基于视觉的动态手势识别概述

李为斌　刘　佳
（南京信息工程大学自动化学院　江苏 南京 210044）

摘　要　基于计算机视觉的动态手势识别是最直观、最自然的人机交互方式之一。简单回顾了手势识别技术的发展历程；从手势检测和分割、手势跟踪、特征提取和识别算法四个方面对动态手势进行阐述，梳理并归纳了几种代表性的算法，详细探讨了各算法的优缺点；介绍了动态手势识别的应用领域，并展望了今后的发展趋势。

关键词　计算机视觉　人机交互　动态手势　综述

中图分类号　TP391　　**文献标志码**　A　　DOI：10.3969/j.issn.1000-386x.2020.03.032

OVERVIEW OF VISION-BASED DYNAMIC GESTURE RECOGNITION

Li Weibin　Liu Jia
（*School of Automation*, *Nanjing University of Information Science and Technology*, *Nanjing 210044*, *Jiangsu*, *China*）

Abstract　Dynamic gesture recognition based on computer vision is one of the most intuitive and natural ways of human-computer interaction. This paper briefly reviews the development of hand gesture recognition technology. It expounded the dynamic gesture from four aspects: gesture detection and segmentation, gesture tracking, feature extraction and recognition algorithm, summarized several representative algorithms, and discussed the advantages and disadvantages of each algorithm in detail. Finally, we introduced the application of dynamic gesture recognition, and looked forward to its future development.

Keywords　Computer vision　Human-computer interaction　Dynamic gesture　Review

0　引　言

随着计算视觉技术的迅速发展，传统的人机交互技术（如键盘、鼠标等）逐渐无法满足人们日常活动中的应用。许多研究工作致力于开发新的人机交互技术，包括手势识别、人脸识别、人体姿态识别[1]等。其中，手势识别是人机交互中一种十分重要的交互方式，它主要由计算机通过视频输入设备（摄像头等）对用户手势进行检测、跟踪与识别，从而理解人的意图。鲁棒的人手识别系统对手语识别、机器人控制、人机交互等多方面应用的发展起到积极影响[2]。

在手势识别技术发展初期，一些研究[3]使用数据手套等传感器设备来直接检测、获取人手及其各个关节的空间信息，以便于精确地提取并识别出特定的手势。一些学者将光学标识引入识别系统中以提高识别准确性，也取得了较好的效果[4]。数据手套和光学标识等外部设备虽然提高了识别的稳定性和准确性，但在一定程度上掩盖了手势自然的表达方式。为了追求用户更舒适的体验，自然手势识别技术逐渐成为当前研究的重心。然而，由于人手结构的复杂性、手势动作的多义性以及图像获取过程的模糊性，使得手势识别成为了一个极具挑战性的研究课题。

手势一般可划分为静态手势和动态手势两类，本文主要对近年来基于视觉的动态手势识别研究进行综述，着重于从手势分割、手势跟踪、特征提取和识别算法四个方面进行梳理和归纳。基于视觉的手势识别系统基本流程如图 1 所示。

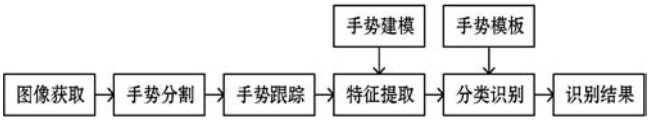

图 1　手势识别流程框图

收稿日期：2019－04－08。国家自然科学基金项目（61773219）。**李为斌**，硕士生，主研领域：增强现实人机交互。**刘佳**，副教授。

1　手势检测和分割

手势分割旨在将图像中手势区域和背景区域分离。手势分割的结果在一定程度上会影响到手势跟踪、特征提取以及手势识别的准确性。目前手势分割技术主要受限于复杂的背景条件,如手势距离视频采集设备不一、光照条件变化频繁等。另外手势运动幅度较大也会影响到手势分割的结果。常用的手势分割方法可分为基于表现特征、基于运动信息、基于深度阈值和混合分割四类。

1.1　基于表现特征的分割方法

对于动态手势识别,主要考虑的是手势的运动而不是手势的姿态,计算复杂度应尽可能低。因此,低层次的特征(如肤色、手形)往往是首选。肤色检测是最常用的分割方法,它利用人手肤色和背景间的差异性以实现分割。肤色检测的关键因素在于颜色空间的选择。RGB 颜色空间[5]在早期手势分割研究中得到了广泛的应用,但其对光照条件敏感,且计算成本相对较高,学者们提出了几种能更好地分离开色度和亮度分量的正交颜色空间,如 HSV[6]、YCrCb[7]、Lab[8]等。在 HSV 中,颜色信息多由色调(Hue)分量和饱和度(Saturation)分量表示,通常不随亮度(Value)分量变化而变化。在 YCrCb 内,Y 表示图像的亮度分量,Cr 和 Cb 表示色度分量。肤色分割方法计算成本低,且对手势的旋转、部分遮挡和尺度变化具有一定的鲁棒性,但它不能有效处理背景中类肤色的区域(如人脸、手臂)。

