

文章编号:1007-5321(2020)05-0091-07

DOI:10.13190/j.jbupt.2020-056

基于视频数据特性的动态手势识别

谢晓燕¹, 赵欢¹, 蒋林²

(1. 西安邮电大学 计算机学院, 西安 710121; 2. 西安科技大学 集成电路设计实验室, 西安 710054)

摘要: 针对现有动态手势识别方法环境适应性低、计算复杂的问题,提出了一种基于视频数据特性的动态手势识别方法. 使用基于密度的聚类算法 DBSCAN 直接从视频编码数据中的运动矢量提取出运动趋势特征,再通过随机森林分类运动趋势,结合卷积神经网络(CNN)提取的手型特征识别动态手势. 实验结果表明,该方法对剑桥大学和美国西北大学数据集中动态手势的平均识别率分别达到 94.22% 和 94.48%,并且与 CNN 结合长短期记忆网络的识别方法相比,手势识别时间减少了 85%. 在背景图像复杂且光照条件不足时,该方法仍然能够维持较高的识别率,表现出较好的鲁棒性.

关键词: 动态手势识别; 运动矢量; DBSCAN; 随机森林; 卷积神经网络

中图分类号: TP391.4

文献标志码: A

Dynamic Gesture Recognition Based on Characteristics of Encoded Video Data

XIE Xiao-yan¹, ZHAO Huan¹, JIANG Lin²

(1. School of Computer, Xi'an University of Posts & Telecommunications, Xi'an 710121, China;

2. Integrated Circuit Design Laboratory, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: Aiming at the challenges to scene adaptability and computational complexity of dynamic gesture recognition, a method based on characteristics of encoded video data is proposed. Firstly, density-based spatial clustering of applications with noise is used to extract motion trend features from motion vectors. Then, the motion trends are classified by random forest. Finally, combined by the hand shape features extracted by convolutional neural network(CNN), the dynamic gesture is recognized. The experiment shows that the proposed method has an average recognition rate of 94.22% and 94.48% respective for university of Cambridge and Northwestern University hand gesture data sets. Compared with the scheme combine of CNN and long short-term memory, the gesture recognition time is reduced by 85%. It can still maintain a higher recognition rate for the complex background with insufficient illumination, represents a higher robustness.

Key words: dynamic gesture recognition; motion vector; density-based spatial clustering of applications with noise; random forest; convolutional neural network

动态手势作为一种重要的人机交互方式,受到越来越多的关注. 近年来,动态手势识别在体感游

收稿日期: 2020-06-07

基金项目: 国家自然科学基金项目(61834005, 61772417, 61602377); 陕西省国际科技合作计划项目(2018KW-006); 榆林市科技计划项目(2019-133)

作者简介: 谢晓燕(1972—), 女, 教授, 硕士生导师.

通信作者: 赵欢(1995—), 男, 硕士生, E-mail:750746730@qq.com.

戏、手语识别、辅助驾驶、医疗器械以及智能家电控制等领域的应用越来越广泛。

动态手势识别一直是计算机视觉领域的研究热点,专家学者在相关领域开展了广泛的研究,取得了大量的成果。最早的动态手势识别采用数据手套^[1]辅助完成,虽然能够获取较丰富的手势信息,但由于穿戴繁琐、价格昂贵,人们又开始采用机器视觉技术^[2],以提高手势识别的自然性和舒适性。机器视觉通常利用肤色、形状等视觉特征进行目标检测,在光照不足和复杂背景条件下很难正确地分割出手势。由于手势外观有可能在多帧内大幅变化,这时数据特征的多样性会导致跟踪困难。此外,动态手势的分类还需要兼顾时序信息。Barros 等^[3]提出了一种凸性方法用于提取特征,使用动态时间规整(DTW, dynamic time warping)分析提取的特征序列,得到手势分类结果。由于 DTW 计算耗时,所以实时性较差。Tang 等^[4]使用图像熵和密度聚类提取手势关键帧,再通过加速稳健特征(SURF, speeded up robust features)和三维尺度不变特征变换(SIFT 3D, scale-invariant feature transform 3D),结合特征袋从关键帧中提取外观和运动特征,最后使用支持向量机(SVM, support vector machine)对融合特征分类得到手势类别,时空特征提取复杂且计算耗时。Patil 等^[5]结合傅里叶描述子和几何特征组成的空间特征和霍夫变换(HT, Hough transform)得到的时间特征形成手势特征向量,并使用多层感知机(MLP, multilayer perceptron)进行分类。该方法基于肤色检测虽然提升了算法对光照的鲁棒性,但增加了对手势前景的检测,加大了算法的复杂度。Gatto 等^[6]提出正交 Hankel 子空间法(OHSM, orthogonal Hankel subspace method),通过紧凑轨迹矩阵表示手势的时空特征,以 k 均值聚类选择样本缓解了计算量的问题,但准确率难以保证。由于传统机器视觉方法需要大量的先验知识来提取特征且准确率难以保证,人们又尝试采用深度学习的方法来自动学习特征,并提高识别率。

