

相似度算法在手写签名质谱成像鉴定中的应用研究

刘亚丽^{1,2}, 贾滨², 丁丽英², 吴佳², 陈焕文², 章文军^{1*}

(1. 河北工业大学化工学院, 天津, 300130;

2. 东华理工大学/江西省质谱科学与仪器重点实验室, 江西, 南昌, 330013)

摘要: 质谱成像技术无需任何样品预处理, 即可获取待测样品的分子信息和分布情况。本文采用表面解吸常压化学电离质谱(SDAPCI-MS)技术对手写签名样品进行检测, 通过对所得的质谱特征峰信号进行成像处理, 获取书写油墨分布的强度信息。实验结果表明真实签名和伪造签名因为笔压轻重不同而油墨分布位置不同, 据此能够区分签名的真伪。应用相似度算法对手写签名的特征成像数据进行分析, 在样品量较少的情况下, 能够客观的比较真迹之间以及真迹和伪迹之间的相似程度, 进而实现对真伪签名的准确鉴定。本方法操作简便, 耗时较短, 对手写笔迹的鉴定具有重要的借鉴意义。

关键词: 质谱成像; 相似度算法; 手写签名

中图分类号: TQ015.9; TP391.9; O6-39

文献标识码: A

文章编号: 1001-4160(2012)05-541-544

1 引言

手写签名的鉴定一般分为在线(on-line)签名鉴定和脱线(off-line)签名鉴定, 手写签名在金融、商务、司法、合同、保险、社会生活的各个方面都有广泛应用, 如签订合同、银行存取款等。一般而言, 签名与个人的体质、幼时训练以及个人性格有关, 而且每个人的签名还会随着心理和生理状态以及书写者所用的工具和所处的环境而变化。脱线签名一般仅仅依靠签名图像得到的静态信息, 从图像外形上看伪造签名与真实签名极为相似, 所以对签名真伪的鉴定难点就是人们签名的可变性、不稳定性以及模仿签名的相似性。对离线手写签名鉴定的研究方法主要有从签名图像中提取静态信息^[1]、BP神经网络、SVM支持向量机方法、小波变换^[2]、遗传算法、隐马尔可夫模型、还有一些对伪动态信息^[3]来进行鉴别、局部动态特征^[4]等等。大多数研究基于提取签名图像的静态特征, 针对手写签名的动态特征和同一个人其本身签名差别的研究很少。

质谱技术在许多领域都有很重要的应用, 例如食品安全检测, 食品中残留农药、苏丹红染料的检测^[5], 药品质量检测, 环境安全的检测、原位活体的分析, 以及代谢组学分析和蛋白质组学分析等^[6]。质谱成像技术(Mass Spectrometry Imaging, 简称 MSI)能够在无损样品的情况下在样品的表面进行取样, 获取样品表面的分子信息和分子空间分布情况, 最终获得样品中待测物的质谱成像图。近年来, 直接质谱成像技术在生物组织成像的应用领域有较大发展, 并且在日常生活中也得到了广泛应用, 例如丁丽英等对衣服上的致癌性物质芳香胺进行质谱成像得到芳香胺在与身体接触与不接触的位置的分布图^[7], 从图像上可以明显看出芳香胺的分布情况。

相似方法一般是用待检测物与已知标准物进行比较得出两者的相似结果, 相似度算法在许多方面有着很重要的应用, 例如在人脸识别^[8]中, 先从人脸图像中提取出识别特征, 利用相似匹配的方法识别。相似度匹配的方法有欧式距离(Euclidean Distance), 余弦距离(Cosine Distance), 马氏距离(Mahalanobis Distance), 文本文档相似度的计算^[9], DNA 编码序列及相似度的计算^[10], 用计算机对 C₁₃NMR(核磁共振)质谱等波谱图进行相似化合物检索^[11]。

本实验通过观察签名的质谱成像图看出, 一个人临摹别人的签名从外观轮廓上看十分相似, 但人的书写习惯不同写字时笔压轻重不同导致油墨分布多少位置不同。本文通过对手写签名的质谱成像数据预处理, 依据笔压轻重位置应用设定阈值的方法提取峰强度得到特征值, 构造合适的相似度算法对不同的笔迹进行相似度计算, 进而根据相似程度大小鉴别签名的真伪。

2 实验部分

2.1 仪器与试剂

SDAPCI 离子源(东华理工大学研制); LTQ-XL 增强型线性离子阱质谱仪(美国 Finnigan 公司产品), 配有 Xcalibur 数据处理系统; 二次蒸馏水。

