

DOI: 10.13382/j.jemi.2017.06.012

非线性线条痕迹小波域特征快速溯源算法研究^{*}

潘楠¹ 阚立峰² 潘地林³

(1. 昆明理工大学 航空学院 昆明 650500; 2. 昆明理工大学 机电工程学院 昆明 650500;
3. 昆明信诺莱伯科技有限公司 昆明 650228)

摘要: 高铁线缆盗割案件案发现场遗留的断头承痕面上存在着大量线条痕迹,其往往呈现非线性形态特征,随机性较强。为了更快速地进行痕迹特征分析及所属工具推断,设计出一种针对非线性线条痕迹的小波域特征快速溯源算法,该算法首先将单点激光位移传感器检测断头表面拾取的一维信号利用小波分解进行降噪,随后使用基于小波特征的动态时间规整算法实现痕迹特征相似重合度匹配,最后使用基于梯度下降法的线性回归机器学习算法不断的迭代使得代价函数值代价最小,从而实现对应工具快速推断。最终通过实际痕迹推断剪切工具试验验证了本算法的实用性和有效性。

关键词: 线条痕迹;小波降噪;动态时间规整;机器学习;快速溯源

中图分类号: TN911.72;TH39 **文献标识码:** A **国家标准学科分类代码:** 510.4010

Study of fast tracing algorithm based on wavelet domain features of nonlinear line trace

Pan Nan¹ Kan Lifeng² Pan Dilin³

(1. Faculty of Aviation, Kunming University of Science & Technology, Kunming 650500, China;
2. Faculty of Mechanical and Electrical Engineering, Kunming University of Science & Technology, Kunming 650500, China;
3. Kunming SNlab Co. Ltd., Kunming 650228, China)

Abstract: There is a large number of line traces on the surface of the broken end which left in the cable cutting case crime scene of high-speed railway, they often present nonlinear morphological feature and has strong randomness. In order to make rapid trace analysis and infer the tools, a fast algorithm based on wavelet domain feature aiming at the nonlinear line traces is presented. The proposed algorithm applies wavelet decomposition to the 1-D signal which picked up by single point laser displacement sensor to partially reduce noises. After that, the dynamic time warping is employed to realize the trace feature similarity matching. Finally, the linear regression machine learning algorithm based on gradient descent method is used to realize the constant iteration. The experiment results of cut traces sample data comparison demonstrate the accuracy and reliability of the proposed algorithm.

Keywords: line trace; wavelet de-noising; dynamic time warping; machine learning; fast tracing

1 引言

我国高速铁路发展迅速,为确保高铁长期稳定运行,铁路沿线分布着大量光缆、漏泄电缆、信号电缆、轨道交通车辆用电缆和贯通地线。由于各类线缆内导体多为铜

质,经济价值较高,逐渐成为了犯罪分子觊觎的对象。近年来针对高铁沿线线缆的割盗案件频发,不但给国家财产造成巨大损失,且易造成铁路信号及通信设备电源供电中断,迫使列车故障停运,进而引发更为严重的铁路交通事故,给人民生命财产安全带来重大损失^[1]。

根据统计,犯罪分子在作案时,多使用断线钳、线缆

收稿日期:2017-01 Received Date: 2017-01

^{*} 基金项目:云南省科技计划(2014SC030,2016RA042,2017EH028)、公安部技术研究计划(2014JSYJA020,2016JSYJA03)、昆明市科技计划(2015-1-S-00284)、昆明理工大学分析测试基金(2016T20130030)、国家留学基金委创新型人才国际合作培养项目(201608740005)资助

剪、破坏钳等大型剪切工具剪切线缆,形成的断头表面线条痕迹是案发现场中最为常出现的一种痕迹,犯罪嫌疑人使用工具加载于承痕体,在接触部位形成的局部物质形态变化。工具痕迹反映的是工具接触部位的外部形态结构,据此通过推断分析作案工具,起到提供侦查线索、缩小侦查范围的作用。其具有不易破坏、难以伪装、出现率频繁,鉴定价值高等特点,对于办案人员认定案件性质,确定作案工具,进而证实犯罪嫌疑人有着十分重要的意义^[2]。

相较于传统通过显微镜观察、人为比对的方式,近年来兴起的图像识别以及3D扫描技术,为非线性剪切痕迹的定量检验提供了一些新的解决方案:美国纽约城市大学的 Gambino 等人^[3]利用白光共聚焦显微镜测取线性痕迹表面波纹数据,随后使用主分量分析方法对采集数据进行降维,最后使用支持向量机和形预测算法进行相似度比对,利用多组擦划线条痕迹样本验证了该方法的有效性。荷兰尼德兰法医研究所的 Baiker 等人^[4]构建了一种针工具线条痕迹的多尺度配准框架,通过实际线条痕迹比对实验验证了此方法的有效性。德国弗劳恩霍夫信息与数据处理研究对工具擦划痕迹进行了研究,提取痕迹中沟槽特征,对沟槽进行定位,采用基于模板的方法,通过相关分析计算两个沟槽的关联程度^[5]。广东警官学院的杨敏等人^[6]对工具痕迹纹理信号进行离散小波变换,构造局部小波能量特征矢量,通过衡量特征标准差实现痕迹相似度检验。深圳大学的王炳成等人^[7]利用分形维数量化线条痕迹表面形态特征,建立分形维数群识别规则,通过断线钳剪切刃口痕迹自动区分实验验证了该方法的有效性。然而,受限于现场痕迹随机性强、检测手段繁琐、算法结构复杂不利于工程化、图片及3D文件体积过大造成计算难度增加等问题,使得上述方法并不适合在办案一线推广,大大降低了其实用价值。

单点激光位移测试具有不损坏被测物体表面、不受光线环境影响、精度高、数据文件体积小、频响特性好等特点,对于线条痕迹细节特征的非接触式精准测量是非常有效的。昆明理工大学的潘楠等人^[8]于2015年6月提出了一种线条痕迹激光检测信号自适应匹配算法,该方法对痕迹检测信号数据中“粗大连贯”趋势特征的识别,通过余弦向量曲线拟合的方法量化特征向量,逐个计算样本空间距离。最后利用动态规划算法批量进行相似度比对,匹配各自最大相似样本。然而,在将该方法应用到实际案发现场剪切痕迹的检验分析时,遇到如下问题。

