

基于梯度阈值的太赫兹时域信号自适应稀疏算法

秦丽丽^{1,2}, 李丽娟^{1,2,3*}, 任姣姣^{1,2,3}, 顾健^{1,2,3}, 熊伟华^{1,2}, 张丹丹^{1,2,3},
祝莉莉^{1,2}, 张霖昉^{1,2,3}, 薛竣文³, 江柏红⁴, 高增华⁴

(1. 长春理工大学光电工程学院 光电测控与光信息传输技术

教育部重点实验室, 吉林 长春 130022;

2. 长春理工大学光电工程学院 光电工程国家级实验教学示范中心, 吉林 长春 130022;

3. 长春理工大学中山研究院, 广东 中山 528400;

4. 航天特种材料及工艺技术研究所, 北京 100074)

摘要: 胶接结构广泛应用于航空航天等国防领域,但在工艺制作及使用过程可能会产生胶接界面脱粘缺陷和损伤,由于太赫兹无损检测技术对非金属材料良好的穿透性能,已被广泛应用于复合材料的无损检测中,太赫兹无损检测技术在多层胶接结构样件胶层内部缺陷的无损检测方面具有较大优势。利用反射式太赫兹时域光谱系统检测多层胶接结构样件,得到的具有样件内部材料信息的太赫兹时域信号,但信号中还包含了大量的冗余特征和噪声等无效信息,这些无效信息大大降低了信号处理和分析效率。针对这一问题,文中提出了基于二阶梯度法提取太赫兹时域信号有效特征,以飞行时间误差为限制条件基于信号的时域特征自适应确定阈值,稀疏太赫兹时域信号,减少信号中冗余无效信息,实现太赫兹时域信号的有效压缩。然后,通过二值化图像分割识别多高斯恢复信号和太赫兹时域光谱系统检测信号的太赫兹图像缺陷区域。最后,制备具有脱粘缺陷的多层胶接结构样件,开展太赫兹无损检测实验。结果表明:文中算法的数据压缩率达到了 81%,相比传统压缩算法离散余弦变换提高了 59%,相比主成分分析算法提高了 75%,相比 K-SVD 字典学习算法提高了 26%,缩短了约 80% 的数据计算时间,减小了约 95% 数据存储空间占用,且缺陷识别偏差不超过 0.05。文中算法极大地提高了数据处理和分析效率,保证了缺陷识别的精度。

关键词: 太赫兹时域光谱; 稀疏表示; 特征提取; 自适应阈值; 多层胶接结构

中图分类号: O433.4 **文献标志码:** A **DOI:** 10.3788/IRLA20220443

0 引言

胶接结构因其具有抗疲劳、耐高温、耐腐蚀、应力均匀分布等优点,广泛应用于航空、航天、兵器等国防领域^[1]。在胶接过程中由于胶接工艺过程粘接剂的控制、服役使用过程环境、载荷或冲击等因素,可能会导致胶接界面产生脱粘缺陷或损伤,这些制造缺陷和使用损伤会严重降低结构的承载能力,影响结构的使用安全性,甚至造成灾难性的事故发生。故需要有效的无损检测技术手段对胶接界面质量进行有效的缺陷检测与识别,以保证胶接质量和使用安全性。

太赫兹波段^[2]的辐射光子能量相对较低,对人体无害,且太赫兹波频率介于 0.1~10 THz 之间,波长介于 3 mm~30 μm ,能够透过高分子复合材料、塑料、泡沫等非金属和非极性材料,可以对这些材料样件内部缺陷与损伤清晰成像,实现无损检测^[2]。太赫兹无损检测技术凭借非金属材料良好的穿透性、人体安全性、复合材料检测有效性和操作简易性等优点在胶接结构胶层内部缺陷的无损检测领域具有很大优势。

太赫兹时域光谱技术^[2](Terahertz-Time Domain Spectroscopy, THz-TDS)在太赫兹检测领域是普遍使用的技术。THz-TDS 技术是太赫兹频率范围内对样

收稿日期:2022-06-27; 修订日期:2022-09-30

基金项目:吉林省科技发展计划项目(20220508032RC);中山市第二批社会公益与基础研究项目(2022B2012)

作者简介:秦丽丽,女,硕士生,主要从事太赫兹无损检测方面的研究。

导师(通讯作者)简介:李丽娟,女,教授,博士,主要从事复合材料太赫兹无损检测与评估技术、结构健康检测与评估方面的研究。

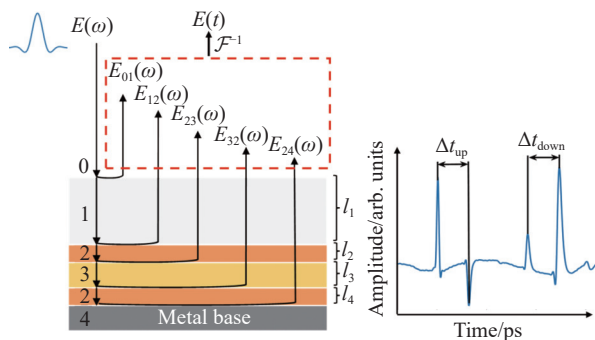


图 1 THz 信号在多层胶接结构样件传播过程

Fig.1 Propagation process of THz signal in multilayer bonded structure sample

品表征的一种强有力的技术,使用超短飞秒激光系统产生的亚皮秒电磁脉冲,能够覆盖 100 GHz~3 THz 的频率范围^[2]。太赫兹时域光谱系统主要由超快脉冲激光器、THz 发射与接收装置、时间延迟控制系统和光学系统组成。利用反射式 THz-TDS 系统探测多层胶接结构样件,采用逐点扫描的方式获取成像信息,每个点采集回来的信息都包含了样品在该点的幅值和相位信息,图像中每个像素点都一一对应 THz 时域信号,包含了被测样件的信号时域特征,同时 THz 信号携带大量无效冗余特征和噪声等无效信息,随着检测数据的逐渐增加,数据中冗余无效信息也在不断增加,数据处理的工作量也越来越大。大量的无效信息不仅消耗大量数据处理和分析时间,也为后续缺陷识别等信号分析工作带来极大的干扰。故通过压缩算法来消除 THz 信号中无意义的冗余数据是十分必要的。除了要实现信号的压缩,对 THz 信号有效特征信息的提取也是十分必要的,包括最大值、最小值、峰峰值、峰谷值以及信号幅值等时域特征,提取 THz 信号有效特征信息是太赫兹检测数据分析的关键。基于 THz 信号的时域特征成像,通过图像可视化可识别检测样件缺陷,提取有效特征,减少信号中大量无效冗余特征信息,对于提高缺陷识别的效率具有重要意义。

传统的数据压缩技术,例如离散余弦变换^[3](Discrete Cosine Transform, DCT)、离散小波变换^[4](Discrete Wavelet Transform, DWT)等信号表示方法,虽然达到了压缩的效果,但为了达到一定的表示精度,数据的压缩率并不高。Mallat^[5]基于小波分析提出用一个超完备字典表示信号,开启了稀疏表示的先河,并发现信号越稀疏则重建后的精度就越高。信号

稀疏表示^[6]是通过超完备字典稀疏表示复杂冗余信号,实现信号信息的便捷提取,有利于进一步的信号处理,如信号压缩、信号编码等。文献 [7] 提出 K-SVD 字典学习算法稀疏信号,该算法虽然实现较高的压缩率,但是仅适用于采样点较少的情况,当采样点超过一定值相对误差明显增大,并且字典更新的计算量较大。

