

DOI: 10.19650/j.cnki.cjsi.J2209320

基于皮尔逊相关系数的动态签名验证方法*

刘若男¹, 辛义忠¹, 李岩²

(1. 沈阳工业大学信息科学与工程学院 沈阳 110142; 2. 沈阳体育学院基础部 沈阳 110102)

摘要:针对动态签名验证中存在的动态特征长度不等、动态签名验证方法较复杂以及识别率较低等问题,提出了一种基于皮尔逊相关系数的动态签名验证方法。首先通过划分原始特征区域,筛选并计算对应区域内的特征权重和,然后利用皮尔逊相关分析法计算各签名特征间的相关系数;再将皮尔逊相关系数作为新特征,分析真伪签名的皮尔逊相关系数分布情况;最后结合高斯密度函数模型,并通过设置个体判别阈值来进行签名验证。实验结果表明,真签名内的皮尔逊相关系数普遍高于真伪签名间的皮尔逊相关系数,且本方法在 SVC 和 xLongSignDB 数据集上均展现了较优的签名验证性能,其中 xLongSignDB 数据集上的误拒率和误识率分别为 2.1% 和 1.7%。

关键词:皮尔逊相关系数;动态签名验证;特征权重;高斯密度函数

中图分类号: TH701 TP391 **文献标识码:** A **国家标准学科分类代码:** 510.40

Dynamic signature verification method based on Pearson correlation coefficient

Liu Ruonan¹, Xin Yizhong¹, Li Yan²

(1. School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110142, China;

2. Department of Basic Courses, Shenyang Sport University, Shenyang 110102, China)

Abstract: The dynamic signature verification has problems of the unequal length of dynamic features, complex dynamic signature verification methods, and low recognition rate. To address these issues, a dynamic signature verification method based on correlation coefficient is proposed. First, the feature weight sum in the corresponding region is filtered and calculated by dividing the original feature region. And the correlation coefficients between signature features are calculated by the Pearson correlation analysis method. Secondly, the Pearson correlation coefficient distribution of genuine and simulated signatures is analyzed with the correlation coefficient as a new feature. Finally, the signature is evaluated by combining the Gaussian density function model and setting an individual discrimination threshold. Experimental results show that the Pearson correlation coefficient inside the genuine signature is generally higher than that between the genuine and simulated signatures. This method shows better signature verification performance on SVC and xLongSignDB data sets. The false rejection rate and false acceptance rate on xLongSignDB data sets are 2.1% and 1.7%, respectively.

Keywords: Pearson correlation coefficient; dynamic signature verification; feature weight; Gaussian density function

0 引言

签名识别是生物特征识别领域的一个重要组成部分,而签名验证是签名识别中的一类重要问题。签名验证是一个二分类问题,通过比较两份签名样本,判断它们是否属于同一书写人^[1]。签名验证又分为静态签名验证和动态签名验证。静态签名验证需要利用扫描仪或相机

等设备将纸质版签名转换为签名图像,再通过提取笔划等静态特征进行签名验证。然而当伪造者对签名进行熟练模仿后,笔划等静态特征会变得极其相似,使得算法复杂度增加,准确度降低。因此,研究人员引入了签名的动态特征,例如书写过程中的笔压、速度和加速度等。相对于静态特征,由于动态特征是隐性的,伪造者完全模仿的可能性极低。例如, Iranmanesh 等^[2]利用主成分分析法提取动态签名特征,并结合多层感知分类器,结果发现误

识率(false acceptance rate, FAR)为4.4%,误拒率(false rejection rate, FRR)为6.4%。Yahyatabar等^[3]通过动态特征稳定性(dynamic feature stability, DFS)实验,提取笔压、速度和加速度等特征函数的稳定点,并使用拉东变换转换签名角度。Hu等^[4]提出了基于单模板匹配的特征曲线相似度模型,并利用进化算法搜索两条特征曲线的最优匹配距离,结果表明萨班哲大学签名数据集(Sabancı University signature database, SUSIG)的签名识别率最高。Tolosana等^[5]创建了一个新的公共在线签名数据集,并利用时间递归神经网络方法进行签名验证,结果发现等错误率(equal error rate, EER)小于2.0%。Chen等^[6]提出了一种基于多模式匹配的动态签名验证方法,通过将动态时间规整算法(dynamic time warping, DTW)和字符串编辑距离算法(string editing distance, SED)相结合,计算待测签名与模板签名间距离的平均值和最大值等,结果发现FAR和FRR均为1.25%。Okawa等^[7]利用平均模板集并结合加权动态时间规整算法进行签名验证。Hefny等^[8]将勒让德多项式系数作为动态签名特征,并利用深度前馈神经网络对签名进行分类,降低了匹配错误率。Zhou等^[9]将签名的动态特征和静态特征相结合,并利用支持向量机和动态时间规整算法进行签名验证。Saleem等^[10]考察了采样频率和采样点数量对签名验证系统准确性的影响,并提出了一种基于采样频率和动态时间规整算法的动态签名验证方法。

通过检索文献发现,目前动态签名验证普遍分为基于模型的动态签名验证方法和基于距离的动态签名验证方法。尽管方法不同,但实际上都是通过直接或间接的计算签名特征之间的相似性来验证真伪签名,且算法的复杂程度较高。鉴于此,本文提出一种更易理解的基于皮尔逊相关系数的动态签名验证方法。