利用手形特征分割手势是另一种简单有效的方法。手势轮廓能有效表示手的形状,它可以通过边缘检测算子(如 Canny、Sobel、Prewitt、Laplace 等)获得。与肤色检测相比,轮廓特征不受光照变化和类肤色背景的干扰,但其有效性可能受到部分遮挡和视点变化的影响。综上所述,肤色和手形特征在手势分割应用中各具优势,一些研究通过将它们结合起来以提高准确性[9-10]。

1.2　基于运动信息的分割方法

在某些特定的交互场景中,人手是唯一移动的对象,利用运动信息来检测和分割手势的方法是可行的。这一领域主要有背景差分法和光流法[11]两类。背景差分法是指将图像序列中的当前帧与预设的背景图像进行比较来检测运动的手势。该方法检测运动目标速度快,易于实现,其关键在于背景图像的选取。

光流法是一种基于像素的目标运动估计方法,它利用图像序列中的像素强度数据的时域变化和相关性来确定各自像素位置的运动。与背景差分法相比,基于光流的分割算法无需预先保存背景图像,它能在复杂环境下清晰地描述手势的运动,具有较好的鲁棒性,但在手势有遮挡时分割效果不佳,且实时性较差。

1.3　基于深度阈值的分割方法

随着深度传感技术迅速兴起,基于深度相机的手势识别的研究得到了更多的关注。与基于单目视觉的方法相比,其主要优势是,能够有效地捕获具有时空信息的深度图像。深度图像中每个像素点代表了在三维场景内目标对象与摄像头间的真实距离,这也为基于深度阈值的分割方法提供了可靠的理论依据。文献[12]假设人手是距离视频设备最近的对象,根据经验设置了一个深度值在 0.8 米至 1 米的阈值区间,通过将人手置于此区间内以实现检测和分割的目的。文献[13]提出了一种动态阈值分割算法,首先利用 NITE 工具包来检测手掌质心并跟踪,然后根据质心的深度值设置一个动态阈值区间,依据此区间分割出动态手势。上述两种方法虽然简单可行,但其缺陷在于无法精确地切除手臂区域。

针对手臂区域的分割问题,Ren 等[14]建议在手腕处系一条黑色的丝带,这是最简单、稳定的优化方案,也在很多识别系统中得以应用。Chen 等[15]将手掌矩心记作圆心来绘制手掌的最大内圆,在内圆的像素点集合内,相距最远的两个连续像素点被记作手腕的两个端点,最后根据手腕线完成手臂和手势区域的分离。该方法具有较强的适用性和准确性,但计算复杂度略高,有待进一步优化。

1.4　混合分割方法

基于深度阈值的分割方法往往假设人手是最接近相机的对象,显然此方案仅适用于非常有限的场景。Tao 等[16]通过结合动态阈值和 YCrCb 颜色空间的方法进行手势识别和分割,取得了较好的效果。Wen 等[8]设计了一种基于深度信息的双手识别系统,它首先基于 Lab 颜色空间完成手势的粗分割,然后使用 K-means 聚类算法进行手部聚类以进一步分割手势区域。该方法能有效解决因双手互相重叠而导致分割失败的问题。综上所述,通过利用彩色图像和深度图像的彼此优势,设计一种具有更高精度的分割算法是值得探究的。

2　手势跟踪

手势跟踪在动态手势的识别中起着关键的作用,其主要目的是确定视频序列中人手的运动轨迹。目前

手势跟踪算法主要分为三类:基于生成模型、基于判别模型和基于混合模型的跟踪方法。

2.1 基于生成模型的跟踪方法

生成模型方法通过在线学习方式建立目标模型,然后使用模型搜索重建误差最小的图像区域,完成目标定位。使用生成模型跟踪手势时,首先需要对当前帧中的手势区域进行建模,然后在下一帧中寻找与该模型最相似的区域就是预测位置。生成模型以统计学和贝叶斯为理论基础,一般涉及的算法有连续自适应均值漂移(Camshift)算法、卡尔曼滤波(Kalman Filter)、粒子滤波(Particle Filter)等。

2.1.1 Camshift 算法

Camshift 算法是一种以颜色概率分布图为基础的目标跟踪算法。文献[17]使用 Camshift 算法和 HSV 颜色空间来跟踪手势,其具体跟踪过程是:首先初始化一个搜索窗口,并计算色度分量概率分布;然后使用 meanshift 算法调整搜索窗口的中心位置和大小;随后再将当前帧迭代得到的搜索窗口用作下一帧搜索窗口的初始值,如此反复直至收敛。该方法具有低复杂性和较高的实用性,但在更复杂的场景中它的鲁棒性不稳定。例如包含手臂区域的图像,由于手臂具有和手势相似的颜色概率分布,会导致搜索窗口漂移。为了解决这一问题,Yang 等[18]寻找并计算了一个手部深度优化概率,然后将其与 Camshift 算法进行结合,通过实时更新直方图以确保手部跟踪的准确性。

2.1.2 卡尔曼滤波

卡尔曼滤波是一种能够对目标位置进行有效预测的算法。当目标的运动满足高斯模型时,对其运动状态进行预测,并与观察模型进行对比,根据误差来更新运动目标的状态。Park 等[19]将状态方程设为三个维度,在各个维度上使用三维向量作为在轴上的速度,并应用卡尔曼滤波算法逐帧预测手掌质心及其周围参考点。该方法能适应人手运动速度的连续变化,跟踪效果较好。Yeo 等[20]利用卡尔曼滤波以消除不稳定的噪声来估计和细化最优位置。该方法能够使手部的位置或轨迹变得更加平滑且跳动更小,更适合控制应用程序。