扩散映射^[8](Diffusion Map, DM)、主成分分析^[9](Principal Component Analysis, PCA)、线性判别分析^[9](Linear Discriminant Analysis, LDA)等降维方法也可以实现信号稀疏,但由于 THz 信号的非线性这些方法的稀疏结果并不理想。所以寻找一种算法有效实现 THz 信号的稀疏,提取信号有效特征,减少大量无效的冗余特征,从根本上有效地减少存储压力,对于提高数据 THz 信号处理和分析效率具有重要意义。

针对以上问题,为了保证不损失 THz 信号有效特征同时最大程度减少冗余特征,文中提出基于梯度阈值的太赫兹时域信号自适应稀疏算法,基于二阶梯度法提取 THz 信号有效特征,通过自适应确定最佳阈值稀疏 THz 信号;然后,利用多高斯函数恢复 THz 信号;最后,通过二值化图像分割识别 THz 图像的缺陷区域。通过压缩率、相对均方根误差、相关系数等指标评价 DCT、PCA、K-SVD 算法及文中算法的压缩性能。

1 理论基础

图 1 所示多层胶接结构样件从上至下是由聚甲基丙烯酸酯酰胺 (Polymethylacrylimide, PMI)、上胶层、缓冲垫、下胶层及金属基体经过胶接技术^[10]得到的。实验采用的反射式 THz-TDS 系统的飞秒激光器为锁模钛蓝宝石飞秒激光器,产生的光脉冲中心波长为 810 nm,重复频率为 80 MHz,脉宽为 100 fs,系统的时间分辨率为 0.1 ps,快速扫描范围为 160 ps,对待测的样件进行逐点扫描,扫描步距为 0.5 mm,探测器焦距为 76.2 mm,有效焦深为±4.8 mm,以胶层信号的最佳信噪比为依据,调整样件在最佳焦深范围处。THz 波经 PMI、上胶层、缓冲垫、下胶层,最后到达多层胶接结构样件的金属基体表面,THz 波在不同介质表面产生的反射信号被 THz 探测器接收。THz 波在不同介质中传播时,在不同介质分界面时产生反射回波,根据先后通过介质的顺序不同,返回的太赫兹时

域信号也会携带胶层信息, THz 探测器会依据 THz 波在通过不同材料到达 THz 探测器时间的先后, 接收不同介质层的反射脉冲, 得到太赫兹时域信号^[11]。由于检测样件的胶层厚度与探测器接收信号的时间延迟有关, 设时间延迟为 Δt , 胶层厚度为 l , 则有 $\Delta t = 2nl/c$ 。式中: n 为介质折射率, c 为光速 ($c = 299\,792\,458\text{ m/s}$)。图 1 中 Δt_{up} 和 Δt_{down} 分别表示上胶层和下胶层的时间延迟, 可计算检测样件的上/下胶层厚度。

图 1 中 $E(\omega)$ 为太赫兹发射器射出太赫兹脉冲信号电场强度, 四层介质材料的厚度分别为 l_1, l_2, l_3 和 l_4 , 第一层介质 PMI 材料上表面返回的太赫兹脉冲信号电场强度记为 $E_{01}(\omega)$, 下表面返回的太赫兹脉冲信号电场强度记为 $E_{12}(\omega)$, 第二层介质上胶层下表面返回的太赫兹脉冲信号电场强度记为 $E_{23}(\omega)$, 第三层介质缓冲垫下表面返回的太赫兹脉冲信号电场强度记为 $E_{32}(\omega)$, 第四层介质下胶层下表面返回的太赫兹脉冲信号电场强度记为 $E_{24}(\omega)$ 。太赫兹接收器收到的太赫兹时域信号 $E(t)$ 为 $E_{01}(\omega)$ 、 $E_{12}(\omega)$ 、 $E_{23}(\omega)$ 、 $E_{32}(\omega)$ 及 $E_{24}(\omega)$ 之和的傅里叶逆变换, 即:

$$E(t) = \mathcal{F}^{-1} [E_{01}(\omega) + E_{12}(\omega) + E_{23}(\omega) + E_{32}(\omega) + E_{24}(\omega)] \quad (1)$$

式中: \mathcal{F}^{-1} 为傅里叶逆变换。

通过 THz-TDS 系统检测样件可得到具有 N 个采样点的太赫兹时域信号, 记为 THz 信号, 如图 2 所示, 图中纵坐标 Amplitude/a.u. 表示 THz 接收器接收的信号振幅大小, 横坐标 Time/ps 表示 THz 接收器采集信号的时间, 单位为 ps。THz 波在 PMI 与上胶层界面反射得到的波峰记为特征峰 1, 其对应时间位置记为 $T_{\text{up}1}$, 上胶层与缓冲垫界面反射得到的波谷记为特征谷 2, 其对应时间位置记为 $T_{\text{up}2}$; 缓冲垫与下胶层界面反射得到的波峰记为特征峰 3, 其对应时间位置记为 $T_{\text{low}1}$; 下胶层与基体界面反射得到的波峰记为特征峰 4, 其对应时间位置记为 $T_{\text{low}2}$ 。用点框将 THz 信号一部分时间区间划分为 4 个区域, 框位置不固定。可以看到 THz 信号的 4 个特征峰/谷的峰/谷值是某段时域内的最值, 最值点一般也是极值点, 故计算 THz 信号的极值点, 然后通过对比获得最值, 进而可确定 THz 信号有效特征区域。

分析 THz 信号的时域特征, 如图 3 所示, 虚线框内表示 THz 信号的有效特征信息所在区域, 即信号

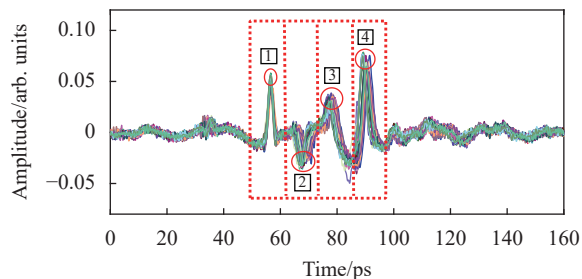


图 2 THz 信号特征峰/谷

Fig.2 THz signal characteristic peak/valley

的 4 个特征峰/谷所在时域区间, 定义为有效特征区域, 第一个有效特征峰对应时间位置^[12]记为 T_{up} , 最后一个有效特征峰对应时间位置记为 T_{low} , 则有效特征区域对应时间区间为 $[T_{\text{up}}, T_{\text{low}}]$ 。特征区域内还包含了 THz 信号上/下层飞行时间^[13], 特征峰 1 对应时间位置 $T_{\text{up}1}$ 与特征谷 2 对应时间位置 $T_{\text{up}2}$ 的差值为上层飞行时间 Δt_{up} , 特征峰 3 对应时间位置 $T_{\text{low}1}$ 与特征峰 4 对应时间位置 $T_{\text{low}2}$ 差值为下层飞行时间 Δt_{low} , 飞行时间是计算多层胶接结构样件上/下胶层厚度的重要参数。