皮尔逊相关系数是英国数学家卡尔·皮尔逊在高尔顿等所研究的回归统计的基础上提出的。在医学领域,Rovetta^[11]认为通过计算皮尔逊相关性可以揭露研究对象间的内在关联。Pugach等^[12]计算了每百万人口的新冠肺炎死亡率和严重维生素D缺乏患病率间的皮尔逊相关系数,发现其相关系数为0.79。根据实验结果,Pugach等建议通过增加维生素D的摄入来预防和治疗新冠肺炎。Grover等^[13]利用皮尔逊相关性来评估认知功能和人口统计学变量间的关联,发现教育程度和认知功能间呈正相关。Koerner等^[14]为考察神经生理学变量之间的内在关系,利用皮尔逊相关分析模型和线性混合效应回归模型对变量进行了分析和预测,发现皮尔逊相关分析模型可能并不适用于神经生理学变量的预测。Walther等^[15]为证明多普勒超声方法在膝关节诊断中具有重要意义,分别使用多普勒超声和专家会诊两种方法对膝关节进行临床诊断,并考察这两种条件下所采集到的膝关

节数据间的皮尔逊相关性,发现多普勒超声方法更适合膝关节的临床诊断。Li等^[16]利用皮尔逊相关性分别考察了和肥胖相关的20项临床指标与血清趋化素浓度间的相关性。Liu等^[17]通过计算皮尔逊相关系数来探讨创伤后成长(posttraumatic growth, PTG)和创伤后应激障碍症状(posttraumatic stress disorder, PTSD)间的相关性,发现PTG和PTSD间呈正相关,且这种相关性可能与年龄和创伤类型等有关。Zhao等^[18]为评估中国成人和儿童的C反应蛋白与肥胖之间的关系,采用随机效应模型对皮尔逊相关系数进行荟萃分析,发现C反应蛋白与肥胖之间存在显著相关性。

除医学领域外,皮尔逊相关系数也被广泛应用于生物信息或特征的检测与识别方向。Nayar等^[19]将60颗牙齿在不同的条件下分为4组,并通过牙齿状态评估牙齿主人的年龄,结果发现评估的年龄和实际年龄间存在强的正相关性。Berus等^[20]利用皮尔逊相关分析法选择语音特征,并将语音特征送入人工神经网络来预测帕金森疾病。Islam等^[21]为从脑电图信号中识别情感,将一维脑电图数据转换为皮尔逊相关系数矩阵,并把相关系数矩阵送入卷积神经网络来进行情感识别。Lee等^[22]提出了一种基于光容积图的深度特征融合方法,通过将卷积神经网络中提取的深度特征和利用皮尔逊相关法选择的统计特征相融合来识别情感。Friedman等^[23]将类内相关系数作为评估眼动、面部和步态等多个生物特征稳定性的指标,并与皮尔逊相关系数进行了对比,发现类内相关系数更适合评估特征的稳定性。Sapey-Triomphe等^[24]探讨了轻度认知障碍和早期阿尔兹海默症对面部情绪识别的影响,并利用皮尔逊相关性进一步考察了不同面部表情和大脑区域之间的相关性,发现每一种情绪都与一个大脑特定区域相关。Rujirakul等^[25]考虑到化妆可能会对面部特征造成影响,利用皮尔逊相关整合统计模型来提高人脸识别系统的精度。Gosciewska等^[26]将从视频序列中截取的动作剪影生成形状凸包,并采用欧氏距离、相关系数等方法对动作进行分类。苑玮琦等^[27]为避免特征间相关性对手形识别的影响,提出了基于相关性的手形特征选择方法,减小了手形特征选择的复杂性。

上述文献表明皮尔逊相关系数可应用在临床医学、人脸识别和情感识别等多个研究方向,然而却很少有研究将皮尔逊相关系数作为动态特征来进行签名验证。检索到的文献中,Iranmanesh等^[28]通过计算横纵坐标和笔压之间的皮尔逊相关系数来选择特征,结果发现FAR为21.35%,FRR为13.81%。文献^[29]则对签名进行了笔划分割,通过计算各笔划内的笔压和,以及对应笔划笔压和之间的皮尔逊相关系数来进行签名验证。然而这种方法只考察了笔压特征,存在一定局限性。

1 动态签名验证过程和特征选择方法

为了解决上述文献中存在的动态特征长度不等、单一特征识别率低和验证算法过于复杂等问题,本文提出了一种基于皮尔逊相关系数的动态签名验证方法。通过划分原始特征空间和特征融合等实现特征长度对齐,并利用皮尔逊相关分析法计算多特征融合后的特征间相关系数,再将相关系数作为新的动态特征,并结合高斯密度函数模型实现签名验证。

1.1 皮尔逊相关系数

皮尔逊相关系数 r 用来描述两变量间线性相关强弱的程度,且 r 的绝对值越大,两变量间线性相关的程度越强。现有两个长度为 n 的特征集合 $\{x_1, x_2, \dots, x_i, \dots, x_n\}$ 和 $\{y_1, y_2, \dots, y_i, \dots, y_n\}$, 则相关系数 r 的计算公式为:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