近年来,随着人工智能技术的飞速发展,卷积神经网络(CNN, convolutional neural network)在图像识别领域的优势突显出来。CNN 具有更强的学习能力和较高的鲁棒性,可以很好地应用于动态手势识别。Molchanov 等^[7]提出了结合高精度网络 and 低精度网络的 3D-CNN,通过 softmax 分类器融合 2 个网络的特征,能够有效提升复杂手势的识别精度和鲁

棒性。Lu 等^[8]使用 CNN 提取帧空间特征,结合长短期记忆网络(LSTM, long short-term memory)时序建模的能力,将 CNN 空间特征输入 LSTM,得到动态手势的时空融合特征,最后使用 softmax 函数得到分类结果。CNN 应用于动态手势识别得到了较好的识别效果,但计算耗时且容易出现过拟合现象,因此需要大量的数据来训练模型,导致其对存储空间和硬件设备性能的要求较高,难以在资源受限的环境中应用。

此外,目前在计算机视觉领域,对于动态手势识别都是采用“源端视频压缩→传输→后端解码重建视频→视频图像分帧→特征提取与分析识别”的处理框架进行的。这样的处理框架将视频图像的分析处理与多媒体数据处理领域割裂开来,没有充分利用多媒体数据处理领域已经挖掘出的视频数据时间及空间特征信息。在现有视频设备普遍采用的视频编码技术中,如数字音视频编解码技术标准(AVS, audio video coding standard)、高效视频编码(HEVC, high efficiency video coding)等,使用运动矢量来表征运动对象在相邻图像帧之间的位移信息^[9]。如果能够充分利用这些信息,不但可以降低用于视频图像分析识别的数据量,还可以简化特征提取的计算复杂度。因此,提出了一种基于视频编码数据特性的动态手势识别框架及视频图像分析识别方法,通过对编码视频数据中的运动矢量使用基于密度的聚类算法 DBSCAN (density-based spatial clustering of applications with noise)提取出手势的运动趋势特征,再采用随机森林(RF, random forest)分类特征,结合 CNN 从编码视频信息帧中提取的手型特征计算得到动态手势的分类结果。该方法降低了视频图像特征提取和分析识别所需的数据量和计算复杂度,可以直接被部署在资源受限的嵌入式设备上。与远程调用云端服务完成高强度神经网络计算的方案相比,该方法在安全性、时延响应和传输开销等方面都有一定的优势。此外,与现有研究水平相比,所提方法对于不同光照条件和复杂背景图像带来的噪声扰动,表现出了较好的鲁棒性,因此应用场景也更加广泛。

1 手势识别框架

为了发挥智能算法的优势,同时又受制于终端的计算和存储资源限制,现有基于动态手势识别应用不得不采用如图 1 所示的系统架构。在本

地获取手势视频数据,编码压缩后通过网络传输到云端,在云端解码视频流得到一系列的静态手势图像帧,再对图像帧序列进行相关预处理后利用智能算法提取特征并分析识别,最后将结果通过网络传回本地。