基于卡尔曼滤波的跟踪算法具有实时性好、无偏和最优的特点,能有效地解决手势部分遮挡和跟踪丢失的问题,但缺点是仅适合于线性系统,适用范围小。为了解决这一问题,学者们提出了粒子滤波算法。

2.1.3 粒子滤波

粒子滤波算法源于蒙特卡洛方法,其核心思想是利用随机搜索以获得目标的后验分布的估计。粒子滤波往往将跟踪问题视为状态估计问题。目前粒子滤波算法大多被用于跟踪手掌质心或手指的空间坐标,在跟踪过程中,通过使用一组随机采样粒子对目标的位置进行建模。文献[21]设计的跟踪算法可分为基于图像的二维位置跟踪和一维深度跟踪两部分。通过将尽可能多的处理放在二维空间中以实现更快的跟踪,指尖点三维坐标则通过合并可用的深度信息来获得。相较于文献[22]中三维跟踪方法,该算法的跟踪精度和稳定性更优异。

在嘈杂环境中,粒子滤波的鲁棒性优于卡尔曼滤波,但其也存在粒子退化现象。许多研究者尝试将粒子滤波与其他算法结合起来,解决样本退化问题[23]。

2.2 基于判别模型的跟踪方法

判别模型方法通常将手势跟踪视作二分类问题,它同时提取手势和背景信息用来训练分类器,将手势从图像序列背景中分离出来,从而得到当前帧的手势位置。目前判别模型在手势跟踪中应用较多。

Keskin 等[24]提出了一种利用多层随机决策树预测手部关节位置的方法,其主要思想是通过聚类将手势数据集划分为更简单的子集,然后为每个子集训练单独的手势估计器。

文献[25]提出了一种用于无标记复杂关节实时连续手势恢复的跟踪算法,主要包括以下阶段:用于图像分割的随机决策树分类器、用于标记数据集生成的鲁棒方法、用于特征提取的卷积神经网络以及用于稳定实时手势恢复的逆运动学算法。

文献[26]提出了一种基于多视图卷积神经网络的三维回归方法,它首先将深度图像点云投影到三个正交平面上,然后为每个投影图像生成一组手关节的热图,最后将三组热图与预先学习的手势先验相融合以得到三维关节位置。

2.3 基于混合模型的跟踪方法

判别模型方法一般具有较好的稳健性,但其缺乏模型拟合中固有的准确性。生成模型方法在初始化时可能会依赖于前一帧,这使得从错误中恢复变得更具挑战性。混合模型通过结合判别模型方法和生成模型方法来提高相邻帧的模型拟合的鲁棒性。

文献[27]提出使用一种混合方法跟踪人手的关节三维运动,其中,判别模型方法采用基于部分的姿势检索算法,通过检测深度图像中的指尖来估计完整或部分手势;生成模型方法则使用颜色信息和特定手势模型来预测最佳的手势位置。

文献[28]设计了一种判别估计器用以预测手势的分布,并从该分布中提取不同的手势假设;然后通过

结合粒子群优化算法和遗传算法来检测三维手势模型和观察图像之间的重建误差;最后将误差最小的结果输出。该方法具有较好的鲁棒性,且在目标跟踪丢失后能确保快速恢复跟踪。

Mueller 等[29]使用两个卷积神经网络实时定位和估计人手的 3D 关节位置。该方法在 SynthHands 数据集和 EgoDexter 基准数据集上测试,结果表明它在目标遮挡、背景变化等环境中能实现低误差。

表 1 列举了近几年主流的动态手势的跟踪算法。

表 1　动态手势跟踪算法及跟踪对象

方法	作者	年份	数据集	跟踪对象
Camshift	Nadgeri 等[17]	2010	ASL	颜色直方图
卡尔曼滤波	Park 等[19]	2012	/	手掌质心
	Yeo 等[20]	2015	/	指尖点
粒子滤波	Alamsyah 等[21]	2013	/	指尖点
判别模型	Keskin 等[24]	2012	ASL	骨骼关节点
	Tompson 等[25]	2014	/	三维关节点
	Ge 等[26]	2016	Cross-dataset	
混合模型	Sridhar 等[27]	2013	Fingerwave	指尖点
	Sharp 等[28]	2015	Synthetic/ FingerPaint	三维手模型
	Mueller 等[29]	2017	SynthHands/ EgoDexter	三维关节点

3　特征提取

特征提取是将输入样本中感兴趣区域转换为特征向量的集合。在手势识别中,提取的特征应包含待测手势的相关信息,并以紧凑的形式表示,作为该手势的标识,将其与其他手势区分开来。结合现有的多数文献来看,动态手势特征主要包括全局特征、局部特征和融合特征。