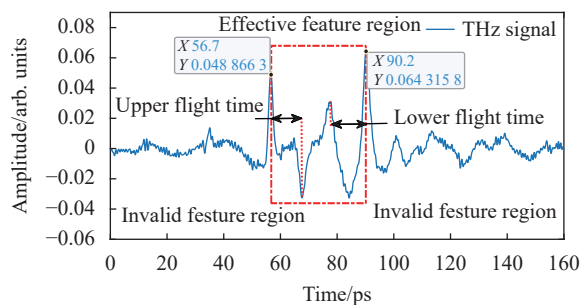


图 3 THz 信号的有效特征区域及无效特征区域

Fig.3 Effective and invalid feature regions of THz signal

THz-TDS 系统的时间分辨率为 0.1 ps, THz 信号的有效特征区域内采样点的数量 $N_1 = 10(T_{\text{low}} - T_{\text{up}}) + 1$, 框线外无效特征区域, 此区域采样点数量为 $N_2 = N - N_1$ 。

如图 3 所示, 经过统计分析, THz 信号上层飞行时间为 10.9 ps, 下层飞行时间为 12.4 ps, 有效特征区域内采样点数量为 335, 无效特征区域采样点数量为 1265, 无效特征区域的采样点数量占据了总采样点数量的 79.06%, 可见 THz 中的无效信息不仅占据极大存储空间, 更降低了数据处理效率, 增加了数据计算时间, 故在不影响信号有效信息的前提下, 对 THz 信

号进行稀疏,减少无效特征并提高数据处理效率,对于 THz-TDS 检测的数据处理工作是十分必要且有意义的。

2 算法分析

2.1 二阶梯度法特征提取

记采样点为 n 的 THz 信号为 X , 采样时间 T , 由于 THz-TDS 系统检测样件时等时间间隔采样, 记时间间隔为 ΔT 。将信号 X 一阶梯度定义为 J , 表达式为:

$$J(i) = \frac{X(i+1) - X(i)}{\Delta T}, \quad i = 1, 2, \dots, n-1 \quad (2)$$

由公式 (2) 可得二阶梯度, 定义为 N ,

$$J_2(j) = \frac{J(j+1) - J(j)}{\Delta T}, \quad j = 1, 2, \dots, n-2 \quad (3)$$

可利用二阶梯度搜索 X 的极值点, 极小值数量记为 n_1 , 极大值数量记为 n_2 , 提取 THz 信号的特征峰, 确定上层特征峰和下层特征峰对应时间位置, 即 T_{up1} 、 T_{up2} 、 T_{low1} 和 T_{low2} , 其中 $T_{up1} = T_{up}$ 和 $T_{low2} = T_{low}$, 进而提取有效特征区域 $[T_{up1}, T_{low2}]$ 。

2.2 自适应阈值 THz 信号稀疏化

将原 THz 信号 T_{up1} 与 T_{up2} 的差值记为上层飞行时间 Δt_{up1} , 稀疏 THz 信号 T_{up1-s} 与 T_{up2-s} 的差值记为上层飞行时间 Δt_{up2} 。文中利用 THz 信号的时域特征自适应确定阈值稀疏信号, 以稀疏前后 THz 信号上层飞行时间误差为限制条件, 假设检测样件 THz 信号的飞行时间能够接受的最大误差为 0.05, 稀疏前后误差 e 表示为 $e = |\Delta t_{up2} - \Delta t_{up1}|$ 。文中基于梯度阈值的太赫兹时域信号自适应稀疏算法步骤如下:

Step 1: 输入: 采样点为 N 的 THz 信号 X , 设置有效特征区域初始阈值 $\tau_0 = \frac{n_1 + n_2}{10(T_l - T_u)}$, 无效特征区域初始阈值 $\eta_0 = \frac{n_1 + n_2}{N - (n_1 + n_2)}$, 阈值精度为 0.01;

Step 2: 通过阈值 τ 对 THz 信号有效特征区域稀疏, 稀疏 THz 信号记为 Y , 采样点数量记为 M_1 , 提取稀疏后 THz 信号的 T_{up1-s} 与 T_{up2-s} ;

Step 3: 计算 $e = |\Delta t_{up2} - \Delta t_{up1}|$;

Step 4: 若 $0 \leq e < 0.05$, 则尝试调小阈值 $\tau_{j+1} = \tau_j - 0.01$, $\tau_j \neq 0$, j 为迭代次数, 返回 Step 2, 若误差仍满足 $0 \leq e < 0.05$ 则继续调整, 直至 $e = 0.05$ 迭代停止, 取此次迭代的阈值为最佳稀疏阈值 τ , 转至 Step 6;

Step 5: 若 $e > 0.05$, 则需要调大阈值 $\tau_{j+1} = \tau_j + 0.01$, 返回 Step 2, 直至满足误差 $e = 0.05$, 则停止, 此次迭代的阈值为最佳稀疏阈值 τ , 转至 Step 6;

Step 6: 通过阈值 η 稀疏 THz 信号无效特征区域, 稀疏后采样点数量记为 M_2 ;

Step 7: 恢复 THz 信号, 记为 \hat{X} , 依据误差分析判断信号是否失真;

Step 8: 若信号未失真, 尝试调小阈值 $\eta_{k+1} = \eta_k - 0.01$, $\eta_k \neq 0$, K 为迭代次数, 返回步骤 6, 若下一次迭代信号失真, 则此次迭代的阈值为无效区域最佳稀疏阈值 η , 转至 Step 10;

Step 9: 若信号失真, 调整阈值 $\eta_{k+1} = \eta_k + 0.01$, 返回步骤 6, 直至信号恢复停止迭代, 此次迭代阈值为最佳稀疏阈值 η , 转至 Step 10;

Step 10: 输出: 恢复 THz 信号 \hat{X} 。

2.3 稀疏化 THz 信号恢复

将 THz 信号划分为若干区间, 通过高斯函数^[14] 分区间拟合恢复 THz 信号, 恢复信号记为 \hat{X} , $\hat{X} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_i, \dots, \hat{x}_m)$, $i = 1, 2, \dots, m$, 高斯拟合函数可以表示为:

$$g_i = f(\hat{x}_i) = a_i \exp(-((\hat{x}_i - b_i)/c_i)^2) \quad (4)$$

式中: a_i , b_i , c_i 为高斯函数在分段区间的拟合系数, 分别表示高斯信号的峰高、峰位置和半宽度。多层胶接结构的 THz 信号拟合过程存在多个高斯峰叠加的情况, 如图 4 所示, 某段时域的 THz 信号在高斯拟合时至少需要 3 个高斯峰进行叠加。

多高斯函数表达式为:

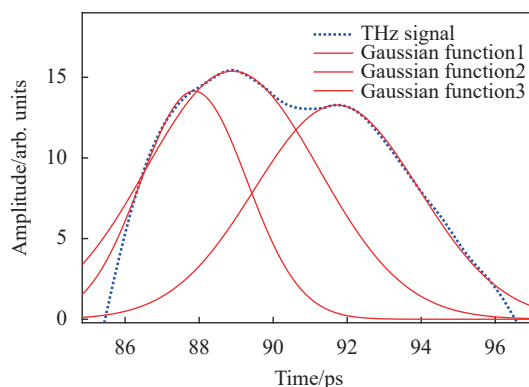


图 4 多高斯峰叠加拟合 THz 信号

Fig.4 Multi-Gaussian peak superposition to fit THz signal

$$g = f(\hat{x}) = \sum_{i=1}^{i=k} a_i \exp(-((\hat{x}_i - b_i)/c_i)^2) \quad (5)$$