1) 信号降噪:受限于实际的测试环境和线缆断头本身的差异性,激光传感器拾取的位移信号中往往夹杂有很多背景噪声,此类噪声来自于检测设备自身的机械误差以及外部干扰。为此,需要在保证信号准确的基础上,最大程度的降低噪声干扰,提高后续工作的质量与准

确性。

2) 信号匹配:实际剪切痕迹的形态千差万别,甚至同一剪切工具经过多次不同方向和位移偏差的剪切而形成的断头痕迹,彼此之间也会产生较大差异。匹配信号和待匹配信号的长度并不相同。同时,两个痕迹检测信号很可能只是在某一部分有重合。现实情况中,往往同一个剪切工具形成痕迹的检测信号在时域无法直接获取其相似特征,需要将数据转换到变换域进行处理。

3) 工具推断:由于犯罪嫌疑人经常连续作案,在案发现场可能遗留成百上千个剪切断头,且剪切工具的剪切面形态基本无法完整反映在线缆的断头上,承痕体上只保留部分剪切面(残缺痕迹)。如进行逐数据点的比对,将大大增加算法计算量,降低比对效率。为此,需要通过动态时间规整算法等手段实现大量痕迹激光检测数据特征的快速比对,以进一步实现痕迹所属剪切工具的溯源推断。

综上,针对实际案发现场剪切工具高效快速判断的要求,设计出一种非线性线条痕迹小波域特征快速溯源算法,该算法首先将激光位移传感器检测断头表面拾取的特征信号利用小波分解进行降噪,随后使用基于小波特征的动态时间规整算法实现痕迹特征相似重合度匹配,最终实现对应工具快速推断。通过实际痕迹推断剪切工具试验验证了本算法的实用性和有效性。

2 检测信号降噪

2.1 小波降噪概述

首先,利用文献[2]所述单点激光检测仪器对断头表面进行非接触数据采集,随后对激光检测信号中含有的大量背景噪声干扰进行去噪,以期获取一个相对稳定的粗大连贯波峰趋势。选用小波降噪,其基本步骤是将信号进行小波变换,把特性分配到各个尺度的小波变换系数上,再根据小波变换系数的分析与处理,选择相关的阈值,以期将背景噪声抑制到最低程度^[9-10]。

2.2 基于小波分解的剪切痕迹检测信号降噪

在降噪之前,首先需要选定小波的类型。实际降噪过程中,需要对原始信号 $f(t)$ 进行多层的分解,层数分解的越多,对细节数据处理的越多,能够消除的噪声更多,但是可能抹平的细节也更多。因此寻找一个平衡的分解层数就显得十分的必要^[11-12]。

在分解过程中,一般采用将原始信号分解为两部分,假设进行 n 层的分解,那么原始信号的构成可如下描述:

$$f = a_n + \sum_i^n d_i \quad (1)$$

其中分层的信息如图1所示,图1中的各节点同原公式的对应如下:

1) a_i 为第 n 层的近似,在图中的表示为(1,0);

- 2) d_i 为第 i 层的细节数据,在图中的表示为 $(i,1)$;
- 3) f 为原始数据,在图中的表示为 $(0,0)$ 。

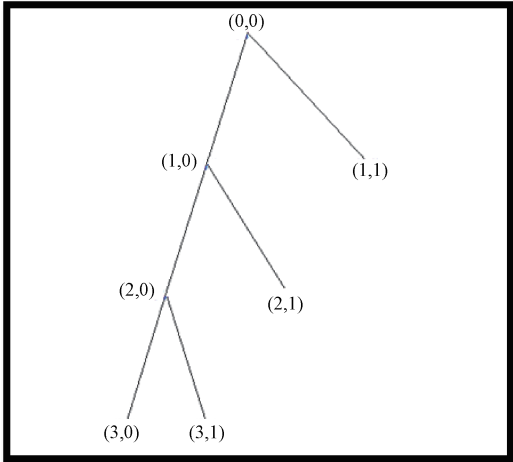


图1 分解信息

Fig.1 Decomposition message

1~3 层的近似数据如图 2 所示,3 层的细节展开如图 3 所示。

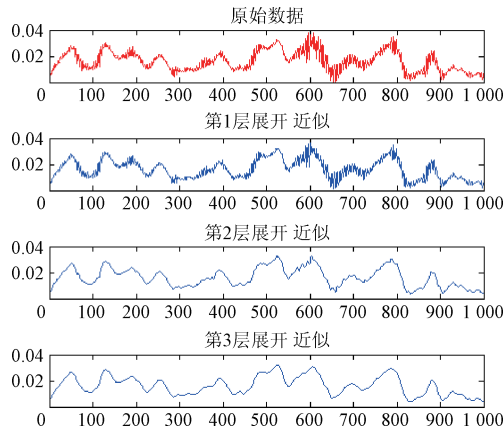


图2 各层级近似信号

Fig.2 Approximate signal at each level

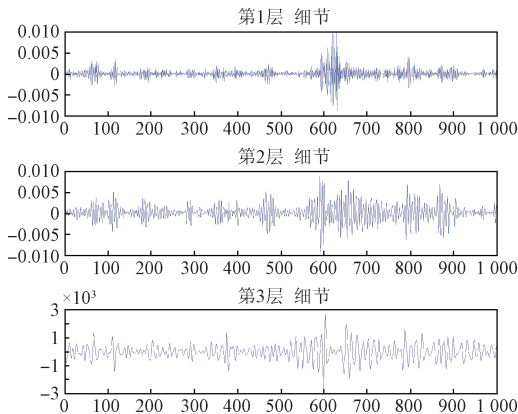


图3 3层细节展开

Fig.3 The three layer details

根据图 3 对示例信号分解得出的近似数据和细节数据可以发现,对于当前的痕迹检测来说,层级 1 的展开得到的近似数据仍然有较多的毛刺,层级 2 的近似信号展开相比于层级 1 的展开来说已经得到了极大的改善,绝大部分的毛刺得到了改善,而层级 3 的近似信号则进一步的改善了此现象,从得到的图形反馈来看,已经十分接近于所期望得到的样式了。因此,一般情况下只需进行到层级 3 的展开。

