轨迹特征融合双流模型的动态手势识别

林 玲[1,2],陈姚节[1,2,3],徐 新[1,2],郭同欢[1,2]
(1. 武汉科技大学 计算机科学与技术学院,湖北 武汉 430070;
2. 智能信息处理与实时工业系统湖北省重点实验室,湖北 武汉 430070;
3. 冶金工业过程国家级虚拟仿真实验教学中心,湖北 武汉 430070)

摘 要:针对现有动态手势识别任务的识别率不高、鲁棒性不强等问题,提出一种新的动态手势识别方法。该方法将轨迹特征与手型时空特征融合到自适应分配权值的双流网络模型中,实现动态手势有效准确的识别。通过 Kinect 采集到整个动态手势的深度图序列和彩色图序列,从中提取出动态手势的轨迹特征曲线图与手型特征变化序列图;而后利用 2D 残差网络对动态手势的轨迹特征曲线图进行识别,得到轨迹信息识别结果;同时采用二模态训练后的 3D 双卷积神经网络对动态手势时空信息进行识别,得到时空网络识别结果;再根据两种网络的识别结果通过自适应分配权值进行融合得到最终的识别结果。实验结果表明,该方法在自制 SKIG 数据集上的识别率平均为 99.52%,相比于其他方法取得了更高的识别精度,体现了该方法的鲁棒性与优越性。

关键词:轨迹识别;时空信息识别;双流网络;自适应分配权值;手势识别
中图分类号:TP391.7 **文献标识码**:A **文章编号**:1673-629X(2020)12-0034-06
doi:10.3969/j.issn.1673-629X.2020.12.006

Dynamic Gesture Recognition Based on Trajectory Feature Recognition and Fusion Dual-flow Model

LIN Ling[1,2],CHEN Yao-jie[1,2,3],XU Xin[1,2],GUO Tong-huan[1,2]
(1. School of Computer Science and Technology,Wuhan University of Science and Technology,
Wuhan 430070,China;
2. Key Laboratory of Intelligent Information Processing and Real-time Industrial System of Hubei Province,
Wuhan 430070,China;
3. Metallurgical Industry Process National Virtual Simulation Experimental Teaching Center,Wuhan 430070,China)

Abstract:Aiming at the problems of low recognition rate and low robustness of existing dynamic gesture recognition tasks,a new method for dynamic gesture recognition is proposed,which integrates trajectory features and hand-space-time features into a dual-flow network model with adaptively assigned weights to achieve effective and accurate recognition of dynamic gestures. The depth map sequence and color map sequence of the entire dynamic gesture are collected by Kinect,and the trajectory characteristic curve diagram and hand shape characteristic sequence diagram of the dynamic gesture are extracted from it. Then,the 2D residual network is used to identify the trajectory characteristic curve diagram of the dynamic gesture and obtain the trajectory information recognition result. At the same time,the 3D dual convolutional neural network after two-modal training is used to recognize the spatiotemporal information of dynamic gestures to obtain the recognition result of the spatiotemporal network,and then based on the recognition results of the two networks,adaptive weight assignment is performed. The final recognition result is obtained by fusion. The experiment shows that the average recognition rate of this method on the homemade SKIG dataset is 99.52%. Compared with other methods,it has achieved higher recognition accuracy,which shows its robustness and superiority.

Key words:trajectory recognition;spatio-temporal information recognition;dual stream network;adaptive weights assigning;gesture recognition

收稿日期:2020-01-02 **修回日期**:2020-05-06
基金项目:国家自然科学基金(U1803262)
作者简介:林 玲(1996-),女,硕士,研究方向为计算机应用与仿真;陈姚节,高级工程师,研究方向为计算机仿真;徐 新,教授,研究方向为人工智能和图像处理。

0　引　言

手势识别作为一种重要的交互方式,由于更自然,直观和易于学习的特点,在虚拟仿真、手语识别等领域得到了大量应用。基于视觉的手势识别主要分为三个阶段:手势分割、特征提取和识别。

手势分割作为手势识别的基础,对后续手势识别工作有着至关重要的影响。传统手势分割利用肤色、轮廓从彩色图像视频中分割出手势,如 Bao 等[1]提出的利用肤色检测与背景差分的方法,Rahmat 等[2]结合人手肤色与光照的实时手势分割,Dawod 等[3]采用自由形式肤色模型进行的手势分割。以上方法进行的手势分割效果较好但易受光照、复杂背景的影响,影响后续的手势识别工作。

