

引用格式: XU Haiyan, KOU Qingkang, XIE Yingjuan, et al. Feature Extraction Algorithm of Optical Fiber Vibration Signal Based on Compensation Distance Estimation[J]. Acta Photonica Sinica, 2022, 51(4):0406001

许海燕,寇庆康,谢迎娟,等. 基于补偿距离估计的光纤振动信号特征提取算法[J]. 光子学报, 2022, 51(4):0406001

基于补偿距离估计的光纤振动信号特征提取 算法

许海燕,寇庆康,谢迎娟,朱俊,李敏

(河海大学 物联网工程学院, 江苏 常州 213022)

摘 要:为进一步提高光纤振动信号模式识别的实时性及准确性,提出一种基于补偿距离估计技术的信号特征提取算法。算法借鉴人类听觉感知机理,对光纤传感振动信号提取 Mel 频率倒谱系数特征向量,采用补偿距离估计技术制定特征筛选策略实现特征评估与优化。实验结果表明,基于补偿距离估计技术的振动信号特征提取算法可以有效减少光纤传感系统中影响分类精度的冗余向量,解决振动信号特征表征及运算复杂等问题,进一步提高光纤传感振动信号模式识别的有效性和实时性,并且具有一定的抗噪性能。

关键词:光纤传感;特征提取;补偿距离估计;特征优化;模式识别

中图分类号:TN911.7

文献标识码:A

doi:10.3788/gzxb20225104.0406001

0 引言

光纤传感系统广泛用于油气长输管道、隧道安全监测、大型结构安全监测、周边安全等诸多领域^[1-5]。光纤传感信号识别在实时监测、异常报警等多个方面起关键作用,其工作性能直接决定了整个监测系统的性能,因此,能否快速、准确地进行识别分类对于保障维护各个领域的安全和减少设备损坏造成的成本损失而言具有重要意义。

经过光纤传感系统提取的特征向量代表了光纤振动信号的特征,提取算法的有效性直接影响事件识别结果的准确性。因此,根据特征提取算法,获得有效的特征参数,对提高识别率尤为关键。国内外学者已经在这方面做出大量研究。LI Junchan 等提出了一种基于相关向量机的光纤振动传感器模式识别方法^[6]。首先对初始振动信号进行小波去噪预处理,然后利用小波分解算法得到各级分解系数,构建特征向量,最后利用相关向量机作为分类器进行分类识别。但是基于小波分解的特征提取方法由于信号的特征容易受到高斯噪声的影响,在小波分解处理过程中存在模糊提取和线性稳态的缺陷。WANG Liang 等提出基于门限过零率和稀疏编码器的算法提取光纤振动信号的两级特征^[7]。第一级特征提取过零率特征,以识别振动是否发生,通过第一级特征提取,降低光纤振动信号数据的维数。在发生振动的情况下,采用稀疏自编码神经网络算法提取振动信号的高维特征。但是采用神经网络算法需要大量数据进行训练。LIU Kun 等提出并论证了一种基于混合特征提取算法和组合分类器的识别方案^[8]。提取过零率、样本熵、小波包能量熵、峰度和多尺度置换熵组合为混合特征向量,然后采用一种支持向量机和径向基函数神经网络相结合的分类器进行识别分类。但是多种特征提取方案会增加特征提取过程的时间成本。

近年来,借鉴听觉感知机理的 Mel 频率倒谱系数(Mel-frequency Cepstral Coefficients, MFCC)特征提取方法在振动信号特征提取中得到了广泛的应用,其算法简单,能有效提取出高度可识别的特征信息,并取得良好的分类效果。MFCC 特征提取方法可以对光纤传感系统采集到的振动信号进行特征提取,进而对模

基金项目:国家自然科学基金(No. 61701169)

第一作者:许海燕(1978—),女,副教授,博士,主要研究方向为光纤传感与智能信息处理。Email: hyxu@hhu.edu.cn

收稿日期:2021-09-14;录用日期:2021-10-26

<http://www.photon.ac.cn>

式进行预测,从而识别干扰信号^[9]。使用MFCC特征提取算法时,提取的特征向量存在维数高、向量冗余的问题。当它被送到分类器进行训练和识别时,不仅会增加时间成本,还会降低识别准确率^[10]。因此,如何实现MFCC特征向量的有效降维是提高光纤传感振动信号实时性和准确性的关键。

目前已知的降维技术分为两类:一类是基于原始特征向量进行选择,另一类是从原始特征中重新提取特征^[11]。主成分分析(Principal Component Analysis, PCA)被广泛用于降低MFCC特征向量的维数。通过从原始特征向量中重新提取特征向量,实现了MFCC特征向量的特征降维,可以提高识别和分类^[12]。但是,由于主成分分析方法是从原始特征向量中重新提取特征矩阵,对原始特征向量的各个特征维度的意义具有一定的模糊性,方差小的非主成分也可能包含不同的信息。对后续的和分类会产生一定的影响。

针对上述问题,本文提出了一种基于补偿距离估计技术(Compensation Distance Estimation Technology, CDET)的特征提取方法,采用对特征向量类内离散性和类间离散性联合评估的方式,对特征矩阵的不同维度进行特征评分,通过在原有特征向量矩阵中删减低评分冗余向量进行特征降维,从而解决冗余向量对分类的影响以及特征提取维数过多造成的运算复杂问题。