3.1　全局特征

全局特征通常用于表示图像序列的整体属性,可以通过运动历史图、运动能量图、时空形状等加以描述。Ren 等[30]在陆地移动距离算法(Earth Mover's Distance,EMD)基础上设计了一种手指陆地移动距离(FEMD)算法。相比于其他距离度量方法,该算法仅考虑手指的外轮廓而不包括手掌区域,它通过一组记录着每个轮廓顶点到手掌中心点之间相对距离的时间序列曲线来表示手部轮廓,其中每个手指对应于曲线的某一段。文献[31]提出超像素陆地移动距离(SPEMD)算法来度量不同手势间的差异性,其中手形和纹理信息以超像素的形式进行描述。该方法对手势的旋转、平移和缩放都有一定的鲁棒性。文献[32]通过组合指尖到手掌质心的距离、手形边缘曲率特征以提高识别准确率,但也导致其特征维数较高,实时性受到影响。

3.2　局部特征

这类特征通过提取图像序列中能有效地表征动态手势的局部特征点,再对这些特征点的各种属性进行统计建模,实现对动态手势的描述。文献[33]提出了 HON4G(Histogram of Oriented 4D Normals)描述符来描述深度序列,该描述符使用了时间、深度和空间坐标的 4D 空间中的表面法线方向直方图来捕获运动和几何数据。在 MSR Gesture 3D 数据集上测试,实验结果验证了该描述子的优越性。

文献[34]提出了一种改进的三维局部稀疏运动尺度不变特征变换(3D SMoSIFT)特征描述子。首先为每个灰度帧和深度帧构建金字塔作为尺度空间,然后使用角点检测算法和稀疏光流法快速检测和跟踪尺度空间中动态手势周围的鲁棒关键点,再基于这些关键点构建三维梯度和运动空间,最后分别在两个空间上计算 SIFT 描述符。在 Chalearn 手势数据集的评估结果表明,该特征能明显提高识别精度。

文献[35]提出了三维面片直方图(Histogram of 3D facets,H3DF)特征描述符来有效捕获和编码三维形状信息。首先将三维云点及其周围点定义为三维面,其中每个面均由小平面建模,然后使用空间中心池策略来组织小平面的集合以描述感兴趣区域,形成最终的 H3DF 描述符。

3.3　融合特征

对于某些特定手势,单一特征通常能给出优异的识别结果,但其适用范围小,且稳定性一般。为了提高稳定性和适用性,学者们考虑了将多个特征相融合的方案。Monnier 等[36]首先从每个视频帧中提取归一化骨架关节位置和 HOG 特征,以产生相应的多维描述符序列,然后对这些特征描述符进行插值以确定其固定宽度,最后合并该序列中的特征形成最终描述符。Liang 等[37]提出了一种融合全局特征运动能量图和局部特征梯度直方图来识别手势的方法。首先将归一化的运动能量矢量分割为一组分段,并利用相应的帧索引将深度视频分割成一组网格;然后使用 HOG 特征从深度序列中提取手势的局部纹理信息和运动信息。

4　手势识别

识别算法是手势识别研究中最后但最重要的一

步,当前主流的识别算法包括基于模板匹配的方法、基于数据分类的方法和基于深度学习的方法。动态时间规整算法(Dynamic Time Warping,DTW)是最常用的模板匹配方法;数据分类方法中应用较为广泛的有支持向量机(Support Vector Machine,SVM)、人工神经网络(Artificial Neural Network,ANN)、隐马尔可夫模型(Hidden Markov Model,HMM)等;深度学习方法有卷积神经网络(Convolutional Neural Network,CNN)。

4.1 动态时间规整

DTW 算法是一种基于动态规划思想的非线性规整方法,它利用规整函数描述待测模板和参考模板的时间对应关系,以求解出两个时间序列的相似度。使用 DTW 算法识别手势时,需预先记录一系列参考模板,然后匹配并计算待测手势和参考模板间的相似度,将其中相似度最高的模板手势记作识别结果。DTW 算法具有训练样本需求少、精度高等特点。Plouffe 等[38]利用 DTW 算法识别动态手势,识别精度达到 96.25%。

然而,DTW 算法受限于计算复杂度高、稳定性差等问题,尤其是在复杂手势、训练样本数量较多等情况下。因此,涌现了很多优化原始 DTW 算法的研究,如 Ruan 等[39]针对搜索路径和匹配过程两个方面进行改进,通过全局约束算法将搜索路径斜率约束至[1/2,2]区间内,再利用失真阈值算法实时地控制待测手势与参考手势匹配的过程。经测试,该方法能显著地减少计算复杂度,提高识别效率。

4.2 支持向量机

支持向量机是一种基于统计学习理论的有监督的学习模型,其基本原理是利用在样本空间或特征空间求出最优超平面,使小同类样本集与超平面之间的距离最大。Huang 等[40]提出了一种基于 Gabor 滤波器和 SVM 分类器的手势识别方法。该方法通过使用自适应肤色模型以克服光照变化对手部检测的干扰,并采用基于 Gabor 滤波器的手势角度估计和校正方法实现对手势变化的鲁棒。Dardas 等[41]利用 SIFT 描述符提取图像中的关键点,并使用 K-means 聚类算法和矢量量化算法将这些关键点映射为直方图向量,将该直方图作为多类 SVM 的输入向量以构建手势分类器。在可变尺度、方向和光照条件以及复杂背景等情况,该方法都能达到令人满意的实时性能,分类精度达到 96.23%。