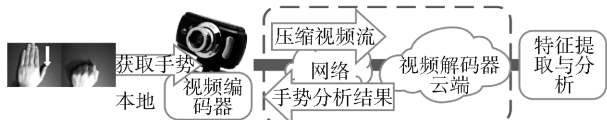


图1 手势识别处理框架

图1所示的处理框架忽略了编码视频中固有的数据特征。在视频编码中,信息帧采用的帧内压缩方式会保留完整的静态图像信息,其中隐藏了动态手势的空间特征。为了降低码率,信息帧之间的预测帧中仅保留了与运动趋势相关的时间特征信息。预测帧编码数据中的运动矢量是通过相邻帧之间的块匹配计算得到的,故这些运动矢量本身就蕴含了视频图像序列中运动对象在时间和空间上的变化特征,即手势变化的趋势信息。可以直接从编码视频数据中提取出来,作为手势的时间特征。这样,一方面可以避免解码带来的量化误差,另一方面还可以大大降低计算复杂度和需要分析处理的数据量,而且对机器学习算法来说运动矢量数据更加规则。此外,由于运动矢量记录的是视频中运动目标相对背景图像的相对位移,光照和背景图像的纹理信息也不会运动矢量中特别明显地表现出来,如图2所示。利用这一特点就能够屏蔽复杂纹理背景和光照条件对运动特征的干扰。依据上述分析,可对图1的处理框架进行简化,略过图中虚线框中的处理过程,直接从编码的视频中提取出特征数据进行分析,极大地减少了分析、处理和存储的数据量,也简化了计算过程。

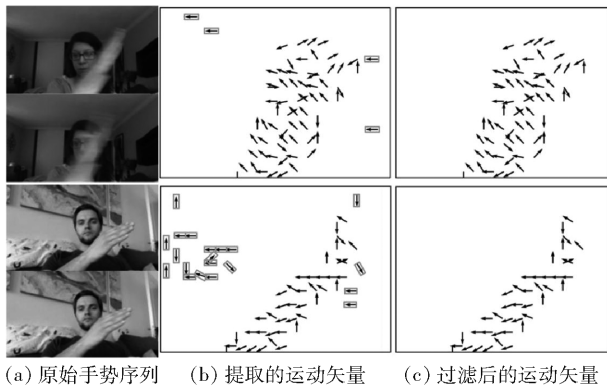


图2 不同光照和背景下的手势及其运动矢量

如果要在边缘设备中完成动态手势识别任务,不但需要高质量的特征数据,还需要轻量的手势特征提取和分析算法。针对简化后的处理框架,提出了一种新的算法框架,如图3所示。手势识别算法框架包括运动趋势分类和手型分类2个处理过程,两部分算法可以并行地对手势特征进行处理。运动趋势分类获取动态手势的时间特征,首先从手势序列的运动矢量中提取出运动趋势特征,再对运动趋势特征分类得到动态手势的运动趋势类别概率。手型分类获取动态手势的空间特征,通过CNN对视频信息帧训练并分类得到手型类别概率。融合运动趋势和手型类别概率得到动态手势的类别概率,概率最大的类别就是预测得到的目标类别。

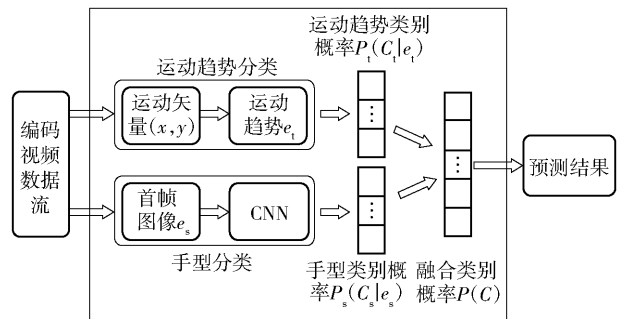


图3 手势识别算法框架

2 运动趋势分类

2.1 运动趋势提取

视频编码中的运动矢量是用于帧间预测的二维矢量,提供编码块从原始帧位置到参考帧位置的偏移量。运动矢量由块匹配计算得到,用参考帧中最佳匹配块相对于原始帧中当前块的位置偏移表示,包含水平方向和垂直方向2个分量。组合视频帧中多个块的运动方位信息,得到视频中物体运动的趋势。手势具有明确的运动趋势轨迹,因而运动矢量的主要趋势可作为手势的运动趋势特征。块匹配可以采用绝对误差和(SAD, sum of absolute differences)作为匹配准则。运动矢量的求取过程为

$$v(i,j) = \arg \min_{(i,j)} \sum_{m=0, n=0}^{N-1} |f_k(m,n) - f_{k-1}(m+i, n+j)| \quad (1)$$

其中: $f_k(m,n)$ 为当前块; $f_{k-1}(m+i, n+j)$ 为参考帧搜索窗内相对位移偏置为 $v(i,j)$ 的一个搜索块, i,j 的变化范围为搜索区域,对所有的 (i,j) 计算误差匹

配函数值;运动矢量 (x, y) 为最小的误差对应的 $v(i, j)$.

