

混合骨骼特征的三帧间差分手势识别方法研究

张永强^{1,2}, 宋美霖¹, 刘天虎¹, 满梦华²

(1. 河北科技大学信息科学与工程学院, 河北 石家庄 050018;

2. 陆军工程大学石家庄校区电磁环境效应国家级重点实验室, 河北 石家庄 050003)

摘要: 三帧间差分检测作为手势识别中的一种优秀算法, 可在一定程度上解决二帧间差分法的“重影”问题。但三帧间差分检测在识别手势时会出现空洞, 无法适应光照骤变等情况。针对此问题, 提出了一种混合骨骼特征的三帧间差分手势识别方法。该方法首先对数据集进行二值化以统一大小, 降低背景颜色干扰和计算量; 然后使用混合骨骼特征的三帧间差分进行手势检测与追踪; 最后通过神经网络进行手势识别分类。该方法可以有效地重点训练手部特征, 显著降低了背景对手势识别的干扰, 同时还提升了识别效率。经实验验证, 该方法在复杂背景下的识别率最低为 93.38%, 最高为 99.99%, 满足鲁棒性的要求, 为在复杂图像背景下的手势识别提供了一个新思路。

关键词: 手势识别; 三帧间差分; 骨骼特征; 神经网络

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202211

Research on three frame difference gesture recognition method based on mixed bone features

ZHANG Yongqiang^{1,2}, SONG Meilin¹, LIU Tianhu¹, MAN Menghua²

1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

2. National Key Laboratory on Electromagnetic Environment Effects, Shijiazhuang Campus, Army Engineering University, Shijiazhuang 050003, China

Abstract: As an excellent algorithm in gesture recognition, three frame difference detection can solve the “double shadow” problem of two frame difference method to a certain extent. But the three frame difference detection will have holes when recognizing gestures, and it cannot adapt to sudden changes in lighting and so on. To solve this problem, the three frame difference gesture recognition method based on mixed bone features was proposed. Firstly, the binary size of the data set was unified, and the background color interference and computation were reduced. Secondly, the three frame difference of mixed bone features was used to detect and track gestures. Finally, the neural network was used for gesture recognition and classification. This method could effectively train hand characteristics, significantly reduce the interference of background on gesture recognition, and improve the recognition efficiency. The experimental results showed that the minimum recognition rate of this method was 93.38% and the maximum was 99.99% in complex background, which could meet the requirement of robustness. This method provides a new idea for gesture recognition in complex image background.

Key words: gesture recognition, three frame difference, bone feature, neural network

收稿日期: 2021-08-20; 修回日期: 2021-10-17

通信作者: 满梦华, manmenghua@126.com

基金项目: 国防基础科研计划 (No.JCKYS2020DC202); 河北省自然科学基金资助项目 (No.F2018208116); 河北省高等学校科学技术研究重点项目 (No.ZD2020176, No.ZD2021048)

Foundation Items: The National Defense Basic Research Program (No.JCKYS2020DC202), The Natural Science Foundation of Hebei Province (No.F2018208116), The Key Project of Science and Technology Research in Colleges and Universities of Hebei Province (No.ZD2020176, No.ZD2021048)

0 引言

随着人工智能技术的迅速发展,智能化设备在生活中被广泛应用,人机交互的需求迅速增加。手势作为一种高效直观的交流方式,比人脸和表情更灵活,因而识别难度更大。在复杂图像背景下进行手势识别是本领域的研究热点。

神经网络应用在手势识别中有很大的优势,识别率高,但直接使用神经网络并不能解决人体肤色的干扰问题。因此本文提出了混合骨骼特征的三帧间差分手势识别方法,采用 VGGNet (visual geometry group network),充分利用手部骨骼特征进行手势识别,在解决复杂背景问题的同时减少花销,并且其识别速度较快,满足实时性的要求。本文将基于视觉的手势识别与利用手部骨骼特征来训练的方法进行结合,其中手部骨骼特征是单独预先训练的,训练出骨骼模型后,将骨骼模型导入神经网络对数据集进行训练,生成最终的模型。该模型可以很好地实现手势与背景的分割,较好地降低背景的影响。该方法利用生物骨骼特性,在面对手部肤色以及手势背景的问题时具有较强的鲁棒性,为手势识别提供了新的研究思路。