式中： k 为 THz 信号高斯峰的总数。不考虑多高斯拟合信号过程存在的误差，根据最小二乘法结合梯度下降法^[15]求解高斯函数的拟合系数，算法的迭代格式为：

$$\hat{x}_{k+1} = P_{\Omega}(t_k - \delta_k \nabla f(\hat{x}_k)) \quad (6)$$

式中： $\delta_k > 0$ 为步长； $\nabla f(\hat{x}_k)$ 为 THz 信号 \hat{x} 的梯度； $P_{\Omega}(\cdot)$ 为点 \hat{x}_k 到 THz 信号时间域 Ω 的投影； k 为迭代次数。

3 分析与讨论

3.1 样件制备

文中设计制作一块 60 mm×60 mm×27 mm 的多层胶接结构样件，其中 PMI 材料厚度为 20 mm，缓冲垫

厚度为 3 mm，金属基体的厚度为 2 mm，上胶层与下胶层材料厚度均为 1 mm，然后在上、下胶层的四边利用抽膜法^[16]模拟约 0.3 mm 厚度空气隙的脱粘缺陷，得到带有预制脱粘缺陷的多层胶接结构样件，如图 5 所示，图 5(b) 圆圈圈出部分是空气隙。通过反射式 THz-TDS 系统分别对制备完成的实验样件进行检测，信号采样时间间隔为 0.1 ps，时间窗口为 160 ps。将检测的 THz 信号分为正常信号与缺陷信号两种，分别记为信号 A、信号 B。

图 6 实线为样件胶层正常所检测到的 THz 信号，即信号 A。图 6 虚线为样件胶层存在缺陷部分检测得到的 THz 信号，即信号 B，图中椭圆圈出的部分为缺陷表征，分别表示的是上层缺陷和下层缺陷，与信号 A 对比，特征峰 1 与特征谷 2 之间突然出现幅值明

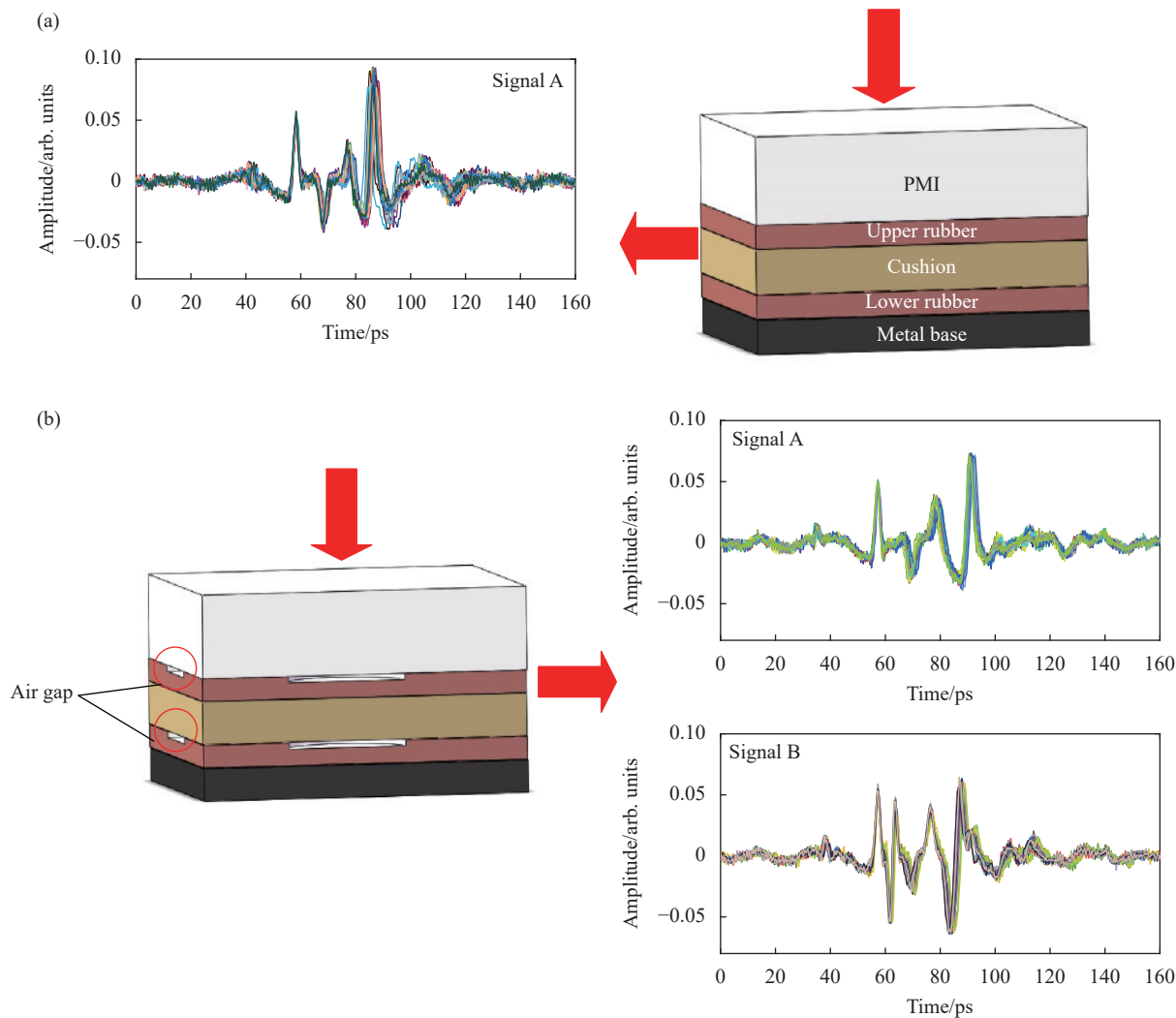


图 5 (a) 正常多层胶接结构样件及 THz-TDS 检测信号；(b) 脱粘缺陷实验样件及 THz-TDS 检测信号

Fig.5 (a) Normal multilayer bonding structure sample and THz-TDS detection signal; (b) Debonding defect test sample and THz-TDS detection signal

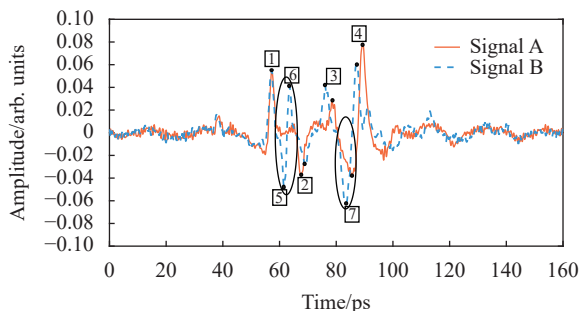


图 6 正常及缺陷 THz 信号

Fig.6 Normal and defect THz signal

显增大的特征谷 5 和特征峰 6; 特征峰 3 与特征峰 4 提前出现, 时间位置减小, 且特征峰 3 与特征峰 4 之

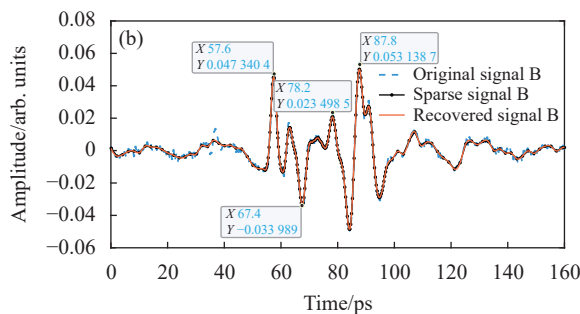
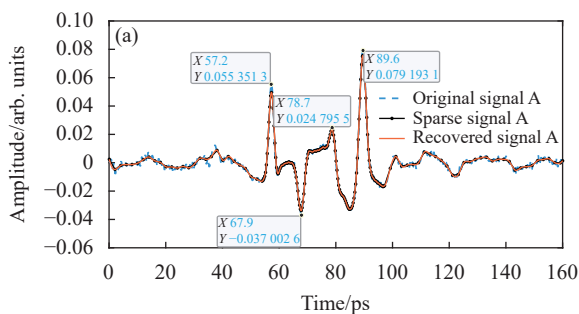


