

杨子恢, 何宏. 基于骨架掩膜图像的动态手势识别算法[J]. 智能计算机与应用, 2024, 14(4): 226–231. DOI: 10.20169/j.issn.2095-2163.240437

基于骨架掩膜图像的动态手势识别算法

杨子恢, 何宏

(上海理工大学 健康科学与工程学院, 上海 200093)

摘要: 动态手势识别是人机交互领域的重要问题, 同时也是虚拟现实、远程医疗、手语识别等领域的关键技术。针对动态手势识别中不同人存在手形差异以及同类动作速度不一致的问题, 本文提出一种基于骨架掩膜图像的动态手势识别算法。首先对源视频的每一帧手势图像进行骨架提取以及阈值化处理得到手部骨架掩膜图像。接着利用卷积神经网络对掩膜后的骨架图像提取静态空间特征, 并输入门控循环单元进行时间上的动态空间特征提取。最后, 在全连接层中实现手势分类。该方法有效解决了传统动态手势识别中的空间关系复杂性和时间可变性问题。实验结果证明, 本文提出的方法在 RWTH 数据集上取得了 95.52% 的准确率, 具有良好的识别性能和应用前景。

关键词: 动态手势识别; 深度学习; 掩膜图像; 时空特征

中图分类号: TP181

文献标志码: A

文章编号: 2095-2163(2024)04-0226-06

Dynamic gesture recognition algorithm based on skeleton mask images

YANG Zihui, HE Hong

(School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Dynamic gesture recognition is a critical issue in the field of human-computer interaction, as well as in domains such as virtual reality, remote healthcare, and sign language recognition. Addressing the challenges posed by variations in hand shapes among different individuals and inconsistencies in the speed of similar actions in dynamic gesture recognition, this paper proposes a dynamic gesture recognition algorithm based on skeletal masked images. Firstly, skeleton extraction and threshold processing are conducted on each frame of the source video to obtain hand skeleton mask images. Subsequently, a Convolutional Neural Network (CNN) is utilized to extract static spatial features from the masked skeleton images, followed by input to Gated Recurrent Units (GRU) for dynamic spatial feature extraction over time. Finally, gesture classification is performed using fully connected layers. This approach effectively resolves spatial complexity and temporal variability issues commonly encountered in traditional dynamic gesture recognition. Experimental results demonstrate that the proposed algorithm achieves 95.52% accuracy on RWTH data set, showcasing promising recognition performance and practical applications.

Key words: dynamic gesture recognition; deep learning; mask image; spatial-temporal features

0 引言

动态手势识别是计算机视觉领域中一项具有重要意义的研究方向。通过分析人体手部姿态和轨迹, 可以在虚拟现实、远程医疗、手语识别等领域中实现自然直观的人机交互, 具有广泛的应用前景。然而, 动态手势识别因其空间关系复杂性和时间可

变性的局限给识别模型提出了挑战。空间关系复杂性涉及手部的形状和姿势多样性等实际问题。在动态手势识别中, 由于不同用户的手形差异较大, 例如手指的长度、手掌的宽度以及手部的整体形状等因素都会产生变化。这些差异使得手势的形状和姿态在三维空间中呈现出多样性, 增加了手势识别的复杂度。而时间可变性指的是不同用户在执行同一类

基金项目: 国家科技部资助项目(G2021013008); 上海市科学技术委员会资助项目(18070503000); 上海市教育委员会 2021 年上海高校大学计算机课程教学改革项目; 上海理工大学医工交叉重点项目(1020308405, 1022308502)。

作者简介: 杨子恢(1999-), 男, 硕士研究生, 主要研究方向: 行为识别与人机交互系统研发。

通讯作者: 何宏(1973-), 女, 博士, 教授, 博士生导师, 主要研究方向: 医学人工智能, 医疗大数据挖掘与人机交互系统研发。Email: hehong

@usst.edu.cn

收稿日期: 2023-01-26

手势时,由于运动速度的不同,手势所耗费的时间也会有所不同。这导致了手势的时间轨迹在不同实例之间存在差异,增加了识别任务的难度。另一方面,复杂的背景干扰、光照环境的变化以及其他客观噪声因素都会影响手势识别的准确性。仅考虑降低客观环境影响的预处理会不同程度地损失识别精度,降低准确率和交互性。因此开发一种能够消除环境对识别的干扰、融合手部时空数据,同时兼具识别精度的手势识别模型,具有十分重要的现实意义。