手势特征的提取是手势识别更为重要的阶段。Asaari 等[4]根据提取的手形特征与纹理特征进行手势识别,由于复杂背景的影响准确率不高,刘富等[5]借助手形轮廓与几何特征提高了手势识别的鲁棒性,但要求手势手指分开,不具有普遍性。

现有的手势识别大多借助模式分类方法对手势进行识别,如 Panwar[6]利用形状参数的位编码序列进行手势分类的方法、杨学文等[7]利用手势主方向和类 Hausdorff 距离模板匹配的手势识别方法等具有一定局限性,鲁棒性较低。近年来动作识别方法的迅速发展和许多大型数据集的引入,使得利用深度神经网络对动态手势进行有效识别成为可能。Molchanov 等[8]引入了一种将归一化深度和图像梯度值结合起来的 3D-CNN 的动态手势识别方法。而后 Molchanov 等[9]又提出了一种 3D-CNN,融合来自多个传感器的数据流进行识别。3D-CNN 模型在视频处理问题上相比于 2D-CNN 更加有效,但是也会存在时间维度上的运动信息的丢失问题。

因此,该文利用 Kinect 深度信息修复后的深度图进行手势精确分割,并由此提取出动态手势的运动轨迹特征,构建一种通过自适应权值分配将动态手势的轨迹识别与手势时空信息识别结合的双流网络模型,利用该模型中的两种网络对动态手势的不同特征的识别优势提高动态手势识别率,并采用 SKIG 数据集测试模型识别性能。

1　动态手势特征提取

实验发现,当手部位置变化较大时就可以通过运动信息来识别,那么这些动态手势的识别就可以转换为对其空间运动轨迹的识别;而当手部位置变化较小时,其轨迹不能明显区分出各个动态手势,此时就需要利用动态手势的手形特征的变化进行动态手势的识别。因此在进行手势识别前,需要对动态手势进行手势分割和轨迹的提取。

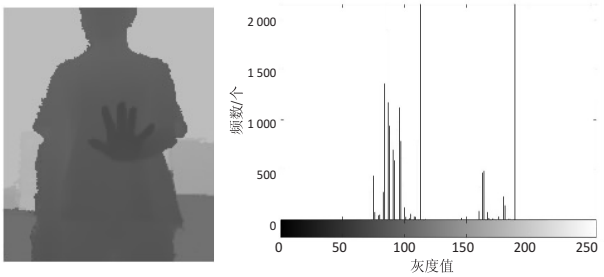
1.1　深度图修复

由于 Kinect 传感器获取的深度图像中存在大量噪声以及深度信息缺失导致的空洞,而动态手势的识别又依赖于手掌在运动过程中的手部形态与精确位置。因此为避免在进行手势分割时,因深度图中的噪声、空洞引起的分割误差进而导致后续的识别误差,笔者首先做了文献[10]中的工作,对采集的深度图像进行初步修复。利用待修复像素点周围时空域的深度数据,对深度图中存在的噪声以及空洞点进行修复,保证后续分割工作中能得到完整的手部形态和精确的空间位置。

1.2　手势分割

手势分割的目的是将手部区域从复杂背景中分离出来。在基于计算机视觉的手势识别技术中,复杂背景下的手势分割非常困难。特别是在单目视觉情况下,这主要是由于背景各种各样,环境因素也不可预见。

修复后的深度图像中手部轮廓完整、没有明显的噪声干扰,因此可以利用深度图中手掌部分的灰度值与深度图中其他位置的灰度值的差异来提取手部感兴趣区域输入网络进行训练,提高动态手势识别准确性。正常情况下,当人位于 Kinect 设备的可视区域内做手势时,手掌部分与 Kinect 相距最近,灰度值与图像中其他部分也会有较大差异,如图 1(a)所示。由此可以借助手势的深度图像,计算生成灰度直方图,如图 1(b)所示。灰度图中横坐标表示灰度级,纵坐标表示各个灰度值的像素在图像中出现的次数。