4.3 人工神经网络

人工神经网络是受生物神经网络启发的一类信息处理模型,它由许多相互连接的并行神经元组成。人工神经网络的模型种类繁多,对于不同的需求可衍生出不同的结构,目前误差反向传播神经网络(BPNN)应用较为广泛。Hasan 等[42]提取了手部轮廓和复数矩来表征手势,并使用 BPNN 对两类特征进行训练,识别正确率分别达到 70.83% 和 86.38%。文献[43]使用量子粒子群算法对 BPNN 进行优化,很大程度上解决了其收敛速度慢、局部极小化导致网络训练失败等问题,提高了识别效率和稳定性。

Tusor 等[44]利用模糊神经网络(Fuzzy Neural Network,FNN)建立了模糊手势模型,并设置了 14 个取值介于“大”、“中”和“小”的模糊特征值用于描述和区分不同的手势。文献[45]提出一种自成长和自组织神经气(Self-Growing and Self-Organized Neural Gas,SGONG)网络,该网络通过拟合人手形状提取出有效的特征。该特征对手势的尺度和旋转变化不敏感。此外,该网络的收敛速度也比其他网络更快。

4.4 隐马尔可夫模型

隐马尔可夫模型是一种典型的概率统计模型,通常用于描述具有隐藏状态的马尔可夫过程。HMM 能有效地捕获时序中的相关性。由于手势动作是一个时间序列,因此 HMM 在手势识别领域中得到了广泛的应用。使用 HMM 识别动态手势时,需预先为每个手势训练一个单独 HMM 模型,然后求解每个 HMM 模型产生待测手势的概率,概率最大的 HMM 模型对应的手势就是识别结果。

文献[46]选用字母轨迹的运动角度作为手势的特征向量,并利用左右拓扑结构的 HMM 对 23 个预设字母轨迹进行识别,识别准确率达到 91.6%。一些研究将 HMM 与其他分类器相结合,如文献[47]利用 AdaBoost 分类器检测用户的手部,再使用粒子滤波进行跟踪,最后基于 HMM 完成手势识别。该方法在识别精度方面有着显著提升,但计算复杂度极大,无法满足实时性的要求。

4.5 深度学习

深度学习是近年来最受学者们关注的研究领域之一,如今已成为大数据和人工智能相结合最成功的典型。深度学习是一种非监督学习,它不仅能自动提取图像中的特征,还能够自动学习更高层次的特征,这克服了人工提取特征的主观性和局限性。根据时间维度处理方式不同,可将基于深度学习的手势识别方法分为卷积神经网络、三维卷积神经网络(3DCNN)和序列模型。

4.5.1 CNN

CNN 又称为二维卷积神经网络(2DCNN),是深度

学习中最基础的网络结构之一,通常由卷积层、池化层和全连接层组成。目前 CNN 已经在人脸识别、目标检测等领域取得了极大的成功,在手势识别领域也迅速兴起。Neverova 等[48]开发了一种手部姿势估计的深度学习模型,该模型关键之处是通过将人手分割成不同的部分来将结构信息编码至训练目标中。Oyedotun 和 Khashman[49]使用 CNN 和堆叠去噪自动编码器来识别 24 种美国手语手势,在公共数据集上取得 91.33% 的识别率。2DCNN 主要提取图像序列中的空间特征,很少涉及到时间维度信息,因此多用于静态手势的分类应用中。

4.5.2　3DCNN

与 2DCNN 相比,3DCNN 可以保持时间结构的同时,还能有效地捕获时空维度上的判别特征。目前一些 3DCNN 已经被提出并应用于手势识别中,其中值得关注的是,Molchanov 等[50]提出了一种基于深度和强度数据的驾驶员手势识别的 3DCNN,该方法为归一化的深度和强度数据分别训练了两个独立的子网络,通过将两个网络输出结果进行融合以实现最终预测。此外,为了避免训练样本集相似度较高而导致的过拟合问题,还采用了时空数据增强技术来改变手势的输入体积。在 VIVA 数据集测试下,该系统的分类率达到了 77.5%。文献[51]设计了一种以多种数据类型为输入的 3DCNN,该结构通过对相邻视频帧进行卷积和子采样,以实现颜色信息、深度信息以及人体骨骼信息的融合。

4.5.3　序列模型

近几年,一些学者采用将 CNN(或 3DCNN)与序列模型相结合的方式来识别动态手势,其基本原理是利用序列模型的结构特性来处理时间维度信息。循环神经网络(Recurrent Neural Network,RNN)是最常用的序列模型之一,它能利用隐藏层中的循环结构来处理时间数据。Molchanov 等[52]利用循环机制对文献[50]中的 3DCNN 进行了扩展,提出了一种循环三维卷积神经网络(R3DCNN)。该网络首先利用 3DCNN 提取视频序列中局部时空特征,然后将这些特征输入到 RNN 中,最后在 Softmax 层中预测待测手势概率。

长短期记忆网络(Long Short-Term Memory,LSTM)是一类特殊的 RNN,它能解决 RNN 内存不足的问题,多被用作 RNN 的隐藏层。为了利用手部骨骼关键点的三维数据序列来解决手势识别问题,文献[53]介绍了一种结合 3DCNN 和 LSTM 循环网络的动态手势识别方法。其中,3DCNN 主要用于检测与骨骼关节位置相关的空间特征,LSTM 循环网络则考虑序列的时间演化。该方法在六种基准数据集上进行了实验,结果表明其在小数据集的性能最佳。