1 相关工作

1.1 手势识别

为了消除手势背景的影响,基于几何特征的识别方法采用提取手势的几何特征信息的方式,具有识别速度较快、误差较小、稳定性较好的优点,但不能通过增加样本量来提高识别率^[1];基于隐马尔可夫模型的识别方法对形状差异较小的手势有较好的识别效果,能够描述手势的时空变化,但识别速度不理想^[2]。Sangi P 等人^[3]通过提取梯度直方图 (histogram of oriented gradient, HOG) 来描述手部特征,并利用神经网络进行手势识别,对手势图像的光照变化和小幅度旋转具有较强的鲁棒性。但是,如果图像有人脸等肤色背景,梯度直方图特征会受到很大影响,因此算法的识别率较低。Mohanty A 等人^[4]采用简单神经网络直接识别原始图像的手势,识别速度快,识别率高。然而,与 HOG 一样,当图像中存在人体肤色等复杂背景干扰时,神经网络无法学习有效的识别模式,识别率大大降低。上述方法虽然对一般的复杂背景有较强的鲁棒性,但是没有很好地解决肤色导致的识别率降低的问题。

R-CNN (region convolutional neural network)^[5]、Faster R-CNN^[6]、YOLO (you only look once)^[7]、SSD (single shot multi box detector)^[8]等大型网络框架能够很好地实现不同肤色背景的手势识别,因为大型网络具有较强的学习能力和抗干扰能力,且能很好地适应各种非线性映射。它们具有计算速度快、鲁棒性强、泛化能力强等优点,充分体现了大型网络的优点。为了减轻计算压力,本文采取轻量型神经网络。

1.2 三帧间差分法

三帧间差分法的原理是在相邻两帧之间得到一张差分图像,然后对差分图像进行“与”运算。这两张图像都包含了运动变化后的区域,对这两张图像进行“与”运算,并取其交点,从而得到与物体相似的轮廓^[8]。传统的三帧间差分法存在“阴影”和“空洞”问题^[9],这类轮廓的噪声不可忽视,因此,去除噪声是图像处理的关键步骤。结合数学形态学理论,腐蚀操作可以降低图像的噪声。对三帧间差分法得到的二值图像进行腐蚀,可以去除提取到的目标图像的多余部分,使目标轮廓更加清晰。

1.3 基于骨骼特征的手势识别

基于骨骼特征比基于二维图像特征进行手势位置定位、手势分类、背景切割的效果好,因此越来越多的学者开始研究骨骼特征以进行分类处理工作。现阶段主要采用体感设备来获取手部骨骼数据。例如皮览月^[10]利用 Leap Motion 和 Intel RealSense 体感设备来获取手势数据,建立了一个大型动态手势数据库,改进了传统的特征提取方法,使用支持向量机 (support vector machine, SVM) 对手势进行分类,准确率达到了 97.5%。王松林等人^[11]采用微软 Kinect 摄像头获取深度信息和人体骨骼信息,并根据人体骨骼信息快速找到手掌的位置,利用轮廓检测算法将手掌轮廓和弯曲手指轮廓从背景图像中分割出来,再检测手指指尖的位置。因为骨骼特征是容易提取的,同时能更好地应对复杂环境尤其是光照影响的问题,因此,本文采用三帧间差分与手势骨骼特征相结合的方法进行手势识别工作,希望在识别率上有所突破。

2 方法

2.1 手势识别流程

手势识别包括手势检测、手势追踪、手势特征

提取和手势分类。在进行手势检测时需要将手势与背景分割开,分割的效果直接影响最终的手势识别效果。为了解决二帧间差分运动对象内部的灰度值相近、不能提取出对象的完整区域等问题,本文使用了三帧间差分法^[12]进行手势识别。手势追踪的本质是对时序上连续的图像进行逐帧分析,在图像变化之间计算出被追踪的目标。由于手势变化非常快,并且目标区域较小,因此对手势的追踪比较困难。本文使用三帧间差分法来实现手势的追踪,三帧间差分法在一定程度上更能与目标的运动速度相匹配,提高了系统的鲁棒性^[13]。手势特征提取是手势识别的重要步骤,在短时间内由动作连续的一系列姿势^[14]提取出视频中的手势,特征提取将输入图像所表达的主要内容转化为特征向量。特征提取的质量将直接影响最终的识别效果。为了降低背景干扰,本文方法充分将手势骨骼作为特征。利用训练好的骨骼模型确定手势范围,着重提取感兴趣的特征。手势分类是对提取的特征进行分类^[15],是手势识别的最后一个关键环节。基于神经网络的方法能够从大量原始输入中自动学习样本的高级特征,具有较强的自组织和自学习能力。本文方法选用在卷积神经网络(convolutional neural network, CNN)基础上发展而来的 VGGNet,使用 Keras 框架搭建了一个轻量级神经网络,除了 Keras 还有许多深度学习编程框架^[16-17],通过 Numpy 进行数据分析,减少了计算量和训练时间,在改进之后能够有效地进行实时分类。手势识别流程如图 1 所示。