由于光照变化的影响,视频编码运动估计也会在非手势区域产生部分与手部运动无关的噪声运动矢量.图 2(b)所示的矩形框内的运动矢量即为图 2(a)中背景区域光照变化产生的.但这些特征与手部运动矢量特征相比位移量没有那么明显,因此采取速度阈值方法^[10]来消除这些无关特征.以运动矢量的相对位移大小作为速度信息,过滤速度小于阈值的特征点,阈值结果为

$$T = \left(2 \times \frac{\sum V}{n} \right) / 3 \quad (2)$$

其中: V 为运动矢量的速度,由相对位移大小计算得出; n 为一个预测帧中所有包含运动矢量的编码块数量.经过速度阈值过滤后的运动矢量如图 2(c)所示,可以看出,背景区域的大多数与手势运动无关的运动矢量都已经被消除.

考虑到手势控制类应用的动作简单,表征其运动趋势的数据维度不高且方向相对集中,但是簇类的形状不确定,因此选择 DBSCAN 对运动矢量的位置进行聚类分析. DBSCAN 不需要预先给出聚类簇的数量,能够以给定的相邻半径(eps)在一帧数据中找出样本数大于给定最少样本数(minpts)的多个密度聚类簇.

以其中样本数量最多的簇作为当前运动手势的主方向簇.对主方向簇的运动矢量进行方向量化,将方向信息转换为二维平面上位于 0° 、 45° 、 90° 、 135° 、 180° 、 225° 、 270° 和 315° 的 8 个角度,运动矢量量化后的角度为

$$r = q \leftarrow \arctan \frac{|x|}{|y|} \quad (3)$$

其中: x, y 为运动矢量 (x, y) 在水平和垂直方向的信息, q 为运动矢量方向角最为趋近的一个量化方向角.

取量化后主方向簇中最多的匹配方向角为当前帧中运动目标的运动趋势 e_t ,其中下标 t 表示时间域.亮度或色度噪声产生的零散运动矢量无法形成密度聚类簇,因此使用 DBSCAN 还可以消除掉这些噪声运动矢量.

采用上述方法从运动矢量中提取手势运动趋势的流程如图 4 所示.

2.2 运动趋势类别概率计算

在得到手势的运动趋势序列后,为了能够与手

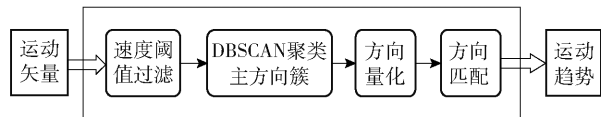


图 4 手势运动趋势提取流程

型特征融合得到动态手势的预测类别,需要使用分类器获取运动趋势的类别概率.对比了 k 最近邻(KNN, k -nearest neighbor)、朴素贝叶斯(NB, naive Bayes)、MLP、RF、SVM 五种分类器对手势运动趋势的分类效果,如表 1 所示.可见在取不同帧数时,RF 分类器对运动趋势的分类效果都是最优的.这是因为 RF 在训练过程中能够检测到特征之间的影响,并且算法拥有很强的抗干扰性,对于不平衡数据集来说,RF 也能够平衡误差,通过平均决策树降低过拟合的风险.因此,采用 RF 对动态手势的运动趋势 e_t 进行分类,可得到 e_t 在类别 C_t 的运动趋势类别概率 $P_t(C_t|e_t)$.

表 1 不同取帧条件下的运动趋势分类算法对比

取帧数/帧	KNN/%	NB/%	MLP/%	RF/%	SVM/%
7	60.37	60.00	44.37	73.41	32.22
19	60.37	75.56	43.41	82.11	57.04
29	61.85	77.04	51.96	85.85	58.52
37	55.93	79.26	50.63	85.93	57.04

3 手型分类与特征融合

动态手势的手型分类相当于静态手势识别,而 CNN 应用于静态图像识别已经非常成熟.杨等^[11]给出了几种 CNN 模型在 ImageNet 数据集上的错误率、网络的时间复杂度和空间复杂度,其中,ResNet 拥有更低的时间复杂度和空间复杂度,并且在图像识别精度上相比其他网络模型更有优势.而且 ResNet 使用残差结构直接将输入信息传到输出,既保护了信息的完整性,又简化了学习目标和难度,同时也缓解了深度神经网络在增加深度时带来的梯度消失问题.因此,可以选择时间复杂度和空间复杂度相对较低的 ResNet-18 作为手型分类的 CNN 模型,通过对手势序列的首帧 e_s 进行分类,得到动态手势的手型类别概率 $P_s(C_s|e_s)$,其中下标 s 表示空间域.