1 算法原理

本文提出一种基于补偿距离估计的特征提取方法。通过该算法可以解决特征向量维数过多导致的复杂运算问题。基于补偿距离估计的特征提取算法可以优化特征向量矩阵,减少影响分类精度的冗余向量。采用基于听觉感知机理的MFCC算法进行初次特征提取,并制定特征筛选策略,采用补偿距离估计技术降低特征维数,解决MFCC提取的特征矩阵维数过多的问题,提高光纤传感信号识别的有效性和实时性。MFCC特征提取算法流程如下:

输入: $\{X_{c,n} | 1 \leq c \leq C, 1 \leq n \leq N_c\}$

输出: $\{C(n), D(n), D_2(n)\}$

1. 预加重:
$$Y(n) = X(n) - 0.97 \times X(n-1) \quad (1)$$

2. 将 $Y(n)$ 以 256 个采样点分帧为一帧;

3. 加汉明窗:
$$T(n) = Y(n) \times W(n) \quad (2)$$

$$W(n) = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1 \quad (3)$$

4. 快速傅里叶变换:
$$T(k) = \sum_{n=0}^{N-1} T(n) e^{-j2\pi nk/N} \quad 0 \leq k \leq K \quad (4)$$

5. 能量谱:
$$P(k) = \frac{1}{N} |T(k)|^2 \quad (5)$$

6. Mel 滤波器组:

频率转换公式:
$$\text{mel} = 1125 \times \ln\left(1 + \frac{f}{700}\right) \quad (6)$$

转换公式:
$$H_m(k) = \begin{cases} 0 & k < f(m-1) \text{ or } k \geq f(m+1) \\ \frac{2(k-f(m-1))}{[f(m+1)-f(m-1)][f(m)-f(m-1)]} & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{[f(m+1)-f(m-1)][f(m)-f(m-1)]} & f(m) \leq k \leq f(m+1) \end{cases} \quad (7)$$

7. 对数能量:
$$S(m) = \ln\left[\sum_{n=0}^{N-1} s(n) \cos\left(\frac{\pi n(m-0.5)}{M}\right)\right] \quad 0 \leq m \leq M \quad (8)$$

8. 离散余弦变换:
$$C(n) = \sum_{m=0}^{N-1} S(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right) \quad n = 1, 2, \dots, L \quad (9)$$

$$9. \text{ 差分系数: } \quad D(n) = \begin{cases} C(n+1) - C(n) & n < K \\ \frac{\sum_{k=1}^K k [C(n+k) - C(n-k)]}{\sqrt{\sum_{k=1}^K k^2}} & K \leq n < L - K \\ C(n) - C(n-k) & n \geq L - K \end{cases} \quad (10)$$

$$D_2(n) = \begin{cases} D(n+1) - D(n) & n < K \\ \frac{\sum_{k=1}^K k [D(n+k) - D(n-k)]}{\sqrt{\sum_{k=1}^K k^2}} & K \leq n < L - K \\ D(n) - D(n-k) & n \geq L - K \end{cases} \quad (11)$$

1.1 光纤传感信号 MFCC 特征提取

MFCC 特征提取方法是语音识别、故障识别和光纤信号识别中常用的一种特征提取方法^[13-15]。MFCC 将信号频率转换为 Mel 频率,设置等距 Mel 频率,再转换为频率,得到 Mel 频率滤波器,然后将每个带通滤波器的输出信号能量作为信号的基本特征^[16]。MFCC 特征提取算法的输入为不同类型的光纤振动信号样本,根据算法流程得到 MFCC 系数与差分系数组合矩阵的输出。MFCC 向量的三维图如图 1。

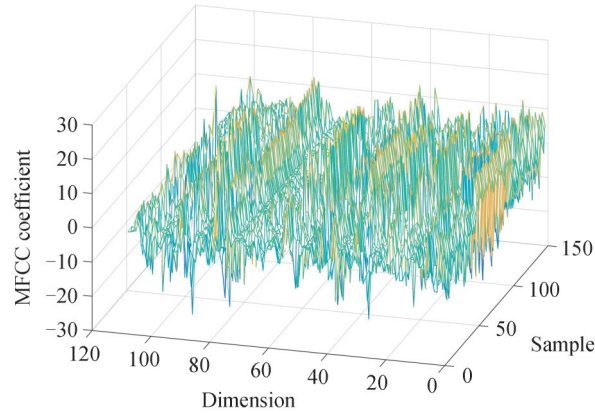


图 1 MFCC 三维特征图

Fig. 1 MFCC 3 dimension feature map

1.2 基于 CDET 的特征筛选算法

补偿距离估计技术(CDET)基于特征筛选策略,通过求取特征类别之间的类间距与特征的类内距,根据其比值获得特征评分,去掉评分低的冗余特征维度从而达到降维的目的。详细步骤如下:

设有 C 个模式类集合为

$$\{f_{c,n,k} | 1 \leq c \leq C, 1 \leq n \leq N_c, 1 \leq k \leq K\} \quad (12)$$