然而,根据公式:

$$f = a_n + \sum_i^n d_i \tag{2}$$

除了近似信号以外,还仍然要有细节数据,虽然根据肉眼观察在层级三的时候近似数据也十分接近波形样式了,但是仍然是缺乏细节的信号,因为这是丢失了很多细小噪声,只能表示大致的趋势,不能准确表示痕迹。

那么根据降噪的公式:

$$c_i = \begin{cases} 0, & |c_i| < r \\ c_i, & |c_i| \geq r \end{cases} \tag{3}$$

其中 c_i 表示第 i 个分解小波系数,为了去除细节里的噪声,需要设定一个阈值,将阈值之下的数据全部清除。因为一般的噪声都是以高频无规律的噪声存在,其转化为细节部分的值都是较小的。因此对每个细节设定一个阈值,进行过滤降噪。

$$f' = a_n + \sum_i^n d'_i \tag{4}$$

式中: d'_i 为经过阈值降噪后的细节数据, f' 为降噪过后的痕迹数据。

考虑到剪切线条痕迹的特性,通过多次实验,最后选用 db4 或 haar 作为母小波,并根据数据检测精度进行 2~5 层的分解重构,分解后进行相应的噪声去除重构得到的波形信号就是降噪过后的数据。数据降噪部分的流程如图 4 所示。

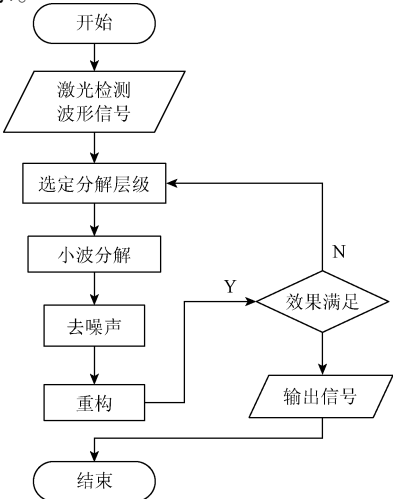


图4 降噪流程

Fig.4 Flowchart of signal denoising

如图5所示,由于机械振动影响,检测信号受到高频噪声的干扰。需进行降噪以凸显实际痕迹特征信号。由图5可以观察到,降噪后基本能够过略掉全部高频噪声,而正常数据部分则不受干扰。

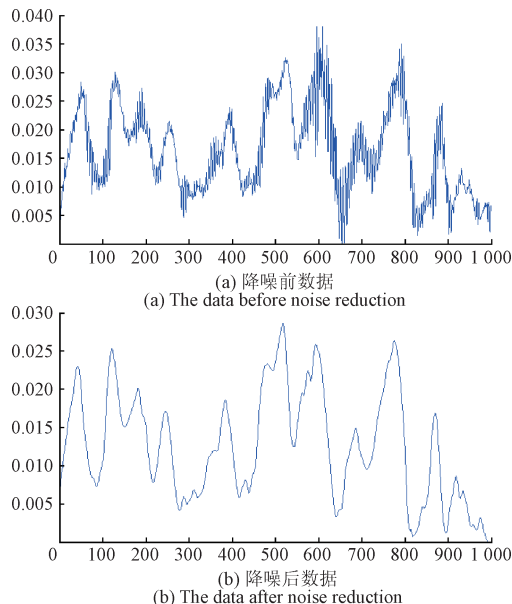


图5 实际检测信号降噪

Fig. 5 The actual detection signal denoising

3 检测信号特征匹配

3.1 长度不定及部分重合问题

在针对降噪后的痕迹的信号进行相似度匹配之前,首先需要处理如下两个^[13]问题。

1) 痕迹长度不定,每次采集得到的痕迹检测信号都长度不经相同,匹配信号数据和待匹配信号的长度绝大部分情况都是不同的。此时,直接使用欧几里得距离、相关系数等方法进行两个离散序列的相似性的度量并不可行,点对点运算变得无意义。

2) 部分重合,即两个痕迹检测信号可能只是在某一部分有重合,这种情况对最后的重合度计算造成很大干扰。

为此,在算法进行匹配的运算过程中,将会针对长度不定和部分重叠的问题进行优化,最终实现针对痕迹信号的匹配的改善,其大致的基本步骤如下。

1) 设输入的数据分别为 A 和 B ,二者均是已经符合上述要求的数据。

3) 设定一个匹配的最小长度 L ,即二者如果重合必须要满足的最小重合长度,从 A 中选择最长长度到最短长度的部分到 B 中进行比对,相当于选择不同的位置进行多次匹配。

3) 迭代执行每个位置的对比,每次比对都比对二者对应位置差异度(对应位置的)的方差大小,如果最小就记录当前状态。

4) 完成了步骤3)的功能后,交换 A 、 B 的角色,继续完成一次步骤2)、3)。

5) 计算方差最小的差异度,输出匹配结果。

信号匹配的处理流程如图6所示。

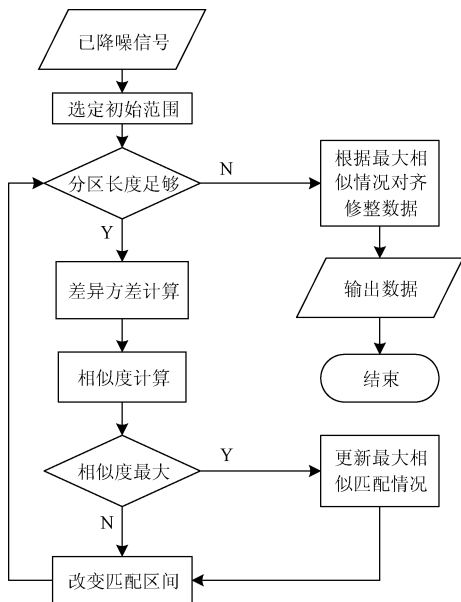


图6 匹配流程

Fig. 6 Flowchart of matching

3.2 动态时间规整算法概述

动态时间规整算法(dynamic time warping, DTW),源于语音信号识别,其基于动态规划的思想,重点解决信号序列长度不一致时的匹配问题^[14]。拓展到所有离散信号序列来说,DTW 是一种衡量两个离散信号序列相似度的方法,主要特点是在序列长度不一或 X 轴无法完全对齐的情况下,用满足一定条件的时间规整函数描述两者之间的时间对应关系^[15]。广泛应用于诸如如语音识别、动态手势识别和信息检索等各种模式匹配任务中。

