



视觉动态手势识别综述

田秋红, 杨慧敏, 梁庆龙, 包嘉欣
(浙江理工大学信息学院, 杭州 310018)

摘要: 基于视觉的动态手势识别是人机交互领域应用较为广泛的技术,其发展对于实现人机自然交互有着重要研究意义。文章介绍并分析了目前常用的手势交互系统。通过对近年来国内外相关文献的领域研究进行梳理,概述了视觉动态手势识别的一般流程,其流程包括检测与分割、追踪、特征提取、分类,并对各流程所涉及的方法及优缺点进行了对比分析;探讨了视觉动态手势识别研究所面临的挑战性问题,和未来可能的研究方向,为推动该领域的进一步研究提供参考。

关键词: 视觉动态手势识别综述;手势分割;手势追踪;特征提取;手势分类

中图分类号: TS195.644

文献标志码: A

文章编号: 1673-3851(2020)07-0557-13

Overview on vision-based dynamic gesture recognition

TIAN QiuHong, YANG Huimin, LIANG Qinglong, BAO Jiaxin

(School of Informatics Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Vision-based dynamic gesture recognition is widely used in the field of human-computer interaction, and its development has important research significance for achieving natural human-computer interaction. This paper introduces and analyzes the current common gesture interaction system. By collating domestic and foreign relevant literatures in recent years, this paper summarizes the general process of vision-based dynamic gesture recognition, including detection and segmentation, tracking, feature extraction and classification. The methods, advantages and disadvantages involved in each process. The challenges and possible future research directions on vision-based dynamic gesture recognition are discussed, which can provide a reference for the further research in this field.

Key words: vision-based dynamic gesture recognition; overview; gesture segmentation; gesture tracking; feature extraction; gesture classification

0 引言

手势是人们进行信息交流的一个重要方式,人们可通过手的动作来表达丰富的语义信息。基于视觉的手势识别(以下简称“手势识别”)融合了先进感知技术与计算机模式识别技术,涉及数学、计算机图形学、生理学、模式识别、医学等多个学科,是多学科交

叉的研究课题,也是实现自然人机交互的重要组成部分,目前已经引起国内外学者的广泛关注。然而,由于手势动作的多样性以及手势语义的复杂性,目前的手势识别技术仍然无法实现人与机器自然交互,因此手势识别研究仍是一项具有挑战性的课题^[1]。

手势识别是跟踪和识别所执行的手势,并将其转换为能表达语义信息的词语或语句的过程。其主

收稿日期:2019-12-28 网络出版日期:2020-04-02

基金项目:国家自然科学基金项目(51405448);浙江理工大学博士科研启动项目(11122932611817);浙江省大学生科技成果推广项目(14530031661961);浙江理工大学国家级大学生创新创业训练计划项目(201910338012)

作者简介:田秋红(1976—),女,辽宁兴城人,博士,主要从事图像处理 and 模式识别方面的研究。

要分为静态手势识别和动态手势识别两种类型。静态手势识别是对手势图像进行识别,只涉及手势的二维空间特征,无需考虑到手势的时序信息。由于静态手势只能表达某一个时刻的状态,并不能检测到状态的改变,所以其表达的语义信息是有限的。动态手势识别处理对象是视频序列,在手势跟踪过程中不仅要考虑颜色、纹理和位置等信息,还需考虑手势运动的轨迹信息,及上下文的语义信息,故动态手势识别比静态识别算法难度大。由于人手是非刚性物体,本身具有多样性、多义性以及差异性等特征,动态手势相较于静态手势,具有语义多样性,更符合实际应用需求,具有更强的实用性,故本文主要聚焦在动态手势识别算法研究上。

从近年来国内外手势识别的相关文献来看,目前常见的手势交互系统大致可分为:基于传感感知的手势交互系统、基于信号的手势交互系统和基于视觉的手势交互系统。

基于传感感知的手势交互系统主要包括基于可穿戴式物理设备的手势交互系统和基于微机电系统(Micro-electro-mechanical system, MEMS)的手势交互系统。基于可穿戴式物理设备的手势交互系统通过用户穿戴内置传感器设备来采集手势运动信息,可同时获得手的位置和手指弯曲度,提供给用户以触觉反馈。常见物理设备有:数据手套、触摸屏、手柄、陀螺仪等^[2]。该类型的手势交互系统灵敏度高、实时性好、沉浸性强,信息通道获取相对独立,能提供最自然的用户反馈,适用于较复杂的手势识别研究。但要求佩戴复杂的物理设备,不仅佩戴不便,而且需要的传感器数量多,增加了额外的成本,难以推广应用。而 MEMS 传感器灵敏度高、体积小、重量轻、功耗低,并且在运动感知、方向定位方面体现出技术先进性和准确性,因此常被用来捕捉物体的运动轨迹^[3]。利用 MEMS 传感器可较方便地捕捉高速手部动作,从而将采集的手势信号集成到移动终端和其他可穿戴设备中,以进行信号分析与手势识别^[4]。但受惯性传感器精度的限制,该类交互系统计算出的运动轨迹精度有待提高^[5]。

基于信号的手势交互系统主要包括基于表面肌信号的手势交互系统和基于无线射频的手势交互系统。基于表面肌信号的手势交互系统计算量小、设备携带方便,能有效解决连续手势动作之间的运动插入问题,具有较高的动作感知可靠性和准确性,但识别的数据量有限,很难投入到实际应用中^[6-7]。为此,有学者开始研究基于无线射频的手势交互系统。

常见的覆盖面较广的无线信号就是 Wi-Fi 信号,通过 Wi-Fi 信号感知手势信息已经成为手势识别中的重要研究方法^[8]。该类交互系统进行的手势识别成本低,且无需考虑光照等背景噪声的影响,但信号的获取效果不稳定,在传输过程中容易丢失部分手势信号,造成手势信息丢失,且只能识别简单的单一动作,无法识别捕捉到的复杂动作^[9]。

基于视觉的手势交互系统通过摄像机捕捉手势运动画面,并对连续画面分析手势动作以识别出手势语义信息,是目前最流行的手势交互系统。该类交互系统输入简单,对设备的依赖性较低,且符合人们的日常交互,因此,基于视觉的动态手势交互系统一直是目前乃至将来的研究热点。

本文对近十年来视觉动态手势识别相关文献进行了研究与分析,概述了动态手势识别的一般流程和技术框架;然后对动态手势识别的各流程涉及的主要技术进行了对比分析,这些流程包括手势检测与分割、手势追踪、手势特征提取以及手势分类。此外,本文还讨论了手势识别研究所面临的挑战和局限,以及未来可能的发展方向,以促进该领域的研究工作。

1 基于视觉的动态手势识别研究现状

目前基于视觉的手势交互系统研究主要有:基于单目视觉的手势交互系统、基于双目视觉的手势交互系统和基于深度相机的手势交互系统。

基于单目视觉的手势交互系统是使用单目 RGB 相机采集手势图像序列,对采集的二维手势图像进行特征分析以识别手势语义信息。Ji 等^[10]提出了一种基于二维图像采样和拼接的手势学习系统,通过对手势演示视频进行采样和串接以构建训练数据,从而对手势进行分类识别。Cambuim 等^[11]提出了极限学习机(Extreme learning machine, ELM)模式识别算法,对基于单目视觉的 LIBRAS 符号进行识别,获得了较高的识别率,但是该方法的准确率会因光照的影响而发生变化。基于单目视觉的手势识别,成本较低、自由度很高、采集的图像和视频比较直观。手势具有多视角性,不同角度下拍摄的手势存在着较大差异,影响手势识别的准确率。