(a)深度图像　　(b)灰度直方图

图 1　深度图像灰度直方图示例

通过观察灰度直方图分析发现,灰度直方图中第一个波峰对应灰度值即手掌部分对应灰度值。为准确把手掌区域和手臂、手腕部分区分开,将在第一个波峰灰度值左右波动 3 以内的像素点保留,其他像素点像素置为 255。由此就得到了分割后的手势图,如图 2 所示。

1.3　轨迹提取

利用 1.1 节分割得到的手势图,计算图中手部质心坐标来代表手在图像坐标系下的坐标。计算采集的

手部质心坐标序列中横坐标的最大值 x_{max}、最小值 x_{min} 和纵坐标的最大值 y_{max}、最小值 y_{min}，给定一个标志 flag 和由实验得到的质心坐标波动阈值 $P = 20$：

当 $x_{max} - x_{min} < P$ 并且 $y_{max} - y_{min} < P$ 时，flag = false，当前动态手势不能用轨迹识别，不进行后续轨迹图的生成。

图 2 分割后的手势图

当 $x_{max} - x_{min} \geqslant P$ 或 $y_{max} - y_{min} \geqslant P$ 时，flag = true，可以用轨迹对动态手势进行识别。此时，为保证轨迹特征具有平移和比例不变性，将手势的运动轨迹，即质心坐标的变化轨迹，整体平移到图像中心位置，并生成动态手势轨迹图。具体过程如下：

(1)计算手势轨迹所占区域的中心位置坐标 (x_0, y_0)。计算公式如下：

$$\begin{cases} x_0 = \frac{1}{2}(x_{max} - x_{min}) + x_{min} \\ y_0 = \frac{1}{2}(y_{max} - y_{min}) + y_{min} \end{cases} \tag{1}$$

(2)由于网络的输入设置为 150×150 大小的图片，故计算 x_0、y_0 与 75 的差值得到对应的轨迹坐标平移距离，即可将轨迹整体平移至图像中心位置。

(3)绘制轨迹序列散点图，拟合轨迹曲线，生成动态手势轨迹图。

采集 8 帧深度图像代表挥手手势一次来回摆动，经分割后的手势图如图 3 所示。

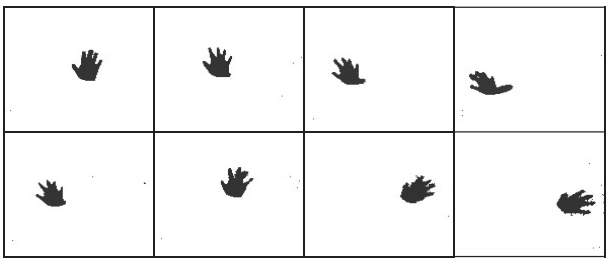
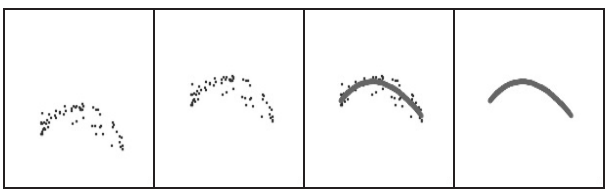


图 3 代表挥手手势一次来回摆动的 8 帧手势图

由整个挥手手势的手势图序列中的手部质心坐标生成轨迹图的过程如图 4 所示。



(a)原质心坐标序列散点图 (b)经过平移后的坐标序列散点图 (c)拟合轨迹图 (d)最终生成的轨迹图

图 4 挥手手势轨迹图生成

2 融合轨迹识别的双流模型

CNN 是一种前馈神经网络[11]，基本结构包括特征提取层和特征映射层。在图像以及视频处理方面，CNN 有明显的优势。相比于静态手势，动态手势还包含了时间维度上的运动信息，因此必须采用 3D-CNN 同时学习手势视频流中的空间特征与时间特征。而一个动态手势从开始到完成的持续时间大约为 2 ~ 3 秒，3D-CNN 并不能将动态手势视频中的每一帧都输入网络进行学习，只能选取一定数量的图像帧代表该动态手势。因此，为防止选取不当导致关键帧信息丢失产生的分类错误，且鉴于 CNN 在提取静态空间结构的优势，该文采用 3D-CNN 对动态手势进行时空信息识别，并采用 2D-ResNet 融合手势轨迹信息识别，构建自适应权值分配的双流网络模型，实现动态手势的识别。网络模型结构如图 5 所示。