DTW 的主要思路如下:

假定两个待匹配的离散数据分别为:

$$A = \{A(1), A(2), \dots, A(m)\}$$

$$B = \{B(1), B(2), \dots, B(m)\} \quad (5)$$

其中下标为 1 的元素为序列的起点,下标为 m/n 的元素为序列终点。为了对齐 A 、 B 两个序列,使用动态规划算法,首先构造一个 $m \times n$ 的矩阵,用于存放两序列点对点之间的距离(如欧氏距离),距离越小则表明两点之间的相似度越高。

此为 DTW 算法的核心,把矩阵看成一个网格,算法的目的可总结为寻找一条通过此矩阵网格的最优路径,

该路径通过的网格点即为两个离散序列经过对齐后的点对。

找到最优路径后,DTW 算法定义了一个归整路径距离(warp path distance, WPD),使用所有相似点之间距离的和来衡量两个时间序列之间的相似性。其路径计算方式如下式:

$$w_{k-1} = (i, j), w_k = (i', j') \quad (6)$$

式中: $i \leq i' \leq i+1, j \leq j' < j+1$ 。

其中路径还需要满足如下条件。

1) 边界条件: $w_1 = (1, 1)$ 和 $w_k = (m, n)$ 。

2) 连续性: 若 $w_{k-1} = (i, j)$, 则路径的下一个点 $w_k = (i', j')$ 需要满足 $i' - i \leq 1$ 和 $j' - j \leq 1$ 。即只能和自己相邻的点对齐。这样可以保证序列 A 和 B 中的每个元素都在规整路径 w 中出现。

3) 单调性: 若 $w_{k-1} = (i, j)$, 则路径的下一个点 $w_k = (i', j')$ 需要满足 $i' - i \geq 0$ 和 $j' - j \geq 0$ 。此处需要假设 A 和 B 的顺序均不可改变。路径 W 在矩阵网格中的走势必须随时间单调递增。

最后,需要定义一个累加距离 $dist$, 即从 $(0, 0)$ 点开始匹配两个序列 A 和 B , 每到一个点,之前所有的点计算的距离都会累加。到达终点 (n, m) 后,这个累积距离就描述了序列 A 和 B 的总体相似程度。累积距离 $dist(i, j)$ 可以表示成以下公式:

$$dist(i, j) = \min \begin{cases} dist(i-1, j-1) + d(A(i), B(j)) \\ dist(i-1, j) + d(A(i), B(j)) \\ dist(i, j-1) + d(A(i), B(j)) \end{cases} \quad (7)$$

式中: 累积距离 $dist(i, j)$ 为当前各点间距离 $d(A(i), B(j))$, 即序列 A 、 B 中对应两点 $A(i)$ 和 $B(j)$ 的欧氏距离与到达该点的最小邻近元素累积距离之和。

3.3 基于小波特征的动态时间规整

由于降噪后的剪切痕迹检测信号在时域表现近似于随机信号,很难利用 DTW 直接进行处理,经过多次实验发现,通过对信号进行小波变换,使得当前在时域范围内较难处理的相似度判定问题够得到一定的改善。

具体处理步骤如下:

1) 将经过降噪的痕迹检测数据进行小波变换,将原有的信号数据变换得到各时间上不同尺度小波的二维矩阵分量,小波变换结果如图 7 所示。

由图 7 可发现,原有信号数据变成三维尺度-时间-分量中的一个结构,通过图中的颜色深浅具体表述为在对应的时间 b 上,对应的小波尺度 a 上的分量大小。

通过提取小波变换后某一尺度,或者某些尺度上的

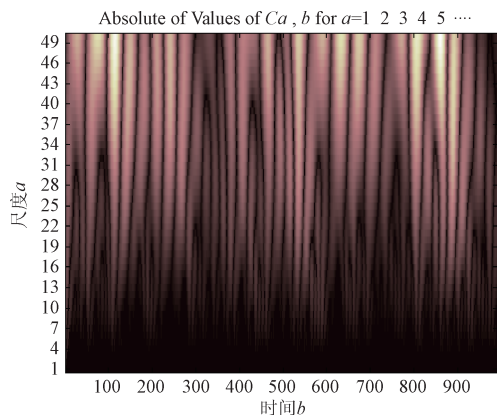


图 7 小波变换结果

Fig. 7 The result of wavelet transform

系数作为对应痕迹信号的特征。选择不同尺度上的对应系数序列,可以观察到其在对应时间上的系数大小,图 8 所示分别表示相同的痕迹检测信号经小波变换后选择不同的尺度而得到的系数。

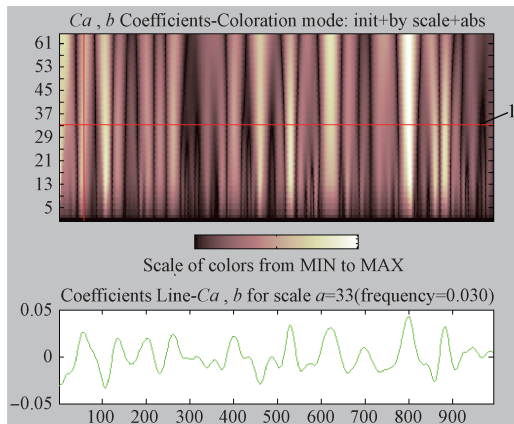


图 8 尺度 33 下系数

Fig. 8 The coefficient under scale 33

图 8 中,上方是经过小波变换的结果,线 1 是所选择的尺度,下方的波形图是该尺度下对应的大小图。可以看出不同尺度在构成原痕迹信息上起到了不同的作用。

针对不同类型的痕迹可以选择不同尺度的系数作为特征,系数与分辨率成反比,系数越大分辨率越小,通常更能表示整体的趋势,而系数越小则分辨率越高、噪声干扰更强,反应局部细节而非整体趋势。针对系数较低的区间的数据的去除,应选择尺度适中的系数作为特征,其原因在于如下的 3 点。