式中, $f_{c,n,k}$ 为第 C 个模式类下第 n 个样本第 k 个特征; N_c 为 c 类模式下的样本总数; K 为每个样本特征数目。

在每类模式下,有 N_c 个样本,因此共有 $\sum_{c=1}^C N_c$ 个样本,则有 $K \sum_{c=1}^C N_c$ 个特征,特征集定义为 $\{f_{c,n,k}\}$ 。

补偿距离评估技术实现步骤具体如下:

1) 计算相同条件样本的平均距离

$$d_{c,k} = \frac{1}{N_c(N_c - 1)} \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} |f_{c,i,k} - f_{c,j,k}| \quad (13)$$

平均化 $d_{c,k}$ 得到平均类内距

$$d_k^a = \frac{1}{C} \sum_{c=1}^C d_{c,k} \quad (14)$$

2) 定义并计算 d_k^a 的方差因子

$$R_k^a = \frac{\max |d_{c,k}|}{\min |d_{c,k}|} \quad (15)$$

3) 计算不同条件样本之间的平均距离

$$P_{c,k} = \frac{1}{N} \sum_{n=1}^{N_c} f_{c,n,k} \quad (16)$$

计算不同条件样本之间的平均类间距离

$$d_k^b = \frac{1}{C(C-1)} \sum_{c=1}^C \sum_{e=1}^C |p_{c,k} - p_{e,k}| \quad c, e = 1, 2, \dots, C; c \neq e \quad (17)$$

4) 定义并计算 d_k^b 的方差因子

$$R_k^b = \frac{\max |p_{c,k} - p_{e,k}|}{\min |p_{c,k} - p_{e,k}|} \quad c, e = 1, 2, \dots, C; c \neq e \quad (18)$$

5) 定义和计算补偿因子为

$$\delta_k = \frac{1}{\frac{R_k^a}{\max |R_k^a|} + \frac{R_k^b}{\max |R_k^b|}} \quad (19)$$

6) 计算 d_k^a 和 d_k^b 的比值并乘以补偿因子

$$g_k = \delta_k \frac{d_k^b}{d_k^a} \quad (20)$$

将 g_k 归一化, 得到距离评价标准

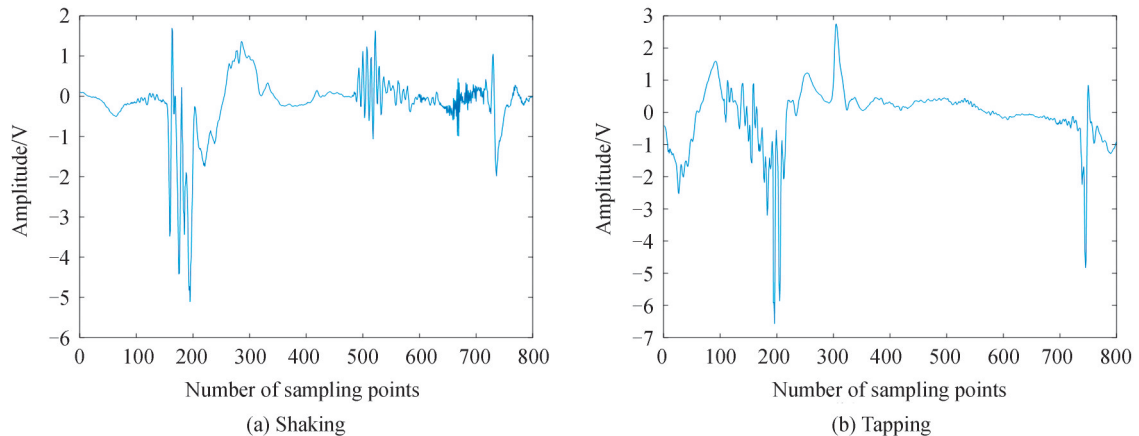
$$\Delta g_k = \frac{g_k}{\max |g_k|} \quad (21)$$

7) 设阈值为 N , 根据评分值确定阈值内的特征数, 选择在阈值范围内的 L 个特征, 将 k 维降至 L 维。

2 实验结果与分析

2.1 实验数据

实验所用的光纤信号数据来源于实验室光纤传感信号采集系统采集的三种信号种类, 分别是晃动、轻敲、重敲, 三种信号的时域图如图 2。经过预处理与端点检测, 将信号分割。实验使用三类信号共 210 个样本, 每类使用 50 个样本, 共 150 个样本作为训练集; 每类使用 20 个样本, 共 60 个样本作为测试集。其中, 样本的采样频率为 10 kHz。