式中: \bar{x} 和 \bar{y} 分别为两个特征集内样本的均值,且 r 的取值为 $[-1, 1]$ 。

1.2 数据预处理

本文使用的数据集为首届国际签名验证比赛 (the first international signature verification competition, SVC2004) 数据集^[30]的“sample”子集和“task2”子集,以及生物特征和数据模式分析实验室 (biometrics and data pattern analytics lab, BiDA Lab) 的数据集 xLongSignDB^[5,31]。其中, SVC2004 数据集保存了 45 位书写人的签名数据,包含 20 个真签名和不少于 4 个伪造者模仿的 20 个伪签名; xLongSignDB 数据集保存了 29 位书写人的签名数据,包含 46 个真签名和 10 个伪签名。两个数据集中的伪签名都是熟练伪签名,相比于其他数据集更难以识别,且两个数据集均包括签名轨迹点的横纵坐标、笔压、笔倾斜角和笔方位角等特征信息。

为消除由于签名大小和位置不同给特征提取带来的影响,通过式(2)和(3)对签名进行归一化处理。数据预处理结果如图 1 和 2 所示。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \times 100 \quad (2)$$

$$y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \times 100 \quad (3)$$

式中: x 和 y 为签名轨迹点的原始坐标; x_{\max} 和 x_{\min} 为签名轨迹点中所有 x 的最大值和最小值; y_{\max} 和 y_{\min} 为签名轨迹点中所有 y 的最大值和最小值; x' 和 y' 为归一化后的值。

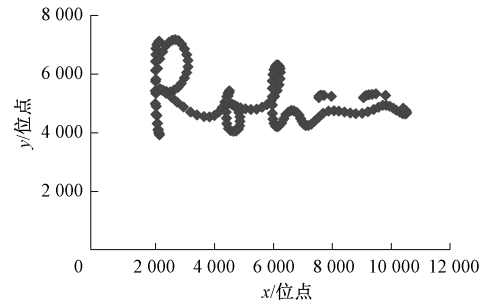


图 1 原始签名

Fig. 1 Original signature

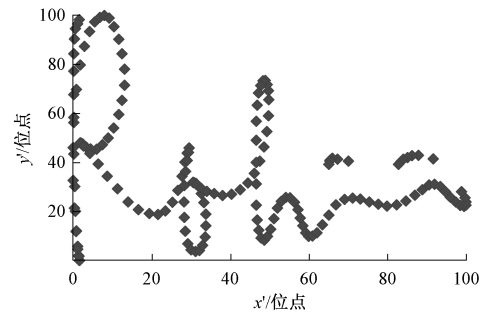


图 2 归一化后的签名

Fig. 2 Normalized signature

1.3 特征选择和融合

除 x 和 y 特征外,实际上每个笔迹点均隐含着 3 个动态特征,即笔压(P)、笔倾斜角(T)和笔方位角(A),它们分别指书写人施加在笔尖上并传递给数位板的压力、笔身与数位板之间的夹角以及笔身在数位板上的垂直投影与数位板正北方向所形成的夹角。

由于在计算相关系数的过程中需要保证特征长度相等,因此本文在归一化的基础上,又将签名轨迹的 x 坐标分割成 $[0, 20)$ 、 $[20, 40)$ 、 $[40, 60)$ 、 $[60, 80)$ 和 $[80, 100]$ 5 个区域,分别筛选出对应区域的特征并对该区域特征求和,再将各区域特征和乘以特征权重,最后将乘以特征权重后的多个特征相加,实现特征选择和多特征的融合。

以 SVC 签名数据集为例,设 User1 的 20 个真签名的特征组合为 $G_1 \sim G_{20}$ 。表 1 为 20 个真签名在不同划分区域下的加权特征组合,表 1 中 u, v, w 分别为笔压、笔倾斜角和笔方位角的特征权重。 $p_{1,2}, t_{1,2}$ 和 $a_{1,2}$ 分别为第 1 个签名在第 2 个区域 $[20, 40)$ 的笔压之和、笔倾斜角之和和笔方位角之和,其他以此类推。

为确定各特征权重,对单特征区分真伪签名的能力进行了考察。图 3 为不同特征在真伪签名下的分布情况。从图 3 可以看出,利用笔压特征区分真伪签名的能力要优于其他特征。且通过统计分析发现,真签名的平均笔压为 628,而伪签名的平均笔压只有 279。但对于笔倾斜角和笔方位角而言,由于真伪签名下的特征分布差

表1 不同区域下各真签名的加权特征组合

Table 1 Weighted feature combination of genuine signatures of different regions

区域	G_1	G_2	...	G_{20}
[0,20)	$(u \times p_{1,1}, v \times t_{1,1}, w \times a_{1,1})$	$(u \times p_{2,1}, v \times t_{2,1}, w \times a_{2,1})$...	$(u \times p_{20,1}, v \times t_{20,1}, w \times a_{20,1})$
[20,40)	$(u \times p_{1,2}, v \times t_{1,2}, w \times a_{1,2})$	$(u \times p_{2,2}, v \times t_{2,2}, w \times a_{2,2})$...	$(u \times p_{20,2}, v \times t_{20,2}, w \times a_{20,2})$
[40,60)	$(u \times p_{1,3}, v \times t_{1,3}, w \times a_{1,3})$	$(u \times p_{2,3}, v \times t_{2,3}, w \times a_{2,3})$...	$(u \times p_{20,3}, v \times t_{20,3}, w \times a_{20,3})$
[60,80)	$(u \times p_{1,4}, v \times t_{1,4}, w \times a_{1,4})$	$(u \times p_{2,4}, v \times t_{2,4}, w \times a_{2,4})$...	$(u \times p_{20,4}, v \times t_{20,4}, w \times a_{20,4})$
[80,100]	$(u \times p_{1,5}, v \times t_{1,5}, w \times a_{1,5})$	$(u \times p_{2,5}, v \times t_{2,5}, w \times a_{2,5})$...	$(u \times p_{20,5}, v \times t_{20,5}, w \times a_{20,5})$