(1) 系数较低的区间蕴藏的数据是较为细节的部分,由于不同痕迹之间本身是存在差异的,而较小尺度的分量通常就是这些个体差异的数据还有机械噪声,所以并不适用。

(2) 尺度较高的区间的数据, 则更加适合描述痕迹表面的高低起伏、剪切的平整程度等的内 容, 缺失了工具的特征。

(3) 算法目的在于剪切工具溯源, 即痕迹间共性最大化, 且不丧失工具痕迹间差异性, 因此中间系数较为合适。根据检测装置和实际实验测试, 选择尺度 35 附近的数据最为合适。

2) 完成了系数的选择之后, 随采用 DTW 的算法将样本特征与待测试痕迹特征进行相似度比对。

设样本为 $S = \{s_1, s_2, \dots, s_n\}$, 设输入的痕迹为 $T = \{t_1, t_2, \dots, t_n\}$, 如此可构建一个 $N \times M$ 大小的二维矩阵 A , 二维矩阵中的元素 $a_{(i,j)}$ 表示样本 $S = \{s_1, s_2, \dots, s_n\}$ 当中第 i 个元素和 $I = \{i_1, i_2, \dots, i_m\}$ 中地 j 个元素的距离。

其中, 矩阵中 $a_{(i,j)}$ 的计算方式如下:

$$a_{(i,j)} = e^{-(s_i - t_j)^2} \quad (8)$$

式(8)为非线性方程, 当二者的匹配距离较为相近时, 其计算相似度能保持在较高水平; 当距离超过一定水平, 趋近于无限远, 都能将相似度控制在较低的范围, 不会出现距离相似度计算不准确的情况。

如图 9 所示, 当 s_i 和 t_j 距离最小的时候, 其相似程度最高为 1, 随着距离越远, 其相似程度越低, 该取值可以无限的趋近于 0。

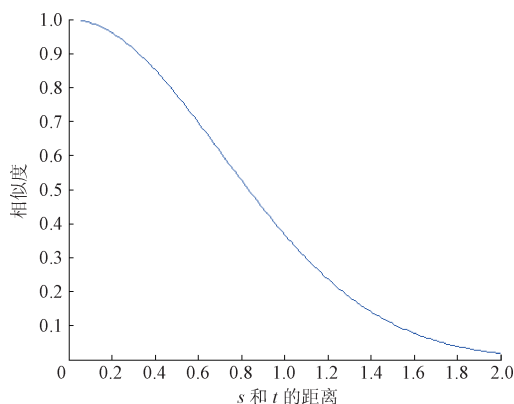


图 9 相似度曲线

Fig. 9 The similarity curve

得到了各点之间的相似程度后, 便采用 DTW 的方式, 寻找最佳匹配的路径, 也就是最终在目的地 (n, m) 处得到最大的相似度值 $sims(n, m)$ 。

图 10 所示为使用两个痕迹信号的小波特征相似矩阵进行 DTW 计算后, 找到最优匹配的示意图。在示意图中, 如果是垂直线条或水平线条则代表了多次匹配, 而如果是斜线则代表是全新匹配。

其中相似度的动态规划方程如下:

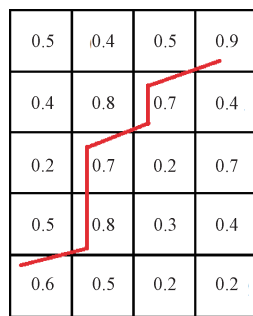


图 10 DTW 匹配示意图

Fig. 10 Schematic diagram of DTW matching

$$sims(i, j) = \max \begin{cases} sims(i-1, j) + a(i, j) \\ sims(i, j-1) + a(i, j) \\ sims(i-1, j-1) + 2 \times a(i, j) \end{cases} \quad (8)$$

其核心思想是动态局部最优(贪心法), 给予全新匹配更高的权重。在 DTW 中, 无论是样本数据中的某一个位置, 还是输入痕迹检测数据中的某一个位置, 只要是和对方的多个位置出现匹配, 那么给予 1/2 与两个都是第一次发生匹配的权重, 如此保证在两个信号长度不一致的情况下, 做到最优匹配。

如图 11 所示, Data 17(图 11(a)) 和 Data 18(图 11(b)) 为同一痕迹的重复检测信号, 经过匹配, 得到二者的最小差异情况下的差异显示在图 11(c)、(d) 叠加显示二者重合后的信号。通过肉眼观测, 可发信其重合程度较高。由于在检测过程当中, 无可避免的会产生一些误差, 二者不可能完全相同。根据测试结果, 其相似度为 90%, 二者得到了有效的对齐匹配。

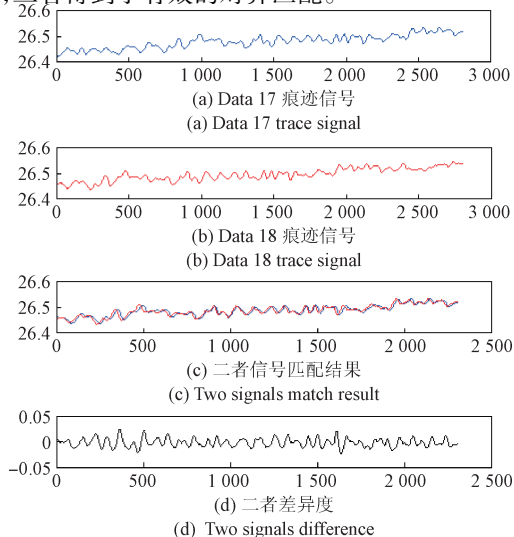


图 11 信号差异匹配

Fig. 11 The signal difference matching

特征的权重设定也会影响识别的准确率,因此使用特定的训练集去做参数训练,降低失败的概率。参数的训练有如下基本步骤。

1) 保证完成样本库建立步骤,样本之间都具有足够的代表性和区分度。

2) 使用梯度下降法进行参数的机器学习,构建相应的代价函数,通过不断的迭代使得代价函数值代价最小^[16-17]。

3) 获得学习的参数后就可以结束参数训练。

4) 当样本库改变或使用场景改变时,建议使用一个更加具有针对性的训练集进行参数的训练。

根据输入的样本痕迹检测信号变化梯度进行相似度的计算,计算得到的相似度数据后进行相应的分组识别并溯源推断可能归属的工具。相似度计算需要计算样本与输入样例的相似程度大小,相似程度的大小可以映射到0~1的区间范围之内,最小值0代表完全不同,最大值1代表完全相同。