1 相关工作

动态手势识别研究涵盖了多种方法和技术,根据数据的获取方式可以划分为基于传感器设备和基于视觉处理两种方法。例如,文献[1]利用内嵌于数据手套中的传感器,将手指的弯曲姿态与手势模板进行对比量化,实现用户手势识别。文献[2]采用贝叶斯线性分类器设计了基于加速度计和表面肌电信号(SEMG)的小尺度手势识别框架。文献[3]开发了一种可穿戴的低成本系统并使用支持向量机(SVM)对手臂肌肉中阻抗值的特征进行手势分类。虽然传感器采集的参数具有较高的精度,但可穿戴式设备刚性强,会阻碍手部的自然运动,无法实现高质量的用户数据交互。为解决该问题,文献[4]利用体感控制器提出了一种基于多特征融合的手势识别方法,实现了较好的鲁棒分类性能,但仍然严重依赖传感器设备。

然而,机器视觉技术的迅猛发展为解决上述问题带来新的研究思路。基于计算机视觉的方法主要分为传统的手势识别方法和基于深度学习技术的手势识别方法。传统的手势识别方法通常包含特征提取与分类器选择两部分。特征提取在手势识别中起着至关重要的作用,因为其有助于将手势的复杂信息转化为计算机能够理解的数字表示。常见的特征提取方法如HOG(梯度方向直方图)、LBP(局部二值模式)、轮廓法、肤色检测等基于几何特征方法可以很好地描述手势中的纹理,形状以及颜色等信息。文献[5]将手势的梯度方向直方图(HOG)作为特征,并使用SVM对特征向量进行分类。该方法虽然对光照具有较好的鲁棒性但只适用于静态手势识别。而动态手势识别涉及到对手势的时序变化进行建模和理解,存在空间动态难捕捉和时序差异性的问题。为了克服这些问题,后续的研究逐渐引入了更加复杂的模型和技术,如动态时间规整(Dynamic Time Warping, DTW)和隐马尔可夫模型(Hidden Markov Model, HMM)以便更好地建模和识别动态

手势。文献[6]采用动态时间规整(DTW)算法对不同类手势样本进行时间对齐和归一化,并使用距离分类器将输入手势的特征向量与所有类原型特征向量进行比较得到识别结果。该方法虽然可行但极大地增加了分类的计算负荷。HMM凭借其强大的时序建模能力^[7-9],能够充分考虑先前手势对后续手势的影响,在处理具有连续动作的手势时表现出色。例如,文献[7]提出了一种基于隐马尔可夫模型的自适应阈值方法,实现了手势与背景的精准分类,但在处理复杂场景时表现较差。

与传统的前馈网络相比,深度学习具有更强的特征表达能力,可以捕获更多抽象和高级的手势特征,在动态手势识别中取得了显著的性能提升。文献[10]使用二维卷积神经网络(C2D)提取手势二值图像的空间信息,并实现相对复杂的手势识别。但该方法受限于C2D的时间建模能力,忽略了手势之间的关联性。为了更好地处理时序数据,文献[11]提出了双流网络模型,通过增加光流信息来弥补C2D的不足。然而,这种方法需要对相应的RGB图像进行像素级计算,导致了内存空间与时间资源的浪费。为了更好地捕捉动态手势中的时序信息,研究人员引入了循环神经网络(RNN)。文献[12]使用RNN来捕捉和识别手语的手势,探索了RNN在动态手势识别领域的潜力。为了增强RNN的特征可表征性,文献[13]使用长短期记忆神经网络RNNs(LSTM-RNNs)和多种运动特征来识别基于骨骼的动态手势,并证明了该方法的有效性。但该方法易受个体差异的影响,因为提取的多种运动特征都是从骨架序列中获得的,其彼此之间具有高相关度。因此,为解决上述问题,本文提出一种基于手部骨架掩膜图像的动态手势识别算法。该方法首先使用Mediapipe技术对手势视频序列的每一帧图像进行骨架提取并利用图像分割技术得到手部骨架掩膜图像。其次,使用VGG-GRU的组合网络提取手势的静态空间特征与动态空间特征,消除不同手势间的时间差异性。最后,在RWTH-Gesture数据集上使用softmax分类并进行实验评估。

2 基于时空融合的动态手势识别算法

本文提出一种基于骨架掩膜图像的动态手势识别算法,如图1所示。旨在更准确地捕捉手部动作的特征和上下文信息,从而提高动态手势识别的性能和鲁棒性。该网络主要由手势预处理、数据准备、特征提取以及手势识别4个阶段构成。