图 7 (a) 正常信号 A; (b) 缺陷信号 B

Fig.7 (a) Normal signal A; (b) Defect signal B

表 1 信号 A 与信号 B 特征峰峰值对应时间位置

Tab.1 Time position corresponding to characteristic peak-to-peak values of signal A and signal B

THz signal	T_{up1}/ps	T_{up2}/ps	T_{low1}/ps	T_{low2}/ps
Signal A	57.2	67.9	78.7	89.6
Signal B	57.6	67.4	78.2	87.8

从表 1 可知, 图 7 缺陷信号 B 特征峰 1 出现时间比正常信号 A 延迟了 0.4 ps, 信号 B 特征谷 2 比信号 A 时间提前了 0.5 ps, 信号 B 特征峰 3 比信号 A 提前了 0.5 ps, 信号 B 特征峰 4 提前了 1.4 ps。基于 THz 信号的特征峰对应时间位置可确定信号 A 和信号 B 有效特征区域分别为 [57.2 ps, 89.6 ps], [57.6 ps, 87.8 ps]。基于自适应阈值稀疏 THz 信号, 设置最大允许误差为 0.05, 表 2 为图 7 信号 A 和信号 B 有效特征区域和无效特征区域的稀疏阈值 τ 和 η 。

由表 2 可知, 信号 A 的有效特征区域稀疏阈值为 0.42, 无效特征区域稀疏阈值为 0.07; 信号 B 有效

间的特征谷 7 幅值绝对值明显增大。

3.2 信号分析及算法对比

如图 7 所示, 预制缺陷样件的下胶层正常信号 A 与缺陷信号 B, 基于二阶梯度提取信号 A 和信号 B 的 4 个特征峰/谷, 特征峰/谷的峰/谷值对应时间位置如表 1 所示。基于梯度阈值的太赫兹时域信号自适应稀疏算法稀疏化处理信号 A 和信号 B, 并通过多高斯函数恢复信号。图 7 中虚线为原始信号; 带“·”标记的实线为稀疏信号, “·”表示稀疏点; 实线为多高斯函数恢复信号, 恢复信号与原始信号时间序列长度一致。

特征区域稀疏阈值为 0.33, 无效特征区域稀疏阈值为 0.09。稀疏阈值的值越小, 信号越稀疏, 在保证信号不失真且保留有效特征峰/谷的前提下, 得到图 7 所示稀疏信号。图 7(a) 中的信号 A 稀疏后采样点数量为 224, 图 7(b) 中的信号 B 稀疏后采样点数量为 216, 与原始信号相比可以发现信号 A 与信号 B 的无效信息大幅度减少, 并且保留了信号的 4 个有效特征峰, 这对于提高数据处理的效率具有极大意义。

表 2 信号 A 与信号 B 的稀疏阈值

Tab.2 Sparse threshold of signal A and signal B

THz signal	τ	η	Original data number	Sparse data number
Signal A	0.42	0.07	1 600	224
Signal B	0.33	0.09	1 600	216

图 7 所示的原始 THz 信号与恢复 THz 信号之间存在微小偏差, 是由稀疏造成的, 计算稀疏信号的幅值误差, 如图 8 所示。

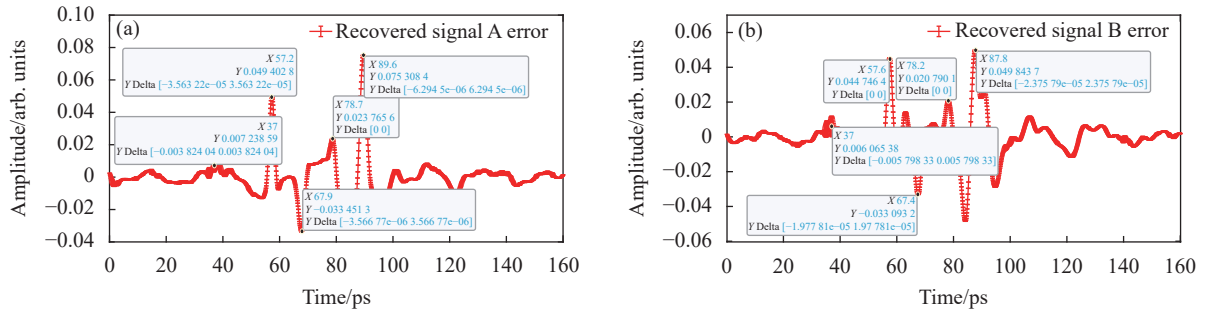


图 8 (a) 恢复信号 A 幅值误差; (b) 恢复信号 B 幅值误差

Fig.8 (a) Amplitude error of recovered signal A; (b) Amplitude error of recovered signal B

图 8(a) 信号 A 稀疏后幅值误差最大值约为 0.003 8, 4 个有效特征峰/谷的幅值误差最大值约为 3.56 e-5; 图 8(b) 为信号 B 稀疏后幅值误差最大值约为 0.005 8, 4 个有效特征峰/谷的幅值误差最大值约为 2.37 e-5。在误差允许的范围, THz 信号经阈值稀疏, 不仅保留了信号有效特征峰/谷, 还减少了无效特征信息, 实现了 THz 信号的有效压缩。

文中通过压缩率^[17](Compression Ratio, CR)、相对均方根误差^[18](Relative Root Mean Squared Error, RMSE)及相系数^[19]3 项指标评价 DCT、PCA、K-SVD 算法与文中算法 4 种算法的压缩性能。

压缩率是评价压缩效果的重要指标, 能较准确地反应信号稀疏程度。用 CR 表示, 计算公式为:

$$CR = \frac{N - M}{N} \times 100\% \quad (7)$$

式中: N 为原 THz 信号采样点数量; M 为稀疏后 THz 信号的采样点数量。CR 越大, 压缩率越高, 表示信号越稀疏。

相对均方根误差与相系数是衡量压缩算法的重要指标。分别记作 RMSE 和 r , 多高斯拟合 THz 信号 $\hat{X} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ 与原信号 $X = (x_1, x_2, \dots, x_n)$ 的 RMSE 和 r 计算公式为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N x_i^2}} \times 100\% \quad (8)$$

$$r = \frac{E(X\hat{X}) - E(X)E(\hat{X})}{\sqrt{D(X)}\sqrt{D(\hat{X})}} \quad (9)$$

式中: $E(X)$ 为信号的数学期望; $D(X)$ 为数学方差。

RMSE 越小, 则表明算法恢复信号与原信号相位差越小; r 越高, 恢复信号与原信号相关度越高。

利用 DCT、PCA、K-SVD 算法与文中算法分别对 THz-TDS 系统检测制备的多层胶接结构样件的检测数据进行压缩, 图 9 为 4 种算法压缩性能的对比。

