

基于自适应多分类器融合的手势识别



刘肖¹ 袁冠^{1,2} 张艳梅¹ 闫秋艳¹ 王志晓¹

1 中国矿业大学计算机科学与技术学院 江苏 徐州 221116

2 教育部矿山数字化工程研究中心 江苏 徐州 221116

(liuxiaocumt2018@163.com)

摘要 为了提高基于可穿戴设备手势识别的性能,针对单分类器在手势识别时会出现偏向性的问题,提出了基于自适应多分类器融合的手势识别方法(Self-adaptive Multi-classifiers Fusion, SAMCF)。首先,针对统计特征无法表征复杂手势之间类内变异性性和相似性的问题, SAMCF 使用卷积神经网络(Convolutional Neural Network, CNN)自动提取具有强表征能力的深度特征;然后,采用多个基本分类器对提取的特征向量进行识别,并通过自适应融合算法决策出最优识别结果,解决了单分类器的偏向性问题;最后,基于数据手套采集的数据集,验证了模型的鲁棒性和泛化能力。实验结果表明, SAMCF 能够有效地提取手势的深度特征,解决单分类器的偏向性问题,提高了手势识别的效率,增强了手势识别的性能,对字符级手势(美国手语和阿拉伯数字)识别的准确率达到 98.23%,较其他算法平均提高了 5%;对单词级手势(中国手语)识别的准确率达到 97.81%,较其他算法平均提高了 4%。

关键词: 手势识别; CNN; 自适应融合算法; 多分类器; 数据手套

中图法分类号 TP391

Hand Gesture Recognition Based on Self-adaptive Multi-classifiers Fusion

LIU Xiao¹, YUAN Guan^{1,2}, ZHANG Yan-mei¹, YAN Qiu-yan¹ and WANG Zhi-xiao¹

1 College of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

2 Digitization of Mine, Engineering Research Center of Ministry of Education, Xuzhou, Jiangsu 221116, China

Abstract In order to improve the performance of hand gesture recognition based on wearable devices, a hand gesture recognition method (SAMCF) based on self-adaptive multi-classifiers fusion is proposed to solve the bias of single classifier in hand gesture recognition. First, for statistical features that cannot characterize intra-class variability and similarity between complex gestures, SAMCF uses a convolutional neural network (CNN) to automatically extract depth features with strong representation capabilities. Then, SAMCF uses multiple basic classifiers to recognize the extracted feature vectors, and determines the optimal recognition result through self-adaptive fusion algorithm, which solves the bias of single classifier. After that, the robustness and generalization ability of the model are verified by using the data set collected by data glove. The experimental results show that SAMCF can effectively extract the depth features of gesture, solve the bias of single classifier, and improve the efficiency of hand gesture recognition and enhance the performance of hand gesture recognition. The recognition accuracy of character level hand gesture (American Sign Language and Arabic numerals) is 98.23%, which is 5% higher than other algorithms on average; the recognition accuracy of word level gesture (Chinese Sign Language) is 97.81%, which is 4% higher than other algorithm on average.

Keywords Hand gesture recognition, Convolutional neural network, Self-adaptive fusion algorithm, Multi-classifiers, Data glove

1 引言

随着人体感应设备的普及和以人为中心的感知计算的快速发展,研究者开始将研究重心转移到基于可穿戴设备的手

势识别,并将其广泛应用于虚拟现实、健康检测和智能家居等领域。相对于基于视觉的手势识别方法^[1],基于可穿戴传感器的手势识别方法^[2]具有以下优点:不受背景环境干扰,数据形式单一,数据处理以及特征提取便捷。因此,本文以基于可

到稿日期:2020-01-13 返修日期:2020-05-17 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(71774159, 61876186, 61977061);中国博士后科学基金(2018M642358);绿色安全管理与政策科学智库(2018WHCC03)

This work was supported by the National Natural Science Foundation of China (71774159, 61876186, 61977061), China Postdoctoral Science Foundation(2018M642358) and Think Tank of Green Safety Management and Policy Science (2018WHCC03).

通信作者:袁冠(yuanguan@cumt.edu.cn)

穿戴传感器的手势数据为研究对象,分析手势数据的特点,解决手势识别问题。

传统的手势识别分类器能够对一些简单手势进行识别,比如贝叶斯模型^[3]、决策树^[4]、支持向量机^[5]和隐马尔可夫模型^[6]。文献^[7]利用手掌伸缩带动前臂伸缩肌变化的特性,通过表面肌电传感器采集前臂伸缩肌数据,并使用核函数将数据映射到高维特征空间,结合支持向量机完成对简单手势(比如握拳、展拳等手势)的有效识别。但是,这些分类器需要手动提取特征,并依赖于研究人员的专业领域经验。当手势数据量过大且种类过多时,识别效果易受人为因素干扰,导致识别精度降低。

相对于传统的手势识别分类器,基于神经网络的分类器具有强大的自动特征表示能力,能够解决手动提取手势数据特征带来的问题。常用的分类器有自编码器^[8]、卷积神经网络^[9]、循环神经网络(Recurrent Neural Network, RNN)^[10]及长短期记忆单元(Long Short-Term Memory, LSTM)^[11]。文献^[12]利用 CNN 参数稀疏连接和权值共享的特性提取手势数据的深度特征,将其作为 LSTM 的输入完成对手势的识别,既提高了手势识别的效率,又解决了动态手势数据的时序性问题。但是,研究者们往往忽略了单分类器在识别过程中对某些手势具有偏向性的问题,比如 LSTM 可以正确识别“惊讶”的手势,而自编码器往往会将“惊讶”手势误识别为“吃饭”手势。因此,本文结合多分类器之间的互补信息,使用多分类器融合的方法弥补单分类器在识别过程中出现的偏向性错误,并通过自适应融合算法解决经典融合算法(如多数投票