2.2 VGGNet 网络设置

VGGNet 网络设置见表 1。VGGNet 采用 1 层卷积+1 层池化+2 层卷积+1 层池化+3 层卷积+1 层池化+1 层全连接的网络结构,是一种轻量级的网络。网络神经元个数为输出图形的神经元各参数相乘的结果,例如第一层卷积层的神经元个数为 $64 \times 64 \times 32$ 。

2.3 三帧间差分法

三帧间差分法步骤如下。

步骤 1: 选取视频序列中某时刻连续的三帧图像,分别记为 $f_{t-1}(x, y)$ 、 $f_t(x, y)$ 、 $f_{t+1}(x, y)$, 其中 (x, y) 为像素坐标, t 为时间。

表 1 VGGNet 网络设置

层	输出格式
conv2d_1	(None, 64, 64, 32)
activation_1	(None, 64, 64, 32)
batch_normalization_1	(None, 64, 64, 32)
max_pooling2d_1	(None, 32, 32, 32)
conv2d_2	(None, 32, 32, 64)
activation_2	(None, 32, 32, 64)
batch_normalization_2	(None, 32, 32, 64)
conv2d_3	(None, 32, 32, 64)
activation_3	(None, 32, 32, 64)
batch_normalization_3	(None, 32, 32, 64)
max_pooling2d_2	(None, 16, 16, 64)
conv2d_4	(None, 16, 16, 128)
activation_4	(None, 16, 16, 128)
batch_normalization_4	(None, 16, 16, 128)
conv2d_5	(None, 16, 16, 128)
activation_5	(None, 16, 16, 128)
batch_normalization_5	(None, 16, 16, 128)
conv2d_6	(None, 16, 16, 128)
activation_6	(None, 16, 16, 128)
batch_normalization_6	(None, 16, 16, 128)
max_pooling2d_3	(None, 8, 8, 128)
flatten_1	(None, 8 192)
dense_1	(None, 512)
activation_7	(None, 512)
batch_normalization_7	(None, 512)
dropout_1	(None, 512)
dense_2	(None, 4)
activation_8	(None, 4)

步骤 2: 分别对相邻两帧进行差分操作,获得差分图像 $D_1(x, y)$ 以及 $D_2(x, y)$, 如式 (1) 所示。

$$\begin{cases} D_1(x, y) = f_t(x, y) - f_{t-1}(x, y) \\ D_2(x, y) = f_{t+1}(x, y) - f_t(x, y) \end{cases} \quad (1)$$

步骤 3: 对 $D_1(x, y)$ 、 $D_2(x, y)$ 进行二值化处理,处理方式如式 (2) 所示。

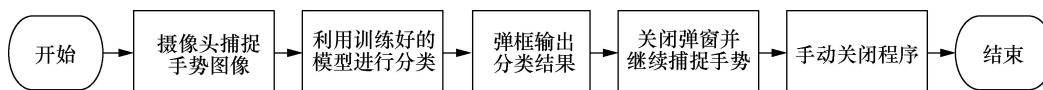


图 1 手势识别流程

$$h_i(x, y) = \begin{cases} 1, D_i(x, y) \geq T \\ 0, D_i(x, y) < T \end{cases} \quad (2)$$

其中, $i=1,2$, T 为预设值的阈值, $h_i(x, y)$ 为二值化处理的结果。

步骤 4: 将得到的两张二值化图像进行“与”运算, 得到手势前景图, 如式 (3) 所示。

$$X = h_1(x, y) \cap h_2(x, y) \quad (3)$$

虽然三帧间差分法可以在一定程度上消除三帧间差分法的“重影”问题, 但是仍然无法避免检测出的运动目标轮廓不完整以及内部有“空洞”的问题, 依然需要与其他方法结合以达到更好的效果。三帧间差分法运动手势检测如图 2 所示。



图2 三帧间差分法运动手势检测

2.4 提取手部骨骼特征

提取手部骨骼特征的关键在于寻找手势关键点。首先, 使用初始关键点检测器在手的多个视图中产生无序的标签。然后, 使用多视图几何图形在 3D 环境中对噪声检测进行三角测量。最后, 将重新投影的三角部分作为新的标记训练数据, 以改进检测器^[18]。本文方法基于 OpenPose^[19], 设置 22 个特征点, 包括手部的 21 个点和表示背景的第 22 个点。手部的 21 个特征点如图 3 所示。



