·多媒体技术及应用·　　文章编号：1000-3428(2018)09-0243-07　　文献标志码：A　　中图分类号：TP393

基于改进型多维卷积神经网络的微动手势识别方法

李玲霞，王　羽，吴金君，王沙沙

（重庆邮电大学 移动通信技术重庆市重点实验室，重庆 400065）

摘　要：传统二维卷积神经网络因遗漏时间维度信息导致不能识别微动手势。为此，提出一种基于视频流的微动手势识别方法。对输入视频流进行简单预处理，利用改进型多维卷积神经网络提取手势的时空特征，融合多传感器信息并通过支持向量机实现微动手势识别。实验结果表明，该方法对手势的背景和光照都具有较好的鲁棒性，且针对各类动态手势数据集能达到 87% 以上的识别准确率。

关键词：计算机视觉；手势识别；二维卷积神经网络；多维卷积神经网络；支持向量机；鲁棒性

中文引用格式：李玲霞，王　羽，吴金君，等. 基于改进型多维卷积神经网络的微动手势识别方法[J]. 计算机工程，2018，44(9)：243-249.

英文引用格式：LI Lingxia，WANG Yu，WU Jinjun，et al. Micro-motion hand gesture recognition method based on improved multiple dimensional convolution neural network[J]. Computer Engineering，2018，44(9)：243-249.

Micro-motion Hand Gesture Recognition Method Based on Improved Multiple Dimensional Convolution Neural Network

LI Lingxia，WANG Yu，WU Jinjun，WANG Shasha

（Chongqing Key Lab of Mobile Communications Technology，
Chongqing University of Posts and Telecommunications，Chongqing 400065，China）

【Abstract】For the traditional Two Dimensional Convolutional Neural Network (2D-CNN)，the time dimension information is lost，and thus the dynamic gesture cannot be recognized. This paper proposes a novel dynamic hand gesture recognition method based on video streams，which can effectively improve the overall performance of hand gesture recognition. The input data is simply preprocessed. The spatio temporal feature extraction operation is performed by using improved Multiple Dimensional Convolutional Neural Network (MD-CNN). A multi-sensor fusion method is provided and the dynamic gesture recognition is realized by using Support Vector Machine (SVM). Experimental results show that the proposed method performs well in robustness with respect to the gesture background and illumination. Furthermore，the method achieves the high recognition accuracy beyond 87% for every kind of dynamic gesture dataset.

【Key words】computer vision；hand gesture recognition；Two Dimensional Convolutional Neural Network (2D-CNN)；Multiple Dimensional Convolutional Neural Network (MD-CNN)；Support Vector Machine (SVM)；robustness

DOI：10.19678/j.issn.1000-3428.0048138

0　概述

计算机的普及使得人机交互方式得以迅速发展。手势作为一种最直接、最方便的人机交互方式受到了越来越多的关注并在各种现实场景中发挥了重要作用，如体感游戏、辅助汽车控制系统、手语识别和个人可穿戴系统等。手势在时空的不确定性、多样性和相似性以及照明条件等的影响都是手势识别过程中需要考虑的重要方面。

目前，比较典型的手势识别方法主要有隐马尔可夫模型(Hidden Markov Model，HMM)[1]、模板匹配[2]和神经网络[3]等。其中，HMM 的手势识别方法能够有效利用手势信号的时序信息来解决微动手势识别问题，但是该模型需要计算大量的状态概率密度，计算量大且识别速度缓慢，不能很好地满足当前应用的需要。基于模板匹配的手势识别方法将手势的轮廓和边缘信息等几何特性作为特征建立手势模板，并通过各种模板匹配算法实现手势识别，具有

基金项目：重庆市基础与前沿研究计划项目(cstc2013jcyjA40032)；重庆邮电大学博士启动基金(A2012-33)；重庆邮电大学青年科学研究项目(A2013-31)。

作者简介：李玲霞(1976—)，女，副教授，主研方向为手势识别、深度学习、宽度无线接入技术；王　羽、吴金君、王沙沙，硕士研究生。

收稿日期：2017-07-27　　**修回日期：**2017-09-11　**E-mail：**lilx@cqupt.edu.cn

较强的稳定性和较高的识别准确率,但是需要根据大量的经验人工构造手势特征,并且这些人工构造的特征在描述手势特性时具有一定的主观性和局限性,使得该方法的学习能力有限且效率不高。基于神经网络的手势识别方法通过神经网络提取手势的拓扑信息作为手势特征,再利用各类分类器对手势进行识别,该方法具有普适性,学习能力强,但是对于微动手势和形态差别不大的手势分类效果不够理想,主要应用于静态手势识别。