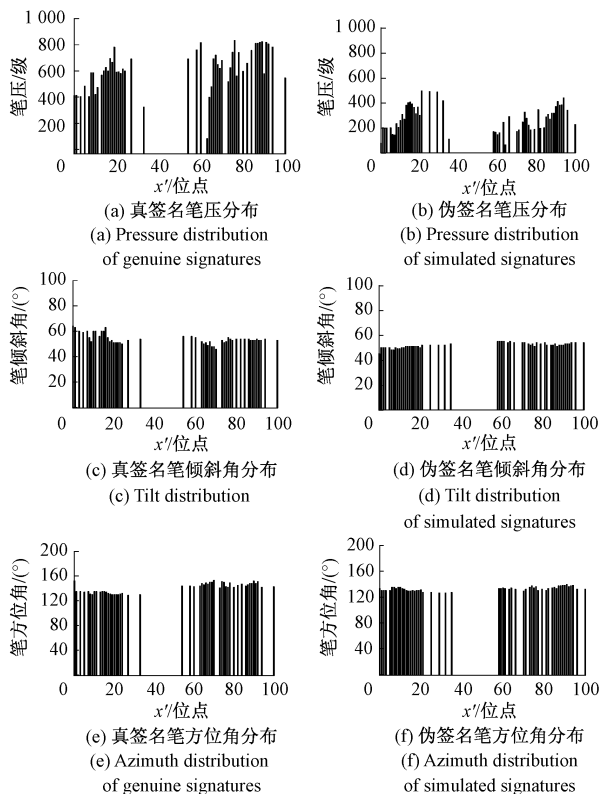


图3 真伪签名的特征分布

Fig. 3 Feature distribution of genuine and simulate signatures

表2 各真签名的新特征序列

Table 2 New feature sequence of various genuine signatures

区域	G'_1	G'_2	...	G'_{20}
[0,20)	$g_{1,1} = 0.6 \times p_{1,1} + 0.2 \times t_{1,1} + 0.2 \times a_{1,1}$	$g_{2,1} = 0.6 \times p_{2,1} + 0.2 \times t_{2,1} + 0.2 \times a_{2,1}$...	$g_{20,1} = 0.6 \times p_{20,1} + 0.2 \times t_{20,1} + 0.2 \times a_{20,1}$
[20,40)	$g_{1,2} = 0.6 \times p_{1,2} + 0.2 \times t_{1,2} + 0.2 \times a_{1,2}$	$g_{2,2} = 0.6 \times p_{2,2} + 0.2 \times t_{2,2} + 0.2 \times a_{2,2}$...	$g_{20,2} = 0.6 \times p_{20,2} + 0.2 \times t_{20,2} + 0.2 \times a_{20,2}$
[40,60)	$g_{1,3} = 0.6 \times p_{1,3} + 0.2 \times t_{1,3} + 0.2 \times a_{1,3}$	$g_{2,3} = 0.6 \times p_{2,3} + 0.2 \times t_{2,3} + 0.2 \times a_{2,3}$...	$g_{20,3} = 0.6 \times p_{20,3} + 0.2 \times t_{20,3} + 0.2 \times a_{20,3}$
[60,80)	$g_{1,4} = 0.6 \times p_{1,4} + 0.2 \times t_{1,4} + 0.2 \times a_{1,4}$	$g_{2,4} = 0.6 \times p_{2,4} + 0.2 \times t_{2,4} + 0.2 \times a_{2,4}$...	$g_{20,4} = 0.6 \times p_{20,4} + 0.2 \times t_{20,4} + 0.2 \times a_{20,4}$
[80,100]	$g_{1,5} = 0.6 \times p_{1,5} + 0.2 \times t_{1,5} + 0.2 \times a_{1,5}$	$g_{2,5} = 0.6 \times p_{2,5} + 0.2 \times t_{2,5} + 0.2 \times a_{2,5}$...	$g_{20,5} = 0.6 \times p_{20,5} + 0.2 \times t_{20,5} + 0.2 \times a_{20,5}$

其中, $g_{k,1} \sim g_{k,5}$ 为加权特征和, 具体计算方法如表2所示。设 G'_m 和 G'_n 是第 m 和第 n 个真签名的特征序列, i 为 G'_m 的第 i 个特征的编号, j 为 G'_n 的第 j 个特征的编号, 则式(1)的相关系数公式变为:

异较小, 因此区分真伪签名的能力也相对有限。鉴于此, 采用3轮询问的德尔菲专家调查法, 确定笔压、笔倾斜角和笔方位角的特征权重 u, v, w 的值。特征权重的确定基于如下原则:

原则1 为避免单一特征识别的弱鲁棒性, 每一种特征都应保证具有一定权重比例;

原则2 特征权重比的选择应尽量使得真伪签名间的相关性属于极弱相关, 又要尽量使得真签名间的相关性属于极强相关;

原则3 对于真伪差异体现较大的特征, 增加该特征的权重比例;