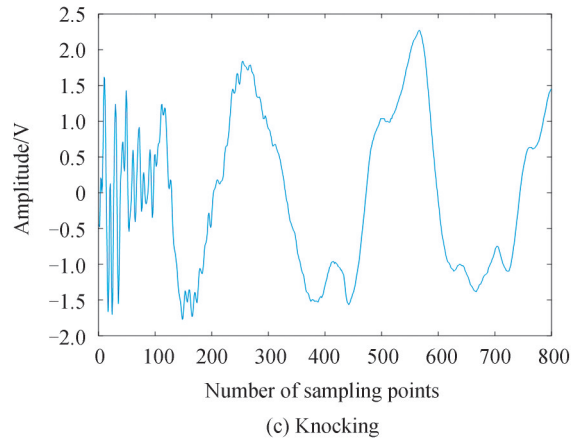


图2 三种信号时域图

Fig. 2 Time domain diagram of three kinds of signals

2.2 结果与分析

为了验证提出的MFCC优化方法的有效性,实验中采用支持向量机(SVM)的方法验证MFCC特征向量优化前与优化后的整个处理系统对三类信号分类识别的准确率,如表1。经过SVM的分类准确率结果与人工标注比较获得。实验流程可描述为3个步骤:

1)MFCC特征向量提取:对每个光纤振动信号样本加窗分帧并做FFT,然后将信号变换到Mel频率刻度下,最后通过倒谱分析得到MFCC向量。

2)特征降维:特征降维技术主要分为两类,一类是通过选择原有特征向量,删减冗余向量,例如本文所提出的补偿距离估计算法;另一类是通过从现有特征重新提取特征来降低维度,例如作为对比实验方法的主成分分析。

3)SVM识别分类:将MFCC特征向量或者经过降维方法处理的特征向量输入训练好的SVM分类器进行识别,标识结果与识别结果比对,计算正确率并输出。

表1 三类信号分类准确率

Table 1 Classification accuracy of three types of signals

Disturbance signal	Number of tests	Correct number	Number of errors	Accuracy
Shaking	20	20	0	100%
Tapping	20	15	5	75%
Knocking	20	18	2	90%

采用本文提出的对MFCC向量进行CDET特征评分算法,评分结果如图3,从图中可以发现1维、41维、

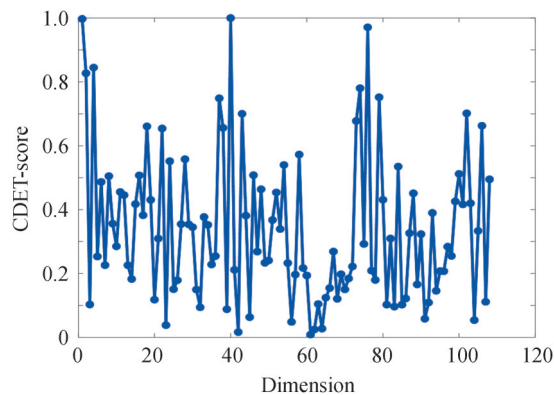


图3 CDET评分表

Fig. 3 CDET-score chart

78维的评分结果最高,根据评分图可以选择高评分维度,通过删减低评分维度即冗余向量进行特征降维优化。为了便于观察并选择合理降维维度,设置了阈值 N ,不同阈值下的维度分布如表2,随着阈值的增大,高分维度相应减少。在SVM分类实验中对于降维维度的选择根据不同阈值下的特征维度进行分类实验。

表2 不同阈值下特征维度

Table 2 Characteristic dimensions at different thresholds

Coefficient N	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.7
Dimension	96	73	61	53	46	37	37	30	24	18	15	10

采用SVM分类器识别三类光纤振动信号的结果如图4。从图4可以看出,使用CDET降维到18维、24维、30维时,分类效果最好,保留了有用特征向量,对所有冗余向量进行删减,此时的特征评分范围为0.45到0.55之间;而PCA在15维、18维、24维时降维效果最好,根据PCA的原理符合低纬度时降维分类效果更好的情况。在10维与15维时,使用PCA降维处理的分类效果要优于使用CDET降维处理的分类效果,因为PCA是通过将原有特征向量集成产生的新的特征向量,而CDET是在原有特征向量上选择高评分特征向量,删减冗余向量,当删减过多时,会造成非冗余特征向量的损失。随着特征维数的增加,使用PCA降维处理的分类效果逐渐下降,甚至低于未经过降维处理的分类效果,而使用CDET降维处理的分类效果仅仅下降了1.5%,保持稳定。