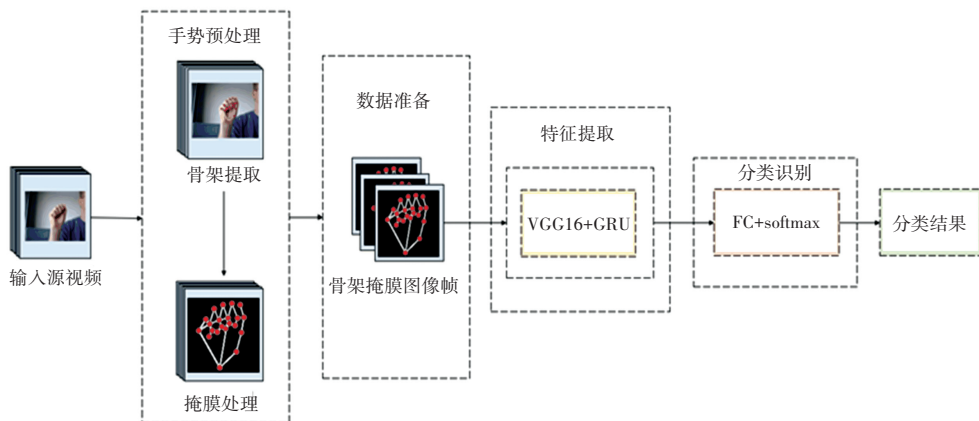


图1 本文方法框架图

Fig. 1 Diagram of the proposed method

2.1 手势预处理

在动态手势识别任务中,原始的手势视频图像通常包含大量与手势无关的冗余信息,如背景环境、用户穿着以及手部表面的纹理等细节。这些冗余信息不仅会对手势识别造成干扰,还会增加算法的复杂性。为解决该问题,本文提出了一种骨架掩膜图像方法。该方法通过提取手部骨架并生成相应的掩膜图像,减少了与手势无关的冗余信息。骨架掩膜图像仅保留手部骨架的信息,聚焦于手势的空间特征。这种图像表示消除了不同人存在手形差异的问题,使得算法能够更专注于手部骨架的形状和动态特征,从而提高手势识别的准确性和效率。与骨架图(graph)的手势识别相比,骨架掩膜图像所含空间信息更加丰富。由于骨架图仅表示手部骨架的拓扑结构,会丢失与手势相关的空间特征。而通过骨架掩膜图像,可以清晰地观察手指的位置和弯曲程度,这些信息在传统的骨架图中是无法直接获取的。基于上述两点,本文对手部骨架掩膜图像进行了构造。

为消除手势中的冗余信息并提高模型的训练速度,本文选取手部姿态估计算法 MediaPipe^[14] 对手势图像进行预处理。MediaPipe 由手掌检测器与手部坐标估计模型两部分组成。手掌检测器通过一个有方向的手部定位框定位手掌的位置,为手势检测与手势跟踪提供支持。而手部坐标估计模型在手掌检测器定位的手部区域内可以精确地回归手部关键点的坐标。本文首先使用 MediaPipe 对源视频中每一帧的手势图像进行手部姿态估计,检测手部关键点的信息。在遍历全部关键点后,根据骨架的连接关系连接关键点,完成骨架提取。最后,采用全局阈值法进行前景和背景的分割,对手势图像进行掩膜处理,整个过程如图2所示。其中,手部骨架掩膜图

像的前景 $f(x,y)$ 包括白色的手部骨架与红色的骨骼关键点。阈值公式如式(1)所示,其中 $G(x,y)$ 表示掩膜后得到的骨架掩膜图像, $g(x,y)$ 表示骨架提取后的手势图像。

$$G(x,y) = \begin{cases} f(x,y), & g(x,y) \in f(x,y) \\ 0, & g(x,y) \notin f(x,y) \end{cases} \quad (1)$$

通过屏蔽冗余的背景有助于模型更好地理解手势的精确形状,使模型专注于手部的空间信息。该数据在减少冗余信息的同时,还能完整地反映各关节点之间的关联信息。这使得算法能够更全面、准确地捕捉手的空间特征。



