

基于信息融合的在线手写签名算法研究

张伟龙, 郑建彬, 詹恩奇

(武汉理工大学 信息工程学院, 武汉 430070)

摘要: 为了进一步提高认证效果,在演化计算、神经网络和离散 F 距手写签名认证算法的基础上,提出了基于信息融合的在线手写签名认证算法。该算法将测试签名和参考签名分别通过三种算法进行认证,得出测试签名为真实签名的置信度,然后对三种认证算法的结果进行加权融合,根据最终的融合结果进行签名真假的判定。实验结果表明,信息融合算法的误拒率和误纳率都有显著的减少。

关键词: 数字签名; 签名认证; 信息融合; 置信度

中图分类号: TP391.41 文献标志码: A 文章编号: 1001-3695(2010)05-1889-03

doi:10.3969/j.issn.1001-3695.2010.05.081

Research of on-line handwriting signature verification algorithm based on information fusion

ZHANG Wei-long, ZHENG Jian-bin, ZHAN En-qi

(School of Communication & Information System, Wuhan University of Technology, Wuhan 430070, China)

Abstract: In order to improve this verification effect, the paper proposed the algorithm of on-line handwriting signature verification based on information fusion. The algorithm fused three kinds of algorithms of on-line handwriting signature verification respectively based on evolution computation, neural network and discrete Fréchet distance. Verified the test signature by the three kinds of algorithms to obtain its confidence level of every algorithm. Fused the last confidence level of the signature by the confidence level of every algorithm by calculating their weights, and it could be used to judge whether the test signature was genuine. The experiment results reveal the remarkably decreasing percentage of false rejection and false acceptance.

Key words: digital signature; signature verification; information fusion; confidence level

0 引言

在线签名认证算法从 20 世纪 80 年代起,就有很多学者致力于这方面的研究。比较有代表性的有: Mohankrishnan 等人^[1]提出的一个基于自回归 (autoregressive) 模型的签名认证方法、Yang 等人^[2]提出的隐马尔可夫模型 (hidden Markov model) 法和 Connell 提出的基于特征矢量匹配的方法等,但单种方法的效果都不理想。最近,郑建彬也分别提出了基于演化计算、BP 网络和离散 F 距的在线手写签名认证算法。

在线手写签名认证是通过手写板实时采集书写人的签名信息,除了可以采集签名位置信息,还可以记录书写时的速度、运笔压力等动态信息。通常在手写板上能采集到 X 坐标、Y 坐标以及压力值,图 1 是某签名及其采集的 X 坐标信号。参考签名和测试签名是否匹配,主要是比较两者所对应的曲线是否能够匹配。

基于演化计算的签名认证算法是将参考签名曲线分割成曲线段,以一定长度的搜索窗在测试签名曲线上进行动态搜索,实现与参考签名曲线段自适应的动态分割与匹配,根据匹配成功的段数与总的匹配段数的比值来判定两个签名是否匹配^[3]。基于 BP 网络的签名认证算法是将参考签名提取的特征输入到 BP 网络中得到 BP 网络各层的权值,然后将测试签

名提取的特征输入到 BP 网络中,根据输出值判定测试签名的真伪^[4]。基于离散 F 距的签名认证算法是提取测试签名曲线和参考签名曲线的波峰点与波谷点,计算波峰点间和波谷点间离散 F 距,将它们的差值的绝对值作为最终的离散 F 距,通过得到的离散 F 距和预先设定的阈值进行比较来判断签名的真伪^[5]。这三种算法各有其优缺点,基于演化计算的签名认证算法的误拒率比较高,误纳率比较低,而基于 BP 网络的签名认证算法有较低的误拒率和较高的误纳率。信息融合技术可以将这三种算法认证结果融合起来,克服它们的缺点,利用好它们的优点,提高签名的识别率。