2.2 样品

真迹样品: 由胡斌本人在不同时间、地点及情绪下写的 5 个签名“binhu”;

伪迹样品: 找其他人模仿胡斌的五个签名“binhu”。

2.3 质谱成像实验

实验设置 SDAPCA 离子源为正离子检测模式, 质量

收稿日期: 2012-01-15; 修回日期: 2012-04-16

基金项目: 国家重大科学仪器设备开发专项(2011YQ170067)和国家自然科学基金资助项目(21175019)

作者简介: 刘亚丽(1985—), 女, 山东人, 硕士研究生, E-mail: helloliuyali@163.com

联系人: 章文军, E-mail: wjzhang@hebut.edu.cn

范围(50~600) u, 电离电压 3.5 KV, 离子传输管温度 200 °C, 解析气(N₂)气压 0.15 MPa, 以 α 为 30 °C 从装有解析液的锥形瓶喷出, N₂ 出口到测定点距离是 1 mm, 二级质谱选定碰撞能量为 31%, 离子的隔离宽度 1.5 u, 其他实验参数由系统自动优化。实验前先将写好签名纸张的图片样品引入 AutoCAD2007 软件中, 定义好原点坐标, 用 Excel 表格记录下来质谱扫描区域内的各个点的坐标值(X,Y), 将纸张放入具有三位调节功能的样品台上, 采用表面解析常压化学电离质谱仪(SDAPCI), 按照事先设定好的点顺次扫描 X、Y 两个方向。测定时线性改变样品台的位置, 每间隔 0.45 mm 采集一次数据。每个坐标点采集一次数据, 并且记录每次扫描的时间和相对应点的坐标, 然后在谱图上找到这个点对应的峰度值(Z 值), 将每一个点对应的 X、Y、Z 值分别对应输入到 Excel 表格中, 把 Excel 数据表格导入到 Surfer 8 软件中成像, 得到手写签名的质谱成像图。从图 1(b)图像上可以看出油墨强弱分布位置的不同。

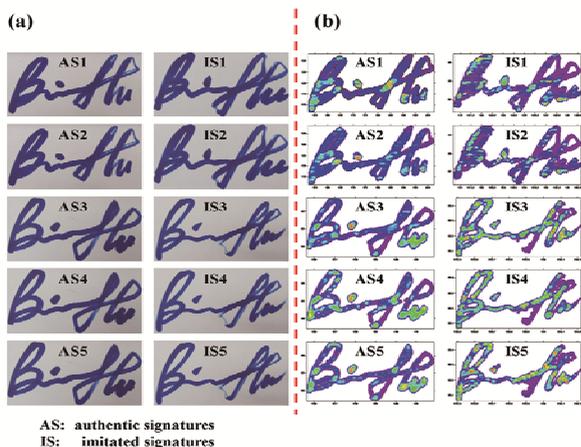


Fig.1 The original figure of the authentic signature and the imitated signature and the corresponding mass spectrometry imaging (MSI).
图 1 由胡斌本人签名和由其他人模仿胡斌的签名图以及签名对应的质谱成像图

2.4 质谱成像数据预处理

通过质谱成像实验得到的三维数据比较复杂, 不能直接用来做相似度的计算, 需要对数据预处理将三维的数据降为二维。通过将 x 轴和 y 轴所在平面的这些点阵均匀划分成 num×num 个网格进行数据降维, 分别找到 x 和 y 的最大和最小值, 记为 xmax、xmin、ymax、ymin, x 轴网格边长为: $dx=(xmax-xmin)/num$, y 轴网格对应边长为: $dy=(ymax-ymin)/num$, 从坐标原点开始先固定 x 所在的网格坐标, 对 y 轴顺次进行数据的处理, 如果原来坐标的 x 坐标值大于等于 xmin 并且小于等于 xmin+dx, 对应的 y 坐标如果大于等于 ymin 小于等于 ymin+dy, z 的值等于此时强度值和上一个强度值的累加和除以 num, 如果 x 和 y 坐标值不满足上述条件, z 值就等于 0, 在实验中我们先固定 x 的值取 1, y 的值依次取: 1、2、3·····num, x 值再依次取 1、2、3·····num, 提取网格对应的强度值作为特征