综上所述,目前常用的手势识别方法都有其自身的优缺点,为了便于选择和应用,有必要对其优缺点进行比较和总结。表 2 对不同的手势识别方法进行了详细的对比。

表 2　常用的手势识别算法比较

识别方法	优点	缺点
DTW	训练样本需求少,识别精度高	计算复杂度高,稳定性较差
SVM	可有效解决小样本、高维度、非线性问题	训练样本数量较大时,效率较低
ANN	自适应性、抗干扰性较强,能有效解决非线性问题	对时间序列的建模能力较差,网络扩充性不佳
HMM	能有效地捕获时序中的相关性	训练过程比较复杂,训练时间长,计算量较大
CNN	无需人为提取特征,权值共享	需要大量的训练数据并且计算成本高

5　结　语

视觉手势识别技术的发展给人机交互带来了一种全新的方式,用户与计算机之间只需徒手便能完成交互功能,这让用户成功摆脱了数据手套和光学标识等外部设备的束缚,从而提高了人机交互的灵活性和自然性。从最初的手语识别应用,到后来体感游戏的创新,手势识别技术正日趋成熟,时至今日手势识别与其他领域也有着密切的结合,主要包括医疗辅助、智能汽车和机器人。

1)医疗系统。在传统外科手术中,医生使用键盘或触摸屏等设备操控计算机以查阅病例资料,该方式不能保证环境无菌,而手势识别系统则可以帮助医生直接使用手势控制计算机,极大提高了手术安全性和效率。伦敦圣托马斯医院和多伦多森尼布鲁克医院[54-55]是该领域的先驱者,他们已成功将手势识别技术应用于医疗实践中。

2)车载系统。自 2013 年谷歌宣布研发利用手势控制汽车的技术至今,手势识别融合车载系统的方案得到了越来越多厂商的关注。2015 年,宝马推出了首款搭载着手势控制功能的智能车,它支持通过手势实现接听电话、控制音量和缩放导航地图等功能。此外,奥迪、福特、谷歌等车企也接连展示过同类型的概念车。

3)机器人控制。对于新兴的社交机器人而言,能够通过理解用户手势来实现人机互动是极其重要的。Fischinger 等[56]发布了一款护理机器人 Hobbit,为了便

于老年人与机器人的日常交互,他新增了一个 3D 手势界面。深圳优必选科技公司推出的多款智能机器人都支持客户利用自然手势完成资料查阅、音乐播放和日程提醒等功能。

尽管已经取得了这么多令人欣喜的进展,但手势识别研究仍然是一个较为新兴的领域,今后其发展将呈现以下趋势:

1)智能学习算法的研究。目前,大多数手势识别系统仍需经历模型训练的阶段,这一过程通常需要大量的训练样本,同时耗费大量的时间。智能学习算法旨在让计算机通过"在线学习"的方式对用户执行的手势动作示范进行学习和记忆,而无需额外的训练操作,从而显著地提高识别效率。

2)多个深度摄像头的融合。不同于普通摄像头,Kinect 和 Leap motion 等新型的深度摄像头能直接获取丰富的手势信息,从而用于精确的手势识别。此前大多数研究者通过单独使用 Kinect 或 Leap motion 进行研究,并取得了较好的结果。通过将多个深度摄像头进行融合以设计出具有更高精度与稳定性的识别系统,是该领域中一个值得研究的潜在方向。

3)现有的大多数手势识别系统无法同时识别多个用户执行的手势。与识别单一用户手势不同,这类问题需要考虑到不同手势特征的提取以及手势模型的构建。此外,由于个人习惯、熟练度和时间的不同,不同用户执行特定的手势也是不同的。因此,开发一种考虑这些个体差异的多用户手势识别技术,是该领域的另一个未来研究方向。

参 考 文 献

[1] Konar A, Saha S. Gesture recognition: Principles, techniques and applications[M]. Springer, 2017.

[2] Chaudhary A. Robust hand gesture recognition for robotic hand control[M]. Springer, 2018.

[3] Premaratne P, Nguyen Q, Premaratne M. Human computer interaction using hand gestures[C]//International Conference on Intelligent Computing. Springer, Berlin, Heidelberg, 2010: 381 - 386.

[4] Zaman M, Rahman S, Rafique T, et al. Hand gesture recognition using color Markers[C]//International Conference on Hybrid Intelligent Systems. Springer, Cham, 2016: 1 - 10.

[5] 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.

[6] Hasan M M, Misra P K. Gesture recognition using modified HSV segmentation[C]//International Conference on Communication Systems and Network Technologies. IEEE, 2011: 328 - 332.

[7] Shaik K B, Ganesan P, Kalist V, et al. Comparative study of skin color detection and segmentation in HSV and YCbCr color space[J]. Procedia Computer Science, 2015, 57: 41 - 48.

[8] Wen Y, Hu C, Yu G, et al. A robust method of detecting hand gestures using depth sensors[C]//IEEE International Workshop on Haptic Audio Visual Environments and Games. IEEE, 2012: 72 - 77.