基于双目视觉的手势交互系统利用仿生学原理,通过 2 个不同定位的摄像机来采集多幅手势图像,可综合分析手势的三维几何信息^[12]。公衍超等^[13]提出一种基于双目视觉的手势识别算法,提高了裸手手势识别算法的准确率,并解决了类肤色物

体干扰的问题。虽然该类型的手势交互系统相比较单目视觉的交互系统获得的手势信息更为丰富,但是该类交互系统需要对双目镜头进行立体匹配和校准,同时需要对畸变的图像进行矫正,计算量较大,影响图像成像质量的干扰因素较多。

近年来,深度相机也应用到手势识别领域,该设备不仅可以获取图像的彩色信息,还可以直接获取手势的三维图像信息,避免了立体视觉算法产生的复杂计算量。Almeida 等^[14]利用 RGB-D 相机获取手势的方向、位置和深度信息,并从 RGB-D 图像中得到 7 个视觉特征,对 34 个巴西手语的平均识别率在 80% 以上。Tiwari 等^[15]利用 Kinect 捕捉了 10 个数字的手势图像,并将手部图像的 DCT 信号送入神经网络进行识别,对 10 个数字的 1236 个静态样本集的识别率为 98%,但是在整个数据集上的平均识别率为 83.5%。该系统仅对 10 个简单的手势进行识别,方法的可扩展性并不高。Sulfayanti 等^[16]

在基于 Kinect 的基础上设计研究了一个手势识别系统,该系统利用手指的长度、方向矢量及手指与手掌之间的矢量角,对 12 个字母进行识别,尽管该系统取得了较高的识别率,但是只对张开的手势形状进行研究,对于手势这种非刚性物体,适用性较差。

2 视觉动态手势识别关键技术

基于视觉的动态手势识别通常包括:基于机器学习的视觉动态手势识别和基于深度学习的视觉动态手势识别。基于机器学习的视觉动态手势识别需要对手势进行检测与分割,去除背景噪声,而后进行手势追踪、特征提取、手势分类及识别等手势建模过程,并选择合适的分类模型进行手势识别,识别过程如图 1 所示。而基于深度学习的视觉动态手势识别是通过组合一些简单非线性单元,将一个级别的特征表示转换到一个更加抽象的特征表示,从而自动学习原始图像的手势特征进行分类,识别过程如图 2 所示。



图1 基于机器学习的手势识别流程

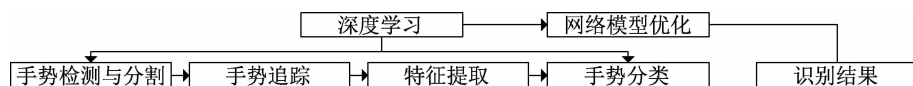


图2 基于深度学习的手势识别流程

综上分析,基于视觉的动态手势识别主要有 4 个关键步骤:手势检测与分割、手势追踪、特征提取和手

势分类。本文将以手势识别步骤为主线来对手势识别关键技术进行分析,所涉及的方法具体如图 3 所示。

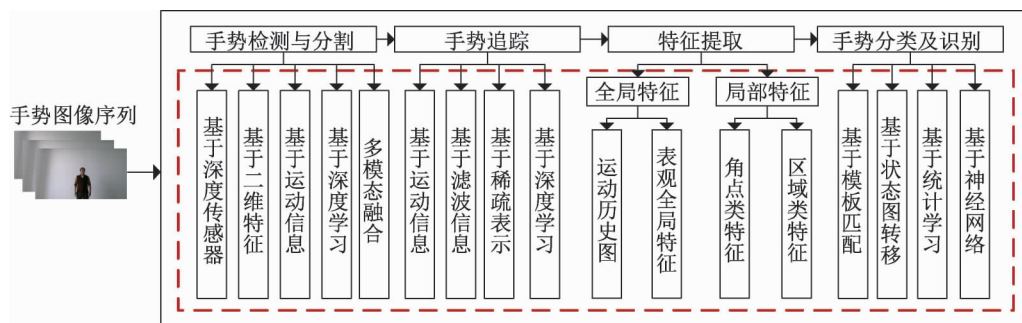


图3 动态手势识别技术框架

2.1 手势检测与分割

手势检测与分割是手势识别的首要步骤,在进行手势分割时需要将手势图像从背景模型中分离出来,手势分割的效果直接关系到手势识别的准确率。传统的手势分割算法主要有:基于深度传感器的手势分割算法、基于二维平面特征的手势分割算法和基于运动信息的手势分割算法。

基于深度传感器的手势分割方法利用深度摄像

机来获取手的深度结构信息,进而对手部区域进行分割。Ren 等^[17]提出了基于 Kinect 传感器的手势识别系统,为了解决 Kinect 产生的手部噪声,提出了一种新的测距仪 FEMD,以减小手形之间的差异,该系统在 10 个手势数据集上的平均识别率为 93.2%,然而采用该方法需要在手腕上佩戴彩色腕带来获取手部区域。为此,Inoue 等^[18]提出了一种基于 TSC 的手部区域自动提取方法,精准地实现了

手掌和手臂的分离,成功地获取了手部区域,并利用深度神经网络对获取到的手部区域进行识别,平均识别率达到88%以上。该方法的不足之处在于,当不同的受试者表示相同手势符号时,会出现符号的颜色信息和深度信息无法进行校准的问题,另外网络模型的训练也有待提高。

基于二维平面特征的手势分割方法是通过手势的表观特征来进行手部区域的提取,常见的二维平面特征包括颜色空间、边缘信息、像素信息等。类肤色区域分割是常用的基于颜色空间的手势分割算法,但该方法不能消除实际应用中背景、光照等条件的影响,无法建立一个具有较高准确度的肤色模型。Santa等^[19]设计了一种识别孟加拉手势单词和句子的系统,将视频生成的静态手势图像翻译成文本,单词识别准确率为94.26%,句子识别准确率为94.49%。但该系统对于低对比度的视频,识别效果不太理想。

基于运动信息的手势分割方法是对图像信息进行建模,将图像信息和背景模型进行比对完成目标物体的检测。Bilal等^[20]为了解决颜色空间中的类肤色问题,采用 Haar-like 和 AdaBoost 技术建立了手部检测模型。为了降低算法复杂度,提出了一种用于皮肤检测的快速搜索方法。该方法计算速度快,不受背景、光照或肤色的影响,提出的动态皮肤