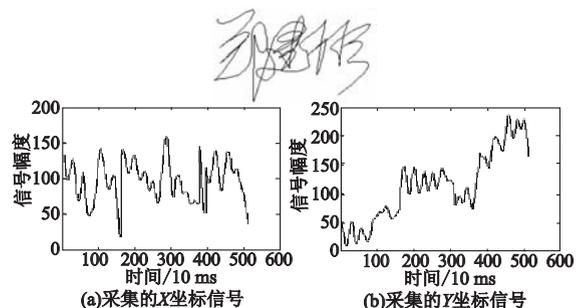


图 1 某签名及其采集的 X 和 Y 坐标信号

早在 20 世纪 70 年代末期,在一些公开出版的文献中就开始出现有关信息融合的概念或名词。在其后的较长一段时期,

收稿日期: 2009-08-22; 修回日期: 2009-10-22

作者简介: 张伟龙(1985-),男,山东东营人,硕士,主要研究方向为在线手写签名应用(zhangweilong1221@yahoo.com.cn); 郑建彬(1966-),男,湖北黄冈人,教授,硕士,博士,主要研究方向为智能信息处理、信息网络集成技术等; 詹恩奇(1972-),男,河南新野人,讲师,博士,主要研究方向为通信与信号处理。

人们普遍使用数据融合这一名词。直到 20 世纪 90 年代,考虑到传感器信息的多样性,信息融合一词才被广泛采用。最初信息融合主要是应用于军事领域,把从敌方多个目标获取的信息进行融合,从而得到比较全面的、准确的情报^[6]。目前对信息融合的研究,主要是指利用计算机进行多传感信息处理,从而得到可综合利用信息的理论和方法,在数字图像处理方面有比较广泛的应用^[7]。

本文提出了一种新的基于信息融合的在线手写签名认证算法。该算法是将测试签名分别通过演化计算、BP 网络和离散 F 距的认证算法,根据认证结果得到测试签名为真实签名的置信度 $YHCL_T$ 、 $BPCL_T$ 、 $LSCL_T$,然后通过融合公式将它们置信度进行加权融合,得到最终的置信度 CL_T ,根据判定规则来判定测试签名的真伪,如图 2 所示。

1 签名认证算法置信度计算

签名数据库的形成:在一个月的时间内,共采集了 1 075 个真实签名、752 个熟练伪造签名和 853 个随机伪造签名。

置信度是判定结果的可信程度。签名判定为真实签名的置信度称为真实签名的置信度。置信度是分布在区间 $[-1, 1]$ 上的一个值。真实签名的置信度的分布区间是 $[0, 1]$,如果得出的置信度在区间 $[-1, 0]$ 上,则签名判定为伪造签名。

1.1 基于演化计算签名认证算法的置信度计算

1.1.1 基于演化计算的签名认证算法的置信度初始值

基于演化计算的签名认证算法的匹配率可以作为它的置信度初始值。该比值越大说明匹配越成功,真实签名的概率就越大^[3]。

1.1.2 基于演化计算的签名认证算法的置信度初始值的统计分布情况

统计数据库中真实签名和真实签名的匹配率与伪造签名和真实签名的匹配率,得到两组数据。对这两组数据作期望和方差计算,得到真实签名匹配率的期望 μ_{yT} 和标准差 δ_{yT} 为 0.792 4 与 0.202 1;伪造签名和真实签名匹配率的期望 μ_{yF} 和标准差 δ_{yF} 为 0.329 9 与 0.319 8。

1.1.3 签名匹配率的最终置信度生成

通过对统计的真实签名和伪造签名的匹配率 M_Y 的分析,得到了匹配率和它的置信度 $YHCL_T$ 的关系曲线如图 3 所示。当匹配率小于 $\mu_{yT} - \delta_{yT}$ 时,真实签名的置信度为 0;当匹配率大于 $\mu_{yT} + \delta_{yT}$ 时,伪造签名的置信度为 0。