原则4 列出满足原则1、2和3的权值比的所有组合, 并在德尔菲调查的基础上, 得到评价最高的权值比组合。

依据上述原则, 本文最终将特征权重 u, v, w 分别设置为0.6、0.2和0.2。

确定特征权重后, 通过计算各签名对应区域的加权特征和得到该区域的新特征, 则每个签名均包含5个新特征, 如表2所示, 在此基础上计算新特征序列间的相关性。

真签名的特征序列表达方式为:

$$G'_k = \{g_{k,1}, g_{k,2}, g_{k,3}, g_{k,4}, g_{k,5}\}, k = 1, 2, \dots, 20 \quad (4)$$

$$r_{G'_m, G'_n} = \frac{\sum_{i=1}^5 (g_{mi} - \bar{g}_m)(g_{ni} - \bar{g}_n)}{\sqrt{\sum_{i=1}^5 (g_{mi} - \bar{g}_m)^2} \sqrt{\sum_{i=1}^5 (g_{ni} - \bar{g}_n)^2}} \quad (5)$$

式中: \bar{g}_m 和 \bar{g}_n 分别为两组真签名特征序列 G'_m 和 G'_n 的均值。通过变换公式(5),分别计算真签名内和真伪签名间的皮尔逊相关系数。

2 实验结果分析

2.1 皮尔逊相关系数分布

实验所使用的 SVC 数据集共包含 45 名书写人的签名数据,每人 20 个真签名和 20 个伪签名,则根据式(5)计算出的真签名的皮尔逊相关系数共有 45×20 (真签名) $\times 20$ (真签名) = 18 000 个,真伪间的皮尔逊相关系数共有 45×20 (真签名) $\times 20$ (伪签名) = 18 000 个;而 xLongSignDB 数据集包含 29 名书写人的签名数据,每人 46 个真签名和 10 个伪签名,则真签名的皮尔逊相关系数共有 29×46 (真签名) $\times 46$ (真签名) = 61 364 个,真伪间的皮尔逊相关系数共有 29×46 (真签名) $\times 10$ (伪签名) = 13 340 个。此外,为便于整理和统计,所计算出的皮尔逊相关系数均以绝对值形式表示。

表 3 为真签名内的皮尔逊相关系数分布情况。从表 3 可以看出 SVC 数据集中有 10 134 个真签名内的皮尔逊相关系数处于 $[0.8, 1.0]$,而 xLongSignDB 数据集中有 29 040 个真签名内的皮尔逊相关系数处于 $[0.8, 1.0]$ 。另外,本文也计算了真伪签名间的皮尔逊相关系数,如表 4 所示。实验结果表明 SVC 数据集中有 7 895 个真伪签名间的皮尔逊相关系数处于 $[0, 0.4]$,而在 xLongSignDB 数据集中有 7 105 个真伪签名间的皮尔逊相关系数处于 $[0, 0.4]$ 。从这些统计结果中可以发现,真签名内的相关性普遍较强,而真伪签名间的相关性普遍较弱。

表 3 真签名内的皮尔逊相关系数分布

Table 3 Pearson correlation coefficient distribution within genuine signatures

区间	SVC 数据集	xLongSignDB 数据集
$[0, 0.2)$	1 364	6 472
$[0.2, 0.4)$	1 130	4 218
$[0.4, 0.6)$	1 818	7 422
$[0.6, 0.8)$	3 554	14 212
$[0.8, 1.0]$	10 134	29 040

为了进一步考察真伪签名的皮尔逊相关系数分布情况,得到真伪签名相关系数分布的普遍规律,本文分别计算了两数据集皮尔逊相关系数的均值和中位数。图 4 为 SVC 数据集中 5 名书写人真伪签名的皮尔逊相关系数分布情况。图 4 中 5 个人的真签名内的皮尔逊相关系数 $RGG_1 \sim RGG_5$ 为每个人的 20 个真签名特征序列分别与这 20 个真签名特征序列求相关系数再取均值,即:

表 4 真伪签名间的皮尔逊相关系数分布

Table 4 Pearson correlation coefficient distribution between genuine and simulated signatures

区间	SVC 数据集	xLongSignDB 数据集
$[0, 0.2)$	3 898	4 682
$[0.2, 0.4)$	3 997	2 423
$[0.4, 0.6)$	2 623	1 949
$[0.6, 0.8)$	1 969	1 528
$[0.8, 1.0]$	5 513	2 758

$RGG = average(r_{c_1, c_1}, \dots, r_{c_1, c_{20}}, \dots, r_{c_{20}, c_1}, \dots, r_{c_{20}, c_{20}})$,真伪签名间的皮尔逊相关系数 $RGS_1 \sim RGS_5$ 为每个人的 20 个真签名特征序列分别与 20 个伪签名特征序列求相关系数再取均值,即: $RGS = average(r_{c_1, s_1}, \dots, r_{c_1, s_{20}}, \dots, r_{c_{20}, s_1}, \dots, r_{c_{20}, s_{20}})$,其中 $S'_1 \sim S'_{20}$ 为伪签名的特征序列。从图 4 中可以看出,真签名内的皮尔逊相关系数均值和中位数明显高于真伪签名间的皮尔逊相关系数均值和中位数。通过统计两个数据集的皮尔逊相关系数发现,SVC 数据集中真签名内皮尔逊相关系数的平均中位数和均值分别为 0.79 和 0.75,而真伪签名间皮尔逊相关系数的平均中位数和均值分别为 0.57 和 0.59。另外,xLongSignDB 数据集中真签名内皮尔逊相关系数的平均中位数和均值分别为 0.74 和 0.71,而真伪签名间皮尔逊相关系数的平均中位数和均值分别为 0.43 和 0.54。

