

## 基于自学习稀疏表示的动态手势识别方法

肖玲<sup>1,2</sup>, 李仁发<sup>1</sup>, 曾凡仔<sup>2</sup>, 屈卫兰<sup>2</sup>

(1. 湖南大学 嵌入式与网络计算湖南省重点实验室, 湖南 长沙 410082; 2. 湖南大学 信息科学与工程学院, 湖南 长沙 410082)

**摘要:** 针对加速度传感器的手势采集方式提出一种基于自学习稀疏表示的动态手势识别方法。该方法将分类识别问题转化为求解待识别样本对于训练样本的稀疏表示问题, 直接对原始加速度信号进行操作, 省去了特征提取过程, 可方便地添加新的手势类别和删除已有的手势类别; 利用面向类别的字典学习, 来寻求一个较小的并经过优化的超完备字典来计算待识别样本的稀疏表示, 从而大大缩减算法的计算复杂度, 满足实时性要求。在包含 18 种手势的 3 000 多个样本的公开数据集上进行测试, 实验结果验证了该方法的有效性。

**关键词:** 手势识别; 稀疏表示; 字典学习; 加速度传感器

中图分类号: TP391

文献标识码: B

文章编号: 1000-436X(2013)06-0128-08

## Gesture recognition approach based on learning sparse representation

XIAO Ling<sup>1,2</sup>, LI Ren-fa<sup>1</sup>, ZENG Fan-zai<sup>2</sup>, QU Wei-lan<sup>2</sup>

(1. Embedded System & Networking Laboratory of Hunan University, Changsha 410082, China;

2. School of Information Science and Engineering, Hunan University, Changsha 410082, China)

**Abstract:** An approach of robust accelerometer-based hand gesture recognition based on self-learning sparse representation was proposed. This method operated directly on the original acceleration signals by sparse representation without feature extraction and used the class-specific dictionary learning for sparse modeling to reduce the computing cost and time of recognition. The proposed system can easily add a novel gesture category as well as remove existing ones. Experiments on real-world database of 18 hand gestures validate the availability of the proposed algorithm.

**Key words:** hand gesture recognition; sparse representation; dictionary learning; accelerometer

### 1 引言

手势作为一种自然而直观、易于学习的人机交互方式一直是人机交互领域的研究热点。手势动作的数据获取方式有基于数据手套、基于视觉、基于肌电信号和基于惯性传感器等多种手段。基于数据手套的手势识别要求用户穿戴复杂的数据手套, 给用户带来较大不便, 另外, 因为输入设备比较昂贵, 难以大范围推广。基于肌电信号的手势识别对采集设备有较高要求, 现在仅限于实验室研究。目前的手势识别研究大多以基于视觉的为主, 但该方法需要处理的数据量大, 处理方法复杂, 不太适合实时

地识别, 且识别性能受光线、背景、拍摄角度的影响较大。随着微型机电系统 MEMS 技术的发展与成熟, 加速度传感器、陀螺仪等惯性测量单元以其低廉的价格、较小的体积和较高的灵敏度广泛地应用于手机、PDA、游戏机等嵌入式手持移动设备中。目前的智能手机上大多都配置了三轴加速度传感器。用户只需要手握嵌入式加速度传感器的移动设备, 就可直接测量手势动作信息。基于加速度传感器的手势识别成为了新一代人机交互的研究热点问题之一。

基于微惯性传感器的手势识别属于模式识别领域的一个新兴分支, 其本质是通过惯性传感器

收稿日期: 2012-11-02; 修回日期: 2013-01-28

基金项目: 国家自然科学基金资助项目(61173036, 61173012, 61272061); 湖南省自然科学基金资助项目(12JJA005)

**Foundation Items:** The National Natural Science Foundation of China (61173036, 61173012, 61272061); The Key Natural Science Foundation of Hunan Province (12JJA005)