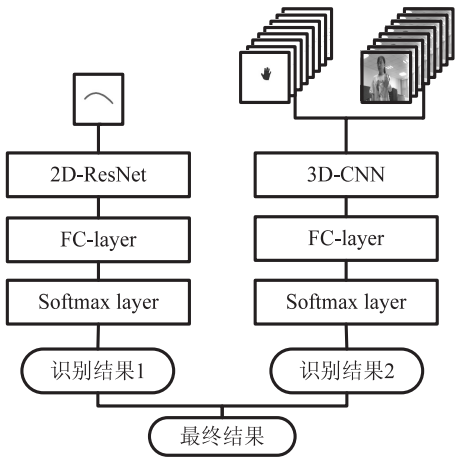


图 5 融合轨迹识别的双流模型结构

2.1 时空信息识别

多模态识别系统使用多个数据流进行训练，并在测试期间对多模态观测结果进行分类，单模态识别系统仅使用一个模态数据进行训练和测试[12]。该文采用了第三种类型，使用一个 3D-CNN 模型接收来自多种模态的数据并融合学习，即利用多模态数据提高单个网络的测试性能。在动态手势识别系统中可用的模式流通常是空间上和时间上对齐的。例如，运动采集设备采集的深度图像和 RGB 图像以及光流通常是对齐的，即使数据以不同的模态出现，但它们代表的语义内容是相同的。

该文引用文献[13]的 3DCNN 模型框架，构建双卷积池化网络。该网络利用两个连续的卷积层保留并传递每个动态手势的特征信息，但 3D 卷积层又是 3D-CNN 中高时空复杂性的主要来源，因此在 3D 卷积核上设置 L2 正则，以避免在神经网络深度有限的前提下，因卷积层密集提取产生过拟合情况。两次卷积操作后添加池化层操作，在保持特征不变性的条件下有

效减少参数数量。在每层卷积之后,设置标准化层实现数据归一化操作。在 3D 卷积之后设计激活函数,激活函数产生非线性操作,进一步增加神经网络的复杂性。由此,利用 Kinect 同时获取彩色数据与深度数据生成图像,对齐裁剪后再对深度图进行手势分割,将分割后的手势图序列与彩色图序列都作为 3D-CNN 的输入数据对网络进行训练,保证网络获得更高识别精度的同时不会带来参数增加的影响。将待识别的手势序列输入训练好的该网络即可得到手势的时空信息识别结果。

2.2　轨迹识别

由于 CNN 模型结构会对网络的特征表达能力产生影响,近年来,用于图像识别的深度网络如 AlexNet、GoogLeNet[14]、VGGNet[15]、ResNet[16] 等被相继提出。卷积核更小化、网络层更深化成为卷积网络结构的一大发展趋势,这种发展趋势使得图像的识别精度更高,模型的计算效率更快。在所有深度网络模型中,残差网络(ResNet)因独特的残差结构,极大地加速了神经网络的训练,模型的准确率有比较大的提升,推广性也非常好,从而得到了广泛的应用。它通过直接将输入信息绕道传到输出,保护信息的完整性,整个网络只需要学习输入、输出差别的那一部分,简化学习目标和难度,一定程度上解决了信息损耗、丢失和梯度消失、梯度爆炸等问题。

引入跳跃连接将目标函数 $F(x)+x$ 的拟合转变为残差函数 $F(x)$ 的拟合,将输入与拟合残差叠加代表网络输出,增强了网络信息流通,降低了数据信息的冗余度。由此,通过训练经典的 ResNet50 网络对动态手势的轨迹图进行识别就得到了该手势轨迹识别的结果。

2.3　融合策略

在经过上述工作后,已经得到了两种网络的最优识别结果,但由于 ResNet 网络只能对产生轨迹的动态手势识别分类,对没有轨迹变化只存在手形变化的动态手势无法识别;而 3D-CNN 虽然可能丢失动态手势时间上的运动信息,但对某些动态手势仍能通过其时空信息进行有效识别。因此这里不宜采用求平均后取概率最大手势的方法得到双流网络的最终识别结果,应根据每个手势样本的具体情况估计出网络识别结果的置信度,依据该置信度计算权值,因此该文提出一种自适应权值分配策略为其分配权值,再由经典的加权平均模型得到识别的最终结果 R。计算公式如式 2 所示,其中 w 为给网络赋予的权值,f 为各个网络的输出。

$$R = w_s f_s + w_e f_e \tag{2}$$

3　双流网络的自适应权值分配

首先根据 1.2 中的 flag 值确定当前动态手势是否产生轨迹:(1)当 flag = false 时,无法通过轨迹直接将动态手势分类,设置 ResNet 网络权值为 0,3D-CNN 的识别结果即为双流网络的最终结果;(2)当 flag = true 时,即两种网络都能对动态手势进行有效识别,此时根据网络识别结果的置信度为其分配权值,方法如下。