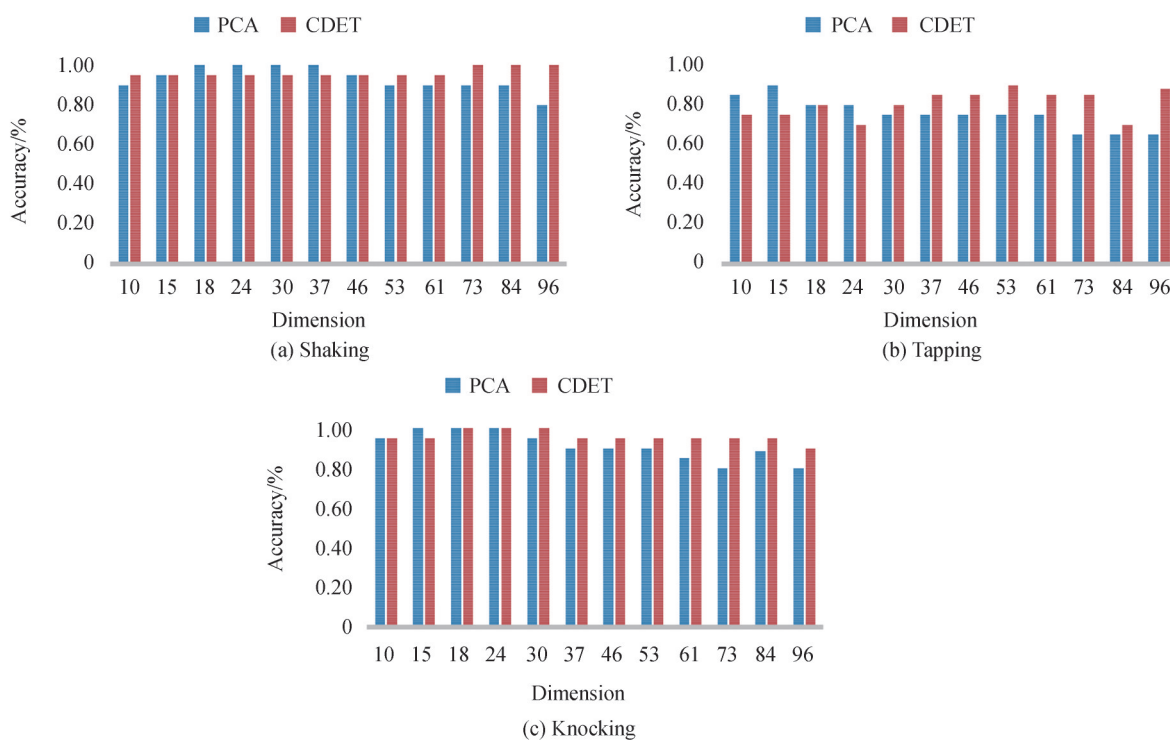


图4 三类信号不同降维维度下的分类准确率

Fig. 4 Classification accuracy under different dimensionality reduction dimensions of three kinds of signals

在传统的MFCC特征提取方法的基础上使用CDET降维处理,识别正确率与未做优化前相比有所提升,平均识别正确率较传统MFCC方法提升了3%左右。

对应图4不同维度下的分类结果,图5显示了不同维度下的时间成本。计算降维优化时间和分类时间。降维优化的时间成本是直接根据训练样本的CDET得分图提取特征向量所需的时间,不同维度下的降维优化时间可以保持在0.5 s左右。没有降维优化的MFCC特征向量需要2.2 s才能通过SVM分类器进行分类。经过降维优化后,SVM分类器对MFCC特征向量分类的时间随着维数的增加而增加,总的的时间消耗也随着维数的增加而增加。在61维中,时间成本等于没有降维优化的时间成本。结合不同维度的分类结果和时间成本,在低维度上可以提高分类结果,降低时间成本。

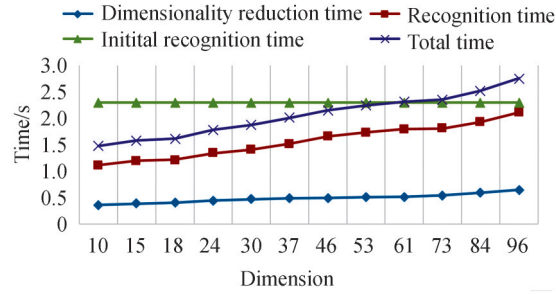
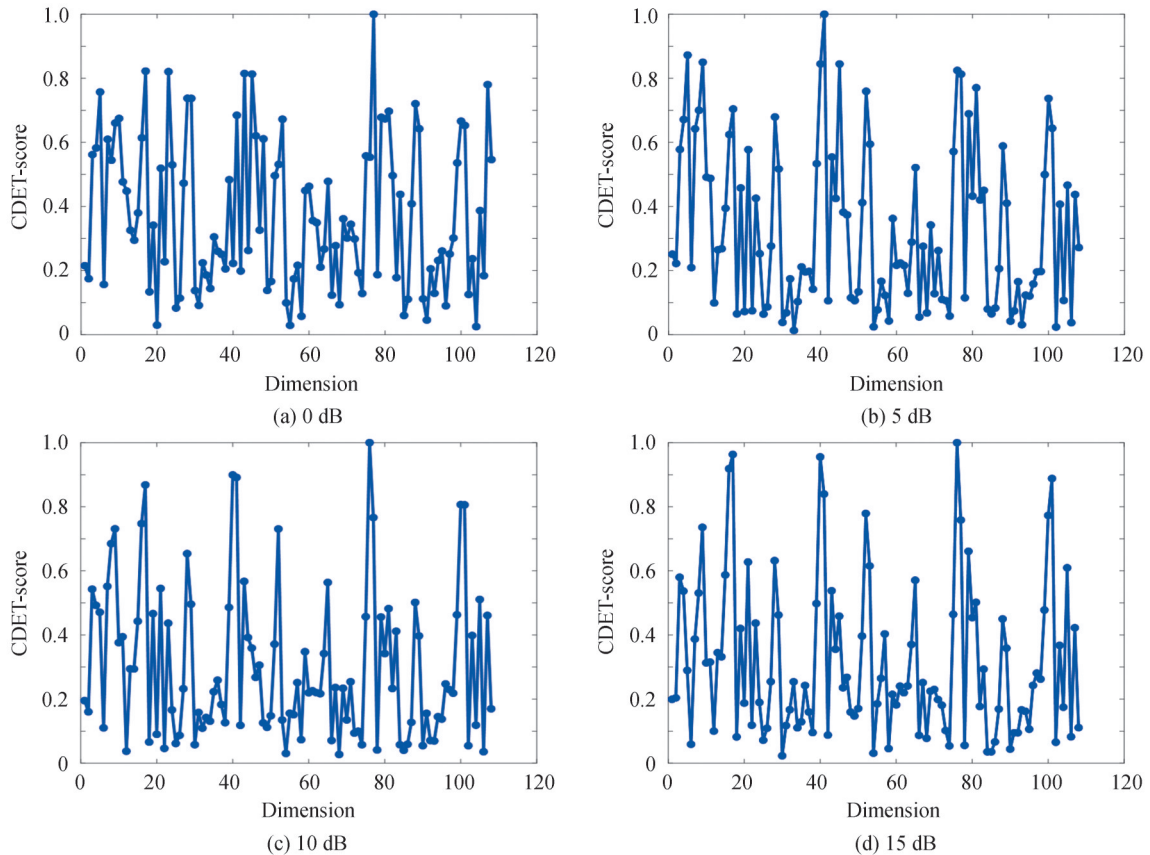