值, 特征向量的维数为网格数 num×num。

2.5 相似度的算法

C₁₃NMR 谱检索算法主要有两类: 特征峰(即单峰)检索和全谱图匹配(相似度)检索。为了比较两张谱图的接近程度可用数学模型计算相似度, 将这种相似度的计算应用于手写签名的真伪笔迹鉴定, 因为每个人的每一次手写签名不可能完全相同, 会随着心情变化而签名的笔压、连笔和省略的部分有所不同, 所以对签名的真伪鉴定算法需要局部相似并具有一定的包容性, 这和 C₁₃NMR 谱检索算法的要求是一致的。在实验中, 将手写签名较强的笔压油墨所在位置的峰强度值提取出来作为特征值, 这个特征值相当于谱图检索中的化学位移值。所谓的相同化学位移是指在一定的允许偏差范围内的相同, 对于偏差的允许数值需要在实验过程中进行验证, 对一维的核磁共振 C₁₃NMR 技术相似检索^[11]进行了改造, 使其适应于二维平面的笔迹特征的相似比较。其算法如下:

$$S = S_1 / S_2$$

$$S_1 = M_3 \cdot \sqrt{M_1 \cdot M_2}$$

$$S_2 = (1 + 1.2F)(1 + 0.8D) \cdot M_1 \cdot M_2$$

$$F = (M_1 \cdot M_2 - M_1 \cdot M_3 - M_2 \cdot M_3 + M_3^2) / (M_1 \cdot M_3)$$

$$D = \sum_{i=1}^{M_3} D_i^2 / (M_3 \cdot \sqrt{M_3} \cdot T^2)$$

M_1 =要比对的笔迹的特征值个数,

M_2 =被比对的笔迹的特征值个数,

M_3 =要比对与被比对的笔迹的相同的特征值个数,

T =给定的最大允许偏差,

D_i =要比对的笔迹的特征值与第 i 个被比对的笔迹的特征值的距离之差,

在实验中, 对 num 赋值进行相似度比较就可以得到相似度数值。

3 结果与讨论

相似检索的相似度算法得到的相似度值如下表 1 所示, 其中比较了真迹和真迹的相似度以及真迹和伪迹的相似度。

在本次实验中, 设定 num=30, T 为 85%, 由实验结果可以看出真迹和真迹的签名相似程度都在 90% 以上, 而真迹和伪迹的相似程度都在 70% 以下。

在实验中, 对笔迹特征值的选取非常关键, 如果对阈值数据设定过大, 可能不是签名笔压较重的位置被认为是特征值所在的位置, 如果设定值过小, 对于签名的整体数据会有所遗漏, 使得签名笔压较重的位置被忽略。数据的预处理对相似度结果有一定的影响, 网格数目的

表 1 真迹和真迹以及真迹和伪迹的相似度数值

Table 1 The similarity of the authentic signatures and the authentic signatures along with the authentic signatures and the imitated signatures.

标准	真迹 1	真迹 2	真迹 3	真迹 4	真迹 5
真迹 1	100	95.76	97.45	96.45	95.61
真迹 2	95.77	100	95.49	94.36	95.10
真迹 3	97.45	95.42	100	96.59	96.92
真迹 4	96.45	94.35	96.59	100	96.92
真迹 5	95.62	95.04	96.92	96.26	100
伪迹 1	66.62	65.31	67.41	66.70	66.11
伪迹 2	66.88	65.56	67.67	66.95	66.37
伪迹 3	67.26	66.07	67.41	66.72	66.27
伪迹 4	66.00	64.20	66.24	66.06	65.54
伪迹 5	67.24	65.40	67.47	66.77	66.19

多少会决定特征值的数目, 在实验中分别对 10×10 , 15×15 , 20×20 , 25×25 , 30×30 , 35×35 个网格的数据进行相似度的比较, 发现随着网格数目的增加真迹与真迹的相似程度随之提高而真迹与伪迹的相似程度随之降低, 考虑到划分太多的网格数目有可能使得原来非笔压重的地方被划分成特征值所在的地方, 就会明显使真迹和伪迹签名相似程度下降, 在本实验中取 $\text{num}=30$ 就已经取得很好的鉴定效果。

实验结果表明改进的相似度算法能够对手写签名的真伪进行有效鉴定, 这个算法对于同一个人的笔迹鉴定具有较好的识别作用和包容性, 对样本数的要求不高, 可以快速准确的鉴别签名真伪, 但是在具体鉴定过程中对参数的设置对鉴定结果会有一定的影响。

References:

- 1 Nemeck W F, Lin W C. Experimental investigation of automatic signature verification. IEEE Trans. on Syst Man Cybernet, 1974, SMC 4(1):121-126.
- 2 Deng P S, Liao H Y M, Ho C W, et al. Offline signature verification and identification by hybrid features and Support Vector Machine Computer. Vision and Image Understanding, 1999, 76(3):173-190.
- 3 Ammar M, Yoshida Y, and Fukumura T. A new effective approach