法^[13])人为干预的问题^[14]。

本文基于可穿戴设备(数据手套),提出了基于自适应多分类器融合的手势识别方法(SAMCF)。主要工作如下:

(1)SAMCF 使用 CNN 自动提取具有强表征能力的深度特征,既解决了手动提取法依赖于人类专业领域知识的局限性,又提高了特征提取的效率;

(2)SAMCF 使用多分类器融合的方法对手势数据进行识别,实现了不同分类器信息的互补,解决了单分类器在识别过程中的偏向性问题;

(3)SAMCF 采用自适应融合算法来融合多个基本分类器,在自动更新各基本分类器权重的过程中,自动抑制识别准确率较低的基本分类器的权重,并提高识别准确率较高的基本分类器的权重,避免了手动更新权重的人为干扰,提高了手势识别的准确率。

2 手势识别模型

基于自适应多分类器融合手势识别方法的研究框架如图 1 所示,主要由以下 3 部分组成:手势数据采集、数据预处理以及手势识别。使用多传感器融合的数据手套采集手势数据,并通过蓝牙或者 USB 串口将其传输到上位机,完成对手势数据的采集。数据预处理包括数据的平滑去噪、数据分割以及数据的归一化处理等步骤,其目的是得到标准的、干净的、连续的手势数据。手势识别包含手势特征的提取、识别模型训练以及手势的识别,即使用 CNN 提取手势数据的特征向量,并结合多分类器对手势数据进行有效识别。

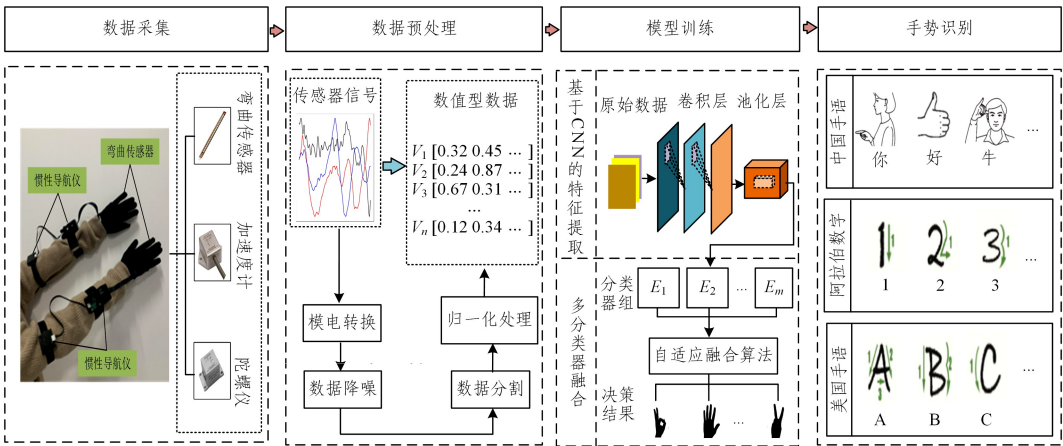


图 1 研究框架

Fig. 1 Research framework

2.1 数据采集

为捕捉动态手势的完整运动轨迹,需要使用多传感器设备(比如手机、数据手套等)同步获取上肢部分的全部姿态信息,因此本文使用数据手套采集手势数据。人体上肢传感器的嵌入如图 2 所示。在数据采集过程中,弯曲传感器采集手指的弯曲程度,磁力陀螺仪采集手部的欧拉角,惯性导航传感器采集前臂和后臂的姿态信息和运动状态信息,以实现对上肢动作数据的全面采集。其中,惯性导航仪集成了磁力陀螺仪、加速度计以及一个可扩展的数字运动处理器。惯性导航仪还免除了组合陀螺仪与加速度器时轴间差的问题,减少了大量的封装空间,可以直接输出四元数和欧拉角格式的数据。

表 1 列出了嵌入在数据手套中各传感器的各项指标^[15]。

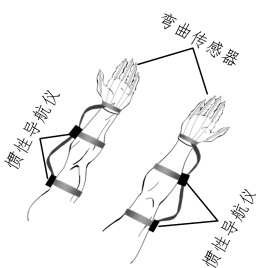


图 2 人体上肢传感器配置图

Fig. 2 Sensor configuration diagram

表 1 数据手套内置传感器的描述

Table 1 Description of data glove’s built-in sensor

传感器	描述	备注
加速度计	(1)加速度计传感器是在加速过程中通过测量物体所受的惯性力来获得加速度值,包含重力加速度 (2)用于分别测量 3 个物理轴(x,y,z)上的加速度的大小,加速度的单位是 m/s^2	(1)不同加速度计之间的差异主要是最大取样频率 (2)加速度计可以测量出运动速度和运动方向,但不能判断空间的南北朝向 (3)加速度计是人体活动识别中应用最广的传感器之一
陀螺仪	(1)陀螺仪通过测量物体运动时的角速度来获取物体旋转的角度和方向 (2)陀螺仪用于测量物体运动时 3 个物理轴(x,y,z)上的角速度大小,单位是 rad/s	(1)角速度就是物体绕自身的 3 个轴的角度变化率,静止时三轴角速度都为 0 (2)角速度可以用于手势识别和游戏中精准的方向控制
弯曲传感器	(1)弯曲传感器通过阻值将弯曲程度数字化 (2)角度量程为 $[0,180]$	(1)用于采集物体的弯曲程度 (2)弯曲程度越大,阻值越高

2.2 数据预处理

基于可穿戴传感器的手势数据容易受噪声、缺失值以及不一致数据的侵扰,直接使用原始数据将影响实验的准确性,从而对实验的结论分析造成误差。为保证数据的准确性、完整性以及一致性,需要对手势数据进行预处理。数据预处理包含 3 个步骤:数据降噪、数据分割以及归一化处理。