获取手势动作时产生的运动信号, 对手势进行分类和识别。基于加速度传感器的手势识别中, 仅有少量研究提到了特征提取和特征选择方法<sup>[1~3]</sup>。文献[1]利用傅立叶变换和小波分析等常用方法来提取手势加速度信号特征, 并考察了不同特征提取方法对识别率的影响。文献[3]针对特定的手势类别分析其特性, 提取了特定的特征, 其优点是识别方法简单, 但是该识别方法只限定在其指定的几种手势类别, 适用性不强。目前大多手势识别研究都不提取任何特征, 而是把原始的手势加速度信号直接送给分类器识别。由于手势动作与语音信号类似, 具有时空变化性和时序相关性, 因此研究者们把语音识别中常用的方法用于基于加速度传感器的手势识别中。具有代表性的有隐式马尔可夫模型(HMM, hidden Markov models)和动态时间规整(DTW, dynamic time warping)技术。基于 HMM 模型的识别方法<sup>[4~7]</sup>在训练阶段需要提供大量的训练数据, 通过反复计算才能得到模型参数, 导致计算量大。而目前的基于加速度的手势识别没有像语音识别那样具有标准的大型语料库, 所以对于研究者而言, HMM 方法很难收集大量的训练样本来得到准确的模型参数。DTW 是一种非线性时间规整模式匹配方法, 它把待识别样本与模版进行比较, 计算两者的相似度, 将与待识别样本最相似的模版类别作为待识别样本的类别。文献[8]在训练和测试阶段都采用了 DTW 技术。在训练阶段, 把 DTW 计算出的样本的相似度利用近邻传播(affinity propagation)进行聚类; 在测试阶段利用 DTW 和压缩感知进行分类。Uwave 手势识别系统<sup>[9]</sup>就是利用 DTW 技术进行模版匹配, 但是它因为是基于个体相关的, 使它的适用范围受限。DTW 的问题是识别性能过分依赖于端点检测, 且不能对样本进行动态训练, 所以 DTW 适合于个体相关的和样本集较小的识别, 当样本数量大时导致 DTW 的运算量很大而不适合实时识别。也有研究者尝试采用了机器学习和模式识别中常用的方法, 如支持向量机(SVM<sup>[2]</sup>, support vector machine)、神经网络(neural network)<sup>[10,11]</sup>、模糊模式识别<sup>[12]</sup>等进行手势识别。

近年来, 信号处理领域的稀疏表示和压缩感知理论在图像处理、模式识别与机器学习等领域得到广泛关注。Wright 等人<sup>[13]</sup>把稀疏表示引入到人脸识别中, 提出了基于稀疏表示分类(SRC, sparse rep-

resentation classification) 方法, 为稀疏表示理论在模式识别的应用树立了一个良好的开端。SRC 的基本思想是: 将所有目标的人脸图像样本构建成一个冗余字典, 通过寻求待识别的人脸图像样本在冗余字典中的稀疏表示(表示系数大部分为零, 不为零的部分应在同类样本上的表示系数), 将人脸识别问题转化为稀疏表示问题, 并运用压缩感知理论对该问题进行求解。但是 SRC 方法采用将所有的训练图像构建成一个冗余字典, 导致了冗余字典的尺寸巨大, 使得 SRC 的稀疏求解时计算非常耗时。

字典学习也是近年稀疏表示领域中的热点问题。字典中的元素被称为原子。字典原子的选择尽可能好地符合被逼近信号。字典学习就是从字典中找到具有最佳线性组合的原子来表示信号。具有代表性的 K-SVD 算法<sup>[14]</sup>可从训练信号集中学习出用于稀疏编码的小型超完备字典。

本文结合 SRC 分类方法和字典学习, 提出了基于自学习稀疏表示的手势识别(L-SRC, learning sparse representation classification)算法, 可有效解决手势识别中的几个问题。

#### 1) 无需特征提取

稀疏表示利用字典的冗余特性捕捉原始信号的自然特性, 直接将传感数据样本作为训练集, 大量的原始特征值将成为算法中可利用的优点, 省去了特征提取过程。基于稀疏表示的手势识别也不需要训练样本进行统计训练, 可以方便地增加或删除手势类别。

#### 2) 有效解决手势信号长度不一

由于人体差异, 不同人执行同一个手势动作, 其动作幅度和执行速度有一定的差别, 即使是同一个人多次执行同一动作也会有差异, 导致采集的手势信号长度不一。本文采用将手势样本向量长度归一化线性插值的方法来解决这一问题。作为构成手势数据的原始特征, 手势样本的数据长度本身也可作为手势的一个重要特征, 而稀疏表示可同时利用这个特征来进行手势识别。

#### 3) 手势识别的实时性

针对每个手势类, 进行单独的字典学习。通过字典学习得到一个较小的超完备字典, 用来计算测试样本的稀疏表示。字典学习是在手势训练阶段离线进行的, 训练的字典可为手势识别阶段缩减识别算法的计算复杂度, 以适应实时性要求。

## 2 基于稀疏表示的手势识别方法

### 2.1 问题描述

从几何的角度看,很多待识别的模式类可以由特定的子空间刻画,每一个子空间代表一个类别。对于动态手势识别,三轴加速度传感器测量的手势所形成的运动轨迹近似分布在不同的子空间中。假设多种不同的手势轨迹空间分布满足混合子空间模型<sup>[15]</sup>,图 1 显示了 8 种不同手势的运动轨迹近似分布在不同的子空间。对于动态手势识别,只要选择适当的手势满足混合子空间模型,每个子空间代表一个手势类别,则手势识别问题可以转化为稀疏表示分类模型。