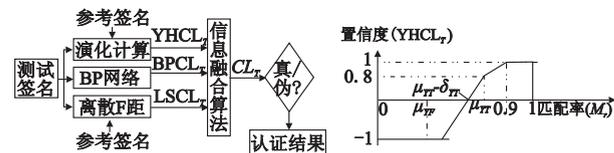


图2 基于信息融合的签名认证流程图

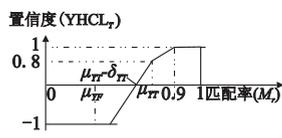


图3 演化计算的匹配率和置信度的关系曲线

1.2 基于 BP 网络签名认证算法的置信度计算

1.2.1 基于 BP 网络的签名认证算法的置信度的初始值

本文选择的 BP 网络输出层神经元只有一个。该输出期望值为 1。输出值越接近 1,签名越真实。所以将输出值作为 BP 网络的置信度初始值。

1.2.2 BP 网络输出值的统计分布情况

对数据库中的签名通过 BP 网络签名认证算法后的输出值

作统计,可得出真实签名输出值的期望 μ_{BT} 与标准差 δ_{BT} 为 0.946 7 和 0.051 2;伪造签名输出值的期望 μ_{BF} 和标准差 δ_{BF} 为 0.725 2 与 0.187 2。

1.2.3 BP 网络输出值的最终置信度生成

通过对真实签名和伪造签名的输出值的统计分析,得出了 BP 网络的输出值 O_{BF} 和它的置信度 $BPCL_T$ 之间的关系曲线,如图 4 所示。当输出值 $O_{BF} < \mu_{BT} - \delta_{BT}$ 时,真实签名的置信度为 0;当输出值 $O_{BF} > \mu_{BT} + \delta_{BT}$ 时,伪造签名的置信度为 0。

1.3 基于离散 F 距签名认证算法的置信度计算

该算法是根据测试签名和参考签名的离散 F 距 D_F 与预先设定的阈值比较来判定签名的真伪的,离散 F 距越小说明测试签名和参考签名越相似^[5]。对数据库中的签名进行分析得到真实签名的离散 F 距的期望 μ_{LT} 和标准差 δ_{LT} 为 87.183 9 与 118.97。伪造签名的离散 F 距的期望 μ_{LF} 和标准差 δ_{LF} 为 978.814 与 120.685。提取数据库中签名人的所有真实签名,每两个签名作为一组,计算每组签名的离散 F 距,然后求得的所有离散 F 距的均值设为 μ ,签名人的离散 F 距 D_F 与其置信度 $LSCL_T$ 之间的关系曲线如图 5 所示。

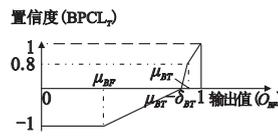


图4 BP网络的输出值及其与置信度之间的关系曲线

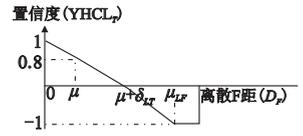


图5 离散F距与其置信度之间的关系曲线

2 三种算法的置信度的融合

2.1 融合公式

信息融合的方法有很多种,包括加权平均法、投票法、贝叶斯估计法、证据推理法、支持向量机、D-S 证据理论等。投票法过于简单,其准确性比较低。贝叶斯估计法、证据推理法和 D-S 理论都是证据推理的方法,利用其先验概率来求后验概率。而支持向量机的方法是将向量从低维空间投射到多维空间,在高维空间对数据进行分析。本文只是将三种算法得出的置信度进行融合,不需要太复杂的推导,所以采用的是加权平均法中的一种——线性加权融合规则^[8],如式(1)所示:

$$CL_T = AW_Y \times YHCL_T + AW_B \times BPCL_T + AW_L \times LSCL_T \quad (1)$$

其中: CL_T 为签名是真实签名的融合后的置信度; AW_Y 、 AW_B 、 AW_L 分别为演化计算、BP 网络和离散 F 距三种算法真实签名的融合权值; $YHCL_T$ 、 $BPCL_T$ 、 $LSCL_T$ 为签名通过三种算法判定后得到的签名为真实签名的置信度。