卷积神经网络是一种典型的特征提取算法,与深度神经网络相比,该网络特有的权值共享、局部连接和下采样方法[4]能够减少网络训练参数,使得网络结构简单,计算量呈指数级降低,并且能大幅降低网络的过拟合风险。近年来,卷积神经网络已成功应用于图像检索[5]、表情识别[6]、行人检测[7]、人体行为检测[8]和手势识别[9-10]中。传统的卷积神经网络利用二维卷积核提取手势特征[9,11],仅对单张图片进行处理,忽略了图片间的时间相关性,遗漏了时间维信息,从而只能识别静态手势。针对这一问题,本文提出一种新的基于视频流的微动手势识别方法。

1 数据采集及预处理

1.1 数据采集

系统采用 Kinect 2.0 采集手势,以 30 frame/s 的速度采集手势数据得到深度和彩色视频并通过 LK 光流法[12]获得手势光流视频,然后建立相应的手势数据库。在本文中,微动手势定义为以手指关节为单位的运动手势,所建立的微动手势数据库包含 10 类动作,分别由 10 个手势用户在多种不同背景和光照条件下采集的手势视频组成,整个手势数据库包含 3 000 个视频。手势动作示意图如图 1 所示。

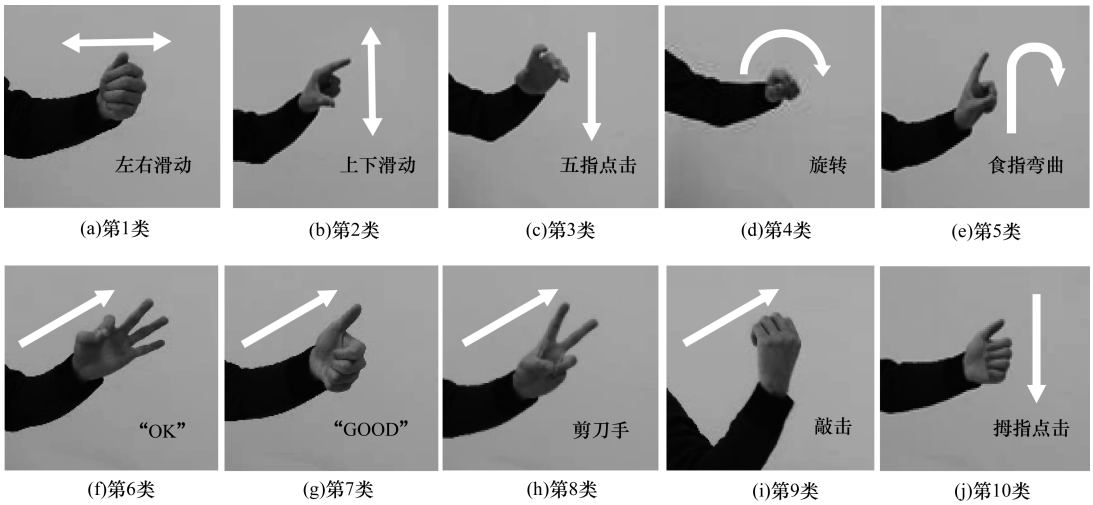

图 1 手势动作示意图

1.2 数据预处理

数据预处理模块主要分为 2 个部分,一是数据增强模块,二是视频流打包模块。数据增强对训练一个好的网络模型十分重要,可以给单幅图片增加多个副本,提高图片的利用率,并防止样本太少而出现的过拟合现象。在执行数据增强之前,首先对手势视频进行提帧操作,将视频表示为多个时间相关的图片序列,然后通过随机裁剪和水平翻转实现数据增强,为原始图片增加多个经过变形的手势副本。本文将原始图片大小随机裁剪为 112 ×112,并用三通道(红绿蓝)对输入的彩色图像进行描述,则输入 MD-CNN 的视频流大小可以表示为 3 ×112 ×112 × K,其中,K 表示每个视频的长度。根据手势的持续时间和设备处理要求,将每个手势剪辑成多个视频块,每个视频块包含 16 帧图片,使得在训练过程中以 16 帧图片组成的视频块为单位进行卷积和池化操作,则最终输入 MD-CNN 的视频块大小可以表示为 3 ×112 ×112 ×16。最后,随机选择 70% 的样本作为训练集并将其送入 MD-CNN 进行训练,而剩下的 30% 作为测试集。

2 手势识别系统

整个手势识别系统的架构如图 2 所示,主要包括数据采集及预处理模块、网络训练模块和手势分类模块。其中,网络训练模块和手势分类模块是关键环节。网络训练模块可以学习到合适的微动手势特征,避免人工设计手势特征的复杂过程;手势分类模块可以建立特征与标签的特殊映射关系,从而实现精准的微动手势识别。