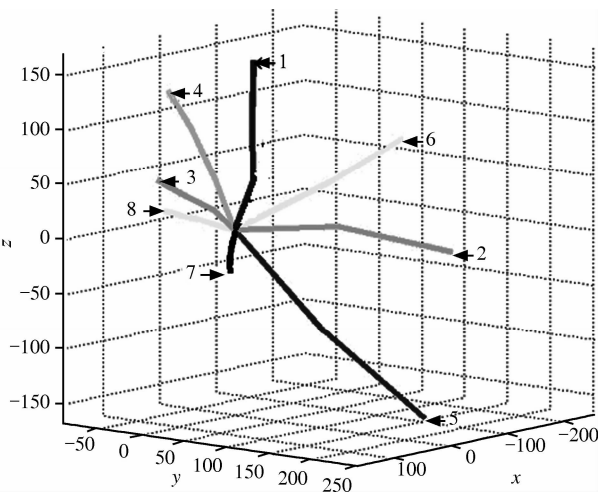


图 1 8 种不同手势的运动轨迹空间分布

采集的一次动态手势样本为三轴加速度数据的时序序列,可用一个高维向量来表示。用 $(x, y, z)$ 表示三轴加速度传感器分别在 $x, y, z$ 3个方向轴测量到的加速度值,则在 $t$ 时刻,加速度传感器的采样值可记为

$$s(t) = (x(t), y(t), z(t)) \quad (1)$$

对应一个完整的手势在持续一小段时长内所获取的 $l$ 次采样值构成了一个一维向量,记为

$$\mathbf{g} = (s(1), s(2), \dots, s(l))^T \quad (2)$$

手势识别问题为在给定的 $K$ 种手势样本的训练集中,判断一个待识别的手势轨迹向量 $\mathbf{g}_{k,\text{test}}$ 所属的手势类别。

### 2.2 稀疏表示分类模型

为识别 $\mathbf{g}_{k,\text{test}}$ 所属的手势类别, SRC 方法将整个训练集中所有 $C$ 种手势的所有轨迹向量一一作为

基向量,组成冗余字典矩阵 $\mathbf{G}$ ,其形式为 $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_k] = (\mathbf{g}_{1,1}, \mathbf{g}_{1,2}, \dots, \mathbf{g}_{c,n_c})^T \in \mathbb{R}^{m \times n}$  (3)

其中,  $n = n_1 + n_2 + \dots + n_K$ 。

考虑到实际测量过程中不可避免地会包含一些误差,当要用字典 $\mathbf{G}$ 来表示出待测的手势轨迹向量 $\mathbf{g}_{k,\text{test}}$ 时,可以用式(4)来表示。

$$\mathbf{g}_{k,\text{test}} = \mathbf{G}\mathbf{a} + \boldsymbol{\varepsilon} \in \mathbb{R}^m \quad (4)$$

其中,  $\mathbf{a} = [0, \dots, 0, a_{k,1}, a_{k,2}, \dots, a_{k,n}, 0, \dots, 0]^T \in \mathbb{R}^n$ ,  $\mathbf{a}$ 的系数中只有与第 $k$ 类对应的不为0,其余都为0,可见 $\mathbf{a}$ 是一个具有 $1/C$ 稀疏度的稀疏向量; $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ 为观测噪声。根据压缩感知理论,式(4)可采用二次约束下极小化 $l_1$ 范数的方法进行求解。

$$\hat{\mathbf{a}} = \arg \min \|\mathbf{a}\|_1, \text{ s. t. } \|\mathbf{g}_{k,\text{test}} - \mathbf{G}\mathbf{a}\|_2 \leq \varepsilon \quad (5)$$

在理想情况下,如果待测动作向量 $\mathbf{g}_{k,\text{test}}$ 是属于第 $k$ 类,则解向量 $\mathbf{a}$ 中只有与第 $k$ 类训练数据对应的元素不为零,其他元素都为0。由于噪声的存在,实际情况中非零元素也可能会发生在其他位置。则计算该待测动作向量与每一类中所有训练动作向量的线性加权的差值 $r_i(\mathbf{g}_{k,\text{test}})$ ,如式(6)。

$$r_i(\mathbf{g}_{k,\text{test}}) = \|\mathbf{g}_{k,\text{test}} - \mathbf{G}\delta_i(\hat{\mathbf{a}})\|_2, i = 1, 2, \dots, k \quad (6)$$

其中,  $\delta_i(\hat{\mathbf{a}})$ 表示提取稀疏表示系数 $\hat{\mathbf{a}}$ 中与第 $i$ 类的所有训练动作向量对应的系数,而其余的系数均为0。选择差值最小的 $r_i(\mathbf{g}_{k,\text{test}})$ 所属的类别即为最终的识别结果。

### 2.3 手势采样数据长度不一的问题

在2.2节中的冗余字典集 $\mathbf{G}$ 中要求所有向量具有相同的维数 $m$ ,但是手势信号具有长度不一的特点。不同的手势,甚至是同一个人做同种手势所持续的时间也可能不同,所以导致获取的加速度采样值的长度 $l$ 也会不同,即训练字典集 $\mathbf{G}$ 中的手势轨迹向量和待识别手势,轨迹向量具有不同的维数。考虑到稀疏表示手势识别算法要求所有手势样本的数据向量具有相同的维数,因此有必要把 $\mathbf{g}_{i,1}, \mathbf{g}_{i,2}, \dots, \mathbf{g}_{i,n_i}$ 和测试向量 $\mathbf{g}_{k,\text{test}}$ 转换成具有相同维数的一维向量。