图5 时间消耗曲线

Fig. 5 Time consumption curve

为了进一步对两种方法做出比较,通过验证不同信噪比下MFCC特征向量的优化效果比较两种方法的抗噪性能。实验中将高斯白噪声以0 dB、5 dB、10 dB、15 dB、20 dB的信噪比叠加到三种光纤振动信号中,对SVM分类器识别正确率进行对比,测量优化方法的抗噪性能。图6为叠加不同信噪比噪声的CDET评分图,随着信噪比减小,噪声比例增大,观察评分图可以发现高评分特征向量减少,特征向量的中间评分值增多,表明特征向量辨识度减小。

叠加噪声后测试的分类正确率结果如表3,参考分类实验的实验结果,选择最优分类结果时的维度进行叠加噪声的实验,不同维度的分类准确率与差值结果如图7。从表3的结果可以发现,在同一信噪比的情况下,随着维数的增多,基于CDET优化的MFCC特征向量通过SVM分类准确率增加,说明在存在噪声的情况下,由于特征辨识度下降,从而需要的特征向量维数增多。从图7的结果可以看出,随着信噪比的降低,传统MFCC特征向量在SVM分类器中的分类结果正确率也随着下降,这是由于噪声的增加掩盖了原有的信号特征;基于PCA和CDET降维的MFCC特征向量在使用SVM分类时,随着信噪比减小,分类结果正确率虽然也随着下降,但是下降幅度不是很大。从图7可以发现在相同信噪比下,使用降维处理的特征向量比传统MFCC特征向量在使用SVM分类时对于分类结果有着提升效果,在18维时,PCA的性能优于CDET的



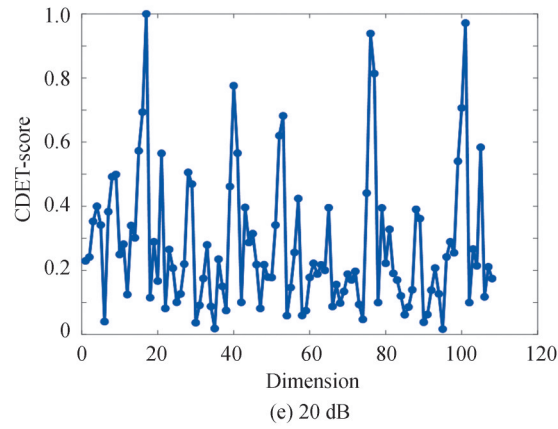
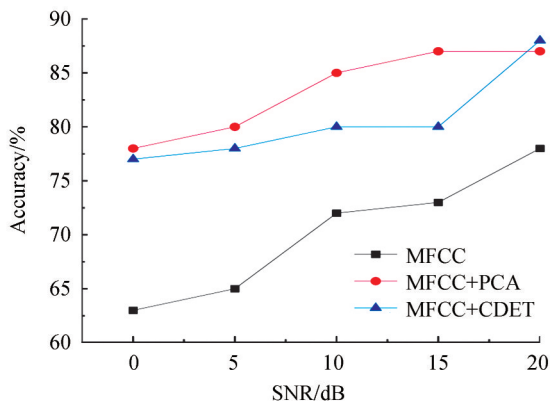