(a) 原始图像 (b) 骨架提取 (c) 掩膜处理

图2 手势图像预处理

Fig. 2 Gesture image preprocessing

2.2 手势骨架掩膜图像的空间特征提取

在执行手势识别任务时,需综合考虑手势的静态空间特征和动态空间特征。静态空间特征反映了手势的形状,姿势以及位置等信息,而动态空间特征则关注手势的运动模式与对时间的依赖关系。这两种特征的结合可以提供更全面、准确的手势表示,从而提高手势识别的性能和鲁棒性。本文提出了一种两阶段的训练策略,首先利用迁移学习在手势数据集上对 CNN 进行训练,其次设计 VGG-GRU 组合网络以提取手势的静态空间特征与动态空间特征,消除不同手势间的时间差异性。

2.2.1 静态空间特征的提取

本文采用 VGG16 网络的卷积层和池化层逐渐

提取骨架掩膜图像的静态空间特征。VGG16网络是一种经典的卷积神经网络,由13个二维卷积层与3个全连接层组成。初始卷积层可以捕获图像的边缘、颜色等低级特征。随着层数的加深,CNN凭借局部感知与参数共享的特性可以提取更高维的空间特征,如骨架的形状和结构。这种分层特征提取使CNN能够逐渐理解输入数据的层次结构。在前向传播的过程中,卷积层对手势骨架掩膜图像序列进行卷积操作,并利用卷积核提取图像的局部特征生成一组特征图。经过卷积操作后,特征图会通过激活函数Relu进行激活,引入非线性变换以增强网络的表达能力和非线性建模能力。在卷积层之后,使用最大池化层减小特征图的尺寸同时提取主要特征。通过重复上述操作,将每一张手势骨架掩膜图像特征编码为一个三维张量,这个三维张量表示了该手势骨架图像序列的静态空间特征。

2.2.2 动态空间特征的提取

针对动态空间特征的提取,本文采用门控循环单元(GRU)对静态空间特征进行时间上的建模。GRU网络是一种循环神经网络,具有记忆单元和门控机制,能够处理序列数据的长期依赖关系。在GRU网络中,时间步是指在时间序列中的一个离散的时间点。在GRU网络中,每个时间步对应于一个静态特征图。在前向传播过程中,GRU将当前时间步的静态特征图 x_t 以及上一个时间步的隐藏状态 h_{t-1} 作为候选隐状态 \tilde{h}_t ,输出当前时间步的新隐藏状态 h_t 。隐藏状态包含了时间序列中之前所有静态特征图的信息,并根据当前时间步的输入进行更新。GRU网络中的门控功能允许网络选择性地更新和遗忘信息,并根据当前的输入来调整隐藏状态。其更新门 z_t 决定了当前时间步的输入对隐藏状态的更新程度,而重置门 r_t 控制了之前状态在当前时间步的遗忘程度。通过门控功能,GRU网络能够自适应地更新隐藏状态,并在不同时间步之间建立长期依赖关系。相关公式如式(2)~式(5)所示。这种机制使得网络能够在不同时间步之间传递信息,逐渐构建对整个动态手势序列的理解。

更新门公式:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (2)$$

重置门公式:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (3)$$

当前时刻的输入:

$$\tilde{h}_t = \tanh(W \cdot [r_t \times h_{t-1}, x_t]) \quad (4)$$

当前时刻的输出:

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (5)$$

通过对静态特征图序列的处理和门控机制的运用,GRU网络能够有效地捕获掩膜图像的动态空间特征。在前向传播的最后一步,将GRU网络的输出隐藏状态传递给全连接层。全连接层的输出经过softmax函数进行归一化,以获得每个手势类别的概率分布。通过将GRU网络提取到的动态空间特征和时间特征应用于手势识别任务,可以实现对手势的准确分类和识别。

3 实验结果与分析

3.1 实验数据集

RWTH-Gesture数据集^[15]是一个德国手语的手指拼写手势数据集,该数据集含有35类不同手势,其中包括30类静态和5类动态手势。本文在5类动态手势上加以实验,将源视频进行预处理后得到的11 503张手部骨架掩膜图像作为样本,并按照4:1比例划分训练集与测试集。5类动态手势的一些细节见表1,图3中的(a),(b),(c),(d),(e)分别代表手语中的‘Ä’,‘J’,‘Ö’,‘Ü’和‘Z’符号。

表1 5类动态手势

Table 1 5 types of dynamic gestures

序号	手势标签	样本数	训练集	测试集
1	Ä	2 468	2 028	440
2	J	1 998	1 645	353
3	Ö	2 588	2 136	452
4	Ü	2 490	2 054	436
5	Z	1 959	1 616	343



