DOI:10.13878/j.cnki.jnuist.2021.03.004

熊鹏文[1]　熊昆[1]　张宇[1]　余斯吉[1]

基于三维骨骼信息的动态手势识别

摘要

手势识别作为人机交互的有效手段,成为当前研究的热点话题.针对动态手势识别存在时空多变性、特征复杂性等问题,本文提出了一种基于三维骨骼信息的动态手势识别方法.动态手势具有时间上的差异性和复杂性,极大地影响了动态手势识别的准确率.因此,本文设计了一种动态手势关键帧提取算法,该算法可以提取动态手势关键部分,用于进一步的特征提取.另外,单独分类器的分类效果存在差异性,本文采用多个分类器同时对手势特征进行分类,充分利用了所提取的特征.同时,本文还提出了一种自适应融合算法,可以根据分类精度有效融合不同分类器,提高最终分类效果.最后,通过实验验证了本文提出的动态手势识别框架和方法的有效性.

关键词

骨骼信息;动态手势识别;关键帧;多分类器融合

中图分类号 TP391

文献标志码 A

收稿日期 2021-03-05

资助项目 国家自然科学基金(61903175,61663027);江西省主要学科学术和技术带头人项目(20204BCJ23006);住房和城乡建设部 2020 年科学技术项目(2020-K-009)

作者简介

熊鹏文,男,博士,副教授,研究方向为机器人传感与控制技术.steven.xpw@ ncu.edu.cn

张宇(通信作者),男,博士,讲师,研究方向为智能感知与智能制造. holy _ duke @ 163.com

1　南昌大学　信息工程学院,南昌,330031

0　引言

随着社会的发展,传统的硬件输入设备已经无法满足人们对于人机交互的需要.手势识别作为人机交互的有效手段,成为当前研究的热点话题[1-2].手势识别的应用非常广泛,从手语识别到手势操控再到虚拟现实,它给人们提供了一种十分有效而且便利的人机交互方法.手势识别分为静态手势识别和动态手势识别.静态手势识别是指识别静止的手势图像,不具备时序性;动态手势由连续的手势序列组成,动态手势识别时需要克服其时空可变的特点.早期,基于视觉的动态手势识别方法是使用摄像头实现的[3-5].这些方法实现简单,但是容易受到光照等外界环境的影响,稳定性较差.近年来,各种低成本深度传感设备的发展给手势识别提供了新的机会.Tiwari 等[6]使用 Kinect 深度相机获取从 0 到 9 的人体深度图像,对手部轮廓进行离散余弦变换(DCT)处理得到符号数据集,再利用神经网络进行分类并识别,平均准确率达到 83.5%.Sun 等[7]介绍了基于 Kinect 和 sEMG 信号的手势识别方法.针对汽车内驾驶员的手势识别,Molchanov 等[8]提出一种基于近程雷达、彩色摄像机和深度摄像机的手势识别系统,该系统对于不同光照情况下的手势识别都有较好的鲁棒性.Marin 等[9]提出联合利用深度相机和 Leap Motion 传感器进行手势识别,两者相结合获取的最佳准确率为 96.5%.Hakim 等[10]收集深度摄像头和 RGB 摄像头的组合数据,实现了高精度的手势识别.

为了充分利用深度传感器获取的三维骨骼信息,本文提出了一种基于三维骨骼信息的动态手势识别方法.该方法先使用动态手势关键帧提取算法,从不同长度的动态手势中提取手势的关键部分;然后利用一组动态手势特征序列来有效表示动态手势的运动特征;最后通过提出的自适应融合算法用于融合多分类器的分类结果,避免了单独分类器效果的差异性,获取最优分类效果.