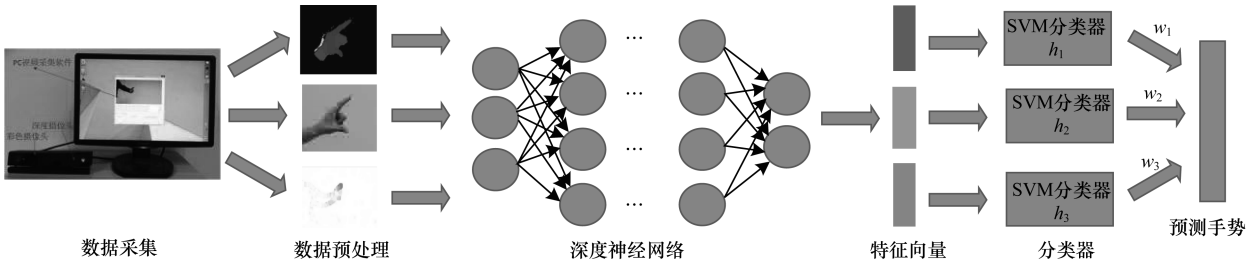

图 2　微动手势识别系统

2.1　网络训练

2.1.1　网络结构

本文采用 MD-CNN 提取手势特征,其网络结构如图 3 所示。整个网络包含 5 个卷积层、5 个池化层和 3 个全连接层。5 个卷积层的滤波器数目从第 1 层到第 5 层分别为 64、128、256、256 和 256。卷积层和池化层交替出现,全连接层紧跟在池化层之后,每个全连接层分别包含 4 096、4 096 和 10 个神经元。由文献[13],将所有多维卷积核的大小设为 3×3×3,步长设为 1×1×1。除第 1 个池化层外,其余所有多维池化层的大小为 2×2×2,步长为 2×2×2。此外,为了保留时序信息,在第1 个池化层中,将卷积核大小设为 2×2×1,步长也设为 2×2×1。与此同时,设输入视频流大小为 $a\times b\times c$,卷积核大小为 $u\times v\times n$,池化层大小为 $p\times q\times r$,则经过卷积层之后的特征图大小为 $(a-u+1)\times(b-v+1)\times(c-n+1)$,再经过池化层之后的特征图大小为 $((a-u+1)/p)\times((b-v+1)/q)\times((c-n+1)/r)$,由此方法可计算出每层特征图的大小。

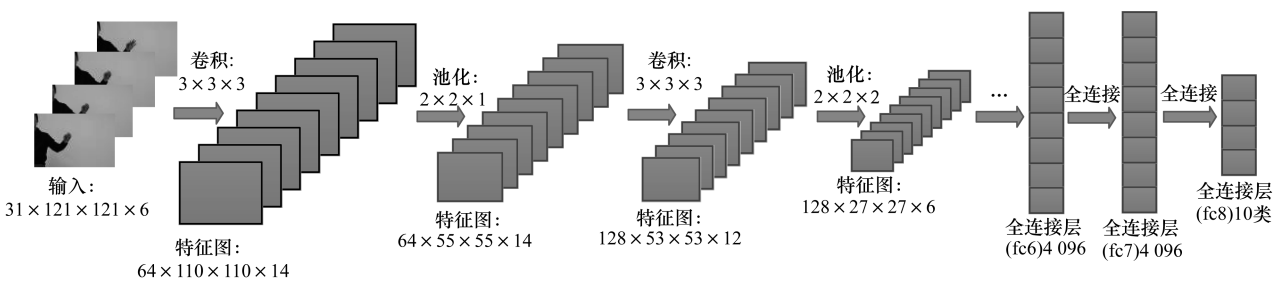

图 3　MD-CNN 网络结构

2.1.2　前向传播

网络的前向传播[10]过程即为卷积层、池化层和全连接层之间的逐层计算过程。在本文中,令 x、y 分别表示该像素在当前层的空间位置,同理 z 表示时间维坐标,此三维坐标唯一对应视频流中的一个像素。式(1)表示第 i 个卷积层在位置(x,y,z)的三维卷积输出。

$$c_i(x,y,z)=\varphi\Big(\sum_{\substack{o\in\{1,2,\cdots,u_i\}\\ l\in\{1,2,\cdots,v_i\}\\ m\in\{1,2,\cdots,n_i\}}} w_{i,o,l,m}I(x+o,y+l,z+m)+b_i\Big) \tag{1}$$

其中,I 表示输入视频块,u_i、v_i 和 n_i 分别表示第 i 个卷积核的长、宽和高,激活函数为 $\varphi(\cdot)=ReLU(\cdot)$,$w_i$ 表示第 i 个卷积核的权重,b_i 为第 i 个卷积核的偏置。

此外,式(2)表示第 j 个池化层在位置(x,y,z)的计算结果。

$$s_j(x,y,z)=\max_{\substack{o\in\{1,2,\cdots,p_j\}\\ l\in\{1,2,\cdots,q_j\}\\ m\in\{1,2,\cdots,r_j\}}}(c_j(xp_j+o,yq_j+l,zr_j+m)) \tag{2}$$