[9] Zhou Y, Jiang G, Xu G, et al. Hand gesture recognition based on the parallel edge finger feature and angular projection[C]//Asian Conference on Computer Vision. Springer, Cham, 2014.

[10] Pisharady P K, Vadakkepat P, Loh A P. Attention based detection and recognition of hand postures against complex backgrounds[J]. International Journal of Computer Vision, 2013, 101(3): 403 - 419.

[11] Hackenberg G, McCall R, Broll W. Lightweight palm and finger tracking for real-time 3D gesture control[C]//IEEE Virtual Reality Conference. IEEE, 2011: 19 - 26.

[12] Sorce S, Gentile V, Gentile A. Real-time hand pose recognition based on a neural network using microsoft kinect[C]//Eighth International Conference on Broadband and Wireless Computing, Communication and Applications. IEEE, 2013: 344 - 350.

[13] Raheja J L, Chaudhary A, Singal K. Tracking of fingertips and centers of palm using kinect[C]//Third International Conference on Computational Intelligence, Modelling & Simulation. IEEE, 2011: 248 - 252.

[14] Ren Z, Yuan J, Zhang Z. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera[C]//Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011: 1093 - 1096.

[15] Chen Z H, Kim J T, Liang J N, et al. Real-time hand gesture recognition using finger segmentation[J]. The Scientific World Journal, 2014, 2014: 267872.

[16] Tao H Y, Yu Y L. Finger tracking and gesture interaction with Kinect[C]//Proceedings of the IEEE 12th international conference on computer and information(CIT). 2012: 214 - 218.

[17] Nadgeri S M, Sawarkar S D, Gawande A D. Hand gesture recognition using CAMSHIFT algorithm[C]//3rd International Conference on Emerging Trends in Engineering and Technology. IEEE, 2010: 37 - 41.

[18] Yang C, Jang Y, Beh J, et al. Gesture recognition using depth-based hand tracking for contactless controller application[C]//IEEE International Conference on Consumer Electronics(ICCE). IEEE, 2012: 297 - 298.

[19] Park S, Yu S, Kim J, et al. 3D hand tracking using Kalman filter in depth space[J]. EURASIP Journal on Advances in Signal Processing, 2012(1): 36 - 54.

[20] Yeo H S, Lee B G, Lim H. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware[J]. Multimedia Tools and Applications, 2015, 74(8): 2687 – 2715.

[21] Alamsyah D, Fanany M I. Particle filter for 3D fingertips tracking from color and depth images with occlusion handling [C]//International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2013: 445 – 449.

[22] Liang H, Yuan J, Thalmann D. 3D fingertip and palm tracking in depth image sequences[C]//Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012: 785 – 788.

[23] Li T, Sun S, Sattar T P, et al. Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches[J]. Expert Systems with applications, 2014, 41(8): 3944 – 3954.

[24] Keskin C, Kıraç F, Kara Y E, et al. Hand pose estimation and hand shape classification using multi-layered randomized decision forests[C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012: 852 – 863.

[25] Tompson J, Stein M, Lecun Y, et al. Real-time continuous pose recovery of human hands using convolutional networks [J]. ACM Transactions on Graphics, 2014, 33(5): 1 – 10.

[26] Ge L, Liang H, Yuan J, et al. Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3593 – 3601.

[27] Sridhar S, Oulasvirta A, Theobalt C. Interactive markerless articulated hand motion tracking using RGB and depth data [C]//Proceedings of the IEEE international conference on computer vision. 2013: 2456 – 2463.

[28] Sharp T, Keskin C, Robertson D, et al. Accurate, robust, and flexible real-time hand tracking[C]//Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015: 3633 – 3642.

[29] Mueller F, Mehta D, Sotnychenko O, et al. Real-time hand tracking under occlusion from an egocentric rgb-d sensor [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1284 – 1293.

[30] Ren Z, Yuan J, Meng J, et al. Robust part-based hand gesture recognition using kinect sensor[J]. IEEE transactions on multimedia, 2013, 15(5): 1110 – 1120.

[31] Wang C, Liu Z, Chan S C. Superpixel-based hand gesture recognition with kinect depth camera[J]. IEEE transactions on multimedia, 2015, 17(1): 29 – 39.

[32] Dominio F, Donadeo M, Zanuttigh P. Combining multiple depth-based descriptors for hand gesture recognition [J]. Pattern Recognition Letters, 2014, 50: 101 – 111.

[33] Oreifej O, Liu Z. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 716 – 723.

[34] Wan J, Ruan Q, Li W, et al. 3D SMoSIFT: three-dimensional sparse motion scale invariant feature transform for activity recognition from RGB-D videos[J]. Journal of Electronic Imaging, 2014, 23(2): 023017.

[35] Zhang C, Tian Y. Histogram of 3D facets: A depth descriptor for human action and hand gesture recognition[J]. Computer Vision and Image Understanding, 2015, 139: 29 – 39.

[36] Monnier C, German S, Ost A. A multi-scale boosted detector for efficient and robust gesture recognition [C]//Workshop at the European Conference on Computer Vision. Springer, Cham, 2014: 491 – 502.

[37] Liang C, Qi L, Chen E, et al. Depth-based action recognition using multiscale sub-actions depth motion maps and local auto-correlation of space-time gradients[C]//IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2016: 1 – 7.