2.2.1 数据降噪

由于传感器本身精度的问题,即使在静止状态下,数据手套仍能采集到细微变化的数据。因此,需要通过噪声滤波器对数据手套采集的原始手势数据进行预处理。同时,针对手势数据低幅噪声的影响,本文采用均值滤波器^[16]对手势数据进行滤波处理。图 3 为“惊讶”手势的小臂滚转角(x 轴)滤波前后的对比图。

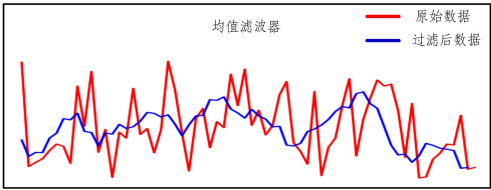


图 3 均值滤波器
Fig. 3 Mean filter

2.2.2 数据分割

数据手套对手势数据的实时采集,意味着手势数据被持续不断地发送到上位机。动态手势数据(比如惊讶、早安、再见等手势)是连续不断的,但手势识别的特征向量是在某一种特定动作之下提取的特征值。使用滑动窗口策略将滤波后的信号数据进行分窗处理,常用的滑动窗口策略有固定时间窗分割、重叠时间窗分割以及固定采样数分割。由于固定时间窗分割和固定采样数分割会忽略窗口边缘的手势,因此本文采用重叠时间窗进行手势数据采样,并将滑动窗口的长度设定为 10,窗口每次向前滑动时,每两个相邻窗口之间有 50% 的重叠^[17],如图 4 所示。

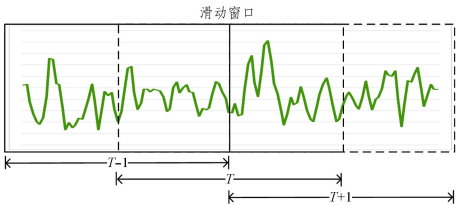


图 4 滑动窗口
Fig. 4 Sliding window

假设 $T-1$ 窗口中的读数是 1—20,则 T 窗口中读数为 11—30, $T+1$ 窗口的读数为 21—40。除第一个窗口外,每个窗口中都包含上一个窗口的后半部分。

2.2.3 归一化处理

在实际应用中,不同传感器的取值范围存在很大的差异,且在手势采集过程中,多次多人采集的同一手势的幅值和速度也会有所不同。为消除幅值和速度的差异性对识别效果的影响,本文使用滑动窗口分割从数据手套捕捉的数据流的同时,在每个窗口中对数据进行归一化处理。

首先,标记每个窗口中每个轴的最值信息(包含最大值和最小值)。 A_i, H, A_i, L 和 A_i 分别表示滑动窗口中加速度计各轴的最大值、最小值和采样值($i=x,y,z$)。 G_i, H, G_i, L 和 G_i 分别表示滑动窗口中陀螺仪各轴的最大值、最小值和采样值($i=x,y,z$)。然后,使用式(1)和式(2)分别对加速度计和陀螺仪的数值进行归一化,得到 a_i 和 g_i 。

$$a_i = \frac{A_i - A_i, L}{A_i, H - A_i, L}, i = x, y, z$$
 (1)

$$g_i = \frac{G_i - G_i, L}{G_i, H - G_i, L}, i = x, y, z$$
 (2)

其中, a_i 表示第 i 轴(x,y,z)的加速度计值, g_i 表示第 i 轴(x,y,z)的陀螺仪值。

数据手套中用于采集手指弯曲程度的弯曲传感器是单极传感器,其特点是电阻值与弯曲传感器的弯曲程度呈线性关系,当且仅当贴合手指正常弯曲方向时有效。因此,本文通过比较采样弯曲阻值与伸直状态和完全弯曲状态的阻值来映射角度,其计算公式如式(3)所示。

$$angle = (1 - \frac{R_c - R_s}{R_b - R_s}) * 180$$
 (3)

其中, $angle$ 表示弯曲传感器的弯曲程度; R_c, R_b 和 R_s 表示采样时刻的电阻值、完全弯曲状态的电阻值和伸直状态下的电阻值,并规定伸直状态下的角度为 180° ,完整弯曲状态下的角度为 0° 。 $flex = 1 - \frac{R_c - R_s}{R_b - R_s}$ 为弯曲传感器的归一化数值。

2.3 识别模型

基于自适应多分类器融合的手势识别模型由特征工程和手势识别两部分组成,如图 5 所示。特征工程的目的是获取具有强表征能力的特征向量,以提高手势识别的准确率。手势识别主要使用自适应融合算法决策出基本分类器的最优结果,以完成对手势的有效识别。

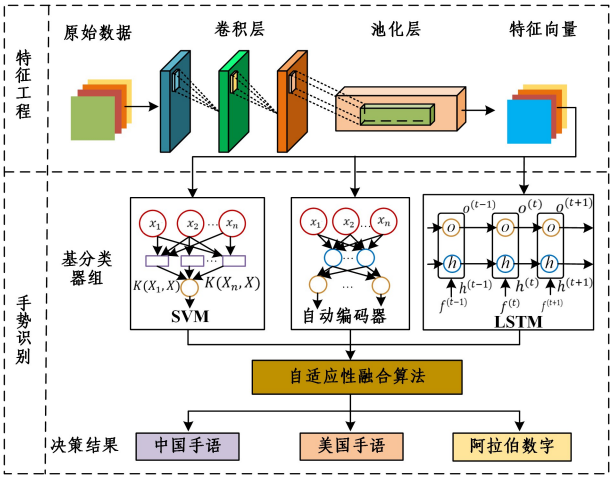


图5 基于自适应多分类器融合的手势识别模型

Fig. 5 Hand gesture recognition model based on self-adaptive multi-classifiers fusion

2.3.1 基于CNN的特征提取

由于CNN具有稀疏连接、参数共享的特点^[18],且在提取手势数据特征向量过程中可以降低手势数据的冗余性并提高提取特征向量的效率,因此SAMCF使用CNN提取手势数据的特征向量。