1　动态手势识别方法

基于三维骨骼信息的动态手势识别方法框架如图 1 所示,主要步骤包括数据预处理、特征提取和多分类器融合识别.该动态手势识别方法是基于三维骨骼信息的,输入为动态手势的三维关节坐标值.数据预处理包括动态手势的无效帧删除、手势关节长度的统一化和动态手势关键帧提取等步骤,目的是使输入的动态手势数据更加标准

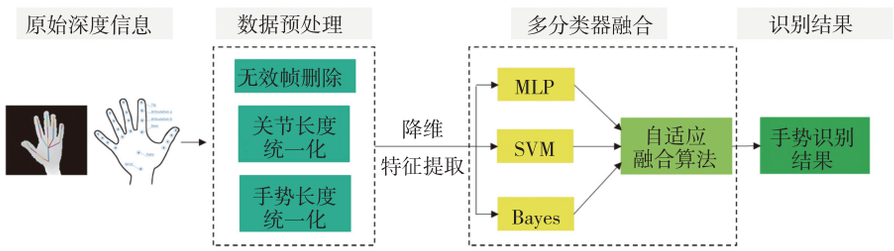


图 1　动态手势识别方法框架

Fig. 1　Framework of dynamic gesture recognition

化;特征提取是对预处理的动态手势数据进行空间特征和局部特征提取,减少冗余信息进而提高分类效果;多分类器融合识别是采用多个分类器对手势进行分类的方法减小单个分类器分类存在的偏差,并使用自适应融合算法对多个分类器分类结果进行自适应加权融合,得到最优的手势识别效果.

2　数据预处理

2.1　无效帧删除和关节长度统一化

动态手势具有复杂的时空可变性[11],同一用户和不同用户做同样的手势都存在不同的范围、幅度和速度的差异.另外,不同用户的手部大小与关节之间的长度也存在差异.为了减少手势无效帧对动态手势识别造成的影响,必须进行无效帧删除和关节长度统一化.

无效帧删除的关键在于动态手势序列起始帧与结束帧的标记,本文所使用的动态手势数据集已经对每个手势序列的起始帧与结束帧进行了有效标记,所以可以直接进行无效帧删除.针对手势骨骼坐标的个体差异性,需要对原始数据进行统一化处理.进行关节长度统一化可以有效提升动态手势识别准确率和识别方法泛化能力.本文采用 Z-score 方法进行统一化处理,将坐标转换到同一量级以消除不同个体之间的差异性.即

$$x' = \frac{x - \mu}{\delta}, \tag{1}$$

其中 μ 表示总体数据的平均值,δ 表示总体数据的标准差.

2.2　关键帧提取

动态手势的时空可变性使得动态手势时间序列存在较大的差异性,手势序列的长度长短不一.经过无效帧删除之后,本文使用的动态手势数据集手势序列长度最小为 7 帧,最大为 149 帧.手势序列长度差异较大,仍然存在冗余信息.因此,本文提出了一种动态手势关键帧提取算法,对动态手势序列进行处理并提取指定数量的关键帧.具体描述如算法 1 所示.

2.3　特征提取

本文使用的动态手势数据集中的三维骨骼信息包含了手掌 22 个关节的坐标数据,如图 2 所示.利用这些坐标数据可以直接得到或者计算出动态手势的掌心位置、手指弯曲、手指间距离和手指间角度等局部特征,也可以计算得到手掌运动速度、手掌关节旋转等全局特征.

算法 1.动态手势关键帧提取算法

输入:数据集 $T = \{ [(x_1^1,y_1^1,z_1^1),(x_2^1,y_2^1,z_2^1),\cdots,(x_{22}^1,y_{22}^1,z_{22}^1)], [(x_1^2,y_1^2,z_1^2),(x_2^2,y_2^2,z_2^2),\cdots,(x_{22}^2,y_{22}^2,z_{22}^2)],\cdots, [(x_1^n,y_1^n,z_1^n),(x_2^n,y_2^n,z_2^n),\cdots,(x_{22}^n,y_{22}^n,z_{22}^n)] \}$ 和提取的关键帧数量 m.

1)初始化提取的每帧信息间距 $t_{\text{step}} = n/m$.

2)对 22 个关节坐标分别计算 $(j = 1,2,\cdots,22)$:

①初始化信息间距 $t = 0$.

②计算关键帧数据,$s = 1,2,\cdots,m-1$:

a.信息间距 t 的整数部分和小数部分分别表示为 t_i 和 t_r;

b.计算平均变化的骨骼关节信息;

$(x_j^s,y_j^s,z_j^s) = (x_j^{t_i},y_j^{t_i},z_j^{t_i}) + [(x_j^{t_i+1},y_j^{t_i+1},z_j^{t_i+1}) - (x_j^{t_i},y_j^{t_i},z_j^{t_i})] \times t_r$;

c.更新信息间距 $t = t + t_{\text{step}}$.