其中,p_j、q_j 和 r_j 分别表示池化核的长、宽和高。在本文中选择最大池化法,则输出 s_j 表示第 j 个池化区域的最大值。在池化层没有可训练的权重和偏置,从而既可以简化卷积层的输出,又可降低网络出现过拟合的概率。

最后,式(3)表示第 n 个全连接层的计算结果。

$$g(n)=\sum_{o,l,m} w_{n,o,l,m}s_j(o,l,m)+b_n \tag{3}$$

其中,w_n 和 b_n 分别表示相应的权重和偏置。

在前向传播阶段从训练样本中选取第 k 个样本,记为 X_k,令 F_n 表示第 n 层的激活函数,则 X_k 从输入层经逐级变换传送到输出层的结果可表示为式(4)。

$$h(X_k)=F_n(\cdots(F_2(F_1(X_k,w_1,b_1),w_2,b_2)\cdots),w_n,b_n) \tag{4}$$

2.1.3　参数更新

原始的网络使用传统的随机梯度下降算法[14]更新网络参数,但是该算法的收敛过程十分缓慢,对内存要求高且容易陷入局部最小值。因此,本文采用自适应矩估计方法[15]求解目标函数的最小值。该算法根据损失函数对每个参数梯度的一阶和二阶

矩估计动态地调整每个参数的学习率。将目标函数定义为交叉熵损失函数[16]。同时,为了避免网络出现过拟合,引入 L2 范数[17]对损失函数进行正则化处理,则最终的损失函数可表示为式(5)。

$$J(\theta)=-\frac{1}{M}\left[\sum_{k=1}^{M}l\{Y_k=h(X_k)\}\log_a\frac{e^{\theta^T x(k)}}{\sum_{k=1}^{N}\theta^T x(k)}\right]+\lambda R(\theta) \tag{5}$$

其中,M 为训练样本个数,Y_k 为第 k 个训练样本所对应的真实手势标签($Y_k\in\{1,2,\cdots,10\}$),$h(X_k)$ 为卷积神经网络最后一层的输出值,函数 $l\{Y_k=h(X_k)\}$ 表示当 $Y_k=h(X_k)$ 时,值为 1,否则为 0,θ 表示所有的网络参数,λ 表示正则化参数,$R(\theta)$ 表示对 θ 求 L2 范数,即 $R(\theta)=\|\theta\|_2^2=\sum_{k=1}^{M}|\theta_k^2|$。

该算法通过最小化代价函数 $J(\theta)$ 更新网络参数,其具体求解过程如下:

$$g_t=\nabla_\theta J_t(\theta) \tag{6}$$

$$P_t=\mu_1 P_{t-1}+(1-\mu_1)g_t \tag{7}$$

$$Q_t=\mu_2 Q_{t-1}+(1-\mu_2)g_t^2 \tag{8}$$

其中,t 表示时刻,g_t 表示目标函数在 t 时刻的梯度,P、Q 分别表示对梯度的一阶矩估计和二阶矩估计,μ_1 和 μ_2 为优化模型的参数,表示指数衰减速率,根据文献[15],将其分别设置为经验值 0.9 和 0.999。

接着,根据式(9)和式(10)校正一阶和二阶矩估计,P_t' 和 Q_t' 分别表示校正后的结果:

$$P'_t=\frac{P_t}{1-\mu_1} \tag{9}$$

$$Q'_t=\frac{Q_t}{1-\mu_2} \tag{10}$$

最后,根据规则更新参数:

$$\theta_t=\theta_{t-1}-\frac{P'_t}{\sqrt{Q'_t}+\varepsilon}\times\eta \tag{11}$$

其中,η 设置为经验值 0.001,ε 为经验值 10^{-8}。

本文所采用的自适应矩估计优化方法的伪代码如下。

算法 自适应矩估计算法

输入 目标函数 $J(\theta)$、学习率 η、初始网络参数 θ_0、停止时间 T

输出 最终参数 θ_t

1. 初始化参数:$P_0=0$,$Q_0=0$,$t=0$,$\mu_1=0.9$,$\mu_2=0.999$
2. for $t\leqslant T$
3. $t=t+1$
4. 计算 t 时刻目标函数的梯度 $g_t=\nabla_\theta J_t(\theta)$
5. 更新有偏差的第 1 矩估计 $P_t=\mu_1 P_{t-1}+(1-\mu_1)g_t$
6. 更新有偏差的第 1 矩估计 $Q_t=\mu_2 Q_{t-1}+(1-\mu_2)g_t^2$
7. 计算偏差矫正后的第 1 矩估计 $P'_t=\frac{P_t}{1-\mu_1}$
8. 计算偏差矫正后的第 2 矩估计 $Q'_t=\frac{Q_t}{1-\mu_2}$
9. 更新网络参数 $\theta_t=\theta_{t-1}-\frac{P'_t}{\sqrt{Q'_t}+\varepsilon}\times\eta$
10. end