斑点检测算法与其他运动检测算法相结合,可较好地解决遮挡问题。

上述的手势分割方法要借助人机交互系统,在简单背景下可以通过设计特征实现手势的分割,但在复杂环境下难以设计有效的手势特征,故难以应用于自然人机交互系统中。深度学习的崛起给手势分割带来新的契机,通过海量的数据来进行训练,自动学习目标手势特征,从而完成目标手势的检测,通过检测出来的目标手势完成相应的手势分割。任博涵^[21]使用百度提供的开源数据集,使用 DeepLab1 和 DeepLab2 网络模型对图像进行分割,并使用了迁徙学习方法,最终获得了较为精确的分割效果。Li等^[22]提出了一种将基于卷积神经网络的手部分割方法从静止图像扩展到视频图像的方法,提出的方法对失真和遮挡问题具有更强的鲁棒性,实现了更好的精度和延迟权衡。基于深度学习的手势分割方法相比较传统方法,无需人工分析手势特征进行分割,使得分割更加便捷,在手势分割方面有较好的应用前景。但就目前发展状况而言,该方法也存在着不足之处:一方面,部分网络层结构较复杂,使得手势分割的速度较慢;另一方面,部分边缘信息检测比较模糊,边缘检测精度仍有待提高^[23]。本文总结了手势检测与分割常见的方法如表1所示。

表1 手势检测与分割方法

	方法	参考文献	说明
基于二维平面信息	颜色空间	颜色直方图 [24]	容易受到类肤色区域的干扰
		肤色聚类 [25-26]	
	边缘信息	Canny 分割 [27]	通常需要进行边界闭合处理来得到更加完整的边缘信息
		Sobel 分割 [28]	
		Log 分割 [29]	
		Prewitt 分割 [29]	
		Robert 分割 [29]	
	像素信息	迭代阈值分割 [30]	更适用于带有噪声的图像
		区域生长分割 [31]	
基于运动信息	光流信息	[32]	背景变化时也能保证良好的检测结果
	背景差分法	[33]	对图像中光照变化比较敏感,噪声的积累可能形成伪目标。
	帧间差分法	[34]	
基于深度信息	提取手势的抽象特征信息,或者利用深度摄像机提取手势的深度信息	[35-37]	边缘信息检测也比较模糊
综合多要素的手势分割方法	基于角度和位置的麦克斯韦-玻尔兹曼分布的手势分割	[38]	能够精确提取手部区域,排除噪声影响
	利用形态学处理所需的不同颜色空间模型进行手部分割	[39]	
	通过深度和颜色信息融合建立的自适应肤色模型	[40]	
	基于深度图像的随机决策森林框架的手工分割策略	[41]	
	基于肤色阈值的自适应背景差分手势分割方法	[42]	
	结合改进的高斯混合模型和 K 均值的手势分割	[43]	

由表1可知,手势是一种非刚性物体,其手势变化的方向、手指的弯曲度等因素都影响着手势的分割效果,通用的区域分割算法虽然能够将图像和背景区域进行分离,但是图像的处理结果会受到两种因素的影响:a)类肤色区域;b)身体其他区域及背景噪声的干扰。通过分析文献[24-37]可发现,如果只提取手部区域,单一的通用方法难以避免背景噪声的影响,也无法保证手部区域分割的完整性。而综合多要素的手势分割方法能排除噪声干扰,实现手势的精准分割,见文献[38-43]。因此,为了保证手势的分割效果,应针对手形的变化,设计更为合理的

手势分割方法,以提高分割算法的鲁棒性。

2.2 手势追踪

手势追踪是手势识别中的重要环节。手势追踪的本质是对时序上连续的图像进行逐帧分析,在图像变化间隔之内计算出被追踪的目标。由于手势的动作变换非常快,自由度比较高,且目标相对较小,所以通常对手部的追踪是困难的。从近年来手势追踪的国内外文献及技术实现上来看,常见的手势追踪算法主要包括:基于运动信息的手势追踪、基于区域的手势跟踪、基于稀疏表示的手势追踪以及基于深度信息的手势追踪。常用的手势追踪方法对比情况如表2所示。

表2 手势追踪方法原理及优缺点

方法	原理	优点	缺点
基于运动信息	帧差法 ^[44]	基于前后两帧图像对应像素点的灰度值相减	实现简单,对背景环境和光照变化的敏感度较低
	背景差分法 ^[45]	建立恰当的背景模型,将当前帧和得到的背景帧对应的像素值进行相减	检测效果较明显,能够较准确的检测提取出前景目标像素点
	光流法 ^[46]	通过手部所有点的运动方向和运动速率的变化来进行手部区域的确定	背景变化易导致差值图像噪声,从而易形成伪目标造成错误检测
基于区域	Meanshift ^[47]	在理想情况下,能够检测出独立运动的目标,无需确定运动场景信息	边缘检测较模糊、运算量巨大、实时性不高
	Camshift ^[48]	比较候选目标可能位置与目标模板进行相似度比较	算法简洁、实时性好
基于稀疏表示	常结合粒子滤波算法 ^[49]	可以自适应改变窗口大小,很好的适应目标形变	目标大小变化时,容易出现跟踪漂移
基于深度学习 ^[50]	—	假设跟踪目标在一个由目标模板构成的子空间内,在此空间内寻求与模板重构误差最小的最佳候选	周围环境复杂时,会导致跟踪窗口发散,造成目标丢失
		计算速度快、解释性好、低储存	目标与背景相似时,容易出现误匹配、容易丢失细节特征
		提取运动目标特征,训练分类器	算法运行速度慢、容易丢失细节信息

由表2可知,在实际应用中,单一的跟踪算法难以准确追踪手势位置。因此,往往需要结合多种算法来提高目标跟踪方法的鲁棒性。Yuan等^[51]提出了一种多特征融合模型,并将其应用于视觉跟踪的相关滤波框架中,该模型可以有效地提高跟踪器的跟踪性能和鲁棒性。Danelljan等^[52]研究了颜色在逐点检测框架中的作用,提出了一种自适应的低维颜色属性变换方法以提高跟踪效果。针对传统Meanshift跟踪算法解决目标遮挡问题时易出现目标丢失的情况,Khan等^[53]将传统Meanshift跟踪算法与运动目标的空间信息相结合,解决了运动目标的遮挡问题。李妍妍等^[54]提出一种自适应跟踪窗口的处理方法,实现目标的有效检测和跟踪。实验结果表明:该方法能有效降低目标跟丢的概率,预防目标的误跟踪,适应目标尺度变化。虽然基于稀疏表示的方法与基于深度学习的方法改善了目标追

踪效果,但是手势特征的细节信息容易丢失。为此,Qian等^[55]提出了一种基于深度学习和自适应粒子滤波的在线视觉跟踪算法,该算法在目标或摄像机快速移动时表现出了先进性。Zhao等^[56]提出一种采用结构化局部稀疏表示方法对目标周围的背景区域进行分析,并设计了一种加权搜索方法寻找最佳候选目标。

2.3 手势特征提取

手势特征提取是手势识别中的关键步骤。特征提取是将输入图像中的感兴趣部分转换成一组紧凑的特征向量。在手势识别中,提取的手势特征应包含从输入手势中获得的相关信息和区别于其他手势的特征表示信息。手势特征的选择和设计不仅关系到手势识别的准确率,还关系到系统的复杂度和实时性。手势特征主要分为全局性特征和局部性特征。常见的手势全局性特征包括颜色、纹理、形状等