③加入手势结束帧 $(x_j^m,y_j^m,z_j^m) = (x_j^n,y_j^n,z_j^n)$.

输出:m 帧 $T' = \{ [(x_1^1,y_1^1,z_1^1),(x_2^1,y_2^1,z_2^1),\cdots,(x_{22}^1,y_{22}^1,z_{22}^1)], [(x_1^2,y_1^2,z_1^2),(x_2^2,y_2^2,z_2^2),\cdots,(x_{22}^2,y_{22}^2,z_{22}^2)],\cdots, [(x_1^m,y_1^m,zm_1),(x_2^m,y_2^m,z_2^m),\cdots,(xm_{22},y_{22}^m,z_{22}^m)] \}$.

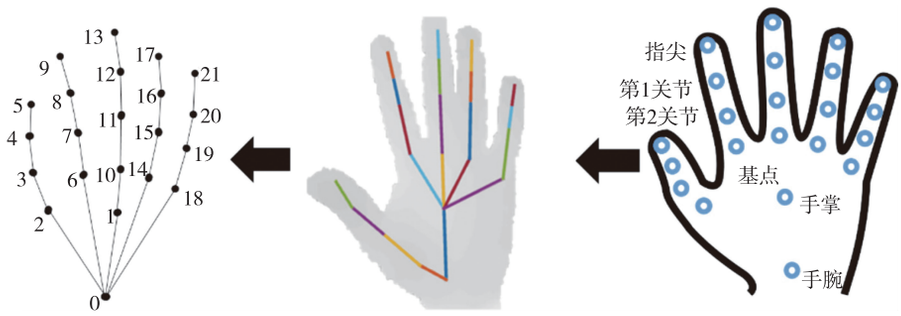


图 2　手部骨骼关节

Fig. 2　The joint diagram of hand skeleton

本文通过手势位移将数据集中手势类别分为滑动类手势和动作类手势两种.滑动类手势主要是手掌的平移而不涉及手指之间的动作,针对这个特点,本文主要提取动态手势的空间特征,包括手掌在空间中平移的距离和手掌平移的方向.动作类手势,它的特点是手掌的移动较少,主要是手指间的动作,包括手指之间的接触与分离.因此,本文选取手掌手指的运动(包括五指指尖与手腕的距离、拇指尖与其他四指指尖距离)和拇指与食指夹角角度作为手势局部特征表示. P_j^i 表示动态手势手掌在第 i 帧中第 j 个关节的坐标.

手掌平移的方向由手势相邻两帧之间手腕位置坐标连成的方向向量来确定.为了简化平移方向的表示,本文利用 8 个标准向量将平面空间平均分为 8 个方向,8 个标准向量分别为 [1,0,0]、[1,1,0]、[0,1,0]、[-1,1,0]、[-1,0,0]、[-1,-1,0]、[0,-1,0]、[1,-1,0],对应的 8 个方向分别用 1~8 的数字表示.分别计算平移方向向量与 8 个标准向量的夹角大小,夹角最小的标准向量方向表示为平移方向.即

$$\boldsymbol{T}=\boldsymbol{P}_0^{i+1}-\boldsymbol{P}_0^i,$$

$$\cos\theta_i=\frac{\boldsymbol{T}\cdot\boldsymbol{E}_i}{|\boldsymbol{T}|\cdot|\boldsymbol{E}_i|},$$

$$V=\underbrace{\min}_{i}\{\cos\theta_1,\cos\theta_2,\cdots,\cos\theta_i\},\tag{2}$$

其中:$\boldsymbol{T}$ 表示平移方向向量;$\boldsymbol{E}_i$ 表示标准向量,$i=1,2,\cdots,8$;V 表示平移方向.