2.2 手势分类

对于 MD-CNN 来说,前面的卷积层和池化层学习到的是图像的浅层特征,如图像的轮廓、角点等,具体特征如图 4 所示。越靠后的层级结构学习到的是越高级的图像特征,如图像的细节、线条组合等,具体特征如图 5 所示。其中,全连接层所学习的特征最高级、表达能力和泛化能力最强,因此,在本文中提取全连接层 fc6 的手势特征用于后续识别。

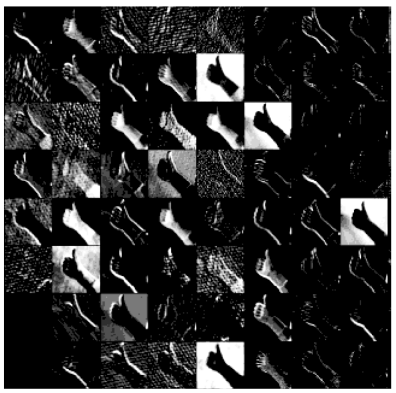

图 4 第一个卷积层的特征

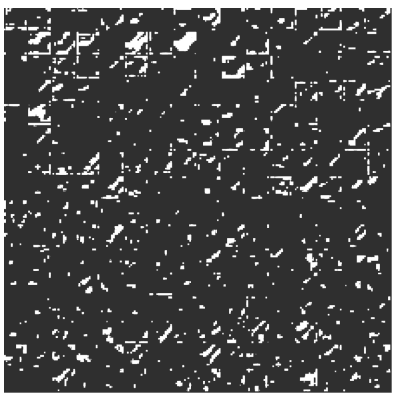

图 5 最后一个卷积层的特征

提取出手势特征后,将其输入合适的分类器建立标签与特征的映射关系,考虑到 SVM 本身具有较强的泛化能力和高效的分类能力[9,18],选择 SVM 进行手势识别。在本文中,将不同的传感器信息输入 MD-CNN 进行训练,可以提取出不同的特征向量,将 3 类传感器对应的特征向量输入 SVM 建立不同的分类器,分别记为 h_1、h_2 和 h_3。与此同时通过交叉验证法[19]计算各个分类器的准确度,并进行归一化处理得到各个分类器的权重,分别记为 w_1、w_2 和 w_3,将预测手势的最终得分记为 s,且 $s=\sum_{i=1}^{3}w_i h_i$,选择得分最高的预测标签作为最终的预测手势类别。

3　实验设置与结果分析

3.1　实验环境

实验地点为典型室内环境,背景包含办公室的白墙背景、书柜背景和门背景;光照设置为正常光照、微弱光照和强光照射,实验场景如图 6 所示。在实验中,手势识别服务器运行在 Linux 操作系统上,硬件配置为:Intel-6700K 处理器,NVIDIA-GTX1080 显卡,显存为 8 GB。

(a)不同背景

(b)不同光照

图 6　实验场景

3.2　实验结果与分析

为了评估系统的性能,利用实际采集的手势数据集和公开的 SKIG[20] 数据集进行多次实验并比较不同输入数据类型对系统性能的影响。此外,通过对比文献[6]、文献[10]、文献[21]和文献[22]所用方法进行手势识别的准确率和处理时间,验证了本文方法的准确性和有效性。

3.2.1　系统实验结果分析

表 2 给出了利用本文方法对微动手势进行分类所得到的混淆矩阵。图 7 ~ 图 9 分别给出了深度、彩色和光流数据的损失曲线,该曲线描述了损失函数与迭代次数的对应关系。图 10 给出了不同数据类型的学习曲线,该曲线描述了准确率与迭代次数的对应关系。由混淆矩阵可知,本文方法对于微动手势具有较高的识别准确率。尤其是第 3 类手势,该类手势在测试数据集上准确率可达 100% 。此外,由于第 2 类手势与第 5 类手势、第 7 类手势与第 10 类手势均具有一定程度的相似性,因此存在一定程度误判的情形,准确率在 90% 以下。由损失函数曲线可知,当迭代次数小于 6 000 时,训练与测试损失均呈下降趋势,即网络处于欠拟合状态,而当迭代次数大于 6 000 时,准确率几乎不变,即网络陷入过拟合状态。根据损失函数曲线的变化规律,设置迭代次数为 6 000 以尽可能避免训练过程中出现的过拟合或欠拟合现象。同时,由学习曲线可知,当网络趋于稳定后,深度数据集和光流数据集的性能相差不大,且略优于彩色数据集。