特征。Ibrahim 等^[57]采用基于人脸色调的动态皮肤检测器进行手部分割,并采用皮肤斑点跟踪技术对包含 30 个孤立单词的数据集进行识别和跟踪,识别率可达 97%。王景中等^[58]采用 Local binary pattern (LBP) 变换与主成分分析 (Principal component analysis, PCA) 进行特征提取与压缩,并对提取出的手势特征进行分类识别,提出的算法有效提高了识别率与速度,识别率达 94.22%。全局类特征计算简单、表示直观,具有良好的不变性。但该类特征多为基于像素点的特征表示,因此存在着特征维数高、计算量大等问题。此外,该类特征描述不适用于图像混叠和有遮挡的情况。

近年来,局部图像特征在手势识别领域得到了广泛的应用。局部特征是从图像局部区域中抽取的特征,包括角点类和区域类两大类。Harris 是最常见的角点类特征。Chen 等^[59]利用 Harris 对二值图像进行特征点检测,并利用特征点的先验信息和向心凸包分析的全局优化方法对指尖进行定位。Hsieh 等^[60]研究了 4 组新的手势 Harr-like 特征,对复杂背景下的手势进行了分类。50 个动态手势的平均准确率为 95.37%。Scale-invariant feature transform (SIFT) 是 Lowe^[61]提出的一种尺度和旋转不变的特征提取技术,目前已广泛用到手势识别中。Tharwat 等^[62]利用 SIFT 技术对阿拉伯手语 (ArSL) 进行识别,系统对任何旋转都具有较强的鲁棒性,识别率接近 99%。方向梯度直方图 (Histogram of oriented gradient, HOG) 特征也是一种常用在计算机视觉和图像处理中的特征描述子^[63]。Gupta 等^[64]提取出手势的 HOG 特征和 SIFT 特征,并组合成特征矩阵来获取识别结果。Dardas 等^[65]将图像中提取的 SIFT 特征映射到特征包 (Bag-of-feature, BoF), 并使用 K -means 算法进行量化。Speeded up robust features (SURF) 是基于 SIFT 发展的描述符,相比较 SIFT, SURF 有着计算迅速的绝对优势。Sykora 等^[66]使用支持向量机 (Support vector machine, SVM) 分类器对 500 个测试图像提取的 SIFT 和 SURF 特征进行分类,识别率分别为 81.2% 和 82.8%。Hartanto 等^[67]从 24 种手语的 288 张图片中提取了 SURF 特征进行识别,总的准确率是 63%。实验结果表明,只有手势的旋转角度小于 15° , SURF 特征才具有旋转不变性。因此, SURF 在旋转不变性方面仍逊色于 SIFT。

以上采用传统的人工设计特征很容易受到光照、背景的影响,且仅用于解决一种特定数据集上的

特定问题。因此,寻求识别多类型数据集的通用方法具有重要意义。利用深度学习的方法来获取手势特征不仅可以避免背景噪声,还可以学习不同维度的手势抽象特征。Simonyan 等^[68]提出了双流卷积神经网络,将图像信息和光流信息分别输入 3D 卷积神经网络中训练,此模型结构很好地考虑了时间信息和空间信息,在公开的数据集上获得了较高的识别率,但是将光流信息输入 3D 卷积神经网络时,由于局部特征集中在以轨迹为中心的时空管中,而该方法并没有考虑轨迹信息,所以获取的局部特征并不充分。Du 等^[69]提出一种基于人体骨架识别的递归神经网络模型,首先将人体骨架分为 5 部分并分别送入 5 个子网,之后将子网提取的特征融合为更高级的特征作为网络模型的输入,实验结果显示:该网络模型有着较高的计算效率和不错的性能,但是在处理非常相似的手势时,仅仅依靠骨骼点很难对其进行区分。

在研究手势图像特征时,由于手势区域在图像中占比较小,利用全局特征难以提取到感兴趣的手部区域,图像表现能力较差。而局部图像特征数量丰富,稳定性好,与全局图像特征相比,特征间相关度小,一定程度上可避免手部区域遮挡的情况。而且,对光照、旋转、视点变化等图像变换具有鲁棒性。因此大多数学者为了丰富手势的特征信息,往往将局部特征和全局特征进行融合,以获得较高的识别率。本文对近年来部分文献中提到的手势特征以及用到的实验数据和准确率进行总结,结果如表 3 所示。

2.4 手势分类

手势分类是对提取的手势时空特征进行分类,是实现手势识别的最后一个核心环节。常见的动态手势分类算法大概可以分为四种:基于模板匹配的算法、基于状态图转移的算法、基于统计学习的算法和基于神经网络的算法。

模板匹配是通过对大量手势样本进行学习得到手势模板,并将目标数据与手势模板进行匹配。动态时间归整算法 (Dynamic time warping, DTW) 是最经典的用于动态手势分类的模板匹配算法,当两个时间序列的长度和速度不相同,该算法可很好地检测出其相似性。郭鑫鹏等^[80]提出一种基于手势轨迹的自适应关键帧提取算法和基于加权关键帧序列的识别算法,该系统可以对 26 个常用词汇的关键帧进行识别,但该方法并不能保证非特定人的识别率,且需要借助体感设备,交互性也有待提高。梁文乐等^[81]提出一种针对动态手势的分级匹配算法,

表3 部分手势特征提取技术

参考文献	特征	实验数据	准确率/%
[70]	混合 Hu 矩、手指角度计数、肤色角度和非肤色角度的融合特征	10 种手势	90.0
[71]	LBP	中国数字手势数据集,孟加拉数字手势数据集	87.3,85.1
[72]	距手端点的距离、角度	10 种手势	92.13
[73]	手势边缘信息、肤色区域边界	带有 10 种手势的 800 个样本	—
[74]	SURF、轨迹方向	26×40 个手势样本	84.6
[75]	PCBR、WPD-2	23×40 个手势样本	86.3(静态),77.2(动态)
[76]	峰度位置、PCA 和运动链码	50 个手势样本	89.1
[77]	边缘信息、手中心位置	20×40 个手势样本	92.7
[78]	SURF、LCS	4 种手势	93.0
[79]	WD、MFCC	23×8 个静态手势样本,19×8 个动态手势样本	—

利用体感设备获得手势轨迹,根据点轨迹密度设计了一种关键帧算法,并利用优化的 DTW 算法实现手势匹配,所提算法解决了 DTW 算法实时性低的问题,准确度高。

隐马尔可夫模型 (Hidden Markov model, HMM) 作为基于状态图转移的典型方法,具有分析时空变异性的时变数据的能力。近年来在手势识别方面得到了广泛应用。Chen 等^[82]利用 HMM 对包含 20 种手势的 1200 个测试序列进行识别,识别结果达到 98.5%。Elmezain 等^[83]使用 200 个手势训练样本来对 10 个动态手势进行分类,98 个测试序列的识别率为 94.29%。Hazmoune 等^[84]将 HMM 集成到 KNN 结构中对阿拉伯数字数据集进行分类,表明该方法有效。

