

文章编号:1671-6833(2021)02-0067-07

基于视觉和肌电信息融合的手势识别方法

彭金柱,董梦超,杨 扬

(郑州大学 电气工程学院,河南 郑州 450001)

摘要:针对人机交互技术对手势识别的可识别种类和识别正确率的问题,提出一种基于视觉方向梯度直方图(HOG)特征和肌电信号(EMG)时域特征融合及支持向量机(SVM)分类器的手势识别方法。利用视觉传感器和智能臂环分别采集手势图像信息和肌电信号,预处理后提取对应的HOG特征和时域特征;采用串行融合的方式将2种特征进行特征级融合;以一对一方式构建组合式SVM多分类器完成手势识别模型的训练和检验,并进行对比分析。结果表明:采用特征级串行融合的36类手势识别模型的总体正确率达到了96%,相较于融合前单一HOG特征和EMG时域特征分别有33%和16%的提升;相较于决策级融合有11%的正确率提升,且计算耗时仅0.274 ms,证明该方法能显著减少特征数据量,有效提高多种类手势识别正确率。

关键词:手势识别;方向梯度直方图;肌电信号;时域特征;特征融合

中图分类号: TP391.4

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2021.02.014

0 引言

近年来,随着人机交互技术的快速发展,利用各类传感器进行手势识别从而控制机器完成指定任务已经变得非常普遍^[1]。手势动作交流具有较高的灵活性且生动形象,当将其应用于人机交互中时,对可识别手势动作的种类和手势动作信息的精准获取提出了更高的要求。

从手势信息获取方式来看,手势识别主要分为2种:通过人体手势视觉图像信息实现手势识别和获取人体手势肌电信号(EMG)进行手势识别^[2]。Saha等^[3]通过Kinect视觉传感器采集人体手势图像信息,利用隐马尔可夫模型(HMM)识别手势,但该方法存在难以调参的问题,对识别时间和效率有一定影响。任彧等^[4]提取方向梯度直方图(HOG)特征作为手势图像的特征描述符,基于支持向量机(SVM)多分类算法完成对手势识别模型的训练,消除了光照不均和手势角度旋转对手势识别的影响。但该方法要求背景简单,且只对有限的9类手势进行识别。Naik等^[5]通过采集到的手臂EMG信号,提取信号的分维数特征,用SVM等方法进行

手部运动模式的分类,识别正确率较高。张启忠等^[6]提出了球均值Lyapunov指数计算的方法提取EMG信号特征,利用SVM作为分类器完成4类手势的分类,准确率达到96%。Lahiani等^[7]开发了一种基于HOG和局部二值模式(LBP)特征的静态手势识别系统,对10种手势的识别正确率达到了92%。但是,受限于当前常用特征的区分能力,目前已有的大多数手势识别算法只能在手势类别数量较少时获得不错的性能。当需要处理的手势种类增多时,不同手势的类间距离迅速变小,使得这些方法的识别正确率难以保证。

针对上述问题,本文提出一种基于视觉和EMG信息融合的手势识别方法,主要分为3个部分:对单一视觉HOG特征和单一肌电信号时域特征的提取;将2种单一特征进行特征融合,得到融合特征;以融合后特征作为SVM多分类器的输入训练手势识别模型。该方法利用融合特征代替单一特征,从视觉和肌电2个角度完成对手势信息的特征描述,所需数据量远小于同等识别效果的单一HOG特征,在手势种类多达36种的情况下识别正确率达到96%。

收稿日期:2020-09-12;修订日期:2020-11-23

基金项目:国家自然科学基金资助项目(61773351);河南省高校科技创新人才支持计划项目(20HASTIT031)

作者简介:彭金柱(1980—),男,河南信阳人,郑州大学副教授,博士,主要从事机器人控制、计算机视觉等方面的研究,E-mail:jzpeng@zzu.edu.cn。

