxPoling。

2 实验结果及分析

2.1 实验结果

对提出的方法在两个数据库上进行验证,第一个数据库是在室内场景采集的手势图像数据库,通过普通的摄像头拍摄不同环境、不同旋转角度下的 3 种类别的手势图像各 100 张,用于算法性能的测试;第二个数据库采用 Thomas Moeslund’s Gesture Recognition Database。同时在两个数据库中对文中所涉及的手势识别模型进行验证,结果如表 1 所示。

表 1 识别性能的比较(1)

算法	平均消耗时间/(幅/ms)	平均识别率/%
SVM	39.335	89.33
DT	0.450	93.11
RF	0.425	96.22
CNN	4.000	98.57

可以看出,平均消耗时间上虽然随机森林(RF)和决策树(DT)比基于卷积神经网络(CNN)的消耗时间过短,这是因为其独特的树形结构在分类过程中会减少算法的时间复杂度,但是在平均识别率上,CNN 却有着天然的优势,而且其消耗时间也在可接受范围之内;而 SVM 无论在消耗时间还是速度上都逊色于 CNN,因此采用深度卷积神经网络进行手势识别可行。

为了更好地验证卷积神经网络训练次数对手势识

别率和误差的影响,从拍摄的各类手势图像中选取 2 000个训练样本和 100 个测试样本进行实验,不同的训练次数与手势识别率和误差的关系如图 4 所示。

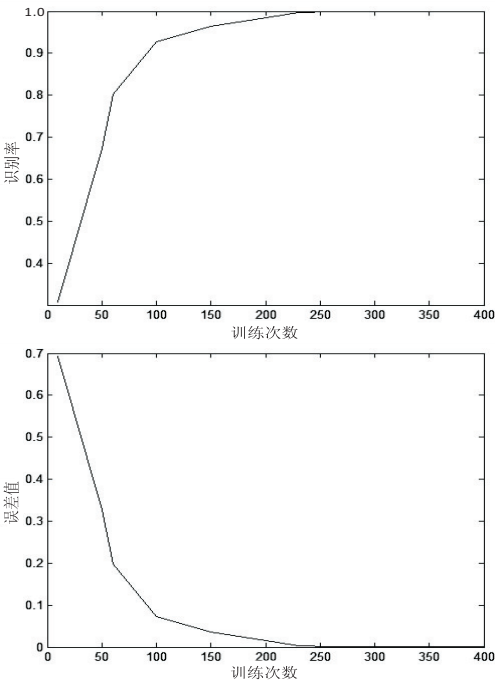


图 4 训练次数与手势识别率和误差的关系

可以看出,训练次数较少时,手势的识别率较低,网络需要训练较多的次数才可以达到较好的识别效果。因为在训练次数较低时,由于手势的复杂性,并不能提取出具有高效分类的网络参数,使得训练误差仍然很高,当训练进行到一定程度时,网络参数不会发生太大变化,误差趋于稳定,网络识别率的提高趋于稳定。

2.2 实验对比分析

为进一步验证文中算法的性能,与国内其他学者提出的算法进行比较,表 2 显示了手势样本在不同方法下的识别率和时间消耗对比。

表 2 识别性能的比较(2)

算法	平均消耗时间/(幅/s)	平均识别率/%
文中算法	0.004	98.57
文献[7]	0.005	95.00
文献[8]	0.008	96.80
文献[1]	0.003	87.25
文献[14]	0.12	84.00

通过对比发现,文中算法获取的识别率相对较高的原因在于将手势图像的骨架和边缘的融合特征图像作为卷积神经网络的输入,能够描述手势所代表的物理含义,从而获得更好的识别效果。而且网络结构更为简单,从一定程度上减少了识别的消耗时间,从而实现快速准确的手势识别。

3 结束语

针对手势的复杂性,通过融合手势的边缘与骨架

特征作为识别算法的输入,提出基于卷积神经网络的手势识别方法。实验结果表明,基于卷积神经网络的手势识别具有较高的准确率,并且识别速度也在可接受范围之内。下一步将通过改进网络结构进一步提高手势识别的速度,实现复杂环境下动态的手势识别。

参考文献:

[1] STERGIOPOULOU E, PAPAMARKOS N. Hand gesture recognition using a neural network shape fitting technique [J]. Engineering Applications of Artificial Intelligence, 2009,22(8):1141-1158.

[2] 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:自然科学版,2006,30(5):32-36.

[3] LIU Yun, YIN Yanmin, ZHANG Shujun. Hand gesture recognition based on HU moments in interaction of virtual reality[C]//4th international conference on intelligent human-machine systems and cybernetics. Nanchang, China; IEEE, 2012:145-148.

[4] 董立峰,阮军,马秋实,等. 基于不变矩和支持向量机的手势识别[J]. 微型机与应用,2012,31(6):32-35.

[5] 隋云衡,郭元术. 融合 Hu 矩与 BoF-SURF 支持向量机的手势识别[J]. 计算机应用研究,2014,31(3):953-956.

[6] MURTHY G R S, JADON R S. Hand gesture recognition using neural networks[C]//Advance computing conference. Patiala, India; IEEE, 2010:134-138.

[7] 王龙,刘辉,王彬,等. 结合肤色模型和卷积神经网络的手势识别方法[J]. 计算机工程与应用,2017,53(6):209-214.

[8] 操小文,薄华. 基于卷积神经网络的手势识别研究[J]. 微型机与应用,2016,35(9):55-57.

[9] SUKITTANON S, SURENDRAN A C, PLATT J C, et al. Convolutional networks for speech detection[C]//International conference on spoken language processing. Jeju Island, Korea; [s. n.], 2004.

[10] CHEN Y N, HAN C C, WANG C T, et al. The application of a convolution neural network on face and license plate detection[C]//18th international conference on pattern recognition. Hong Kong, China; IEEE, 2006:552-555.

[11] LAUER F, SUEN C Y, BLOCH G, et al. A trainable feature extractor for handwritten digit recognition[J]. Pattern Recognition, 2007,40(6):1816-1824.

[12] SUN Yi, WANG Xiaogang, TANG Xiaou. Deep convolutional network cascade for facial point detection[C]//IEEE conference on computer vision and pattern recognition. Portland, OR, USA; IEEE, 2013:3476-3483.

[13] 常亮,邓小明,周明全,等. 图像理解中的卷积神经网络[J]. 自动化学报,2016,42(9):1300-1312.

[14] 蔡娟,蔡坚勇,廖晓东,等. 基于卷积神经网络的手势识别初探[J]. 计算机系统应用,2015,24(4):113-117.