图6 不同信噪比下CDET评分图
Fig.6 CDET-score chart under different signal-to-noise ratio

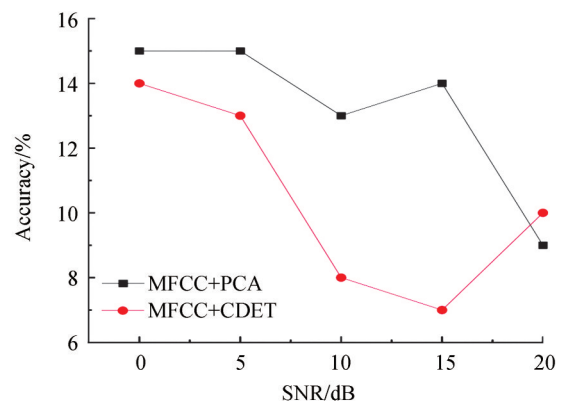
表3 不同信噪比下的分类精度

Table 3 Classification accuracy under different signal-to-noise ratio

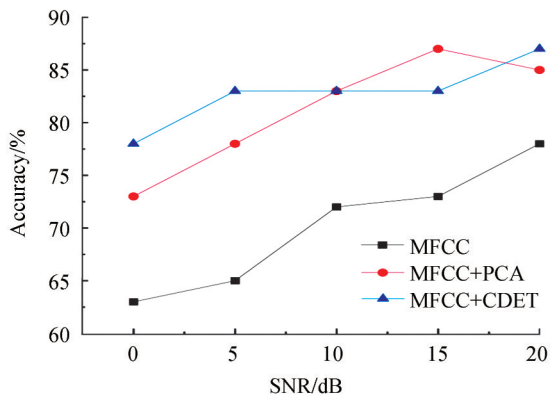
SNR/ dB	MFCC	MFCC+PCA			Proposed algorithm		
		18	24	30	18	24	30
20	78%	87%	85%	82%	88%	87%	90%
15	73%	87%	87%	85%	80%	83%	85%
10	72%	85%	83%	80%	80%	83%	87%
5	65%	80%	78%	75%	78%	83%	87%
0	63%	78%	73%	72%	77%	78%	82%



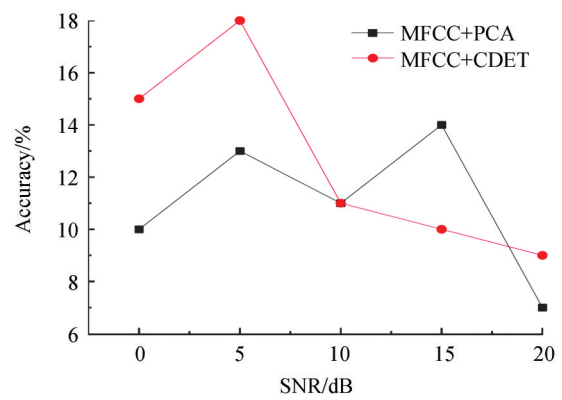
(a) 18 dimensions



(b) Difference under 18 dimensions



(c) 24 dimensions



(d) Difference under 24 dimensions

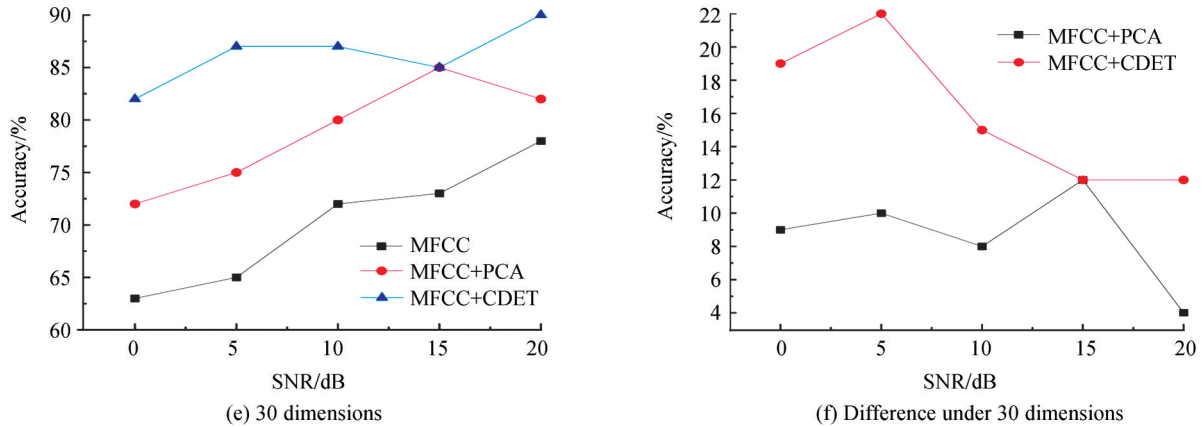


图7 不同信噪比下的准确率与差值图

Fig.7 Accuracy and accuracy difference under different signal-to-noise ratio

性能,根据差值图可以发现,随着信噪比的减小,差值逐渐接近;在24维与30维时CDET的性能要优于PCA的性能,此时根据差值图可以发现,随着信噪比的减小,差值逐渐增大,这表明CDET在低信噪比时有更优性能。补偿距离估计技术是根据特征向量的类间距离与类内距离的比值进行评估,虽然叠加噪声会对信号产生影响,但是通过删减冗余向量而降低叠加噪声产生的影响,基于CDET的特征提取方法会对特征辨识度高的特征向量进行提取,因此基于CDET的特征提取方法具备一定的抗噪性能。