图3 手部的 21 个特征点

2.5 三帧间差分法与骨骼特征融合

(1) 视频数据读入: 将输入的监控视频读为一帧一帧的图像并保存; 在对视频进行处理时, 选取视频序列中某时刻连续的 3 帧图像, 分别记为 $f_{t-1}(x, y)$ 、 $f_t(x, y)$ 、 $f_{t+1}(x, y)$ 。

(2) 骨骼特征提取: 对读取的 3 帧图像进行骨骼特征提取; 得到有骨骼特征的灰度图像, 分别记作 $I_{t-1}(x, y)$ 、 $I_t(x, y)$ 、 $I_{t+1}(x, y)$ 。

(3) 三帧间差分法: 分别对 $I_{t-1}(x, y)$ 、 $I_t(x, y)$ 、 $I_{t+1}(x, y)$ 进行帧间差分, 得到差分图像 $G_1(x, y)$ 、 $G_2(x, y)$ 。

(4) 二值化处理: 根据式 (2) 将 $G_1(x, y)$ 、 $G_2(x, y)$ 转化为二值化图像, 记作 $h_i(x, y)$ 。

(5) 将两张二值化图像进行式 (3) 的“与”运算得到手势前景图。

基于手部骨骼特征的训练过程如图 4 所示。

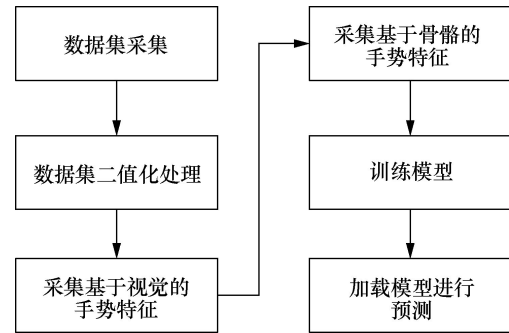


图4 基于手部骨骼特征的训练过程

3 实验

首先进行基于视觉的手势识别训练, 对数据进行切分及增强处理, 然后加载神经网络进行训练。为了避免因过度提取背景颜色特征而对实验产生负影响, 在加载数据时进行二值化处理, 并在此基础上进行基于手部骨骼特征的训练, 提取设置的 22 个特征点, 继而进行训练生成模型, 最后进行预测。

本文使用的数据集为 256×256 像素的手势数据集, 共分为 4 类: 剪刀、石头、布和其他, 将两个男生和两个女生的手作为手势采集的对象, 本数据集为实验室内部采集, 数据集通过 OpenCV 调用摄像头获取, 将标签数据经过 one-hot 编码处理成标签向量。原始数据集共 4 000 张图像, 每类图像 1 000 张, 每张图像的大小是 256×256 像素, 本文将数据集命名为 Gesture 数据集, 数据集部分图像如图 5 所示。

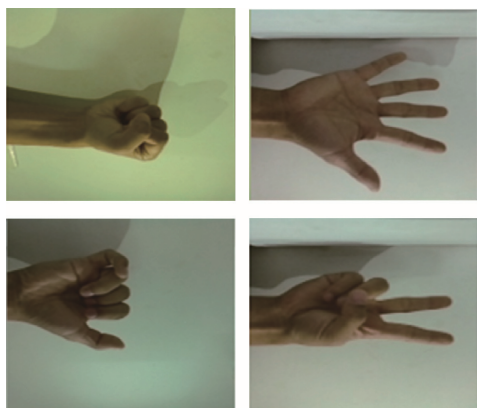


图5 数据集部分图像

3.1 数据预处理

Gesture 数据集中每类包含 1 000 张图像, 数据量少, 而且手势方向基本一致, 因此需要对数据进行图像增强处理。对每张图像进行旋转, 旋转方式分别为绕 X 轴、 Y 轴和同时绕 X 和 Y 轴旋转, 处理后得到多个方向的手势数据, 每类数据量达到 4 000 张图像, 可以更好地应对手势方向带来的负影响。

光照的突然变化及其他噪声引发的像素变化会导致图像被包含在差分图像中。因此, 进行图像相邻帧间差分运算及二值化处理, 得到差值图像并去噪。确定二值化图像的边界, 并从相应的位置提取目标图形^[20]。处理后的数据集, 即基于视觉的手势识别数据集如图 6 所示。