1 特征提取

1.1 视觉 HOG 特征提取

HOG 最早由 Dalal 于 2005 年提出,具有几何不变矩的特性,多用于行人检测中,并取得了较好的效果^[8]。HOG 依据一幅图像的形状、外观等特征能够被梯度或边缘的方向密度分布很好地描述的原理,来适应光照变化和目标旋转。

视觉 HOG 特征提取过程:首先利用 Kinect 视觉传感器获取手势图像并对其进行图像中值滤波、颜色空间转换等预处理操作;然后基于 YCbCr 颜色空间的椭圆模型完成对手势区域的分割;最后对图像进行灰度化操作,计算并提取手势 HOG 特征。HOG 特征提取过程如图 1 所示。

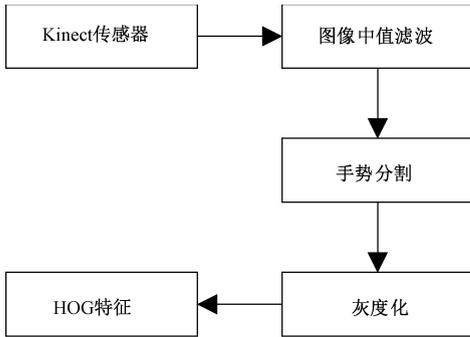


图 1 HOG 特征提取流程

Figure 1 Flow chart of HOG feature extraction

1.1.1 手势分割

手势分割是把人的手势区域从完整的图像中分割出来,减少背景带来的训练识别过程的计算量冗余。

椭圆模型是一种在 YCbCr 颜色空间下的手势分割模型。研究表明:把人体手势肤色像素投影到 CbCr 子空间上,其投影近似在一个椭圆范围内^[9]。通过计算图像中像素点的投影位置是否在椭圆范围内,判定当前像素点是否是人体手势像素点^[10]。椭圆模型的表达式为:

$$\frac{(x - eC_x)^2}{a^2} + \frac{(y - eC_y)^2}{b^2} = 1; \quad (1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} C_b & -C_x \\ C_r & -C_y \end{bmatrix}. \quad (2)$$

式中: $C_x = 109.38$; $C_y = 152.02$; $\theta = 2.53 \text{ rad}$; $eC_x = 1.60$; $eC_y = 2.41$; $a = 25.39$; $b = 14.03$ 。

椭圆模型具有计算速度快和检测准确率高等特点,所以本文选用椭圆模型完成对手势的分割。图 2 为基于 YCbCr 颜色空间的椭圆模型手势分割效果。

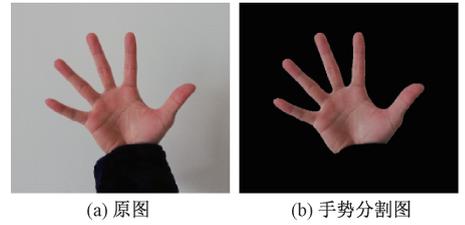


图 2 椭圆模型手势分割效果

Figure 2 Ellipse model gesture segmentation renderings

1.1.2 HOG 特征计算

与其他图像几何特征不同,HOG 特征将图像细分为多个小的细胞单元 (cell),然后计算所有细胞单元中各像素点的梯度或边缘的方向直方图。为提高性能,将若干个细胞单元组成一个区间 (block),图像变为由若干个区间组成的连通图,然后在这些区间内对各细胞单元的梯度进行归一化,取得最终的梯度方向向量。

图像的梯度方向向量数由图像大小、cell 大小、每个 cell 所取梯度维数等因素决定。计算式为:

$$D_{\text{HOG}} = C_d \cdot \left(\frac{B_s}{C_s}\right)^2 \cdot \frac{(w - B_n) \times (h - B_n)}{B_s^2}. \quad (3)$$

式中: $B_n = (B_s - B_t)$, B_t 表示 block 移动步长, B_s 表示 block 大小; C_s 表示 cell 大小; C_d 表示 cell 维数; w 与 h 分别表示图像宽与高。