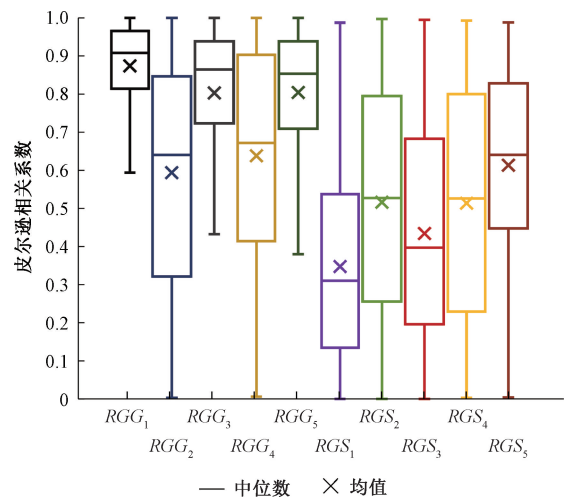


图 4 真伪签名的皮尔逊相关系数分布

Fig. 4 Pearson correlation coefficient distribution of genuine and simulate signatures

2.2 签名验证方法

由于真签名内和真伪签名间的皮尔逊相关系数分布不同,因此本文考虑将皮尔逊相关系数作为新的动态特

征,并通过设置皮尔逊相关系数阈值来区分真伪签名。考虑到各书写人的书写方式和习惯均不同,若设置整体相关系数阈值,可能会影响真伪签名验证结果。因此为提高签名验证率,本文将根据各书写人的书写特征设置个体相关系数阈值。实验分别取每名书写人的前10个真签名作为训练集,则 SVC 数据集中每名书写人的后10个真签名和20个伪签名将作为测试集,而 xLongSignDB 数据集中每名书写人的后36个真签名和10个伪签名将作为测试集。表5为训练集和测试集的真签名内与真伪签名间的皮尔逊相关系数。对于 SVC

数据集,表中前10列为训练样本与训练样本间的皮尔逊相关系数(包括自身),共有100个值,这100个值将作为新的训练样本(记为 $\{x_1, \dots, x_{100}\}$)参与后继运算;表5中第11~20列为训练样本与真签名测试样本间的皮尔逊相关系数,共有100个值;第21~40列为训练样本与伪签名测试样本的皮尔逊相关系数,共有200个值。第11~40列的300个值将作为新的测试样本参与后继运算。同理 xLongSignDB 数据集中每名书写人的新训练样本数量为100个,新测试样本数量为460个。

表5 训练集和测试集的真签名内与真伪签名间的皮尔逊相关系数

Table 5 Pearson correlation coefficient within genuine signatures and between genuine and simulate signatures according to training set and test set

特征序列	G'_1	...	G'_{10}	G'_{11}	...	G'_{20}	S'_1	...	S'_{20}
G'_1	$r_{G'_1, G'_1}$...	$r_{G'_1, G'_{10}}$	$r_{G'_1, G'_{11}}$...	$r_{G'_1, G'_{20}}$	$r_{G'_1, S'_1}$...	$r_{G'_1, S'_{20}}$
G'_2	$r_{G'_2, G'_1}$...	$r_{G'_2, G'_{10}}$	$r_{G'_2, G'_{11}}$...	$r_{G'_2, G'_{20}}$	$r_{G'_2, S'_1}$...	$r_{G'_2, S'_{20}}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
G'_{10}	$r_{G'_{10}, G'_1}$...	$r_{G'_{10}, G'_{10}}$	$r_{G'_{10}, G'_{11}}$...	$r_{G'_{10}, G'_{20}}$	$r_{G'_{10}, S'_1}$...	$r_{G'_{10}, S'_{20}}$

在数据的处理和分析过程中,本文发现特征值总是围绕着均值在标准差附近波动,因此为了对这些特征进行建模并最终实现签名验证,引入了高斯密度函数。高斯密度函数的公式如下:

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right), i = 1, 2, \dots, 400 \tag{6}$$

仍然以 SVC 数据集中的 User1 为例,式中 x_i 为新训练样本的第 i 个皮尔逊相关系数。 μ 和 σ 则分别为100个新的训练样本的皮尔逊相关系数的均值和标准差,并将其作为模板。为进行签名验证,需要先计算100个新的训练样本的均值和标准差,然后分别将 $\{x_1, \dots, x_{100}\}$ 代入式(6),得到对应的特征函数值,最后再计算这100个特征函数值的均值,并作为 User1 的个体判别阈值。再将作为新测试样本的300个皮尔逊相关系数分别代入式(6),得到对应的特征函数值。为进一

步提高签名验证速度,本文在签名验证时将新测试样本的特征函数值按列取均值,并将此均值作为对应签名的最终特征值,进而与个体判别阈值相比较。若各签名的最终特征值大于当前签名的个体判别阈值时,判断为真签名,否则判断为伪签名。通过计算 FRR 和 FAR 来衡量方法的有效性,公式如下:

$$FRR = \frac{\text{误拒的真签名个数}}{\text{真签名总数}} \times 100\% \tag{7}$$

$$FAR = \frac{\text{误识的伪签名个数}}{\text{伪签名总数}} \times 100\% \tag{8}$$

在签名验证中,应当尽量降低伪签名的误识率,也就是 FAR 的值。表6将本文方法与其他方法进行了对比总结,发现本方法在 xLongSignDB 数据集上的性能最优,识别率达到了97.8%,且在 SVC 数据集上的识别率也达到了96.0%,处于中等水平。这可能是由于本文采用了个体阈值法,可根据每名书写人的特性设置对应的阈值,提高了