2.2 权值计算

由于签名算法的准确度是以它的误拒率和误纳率来判定的,误拒率和误纳率越低,说明算法的准确度越高,那么它的判定结果在融合过程中所占的比重就越大,权值就越大。

当理想情况下,三种算法对某个签名判定的结果是签名为真实签名的置信度均为 1,那么融合后的签名为真实签名的置信度应该也为 1,将 $YHCL_T = BPCL_T = LSCL_T = 1$ 和 $CL_T = 1$ 代入到式(1)中可得:

$$AW_Y + AW_B + AW_L = 1 \quad (2)$$

演化计算、BP 网络和离散 F 距三种算法的平均误拒率分别设为 $APFR_i (i = 1, 2, 3)$,根据线性加权融合规则可得权值的计算公式为

$$AW_i = \frac{1}{\gamma_{a,i}^2 \times \sum_{j=1}^3 \gamma_{a,j}^{-2}} \quad (3)$$

其中: $\gamma_{a,i}$ 可根据式(4)得到:

$$\gamma_{a,i} = \frac{1}{100 - 100 \times APFR_i} \quad (4)$$

根据三种算法的误拒率和误纳率可以求得它们的权值分别为 $AW_Y = 0.3274$ 、 $AW_B = 0.3427$ 、 $AW_L = 0.3299$ 。将这些权值代入到式(1)中就可以求得签名为真实签名的置信度 CL_T 了。

2.3 判定规则

如果 $CL_T > 0$, 则签名判定为真实签名; 如果 $CL_T < 0$, 则签名判定为伪造签名; $CL_T = 0$, 则签名无法判定真伪。

3 实验结果分析

图 6 是用手写板采集的张伟龙的真实签名的 X 和 Y 坐标信息, 这些坐标信息被保存到一个后缀为 .h 的文件, 并存放到签名数据库中。

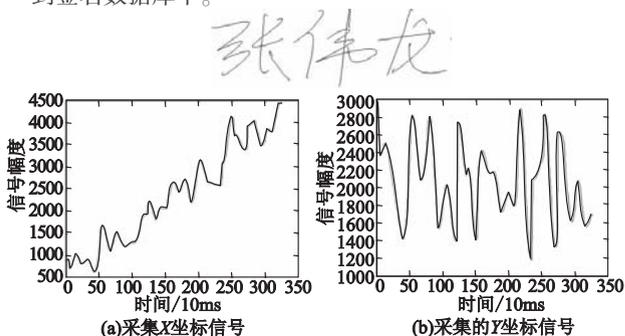


图6 张伟龙的真实签名及采集的X和Y坐标信号

在签名数据库中提取“张伟龙”的五个真实签名 ZWL01.h、ZWL02.h、ZWL03.h、ZWL04.h 和 ZWL05.h 作为参考签名, 五个真实签名 ZWL06.h、ZWL07.h、ZWL08.h、ZWL09.h、ZWL10.h 和五个熟练伪造签名 Zs01.h、Zs02.h、Zs03.h、Zs04.h、Zs05.h 作为测试签名。将测试签名分为真实签名和伪造签名两组, 并与参考签名作实验分析。测试签名和参考签名可通过三种算法得到演化计算的匹配率、BP 网络的输出值和离散 F 距的输出值, 再根据它们和各自的置信度的关系曲线, 如图 3~5 所示, 可以得到三种算法的置信度和融合后的置信度, 如表 1 和 2 所示。

表 1 张伟龙的真实签名经过三种算法和融合算法后的匹配情况

测试签名	参考签名	YH_T	BP_T	LS_T	CL_T	结果
ZWL06.h	ZWL01.h	0.87	0.99	0.84	0.90	真
ZWL07.h	ZWL02.h	0.92	0.85	0.95	0.91	真
ZWL08.h	ZWL03.h	0.99	0.87	-0.99	0.29	真
ZWL09.h	ZWL04.h	0.72	0.88	0.74	0.78	真
ZWL10.h	ZWL05.h	0.99	0.65	0.99	0.87	真