手掌拇指与食指夹角角度可以通过计算拇指方向向量和食指方向向量得到,拇指方向向量由拇指关节 3 和 5 连接表示,食指方向向量由食指关节 7 和 9 连接表示.即

$$\boldsymbol{T}_1=\boldsymbol{P}_5^i-\boldsymbol{P}_3^i,\quad \boldsymbol{T}_2=\boldsymbol{P}_9^i-\boldsymbol{P}_7^i,$$

$$A=\arccos\left(\frac{\boldsymbol{T}_1\cdot\boldsymbol{T}_2}{|\boldsymbol{T}_1|\cdot|\boldsymbol{T}_2|}\right).\tag{3}$$

3　多分类器融合识别

3.1　基本分类器

原始骨骼信息经过数据预处理和特征提取以后输入多个基本分类器进行分类,再利用自适应融合算法对多个分类结果进行加权融合,得到最优分类结果.利用多个分类器进行分类可以充分利用不同特点的基本分类器,不仅可以避免单个分类器的偏差性问题,还可以提高分类算法的泛化能力,提高手势识别准确率.基本分类器选取原则是选取不同分类原理的分类器,以便更好地提高融合效果,增加手势分类泛化能力.本文选择的 3 个基本分类器分别为支持向量机(Support Vector Machine,SVM)、朴素贝叶斯(Naive Bayes,NB)和多层感知器(Multi-Layer Perception,MLP).

3.2　自适应融合算法

传统的融合算法采用人工设定多分类器的权重系数进行加权融合,容易受到经验的影响,降低分类效果.因此,本文提出一种自适应融合算法实现多分类器融合,可以根据不同分类器分类结果自动更新融合权重,有效减少了人为因素的干扰,获取最优识别结果,结构如图 3 所示.自适应融合算法的具体描述见算法 2.

4　实验结果与分析

4.1　实验环境

本文实验运行环境为 Intel(R) Core(TM) i5-6300HQ CPU @ 2.3GHz,运行内存 8 GB,操作系统为 Windows10(64 位),主要识别模型由 Python 语言编写,使用的开发软件为 PyCharm2019.本文采用的

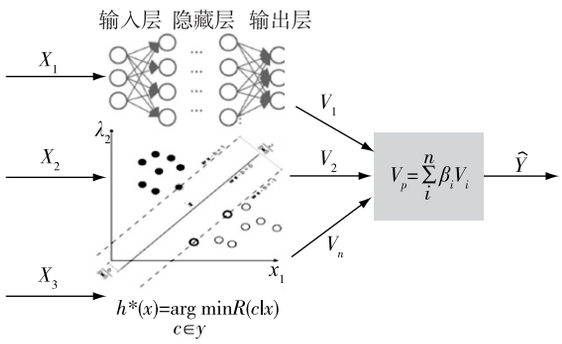


图 3　自适应融合算法结构

Fig. 3　Structure of adaptive fusion algorithm

实验数据集为 DHG-14/28 动态手势数据集，该数据集共有抓握、按等 14 种手势类别，包括 Fine 型和 Coarse 型两大类.每种手势类别又可以根据执行方式细分为两类：使用一根手指和整只手.数据集中每种动态手势都由 20 位参与者按照上述两种方式执行 5 次，总共包括 2 800 个手势序列.

4.2　动态手势关键帧帧数选取

进行动态手势关键帧提取之前必须确定关键帧帧数，不同的关键帧数对于动态手势识别效果有不同的影响.因此，使用不同的关键帧数 n 进行实验对比，选择识别准确率最高的帧数进行进一步分类.如图 4 所示，在不同关键帧下不同种类手势的识别准确率出现了明显变化.实验结果表明，当关键帧数不断增加，识别准确率呈现下降趋势.当 $n=10$ 时，对于数据集中 28 种手势分类准确率最高达到 81.57%.

算法 2.自适应融合算法

输入：3 个类别概率向量 $\boldsymbol{P}_1,\boldsymbol{P}_2,\boldsymbol{P}_3$，分别代表 3 个基本分类器分类预测的类别概率向量.