本文采用将手势样本向量线性插值的方法来解决样本采集时长不一致的问题。首先找出所有手势轨迹向量中的最大采样长度 $l_{\max}$

$$l_{\max} = \max \{l_{\text{test}}, l_1, \dots, l_L\} \quad (7)$$

其中,  $L$ 表示在训练集 $\mathbf{G}$ 中所有手势轨迹向量的数目,  $l_1, \dots, l_L$ 表示每个手势轨迹向量的采样值长度。

接下来,对所有采样值长度小于  $l_{\max}$  的手势样本进行长度归一化线性插值。因为一次采样值有  $x$ 、 $y$ 、 $z$  3 个方向的加速度值,对每个方向的加速度值分别进行线性插值,即

$$g(x) = f(x_0) \frac{x - x_1}{x_0 - x_1} + f(x_1) \frac{x - x_0}{x_1 - x_0} \quad (8)$$

其中,  $f(x_0)$ 、 $f(x_1)$  分别表示第  $x_0$  和第  $x_1$  时刻的采样值,  $g(x)$  表示  $f(x_0)$  和  $f(x_1)$  内插得到的第  $x$  时刻的采样值。

最终使得训练字典集  $\mathbf{G}$  中的手势轨迹向量和待识别手势轨迹向量都具有相同的长度。

### 3 字典学习

在 2.2 节中基于稀疏表示的手势识别是将所有训练样本作为稀疏表示的冗余字典。在理论上,对于每个测试样本,计算其稀疏表示的时间复杂度为  $O(t^2n)$ ,  $n$  为训练样本数,  $t$  为所求系数向量中非零元素的个数。但在实际中,所求的系数向量并非最佳的稀疏表示向量,其中包含很多数值很小的非零向量,使得时间复杂度趋近  $O(n^3)$ 。所以当手势数据集中训练样本的数目比较大时, SRC 算法的计算复杂度成为限制其优良性能的瓶颈,致使 SRC 无法用于实时手势识别。

为克服 SRC 所面对的计算复杂度问题, Li 等<sup>[16]</sup>提出了一种利用测试样本的  $k$  个邻近样本来计算测试样本的局部稀疏表示分类器算法。但是当邻近的样本数较少时,邻近样本集并不满足超完备的要求。最近提出的一种字典训练方法 K-SVD<sup>[14]</sup>,通过迭代不断修正样本的稀疏编码,实现字典的动态更新,得到能更好地表示样本的字典。

本文根据手势识别的实时性要求,采用字典学习方法寻求一个较小且满足条件的超完备字典来计算测试样本的稀疏表示,来减少算法的计算时间。

#### 3.1 基于类的字典学习

采用基于类的字典学习方法,对同一类的训练样本通过 K-SVD 算法得到该类别的超完备字典,将字典组成元素进一步压缩为更小的集合。针对每个类进行字典学习后构成的超完备字典能更有效地表示该类。针对每个类进行字典学习还有一个优势就是每个类别的训练过程可以并行来完成,并且当系统要增加一个新的类别时,不需要对整个训练数据进行重新训练。

针对  $C$  个不同的手势类别,分别通过 K-SVD 算法构建相应的字典  $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C$ 。

K-SVD 训练算法的目标方程可表示为

$$\langle \mathbf{D}_i, \mathbf{X}_i \rangle = \arg \min_{\mathbf{D}_i, \mathbf{X}_i} \|\mathbf{G}_i - \mathbf{D}_i \mathbf{X}_i\|_2^2, \quad \text{s.t. } \forall n, \|x_n\|_0 \leq T \quad (9)$$

其中,矩阵  $\mathbf{G}_i \in \mathbb{R}^{m \times N}$  中的列为训练集中第  $i$  类的所有样本;  $d_j^i$  为第  $i$  类的子字典  $\mathbf{D}_i = [d_1^i, \dots, d_K^i] \in \mathbb{R}^{m \times K}$  中的第  $j$  列;  $T$  为稀疏表示系数中非零分量的数目的上限。

式 (9) 的求解是一个迭代过程,即在得到的字典  $\mathbf{D}$  上求稀疏矩阵  $\mathbf{X}$ , 然后根据系数矩阵  $\mathbf{X}$  找到更好的字典  $\mathbf{D}$ , 字典的更新是逐列进行的,直到最后收敛。

给定数据集训练样本  $\{g_i\}_{i=1}^N$ , 目标字典原子数  $K$  和收敛条件(稀疏限定因子  $T$ )。K-SVD 通过以下步骤找到最佳的超完备字典  $\mathbf{D} \in \mathbb{R}^{m \times K}$ :