图6 基于视觉的手势识别数据集

在基于骨骼特征的手势识别中, 首先对数据集图像中的手势提取设置的 22 个特征点, 对提取出的特征点的数据进行训练, 基于骨骼特征的手势识别数据集如图 7 所示。

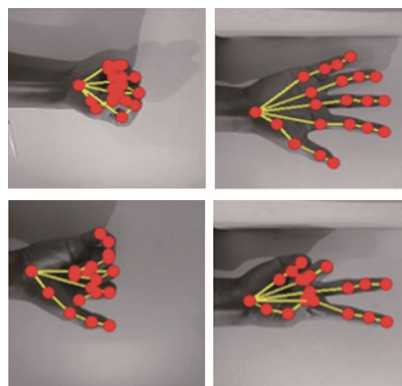


图7 基于骨骼特征的手势识别数据集

3.2 训练模型

首先对图像数据做比例缩放操作, 实现图像的线性变换。其次将数据集切分为 25% 的测试集、75% 的训练集。随后设置初始化超参数, 具体见表 2。超参数设置采用控制变量的方法, 直至达到最优的设置方式。

表2 超参数设置

超参数	设置
初始化学率	0.01
训练总轮数 (epoch)	30
每轮训练次数	128
批次大小	32
激活函数	ReLU
dropout	0.6

基于视觉的手势识别每一轮的训练时间大约为 79 s, 而基于骨骼特征的手势识别每一轮的训练时间大约为 82 s。基于视觉的手势识别损失函数和准确率如图 8 所示, 基于骨骼特征的手势识别损失函数和准确率如图 9 所示。其中 `train_loss` 代表训练时的损失函数; `val_loss` 代表测试时的损失函数; `train_acc` 代表训练时的准确率; `val_acc` 代表测试时的准确率。

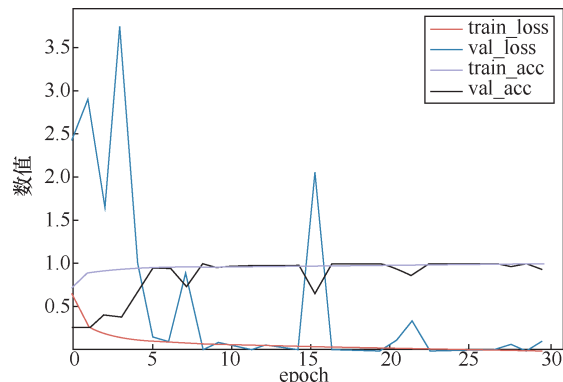


图8 基于视觉的手势识别损失函数和准确率

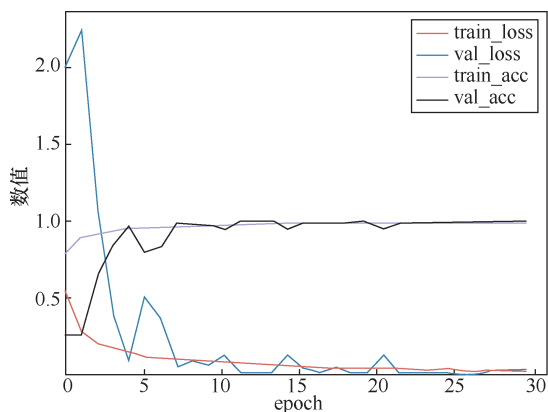


图9 基于骨骼特征的手势识别损失函数和准确率

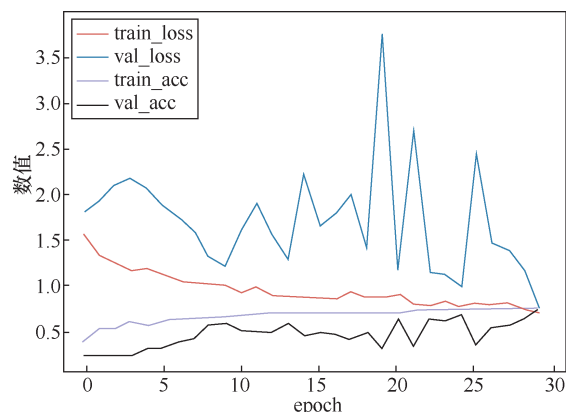


图10 基于骨骼特征的手势识别在 dropout=0.5 时的结果

对比图8和图9,测试集准确率和训练集准确率均趋于1,同时损失函数趋于0,达到了预期的效果。

为了测试 dropout 对结果的影响,在查阅参考文献[21]后,得知 dropout 一般选择 0.5,为了验证效果,设置了 dropout 为 0.5 时的实验,结果如图 10 所示,训练效果并不理想,此时过拟合比较严重。继而进行其他数值测试,经实验验证,dropout 为 0.6 时效果最好。