1）建立 3 个概率矩阵 $\boldsymbol{E}_1,\boldsymbol{E}_2,\boldsymbol{E}_3$，分别代表 3 个基本分类器的分类预测概率矩阵；矩阵大小为 $M\times N$，M 代表分类的动态手势样本总数，N 代表动态手势类别总数.

$$\boldsymbol{E}_1=\begin{bmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & & \vdots \\ a_{M,1} & \cdots & a_{M,N} \end{bmatrix},\quad \boldsymbol{E}_2=\begin{bmatrix} b_{1,1} & \cdots & b_{1,N} \\ \vdots & & \vdots \\ b_{M,1} & \cdots & b_{M,N} \end{bmatrix},\quad \boldsymbol{E}_3=\begin{bmatrix} c_{1,1} & \cdots & c_{1,N} \\ \vdots & & \vdots \\ c_{M,1} & \cdots & c_{M,N} \end{bmatrix}.$$

2）利用概率矩阵计算得出各分类器的类别权重向量 $\boldsymbol{w}_1,\boldsymbol{w}_2,\boldsymbol{w}_3$：

$$\boldsymbol{w}_1=[w_1^1,w_2^1,\cdots,w_N^1],$$
$$\boldsymbol{w}_2=[w_1^2,w_2^2,\cdots,w_N^2],$$
$$\boldsymbol{w}_3=[w_1^3,w_2^3,\cdots,w_N^3],$$

其中 w_j^1,w_j^2,w_j^3 分别代表 3 个分类器在类别 j 的类别权重，类别权重由下式可得：

$$w_j^1=\frac{\sum_{i=1}^{M}\sum_{k=1}^{N} a_{i,k}/(MN)}{\sum_{i=1}^{M} a_{i,j}/M},\quad w_j^2=\frac{\sum_{i=1}^{M}\sum_{k=1}^{N} b_{i,k}/(MN)}{\sum_{i=1}^{M} b_{i,j}/M},\quad w_j^2=\frac{\sum_{i=1}^{M}\sum_{k=1}^{N} c_{i,k}/(MN)}{\sum_{i=1}^{M} c_{i,j}/M}.$$

3）计算标准化类别权重向量 $\boldsymbol{W}_1,\boldsymbol{W}_2,\boldsymbol{W}_3$，即

$$\boldsymbol{W}_i=\frac{\boldsymbol{w}_i}{\boldsymbol{w}_1+\boldsymbol{w}_2+\boldsymbol{w}_3},i=1,2,3,$$

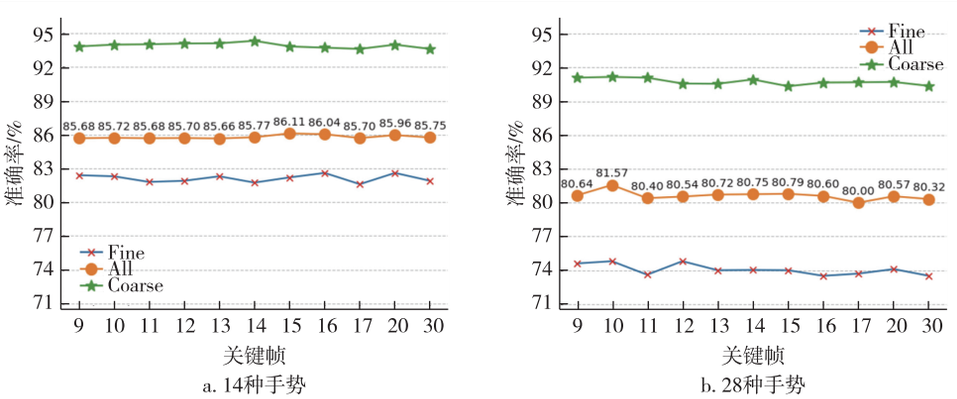
其中向量之间的除法代表向量对于元素的除法.

4）计算融合概率向量 $\boldsymbol{P}$，即

$$\boldsymbol{P}=\boldsymbol{W}_1\times S_1\times\boldsymbol{P}_1+\boldsymbol{W}_2\times S_2\times\boldsymbol{P}_2+\boldsymbol{W}_3\times S_3\times\boldsymbol{P}_3,$$

其中 S_1,S_2,S_3 分别代表 3 个基本分类器的动态手势样本分类准确率，向量之间的乘法代表向量对于元素相乘.

输出：最优分类结果 F，即融合概率向量 $\boldsymbol{P}$ 中概率最大值对应的动态手势类别.



a. 14种手势　　b. 28种手势

图 4　14/28 种不同手势在 n 个关键帧时的识别准确率

Fig. 4　Recognition accuracy of 14 (a) and 28 (b) gestures at n keyframes

因此,根据准确率的变化,可以确定本文中关键帧提取帧数选取为 10 帧.