[38] Plouffe G, Cretu A M. Static and dynamic hand gesture recognition in depth data using dynamic time warping[J]. IEEE transactions on instrumentation and measurement, 2016, 65(2): 305 – 316.

[39] Ruan X, Tian C. Dynamic gesture recognition based on improved DTW algorithm[C]//IEEE International Conference on Mechatronics & Automation. IEEE, 2015: 2134 – 2138.

[40] Huang D Y, Hu W C, Chang S H. Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination [J]. Expert Systems with Applications, 2011, 38(5): 6031 – 6042.

[41] 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.

[42] Hasan H, Abdul-Kareem S. RETRACTED ARTICLE: Static hand gesture recognition using neural networks[J]. Artificial Intelligence Review, 2014, 41(2): 147 – 181.

[43] 杨志奇,孙罡. 基于量子粒子群优化反向传播神经网络的手势识别[J]. 计算机应用,2014,34(A01):137 – 140.

[44] Tusor B, Varkonyi-Koczy A R. Circular fuzzy neural network based hand gesture and posture modeling[C]//Instrumentation & Measurement Technology Conference. IEEE, 2010: 815 – 820.

[45] Stergiopoulou E, Papamarkos N. Hand gesture recognition using a neural network shape fitting technique[J]. Engineering Applications of Artificial Intelligence, 2009, 22(8): 1141 – 1158.

(下转第 208 页)

文档长度的数据集上的表现都优于目前最先进的文摘系统。

参 考 文 献

[1] 王萌,唐新来,何婷婷. 一种文本分割技术的多文档文摘方法研究[J]. 计算机应用与软件,2014,31(9):40 - 44.

[2] 胡迁,黄青松,刘利军,等. 基于自动文摘的答案生成方法研究[J]. 计算机应用与软件,2018,35(12):187 - 192,307.

[3] Wan X, Yang J. Multi-document summarization using cluster-based link analysis[C]//Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM,2008:299 - 306.

[4] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014: 3104 - 3112.

[5] Rush A M, Chopra S, Weston J. A neural attention model for abstractive sentence summarization[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 379 - 389.

[6] Chopra S, Auli M, Rush A M. Abstractive sentence summarization with attentive recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,2016: 93 - 98.

[7] Nallapati R, Zhou B, Dos Santos C N, et al. Abstractive text summarization using sequence-to-sequence RNNs and beyond[C]//Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning. 2016:280 - 290.

[8] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]//29th AAAI Conference on Artificial Intelligence. 2015.

[9] Ma S, Sun X. A semantic relevance based neural network for text summarization and text simplification[EB]. arXiv preprint arXiv:1710.02318, 2017.

[10] Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units[EB]. arXiv preprint arXiv:1508.07909, 2015.

[11] Mihalcea R, Tarau P. Textrank: Bringing order into text[C]//Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. 2004.

[12] Greene D, Cunningham P. Practical solutions to the problem of diagonal dominance in kernel document clustering[C]// Proceedings of the 23rd International Conference on Machine Learning. Association for Computing Machinery, 2006: 377 - 384.

[13] Hulth A. Improved automatic keyword extraction given more linguistic knowledge[C]//Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2003: 216 - 223.

[14] See A, Liu P J, Manning C D. Get to the point: Summarization with pointer-generator networks[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 1073 - 1086.

[15] Lin C Y. ROUGE: A package for automatic evaluation of summaries[C]//Proceedings of the Workshop on Text Summarization Branches Out(WAS 2004). 2004.

(上接第 197 页)

[46] Premaratne P, Yang S, Vial P, et al. Centroid tracking based dynamic hand gesture recognition using discrete Hidden Markov Models[J]. Neurocomputing, 2017, 228: 79 - 83.

[47] Wang X, Xia M, Cai H, et al. Hidden-Markov-Models-based dynamic hand gesture recognition[J]. Mathematical Problems in Engineering, 2012, 2012:986134.

[48] Neverova N, Wolf C, Taylor G W, et al. Hand segmentation with structured convolutional learning[C]//Asian Conference on Computer Vision. Springer, Cham, 2014:687 - 702.

[49] Oyedotun O K, Khashman A. Deep learning in vision-based static hand gesture recognition[J]. Neural Computing and Applications, 2017, 28(12): 3941 - 3951.

[50] 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. 2015: 1 - 7.

[51] Huang J, Zhou W, Li H, et al. Sign language recognition using 3d convolutional neural networks[C]//2015 IEEE international conference on multimedia and expo(ICME). IEEE, 2015: 1 - 6.

[52] Molchanov P, Yang X, Gupta S, et al. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4207 - 4215.

[53] 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.

[54] O'Hara K, Gonzalez G, Sellen A, et al. Touchless interaction in surgery[J]. Communications of the ACM, 2014, 57(1): 70 - 77.

[55] Strickland M, Tremaine J, Brigley G, et al. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field[J]. Canadian Journal of Surgery, 2013, 56(3): E1 - E6.

[56] Fischinger D, Einramhof P, Papoutsakis K, et al. Hobbit, a care robot supporting independent living at home: First prototype and lessons learned[J]. Robotics and Autonomous Systems, 2016, 75: 60 - 78.