(1)卷积层

卷积层通过卷积核的大小控制局部感受野的范围,以达到稀疏连接的目的。同时,卷积运算通过参数共享,减少了网络自由参数的数量。参数共享的机制使得神经网络的特征映射具有位移不变性,图6为卷积运算示意图。假设手势数据的输入向量为 $\mathbf{X}=(x_1, x_2, \dots, x_n)$,其大小为 8×8 。过滤器由权重向量 $\mathbf{W}=(w_1, w_2, \dots, w_n)$ 和偏置参数向量 $\mathbf{B}=(b_1, b_2, \dots, b_n)$ 组成,卷积核 K 的大小为 2×2 ,则经过4层卷积之后得到的第 j 个特征向量为 $\mathbf{z}_j$,其大小为 $4 \times 4 \times 128$,计算公式为:

$$\mathbf{z}_j = \sum_{k=1}^n ReLu(x_k * \mathbf{w}_j + \mathbf{b}_j) \tag{4}$$

其中, $ReLu()$ 为激活函数,用于优化神经网络的非线性能力。

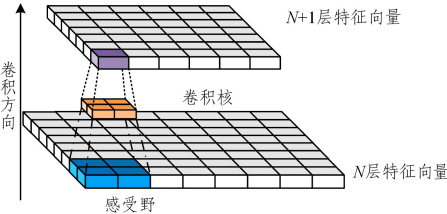


图6 卷积运算

Fig. 6 Convolution operation

(2)池化层

池化层的作用是对输入的特征向量进行压缩,一方面简化网络的计算复杂度,另一方面提取手势数据的主要特征,在一定区域内将极有可能具有相似性的特征进行聚类统计,从而实现对特征向量的压缩。常用的池化方式有:均值池化、最大值池化和随机池化,其计算过程如图7所示。SAMCF采用max pooling作为池化函数,经过池化后得到的特征向量为 $\mathbf{M}=(m_1, m_2, \dots, m_n)$,其大小为 $2 \times 2 \times 64$,其计算公式为:

$$\mathbf{M}_i = \max_{j=1}^k (\mathbf{z}_i), i=1, 2, \dots, n \tag{5}$$

其中, k 表示池化区域的大小。

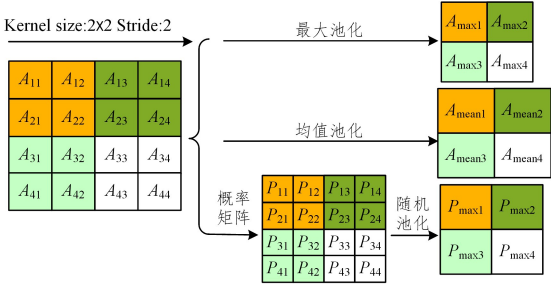


图7 3种池化方式的计算过程

Fig. 7 Calculation process of three pooling methods

2.3.2 基本分类器的模型训练

SAMCF将通过CNN提取的深度特征向量作为各基本分类器的输入,结合各基本分类器的特点得到手势识别结果,并使用自适应融合算法决策出最优结果,以提高手势识别的准确率。多分类器融合能够减小单分类器因偏向性导致的泛化能力不佳的风险,降低陷入局部收敛的可能性,扩大了假设空间以保障模型的鲁棒性。损失函数能够度量模型分类结果的好坏。模型训练的过程就是从模型的假设空间中选取最优模型的过程,即采用合适的算法使损失函数全局收敛。

(1)基于统计模型的分类器的模型训练

支持向量机(Support Vector Machine, SVM)能够有效地识别非线性、高维度的模式识别问题,其主要思想是通过核技巧将输入空间映射到高维度的特征空间,寻找间隔最大化的决策曲面以对手势的分类。已知 $(m_1, y_1), \dots, (m_i, y_i), \dots, (m_N, y_N)$ 为通过CNN提取的特征,其中 $y_i \in \{1, 2, \dots, N_c\}$, $\forall i \in \{1, 2, \dots, N\}$, N_c 为手势种类,则SVM的最优化问题表示为:

$$\min L_{svm} = \frac{1}{N} \sum_{i=1}^N \sum_{j \neq y_i} [\max(0, \mathbf{w}_c^T m_i + b_c - \mathbf{w}_{y_i}^T m_i - b_{y_i} + 1)] + \lambda \sum_{c=1}^{N_c} \sum_{i=1}^N \mathbf{w}_c^T \mathbf{w}_{y_i}^T \tag{6}$$

其中, $\mathbf{w}_c^T$ 表示预测错误类别的权重, $\mathbf{w}_{y_i}^T$ 表示真实类别的权重, b_c 和 b_{y_i} 表示偏置值, λ 为惩罚系数。

梯度矩阵 $\mathbf{d}_w$ 由各样本 $m_i (i=1, 2, \dots, N)$ 在损失函数 L_{svm} 各个分类方向上的梯度构成,其指明了各样本在各个分类方向上的下降速度,计算公式如式(7)所示:

$$\nabla_{\mathbf{w}} L_{svm} = \frac{\partial}{\partial \mathbf{w}} L_{svm} = \begin{cases} \frac{\partial}{\partial \mathbf{w}_c} L_{svm} (j \neq y_i) \\ \frac{\partial}{\partial \mathbf{w}_{y_i}} L_{svm} (j = y_i) \end{cases} \tag{7}$$

结合式(6)和式(7)可以得到梯度的计算公式:

$$\nabla_{\mathbf{w}_{y_i}} L_{svm} = -(\sum_{c \neq y_i} 1(\mathbf{w}_c^T m_i + b_c - \mathbf{w}_{y_i}^T m_i - b_{y_i} + 1 > 0)) m_i \tag{8}$$

$$\nabla_{\mathbf{w}_c} L_{svm} = 1(\mathbf{w}_c^T m_i + b_c - \mathbf{w}_{y_i}^T m_i - b_{y_i} + 1 > 0) m_i \tag{9}$$