表6 签名验证方法对比

Table 6 Comparison of signature verification methods

文献	数据集	方法	FRR	FAR
本文	SVC	皮尔逊相关系数+高斯密度函数	4.9	3.6
本文	xLongSignDB	皮尔逊相关系数+高斯密度函数	2.1	1.7
文献[32]	SVC	小波变换+神经网络	9.5	10.7
文献[33]	SVC	基因表达式编程	5.2	2.23
文献[34]	SVC	约束特征曲线的动态时间规整+高斯混合模型	2.3	2.9
文献[35]	xLongSignDB	特征分区+模糊分类器	11.8	5.4

识别率。此外,本方法在 xLongSignDB 数据集上的性能最优也可能与该数据集的真伪签名数量差异较大有关。由于每名书写人中真签名的数量为 46 个,而伪签名只有 10 个,因此可能会间接影响签名验证性能。

3 结 论

考虑到动态签名验证中存在的单动态特征识别率较低、特征长度不等和签名验证方法较复杂等问题,提出了一种基于皮尔逊相关系数的动态签名验证方法。首先通过重新划分原始特征区域,筛选并融合对应区域特征来实现特征长度对齐,再根据皮尔逊相关系数公式计算签名特征间的皮尔逊相关系数,并将其作为新特征,最后结合高斯密度函数模型进行签名验证。结果表明,SVC 和 xLongSignDB 数据集中真签名内的皮尔逊相关系数均普遍高于真伪签名间的皮尔逊相关系数,且皮尔逊相关系数普遍处于[0.8, 1.0]。另外,通过与其他文献对比发现,本方法在两个数据集上均取得了较好的签名验证性能。

考虑到签名验证结果可能与特征提取、测试集、训练集数量的划分以及数据集的选择等相关,因此本文接下来可能会针对这些不同情况进行签名验证并考察性能。

参考文献

- [1] 丁晓青,李昕. 计算机笔迹鉴别与验证的理论和方法[M]. 北京:清华大学出版社,2012.
DING X Q, LI X. Computer writer identification and verification theory and method [M]. Beijing: Tsinghua University Press, 2012.
- [2] IRANMANESH V, AHMAD S M S, ADNAN W A W, et al. Online handwritten signature verification using neural network classifier based on principal component analysis[J]. The Scientific World Journal, 2014, 2014: 381469.
- [3] YAHYATABAR M E, GHASEMI J. Online signature verification using double-stage feature extraction modelled by dynamic feature stability experiment [J]. IET Biometrics, 2017, 6(6): 393-401.
- [4] HU H C, ZHENG J B, ZHAN E Q, et al. Online signature verification based on a single template via elastic curve matching [J]. Sensors, 2019, 19(22): 4858.
- [5] TOLOSANA R, RODRIGUEZ R V, FIERREZ J, et al.

- Reducing the template ageing effect in on-line signature biometrics[J]. IET Biometrics, 2019, 8(6): 422-430.
- [6] CHEN Y H, MAHPIRAT, UBUL K. Online signature verification based on multi-mode matching [C]. International Conference on Image Processing and Machine Vision, 2020: 190-194.
- [7] OKAWA M. Online signature verification using single-template matching with time-series averaging and gradient boosting[J]. Pattern Recognition, 2020, 102: 107227.
- [8] HEFNY A, MOUSTAFA M. Online signature verification using deep learning and feature representation using Legendre polynomial coefficients [C]. International Conference on Advanced Machine Learning Technologies and Applications, 2019: 689-697.
- [9] ZHOU Y W, ZHENG J B, HU H C, et al. Handwritten signature verification method based on improved combined features[J]. Applied Sciences, 2021, 11(13): 5867.
- [10] SALEEM M, KOVARI B. Online signature verification using signature down-sampling and signer-dependent sampling frequency [J]. Neural Computing and Applications, 2021:1-13.
- [11] ROVETTA A. Raiders of the lost correlation: A guide on using Pearson and spearman coefficients to detect hidden correlations in medical sciences [J]. Cureus, 2020, 12(11): e11794.
- [12] PUGACH I Z, PUGACH S. Strong correlation between prevalence of severe vitamin D deficiency and population mortality rate from COVID-19 in Europe [J]. Wiener Klinische Wochenschrift, 2021, 133(7): 403-405.
- [13] GROVER S, NEHRA R, BHATEJA G, et al. A comparative study of cognitive deficits in patients with delusional disorder and paranoid schizophrenia [J]. Industrial Psychiatry Journal, 2011, 20(2): 107-114.
- [14] KOERNER T K, ZHANG Y. Application of linear mixed-effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies [J]. Brain Sciences, 2017, 7(3): 26.
- [15] WALTHER M, HARMS H, KRENN V, et al. Correlation of power doppler sonography with vascularity of the synovial tissue of the knee joint in patients with osteoarthritis and rheumatoid arthritis [J]. Arthritis & Rheumatism, 2001, 44(2): 331-338.