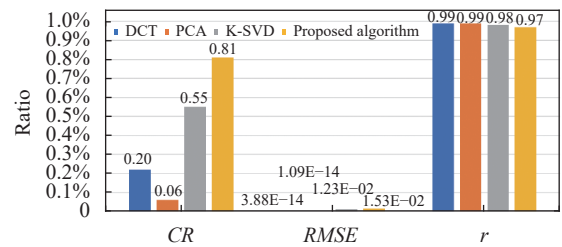


图 9 信号压缩算法压缩性能对比

Fig.9 Compression performance comparison of signal compression algorithms

如图 9 所示, THz 信号经过 4 种算法压缩, 文中算法压缩率最高, 较 DCT 提高了 59%, 较 PCA 提高了 75%, 较 K-SVD 算法提高了 26%。虽然文中算法相比其他 3 种算法相对均方根误差的值较高, 但误差低于 2%, 且 4 种算法的相系数均超过了 0.97, 均呈高度相关。

3.3 THz 图像分析讨论

通过对 THz-TDS 系统检测预制脱粘缺陷的多层胶接结构样件, 分别对其 THz 信号的飞行时间, 峰峰值及峰值成像^[20], 得到 124 pixel×128 pixel 大小的图像, 图像的每一个像素点均对应一组 1600 个采样点的 THz 信号, 其中, 峰峰值图像与峰值图像表示 THz 接收器接收的反射信号强度, 表征材料的反射性能, 如图 10 所示, 缺陷样件的图像部分区域呈现颜色突变, 该区域为样件缺陷区域。利用文中算法稀疏恢

复 THz 信号的飞行时间图像,峰峰值图像和峰谷值图像如图 11 所示。图 10~11 中 (a)~(d) 为缺陷样件的 THz 信号上、下层飞行时间图像,上层峰峰值图像和

下层峰谷值图像,可以看到 (c)~(d) 中虚线框内明显的上、下层缺陷区域,分别标记为 1、2、3、4。

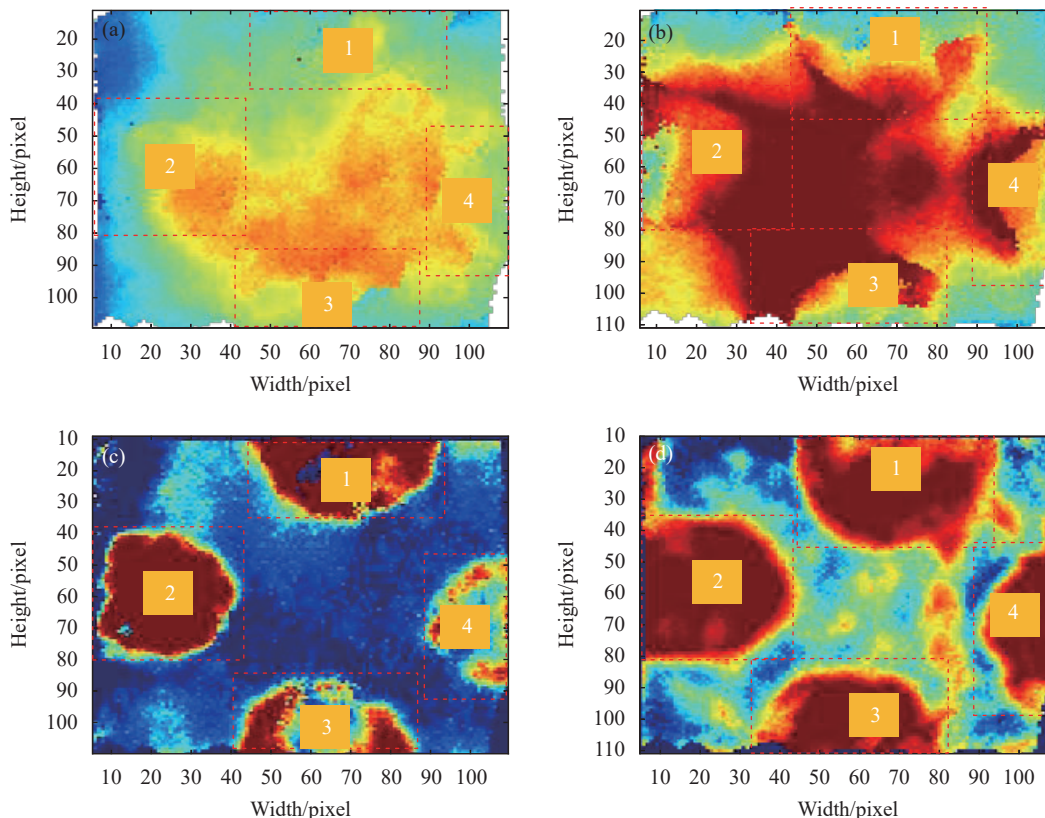


图 10 原始信号的 THz 图像。(a) 上层飞行时间图像; (b) 下层飞行时间图像; (c) 上层峰峰值图像; (d) 下层峰谷值图像

Fig.10 THz images of the original signal. (a) Upper flight time image; (b) Lower flight time image; (c) Upper peak-to-peak image; (d) Lower peak-to-peak image

稀疏恢复信号的 THz 图像与原信号的 THz 图像存在微小偏差,利用二值化进行图像分割^[21],识别缺陷样件 THz 图像的缺陷区域,分别计算 1、2、3、4 个缺陷区域面积,将稀疏恢复信号的 THz 信号缺陷区域的面积与原始信号的 THz 图像对应区域面积比值记为 ρ :

$$\rho = \frac{S_{\text{recover}}}{S_{\text{detect}}} \quad (10)$$

式中: S_{recover} 为稀疏恢复信号的 THz 图像某个缺陷区域面积大小; S_{detect} 为 THz-TDS 系统检测信号的 THz 图像对应缺陷区域面积大小。计算结果如表 3 所示。

由表 3 可以看到, THz 图像上层的缺陷区域 1 和

4 面积稀疏后变大,表明有少部分正常信号被错误识别为缺陷信号,而区域 3 的面积变小,则部分缺陷信号未识别, THz 图像上层缺陷区域面积最大偏差为 0.05; 同理, THz 图像下层的缺陷区域 1 面积变大,表明部分信号错误识别,区域 3 面积变小,则部分缺陷信号未识别, THz 图像下层缺陷区域面积最大偏差为 0.03。结果表明,经文中算法稀疏恢复的 THz 图像,虽然缺陷区域会存在微小偏差,但并不影响对样件中缺陷区域的识别,保证了缺陷识别精度。

由图 10~11 可以看到,经文中算法稀疏恢复信号的 THz 图像与原信号的 THz 图像对比,减少了大量的无效特征,提高了图像清晰度,不影响样件缺陷区域的识别。