1.2 肌电信号时域特征提取

EMG 的时域特征提取过程:首先利用 Myo 臂环获取手势动作信号并进行去噪等预处理操作;其次,对预处理后的肌电信号进行有效活动段检测;最后完成对 EMG 的时域特征提取。EMG 信号的时域特征提取过程如图 3 所示。

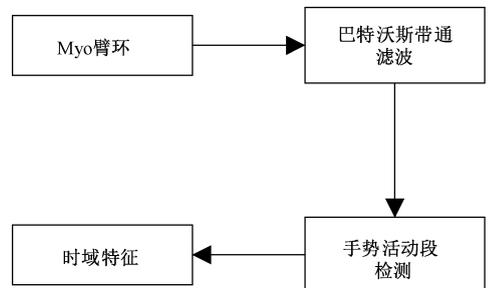


图 3 EMG 时域特征提取流程

Figure 3 Flow chart of EMG time domain feature extraction

1.2.1 有效活动段检测

使用 4 阶巴特沃斯带通滤波器对 EMG 进行滤波处理^[11],然后,对手势动作的肌电信号进行活动段检测,找到手势动作的有效起止点,从而能快速完成手势动作的特征提取,且能够提高特征

提取的准确度。

移动平均法可以在一定程度上减轻随机波对时间序列信号整体走势的影响,因此,选用移动平均法进行有效活动段检测。对获取到的时间序列信号设定一定宽度的时间窗和滑动步长,按顺序叠加平均,计算此时间窗内的采样数据点,进而获得一个新的时间序列信号^[12]。对手势肌电信号进行有效活动段检测时,分别对8通道肌电信号数据进行移动平均处理,具体操作步骤如下。

(1)定义每个通道表面肌电信号 $S_k(t)$, $k=1, 2, \dots, 8$ 。对各通道表面肌电信号 $S_k(t)$ 进行平方处理,从而获得EMG瞬时能量的时间序列 $E(t)$:

$$E(t) = S_k^2(t)。 \quad (4)$$

(2)在采样频率为1 000时,设定窗口大小 $W=64$,在滑动步长为16时,对EMG能量时间序列 $E(t)$ 进行滑动平均计算,进而获得滑动平均后的时间序列 $E_M(t)$:

$$E_M(t) = \frac{1}{W} \sum_{i=t-W+1}^t E(i)。 \quad (5)$$

通过观察,设置合适的阈值,比较 $E_M(t)$ 和阈值的大小,从而确定手势动作肌电信号有效活动段的起止点。具体方法是:将时间序列 $E_M(t)$ 大于阈值且之后的2个连续点数据值都大于阈值的点确认为有效活动段的起始点;将时间序列 $E_M(t)$ 小于阈值且之后的2个连续点数据值小于阈值的点确认为有效活动段的结束点。

1.2.2 时域特征提取

特征提取的目的是利用手势携带的特定信息将不同手势动作区别开来,这对后续不同手势肌电信号的分类有至关重要的作用。目前,肌电信号特征提取在信号分类方面具有较好的效果,其方法包括:时域分析方法、频域分析方法、时频域分析方法及非线性动力学方法。从国内外学者的研究来看,相较于其他分析方法,时域分析方法在表征手势特征方面效果更好,且计算简单、获取迅速、实时性高。

由于EMG属于生理信号,是一种随机性较大、不易重复的生物电信号,所以一种特征参数并不能实现肌电信号的全面描述。本文选用时域统计学特征来作为不同手势的分类标准,5个时域统计学特征如下:均方根值(RMS)、波形长度(WL)、平均绝对值(MAV)、过零点数(ZC)、斜率变化数(SSC)。5个时域统计学特征的数学定义如表1所示。

表1 时域特征的定义

Table 1 Definition of time domain feature

时域特征	数学定义
RMS	$F_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{k=1}^N x_k^2}$
WL	$F_{\text{WL}} = \frac{1}{N} \sum_{k=1}^{N-1} x_{k+1} - x_k $
MAV	$F_{\text{MAV}} = \frac{1}{N} \sum_{k=1}^{N-1} x_k $
ZC	$F_{\text{ZC}} = \frac{1}{N} \sum_{k=1}^{N-1} \text{sgn}(-x_k x_{k+1})$
SSC	$F_{\text{SSC}} = \frac{1}{N} \sum_{i=1}^{N-1} f_i,$ $f_i = \begin{cases} 1, & (x_k - x_{k-1})(x_k - x_{k+1}) > \delta \\ 0, & \text{其他} \end{cases}$