常见的基于统计学习的方法主要有支持向量机和 K 最近邻。Chen 等^[85]采用了 SVM 对包含了数字的三维运动轨迹和字母表 (共 36 个手势) 进行识别。实验结果表明,该系统具有较高的识别率和准确率。Yu 等^[86]利用 SVM 进行姿态识别,同时利用高斯金字塔光流算法进行动态姿态跟踪特征的提取和识别,获得了较高的识别率。Likhitha 等^[87]利用 PCA、KNN 和 SVM 提出一种手势识别系统方案,成功应用在机器人领域中。

近年来深度神经网络在目标识别、图像分类等领域取得了重大突破,由于其具有强大的自主学习能力和自适应性等优点,目前已普遍应用于手势识别领域。常见的神经网络有人工神经网络 (Artificial neural network, ANN)、卷积神经网络 (Convolutional neural networks, CNN)、深度神经网络 (Deep neural networks, DNN)。ElBadawy 等^[88]通过从视频流里获取图像帧,并将图像帧归一化后送入 3D 卷积神经网络完成对手势的识别,识别率平均可达 90%,但该

方法只在有限的样本集上进行训练,识别的精确率也有待提高。Carreira 等^[89]提出了一种新的基于 2D 卷积网的网络模型 (I3D) 来进行动作识别,将深层网络的过滤器和卷积核扩展到三维,从而能够从视频中完整地提取视频特征,该模型在 miniKinetics 上进行预训练后,虽然相对于最先进的动态分类有了很大的改进,但是在 HMDB-51 上识别率达到 80.2%,在 UCF-101 上识别率却达到了 97.9%,在不同的数据集上体现出了很大的差异性,这也是该模型的局限所在。

传统的手势识别方法需要针对手势数据集来训练特定的分类器,而且训练的数据量有限,不适用于大型数据集,难以推广使用。而基于神经网络的方法由于其强大的自学习能力,能够自动获取手势样本的高级特征,能适用于通用数据集。在训练神经网络时,低层特征分辨率较高,主要对应于手势的细节信息,高层特征对应于抽象的语义信息。但对于手势图像,由于手势在图像中占比较小,大多数方法为了丰富手势的特征信息,实现较高的识别精度,需要设计较深的网络层,消耗大量训练时间和运行内存,难以保证手势识别的实时性。

手势识别的常用方法对比情况如表 4 所示。

3 视觉动态手势识别面临的挑战性问题及发展方向

近年来,手势识别技术经过理论探索和不断的实践创新,研究结果已经取得了巨大的进步,目前成熟的技术已经可以对简单的动态手势进行识别,在特定的应用场景下,已经可以满足实际需求。然而,要建立一个稳健的手势识别系统,仍然存在一些障碍,如手部遮挡、手势的误追踪、手势数据库的可扩展性、手势动作的不可预测性、不同的背景光照以及

表 4 手势识别常用方法的基本思想及优缺点

识别方法	涉及算法	基本思想	优点	缺点
基于模板匹配的方法	DTW	比对输入手势与模板手势之间的相似度	识别精度高、复杂度低	没有统一的模型、实时性差
基于状态转移的图模型方法	HMM	通过状态转移的随机过程和观察值输出的过程来确定手势信号	识别效率高、易于训练、较强的时空变化能力	算法复杂、训练和识别计算量大、识别速度慢
基于统计学习的方法	SVM	在特征空间中寻找一个不同类别之间具有最大间隔的超平面进行分类	对噪声的抵抗能力好	训练数据少,只是用于小型样本集
	KNN	比较训练样本在测试空间中最近的 K 个样本类别	分类效果好	消耗大量的内存空间和计算资源
基于神经网络的方法	ANN、DNN、CNN、3D-CNN	从大量原始输入中自动学习样本的高级特征	并行性高、自组织和自学习能力强	学习训练时间长、数据量大、不能实时识别

高昂的计算成本,使得现有的算法在实时性、运算速度、识别率等方面还有诸多问题有待解决。

3.1 面临的挑战性问题

现有的视觉动态手势识别方法存在的主要问题如下:

a) 手势分割问题。手势分割的准确性影响到手势识别的效果,但手势分割比较复杂,其复杂性主要表现在两个方面:一方面为背景环境的复杂性,手势的环境很容易受到颜色、亮度、光照等背景噪声的影响,从而影响到手势的准确分割;另一方面为手势的复杂性,人手的自由度很高,手势动作存在细微性,人手的弯曲程度、手和身体其他部位的遮挡问题等都会影响到手势的准确分割。

b) 动态手势的追踪与匹配问题。在动态手势追踪过程中,手势的外观变化,变化的杂乱背景和对实时速度的要求,使得视觉跟踪一直是一个具有挑战性的研究课题。在对动态手势进行追踪检测时,存在的问题主要有三个方面:一,在对手势的运动轨迹曲线进行分类时,由于不同的手势者描述手势的习惯不同,所呈现出来的手势轨迹曲线也存在差别;二,目标手势在跟踪过程中,跟踪窗口大小可能会发生变化,很容易出现跟踪漂移问题;三,在跟踪过程中,目标手势的姿态会不断发生变化,导致跟踪精度降低。

c) 实时性与计算量问题。对于手势图像,手势在图像中占比较小,大多数方法为了丰富手势的特征信息,实现较高的识别精度,需要设计较深的网络层,然而训练和测试大型深度神经网络存在几点限制:一,昂贵的计算成本。训练和测试深度神经网络不但要求较强的计算能力,而且需要花费大量的运行时间,必须使用高性能的计算设备来加快训练和测试速度;二,数据量。训练深度神经网络需要大量

的标记本来保证识别率,而人工标记样本也是巨大的工程量;三,内存存储。深度神经网络模型通常包含大量的参数,需要大容量的内存来存储模型。

d) 手势实际应用问题。目前大多数对手势识别的研究往往停留在小型的简单数据库或者自建的数据层面,不具有泛化性,不易于推广和使用。如何保证手势识别的通用性,提高手势识别的应用性也是值得研究的问题。另一方面,目前的手势交互系统很少集成在移动终端设备,也是交互性方面的一个短板。

3.2 可能的研究方向

针对以上问题,为了提高手势识别的准确率,未来可着眼于以下研究方向:

a) 基于点云的三维手势重建。目前对视觉动态手势的识别往往由于光照、旋转等问题导致识别率低下。传统的手势建模往往针对手势的运动信息或表观特征,对手势进行建模分析。基于运动信息的手势建模考虑了手势的时空特征,但得到的手势边缘信息较模糊。基于表观特征的手势建模计算量较小,但未充分考虑到手势的形变能力。目前的RGB-D摄像机将手势视为二维结构,使得手势模型的外观变化与手势旋转问题面临挑战性。另一方面,RGB-D摄像机成本高,且要求手势图像的背景模型简单,实用性不强。基于点云的三维目标重建,可生成一个精确的空间支持,也可从手势的二维投影中学习受限的特征表示,从而稳定地定位在视野外旋转或自遮挡问题严重的手势目标。

b) 基于多区域检测的手势追踪。光照变化是视觉手势跟踪中的常见问题,严重影响着系统性能。多区域检测策略可在并行计算的基础上,对多个区域进行同时检测,等效地扩大搜索区域,有效地考虑了不同布局和区域变化的视觉感知,提高目标手势