一类动态手势可以用一组特征的组合来代表,每种特征又单独形成特征空间,而不同类别的手势又可能出现相同特征,因此形成了特征重叠的区域。当一个手势样本被网络识别后,识别结果中各个类别的概率相差不大时,认为该手势样本处于特征重叠区域;而当识别结果中概率相差较大、较为分散时,认为该手势样本属于非特征重叠区域。这样,就将样本空间分成了特征重叠区域和非特征重叠区域两部分。

若共有 J 个手势类,根据训练集数据估计每种类别的均值 ζ_j 与协方差 $\sum_j$。N_j 为第 j 类中的手势个数,x_k^j 表示类 j 中的第 k 个手势的特征向量,$j = 1,2,\cdots,J$。

$$\zeta_j = \frac{1}{N}\sum_{k=1}^{N_j} x_k^i \tag{3}$$

$$\sum\nolimits_j = \frac{1}{N}\sum_{k=1}^{N_j} x_k^j (x_k^j)^{\mathrm{T}} - \zeta_j (\zeta_j)^{\mathrm{T}} \tag{4}$$

由高斯参数估计手势样本属于每种手势类别的后验概率 $p_j(j = 1,2,\cdots,J)$,将它们组成向量 $P = \{p_j \mid j = 1,2,\cdots,J\}$,其中 J 为手势类别数。这样,就生成了由后验概率估计值组成的 J 维欧氏空间。对每一个特征向量 P,都有一个欧氏空间中的点与其对应。当 P 越接近 $P_{1/J} = \{(p_1,p_2,\cdots,p_J) \mid p_j = 1/J, \forall j\}$ 时,手势样本位于特征重叠区域的可能性越大,对应识别网络的权值越小;P 越远离 $P_{1/J}$ 时,例如当某一 p_j 接近于 1,而其他概率接近 0 时,手势样本位于特征重叠区域的可能性越小,对应识别网络的权值越大。对各个网络识别结果都利用上述方法计算 P 与 $P_{1/J}$ 的欧氏距离 d_n,融合时就可以根据 d_n 给网络分配不同的权值,而后加权融合即可得到双流网络的识别结果。权值计算公式如下:

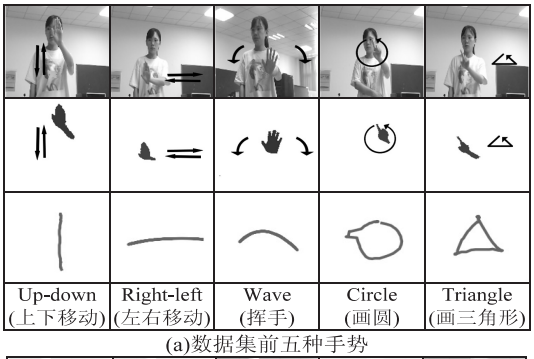
$$w_n = d_n(P, P_{1/J}) \tag{5}$$

4　实验及结果分析

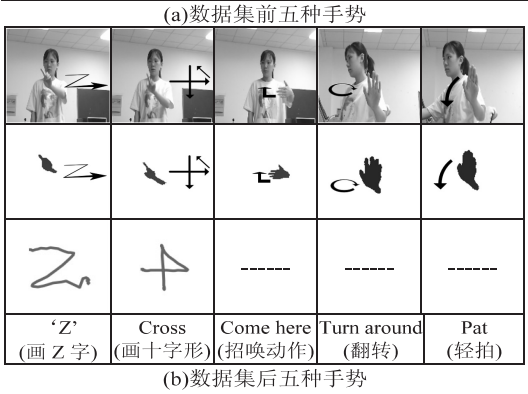
4.1　数据集

由于加入了 ResNet 网络对动态手势轨迹进行识别,并且将分割后的手势深度图处理后与彩色图两种模态的数据同时训练 3D 卷积网络,因此该文采用 Sheffield Kinect Gesture (SKIG) Dataset[17] RGB-D 手