注: $k=1, 2, \dots, N$ 是信号的时间序列。

2 特征融合及分类器设计

2.1 特征融合

相比其他技术,多传感器信息融合技术最大的优势在于能够获得其他单一传感器不能得到的被感知物体的完善的基本特性,并且多传感器信息融合技术在信息融合后具有较强的可靠性、时效性、鲁棒性^[13]。同一被感知物体在不同背景下,表现出来的特征信息其实都不相同,但是对于多传感器信息融合来说,一般都是在同一个级别下完成的。根据被感知物体信息抽象程度,将多传感器信息融合分为像素级数据融合、特征级数据融合和决策级数据融合^[14]。

选择特征级数据融合作为多传感器信息融合的融合结构,串行特征融合是直接来自2类传感器的特征向量进行首尾相接,从而获得包含不同特征类型的高维特征向量。假设样本空间为 Ω 且 A, B 对应不同的特征空间类型,任意选择一个样本 $\zeta \in \Omega$,其所对应的2类特征向量分别是 $\alpha \in A, \beta \in B$,那么 ζ 的串行融合特征向量 φ 定义为:

$$\varphi = \alpha\beta。 \quad (6)$$

提高特征向量有效数据维数是对当前样本更细致的特征描述,可以增加不同样本之间的可区分性。提高单一特征向量的有效数据维数从而细致刻画不同样本在当前特征下的差异,是一种提高识别正确率的方法^[15]。但是该方法在数据量上的增加,会给计算量带来量级上的增大,甚至带来维数灾难问题;其次,有效数据维数提高存在瓶

预期,当数据维数高到一定量级时,再持续增加特征向量维数,对结果提升不明显。

融合特征向量有效数据维数的提高是基于增加样本特征描述的类别,通过不同特征下样本差异性的累加来提高其可区分性。考虑到2种特征融合占比的均衡性,为了更直观地展现融合特征给正确率带来的提升,本文融合特征向量 γ (184 维)由 HOG 特征向量 α (144 维)和 EMG 时域特征向量 β (40 维)串行融合而成。相比于对单一特征细致描述产生的数据量,融合特征数据量增加较少且给样本之间加入另一特征下的差异描述,性价比更高,且多特征的提升空间更大,能有效解决瓶颈期问题。

2.2 SVM 多分类器设计

SVM 通过建立一个分类超平面作为决策面,使不同类之间的隔离边缘最大化,从而实现分类。基础的 SVM 仅针对二分类问题,不能满足实际问题中多分类问题的需要。为将 SVM 良好的分类性能运用到多分类问题,专家提出一对多^[16]、一对一^[17]、基于有向无环图^[18]以及基于决策树^[19]等方法构建组合式 SVM 多分类器。本文采用基于一对一方式实现的组合式 SVM 多分类器解决手势识别问题。

对于一个具有 $N(N \geq 2)$ 个类别的分类问题,此方法共构建 $N(N-1)/2$ 个 SVM 分类器,每个分类器实现对 2 个类别的分类,使得 N 个类别中任意 2 类都有其相对应的分类器。假定 $N=3$,对应的 3 分类组合式 SVM 分类器结构如图 4 所示。

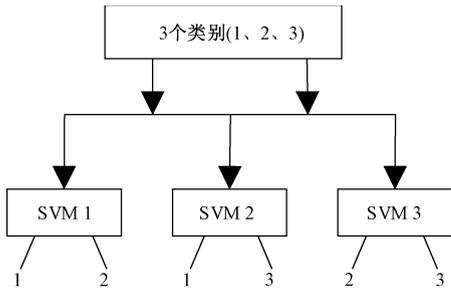


图 4 3 分类 SVM 结构

Figure 4 3 classification SVM structure chart

每个分类器对输入样本进行非此即彼的分类判定,采用投票策略对输入样本进行最终识别,即票数最多的一类为输入样本所属类别。当出现多个类别票数相同的情况时,由算法判定类别序号最小的类别为输入样本最终所属类别。