4.3 多分类器融合对比

为了验证多分类融合识别方法可以有效减少单个分类器偏差性和增强手势识别泛化能力,本文在相同条件下对比了单个分类器与多分类器融合之间手势识别效果.

由图 5 可以看出,单分类器中 SVM 分类器效果优于 MLP 分类器和 Bayes 分类器,主要原因在于 SVM 是一种基于统计理论的分类方法,对于选取的各类手势特征数据有更好的分割效果.多分类器融合分类准确率普遍高于单分类器,说明多分类器融合分类可以减少偏差性,从而提高分类效果.

4.4 不同动态手势识别方法对比

为了验证本文提出的动态手势识别方法的有效性,将本文的方法与其他手工提取特征的方法进行了比较,得出的实验结果如表 1 所示.由表 1 可以看出,本文方法与文献[11]同样基于骨骼信息进行手势识别,但是分类性能平均增加 5.19 个百分点.实验结果表明,本文方法相比传统手工特征方法在 Fine、Coarse 和所有手势方面都实现了更高的识别准确率,其中 14 种手势识别准确率达到 85.91%,28 种手势识别准确率达到 81.57%.

表 1 实验结果比较

Table 1 Recognition accuracy comparison between Skeleton-Based method and the proposed method %

方法	DHG-14			DHG-28		
	Fine	Coarse	All	Fine	Coarse	All
Skeleton-Based[11]	73.60	88.30	83.07	68.20	86.30	79.14
本文方法	82.30	94.00	85.91	74.80	91.17	81.57

为了进一步说明本文提出的方法的性能,图 6 中显示了 28 种手势分类结果混淆矩阵.在矩阵当中,x 轴代表分类结果,而 y 轴代表真实结果.序号 1 到 28 分别代表使用一根手指和整只手进行 14 种手势的 28 种结果,其中 1 和 2 分别代表使用一根手指和整只手进行的抓握动作,其他序号类似.由图 6 可以看出,序号 1 和 7、2 和 8 之间混淆最为严重,有 27%的手势 1 被分类为手势 7,33%的手势 2 被分类为手势 8,在文献[12-14]中也存在类似结果,这说明抓握手势和捏手势运动特征较为相似导致难以分类.另外,文献[12-13]存在将向上滑动手势与展开手势混淆的现象,表明其对于手部展开的局部特征提取存在缺失.本文提出了手掌手指运动和拇指与食指夹角角度等特征,可以有效表示手部的局部运动.同时,由图 6 可以看出,本文提出的方法对于同一种手势不同执行方式之间的分类误差较小,这表明局部特征的表示可以避免一根手指和整只手两种执行方式之间的混淆.对于其他手势,结果显示没有较为严重的混淆现象,最终验证了本文所提出动态手势运动特征以及多分类器融合分类的有效性.

5 结论

本文提出了一种基于三维骨骼信息的动态手势识别方法,该方法包含动态手势关键帧提取和多分类器自适应融合算法.动态手势关键帧提取可以有效提取出动态手势的关键部分,避免了动态手势信息的冗余.多分类器自适应融合算法可以充分利用多个基本分类器的信息,减少个体差异性,提高分类方法的泛化能力.实验结果表明,本文方法比现有的方法具有更好的性能,验证了本文所提出的动态手势识别方法的有效性.

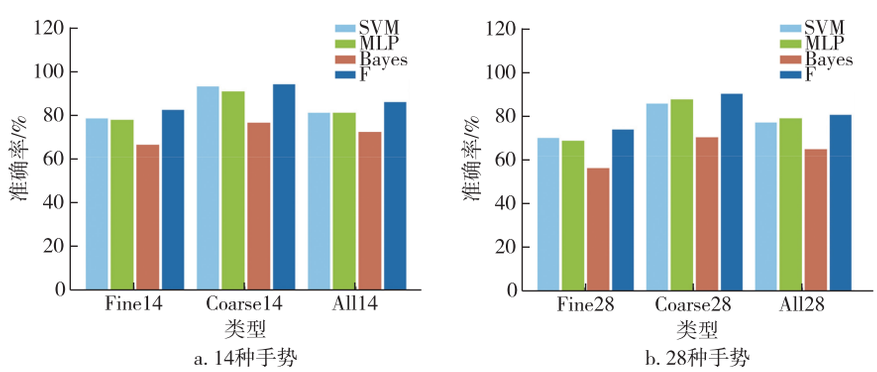


图 5 单分类器与多分类器融合识别效果对比

Fig. 5 Comparison of 14 (a) and 28 (b) gestures recognition accuracy between single classifier and multi-classifier fusion