图3 RWTH数据集示意图

Fig. 3 Diagram of actions of RWTH dataset

3.2 实验细节

本文基于 Pytorch 深度学习框架进行实验。设置 5 个卷积块的卷积核的个数由 64、128、256、512、512 递增,且在每层卷积层之后添加一层批归一化层,在每个卷积块后添加一层最大池化层。GRU 层的隐藏单元个数设为 64,批处理大小设置为 32,学习率设为 0.001 迭代 50 轮。采用 Adam 作为训练的优化算法并使用交叉熵损失函数进行验证。为避免过拟合,在训练过程中引入了随机失活操作。

3.3 评价指标

本文以手势样本的真实类别和模型预测结果为依据,将全部样本分为正样本预测正确 (True Positives, TP)、负样本预测正确 (True Negatives, TN)、正样本预测错误 (False Positives, FP)、负样本预测错误 (False Negatives, FN) 4 种情况。基于以上 4 类情况,使用 4 种评价指标衡量模型的性能,分别是准确度 (Accuracy)、精确度 (Precision)、召回

率 (Recall) 以及 $F1$, 如式(6)~式(9)所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

3.4 消融实验

为验证掩膜预处理功能与时空特征融合模块的有效性,本文以使用原始手势图像数据的 VGG16 模型作为基准在 RWTH 上进行消融实验。建立用于验证消融实验的模型,分别为模型 1:baseline;模型 2:使用掩膜预处理后的 VGG16;模型 3:本文方法。在 RWTH 数据集上的消融实验结果见表 2,在 RWTH 上的分类结果混淆矩阵如图 4 所示。

表 2 RWTH 上消融实验结果

Table 2 Results of ablation experiments on RWTH

模型	掩膜处理	静态空间特征	动态空间特征	Accuracy / %	Precision	Recall	F1
模型 1		✓		90.51	0.90	0.90	0.90
模型 2	✓	✓		92.56	0.93	0.92	0.92
模型 3	✓	✓	✓	95.52	0.96	0.95	0.95

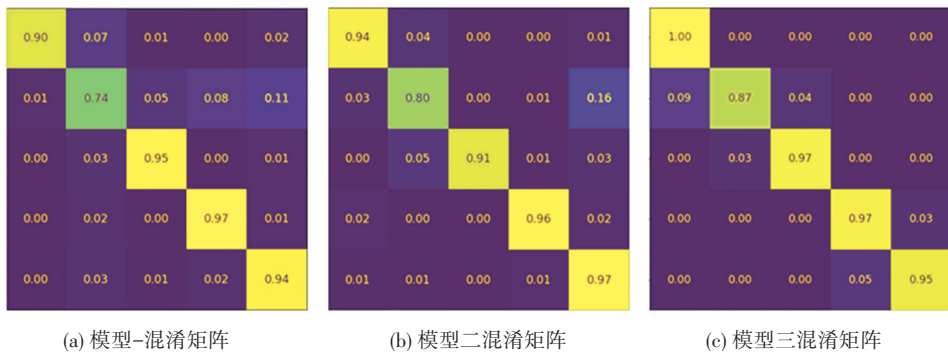


图 4 三个模型混淆矩阵图

Fig. 4 Diagram of confusion matrix for three models

图 4 清楚表示各因素对 5 种手势识别正确率的影响,其中手势 1 的识别率最高为 100%,而部分手势 2 被识别混淆成了手势 1,可能是由于两个手势都与握拳比较相似。实验中,模型一的平均准确率为 90.51%,在加入掩膜预处理技术后,模型二的平均准确率提升了 2.05%。每一轮的训练时间平均加快了 10%,模型收敛的更快。说明相比原始数据,经过掩膜处理的图像数据有以下几点优势:

(1) 消除了背景噪声和冗余信息,有效缓解了传统手势识别中的非均匀照明条件以及空间复杂性

等问题;

(2) 降低了图像处理和分析的计算复杂性,提高了算法的性能和效率;

(3) 有助于深度学习模型更好地理解手的形状和动作,从而提高手势识别的准确性。

在模型三中,通过结合掩膜图像的静态空间特征与动态空间特征,将准确率进一步提升到了 95.52%。相比模型一和模型二精度分别提升了 5.01%和 2.96%。该模型利用数据的时空互补性可以为模型训练提供更丰富的手势信息,具有更高的