- for off-line verification of signatures by using pressure features. Proceedings of The 8th International Conference on Pattern Recognition. Paris, France, 1986:176-184.
- 4 Lin Jun and Li Jiegu. Feature extraction and preprocessing of off-line Chinese signature verification. Journal of Shanghai Jiaotong University, 1996, 30(9):112-118.
- 5 Chen Huanwen, Zhang Xie and Luo Mingbiao. Desorption electrospray ionization mass spectrometry for fast detection of sudan dyes in food without sample pretreatment. Chinese J Anal Chem, 2006, 34(4):464-468.
- 6 Chen Huanwen, Hu Bin, and Zhang Xie. Fundamental principles and practical applications of ambient ionization mass spectrometry for direct analysis of complex samples. Chinese J Anal Chem, 2010, 38(8):1069-1088.
- 7 Ding Lingying, Hu Bin and Yang Shuiping, et al. Imaging carcinogenic aromatic amines in textiles by surface desorption atmospheric pressure chemical ionization tandem mass spectrometry. J Chinese Mass Spectrom Som, 2010, 31:137-140.
- 8 Gu Rui, Zhang Shuling and Wang Xiaofen. Study of feature extraction and similarity match method on face recognition. Chinese J Computer Engineering, 2006, 32(11):225-228.
- 9 Cai Yuanzhe, Li Pei and Liu Hongyan, et al. S-SimRank: combining content and link information to cluster papers effectively and efficiently. Journal of Frontiers of Computer Science and Technology, 2009, 3(4):379-391.
- 10 Zhang Baohua, Wang Haishui and Xu Lu. Codes of DNA primary sequences and the similarity calculation. Chemical Journal Of Chinese Universities, 2006, 27(12):2277-2280.
- 11 Xu Lu, Li Guoquan. Computer analysis of C₁₃NMR spectrum -similar compounds search. Chemistry Bulletin, 1987:44-47.

中文参考文献

- 4 林峻, 李介谷. 离线中文签名鉴别的特征提取及预处理[J]. 上海交通大学学报, 1996, 30(9):112-118.
- 5 陈焕文, 张燮, 罗明标. 电喷雾解析电离质谱法对食品中苏丹红染料的快速检测[J]. 分析化学, 2006, 34(4):464-468.
- 6 陈焕文, 胡斌, 张燮. 复杂样品质谱分析技术的原理与应用[J]. 分析化学, 2010, 38(8):1069-1088.
- 7 丁丽英, 胡斌, 杨水平, 等. 纺织品中致癌性芳香胺的质谱成像[J]. 质谱学报, 2010, 31:137-140.
- 8 郭瑞, 张淑玲, 汪小芬. 人脸识别特征提取方法和相似度匹配方法研究[J]. 计算机工程, 2006, 32(11):225-228.
- 9 蔡元哲, 李佩, 刘红岩, 等. S-SimRank:结合内容和链接信息的文档相似度计算方法[J]. 计算机科学与探索, 2009, 3(4):379-391.
- 10 张宝华, 王海水, 许禄. DNA 序列编码及相似度计算[J]. 高等学校化学学报, 2006, 27(12):2277-2280.
- 11 许禄, 李国权. 碳₁₃NMR 波谱谱图的计算机解析-相似化合物检索法[J]. 化学通报, 1987:44-47.

Studies on verification of handwritten signatures by mass spectrometry imaging with the Similarity Algorithm

LiuYali^{1,2}, Jia Bin², Ding Liying², Wu Jia², Chen Huanwen² and Zhang Wenjun^{1*}

(1. Hebei University of Technology, Tianjin, 300130, China)

(2. East China Institute of Technology, Jiangxi Key Laboratory for Mass Spectrometry and Instrumentation, Jiangxi, 330013, Nanchang, China)

Abstract: Without any sample pretreatment, mass spectrometry imaging (MSI) can provide molecular information and the molecular distribution information on the sample. Handwritten signature samples were directly detected by surface desorption atmospheric pressure chemical ionization mass spectrometry (SDAPCI-MS). The mass spectra recorded by SDAPCI-MS were subjected to data processing. It is easy to distinguish the authentic signature and the forged signature by observing the strength of the corresponding mass spectrometry imaging of the handwritten signature. Considering the relative ink concentration feature, which is extracted from the data of the handwritten signature, as a feature of the signature, we