表 2　微动手势混淆矩阵

真实手势	预测手势									
	1	2	3	4	5	6	7	8	9	10
1	93.4	0	0	3.3	0	0	3.3	0	0	0
2	0	83.4	0	0	13.3	0	0	3.3	0	0
3	0	0	100	0	0	0	0	0	0	0
4	0	3.3	0	96.7	0	0	0	0	0	0
5	13.3	0	0	0	86.7	0	0	0	0	0
6	0	0	0	0	3.3	93.4	0	0	3.3	0
7	0	0	0	0	0	0	83.4	0	0	16.6
8	3.3	0	0	0	0	0	0	96.7	0	0
9	0	0	0	9.9	0	0	0	0	90.1	0
10	6.6	0	0	3.3	0	0	9.9	0	0	81.2

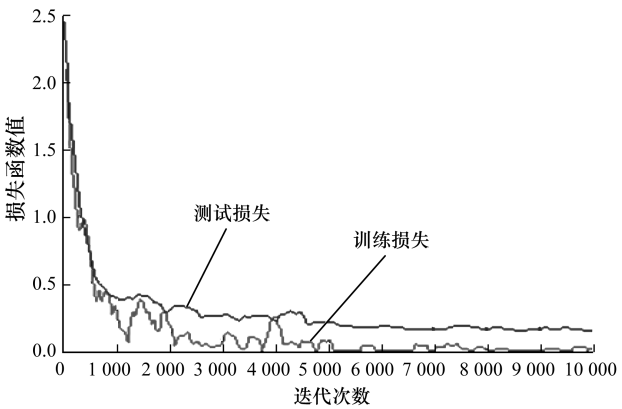

图 7　深度数据损失函数曲线

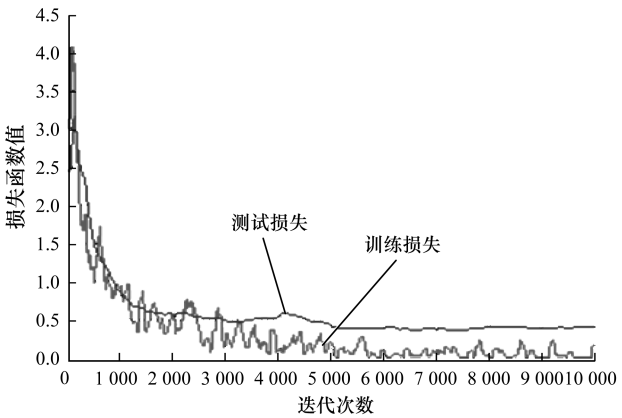

图 8　彩色数据损失函数曲线

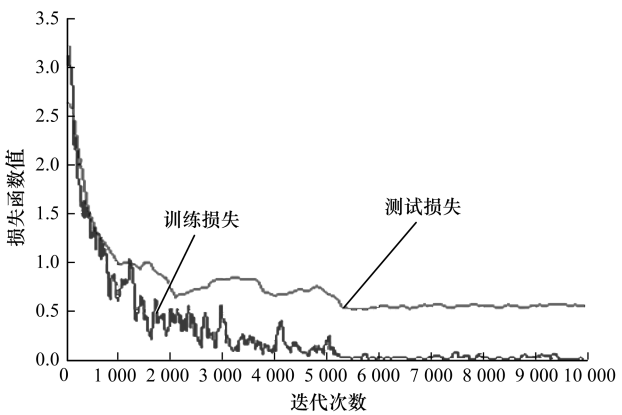

图 9　光流数据损失函数曲线

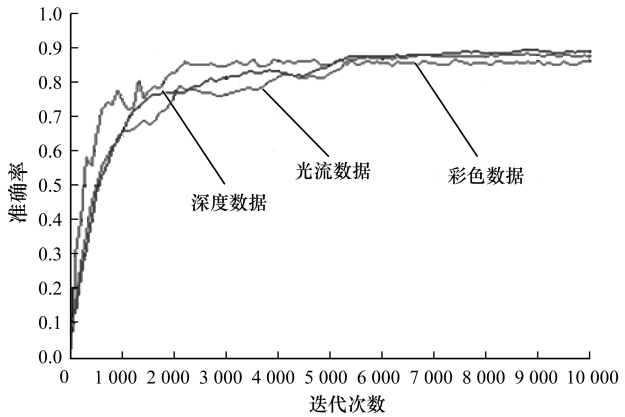

图 10 不同数据类型的学习曲线

3.2.2 对比分析

图 11 给出了不同数据类型的准确率。由图 11 可知,无论是对于 SKIG 数据集还是微动手势数据集,经多传感器融合的手势识别准确率比单一传感器的识别准确率要高出 3% 左右。图 12 ~ 图 14 分别给出了在不同优化方法下深度、彩色和光流数据的损失曲线。由图 12 ~ 图 14 可知,与传统的梯度下降算法相比,本文方法的收敛速度更快,并且损失函数的波动更小,由此说明本文方法能够以更快的速度找到更优的结果。表 3 给出了本文方法与其他方法关于手势识别准确率和速率的对比。与本文方法不同,文献[6]使用传统的卷积神经网络和 SVM 识别微动动作,文献[10]基于神经网络与 softmax 分类器完成微动手势识别,文献[21]提出利用融合的空间域网络和时间域网络实现手势识别的方法,文献[22]基于深度神经网络实现动态手势识别。可以看出,本文方法的识别速率及识别准确率相比于其他方法都有显著提升,尤其是与文献[21]中的融合网络相比,识别准确率提高了 10% 左右。综上所述,本文方法相比于传统方法,大幅提高了每个动作的识别速率和准确率,对手势识别的整体性能有显著改善。