**Step1** 设置初始字典矩阵  $\mathbf{D}^{(0)} \in \mathbb{R}^{m \times K}$  且列向量已进行了归一化, 设  $J=1$ 。

**Step2** 稀疏编码阶段。对每个样本  $g_i$  用任一追踪算法计算出其表示向量  $x_i$ , 即对  $i=1, 2, \dots, N$ , 求解如下最优化问题

$$\min_{x_i} \{\|g_i - \mathbf{D}x_i\|_2^2\}, \quad \text{s.t. } \|x_i\|_0 \leq T$$

**Step3** 字典更新阶段。对字典矩阵  $\mathbf{D}^{(J-1)}$  的每一列按以下方式逐列更新。

1) 定义一组使用了该字典原子的数据样本

$$\omega_k = \{i | 1 \leq i \leq N, x_i^k(i) \neq 0\}$$

2) 按下式计算表示误差矩阵

$$\mathbf{E}_k = \mathbf{G} - \sum_{j \neq k} d_j x_j^T$$

3) 由误差矩阵  $\mathbf{E}_k$  选出仅和  $\omega_k$  相对应的列, 得到  $\mathbf{E}_k^R$ 。

4) 对  $\mathbf{E}_k^R$  进行 SVD 分解  $\mathbf{E}_k^R = \mathbf{U} \Delta \mathbf{V}^T$ , 更新的字典列  $\bar{d}_k$  为  $\mathbf{U}$  的第一列, 用  $\mathbf{V}$  的第一列和  $\Delta(1,1)$  的乘积更新  $x_k^R$ 。

**Step4** 如果满足收敛条件则停止, 否则  $J=J+1$ , 转到 Step2。

K-SVD 算法非常灵活, 可以和常见的稀疏分解的最优原子搜索算法结合使用, 本文选用的是正交匹配跟踪(OMP, orthogonal matching pursuit)算法。

#### 3.2 自学习稀疏表示分类

把各个子字典拼接成一个冗余字典  $\mathbf{D} = [\mathbf{D}_1, \dots,$

$D_C] \in \mathbb{R}^{m \times CK}$ 。用  $D$  取代式(5)最小化  $l_1$  范数问题中的  $G$ , 表示为

$$\hat{\beta} = \arg \min \|\beta\|_1, \text{ s. t. } \|\mathbf{g}_{k,\text{test}} - D\beta\|_2 \leq \varepsilon \quad (10)$$

通过字典学习将稀疏表示问题简化为一个小型的最小化  $l_1$  范数问题, 减少了计算的时间和空间复杂度。得到稀疏表示系数  $\beta$  后, 接下来的处理同第 2 节的一样。自学习稀疏表示分类动作识别 (L-SRC, learning sparse representation classification) 算法的过程描述如下。

**Step1** 初始化参数, 每种类字典的大小  $K$  和稀疏限定因子  $T$ 、误差容忍参数  $\varepsilon$ 。

**Step2** 针对每种手势类别的训练样本利用 K-SVD 算法得到每类手势的超完备字典。

**Step3** 把自学习后的由每种手势类别的超完备字典拼接成一个冗余字典, 并对字典的各列进行归一化处理。

**Step4** 通过求解最小化  $l_1$  范数问题得到稀疏系数。

**Step5** 计算残差, 选择差值最小的为最终的识别结果。

### 4 实验结果与分析

将本文提出的基于稀疏表示的手势识别算法在公开的数据集上进行实验, 测试算法的有效性。

#### 4.1 手势数据集

本文采用由多伦多大学的 Ahmad Akl 提供的手势数据集<sup>[8]</sup>, 它是目前包含手势种类和样本数最多的、基于加速度传感器的公开数据集, 共收集了 7 个用户的 18 种手势的 3 000 多个样本, 每个用户每种手势平均采集 25 次。数据集选用任天堂公司的游戏手柄 Wiimote 作为手势动作采集的输入设备, 内嵌有一个三轴加速度传感器。用户手持 Wiimote 为每种手势采集样本, 通过按下按钮标识手势动作的起始与结束。加速度传感器的采样频率为 100 Hz。18 种手势如图 2 所示。

#### 4.2 预处理

进行手势识别之前, 先对采集到的原始加速度信号进行平滑滤波。因为加速度信号在激励、监测和传输过程中会不同程度地受到随机噪声的影响, 另外手的抖动也会引入噪声, 因此必须对手势加速度信号进行滤波去噪, 本文采用 5 阶平滑滤波来滤除干扰噪声。