准确性。

3.5 不同方法实验结果比较

为验证本文方法的高识别准确度,本文与 VGG19^[16], ResNet^[17], CNN-LSTM^[18] 模型在 RWTH 手势数据集上进行了比较,结果见表 3。比较结果可知:相较于其它未经过掩膜处理的分类模型,基于骨架掩膜图像的动态手势识别算法能够达到更高的准确率,说明时空特征融合能够提供更丰富的信息,体现了本文方法的优越性。该方法可以很好地解决时间可变性与空间复杂性问题,相较于经典的 CNN-LSTM 网络提升了近 1 个百分点。

表 3 本文方法与其他方法在 RWTH 上结果对比

Table 3 Comparison table of classification of proposed method and others on RWTH

模型	Accuracy/%	Precision	Recall	F1
VGG19 ^[16]	90.12	0.90	0.89	0.89
ResNet ^[17]	91.37	0.92	0.91	0.90
CNN-LSTM ^[18]	94.82	0.95	0.94	0.95
本文方法	95.52	0.96	0.95	0.95

4 结束语

本文研究了一种基于骨架掩膜图像的动态手势识别算法。利用手部姿态估计算法 MediaPipe 对动态手势视频进行预处理后,获得掩膜后的手部骨架图像。该方法解决了传统手势视频中存在大量冗余信息以及空间复杂性问题。此外,本文通过 VGG-GRU 的融合网络先后提取手势图像的静态与动态空间信息,在解决了时间可变性问题的同时提高了识别准确性。该方法在 RWTH 数据集上的平均识别准确率达到 95.52%,验证了本文方法的有效性。本文研究为动态手势识别领域提供了一种有前景的方法,并有望对人体行为识别领域提供有益启发。

参考文献

[1] ZIMMERMAN T G, LANIER J, BLANCHARD C, et al. A hand gesture interface device[J]. ACM Sigchi Bulletin, 1986, 18(4): 189-192.

[2] LU Z, CHEN X, LI Q, et al. A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices[J]. IEEE Transactions on Human-Machine Systems, 2014, 44(2): 293-299.

[3] ZHANG Y, HARRISON C. Tomo: Wearable, low-cost electrical

impedance tomography for hand gesture recognition [C]// Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology. IEEE, 2015: 167-173.

[4] 陈国良,葛凯凯,李聪浩. 基于多特征 HMM 融合的复杂动态手势识别[J]. 华中科技大学学报(自然科学版), 2018, 46(12): 6. DOI:CNKI:SUN:HZLG.0.2018-12-008.

[5] FENG K, YUAN F. Static hand gesture recognition based on HOG characters and support vector machines [C]// Proceedings of 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA). IEEE, 2013: 936-938.

[6] BHUYAN M K, BORA P K, GHOSH D. Trajectory guided recognition of hand gestures having only global motions [J]. World Academy of Science, Engineering and Technology, 2008, 21: 753-764.

[7] LEE H K, KIM J H. An HMM-based threshold model approach for gesture recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(10): 961-973.

[8] KESKIN C, ERKAN A, AKARUN L. Real time hand tracking and 3D gesture recognition for interactive interfaces using HMM [J]. ICANN/Iconipp, 2003, 2003: 26-29.

[9] LIU N, LOVELL B C, KOOTSOKOS P J. Evaluation of HMM training algorithms for letter hand gesture recognition [C]// Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology. IEEE, 2003: 648-651.

[10] XU P. A real-time hand gesture recognition and human-computer interaction system [J]. arXiv preprint arXiv:1704.07296, 2017.

[11] SIMON YAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos [C]. Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS). IEEE, 2014: 568-576.

[12] MURAKAMI K, TAGUCHI H. Gesture recognition using recurrent neural networks [C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. IEEE, 1991: 237-242.

[13] CHEN X, GUO H, WANG G, et al. Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition [C]// Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 2881-2885.

[14] LUGARESI C, TANG J, NASH H, et al. Mediapipe: A framework for building perception pipelines [J]. arXiv preprint arXiv:1906.08172, 2019.

[15] DREUW P, DESELAERS T, KEYSERS D, et al. Modeling image variability in appearance-based gesture recognition [C]// Proceedings of ECCV Workshop on Statistical Methods in Multi-Image and Video Processing. IEEE, 2006: 7-18.

[16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.

[17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[18] 谷学静,周自朋,郭宇承,等. 基于 CNN-LSTM 混合模型的动态手势识别方法 [J]. 计算机应用与软件, 2021, 38(11): 5.