3.4 样本库建立

作为鉴别痕迹信号所属工具的根本依据,样本库的质量与建立方式直接影响最终结果。其建立通常遵循如下步骤。

1) 确定需要鉴别的工具种类、范围、类型,用统一规则每个工具进行编号,详细记录其参数信息。

2) 为了排除检测时的偶然性,单个断头痕迹需要至少检测两次,当两次的信号数据重合度能够达到99%以上时,才能认定为合格数据。同一个断头建议记录多个位置的不同数据,每个位置都采用如上的方式。必要的时候可以加入人的手工修正,以及一切有利于提高痕迹样本质量和代表性的手段。

3) 完成所有数据的采集后,同时采用此样本库的数据作为测试数据,针对每个特征进行测试,测试结果应当是每个样本数据和自己组别数据的数据相似程度明显高于其他组别,如果是混杂在一起的,则数据无效,重新进行步骤1)~2)。

3.5 分组识别

完成了相似计算工作之后,将会得到待溯源推断痕迹数据同样本库中所有样本的比对结果,这些结果是一系列的样本同输入痕迹的相似程度的大小,不同种类工具、不同位置的剪切都会影响结果,而分组识别则是将此类数据进行分组聚合,找到最有可能的工具种类。

分组识别通常有如下的基本步骤。

1) 对输入痕迹数据和样本库中样本进行比对过的得到的相似度进行排序,通常排序后可以仅适用前10%~20%的数据进行比对。

2) 对前10%~20%的数据进行加权处理,加权规则一般是排序越靠前的,其权重越大,反之则是权重低。

3) 根据加权后的成绩进行分组归类,计算出每一个组别在这前10%~20%数据当中所占比重的大小。

4) 根据每个分组所占的比例大小,进行排序,选择排序较为靠前的作为推倒的结果。

整个工具溯源流程如图12所示。

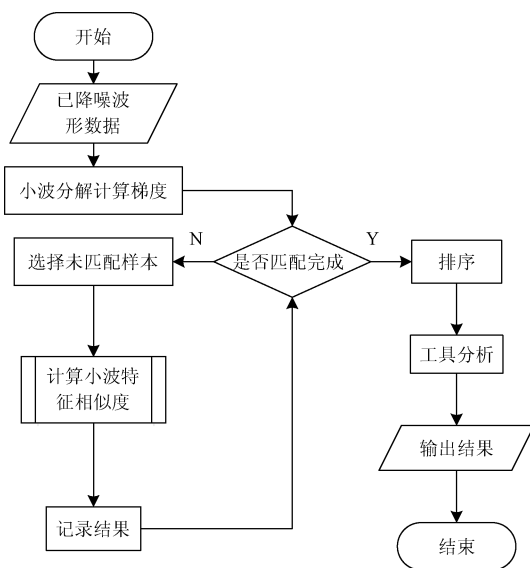


图12 工具溯源流程

Fig. 12 Flowchart of tools tracing

4 实验验证

通过实际剪切工具溯源实验验证本文算法效果,搭建实验环境如下。

一共选取3种工具:断线钳(A)、破坏钳(B)、钢丝钳(C)分别剪切1 cm直径铜制棒材形成断头,所有断头表面均使用文献[2]所述仪器进行检测。采样参数设置:激光斑点直径2.5 μm,细分数3 200 step/s,同频率脉冲1 000,采样间隔50 ms,采样频率20 Hz,采样点数根据断头横截面积大小确定。匹配程序相关算法利用MATLAB 2015b验证后,采用Python进行编写。运行在CPU为Intel Core i5 2.9 GHz,内存8 G DDR3的ThinkPad笔记本电脑上。

使用数据标号T16~T45(序号16~45顺序编号)共计29组数据(T26、T27为基本相同数据)的数据作为测试数据进行测试,样本库中存有10种不同工具剪切形成的500组数据,测试时除了自身以外的数据作为样本库,并显示出排序值前5的结果值。取匹配率最高的样本编号为最高匹配样本。

T16~T45的数据当中,共有3组数据,T16~T25来自工具A形成的痕迹,T26~T38来自工具B1和B2形成的痕迹,T39~T45来自工具C形成的痕迹,B1、B2为同属工具B的两把工具。

为了更为真实的模拟案发现场数据采集,每组测试数据在基准痕迹数据的位置基础上进行位移后再次检测而形成的新数据,其中 A 中的数据主要是横向位移,也就是在原有痕迹所在的直线上移动,移动后有部分数据与原数据重合,同时 A、B、C 都有 U 方向(向上)和 D 方向(向下)的移动,和原来的基准痕迹形成一定程度的错位。实际测试场景如图 13 所示。

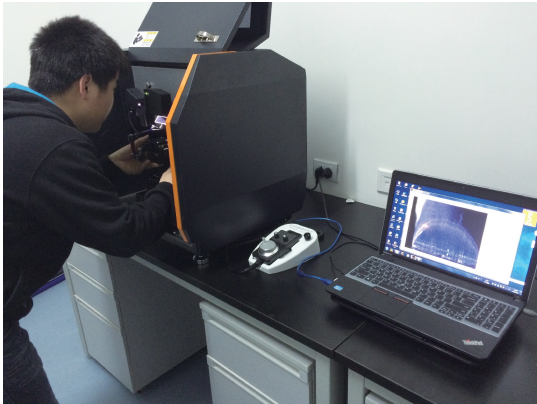


图 13 实际测试
Fig. 13 The actual test

定义工具溯源情况如下。
准确溯源:对于是相同的痕迹(或工具),但是检测的位置、数据长短等都存在一定的偏移其排序较为靠前,且能和其他类型的工具相似度有区分度。溯源失败:与原痕迹并无关联的数据。本文算法样本匹配结果如表 1 所示。