3.3 实验结果分析

图 11、图 12 给出了基于视觉和骨骼特征的手势识别准确率,训练时间增加,但是准确率有所提升,识别时间由原来的 1.94 s 降为 1.02 s。通过对结果进行分析,证明了本文方法技术可行,稳定高效,可以有效解决在复杂图像背景下的手势识别准确率较差的问题。

3.4 对比试验

基于 Gesture 数据集,分别使用 ResNet、AlexNet 进行手势分类训练,训练过程中未使用骨骼特征,训练结果中损失函数为 0.4356,准确率为 0.875, AlexNet 测试结果如图 13 所示。

从图 13 可以看出,在不使用骨骼特征的情况下,对剪刀这一类别的识别错误率高,且对于简单背景具有很好的识别效果,但是对光线复杂背景的数据识别成功率低,鲁棒性差。为了验证基于骨骼特征的手势识别的可行性,笔者在 AlexNet 基础上进行了改进,通过判断骨骼特征进行手势的识别,基于骨骼特征的 AlexNet 测试结果如图 14 所示。从图 14 中可以看出,使用骨骼特征后可以避免不使用骨骼特征时的严重误判情况,对识别效果的提升有显著的帮助。

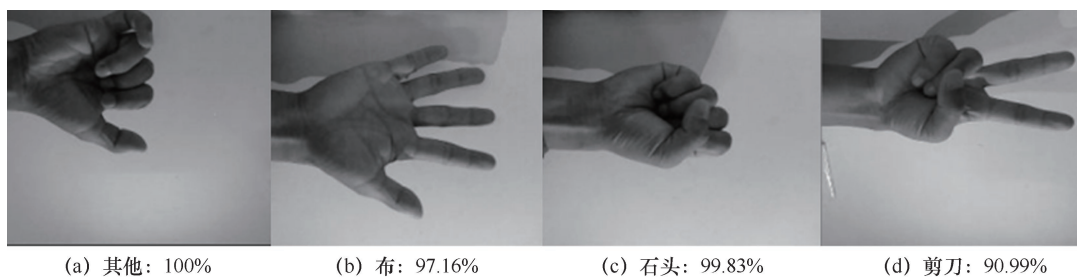


图11 基于视觉的手势识别准确率

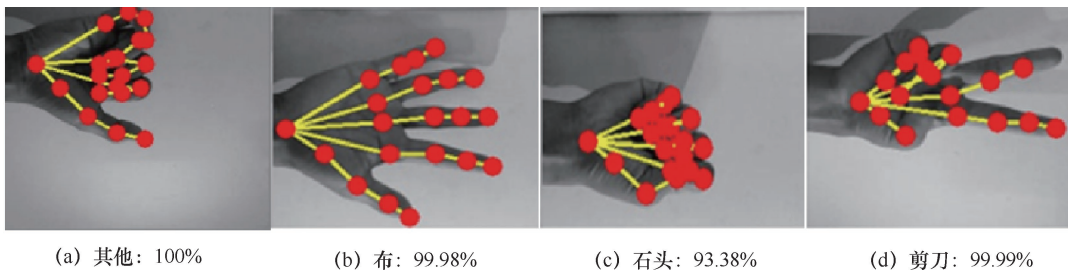


图12 基于骨骼特征的手势识别准确率

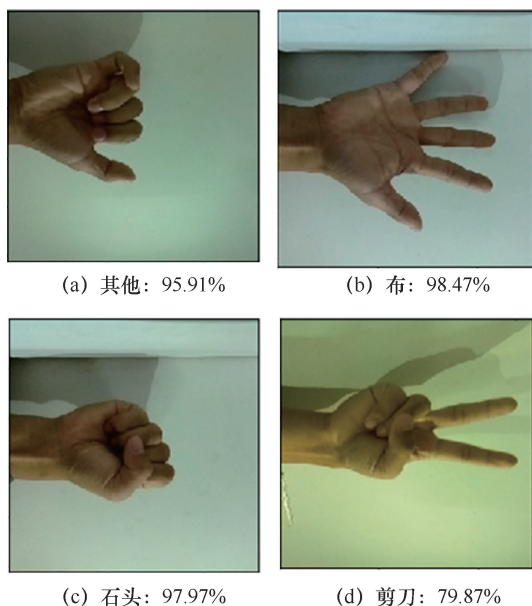


图13 AlexNet 测试结果

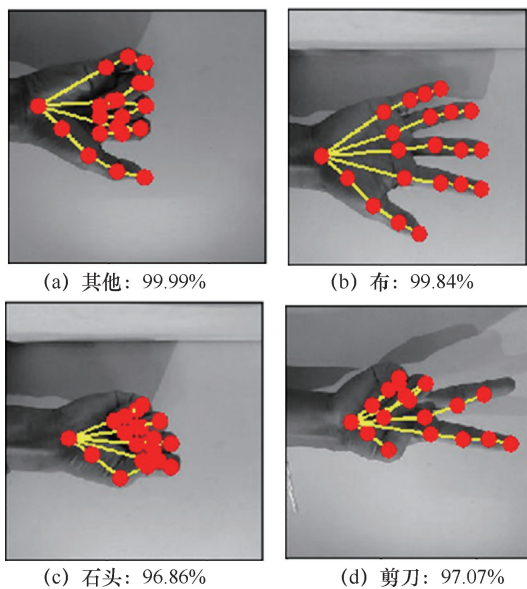


图14 基于骨骼特征的 AlexNet 测试结果

为了进一步进行对比试验,笔者使用 ResNet 进行了手势识别的实验,其中类别是其他的编号为 0,布为 1,石头为 2,剪刀为 3。不使用骨骼特征的 ResNet 测试结果如图 15 所示。在使用骨骼特征后,可以更加准确地判断类别。