势数据集中的 10 种动态手势类型,利用 Kinect 2 同步获得彩色数据与深度数据,重新制作数据集。数据采集由 6 人完成,每人每种手势执行 10 次,每种模态各 600 个动态手势视频,并按照 8∶1∶1 的比例将数据集随机划分为训练集、验证集、测试集。对除测试集外的深度视频,按照 1.1 的手势分割方法将手掌部分分割出来,然后平均选取 8 帧图像代表该动态手势。再按照 1.2 中所提方法从分割后的手势图序列中计算质心坐标得到轨迹序列并生成轨迹图。数据集样例如图 6 所示。



(a)数据集前五种手势



(b)数据集后五种手势

图 6 数据集样例

4.2 数据扩充与训练

为防止网络在训练过程中出现过拟合现象,有必要对数据集进行数据扩充。分别对 ResNet 网络和 3D-CNN 的输入数据进行扩充。对 3D-CNN 的输入数据采用以下两种数据扩充策略:(1)在同一个手势视频的完整帧序列中,选用不同的帧作为采集的第一帧,平均采集 8 帧图像代表该手势;(2)将代表一个手势的 8 帧图像进行相同方向相同角度的旋转。以上两种方法扩充后共 2 160 个手势。对 ResNet 网络的输入数据即动态手势轨迹图进行一定比例的放大与缩小,最终动态手势轨迹图包含 1 080 张。实验结果表明,利用数据扩充后的数据集对网络模型进行训练,增强了网络的泛化能力,提高了网络的识别率。

该文基于 Keras 深度学习开发框架,利用 GPU 并行加速对两个网络单独进行训练。数据集中 80% 作为训练集,剩余的 20% 作为验证集,并且将训练集随机打乱。在 ResNet 网络中,网络的输入为根据动态手势运动轨迹生成的大小为 150×150×3 的图像,调整大小至 224×224×3。在 3D-CNN 中,将采集的代表一个手势的 8 帧 150×150 的图像序列作为输入数据,网络每次迭代分批次处理大小为 32,并采用 Adam 方法对网络进行优化。训练周期设为 128,每迭代 5 个批次就对测试集进行一次测试,待网络训练至最优时,将 2 个网络的识别结果,在决策级以加权融合的方式判定所属的动态手势类别。

4.3 实验结果分析

实验计算机配置为 Intel Core i5,内存 32 GB RAM,环境配置 Windows10 + python3.6.8 + Tensorflow1.8.0+CUDA9.0,训练使用显卡 NVIDIA GeForce GTX 980Ti,并采用 Kinect 2.0 设备采集手势数据。实验分为两部分:

(1)用测试集中 60 组动态手势单独测试训练好的两个网络的识别效果。其中,ResNet 网络对除 Come here、Turn around、Pat 以外的 7 种动态手势识别进行测试,结果如表 1 所示;3D-CNN 对数据集中的 10 种动态手势识别结果如表 2 所示。

表 1 ResNet 网络识别结果

手势类别	正确识别次数/次	识别率/%
Up-down	60	100
Right-left	55	91.67
Wave	57	95.00
Circle	58	96.67
Triangle	59	98.33
'Z'	60	100
Cross	60	100
平均值	58.43	97.38

表 2 3D-CNN 识别结果

手势类别	正确识别次数/次	识别率/%
Up-down	60	100
Right-left	59	98.33
Wave	60	100
Circle	55	91.67
Triangle	54	90.00
'Z'	59	98.33
Cross	58	96.67
Come here	58	96.67
Turn around	59	98.33
Pat	58	96.67
平均值	58	96.67

由表 1 可以看出,Resnet50 因其强大的学习能力使得在文中自制的轨迹图像数据集上的平均识别率达

到了 97.38%。其中,当 Right-left 手势执行不规范时,轨迹与 Wave 手势有一定的相似性,正确率略微低于其他手势。同时,3D-CNN 对数据集中 10 种动态手势的平均识别率也达到了 96.67%。其中,Circle、Triangle 两种手势因手型一致,在只提取 8 帧代表该动态手势的情况下存在误识别,故正确率低于其他手势。