表 1 本文提出算法样本匹配结果
Table 1 The sample matching results by the algorithm proposed in this paper

工具 序号	最高 匹配	准确溯 源率/%	数据说明
T16	T17	100	组别 A 基准数据
T17	T16	100	组别 A 基准数据 重复检测
T18	T20	100	组别 A 基准数据 位移 1/6
T19	T20	100	组别 A 基准数据 位移 1/3
T20	T21	100	组别 A 基准数据 位移 1/3
T21	T20	100	组别 A 基准数据 位移 1/2
T22	T23	100	同 A 断头 另一条线
T23	T18	100	T22 基础上 D 方向平移 1/6 的另一条线
T24	T39	40	T22 基础上 U 方向平移 1/6 的另一条线
T25	T41	40	T22 基础上错位 1/3 并缺失 2/3
T26	T27	100	组别 B 基准数据
T27	T26	100	组别 B 基准数据 重复检测
T28	T27	100	组别 B 基准数据 重复检测
T29	T30	100	组别 B 基础上 D 方向错位 1/10 另一条线
T30	T29	100	T29 的重复检测
T31	T32	100	组别 B 基础上 U 方向错位 1/10 另一条线
T32	T31	100	T31 的重复检测
T33	T34	100	组别 B 同类工具,不同断头

工具 序号	最高 匹配	准确溯 源率/%	数据说明
T34	T33	100	T33 的重复检测
T35	T36	100	T33 基础上 D 方向错位 1/10 另一条线
T36	T35	100	T35 重复检测
T37	T38	80	T33 基础上 D 方向错位 1/10 另一条线
T38	T37	80	T38 重复检测
T39	T40	80	组别 C 基准数据
T40	T39	80	组别 C 基准数据 重复检测
T41	T25	60	组别 C 基础上 D 方向位移 1/10 另一条线
T42	T43	80	T41 重复检测
T43	T42	80	T41 重复检测
T44	T45	80	组别 C 基础上 U 方向位移 1/10 另一条线
T45	T44	80	T44 重复检测

计算耗时 15 s,其中匹配成功率为 88.9%,失败率为 11.1%。前期大样本比对实验表明,待检测断头摆放角度、激光信号检测角度、剪切工具类型、剪切角度、断头材质等因素对最终匹配结果的影响不大,误差率约在2%~3%。

为了与本文提出技术进行比对,利用文献[8]算法同样使用数据标号 T16~T45 29 组检测数据进行总计 145 次的测试,其结果为如表 2 所示。

表 2 文献[8]提出算法匹配结果
Table 2 The sample matching results by the algorithm proposed in the reference [8]

工具 序号	最高 匹配	准确溯 源率/%	数据说明
T16	T18	80	组别 A 基准数据
T17	T16	60	组别 A 基准数据 重复检测
T18	else	20	组别 A 基准数据 位移 1/6
T19	else	20	组别 A 基准数据 位移 1/3
T20	else	20	组别 A 基准数据 位移 1/3
T21	T16	60	组别 A 基准数据 位移 1/2
T22	T23	60	同 A 断头 另一条线
T23	T22	80	T22 基础上 D 方向平移 1/6 的另一条线
T24	其他	0	T22 基础上 U 方向平移 1/6 的另一条线
T25	其他	0	T22 基础上错位 1/3 并缺失 2/3
T26	T27	80	组别 B 基准数据
T27	T26	80	组别 B 基准数据 重复检测
T28	T27	80	组别 B 基准数据 重复检测
T29	T26	60	组别 B 基础上 D 方向错位 1/10 另一条线
T30	T26	60	T29 的重复检测
T31	else	0	组别 B 基础上 U 方向错位 1/10 另一条线
T32	T31	100	T31 的重复检测
T33	T34	60	组别 B 同类工具,不同断头
T34	T33	60	T33 的重复检测
T35	T37	20	T33 基础上 D 方向错位 1/10 另一条线
T36	其他	0	T35 重复检测
T37	其他	0	T33 基础上 D 方向错位 1/10 另一条线
T38	T37	60	T38 重复检测
T39	T40	80	组别 C 基准数据
T40	T39	80	组别 C 基准数据 重复检测
T41	T25	20	组别 C 基础上 D 方向位移 1/10 另一条线
T42	T43	60	T41 重复检测
T43	T42	60	T41 重复检测
T44	T45	60	组别 C 基础上 U 方向位移 1/10 另一条线
T45	T44	60	T44 重复检测

计算耗时 18 s, 其中匹配成功率为 51%, 失败率为 49%。由此可见, 本文提出溯源技术对比文献[8]提出的方法, 虽在运算速度上差别不大, 但在运算精度、稳定性数量级上具有明显优势, 更适用于实际案发现场痕迹检测数据。

5 结 论

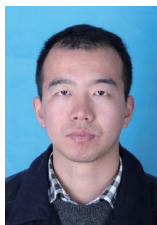
本文提出一种针对非线性剪切痕迹的小波域特征快速溯源算法, 该算法结合基于小波特征的动态时间规整算法实现痕迹特征相似重合度匹配, 从而实现对应工具快速推断。本文算法结构复杂度较低, 可以直接通过 Python 语言编程实现, 生成可执行文件运行在中等配置计算机上。通过某公安部痕迹鉴定专家实际测试及光学比对, 证明其能够为有效的应用到基于线条痕迹的剪切工具溯源工作中, 为后续鉴定工作提供可靠参考。为了进一步贴近实际办案环境, 如何提高更大数量级数据的运算速度, 以较快速完成跨区域的串并案比对溯源, 将是下一步研究的方向。