基于一对一方式实现的组合式 SVM 多分类器中每个 SVM 面对的都是 2 个类别的分类问题,易于训练,充分发挥了 SVM 解决二分类问题的优

势。本文使用的 SVM 分类器的类型是 C-SVC。选用径向基核函数(RBF)作为核函数,其具有较宽的收敛性,易于调参,可以在一定程度上消除高维空间中的复杂内积运算。通过网格搜索法对惩罚系数以及核参数进行寻优。

3 实验结果及对比

本文以图像 HOG 特征与肌电信号时域特征组成的融合特征作为手势的特征描述符,基于 SVM 多分类算法实现对 36 种手势(包含 10 个数字和 26 个英文字母)的识别。由于识别手势种类的多样性以及多特征融合的方法在当下没有合适的公开数据集供研究使用,所以采用自制的手势数据集完成相关研究工作。

自制手势数据集共由 10 位实验者提供组成,其中男女人数比例为 1:1。36 种手势与所代表含义的对应关系如图 5 所示,每人每种手势采集 3 张图片以及相对应的 3 组手势肌电信号,即每种手势共分别得到 30 组有效数据。

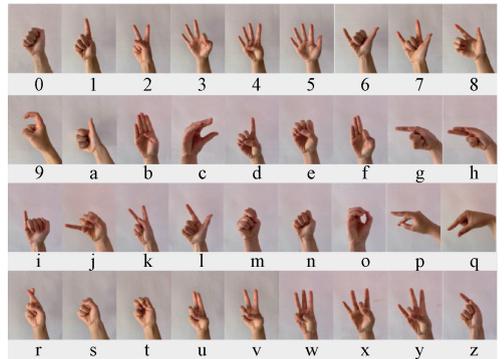


图 5 手势对应关系

Figure 5 Gesture corresponding chart

本文实验硬件平台配置:Inter i5-4460 处理器,8 G 内存(1 600 MHz)。SVM 多分类算法在 MATLAB 2020b 平台基于 Libsvm 功能包实现。

使用十折交叉验证评价算法的分类性能,并记录每个类别手势的识别正确率,总体正确率为每个类别的正确率之和与总类别数的比值。采用融合特征的 36 种手势识别的正确率如表 2 所示。

相同实验条件下,分别对单独 HOG 特征(144 维)、单独 EMG 时域特征以及基于 SLMF(score-level multifeature fusion)方式^[20]的特征融合(144 维 HOG 特征与 EMG 时域特征的融合)进行 36 种手势的识别实验,并将得到的手势识别正确率与本文提出的特征融合做对比,对比情况如图 6 所示。该方法的实验结果在图 6、8 和表 3 中由 HOG144+EMG(S)表示。

表 2 特征融合识别正确率

Table 2 Feature fusion recognition accuracy

手势	识别正确率/%	手势	识别正确率/%	手势	识别正确率/%
0	90	c	100	o	97
1	97	d	97	p	97
2	97	e	100	q	90
3	100	f	97	r	100
4	97	g	100	s	93
5	100	h	90	t	100
6	97	i	100	u	77
7	97	j	100	v	90
8	97	k	100	w	93
9	90	l	97	x	87
a	100	m	90	y	100
b	87	n	100	z	100

势识别正确率总体上高于融合前单一特征。由图 6 可以看出,该方法虽然部分字符识别的正确率较高,但是有 11 种手势的正确率下降明显,如手势 1 的正确率比融合前单一 EMG 时域特征识别方法下降了 40%,手势 b 的正确率下降了 16%。其原因是,当分类类别过多以及类别之间差异性不大时,分类器无法对正确的类别做出坚定的选择,使得正确类别与相似类别的评判分数相近。以 SLMF 为代表基于权重系数的决策级特征融合无法有效处理上述情况,甚至会因为权重而提高相似类别的评判分数,使其大于正确类别,导致融合后正确率的降低。