区域的显著性,在此基础上,可采集目标手势模板作为备用模板,并利用模板更新机制保存备用模板,以提高相关跟踪器在强光照条件下的准确性和鲁棒性。

c)基于多粒度稀疏表示的手势追踪。通常手势区域在手势图像中所占比例较小,在分析手势图像时,应充分考虑到手势的整体特征和局部特征,以提高手势特征的表达能力。基于多粒度稀疏表示的手势跟踪方法,既充分利用了整体特征和局部特征的有效性,又充分利用了不同粒度下多个补丁的表示能力。通过对手势图像进行多粒度稀疏性分析,对不同尺寸目标手势的联合表示能力进行充分考虑,以提高手势特征的表达性能。

d)基于多流深度相似性学习网络的手势追踪。目前,CNN被广泛地应用于从大规模数据集中学习健壮的特征表示和复杂的模型。在对象分类^[90]、目标检测分类^[91]、语义分割^[92]等各种任务上取得了相应的进步,但目前的方法大多将跟踪问题视为一个二分类问题,通过训练分类器从背景中识别出目标。在测试过程中,采用梯度下降算法对网络进行更新,但由于模型更新速度太慢,无法进行实时操作。另外,一般的分类器如CNN模型,通过预先训练数据集来区分不同类别的对象,对于区分同一类对象之间的差异并未显示出优越性。因此,该分类器所获得的特征不足以捕获目标手势的外观变化。为了更好地处理跟踪特定目标的任务,多流深度相似性学习网络把分类问题当作实例级的验证问题。通过离线学习一个通用的深度相似性比较模型来追踪后续图像帧中的目标手势。该模型不需要从一般的分类网络中转移特征空间,而是有效地学习与手势实例进行相似性比较的特征空间,避免了同类目标之间的相似性比较。学习模型直接用于定位每帧目标手势,不需要在线学习或参数微调,实现对目标手势的高效判别。

e)基于自适应卷积神经网络的手势分类。CNN已被证明可以在视觉目标跟踪任务中获得最佳的性能。然而,现有的基于CNN的跟踪器通常使用整体目标样本来训练网络,一旦目标经历了复杂的情况(如遮挡、背景杂乱和变形),跟踪性能就会严重下降。自适应卷积核模型,通过设计一个掩码集来生成卷积核来捕获目标手势结构。同时,对卷积核采用自适应加权融合策略,以适应目标手势外观的变化,可有效地增强跟踪器的鲁棒性。对于手势这种非刚性目标,可提高其显著性检测能力和分

类的准确率。

4 结 语

手势识别具有广泛的应用潜力,如手语识别、遥控机器人、3D游戏和虚拟现实等,是一个持续的研究热点。本文从手势检测与分割、手势追踪、特征提取、手势分类等方面对视觉动态手势识别技术进行分析,总结了各类方法的优缺点。就当前研究情况来看,传统的基于机器学习的动态手势识别仍然存在着较多的局限性,主要原因在于手势动作变换的多样性、手势语义的复杂性以及背景噪声的干扰,在数据量较少时表现出其自身优越性。基于深度学习的动态手势识别发展较为迅速,应用较广,是当前研究的热点。而如何在保证识别准确率的情况下,有效地减少高计算成本和运行时间,提高手势识别的实时性是未来研究的重点。另外,现有手势识别系统中使用的手势仅限于精心选择的手势词汇(象征性手势),数据库的可扩展性也是在实际应用中值得研究的问题。

参考文献:

- [1] 武霞, 张崎, 许艳旭. 手势识别研究发展现状综述[J]. 电子科技, 2013, 26(6): 171-174.
- [2] Hassan M, Assaleh K, Shanableh T. User-dependent sign language recognition using motion detection[C]// 2016 International Conference on Computational Science and Computational Intelligence (CSCI). Las Vegas: IEEE, 2016: 852-856.
- [3] Hyun D, Jegal M, Yang H S. Compact self-contained navigation system with MEMS inertial sensor and optical navigation sensor for 3-D pipeline mapping[C]// 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei: IEEE, 2010: 1488-1493.
- [4] Cao Y, Xi Z W. A review of MEMS inertial switches [J]. Microsystem Technologies, 2019, 25(12): 4405-4425.
- [5] Ciuti G, Ricotti L, Menciassi A, et al. MEMS sensor technologies for human centred applications in healthcare, physical activities, safety and environmental sensing: a review on research activities in Italy [J]. Sensors, 2015, 15(3): 6441-6468.
- [6] Zhang X, Chen X, Li Y, et al. A framework for hand gesture recognition based on accelerometer and EMG sensors[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2011, 41(6): 1064-1076.

- [7] Moin A, Zhou A, Benatti S, et al. Adaptive EMG-based hand gesture recognition using hyperdimensional computing [EB/OL]. (2019-08-30) [2019-12-28]. <https://arxiv.org/abs/1901.00234>.
- [8] Li H, Yang W, Wang J, et al. WiFinger: talk to your smart devices with finger-grained gesture [C]// Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Heidelberg: 2016: 250-261.
- [9] Wang F X, Gong W, Liu J C, et al. Channel selective activity recognition with WiFi: A deep learning approach exploring wideband information[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(1): 181-192.
- [10] Ji Y, Kim S, Lee K B. Sign language learning system with image sampling and convolutional neural network [C]//2017 First IEEE International Conference on Robotic Computing (IRC), Taiwan: IEEE, 2017: 371-375.
- [11] Cambuim L F S, Macieira R M, Neto F M P, et al. An efficient static gesture recognizer embedded system based on ELM pattern recognition algorithm [J]. Journal of Systems Architecture, 2016, 68: 1-16.
- [12] Stidwill D, Fletcher R. Normal binocular vision: Theory, investigation and practical aspects[M]. West Sussex: John Wiley & Sons, 2017.
- [13] 公衍超, 万帅, 杨楷芳, 等. 利用双目视觉视频的实时三维裸手势识别[J]. 西安电子科技大学学报(自然科学版), 2014, 41(4): 130-136.
- [14] Moreira Almeida S G, Guimarães F G, Arturo Ramírez J. Feature extraction in Brazilian sign language recognition based on phonological structure and using RGB-D sensors[J]. Expert Systems with Applications, 2014, 41(16): 7259-7271.
- [15] Tiwari V, Anand V, Keskar A G, et al. Sign language recognition through kinect based depth images and neural network[C]//2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). Kochi, India: IEEE, 2015: 194-198.
- [16] Sulfayanti, Dewiani, Lawi A. A real time alphabets sign language recognition system using hands tracking [C]//2016 International Conference on Computational Intelligence and Cybernetics. Makassar, Indonesia: IEEE, 2016: 69-72.
- [17] Ren Z, Yuan J S, Meng J J, et al. Robust part-based hand gesture recognition using kinect sensor[J]. IEEE Transactions on Multimedia, 2013, 15(5): 1110-1120.
- [18] Inoue K, Shiraishi T, Yoshioka M, et al. Depth sensor based automatic hand region extraction by using time-series curve and its application to Japanese finger-spelled sign language recognition [J]. Procedia Computer Science, 2015, 60: 371-380.
- [19] Santa U, Tazreen F, Chowdhury S A. Bangladeshi hand sign language recognition from video[C]//2017 20th International Conference of Computer and Information Technology (ICCIT). Dhaka: IEEE, 2017: 1-4.
- [20] Bilal S, Akmeliawati R, Salami M J E, et al. Dynamic approach for real-time skin detection[J]. Journal of Real-Time Image Processing, 2015, 10(2): 371-385.
- [21] 任博涵. 利用深度学习进行图像分割[J]. 中国设备工程, 2018(23): 79-81.
- [22] Li M L, Sun L, Huo Q. Flow-guided feature propagation with occlusion aware detail enhancement for hand segmentation in egocentric videos [J]. Computer Vision and Image Understanding, 2019, 187: 102785.
- [23] 蒋昀昊. 利用深度学习进行图像分割[J]. 通讯世界, 2018(6): 243-244.
- [24] Paul S, Bhattacharyya A, Mollah A F, et al. Hand segmentation from complex background for gesture recognition [M]// Mandal J, Bhattacharya D. Emerging Technology in Modelling and Graphics. Singapore: Springer International Publishing, 2020: 775-782.
- [25] Nallaperumal K, Ravi S, Babu C N K, et al. Skin detection using color pixel classification with application to face detection: A comparative study [C]//International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007). Sivakasi, Tamil Nadu, India: IEEE, 2007: 436-441.
- [26] Chandrappa D N, Ravishankar M, Ramesh Babu D R. Face detection in color images using skin color model algorithm based on skin color information[C]//2011 3rd International Conference on Electronics Computer Technology. Kanyakumari, India: IEEE, 2011: 254-258.
- [27] Yan B, Li Y, Ren S T, et al. Recognition and evaluation of corrosion profile via pulse-modulation eddy current inspection in conjunction with improved Canny algorithm[J]. NDT & E International, 2019, 106: 18-28.
- [28] Zhou R G, Liu D Q. Quantum image edge extraction based on improved sobel operator [J]. International