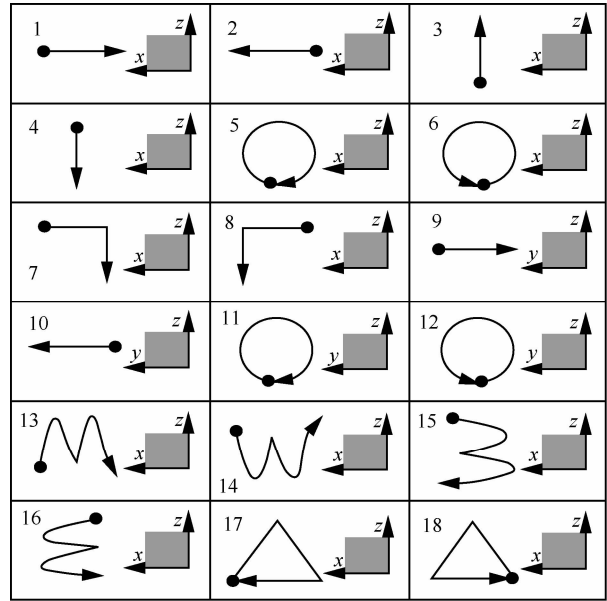


图 2 18 种手势

对手势动作的加速度信号采样是等间隔采样, 但因为执行不同的手势动作所完成的快慢不同, 因此数据集中采样到的各个手势动作样本的采样点个数也各不相同, 加速度采样值  $l$  的长度从 35 到 200 不等。实验中取  $l_{\max}=200$ , 则  $m=600$ 。对于小于  $l_{\max}$  的手势样本分别进行长度归一化插值计算。

#### 4.3 实验结果

测试 SRC 算法和基于自学习稀疏表示的手势识别算法 L-SRC 算法分别在个体相关(user-dependent)、个体无关(user-independent)情况下的识别性能, 并与现有的算法进行了性能比较。实验中 L-SRC 算法的求解利用了 MATLAB K-SVD 字典学习的工具包<sup>[16]</sup>和求解优化问题的 SPGL<sub>1</sub> 工具包<sup>[17]</sup>。

测试时将数据集分为训练集和测试集 2 部分, 测试集中的数据不会包含在训练集中。个体相关的手势识别测试是针对单个测试者的所有样本数据进行, 随机选 15 个组成训练集, 即  $n=270$ , 剩下 10 个做测试。分别得到 7 个人的个体相关的识别率。个体无关测试是采用留一交叉验证 (leave-one-user-out-validation) 法, 从 7 个采集者的手势样本中选出 6 个人的手势样本做训练集, 即  $n=2\ 597$ , 然后用剩下 1 个人的手势样本做测试集。分别得到 7 个人的个体无关的识别率。

首先测试 SRC 算法的识别性能。7 个人在 SRC 算法下对应的手势识别准确率如图 3 所示。结果显示通过针对手势数据的样本长度不同进行长度归一化插值处理后, SRC 算法能很好地运用到手势识

别中，并取得了较高的手势识别率，其中，个体无关的平均识别率为 99.4%，个体相关的平均识别率为 98.7%。但是因为 SRC 算法在进行个体无关的手势识别时是把训练集中的所有样本作为冗余字典（字典的大小  $n=2\ 597$ ），其识别结果比较耗时。

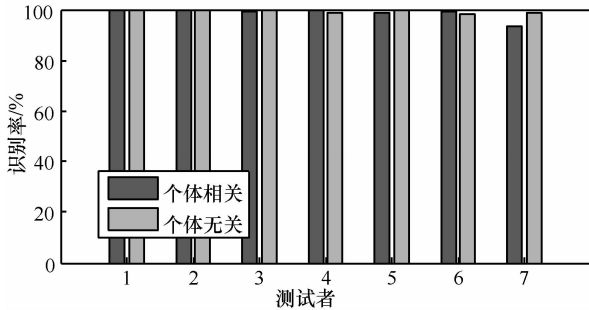


图 3 7 个人在 SRC 算法下的手势识别率

为提高手势识别的实时性，接下来测试本文提出的 L-SRC 算法的识别性能。针对手势数据集中 18 个不同的手势类别，分别对每种手势所构成的训练集样本进行字典学习，构建每种手势的超完备字典  $D_1, D_2, \dots, D_{18}$ 。考察 L-SRC 方法的字典学习时参数选取对识别算法的影响，分别选择不同字典尺寸大小  $K$  的超完备字典和稀疏限定因子  $T$  进行 L-SRC 验证，如表 1 所示。表 1 显示当  $K$  和  $T$  取不同值时的识别准确率差别不大。 $K$  取值越小，所构建的冗余字典也越小，越有利于实时识别。图 4 为 7 个人在 L-SRC 算法下的手势识别率。

表 2 给出了 SRC 和 L-SRC 算法分别在测试样本集相同情况下，个体无关的识别率和计算时间的比较。表中列出的时间为测试一个样本所用的平均时间，通

过测试样本集运行的总时间除以测试样本数（测试样本个数为 450）得到。测试时所用的笔记本电脑配置为 2.16 GHz 的 CPU 和 1 GB 的内存。表 2 显示 L-SRC 算法通过字典学习后的字典尺寸越小，所用的识别时间也越少。当  $K=20$  时，即冗余字典的大小从 SRC 算法中的  $n=2\ 597$  减小到  $n=360$ ，所用的识别时间减少到原来的 1/4，而手势识别率仅降低了 0.2%。