串行融合方式的特征融合则会提供更多的特征,增加相似样本之间的差异性,促使分类器做出正确的选择。串行融合特征下唯有手势 q 的正确率相对于单一特征下降了 7%,正确率降到 90%,特征的融合对极个别手势的差异性带来一定损失。

特征融合后与特征融合前的总体正确率对比如表 3 所示。本文串行融合特征总体识别率相对于单一 HOG 特征提高了 33%,相对于 EMG 时域特征提高了 16%。相对于 SLMF 融合方式提高了 9%。基于串行融合方式,融合特征的手势识别方法对手势识别正确率有明显提升。

表 3 总体正确率对比

Table 3 Comparison of overall accuracy

特征类型	总体正确率/%
EMG	80
HOG144	63
HOG576	84
HOG2304	92
HOG144+EMG(S)	87
HOG144+EMG	96

为验证串行融合特征的高效性,提高手势图像信息单一 HOG 特征的有效维数到 576 维和 2 304 维,与串行融合特征进行对比实验,不同种类手势的识别正确率和总体正确率如图 7、表 3 所示。

对比上述实验结果可知,随着 HOG 特征数据维数的增加,识别正确率虽然也在提高,但是提高幅度越来越小。相较于 576 维的 HOG 特征,串行融合特征向量只有 184 维,分类别手势识别正确率普遍有提高,总体正确率比其要高 12%。直至 HOG 特征向量达到 2 304 维,此时 HOG 特征的数据量是串行融合特征的 12 倍,其总体正确率

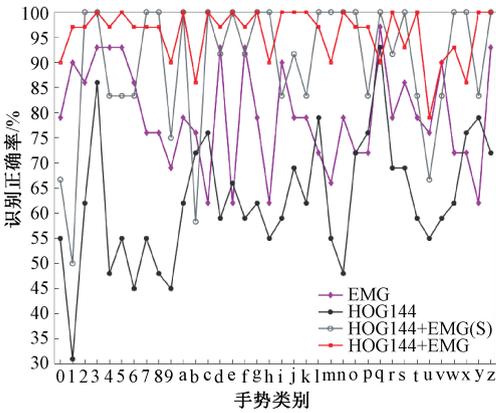


图 6 特征融合前后分类识别正确率比较

Figure 6 Comparison of classification accuracy before and after feature fusion

本文在基于 SLMF 方式的特征融合实验中,自制数据集的 60%作为训练集,40%作为测试集;由训练集样本训练不同特征类型对应的模型,记录样本识别分数并计算不同特征类型对应的权重;测试得到测试集样本的识别分数,结合不同特征对应的权重,完成特征融合过程。

对比上述实验结果发现,本文提出的特征融合的手势识别方法在 36 种手势识别正确率方面明显高于单一特征,其中 33 种手势识别正确率在 90%以上,更有 14 种达到了 100%。特征融合对视觉上相似手势 0、m、n、s、t 的识别正确率有 24%~53%不等的提升,表明 EMG 时域特征的融合增加了相似手势之间的差异性;对 EMG 时域特征识别正确率较差的手势 c、e、h、y 等有良好的识别效果,正确率都在 90%以上,其中 c、e、y 等的正确率提高到了 100%。

与串行融合特征相似,基于 SLMF 方式的手

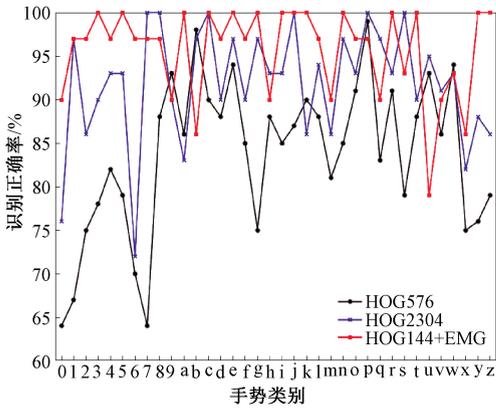


图7 与高维 HOG 特征的分类识别正确率比较
Figure 7 Comparison of classification accuracy with high-dimensional HOG features