将 2.2 节得到的运动趋势类别概率和 ResNet-18 计算的手型类别概率融合,就可以得到动态手势的类别概率为

$$P(C) = P_t(C_t|e_t)P_s(C_s|e_s) \quad (4)$$

其中： $P_l(C_l|e_l)$ 为运动趋势类别概率， e_l 为运动趋势特征。 $P_s(C_s|e_s)$ 为手型类别概率， e_s 为首帧图像特征。将运动趋势所有类别 C_l 和手型所有类别 C_s 对应的类别概率两两相乘得到动态手势的类别概率 $P(C)$ ，再取动态手势概率最大的类别为目标预测类别。

4 实验仿真与结果分析

4.1 实验条件

为了对功能和性能进行评估，在 i7-3770 3.40 GHz CPU、16 GB 内存、Windows 7 操作系统的

平台上，采用 PyTorch、scikit-learn 框架，用 Python3 语言编程实现了所提出的动态手势识别算法框架。

为了方便与同类工作进行对比，使用了英国剑桥大学手势数据集^[12] 和美国西北大学手势数据集^[13]。2 个数据集的具体指标和参数如表 2 所示。剑桥大学手势数据集拥有 5 种不同的光照条件。由于在同一种照明条件下的数据具有相似性，如果将所有数据混合在一起，随机选取 20% 数据做测试，有可能获得很高的精度，但不能正常反映算法的性能。因此，将数据集按 5 种照明条件进行划分，使用五折交叉验证进行评估，以得到算法的泛化能力。

表 2 数据集信息

数据集	手势视频数量	类别数量	动作数量	手型数量	受试者数量	图像大小/像素	手势动作	手型姿势
剑桥大学手势数据集	900	9	3	3	2	320 × 240 × 3	向左、向右、收缩	扁平、伸展、V 型
美国西北大学手势数据集	1 050	10	10	7	15	640 × 480 × 3	向右、向左、向上旋转、向下旋转、下右移动、右下移动、顺时针旋转、逆时针旋转、“Z”动作、交叉动作	拇指、食指、伸展、“OK”、握拳、侧食指、侧扁平

运动趋势分类中 DBSCAN 取相邻半径为 60，一个簇的最少样本数取 20。使用 scikit-learn 提供的随机森林分类器，参数 $n_estimators$ 取 50， $criterion$ 选择 $entropy$ ， max_depth 取 20，其余参数使用默认值。

手型分类通过从视频中提取首帧 RGB 图像数据输入 ResNet-18 进行训练。训练执行 100 个 epoch，batch 大小为 16，损失函数使用交叉熵损失函数，优化器使用随机梯度下降法，学习率设置为 1×10^{-3} ，动量系数为 0.9。

美国西北大学手势数据集的动作类型均是二维平移动作，因此仅以所提算法中的运动趋势分类算法即可完成该类型的动态手势识别任务，可完全消除对于图像数据的依赖。DBSCAN 取相邻半径为 45，其他配置与剑桥大学手势数据集配置相同。

4.2 实验结果和分析

图 5 所示为剑桥大学手势数据集的五折交叉验证下不同手势的识别结果混淆矩阵。可以看出，扁平/向左、V 型/向左 2 种手势识别率在 97% 及以上，识别效果最好。扁平/向左和扁平/向右容易混淆为扁平/收缩。图 6 所示为美国西北大学手势数据集以受试者进行留一法交叉验证下不同手势的识别结果混淆矩阵。可以看出，算法对向左、下右移动、逆时针旋转、“Z”动作识别效果最好，因为这些动作包含的运动趋势特征相对其他手势区分度较