表 1 L-SRC 在不同  $T$  和  $K$  取值下的个体无关的识别率/%

L-SRC	K		
	20	40	60
T=10	99.2	99.3	99.1
T=20	99.2	99.2	99.2
T=40	—	99.1	99.2

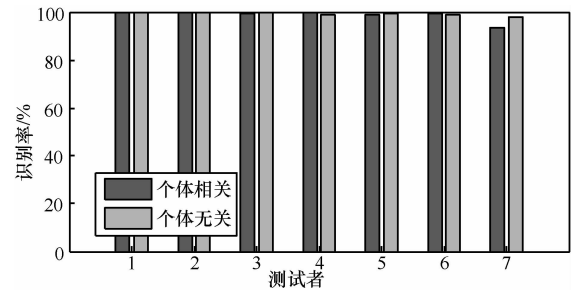


图 4 7 个人在 L-SRC 算法下的手势识别率

表 2 L-SRC 和 SRC 的性能比较

分类方法	识别率/%	所用时间/ms
SRC	99.4	280
L-SRC(K=60)	99.3	145
L-SRC(K=40)	99.3	106
L-SRC(K=20)	99.2	70

表 3 L-SRC 算法 18 种手势的混淆矩阵

类别	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	识别率/%
1	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
2	2	173	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	99.8
3	0	0	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
4	1	0	0	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
5	0	0	0	0	175	0	0	0	0	0	0	0	0	0	1	2	0	0	98.3
6	0	0	0	1	0	168	0	0	0	0	0	0	0	0	0	1	0	5	96.0
7	0	0	0	3	0	0	172	0	0	0	0	0	0	0	0	0	0	0	98.3
8	0	0	0	0	0	0	0	175	0	0	1	0	0	0	0	0	0	0	100
9	0	0	0	0	0	0	0	0	175	0	0	0	0	0	0	0	0	0	100
10	0	0	0	0	0	0	0	0	0	175	0	0	0	0	0	0	0	0	100
11	0	0	0	0	0	1	0	0	0	1	174	0	0	0	0	0	0	0	99.4
12	0	0	0	0	0	0	0	0	1	0	0	174	0	0	0	0	0	0	99.4
13	0	0	0	0	0	0	0	0	0	0	0	0	173	0	0	0	2	0	98.9
14	0	0	0	0	0	0	0	0	0	0	0	0	0	175	0	0	0	0	100
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	175	0	0	0	100
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	175	0	0	100
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	175	0	100
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	174	99.4

表 3 给出了 18 种手势在个体无关下的混淆矩阵。在留一交叉验证中, 测试数据集中平均每种手势的样本个数为 175。混淆矩阵  $R$  为  $18 \times 18$  的矩阵,  $r_{ij}$  对应的数字表示第  $i$  类动作误识别为第  $j$  类动作的个数, 对角线上的数字代表对应手势正确识别的个数。由表可看出, 各种手势都获得了较高的正确识别率, 但也有少量误识别, 如手势 6 容易混淆成手势 18。L-SRC 算法只要求不同种类的手势具有一定的区分度, 即不同种手势所形成的运动轨迹近似分布在不同的子空间, 就能正确的识别。

为进一步考察不同手势对算法性能的影响, 分别选择不同手势种类的手势集  $C=\{8, 14, 18\}$ , 其中,  $C=8$  为选取 18 种手势的前 8 种手势,  $C=14$  为选取 18 种手势中去掉  $yz$  平面的 4 种手势剩下的 14 种手势。表 4 给出了不同的手势识别算法的性能比较。表 4 显示 L-SRC 算法取得了与其他算法相比拟的个体相关的手势识别率, 而个体无关的手势识别率要明显优于其他识别方法。此外 L-SRC 算法可方便地添加新的手势类别和删除已有的手势类别。

表 4 L-SRC 算法与其他手势识别系统的识别率比较

识别方法	手势种类	识别率/%	
		个体相关	个体无关
L-SRC	8~18	100~98.7	99.6~99.3
NN	8~18	99.3~97.7	91.2~88.9
文献[8]	8~18	100~99.8	96.8~94.6
HMM <sup>[6]</sup>	8~18	99.9~99.5	75.9~71.5
Uwave <sup>[9]</sup>	8	98.6	—

## 5 结束语

本文针对动态手势识别中的手势信号长度不一、特征提取与选择对手势识别率影响大和实时性的问题, 提出了一种基于自学习稀疏表示的动态手势识别方法。该方法将分类识别问题转化为求解待识别样本对于整体训练样本的稀疏表示问题, 直接对原始加速度信号进行操作, 省去了特征提取过程。该方法利用面向类别的字典学习, 来寻求一个较小的超完备字典来计算测试样本的稀疏表示, 从而缩减算法的计算时间, 满足实时性要求。该方法可以方便地添加新的手势类别和删除已有的手势类别。在包含 18 种手势的 3 000 多个样本的公开数据集上进行测试, 结果显示 L-SRC 算法获得了较高