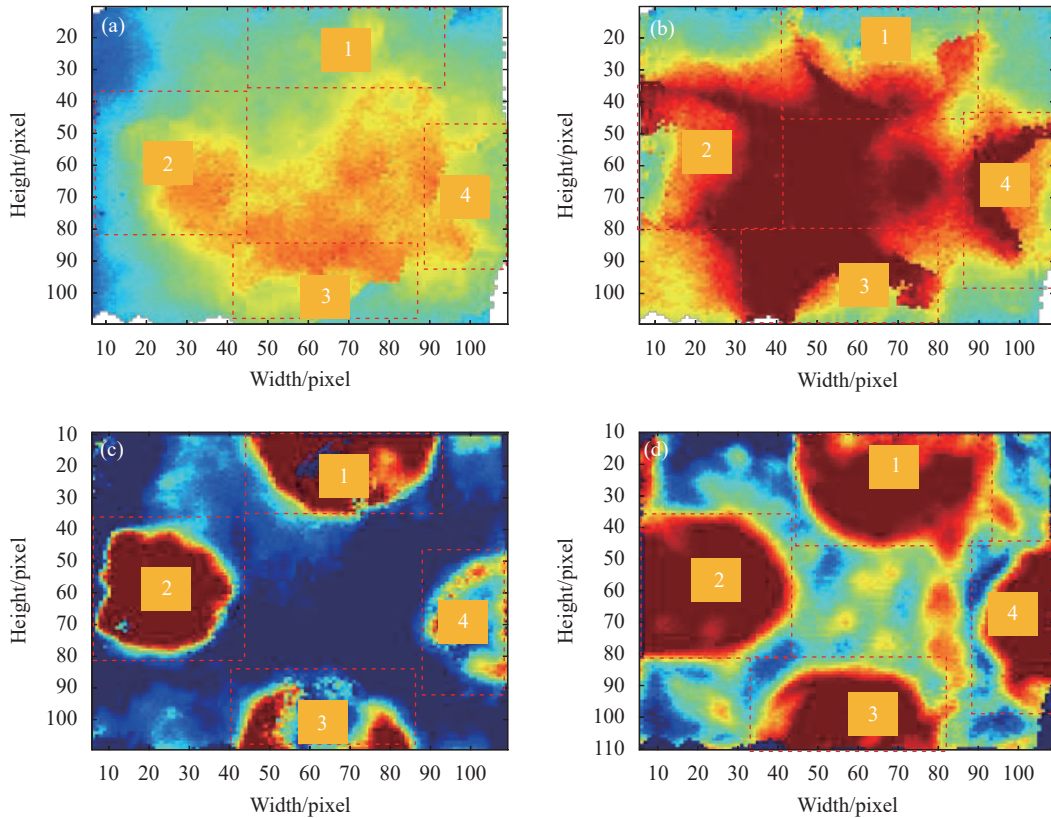


图 11 稀疏恢复信号的 THz 图像。(a) 上层飞行时间图像; (b) 下层飞行时间图像; (c) 上层峰峰值图像; (d) 下层峰峰值图像

Fig.11 THz images of sparse recovery signal. (a) Upper flight time image; (b) Lower flight time image; (c) Upper peak-to-peak image; (d) Lower peak-to-peak image

表 3 上、下层 THz 信号图像缺陷区域面积比

Tab.3 Area ratio of defect regions in upper and lower THz signal images

	Defect region 1	Defect region 2	Defect region 3	Defect region 4
Upper	1.01	1.00	0.96	1.05
Lower	1.03	1.00	0.99	1.00

3.4 算法效率分析

实验所用 CPU 为 2.4 GHz, 内存空间为 16 GB, 通过算法压缩后数据计算时间与数据存储空间表征算法效率, 图 12 为 DCT、PCA、K-SVD 算法及文中算法压缩后数据计算时间和存储空间的对比结果。

由图 12 可以看到, 文中算法压缩的数据计算时间最短, 仅需 0.25 s, 占用存储空间最少, 仅占 0.0149 GB 空间, 相比其他算法缩短了大约 80% 的时间, 减小了大约 95% 空间占用, 极大地提高了数据处理效率。

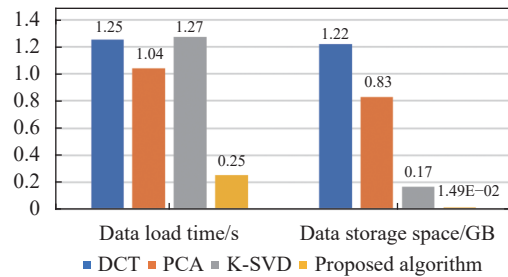


图 12 数据计算时间及存储空间

Fig.12 Data calculation time and storage space

4 结 论

文中针对具有大量复杂和冗余特征信息的 THz 信号的压缩问题, 提出基于梯度阈值的太赫兹时域信号自适应稀疏算法。利用二阶梯度提取信号的有效特征, 确定有效特征区域, 再基于信号的时域特征自适应确定有效特征区域及无效特征区域的稀疏阈值, 利用阈值稀疏 THz 信号, 最后通过多高斯函数恢复 THz 信号。文中算法实现了 81% 以上的压缩率, 较

DCT 提高了 59%, 较 PCA 提高了 75%, 较 K-SVD 算法提高了 26%, 算法的相对均方根误差小于 2%, 相关系数大于 0.97, 并且相比传统信号压缩算法数据计算时间缩短了 20%, 空间占用减少了 95%。文中算法减少了大量无效特征, 保留有效特征, 保证 THz 图像缺陷识别的精度, 适用于具有复杂冗余特征信息的正常 THz 信号及缺陷 THz 信号的压缩。

参考文献:

- [1] Li Wenjun, Wang Tianyi, Zhou Yu, et al. Terahertz non-destructive inspection of air defect within adhesive layers of multi-layer bonded structure [J]. *Acta Optica Sinica*, 2017, 37(1): 0111002. (in Chinese)
- [2] Zhou Xiaodan, Li Lijuan, Zhao Duo, et al. Application of terahertz technology in nondestructive testing of ceramic matrix composite defects [J]. *Infrared and Laser Engineering*, 2016, 45(8): 0825001. (in Chinese)
- [3] Ma W W, Gao J X, Yuan Y N, et al. Suppression of continuous wave interference in loran-c signal based on sparse optimization using tunable Q-factor wavelet transform and discrete cosine transform [J]. *Sensors*, 2021, 21(21): 86-90.
- [4] Miao F, Zhao R Z. A new method of vibration signal denoising based on improved wavelet [J]. *Journal of Low Frequency Noise, Vibration and Active Control*, 2022, 41(2): 637-645.
- [5] Mallat S, Zhang Z. Matching pursuits with time-frequency dictionaries [J]. *IEEE Transaction on Signal Process*, 1993, 41(12): 3397-3415.
- [6] Rubnstein R, Member S, Michael E, et al. Dictionaries for Sparse Representation Modeling[C]//Proceedings of the IEEE, 2010, 98(6): 1045-1057.
- [7] Guo Junfeng, Shi Bin, Wei Xingchun, et al. A Method of reconstruction of compressed measuring for mechanical vibration signals based on k-svd dictionary-training algorithm sparse representation [J]. *Journal of Mechanical Engineering*, 2018, 54(7): 97-106. (in Chinese)
- [8] Ni Jiapeng, Shen Tao, Zhu Yan, et al. Terahertz spectrum identification with diffusion map [J]. *Spectroscopy and Spectral Analysis*, 2017(8): 2360-2364. (in Chinese)
- [9] Liu Junxiu, Du Bin, Deng Yuqiang, et al. Terahertz-spectral identification of organic compounds based on differential pca-svm method [J]. *Chinese Journal of Lasers*, 2019, 46(6): 0614039. (in Chinese)
- [10] Cao Binghua, Zheng Dedong, Fan Mengbao, et al. Efficient and reliable thickness measurement method for multilayer coatings based on terahertz time-domain spectroscopy technology [J]. *Acta Optica Sinica*, 2022, 42(1): 0112001. (in Chinese)
- [11] Jia Meihui, Li Lijuan, Ren Jiaojiao, et al. Terahertz nondestructive testing signal recognition based on PSO-BP neural network [J]. *Acta Photonica Sinica*, 2021, 50(9): 0930004. (in Chinese)
- [12] 高翔, 张涛. 视觉SLAM十四讲: 从理论到实践[M]. 第2版. 北京: 电子工业出版社, 2019.
- [13] Ren Jiaojiao, Li Lijuan, Zhang Dandan, et al. Multi-feature parameter neural network analysis technique based on terahertz nondestructive testing [J]. *Acta Photonica Sinica*, 2017, 46(4): 210-216. (in Chinese)
- [14] Yang Jiayi, Xiong Yongqian. Gaussian fitting algorithm for spectral overlapping peaks and terahertz time-domain curves [J]. *Laser & Optoelectronics Progress*, 2021, 58(4): 0407002. (in Chinese)
- [15] Chajrabarti Kushal, Gupta Nirupam, Chopra Nikhil. Iterative pre-conditioning for expediting the distributed gradient-descent method: The case of linear least-squares problem [J]. *Automatica*, 2022, 137: 110095.
- [16] Zhou Tongyu, Li Lijuan, Ren Jiaojiao, et al. Simulation study on bonding defects of pulsed terahertz rubber based on finite difference time domain [J]. *Journal of Changchun University of Science and Technology (Natural Science Edition)*, 2021, 44(1): 36-41. (in Chinese)
- [17] Hla Myo Tun, Win Khaing Moe, Zaw Min Nang. Analysis on ECG data compression using wavelet transform technique [J]. *International Journal of Psychological and Brain Sciences*, 2017, 2(6): 127-140.
- [18] Li Shunshan, Zhuang Tiange, Li Gaoping, et al. Study of performance evaluation of electrocardiogram compression [J]. *Journal of Shanghai Jiaotong University*, 2001, 35(5): 706-709. (in Chinese)
- [19] Wu Qian, Zhang Rong, Xu Da-wei. Hyperspectral data compression based on sparse representation [J]. *Journal of Electronics & Information Technology*, 2015, 37(1): 78-84. (in Chinese)
- [20] Xiong Weihua, Li Lijuan, Ren Jiaojiao, et al. THz multi-feature parameter imaging of bonding defects of high temperature composite materials [J]. *Acta Optica Sinica*, 2020, 40(17): 1711001. (in Chinese)
- [21] Wang Zijun, Qiu Yanrui, Yang Hongxiao, et al. Algorithm for defect segmentation in infrared nondestructive testing based on robust Otsu [J]. *Infrared and Laser Engineering*, 2019, 48(2): 0204004. (in Chinese)

Adaptive sparse algorithm for terahertz time domain signals based on gradient threshold

Qin Lili^{1,2}, Li Lijuan^{1,2,3*}, Ren Jiaojiao^{1,2,3}, Gu Jian^{1,2,3}, Xiong Weihua^{1,2}, Zhang Dandan^{1,2,3}, Zhu Lili^{1,2},
Zhang Jiyang^{1,2,3}, Xue Junwen³, Jiang Baihong⁴, Gao Zenghua⁴

(1. Key Laboratory of Optoelectronic Measurement and Optical Information Transmission Technology, Ministry of Education, School of Opto-Electronic Engineering, Changchun University of Science and Technology, Changchun 130022, China;

2. National Demonstration Center for Experimental Opto-Electronic Engineering Education, School of Opto-Electronic Engineering, Changchun University of Science and Technology, Changchun 130022, China;

3. Zhongshan Research Institute, Changchun University of Science and Technology, Zhongshan 528400, China;

4. Aerospace Special Materials and Processing Technology Institute, Beijing 100074, China)

Abstract:

Objective Bonding structure is widely used in aviation, aerospace, national defense and other fields. But during service, the bond interface may appear disbonding defects or damage, seriously reducing the bearing capacity of the structure and affecting the structure safety. Terahertz nondestructive testing technology is widely used in the nondestructive test of composite materials. Terahertz time-domain spectroscopy technology can effectively realize the nondestructive test and identification of internal defects of multilayer adhesives. However, the terahertz detection signal carries a large number of invalid redundant features, noise and other invalid information. With the gradual increase of detection data, the redundant and invalid information in the data, and the workload of data processing are also increasing. A large amount of invalid information not only consumes a lot of data processing and analysis time, but also brings great interference to the subsequent signal analysis work such as defect identification. To solve this problem, a gradient threshold adaptive sparse compression algorithm is proposed based on time-domain characteristics of terahertz signals with multi-layer adhesive structures.

Methods The gradient threshold adaptive sparse model is established. Effective time-domain features of terahertz signals were extracted using the second-order gradient (Fig.3), and the time-domain features of signals were used as constraints to determine the threshold sparse time-domain signals based on the time-domain features of signals, and the terahertz signals were recovered by the multi-Gaussian fitting function (Fig.4). The compression performance of the algorithm was evaluated according to the compression ratio, relative root mean variance and correlation coefficient, and the data processing time and memory occupied space were used to characterize the compression efficiency of the algorithm.

Results and Discussions Terahertz detection signals were divided into normal signals and defective signals (Fig.5), and signal characteristic peaks were extracted (Tab.1) to determine effective feature intervals. The maximum allowable error was set as 0.05, and the threshold was determined adaptively. The sparse recovery results of terahertz signals were shown (Fig.7). The reconstruction error of the recovered signal is less than 0.006 (Fig.8). The compression rate of this algorithm reaches 81%, which is 59% higher than that of discrete cosine transform, 75% higher than that of principal component analysis, and 26% higher than that of K-singular value decomposition. The relative root-mean-square error of the algorithm is less than 2%, and the correlation coefficient is greater than 0.97 (Fig.9). Compared with the traditional signal compression algorithm, the data processing time is reduced by 20%. Space utilization is reduced by 95% (Fig.12). This algorithm achieves effective compression of terahertz signal. Combined with terahertz imaging technology and binarized threshold

segmentation method, the debonding defects of the sample were identified, and the identification deviation was less than 0.05 (Tab.3). The results show that the algorithm improves the efficiency of data analysis and guarantees the accuracy of defect identification.

Conclusions A gradient threshold adaptive sparse algorithm is proposed to solve the problem of terahertz signal feature redundancy and low processing efficiency. The algorithm has the advantages of strong adaptive ability, high compression rate, fast running speed and low complexity. The second order gradient is used to extract the signal feature peak and determine the effective feature region. Then, according to the time-domain characteristics of terahertz signals, sparse thresholds and sparse signals of effective and invalid feature regions are determined. Finally, signals are restored by using multiple Gaussian functions. The compression ratio of the algorithm is greater than 81%, the relative root mean square error is less than 2%, the correlation coefficient is greater than 0.97, and the defect identification error is less than 5%. Compared with the traditional signal compression algorithm, the data computation time is reduced by 20% and the space is reduced by 95%. The algorithm reduces a large number of invalid features and retains effective features, ensuring the accuracy of terahertz image defect recognition. It is suitable for compression of normal terahertz signals and defective terahertz signals with complex redundant characteristic information.

Key words: terahertz time domain spectroscopy; sparse representation; feature extraction; adaptive threshold; multilayer bonding structure

Funding projects: Science and Technology Development Plan of Jilin Province (20220508032RC);
Second Batch of Social Welfare and Basic Research Project of Zhongshan City (2022B2012)