其中, $1(x)$ 为示性函数,当 x 为真时取值为1,当 x 为假时取值为0。

(2)基于神经网络的分类器的模型训练

自编码器(AutoEncoder, AE)具有良好的特征提取、数据

降维以及信息补全功能,其神经元之间全连接的方式为深层网络确定了相对合理的网络初始化参数,促使自编码器能够有效地避免梯度消失和梯度爆炸问题。LSTM 通过输入门、遗忘门以及输出门选择性地丢弃和保留手势数据的信息量,能够解决手势数据的时序性和长距离依赖的问题。自编码器和 LSTM 属于基于神经网络的模型,通常使用 BP 算法对交叉熵损失函数进行训练。已知通过 CNN 提取的特征数据为 $(m_1, y_1), \dots, (m_i, y_i), \dots, (m_N, y_N)$, 在这些数据中, $y_i \in \{1, 2, \dots, N_c\}$, $\forall i \in \{1, 2, \dots, N\}$, N_c 为手势种类, 则自编码器和 LSTM 的最优化问题表示为:

$$\min L_{NN} = \sum_{i=1}^N [y_i - f_{\theta}(x_i)]^2 + \lambda \theta^T \theta \tag{10}$$

其中, $f_{\theta}(x)$ 为神经网络模型, θ 为模型的参数, λ 为惩罚系数。BP 算法是基于梯度下降的策略, 依目标函数的负梯度方向对参数进行更新, 使损失函数收敛于全局极小点。根据链式法则, 误差函数 L_{NN} 的第 k 层隐藏层的梯度如式(11)所示:

$$\Delta w_i^k = \frac{\partial L_{NN}}{\partial w_i^k} = \frac{\partial L_{NN}}{\partial y_i^n} * \frac{\partial y_i^n}{\partial y_i^{n-1}} * \dots * \frac{\partial y_i^{k+1}}{\partial w_i^k} \tag{11}$$

其中, w_i^k 表示第 k 层第 i 个权值, y_i^j 表示第 j 层第 i 个输出值。给定学习率 η , 各个神经单元的权重更新公式如下:

$$w_i^{k+1} = w_i^k - \eta \Delta w_i^k \tag{12}$$

2.3.3 自适应融合算法

为解决单分类器识别的偏向性问题, SAMCF 采用多分类器融合的识别方法, 但是常见的融合算法通常采用手动赋值基本分类器的加权系数, 极易受人为因素干扰, 影响最后的识别精度。因此, SAMCF 使用自适应融合算法既实现了多分类器的融合, 又能够自动更新基本分类器的加权系数, 最终决策出最优识别结果, 完成对手势数据的有效识别。自适应融合算法的具体描述如算法 1 所示。

算法 1 自适应融合算法

输入: 手势数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 其中 $x_i \in \mathcal{X} \subseteq \mathcal{R}^n, y_i = \{y_1, y_2, \dots, y_c\}$, C 为手势种类
输出: 最优识别结果 $y_j = \{y_1, y_2, \dots, y_N\}$

1. 初始化权值分布 $D_1 = (w_{11}, w_{12}, \dots, w_{1n}), w_{1i} = \frac{1}{n}, i = 1, 2, \dots, N$ 。
2. 使用 CNN 提取手势数据集的特征向量 $\mathbf{M} = (m_1, m_2, \dots, m_n)$ 。
3. 使用特征向量 $\mathbf{M}$ 学习各个基本分类器模型, $G_i^j(x): \mathcal{X} \rightarrow \{y_1, y_2, \dots, y_n\}$ 。
4. 计算 $G_i^j(x)$ 在手势数据集 T 上的分类误差率 $e_i = \frac{\sum_{i=1}^N P(G_i(x_i) \neq y_i)}{\sum_{i=1}^N w_{ti}} * I(G_i(x_i) \neq y_i)$ 。
5. 计算 $G_i^j(x)$ 的系数 $\alpha_i = \frac{1}{2} \log \frac{1 - e_i}{e_i}$, 这里的对数是自然对数 e 。
6. 对 $t = 1, 2, \dots, T //$ 迭代次数

6.1. 构造具有权值分布的基本分类器的线性组合 $f(x) = \sum_{j=1}^3 w_{ij} * G_i^j(x)$, 得到最终分类器 $G(x) = \text{softmax}(f(x)) = \text{softmax}(\sum_{j=1}^3 w_{ij} * G_i^j(x))$;

6.2. 得到决策结果 $y_j = \arg \max(G(x))$, 并输出识别准确率;

6.3. 更新权值分布 $D_{t+1} = (w_{t+1,1}, w_{t+1,2}, w_{t+1,3}), w_{t+1,i} = \frac{w_{ti}}{Z_t} \exp(-\alpha_i * y_i * G_i(x_i)), i = 1, 2, \dots, N$, 其中 $Z_t = \sum_{i=1}^N W_{ti} * \exp(-\alpha_i *$
- $y_i * G_i(x_i))$ 。

7. 输出最优决策结果 $y_j = \{y_1, y_2, \dots, y_N\}$ 以及最优识别准确率。
- 3 实验结果与分析
- 3.1 实验环境
- 本文使用数据手套(含有大小臂导航仪)采集完整的手势数据, 手势数据采集平台由 C# 编写而成, 主要识别模型使用 Python 编写, 实验运行环境为微软 Windows10(64 位)操作系统, 使用的开发软件包含 PyCharm, Anaconda, Visual studio 2017。具体的实验环境如表 2 所列。
- 表 2 实验环境
- Table 2 Experimental environment