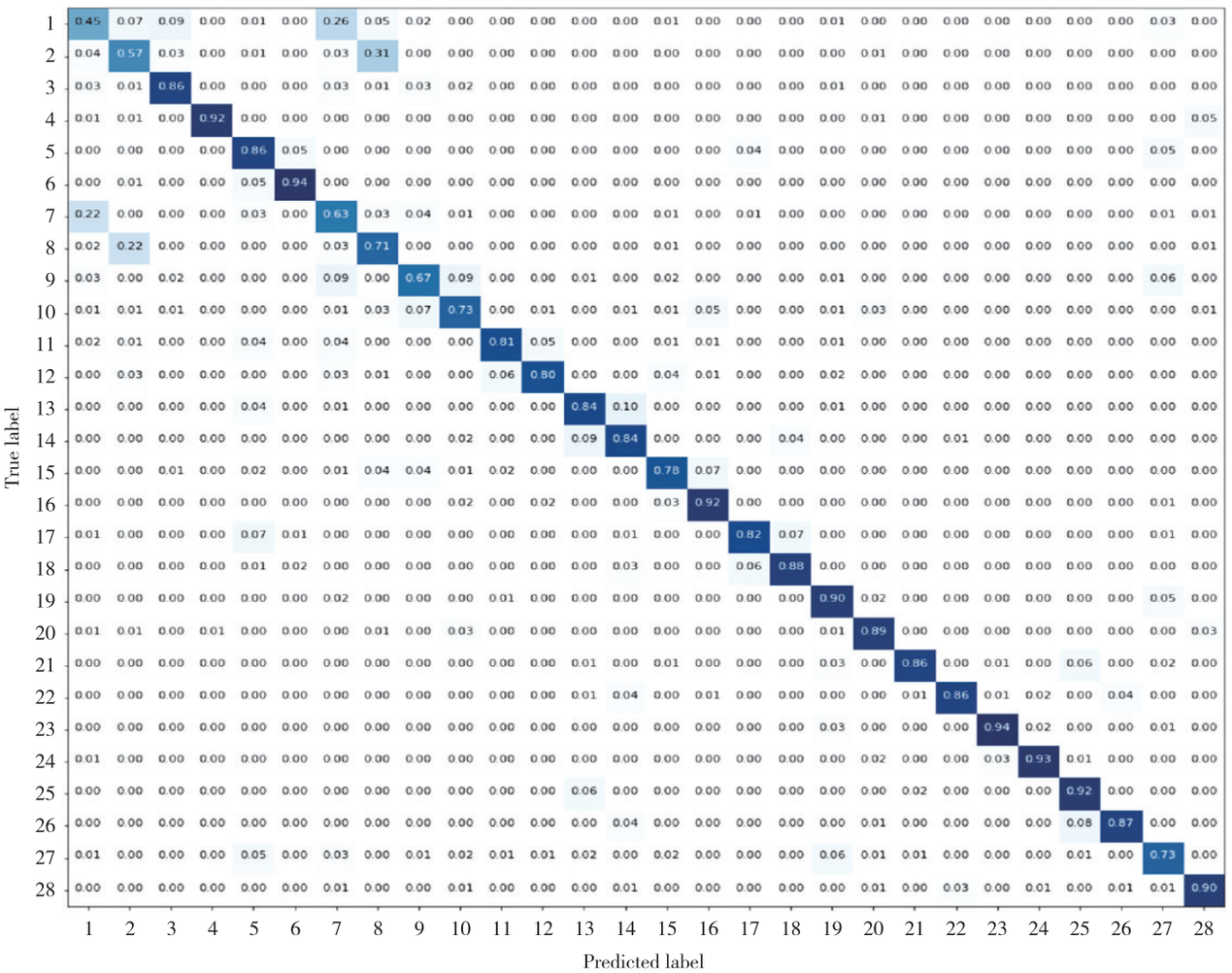


图 6　28 种手势分类混淆矩阵

Fig. 6　Confusion matrices of 28 gesture classification

参考文献

References

[1] Narayana P, Beveridge J R, Draper B A. Gesture recognition: focus on the hands [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 5235-5244