表 2 张伟龙的伪造签名经过三种算法和融合算法后的匹配情况

测试签名	参考签名	YH_T	BP_T	LS_T	CL_T	结果
Zs01.h	ZWL01.h	-0.86	0.91	-0.97	-0.29	伪
Zs02.h	ZWL02.h	-0.99	0.88	-0.99	-0.35	伪
Zs03.h	ZWL03.h	-0.91	-0.99	-0.91	-0.94	伪
Zs04.h	ZWL04.h	-0.92	-1.00	-0.88	-0.93	伪
Zs05.h	ZWL05.h	0.92	-0.99	-0.91	-0.34	伪

将签名数据库中的部分真实签名作为参考签名, 剩余的真实签名及伪造签名作为测试签名。将测试签名和与之对应签名者的参考签名分别通过三种签名认证算法和融合算法。统计签名的匹配结果计算它们的误拒率和误纳率, 如表 3 所示。

表 3 融合后算法及三种认证算法误拒率和误纳率统计表 %

算法名称	误拒率	误纳率
基于演化计算的签名认证算法	8.43	5.67
基于神经网络的签名认证算法	6.30	10.71
基于离散 F 距的签名认证算法	8.07	7.75
融合算法	4.87	5.93

从表 3 中可以看出, 三种认证算法融合后的误拒率比融合前各自的误拒率降低了, 误纳率虽然高于基于演化计算的签名认证算法的误纳率, 但相比其他两种算法的误纳率降低了很多, 是可以接受的, 这样就实现了三种签名认证算法的优缺点的互补, 提高了签名认证的准确度。

4 结束语

本文将数据融合的思想应用到了在线手写签名认证中, 在原有的三种签名认证算法的基础上, 对其进行加权融合。由实验分析, 把数据融合应用到签名认证中, 在保证认证速度的基础上, 使签名认证算法的准确度有了提高。实验结果证明了该方法的可行性。但是, 该算法还有一些问题, 比如在认证过程中需要某人的离散 F 距, 得到的是根据他的几次真实签名进行匹配得到的离散 F 距的均值, 这个值可能存在一些误差给判定结果带来错误。利用线性加权融合规则来进行融合, 三种算法的连接方式是并联的, 利用串联或混联的连接方式的准确度可能比并联的更高, 但线性加权融合规则不再适用, 而且理论分析更为复杂, 本文不予以讨论。该算法需要大量的数据作为基础, 随着时间的推移, 人的签名习惯也可能发生变化, 所以数据库要及时更新。

参考文献:

- [1] MOHANKRISHNAN N, PAULIK M J, KHALIL M. On-line signature verification using a nonstationary autoregressive model representation [J]. IEEE Int Sym on Circuits and System, 1993(2): 2303-2306.
- [2] YANG L, WIDJAJA B R, PRASAD R. Application of hidden Markov model for signature verification [J]. Pattern Recognition, 1996, 28(2): 180-188.
- [3] ZHENG Jian-bin, ZHU Guang-xi. A new algorithm for on-line handwriting signature verification based on evolutionary computation [J]. Wuhan University Journal of Natural Science, 2006, 11(3): 596-600(EI).
- [4] ZHENG Jian-bin, MA Chan. Applied research of on-line handwriting signature verification in electronic document security [C]// Proc of the 3rd International Symposium on Progress in Intelligence Computation and Applications. 2008: 91-95.
- [5] ZHENG Jian-bin, GAO Xiao-lei, ZHAN En-qi, et al. Algorithm of on-line handwriting signature verification based on discrete Fréchet distance [C]// Proc of the 3rd International Symposium on Advance in Computation and Intelligence. 2008: 461-469.
- [6] PAN Quan, YU Xin, CHENG Yong-mei, et al. Essential methods and progress of information fusion theory [J]. Acta Automatica Sinica, 2003, 29(4): 600-609.
- [7] FREEDMAN D D. Overview of decision level fusion techniques for identification and their application [C]// Proc of American Control Conference on American Automatic Control Council, International Federation of Automatic Control. 1994: 1299-1303.
- [8] SCOTT D C. A data fusion approach to verifying hand-writing signature on bank cheques [C]// Procs of Applied Imagery Pattern Recognition Workshop N^o27 1998. Washington DC: Society of Photo-Optical Instrumentation Engineers, 1999: 84-95.