- Journal of Theoretical Physics, 2019, 58(9): 2969-2985.
- [29] Muthukrishnan R, Radha M. Edge detection techniques for image segmentation[J]. International Journal of Computer Science and Information Technology, 2011, 3(6): 259-267.
- [30] Zhu Q, Jing L, Bi R. Exploration and improvement of Otsu threshold segmentation algorithm[C]//2010 8th World Congress on Intelligent Control and Automation. Jinan, China: IEEE, 2010: 6183-6188.
- [31] 严静, 刘启榆, 周莹, 等. 基于显著性定位和动态自适应区域生长的 HRMR 图像斑块分割[J]. 计算机应用研究, 2019, 36(11): 3499-3503.
- [32] Narayana M, Hanson A, Learned-Miller E. Coherent motion segmentation in moving camera videos using optical flow orientations[C]//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 1577-1584.
- [33] Sun P, Lv L, Qin J, et al. Moving target detection based on multi-feature adaptive background model [C]//2019 IEEE International Conference on Mechatronics and Automation (ICMA). Tianjin, China: IEEE, 2019: 1610-1614.
- [34] Gao F, Lu Y. Moving target detection using inter-frame difference methods combined with texture features and lab color space[C]//2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). Dublin, Ireland: IEEE, 2019: 76-81.
- [35] Vodopivec T, Lepetit V, Peer P. Fine hand segmentation using convolutional neural networks[EB/OL]. (2016-08-26) [2019-12-28]. <https://arxiv.org/abs/1608.07454>.
- [36] Nguyen H D, Kim S. H Hand segmentation and fingertip tracking from depth camera images using deep convolutional neural network and multi-task SegNet [EB/OL]. (2019-03-05) [2019-12-28]. <https://arxiv.org/abs/1901.03465>.
- [37] Dadashzadeh A, Targhi A T, Tahmasbi M, et al. HGR-Net: a fusion network for hand gesture segmentation and recognition [J]. IET Computer Vision, 2019, 13(8): 700-707.
- [38] Betancourt A, Morerio P, Barakova E, et al. Left/right hand segmentation in egocentric videos [J]. Computer Vision and Image Understanding, 2017, 154: 73-81.
- [39] Mazumdar D, Nayak M K, Talukdar A K. Adaptive hand segmentation and tracking for application in continuous hand gesture recognition [M]// Recent Trends in Intelligent and Emerging Systems. New Delhi: Springer International Publishing, 2015: 115-124.
- [40] Heo G, Lee D W, Shin H C, et al. Hand segmentation and fingertip detection for interfacing of stereo vision-based smart glasses [C]//2015 IEEE International Conference on Consumer Electronics (ICCE). Las Vegas, USA: IEEE, 2015: 585-586.
- [41] Zhao M Y, Jia Q X. Hand segmentation using randomized decision forest based on depth images[C]//2016 International Conference on Virtual Reality and Visualization (ICVRV). Hangzhou, China: IEEE, 2016: 110-113.
- [42] Elsayed R A, Sayed M S, Abdalla M I. Skin-based adaptive background subtraction for hand gesture segmentation[C]//2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS). Cairo, Egypt: IEEE, 2015: 33-36.
- [43] Zheng Y, Zheng P. Hand Segmentation Based on Improved Gaussian Mixture Model [C]//2015 International Conference on Computer Science and Applications (CSA). Wuhan, China: IEEE, 2015: 168-171.
- [44] 岳昊恩, 袁亮, 吕凯. 结合帧差法与 Mean Shift 的抗遮挡跟踪算法[J]. 现代电子技术, 2019, 42(12): 180-182, 186.
- [45] Garcia-Garcia B, Bouwmans T, Rosales Silva A J. Background subtraction in real applications: Challenges, current models and future directions[J]. Computer Science Review, 2020, 35: 100204.
- [46] Ranjan A, Romero J, Black M. Learning human optical flow [EB/OL]. (2018-07-22) [2019-12-28]. <https://arxiv.org/abs/1806.05666>.
- [47] Guo Y H, Şengür A, Akbulut Y, et al. An effective color image segmentation approach using neutrosophic adaptive mean shift clustering [J]. Measurement, 2018, 119: 28-40.
- [48] Allen J G, Xu R Y D, Jin J S. Object tracking using camshift algorithm and multiple quantized feature spaces [C]// Proceedings of the Pan-Sydney area workshop on Visual information processing. Sydney, Australia: Australian Computer Society, 2004: 3-7.
- [49] 孙玉宝, 肖亮, 韦志辉, 等. 图像稀疏表示的结构自适应子空间匹配追踪算法研究[J]. 计算机学报, 2012 (8): 191-198.
- [50] Cheng J, Tsai Y H, Hung W C, et al. Fast and accurate online video object segmentation via tracking