编程语言	Python, C#	操作系统	Windows10(64 位)
软件环境	PyCharm, Anaconda, Visual studio 2017		
数据手套	加速度计、陀螺仪、弯曲传感器		
硬件环境	Intel(R) Core(TM) i5-8300 CPU @2.3 GHz, 运行内存 16 GB		
- 3.2 数据集
- 基于可穿戴传感器的手势数据集由 6 名志愿者(3 名男性和 3 名女性)完成, 他们按照活动的协议完成数据的采集。所有的实验参与者在实验过程中佩戴数据手套, 以 20Hz 的固定频率, 利用嵌入在数据手套中的弯曲传感器、陀螺仪以及加速度计采集手指的弯曲信息、手掌以及大小臂的姿态信息, 并手动对数据进行标记。得到的数据集被随机分为两组, 其中 70% 的数据用于生成训练数据, 30% 用于测试数据。
- (1) 字符级数据集(美国手语数据集和阿拉伯数字数据集)。为测试字符级手势的识别准确性, 实验按照文献[19]的动作协议, 选取并采集标准的美国手语数据集的 6 个字母(A—F)和 6 个阿拉伯数字(1—6)。该数据集的具体描述如表 3 所列。
- 表 3 字符手势数据集
- Table 3 Character hand gesture dataset

样本总数	训练集	测试集	标注方式	实验设备	传感器	采样频率	手势种类数
5547	3882	1665	后期手动标注	数据手套	加速度计、陀螺仪、弯曲传感器	20 Hz	12
- (2) 单词级数据集(中国手语数据集)。为测试动态复杂手势的识别准确性, 实验按照文献[20]的动作协议, 采集了基于中国手语的日常手势数据集。该数据集主要包含复合动态手势(比如“你好”复合手势由“你”和“好”两个基本手势组成)和基本动态手势(比如再见、惊讶等手势)。本文实验从中国手语数据集中选取了 10 个手势进行分析和研究。该数据集的具体描述如表 4 所列。
- 表 4 中国手语数据集
- Table 4 Chinese sign language(CSL) dataset

样本总数	训练集	测试集	标注方式	实验设备	传感器	采样频率	手势种类数
34452	24116	10336	后期手动标注	数据手套	加速度计、陀螺仪、弯曲传感器	20 Hz	72

3.3 实验分析

3.3.1 对比算法

为了验证 SAMCF 能够自动提取具有强表征能力的深度特征,将 SAMCF 与基于统计模型的识别方法 (Statistical Model-based Recognition Method, SMRM) 进行对比,其中包含 K 最近邻 (K-Nearest Neighbor, KNN)^[21] 和 SVM^[5]。此外,为了验证 SAMCF 能够解决单分类器偏向性问题,实现多分类器的信息互补,增强手势识别的性能,将 SAMCF 与基于神经网络的识别方法 (Neural Network-based Recognition Method, NNRM) 和深度混合模型 (Deep Hybrid Model, DHM) 进行对比。基于神经网络的手势识别方法包含 CNN^[9] 和 LSTM^[22];深度混合模型包含 CNN-LSTM^[12] 和 CNN-SVM^[23]。基于统计模型的识别方法、基于神经网络的识别方法以及深度混合模型的具体描述如表 5 所列。

表 5 对比算法		
Table 5 Comparison algorithms		
算法分类	算法名称	特点
SMRM	KNN	利用周围有限的邻近样本对类域进行判定
	SVM	利用内积核函数解决非线性分类问题
NNRM	CNN	能够提取手势数据的深度特征,增强手势识别的性能
	LSTM	解决动态手势数据的长距离依赖问题和时序性问题
DHM	CNN-LSTM	使用 CNN 提取手势特征向量,LSTM 解决动态手势的长短期依赖问题
	CNN-SVM	使用 CNN 提取手势数据深度特征,结合 SVM 完成对手势的识别
本文方法	SAMCF	使用 CNN 提取手势数据的深度特征,结合自适应融合算法决策出多个基本分类器的最优手势识别结果

3.3.2 评价指标

为了评估 SAMCF 在手势识别上的性能,本文采用准确率、精确率、召回率以及 F 值作为评价标准。

定义 1(准确率) 所有被预测正确的样本数占样本总数的比值。其计算公式如下:

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN}$$
 (13)

其中,TP 表示真实样本被正确预测为正的样本数,TN 表示真实样本被正确预测为负的样本数,FP 表示真实样本被错误预测为正的样本数,FN 表示真实样本被错误预测为负的样本数。

定义 2(精确率) 真实样本种类被正确预测的样本数占所有被预测为正的样本数的比值。其计算公式如下:

$$Precision = \frac{TP}{TP + FP}$$
 (14)

定义 3(召回率) 真实样本种类被正确预测为正的样本数占所有被预测为真的样本数的比值。其计算公式如下:

$$Recall = \frac{TP}{TP + FN}$$
 (15)

定义 4(F 值) 综合考虑精确率和召回率的指标,是具有相同权重的精度和召回的加权平均值。其计算公式如下:

$$F-Score = \frac{2 * P * R}{P + R}$$
 (16)

其中,P 表示精确率,R 表示召回率。

3.3.3 参数分析

(1)卷积层数分析

本文使用 CNN 提取手势数据的深度特征,不同层次的特征图从输入数据中提取不同层次的信息。浅层特征包含更多的原始数据信息,但是语义歧义的问题更为突出;深层特征具有高语义性,能够有效地表征原始数据。因此,提取特征向量的卷积层深度影响着手势识别的性能。图 8 给出了基于不同卷积层深度的手势识别精度。可以看出,卷积层越深,手势识别的准确率就越高;但是当卷积层的深度增加到 4 层时,准确率逐渐趋于平稳,变化不明显,甚至出现下降。下降的原因在于提取的深度特征维度太低,无法全面表征原始手势数据。因此,本文使用深度为 4 的卷积神经网络提取手势的特征向量。