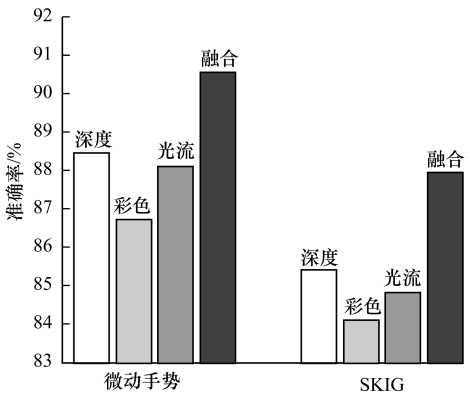

图 11 不同数据类型对准确率的影响

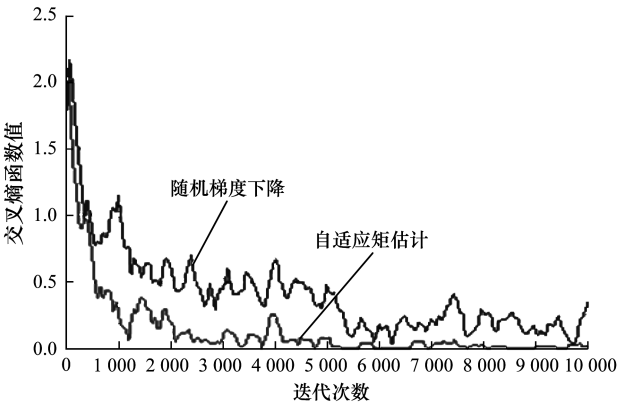

图 12 深度数据损失函数曲线

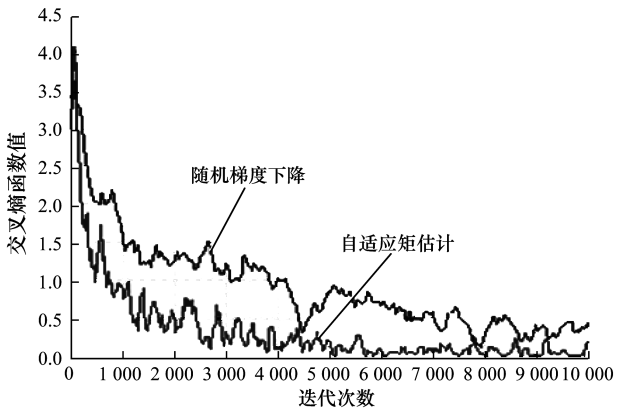

图 13 彩色数据损失函数曲线

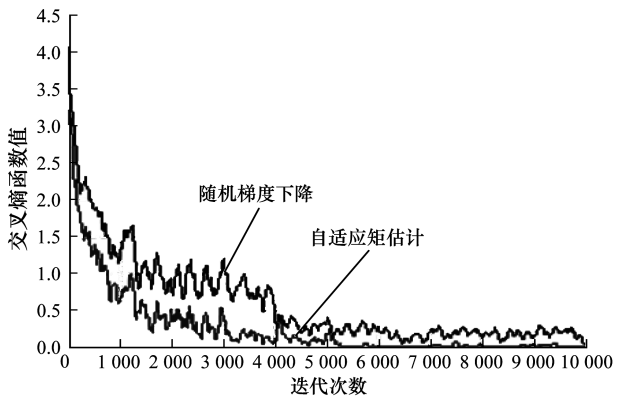

图 14 光流数据损失函数曲线

表 3 各方法性能对比

方法	平均识别时间/s	平均准确率/%
改进型 MD-CNN + 组合分类器	2.11	90.5
传统 MD-CNN + SVM	2.89	89.1
传统 MD-CNN + Softmax	2.47	88.2
时空域网络 + Softmax	1.01	80.4
深度神经网络	3.24	89.8

4 结束语

本文提出了一种基于改进型 MD-CNN 的微动手势识别方法。该方法首先对原始视频流进行提帧

和数据增强处理;然后利用深度神经网络提取手势的时空特征,避免了根据手势的轮廓和几何特性人为设计特征的复杂过程;最后通过组合分类器实现微动手势识别。实验结果表明,本文方法不仅能够有效识别在多光照、复杂背景下的微动手势,而且具有较高的准确性、鲁棒性和普适性。下一步将优化算法,提高微动手势识别的准确率和实时性。

参考文献

[1] JIANG Yongsen. An HMM based approach for video action recognition using motion trajectories [C]// Proceedings of IEEE International Conference on Intelligent Control and Information Processing. Washington D. C. ,USA:IEEE Press,2010:359-364.