的手势识别率, 个体无关的平均识别率为 99.3%, 个体相关的平均识别率为 98.7%。

下一步的工作是在自己的加速度传感器硬件平台上进行实时的动态手势识别; 探索字典学习中的参数选取与识别准确率之间的关系; 进一步探索在稀疏表示分类模型下的长度不一的手势样本的匹配方法, 使之更好地构建冗余字典。

## 参考文献:

- [1] HE Z Y, JIN L W, ZHEN L X, *et al.* Gesture recognition based on 3D accelerometer for cell phones[A]. IEEE Asia Pacific Conference on Circuits and Systems[C]. Macao, China, 2008. 217-22.
- [2] WU J H, PAN G, ZHANG D Q, *et al.* Gesture recognition with a 3-D accelerometer[A]. Proceedings of the 6th International Conf on Ubiquitous Intelligence and Computing[C]. Berlin, 2009. 25-38.
- [3] XU R, ZHOU S, LI W J. MEMS accelerometer based non-specific user hand gesture recognition[J]. IEEE Trans on Sensors, 2012, 12(5): 1166-1173.
- [4] 梁秀波, 张顺, 李启雷等. 运动传感驱动的 3D 直观手势交互[J]. 计算机辅助设计与图形学学报, 2010, 22(3):521-533.  
LIANG X B, ZHANG S, LI Q L, *et al.* 3D intuitive gesture interaction via motion sensing[J]. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(3):521-533.
- [5] 王万良, 杨经纬, 蒋一波. 基于运动传感器的手势识别[J]. 传感技术学报, 2011, 24(12):1723-1727.  
WANG W L, YANG J W, JIANG Y B. Motion sensor based gesture recognition[J]. Chinese Journal of Sensors and Actuators, 2011, 24(12): 1723-1727.
- [6] THOMAS S, BENJAMIN P. Gesture recognition with a Wii controller[A]. Proc of the 2nd International Conference on Tangible and Embedded Interaction[C]. New York, USA, 2008. 11-14.
- [7] TAIWOO P, JINWON L, INSEOK H. E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices[A]. Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems(SenSys'11)[C]. 2011. 260-273.
- [8] AKL A, FENG C, VALAEE S. A novel accelerometer-based gesture recognition system[J]. IEEE Trans on Signal Processing, 2011, 59(12): 6197-6205.
- [9] LIU J, ZHONG L, Wickramasuriya, *et al.* uWave: accelerometer-based personalized gesture recognition and its applications[J]. Pervasive and Mobile Computing, 2009, 5(6):657-675.
- [10] MOIZ F, NATOO P, DERAKHSHANI R, *et al.* A comparative study of classification methods for gesture recognition using a 3-axis accelerometer[A]. Proc of the International Conference on Neural Networks[C]. California, USA, 2011. 2479-2486.
- [11] ARCE F, VALDEZ J. Accelerometer-based hand gesture recognition using artificial neural networks[J]. Trans on Soft Computing for Intelligent Control and Mobile Robotics, 2011, 318: 67-77.
- [12] YUAN T, WANG B. Accelerometer-based Chinese traffic police gesture recognition system[J]. Chinese Journal of Electronics, 2010, 19(2): 270-274.

- [13] WRIGHT J, YANG Y, GANESH S, *et al.* Robust face recognition via sparse representation[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227.
- [14] AHARON M, ELAD M, BRUCKSTEIN A M. The K-SVD: a algorithm for designing of over-complete dictionaries for sparse representation[J]. *IEEE Trans on Signal Processing*, 2006, 54(11):4311-4312.
- [15] WANG L, GU T, TAO X P, *et al.* A hierarchical approach to real-time activity recognition in body sensor networks[J]. *Pervasive and Mobile Computing*, 2012, 8(1):115-130.
- [16] LI C G, GUO J, ZHANG H G. Local sparse representation based classification[A]. *Proceedings of the 2010 International Conference on Pattern Recognition*[C]. Istanbul, Turkey, 2010. 649-652.
- [17] BERG V D, FRIEDLANDER M P. Sparse Optimization with Least-squares Constraints E[R]. Dept of Computer Science, Univ of British Columbia, 2010.

#### 作者简介:



肖玲 (1977-), 女, 湖南衡阳人, 湖南大学博士生、讲师, 主要研究方向为无线传感器网络、移动计算。



李仁发 (1957-), 男, 湖南郴州人, 博士, 湖南大学教授、博士生导师, 主要研究方向为嵌入式计算和 CPS。



曾凡仔 (1971-), 男, 湖南郴州人, 博士, 湖南大学副教授, 主要研究方向为压缩感知、信号处理。



屈卫兰 (1977-), 女, 湖南娄底人, 湖南大学博士生、讲师, 主要研究方向为移动计算。