参考文献

- [1] BUNCH S, WEVERS G. Application of likelihood ratios for firearm and toolmark analysis [J]. Science and Justice, 2013, 53 (2) : 223-229.
- [2] 刘益, 潘楠, 沙生, 等. 一种警用线性痕迹激光检测系统[P]. 中国专利: CN 105300996 A, 2016.
LIU Y, PAN N, SHA SH, et al. One police laser detection system for linear traces [P]. China Patent: CN 105300996 A, 2016.
- [3] GAMBINO C, MCLAUGHLIN P, KUO L, et al. Forensic surface metrology: Tool mark evidence [J]. Scanning, 2011, 33 (5) : 272-278.
- [4] BAIKER M, KEEREWEER I, R PIETERMAN R, et al. Quantitative comparison of striated toolmarks [J]. Forensic Science International, 2014, 242: 186-199.
- [5] KASSAMAKOV I, BARBEAU C, LEHTO S, et al. CSI Helsinki: Comparing three-dimensional imaging of diagonal cutter toolmarks using confocal microscopy and SWLI [C]. Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV, 2010: 76900Y.
- [6] 杨敏, 李冬云, 王纬东. 基于局部小波能量工具痕迹识别的研究[J]. 中国人民公安大学学报: 自然科学版, 2008, 56(2) : 73-75.
YANG M, LI D Y, WANG W D. The study of tool mark identification based on local wavelet energy [J]. Journal of People's Public Security University of China: Science and Technology, 2008, 56(2) : 73-75.
- [7] 王炳成, 任朝晖, 任立义, 等. 剪切痕迹表面形貌的分形描述[J]. 机械设计与制造, 2003, 160(6) : 106-107.
WANG B CH, REN CH H, REN L Y, et al. Fractal geometry description of the surface appearance of the shearing trace [J]. Machinery Design & Manufacture, 2003, 160(6) : 106-107.
- [8] 潘楠, 伍星, 刘益, 等. 线性痕迹激光检测信号自适应匹配算法研究[J]. 仪器仪表学报, 2015, 36(6) : 1372-1380.
PAN N, WU X, LIU Y, et al. Research on the adaptive matching algorithm for laser linear mark detection signals [J]. Chinese Journal of Scientific Instrument, 2015, 36(6) : 1372-1380.
- [9] 赵新中, 陶永耀, 贺佩, 等. 一种基于小波邻域的半软阈值去噪算法[J]. 国外电子测量技术, 2016, 35(4) : 42-45.
ZHAO X ZH, TAO Y Y, HE P, et al. Denoising algorithm of semi soft threshold based on wavelet neighborhood [J]. Foreign Electronic Measurement Technology, 2016, 35(4) : 42-45.
- [10] 张晖, 鲁明明, 丁华平, 等. 基于经验模态分解法的桥梁应变信号降噪方法研究[J]. 电子测量技术, 2016, 39(10) : 166-170.
ZHANG H, LU M M, DING H P, et al. Research of denoising method of bridge strain signal based on empirical mode decomposition method [J]. Electronic Measurement Technology, 2016, 39(10) : 166-170.
- [11] 惠巧娟, 李洪安, 陆焱. 一种基于小波变换和人类视觉系统的图像压缩算法[J]. 电子测量与仪器学报, 2016, 30(12) : 1838-1844.
HUI Q J, LI H AN, LU Y. Image compression algorithm based on wavelet transform and HVS [J]. Journal of electronic Measurement and Instrumentation, 2016, 30(12) : 1838-1844.
- [12] OGAWA K, SAKATA M, LI Y. Adaptive noise reduction of scintigrams with a wavelet transform [J]. International Journal of Biomedical Imaging, 2012, 2012(11) : 130482.
- [13] ZHENG Y, XIAO L M, TANG W Q, et al. Parallelization of dynamic time warping on a heterogeneous platform [J]. IEICE Transactions on Fundamentals of Electronics Communications & Computer Sciences, 2014, E97A(11) : 2258-2262.
- [14] LI H L. On-line and dynamic time warping for time series data mining [J]. International Journal of Machine

- Learning and Cybernetics, 2014, 6(1): 145-153.
- [15] DUPAS R, TAVENARD R, OPHÉLIE F. Identifying seasonal patterns of phosphorus storm dynamics with dynamic time warping [J]. Water Resources Research, 2015, 51(11): 8868-8882.
- [16] JIA N N, LAM E Y. Machine learning for inverse lithography: Using stochastic gradient descent for robust photomask synthesis [J]. Journal of Optics, 2010, 12(4): 45601-45609.
- [17] SHAO J, YANG J Y. Multi-robot reinforcement learning based on learning classifier system with gradient descent methods [J]. Journal of Computational Information Systems, 2010, 8(6): 2449-2455.

作者简介



潘楠(通讯作者), 1986年出生, 2008年于安徽理工大学获得学士学位, 2012年于昆明理工大学获得博士学位(硕博连读), 现为昆明理工大学航空学院讲师、硕导, 主要研究方向为信号处理、刑侦特种设备研发。

E-mail: nanpan@kmust.edu.cn

Pan Nan (Corresponding author) was born in 1986, received B. Sc. from Anhui University of Science and Technology in 2008 and Ph. D. from Kunming University of Science and Technology in 2012, respectively. Now, he is a lecturer and

M. Sc. supervisor in Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His main research directions include signal processing and criminal investigation special equipment R&D.



阚立峰, 1991年出生, 硕士研究生, 2015年于青岛理工大学获得学士学位, 现于昆明理工大学攻读硕士学位, 主要研究方向为信号处理、刑侦特种设备研发。

E-mail: kanlifeng@ldacn.com

Kan Lifeng was born in 1991, received B. Sc. from Qingdao Technological University in 2015. Now, he is a M. Sc. candidate in Kunming University. His main research direction includes signal processing, criminal investigation special equipment R&D.



潘地林, 1955年出生, 分别在1984年, 1986年于安徽理工大学获得学士学位和硕士学位, 现任昆明信诺莱伯科技有限公司总工程师, 主要研究方向为刑侦特种设备研发。

E-mail: dilinpan@camsonar.com

Pan Dilin was born in 1955, received B. Sc., M. Sc. from Anhui University of Science and Technology in 1982 and 1986, respectively. Now, he is the chief engineer of Kunming SNLab Technology Co. Ltd. His main research directions include signal processing and criminal investigation special equipment R&D.