[2] Wang T Q, Li Y D, Hu J F, et al. A survey on vision-based hand gesture recognition [M] // Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 219-231

[3] Yoon H S, Soh J, Bae Y J, et al. Hand gesture recognition using combined features of location, angle and velocity [J]. Pattern Recognition, 2001, 34(7): 1491-1501

[4] 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). April 10 - 12, 2017, Taichung, Taiwan, China. IEEE, 2017: 371-375. DOI: 10.1109/IRC.2017.40

[5] Tan T D, Guo Z M. Research of hand positioning and gesture recognition based on binocular vision [C] // 2011 IEEE International Symposium on VR Innovation. March 19-20, 2011, Singapore. IEEE, 2011: 311-315. DOI: 10.1109/ISVRI.2011.5759657

[6] 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). August 10 - 13, 2015, Kochi, India. IEEE, 2015: 194-198. DOI: 10.1109/ICACCI.2015.7275608

[7] Sun Y, Li C Q, Li G F, et al. Gesture recognition based on kinect and sEMG signal fusion [J]. Mobile Networks and Applications, 2018, 23(4): 797-805. DOI: 10.1007/s11036-018-1008-0

[8] Molchanov P, Gupta S, Kim K, et al. Multi-sensor system for driver's hand-gesture recognition [C] // 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). May 4-8, 2015, Ljubljana, Slovenia. IEEE, 2015: 1-8. DOI: 10.

1109/FG.2015.7163132
[9] Marin G, Dominio F, Zanuttigh P. Hand gesture recognition with jointly calibrated leap motion and depth sensor[J]. Multimedia Tools and Applications, 2016, 75(22):14991-15015. DOI:10.1007/s11042-015-2451-6
[10] Hakim N L, Shih T K, Kasthuri Arachchi S P, et al. Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model[J]. Sensors, 2019, 19(24):5429. DOI:10.3390/s19245429
[11] De Smedt Q, Wannous H, Vandeborre J P. Skeleton-based dynamic hand gesture recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). June 26 - July 1, 2016, Las Vegas, NV, USA. IEEE, 2016: 1206-1214. DOI: 10.1109/CVPRW.2016.153
[12] Núñez 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. DOI: 10.1016/j.patcog.2017.10.033
[13] Chen X H, Guo H K, Wang G J, et al. Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition[C]//2017 IEEE International Conference on Image Processing (ICIP). September 17 - 20, 2017, Beijing, China. IEEE, 2017: 2881-2885. DOI: 10.1109/ICIP.2017.8296809
[14] Lai K, Yanushkevich S N. CNN+RNN depth and skeleton based dynamic hand gesture recognition[C]//2018 24th International Conference on Pattern Recognition (ICPR). August 20 - 24, 2018, Beijing, China. IEEE, 2018: 3451-3456. DOI: 10.1109/ICPR.2018.8545718

Dynamic gesture recognition based on 3D skeleton information

XIONG Pengwen[1] XIONG Kun[1] ZHANG Yu[1] YU Siji[1]
1 School of Information Engineering, Nanchang University, Nanchang 330031

Abstract As an effective means of human-computer interaction, gesture recognition has become a hot topic in current research. In order to solve the problems of spatio-temporal variability and feature complexity concerning dynamic gestures, we propose a dynamic gesture recognition solution based on 3D skeleton features. The accuracy of dynamic gesture recognition is greatly impaired due to the temporal differences and complexity of dynamic gestures, thus a key frame extraction algorithm is designed to extract key features of dynamic gestures for further feature extraction. To overcome the difference in classification performance between single classifiers, multiple classifiers are used to simultaneously classify and fully exploit gesture features. We also propose an adaptive fusion algorithm to effectively fuse different classifiers according to their classification performances thus improve the final classification accuracy. Finally, experiments are carried out, and results verify the effectiveness of the proposed dynamic gesture recognition approach.

Key words skeleton information; dynamic gesture recognition; key frame; multi-classifier fusion