仍比特征融合低 4%。

在各类别手势中随机选取 10 组样本数据, 共计 360 组样本数据, 分别放入 6 种不同特征类型对应的模型中, 记录不同特征类型下 360 组样本数据计算识别过程所需时间, 每组样本数据的平均计算耗时如图 8 所示。

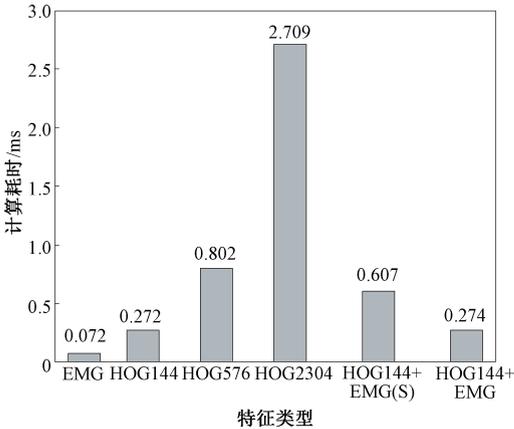


图8 计算耗时比较

Figure 8 Comparison of calculation time

图 8 实验结果中, 本文串行融合特征下对每个样本的平均识别时间是 0.274 ms, 融合前 144 维单一 HOG 特征是 0.272 ms, 2 304 维 HOG 特征是 2.709 ms, 基于 SLMF 方式的特征融合的平均耗时是 0.607 ms。结合图 8 计算耗时与表 3 中总体正确率, 有以下分析: ①本文串行融合方法的计算耗时与融合前 144 维单一 HOG 特征相近, 总体正确率却由 63% 提升至 96%; ②提升单一 HOG 特征至 2 304 维, 单一 HOG 特征比串行融合特征总体正确率低 4%, 计算耗时是本文串行融合特征的将近 10 倍; ③以基于 SLMF 方式为代表的决策级特征融合由算法特性决定, 需要多个模型参与识别, 其计算耗时是多个模型计算识别时间的

总和, 故该算法虽然与本文串行融合特征维数相同, 但计算耗时是本文方法的 2 倍多。相对于本文串行融合特征 96% 的总体正确率, 该方法为 87%, 比本文方法低 9%。

4 结论

本文提出一种融合视觉 HOG 特征和肌电信号时域特征进行手势识别的方法, 实验结果表明: 在识别手势种类多达 36 种的情况下, 特征融合后的识别正确率明显高于融合前单一特征。与高维 HOG 特征对比表明, 特征融合能够以较少的特征数据量完成对手势信息更完善的描述, 减少计算量, 有效提高正确率。

参考文献:

- [1] 于泽升, 崔文华, 史添玮. 基于 Kinect 手势识别的应用与研究[J]. 计算机科学, 2016, 43(增刊 2): 568-571.
- [2] 张风军, 戴国忠, 彭晓兰. 虚拟现实的人机交互综述[J]. 中国科学: 信息科学, 2016, 46(12): 1711-1736.
- [3] SAHA S, LAHIRI R, KONAR A, et al. HMM-based gesture recognition system using kinect sensor for improvised human-computer interaction[C]//2017 International Joint Conference on Neural Networks (IJCNN). New York: IEEE, 2017: 2776-2783.
- [4] 任彧, 顾成成. 基于 HOG 特征和 SVM 的手势识别[J]. 科技通报, 2011, 27(2): 211-214.
- [5] NAIK G R, KUMAR D K, ARJUNAN S P. Towards classification of low-level finger movements using forearm muscle activation: a comparative study based on ICA and Fractal theory[J]. International journal of biomedical engineering and technology, 2011, 6(2): 150-162.
- [6] 张启忠, 席旭刚, 罗志增. 基于非线性特征的表面肌电信号模式识别方法[J]. 电子与信息学报, 2013, 35(9): 2054-2058.
- [7] LAHIANI H, KHERALLAH M, NEJI M. Hand gesture recognition system based on local binary pattern approach for mobile devices[C]//Intelligent systems design and applications. Berlin: Springer, 2018: 619-629.
- [8] 曾庆山, 宋庆祥, 范明莉. 基于光流共生矩阵的人群行为异常检测[J]. 郑州大学学报(工学版), 2018, 39(3): 29-33, 39.
- [9] 薛俊韬, 纵蕴瑞, 杨正瓩. 基于改进的 YCbCr 空间及多特征融合的手势识别[J]. 计算机应用与软件, 2016, 33(1): 151-155.
- [10] CHAKRABORTY B K, BHUYAN M K, KUMAR S. Combining image and global pixel distribution model