[2] REYES M, DOMÍNGUEZ G, ESCALERA S. Feature weighting in dynamic time warping for gesture recognition in depth data [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ,USA:IEEE Press,2011:1182-1188.

[3] VALENCIA C R,GARCIA-BERMEJO J G,CASANOVA E Z. Combined gesture-speech recognition and synthesis using neural networks [J]. IFAC Proceedings Volumes, 2008, 41(2): 2968-2973.

[4] 陈祖雪. 基于深度卷积神经网络的手势识别研究[D]. 西安:陕西师范大学,2016.

[5] 柯圣财,赵永威,李弼程,等. 基于卷积神经网络和监督核哈希的图像检索方法[J]. 电子学报,2017,45(1):157-163.

[6] FAN Yin, LU Xiangju, LI Dian, et al. Video-based emotion recognition using CNN-RNN and C3D hybrid networks [C]//Proceedings of the 18th ACM International Conference on Multimodal Interaction. New York,USA:ACM Press,2016:445-450.

[7] 左艳丽,马志强,左宪禹. 基于改进卷积神经网络的人体检测研究[J]. 现代电子技术,2017,40(4):12-15.

[8] IJJINA E P, CHALAVADI K M. Human action recognition using genetic algorithms and convolutional neural networks[J]. Pattern Recognition,2016,59(1): 199-212.

[9] NAGI J, DUCATELLE F, DI CAROG A, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition [C]//Proceedings of IEEE International Conference on Signal and Image Processing Applications. Washington D. C., USA: IEEE Press, 2011:342-347.

[10] DUFFNER S, BERLEMONT S, LEFEBVREG, et al. 3D gesture classification with convolutional neural networks[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Washington D. C. ,USA:IEEE Press,2014:5432-5436.

[11] KONDA K R,KÖNIGS A,SCHULZ H,et al. Real time interaction with mobile robots using hand gestures[C]// Proceedings of International Conference on Human-robot Interaction. New York, USA: ACM Press, 2012: 177-178.

[12] BAKER S,SCHARSTEIN D,LEWISJ P,et al. A database and evaluation methodology for optical flow [J]. International Journal of Computer Vision, 2011, 92(1): 1-31.

[13] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatio temporal features with 3D convolutional networks[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA:IEEE Press,2015:4489-4497.

[14] OYAMA Y, NOMURA A, SATO I, et al. Predicting statistics of asynchronous SGD parameters for a large-scale distributed deep learning system on GPU supercomputers[C]//Proceedings of IEEE International Conference on Big Data. Washington D. C. ,USA:IEEE Press,2016:66-75.

[15] KINGMA D P, BA J L. Adam: a method for stochastic optimization[C]//Proceedings of International Conference on Learning Representations. Berlin, Germany: Springer, 2015:1-13.

[16] SONG Yu, WU Yiquan, DAI Yimian, et al. A new active contour remote sensing river image segmentation algorithm inspired from the cross entropy [J]. Digital Signal Processing,2016,48(3):322-332.

[17] MARCELO V W, ZIBETTI D R, PIPA A R, et al. Fast and exact unidimensional L2-L1 optimization as an accelerator for iterative reconstruction algorithms [J]. Digital Signal Processing,2016,48(3):178-187.

[18] SAVARIS A, WANGENHEIM A. Comparative evaluation of static gesture recognition techniques based on nearest neighbor, neural networks and support vector machines[J]. Journal of Brazilian Computer Society, 2010,16(2):147-162.

[19] 高　林,盛子豪,刘　英. 基于交叉验证支持向量机算法的交通状态判别研究[J]. 青岛科技大学学报(自然科学版),2017,38(1):105-108.

[20] LIU Li, SHAO Ling. Learning discriminative representations from RGB-D video data [C]// Proceedings of International Joint Conference on Artificial Intelligence. Berlin, Germany: Springer, 2013: 1493-1500.

[21] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Proceedings of International Conference on Neural Information Processing Systems. New York,USA:ACM Press,2014:568-576.

[22] 卓少伟,柳培忠,黄德天,等. 基于 CW-RNNs 网络的手势识别算法[J]. 海峡科学,2016(7):51-56.

编辑　顾逸斐