- parts [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City; IEEE, 2018; 7415-7424.
- [51] Yuan D, Zhang X M, Liu J Q, et al. A multiple feature fused model for visual object tracking via correlation filters [J]. *Multimedia Tools and Applications*, 2019, 78(19): 27271-27290.
- [52] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE, 2014; 1090-1097.
- [53] Khan B, Khan A K, Raja G, et al. Implementation of modified mean-shift tracking algorithm for occlusion handling[J]. *Life Science Journal*, 2013, 10(11): 337-342.
- [54] 李妍妍, 田瑞娟, 张弦弦. 一种基于帧差法结合 Kalman 滤波的运动目标跟踪方法[J]. *兵工自动化*, 2019, 38(4): 24-27.
- [55] Qian X Y, Han L, Zhang Y L, et al. An object tracking method using deep learning and adaptive particle filter for night fusion image [C]//2017 International Conference on Progress in Informatics and Computing (PIC). Nanjing, China; IEEE, 2017; 138-142.
- [56] Zhao H Q, Xiang K, Cao S X, et al. Robust visual tracking via CAMShift and structural local sparse appearance model[J]. *Journal of Visual Communication and Image Representation*, 2016, 34: 176-186.
- [57] Ibrahim N B, Selim M M, Zayed H H. An Automatic Arabic Sign Language Recognition System (ArSLRS) [J]. *Journal of King Saud University-Computer and Information Sciences*, 2018, 30(4): 470-477.
- [58] 王景中, 李萌. 基于 LBP 和 PCA 机器学习的手势识别算法[J]. *计算机测量与控制*, 2015, 23(4): 1320-1322, 1326.
- [59] Chen J J, Han M M, Yang S, et al. A fingertips detection method based on the combination of centroid and Harris corner algorithm [C]//2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Shanghai, China; IEEE, 2016; 225-230.
- [60] Hsieh C C, Liou D H. Novel Haar features for real-time hand gesture recognition using SVM[J]. *Journal of Real-Time Image Processing*, 2015, 10(2): 357-370.
- [61] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International journal of computer vision*, 2004, 60(2): 91-110.
- [62] Tharwat A, Gaber T, Hassanien A E, et al. Sift-based arabic sign language recognition system [C]//Afro-European Conference for Industrial Advancement. Cham; Springer International Publishing, 2015; 359-370.
- [63] 程德强, 唐世轩, 冯晨晨, 等. 改进的 HOG-CLBC 的行人检测方法[J]. *光电工程*, 2018, 45(8): 72-80.
- [64] Gupta B, Shukla P, Mittal A. K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion[C]//2016 International Conference on Computer Communication and Informatics (ICCCI). Coimbatore, India; IEEE, 2016; 1-5.
- [65] Dardas N H, Georganas N D. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques [J]. *IEEE Transactions on Instrumentation and Measurement*, 2011, 60(11): 3592-3607.
- [66] Sykora P, Kamencay P, Hudec R. Comparison of SIFT and SURF methods for use on hand gesture recognition based on depth map[J]. *AASRI Procedia*, 2014, 9: 19-24.
- [67] Hartanto R, Susanto A, Santosa P I. Real time static hand gesture recognition system prototype for Indonesian sign language[C]//2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE). Yogyakarta, Indonesia; IEEE, 2014; 1-6.
- [68] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[J] *Advances in Neural Information Processing Systems*, 2014; 568-576.
- [69] Du Y, Wang W, Wang L. Hierarchical recurrent neural network for skeleton based action recognition [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. IEEE, 2015; 1110-1118.
- [70] Liu Y, Zhang L F, Zhang S J. A hand gesture recognition method based on multi-feature fusion and template matching[J]. *Procedia Engineering*, 2012, 29: 1678-1684.
- [71] Pan T Y, Lo L Y, Yeh C W, et al. Real-time sign language recognition in complex background scene based on a hierarchical clustering classification method [C]//2016 IEEE Second International Conference on Multimedia Big Data (BigMM). Taipei, Taiwan,

- China. IEEE, 2016: 64-67.
- [72] Rokade U S, Doye D, Kokare M. Hand gesture recognition using object based key frame selection [C]//2009 International Conference on Digital Image Processing. Bangkok, Thailand: IEEE, 2009: 288-291.
- [73] Wang X Y, Xia M, Cai H W, et al. Hidden-markov-models-based dynamic hand gesture recognition [J]. Mathematical Problems in Engineering, 2012: 1-11.
- [74] Bao J T, Song A G, Guo Y, et al. Dynamic hand gesture recognition based on SURF tracking [C]//2011 international conference on electric information and control engineering. Wuhan, China: IEEE, 2011: 338-341.
- [75] Rekha J, Bhattacharya J, Majumder S. Shape, texture and local movement hand gesture features for Indian sign language recognition [C]//3rd International Conference on Trends in Information Sciences & Computing (TISC2011). Chennai, India: IEEE, 2011: 30-35.
- [76] Zaki M M, Shaheen S I. Sign language recognition using a combination of new vision based features [J]. Pattern Recognition Letters, 2011, 32(4): 572-577.
- [77] Zhang H, Wang Y Q, Deng C. Application of gesture recognition based on simulated annealing BP neural network [C]//Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology. Heilongjiang, China: IEEE, 2011, 1: 178-181.
- [78] Barros P V A, Júnior N T M, Bisneto J M M, et al. An effective dynamic gesture recognition system based on the feature vector reduction for SURF and LCS [M]//Artificial Neural Networks and Machine Learning ICANN 2013. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 412-419.
- [79] Baranwal N, Nandi G C. An efficient gesture based humanoid learning using wavelet descriptor and MFCC techniques [J]. International Journal of Machine Learning and Cybernetics, 2017, 8(4): 1369-1388.
- [80] 郭鑫鹏, 黄元元, 胡作进. 基于关键帧的连续手语语句识别算法研究 [J]. 计算机科学, 2017, 44(z2): 178-183.
- [81] 梁文乐, 黄元元, 胡作进. 基于二级匹配策略的实时动态手语识别 [J]. 计算机科学, 2017, 44(7): 299-303.
- [82] Chen F S, Fu C M, Huang C L. Hand gesture recognition using a real-time tracking method and hidden Markov models [J]. Image and Vision Computing, 2003, 21(8): 745-758.
- [83] Elmezain M, Al-Hamadi A, Appenrodt J, et al. A hidden markov model-based isolated and meaningful hand gesture recognition [J]. International Journal of Electrical, Computer, and Systems Engineering, 2009, 3(3): 156-163.
- [84] Hazmoune S, Bougamouza F, Mazouzi S, et al. A new hybrid framework based on Hidden Markov models and K-nearest neighbors for speech recognition [J]. International Journal of Speech Technology, 2018, 21(3): 689-704.
- [85] Chen Y M, Ding Z Y, Chen Y L, et al. Rapid recognition of dynamic hand gestures using leap motion [C]//2015 IEEE International Conference on Information and Automation. Lijiang, China: IEEE, 2015: 1419-1424.
- [86] Yu M, Liu Y. Research on dynamic gesture recognition based on multi feature fusion [C]//International Conference in Communications, Signal Processing, and Systems. Singapore: Springer International Publishing, 2016: 733-738.
- [87] Likhitha Reddy J, Mallela B, Lavanya Bannaravuri L, et al. Hand gesture recognition using PCA, KNN and SVM [J]. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2019: 547-550.
- [88] ElBadawy M, Elons A S, Shedeed H A, et al. Arabic sign language recognition with 3D convolutional neural networks [C]//2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo: IEEE, 2017: 66-71.
- [89] Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 6299-6308.
- [90] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [91] 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.
- [92] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

(责任编辑:康 锋)