(2)对由两种网络构成的双流网络模型进行测试,并将文中方法与近几年相关方法在 SKIG 数据集上的识别准确率与平均消耗时间进行对比,如表 3 所示。

表 3　不同方法在 SKIG 上的准确率对比

方法	识别率/%	平均消耗时间/s
3DCNN+CLSTM	98.89	0.053 4
稠密 3DCNN+GRU	99.07	0.062 9
文中方法	99.52	0.070 4

由表 3 可以看出,文中方法不仅在 SKIG 数据集上的识别率达到 99.52%,相比于现有识别率最高的方法提升了 0.45%,也能较快地识别出动态手势。

5　结束语

为避免由于单个 3D 卷积网络特征提取不充分而导致的误分类,且鉴于 CNN 在提取静态空间结构的优势,引入 ResNet 网络从合成的轨迹图像中提取动态手势运动信息,与二模态训练的 3D 卷积网络构成一种更加复杂的双流网络结构来提高动态手势识别的准确性与鲁棒性。实验结果表明,与现有的在 SKIG 数据集上的方法相比,该方法的识别率更高、鲁棒性更强。虽然提出的双流网络提升了一定的识别率,但识别速度仍需要进一步提高。

参考文献:

[1] BAO H, ZHAO X G. Study on hand gesture segmentation [C]//Proceedings of the 2010 international conference on multimedia technology (ICMT). Ningbo: IEEE, 2010: 1-4.

[2] RAHMAT R W, ALTAIRI Z H, SARIPAN M I, et al. Removing shadow for hand segmentation based on background subtraction[C]//Proceedings of the 2012 international conference on advanced computer science applications and technologies (ACSAT). Kuala Lumpur, Malaysia: IEEE, 2012: 481-485.

[3] DAWOD A Y, ABDULLAH J, ALAM M J. A new method for hand segmentation using free-form skin color model [C]//Proceedings of the 3rd international conference on advanced computer theory and engineering (ICACTE). Chengdu: IEEE, 2010: V2-562-V2-566.

[4] ASAARI M S M, SUANDI S A, ROSDI B A. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics[J]. Expert Systems with Applications, 2014, 41(7): 3367-3382.

[5] 刘　富,刘惠影,高　雷,等. 基于手指融合特征和粒子群优化的手形识别[J]. 光学精密工程,2015,23(6):1774-1782.

[6] PANWAR M. Hand gesture recognition based on shape parameters[C]//Proceedings of the international conference on computing, communication and applications. Los Alamitos: IEEE, 2012: 1-6.

[7] 杨学文,冯志全,黄忠柱,等. 结合手势主方向和类-Hausdorff 距离的手势识别[J]. 计算机辅助设计与图形学学报,2016,28(1):75-81.

[8] MOLCHANOV P, GUPTA S, KIM K, et al. Hand gesture recognition with 3d convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW). Boston, MA: IEEE, 2015: 1-7.

[9] MOLCHANOV P, GUPTA P, KIM K, et al. Multi-sensor system for driver's hand-gesture recognition[C]//Automatic face and gesture recognition (FG). Ljubljana: IEEE, 2015: 1-8.

[10] 林　玲,陈姚节,郭同欢. 基于时空域数据融合的 Kinect 深度图像修复算法[J]. 科学技术与工程,2019,19(30):215-220.

[11] 周飞燕,金林鹏,董　军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.

[12] ABAVISANI M, JOZE H R V, PATEL V M. Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training[C]//The IEEE conference on computer vision and pattern recognition (CVPR). Long Beach, CA: IEEE, 2019: 1165-1174.

[13] 李冠东,张春菊,高　飞,等. 双卷积池化结构的 3D-CNN 高光谱遥感影像分类方法[J]. 中国图象图形学报,2019,24(4):639-654.

[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[15] NG J Y H, HAUSKNECHT M, VIJAYANARASIMHAN S, et al. Beyond short snippets: deep networks for video classification[C]//Proceedings of the IEEE computer society conference on computer vision and pattern recognition. Los Alamitos: IEEE, 2015: 4694-4702.

[16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE computer society conference on computer vision and pattern recognition. Los Alamitos: IEEE, 2016: 770-778.

[17] LIU L, SHAO L. Learning discriminative representations from RGB-D video data[C]//Proceedings of the 23rd international joint conference on artificial intelligence. Beijing: [s. n.], 2013: 1493-1500.