高。对于向右动作识别混淆的其余几个动作大都包含向右的运动趋势，受试者在执行向右的动作时，若角度偏差较大则会导致识别错误。

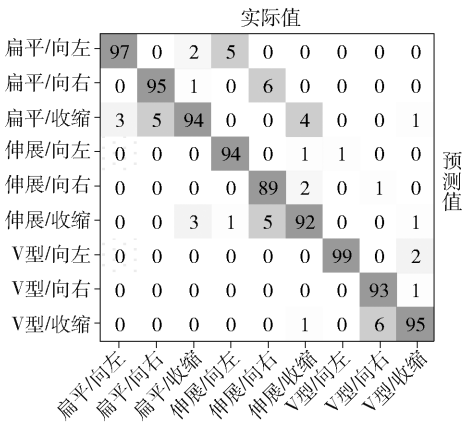


图 5 剑桥大学手势数据集混淆矩阵

表 3 所示为在相同计算资源的条件下，所提算法与其他文献中的算法对剑桥大学手势数据集的识别正确率和计算时间。Tang 等^[4]首先从数据中按照 30 帧/s 的采样频率提取关键帧图像序列，再从各关键帧图像中提取出外观和运动特征数据，算法计算复杂，耗时较长。文献[3,5-6]中所述为传统机器视觉方法，其最优正确率均低于所提算法的平均正确率，且识别时间没有优势。在同样的运行环境下，将文献[8]中所述的 CNN 复现为

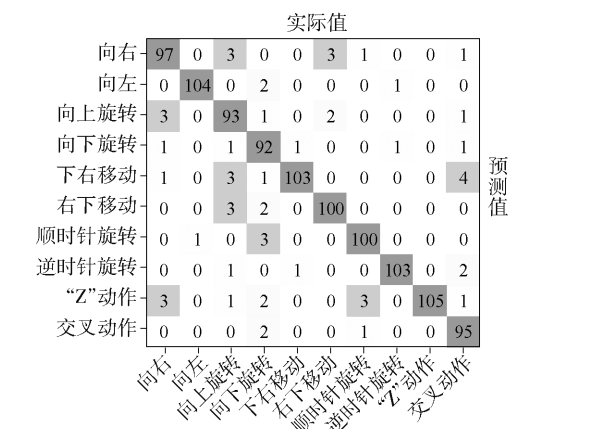


图 6 美国西北大学手势数据集混淆矩阵

ResNet-18,用 6 个关键帧提取特征.可以看出,虽然文献[8]中算法的识别率高于所提算法的识别率,但所提算法的手势识别计算时间减少了 85%.若使用文献[8]中的 ResNet-50,其算法的执行时间会更长,存储资源消耗也会更多.

表 3 剑桥大学数据集上不同手势识别算法的性能对比

算法	识别率/%	时间/ms
IE + DC ^[4]	98.23	4 310
CNN + LSTM ^[8]	97.78	1 020
DTW + CSURF ^[3]	94.08	1 316
Hough + MLP ^[5]	94.00	1 549
gOHSM ^[6]	85.50	930
本文	94.22	149

表 4 所示为针对不同数据集所提出的算法和文献[8]的方法取不同关键帧数量时的性能对比.可以看出,对于剑桥大学数据集,所提算法的识别率比文献[8]中的方法略低,但所提算法只用到了第 1 帧的图像数据,其余帧使用的是运动矢量数据.虽然需要较多帧的矢量信息,但是输入算法的数据量还是远远少于文献[8],相应的特征提取和识别的计算时间也明显优于文献[8]中的算法.

为了验证所提算法对不同光照和背景纹理的鲁棒性和泛化能力,在美国西北大学手势数据集的基础上增加了两组自建数据.如图 7 所示,在自然室内环境下,采集了同一受试者在普通光照和暗光条件下执行美国西北大学数据集中 7 种手型的 10 个动作.以原数据集中 15 个受试者的手势作训练集,以自建数据为测试集,与文献[8]中的方法进行对比,实验结果如表 5 所示.可以看出,所提算法在背

景纹理和光照有较大差异的情况下仍然能够保持相对稳定的识别效果.