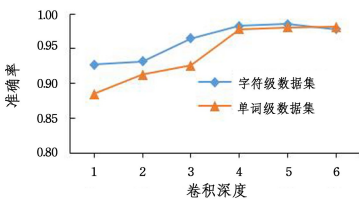


图 8 基于不同卷积层深度的识别准确率

Fig. 8 Recognition accuracy with different convolution depth

(2)基本分类器的选取

基本分类器的选择是多分类器融合的基础。为了获取更好的手势识别效果,本文将以下 7 种常用的基本分类器作为备选基本分类器:KNN、朴素贝叶斯 (Bayes)、SVM、决策树 (Decision Tree,DT)、AE、RNN 和 LSTM。通过实验数据集来验证各个基本分类器的有效性,选取识别准确率最高的 3 个基本分类器,实验结果如图 9 和图 10 所示。从图中可以选择需要融合的 3 个基本分类器:SVM,AE 和 LSTM。

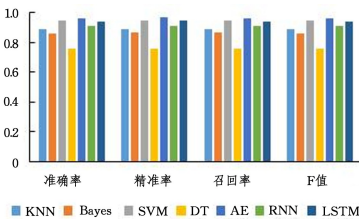


图 9 基于字符级数据集的各分类器的识别效果

Fig. 9 Recognition performance of each classifier on character level dataset

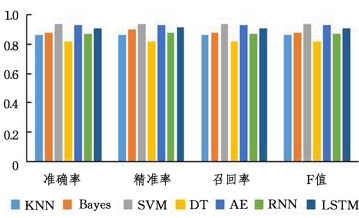


图 10 基于单词级数据集的各分类器的识别效果

Fig. 10 Recognition performance of each classifier on word level dataset

(3)特征选取

为了验证深度特征对手势识别效果的增益,本文将基于

深度特征的多分类器融合模型与基于统计特征的多分类器融合模型进行比较。常见的统计特征有均值特征、最值特征、范围特征以及方差特征,具体描述如表 6 所列^[15]。

表 6 常见的统计特征

Table 6 Common statistical characteristics

统计特征	描述	计算公式
均值特征	一个窗口中信号的平均值	$mean = \frac{1}{n} \sum_{i=1}^n a_i$
最值特征	一个窗口中信号的最大值和最小值	$max = \max(a_i), i \in \{1, 2, \dots, n\}$ $min = \min(a_i), i \in \{1, 2, \dots, n\}$
范围特征	一个窗口中信号的最大值和最小值的差值	$range = max - min $
方差特征	一个窗口中一组数据的分散程度的度量	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - mean)^2$

常见的统计特征能够有效地度量不同手势的间距。均值特征能够减小类内间距;最值特征能够体现出不同手势之间的幅值特征;范围特征能够体现手势之间变化的幅度范围;方差特征能够表征手势之间的离散程度。统计特征需要手动提取原始数据的特征,虽然保留了原始数据细节性信息,但是无法提取原始数据的高语义特征,导致其识别效果远差于深度特征的高语义信息的识别效果。如图 11 和图 12 所示,基于深度特征的识别准确率高于基于统计特征的识别准确率,因此深度特征能够更好地表征手势数据。

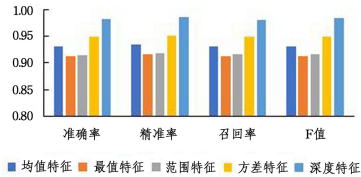


图 11 基于不同特征的字符级手势的识别效果

Fig. 11 Character level hand gesture recognition with different features

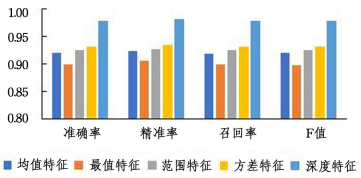


图 12 基于不同特征的单词级手势的识别效果

Fig. 12 Word level hand gesture recognition with different features

3.3.4 自适应多分类器融合算法

字符级手势识别主要针对美国手语和阿拉伯数字的识别,其在数据采集过程中侧重于大小臂的姿态信息。从图 13 可以看出,基于统计模型的识别算法的精度普遍低于基于神经网络的识别模型的精度,主要原因在于深度特征的表征能力强于统计特征。此外,基于神经网络的手势识别模型中,CNN 识别准确率大于 LSTM,说明字符级手势的时序性不强,进一步验证了 CNN 特征提取的能力强于 LSTM。深度混合模型的识别准确率普遍高于基于统计模型的识别方法和基于神经网络的识别方法,说明模型之间能够实现信息互补,从而提高识别准确率。总体来说,SAMCF 能够有效地识别字符级手势,其识别准确率达到 98.23%,精准率达到

98.57%,召回率达到 98.12%,F 值达到 98.32%,识别性能较其他算法平均提升了 5%左右。

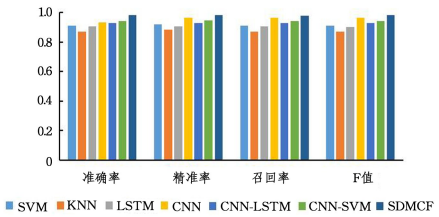


图 13 字符级手势的识别效果

Fig. 13 Recognition effect of character level hand gesture

单词级手势识别主要针对中国手语的识别,在数据采集过程中,既关注手掌的姿态信息以及手指的弯曲程度,又关注大小臂的姿态信息。从图 14 可以看出,各算法的识别精度略低于字符级手势识别精度,主要原因在于单词级手势数据相对于字符级数据集来说,更具时序性,手势之间的干扰性强。基于神经网络的识别方法中的 LSTM 识别准确率高于 CNN,说明 LSTM 能够很好地识别具有时序性的手势数据。总体来说,SAMCF 能够有效地识别单词级手势,充分提取手势数据的深度特征,实现多分类器之间的信息互补,其识别准确率达到 97.81%,精准率达到 98.21%,召回率达到 97.83%,F 值达到 97.9%,识别性能较其他算法平均提升了 4%左右。