对于比较简单的背景,ResNet 在不使用骨骼特征时准确率非常高,此时识别时间为 2.57 s。使用骨骼特征识别时间是 1.56 s,识别速度得到提升。在 AlexNet 基础上不使用骨骼特征时识别时间为 2.31 s,使用骨骼特征时识别时间为 1.25 s,识别速度也得到提升。对比实验识别时间结果见表 3。

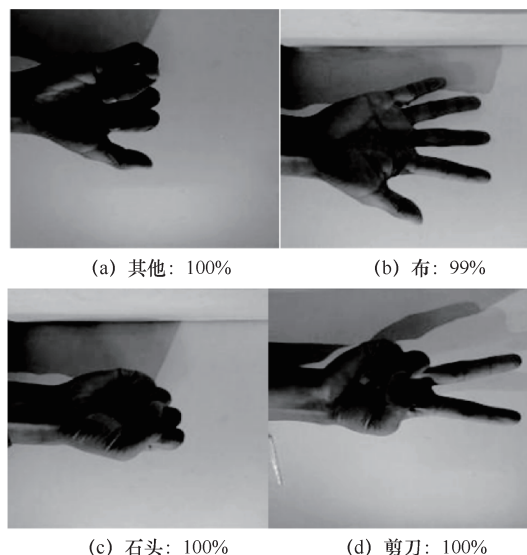


图15 不使用骨骼特征的 ResNet 测试结果

表3 对比实验识别时间结果

网络	不使用骨骼特征识别时间/s	使用骨骼特征识别时间/s
AlexNet	2.31	1.25
ResNet	2.57	1.56

4 结束语

本文使用了混合骨骼特征的三帧间差分法进行石头剪刀布的手势识别。先提取出手势骨骼特征,然后利用三帧间差分法进行手势识别。通过实验可以看出,在手势图像的复杂背景下,单独的基于视觉的手势识别效率低于基于骨骼特征的手势识别,基于骨骼特征的手势识别在面对复杂背景,尤其是在光照强度的影响下有更好的效果。同时该方法降低了对硬件计算性能的要求,识别效率得到了提升,使用场景得到了扩展。

随着机器视觉不断发展,手势的精确快速识别将拥有更广阔的前景,例如汽车手势操控的驾驶技术可在一定程度上提高驾驶员的安全性。本文将深度学习和手势识别结合,提出了基于骨骼特征的手势识别的方法,由实验可以看出该方法在应对此类问题时具有更好的效果。该方法利用了生物骨骼特性,面对手部肤色各异以及手势背景复杂的问题时具有较强的鲁棒性,为研究者提供了一个新的研究思路。

参考文献:

- [1] HAMDA M, MAHMOUDI A. Hand gesture recognition using Kinect's geometric and HOG features[C]//Proceedings of the 2nd International Conference on Big Data, Cloud and Applications. New York: ACM Press, 2017: 1-5.