表 4 不同数据集上的算法性能对比

算法	数据集	关键帧数	识别率/%	时间/ms	数据量/KB
本文	剑桥大学手势	7	90.22	154	151
		19	93.78	146	158
		29	94.22	149	164
	美国西北大学手势	10	92.95	101	23
		15	93.81	102	35
		20	94.48	110	47
CNN + LSTM ^[8]	剑桥大学手势	6	97.78	1 020	882
		9	98.22	1 480	1 323
		12	98.33	1 890	1 764
	美国西北大学手势	6	90.58	1 160	882
		9	92.48	1 640	1 323
		12	92.76	2 270	1 764

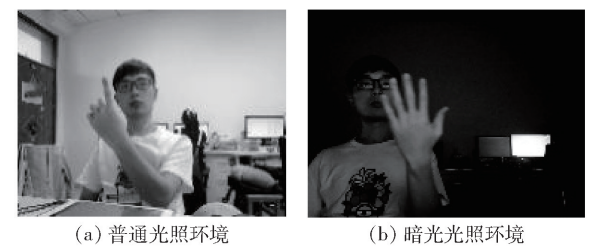


图 7 复杂背景纹理下的光照环境

表 5 不同光照条件下的识别效果对比

算法	普通光照识别率/%	暗光光照识别率/%
本文	98.57	97.14
CNN + LSTM ^[8]	28.57	15.71

5 结束语

为解决现有动态手势识别方法存储占用多、计算复杂的问题,提出了一种利用视频编码数据特征实现动态手势识别的方法.以手势视频编码数据中的运动矢量和首帧图像作为时空特征,避免解码重建视频带来的视觉特征损失.实验结果表明,该方法在剑桥大学和美国西北大学的数据集上能将识别精度维持在 94% 以上,同时还大大减少了算法的执行时间,满足了实时响应的需要.运动矢量只关注手势视频中的运动信息,因此不需要对手势进行定位操作,并且对光照和复杂背景也呈现出较好的鲁棒性.笔者所提的算法框架同样适用于微表情、肢

体等视频中的行为特征提取,后期将尝试探索该方法在相近领域的应用效果。

参考文献:

- [1] Deller M, Ebert A, Bender M, et al. Flexible gesture recognition for immersive virtual environments [C] // IV'06. London: IEEE, 2006: 563-568.
- [2] Elmezain M, Al-Hamadi A, Michaelis B. Hand trajectory-based gesture spotting and recognition using HMM[C] // ICIP. Cairo: IEEE, 2009: 3577-3580.
- [3] Barros P, Maciel-Junior N T, Fernandes B J T, et al. A dynamic gesture recognition and prediction system using the convexity approach [J]. CVIU, 2017, 155: 139-149.
- [4] Tang Hao, Liu Hong, Xiao Wei, et al. Fast and robust dynamic hand gesture recognition via key frames extraction and feature fusion[J]. Neurocomputing, 2019, 331: 424-433.
- [5] Patil A R, Subbaraman S. A spatiotemporal approach for vision-based hand gesture recognition using Hough transform and neural network[J]. Signal, Image and Video Processing, 2019, 13(2): 413-421.
- [6] Gatto B B, dos Santos E M, Da Silva W S. Orthogonal Hankel subspaces for applications in gesture recognition [C] // SIBGRAPI. Niteroi: IEEE, 2017: 429-435.
- [7] Molchanov P, Gupta S, Kim K, et al. Hand gesture recognition with 3D convolutional neural networks [C] // CVPR. Boston: IEEE, 2015: 1-7.
- [8] Lu Dongwei, Qiu Chu, Xiao Yi. Temporal convolutional neural network for gesture recognition [C] // ICIS. Singapore: IEEE, 2018: 367-371.
- [9] Muralidhar P, Rao C B R. Analysis of block matching motion estimation algorithms [C] // ICCNT. Tiruchengode: IEEE, 2013: 1-4.
- [10] Singha J, Roy A, Laskar R H. Dynamic hand gesture recognition using vision based approach for human-computer interaction [J]. Neural Computing and Applications, 2018, 29(4): 1129-1141.
- [11] 杨真真, 匡楠, 范露, 等. 基于卷积神经网络的图像分类算法综述[J]. 信号处理, 2018, 34(12): 1474-1489.
- Yang Zhenzhen, Kuang Nan, Fan Lu, et al. Review of image classification algorithms based on convolutional neural networks [J]. Journal of Signal Processing, 2018, 34(12): 1474-1489.
- [12] Kim T K, Cipolla R. Canonical correlation analysis of video volume tensors for action categorization and detection[J]. TPAMI, 2009, 31(8): 1415-1428.
- [13] Shen Xiaohui, Hua Gang, Williams L, et al. Dynamic hand gesture recognition: an exemplar-based approach from motion divergence fields [J]. Image and Vision Computing, 2012, 30(3): 227-235.