- [16] LI Y, SHI B Y, LI S L. Association between serum chemerin concentrations and clinical indices in obesity or metabolic syndrome: A meta-analysis [J]. *Plos One*, 2014, 9(12): e113915.
- [17] LIU A N, WANG L L, LI H P, et al. Correlation between posttraumatic growth and posttraumatic stress disorder symptoms based on Pearson correlation coefficient: A meta-analysis[J]. *The Journal of Nervous and Mental Disease*, 2017, 205(5): 380-389.
- [18] ZHAO Y, LV G. Correlation of C-reactive protein level and obesity in Chinese adults and children: A meta-analysis[J]. *Journal of Endocrinological Investigation*, 2013, 36(8): 642-647.
- [19] NAYAR A K, PARHAR S, THIND G, et al. Determination of age, sex, and blood group from a single tooth[J]. *Journal of Forensic Dental Sciences*, 2017, 9(1): 10-14.
- [20] BERUS L, KLANCNIK S, BREZOCNIK M, et al. Classifying Parkinson's disease based on acoustic measures using artificial neural networks[J]. *Sensors*, 2019, 19(1): 16.
- [21] ISLAM M R, ISLAM M M, RAHMAN M M, et al. EEG channel correlation based model for emotion recognition[J]. *Computers in Biology and Medicine*, 2021, 136: 104757.
- [22] LEE M S, LEE Y K, LIM M T, et al. Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features [J]. *Applied Sciences*, 2020, 10(10): 3501.
- [23] FRIEDMAN L, NIXON M S, KOMOGORTSEV O V. Method to assess the temporal persistence of potential biometric features: Application to oculomotor, gait, face and brain structure databases [J]. *Plos One*, 2017, 12(6): e0178501.
- [24] SAPEY-TRIOMPHE L A, HECKEMANN R A, BOUBLAY N, et al. Neuroanatomical correlates of recognizing face expressions in mild stages of Alzheimer's disease[J]. *Plos One*, 2015, 10(12): e0143586.
- [25] RUJIRAKUL K, SO-IN C. Parallel optimized Pearson correlation condition (po-pcc) for robust cosmetic makeup facial recognition [J]. *The International Arab Journal of Information Technology*, 2019, 16(3): 442-453.
- [26] GOSCIIEWSKA K, FREJLICHOWSKI D. The analysis of shape features for the purpose of exercise types classification using silhouette sequences [J]. *Applied Sciences*, 2020, 10(19): 6728.
- [27] 苑玮琦, 荆澜涛, 林森, 等. 基于分类区分度和相关性的手形特征选择方法[J]. *仪器仪表学报*, 2013, 34(8): 1787-1794.
- YUAN W Q, JING L T, LIN S, et al. Hand-shape feature selection method based on class distinctive degree and correlation [J]. *Chinese Journal of Scientific Instrument*, 2013, 34(8): 1787-1794.
- [28] IRANMANESH V, AHMAD S M S, ADNAN W A W, et al. Online signature verification using neural network and Pearson correlation features [C]. *IEEE Conference on Open Systems*, 2013: 18-21.
- [29] LI C K, WONG S K, CHIM L C J. A prototype of mathematical treatment of pen pressure data for signature verification [J]. *Journal of Forensic Sciences*, 2018, 63(1): 275-284.
- [30] YEUNG D Y, CHANG H, XIONG Y, et al. SVC2004: First international signature verification competition [C]. *International Conference on Biometric Authentication*, 2004: 16-22.
- [31] GALBALLY J, MARTINEZ-DIAZ M, FIERREZ J. Aging in biometrics: An experimental analysis on on-line signature [J]. *Plos One*, 2013, 8(7): e69897.
- [32] TAMILARASI K, KALYANI S N. Design and implementation of deep learning strategy based smart signature verification system [J]. *Microprocessors and Microsystems*, 2020, 77: 103119.
- [33] TAN H, HE L, HUANG Z C, et al. Online signature verification based on dynamic features from gene expression programming [J]. *Multimedia Tools and Applications*, 2021: 1-27.
- [34] XIA X H, SONG X Y, LUAN F G, et al. Discriminative feature selection for on-line signature verification [J]. *Pattern Recognition*, 2018, 74: 422-433.
- [35] ZALASINSKI M, CPALKA K, ER M J. Stability evaluation of the dynamic signature partitions over time [C]. *International Conference on Artificial Intelligence and Soft Computing*, 2017: 733-746.

作者简介



刘若男,分别在2015年和2018年于沈阳工业大学获得学士和硕士学位,现为沈阳工业大学博士研究生,主要研究方向为生物特征识别和人机交互。

E-mail: lrn_123@sina.com

Liu Ruonan received her B. Sc. and M. Sc. degrees both from Shenyang University of Technology in 2015 and 2018, respectively. She is currently a Ph. D. candidate at Shenyang University of Technology. Her main research interests include biometric identification and human-computer interaction.



辛义忠(通信作者),分别在1997年和2004年于沈阳工业大学获得学士和硕士学位,2010年于日本高知工科大学获得博士学位,现为沈阳工业大学教授,博士生导师,主要研究方向为生物特征识别、机器视觉检测和

和人机交互。

E-mail: xyz@sut.edu.cn

Xin Yizhong (Corresponding author) received his B. Sc. and M. Sc. degrees both from Shenyang University of Technology in 1997 and 2004, respectively. He received his Ph. D. degree from Kochi University of Technology, Japan in 2010. He is currently a professor and a Ph. D. advisor at Shenyang University of Technology. His main research interests include biometric identification, machine vision and human-computer interaction.



李岩,分别在1998年和2007年于沈阳工业大学获得学士和硕士学位。现为沈阳体育学院讲师,主要研究方向为人机交互和生物特征识别。

E-mail: liyan@syty.edu.cn

Li Yan received her B. Sc. and M. Sc. degrees from Shenyang University of Technology in 1998 and 2007, respectively. She is currently a lecturer at Shenyang Sport University. Her main research interests include human computer interaction and biometric identification.