- [2] MALYSA G, WANG D, NETSCH L, et al. Hidden Markov model-based gesture recognition with FMCW radar[C]//Proceedings of 2016 IEEE Global Conference on Signal and Information Processing. Piscataway: IEEE Press, 2016: 1017-1021.
- [3] SANGI P, MATILAINEN M, SILVÉN O. Rotation tolerant hand pose recognition using aggregation of gradient orientations[C]//Proceedings of 2016 International Conference on Image Analysis and Recognition. Cham: Springer, 2016.
- [4] MOHANTY A, RAMBHATLA S S, SAHAY R R. Deep gesture: static hand gesture recognition using CNN[M]//Advances in intelligent systems and computing. Singapore: Springer Singapore, 2016: 449-461.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.
- [8] ZHANG Z G, ZHANG H M, ZHANG Z F. Using three frame difference algorithm to detect moving objects[C]//Proceedings of the International Conference on Cyber Security Intelligence and Analytics. [S.l.:s.n.], 2019.
- [9] JU J G, XING J S. Moving object detection based on smoothing three frame difference method fused with RPCA[J]. Multimedia Tools and Applications, 2019, 78(21): 29937-29951.
- [10] 皮览月. 基于 3D 骨骼数据的动态手势识别研究[D]. 郑州: 郑州大学, 2019.
PI L Y. Research on dynamic gesture recognition based on 3D bone data[D]. Zhengzhou: Zhengzhou University, 2019.
- [11] 王松林, 徐文胜. 基于 Kinect 深度信息与骨骼信息的手指尖识别方法[J]. 计算机工程与应用, 2016, 52(3): 169-173, 235.
WANG S L, XU W S. Fingertips recognition based on Kinect depth and skeleton data[J]. Computer Engineering and Applications, 2016, 52(3): 169-173, 235.
- [12] 高远, 路杨, 李佩琛. 一种基于三帧差分混合高斯模型的运动目标检测方法[J]. 科技创新与应用, 2017(31): 21-23.
GAO Y, LU Y, LI P C. A moving target detection method based on three-frame difference and mixed gaussian model[J]. Technology Innovation and Application, 2017(31): 21-23.
- [13] 尹红娟, 栾帅. 三帧差分运动目标检测算法分析与验证[J]. 计算机与数字工程, 2017, 45(1): 69-71, 87.
YIN H J, LUAN S. Three-image difference algorithm for moving target detection[J]. Computer & Digital Engineering, 2017, 45(1): 69-71, 87.
- [14] BADI H, HUSSEIN S H, KAREEM S A. Retraction note to: feature extraction and ML techniques for static gesture recognition[J]. Neural Computing and Applications, 2017, 28(4): 851.
- [15] JADOOKI S, MOHAMAD D, SABA T, et al. Fused features mining for depth-based hand gesture recognition to classify blind human communication[J]. Neural Computing and Applications, 2017, 28(11): 3285-3294.
- [16] 王秉睿, 兰慧盈, 陈云霁. 深度学习编程框架[J]. 大数据, 2018, 4(4): 56-63.
WANG B R, LAN H Y, CHEN Y J. Programming frameworks for deep learning algorithms[J]. Big Data Research, 2018, 4(4): 56-63.
- [17] 于璠. 新一代深度学习框架研究[J]. 大数据, 2020, 6(4): 69-80.
YU F. Research on the next-generation deep learning framework[J]. Big Data Research, 2020, 6(4): 69-80.
- [18] SIMON T, JOO H, MATTHEWS I, et al. Hand keypoint detection in single images using multiview bootstrapping[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 4645-4653.
- [19] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172-186.
- [20] 谢之宇, 蒋晓瑜, 汪熙, 等. 改进的帧差法在目标匹配中的应用[J]. 计算机工程与应用, 2011, 47(34): 208-211.
XIE Z Y, JIANG X Y, WANG X, et al. Improved frame difference algorithm in application of target-matching[J]. Computer Engineering and Applications, 2011, 47(34): 208-211.
- [21] BOUTHILLIER X, KONDA K, VINCENT P, et al. Dropout as data augmentation[J]. arXiv preprint, 2015, arXiv:1506.08700.

[作者简介]



张永强(1981—), 男, 博士, 河北科技大学信息科学与工程学院副教授, 主要研究方向为人工智能、物联网。



宋美霖(1997—), 女, 河北科技大学信息科学与工程学院硕士生, 主要研究方向为人工智能和图像处理。



刘天虎(1995—), 男, 河北科技大学信息科学与工程学院硕士生, 主要研究方向为人工智能和图像处理。



满梦华(1984—), 男, 博士, 陆军工程大学石家庄校区电磁环境效应国家级重点讲师, 主要研究方向为电磁防护理论与技术。