- for skin colour segmentation [J]. Pattern recognition letters, 2017, 88: 33-40.
- [11] TSENG C C, LEE S L. Closed-form designs of digital fractional order Butterworth filters using discrete transforms [J]. Signal processing, 2017, 137: 80-97.
- [12] 王立玲, 杨铮, 刘瑾. 基于积分肌电值结合双阈值的 s-EMG 活动段的检测 [J]. 东北师大学报 (自然科学版), 2018, 50(3): 65-71.
- [13] 王晶晶. 多传感器信息融合技术在机器人导航中的应用研究 [J]. 电声技术, 2019, 43(11): 23-24, 50.
- [14] LU Z Y, 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.
- [15] 王艳丽, 梁静, 薛冰, 等. 基于进化计算的特征选择方法研究概述 [J]. 郑州大学学报 (工学版), 2020, 41(1): 49-57.
- [16] BOTTOU L, CORTES C, DENKER J S, et al. Comparison of classifier methods; a case study in handwritten digit recognition [C] // Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5). Piscataway: IEEE, 1994: 77-82.
- [17] KNERR S, PERSONNAZ L, DREYFUS G. Single-layer learning revisited: a stepwise procedure for building and training a neural network [J]. Neurocomputing, 1990, 68: 41-50.
- [18] PLATT J, CRISTIANINI N, SHAWE-TAYLOR J. Large margin DAG's for multiclass classification [J]. Neural information processing systems, 2000, 12: 547-553.
- [19] LU S B, ZHOU L. Fatigue crack monitoring of aerospace structure based on binary tree support vector machines [J]. Proceedings of the SPIE, 2017, 10170: 1017032.
- [20] KUANG H L, CHEN L, GU F, et al. Combining region-of-interest extraction and image enhancement for nighttime vehicle detection [J]. IEEE intelligent systems, 2016, 31(3): 57-65.

Human Gesture Recognition Method Based on Vision and EMG Signal Information

PENG Jinzhu, DONG Mengchao, YANG Yang

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Aiming at the problem of variety and accuracy of gesture recognition in human-computer interaction, a gesture recognition method was proposed based on fusion features of visual histogram of orientation gradient (HOG) and time-domain features of electromyography (EMG), where support vector machine (SVM) was used as a classifier. Visual sensors and smart armbands were used to collect gesture image information and EMG signals respectively, and the corresponding HOG features and time domain features were then extracted after preprocessing. These two features were fused at feature level by employing serial fusion. The combined SVM multiple classifier constructed by one Vs one was used to train and verify the gesture recognition model. Experimental results showed that the overall accuracy of 36 types of gesture recognition models reached 96% by using the proposed fusion features, which is 33% and 16% higher than the single hog feature and EMG time-domain feature before fusion, respectively. Compared with decision-level fusion, the accuracy of the proposed method was increased by 11%, and the calculation time was only 0.274 ms, which could effectively reduce the amount of feature data and significantly improve the accuracy of multiple types of gesture recognition.

Key words: gesture recognition; HOG; EMG; time-domain features; fusion features