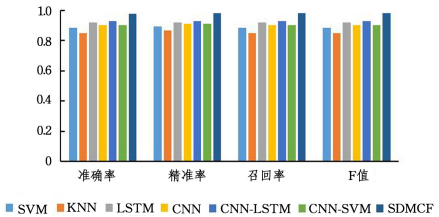


图 14 单词级手势的识别效果

Fig. 14 Recognition effect of word level hand gesture

结束语 本文采用一种新型的可穿戴数据手套来采集手势数据,同时为了消除单分类器的偏向性问题,提出了自适应多分类器融合的方法;使用深度混合模型(SAMCF)分别对字符级手势(美国手语和阿拉伯数字)和单词级手势(中国手语)进行识别。此外,为了获取强表征性的手势运动特征,SAMCF 使用多层卷积的方式提取手势数据的高语义深度特征,并通过自适应融合算法融合多分类器,增强手势识别的性能。实验结果表明,本文算法的性能优于其他同类算法,对字符级手势识别准确率达到 98.23%,对单词级手势识别准确率达到 97.81%。

日常生活中手势数据复杂多样,大致可分为字符级(美国手语、阿拉伯数字)、单词级(中国手语)以及短语级(日常对话短语)。本文只是对字符级、单词级进行识别,没有对短语级数据进行识别;同时,在识别字符级和单词级手势时,忽略了情景环境的干扰,因为同一手势在不同情景环境下具有歧义性。因此,未来研究手势识别方法时,首先应该融合手势的情景环境特征,增强手势识别的性能;然后针对日常交流的需要,实现对短语级手势的识别;最后,将虚拟现实技术和识别模型相结合,实现对手势的在线识别。

参 考 文 献

- [1] JIANG D,ZHENG Z J,LI G F,et al. Gesture recognition based on binocular vision[J]. Cluster Computing,2019,22(6):13261-13271.
- [2] ZHANG X,YANG Z,CHEN T,et al. Cooperative sensing and wearable computing for sequential hand gesture recognition[J]. IEEE Sensors,2019,(19):5775-5783.
- [3] KUMAR P,ROY P P,DOGRA D P. Independent Bayesian classifier combination based sign language recognition using facial expression[J]. Information Sciences,2018,428:30-48.
- [4] CHANG G C,PARK J,OH C,et al. A decision tree based real-time hand gesture recognition method using kinect[J]. Journal of Korea Multimedia Society,2013,16(12):1393-1402.
- [5] SHIFALI S,SHATRUGHAN M,PRASHANT S R,et al. Hand gesture recognition using gaussian threshold and different SVM kernels[C]// Advances in Computing and Data Sciences. 2018:138-147.
- [6] SAGAYAM K M,HEMANTH D J. A probabilistic model for state sequence analysis in hidden markov model for hand gesture recognition[J]. Computational Intelligence,2019,35(1):59-81.
- [7] TAVAKOLI M,BENUSSI C,ALHAIS L P,et al. Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier[J]. Biomedical Signal Processing and Control,2018:121-130.
- [8] ZHANG Z,TIAN Z,ZHANG Y,et al. U-DeepHand:FMCW Radar-Based Unsupervised Hand Gesture Feature Learning Using Deep Convolutional Auto-Encoder Network[J]. IEEE Sensors Journal,2019,19(16):6811-6821.
- [9] LI G F,TANGH,SUN Y,et al. Hand gesture recognition based on convolution neural network[J]. Cluster Computing,2019,22(2):2719-2729.
- [10] KRZYSZTOF C,JACEK R,ALICJA K. Gesture recognition with the linear optical sensor and recurrent neural networks[J]. IEEE Sensors Journal,2018,18(13):5429-5438.
- [11] MA C,WANG A,CHEN G,et al. Hand joints-based gesture recognition for noisy dataset using nested interval unscented Kalman filter with LSTM network[J]. Visual Computer,2018,34(6/7/8):1053-1063.
- [12] NUNEZ J C,CABIDO R,PANTRIGO J J,et al. Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition[J]. Pattern Recognition,2018,76:80-94.
- [13] PAUDYAL P,LEE J,BANERJEE A. DyFAV:Dynamic feature selection and voting for real-time recognition of fingerspelled alphabet using wearables[C]// Proceedings of The 22nd International Conference on Intelligent User Interfaces. 2017:457-467.
- [14] KITTLER J,HATEF M,DUIN R P W,et al. On combining classifiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1998,20(3):226-239.
- [15] YUAN G,WANG Z H,MENG F R,et al. An overview of human activity recognition based on smartphone[J]. Sensor Review,2019,39(2):288-306.
- [16] WANG X,LI B Z. A novel median filtering forensics based on principal component analysis network[J]. International Journal of Electronic Security and Digital Forensics,2019,11(2):145-159.
- [17] DERNBACH S,DAS B,KRISHNAN N C,et al. Simple and Complex Activity Recognition through Smart Phones[C]// Intelligent Environments. 2012:214-221.
- [18] SAINATH T N,VINYALS O,SENIOR A,et al. Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks[C]// 2015 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). 2015:4580-4584.
- [19] LI C,XIE C,ZHANG B C. Deep Fisher discriminant learning for mobile hand gesture recognition[J]. Pattern Recognition,2018,77:276-288.
- [20] ZHANG X,CHEN X,LI Y. A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors[J]. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans,2011,41(6):1064-1076.
- [21] WAHID M F,TAFRESHI R,AL-SOWAIDI M,et al. Subject-independent hand gesture recognition using normalization and machine learning algorithms[J]. Journal of Computational Science,2018,27:69-76.
- [22] FRANCISCO O,DANIEL R. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity Recognition[J]. Sensors,2016,16(1):115.
- [23] MA M,CHEN Z,WU J. A Recognition Method of Hand Gesture with CNN-SVM Model[C]// International Conference on Bio-Inspired Computing: Theories and Applications. Singapore: Springer,2016:399-404.



LIU Xiao, born in 1994, postgraduate. His main research interests include data mining, pattern recognition and perceptual computing.



YUAN GUAN, born in 1982, Ph.D, professor, is a member of China Computer Federation. His main research interests include artificial intelligence, big data technology, machine learning and computational intelligence.