3 结论

本文针对光纤传感振动信号易被噪声淹没,特征弱且难以表征等问题进行研究,提出了一种基于补偿距离估计技术的特征提取方法,以实现光纤传感振动信号的有效识别。该方法在从光纤振动信号中借鉴人类听觉认知机理,提取MFCC特征向量的基础上,并利用CDET对特征进行评估和优化。在光纤传感振动信号识别中,该算法较传统MFCC特征提取方法而言,通过删减冗余信息,突出有效性强、贡献率大的特征向量,解决维数过多导致的复杂运算问题并提高分类准确性。实验结果表明,该方法提高了光纤传感振动信号模式识别的有效性和实时性,对比PCA降维处理方法而言,低维度时性能相同,随着维数的增多,CDET的性能更加稳定。通过本文所提方法能够有效提高光纤振动信号的MFCC特征向量在SVM分类器中的识别正确率,提升分类系统性能,且算法具备一定的抗噪性能。

参考文献

- [1] WANG Yun, JIN Baoquan, WANG Yuncai, et al. Real-time distributed vibration monitoring system using Φ -OTDR[J]. IEEE Sensors Journal, 2017, 17(5): 1333-1341.
- [2] HUANG Wei, PAN Junyu. Application of distributed fiber optic warning system in the refined oil pipeline [J]. Laser Technology, 2016, 40(6): 897-901.
黄伟,潘俊宇.分布式光纤预警系统在成品油管道安防中的应用[J].激光技术, 2016, 40(6): 897-901.
- [3] LI Yunjie, LI Qiang, SHEN Wenyong, et al. Research on the layout and data processing method of distributed optical fiber in shield tunnel monitoring[J]. Journal of Physics: Conference Series, 2020, 1626(1): 012012.
- [4] YANG Yan, LIU Fang, FU Jinghua, et al. Application of bridge continuous shape measurement system based on optical fiber sensing technology in bridge post-seismic detection [C]. International Conference on Optical Fiber Sensors. International Society for Optics and Photonics, 2017: 103236J.
- [5] HUANG Xiangdong, WANG Biyao, LIU Kun, et al. An event recognition scheme aiming to improve both accuracy and efficiency in optical fiber perimeter security system[J]. Journal of Lightwave Technology, 2020, 38(20): 5783-5790.
- [6] LI Junchan, LU Bo, WANG Yu, et al. Distributed optical fiber vibration sensor for the identification of pipeline leakage using relevant vector machine [C]. Optics Frontiers Online 2020: Distributed Optical Fiber Sensing Technology and Applications, 2021: 116070J.
- [7] WANG Liang, GUO Yubin, SUN Tiegang, et al. Signal recognition of the optical fiber vibration sensor based on two-level feature extraction[C]. 2015 8th International Congress on Image and Signal Processing (CISP), 2015: 1484-1488.
- [8] LIU Kun, SUN Zhenshi, JIANG Junfeng, et al. A combined events recognition scheme using hybrid features in distributed optical fiber vibration sensing system[J]. IEEE Access, 2019, (7): 105609-105616.

- [9] SHUAI Shi, WANG Jian, WU Hongyan, et al. A signal recognition method for distributed optical fiber sensor system[J]. Journal of Fudan University(Natural Science), 2018, 57(5):611-618.
帅师,王翦,吴红艳,等.一种分布式光纤传感系统的信号识别方法[J].复旦学报(自然科学版),2018,57(5):611-618.
- [10] XU Chongdong, ZHOU Jing, YING Dongwen, et al. Optimization of the extraction method of MFCC feature vectors for heart sound signals[J]. Journal of Signal Processing, 2019, 35(3):410-418.
许春冬,周静,应冬文,等.心音信号MFCC特征向量提取方法的优化[J].信号处理,2019,35(3):410-418.
- [11] SHARMA S, KUMAR M, DAS P K. A technique for dimension reduction of MFCC spectral features for speech recognition[C].2015 International Conference on Industrial Instrumentation and Control (ICIC), 2015:99-104.
- [12] WINURSITO A, HIDAYAT R, BEJO A. Improvement of MFCC feature extra [C]. IEEE 2018 International Conference on Information and Communications Technology (ICOIACT) 2018:379-383.
- [13] SANGEETHA R. Singer identification using MFCC and CRP features with support vector machines[M]. Computational Intelligence in Pattern Recognition: Advances in Intelligent Systems and Computing, 2020.
- [14] CAO Jiuwen, ZHAO Tuo, WANG Jianzhong, et al. Excavation equipment classification based on improved MFCC features and ELM[J]. Neurocomputing, 2017, 261(25):231-241.
- [15] ZHANG Yuan, ZHAO Lu, TIAN Qian, et al. Optical fiber intrusion signal recognition based on improved mel frequency cepstrum coefficient[C]. 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2018: 1-9.
- [16] CHEN Song, LI Yingchun, HUANG Liangliang, et al. Vehicle identification based on variational mode decomposition in phase sensitive optical time-domain reflectometer[J]. Optical Fiber Technology, 2020, 60(7):102374.

Feature Extraction Algorithm of Optical Fiber Vibration Signal Based on Compensation Distance Estimation

XU Haiyan, KOU Qingkang, XIE Yingjuan, ZHU Jun, LI Min

(College of Internet of Things, Hohai University, Changzhou, Jiangsu 430048, China)

Abstract: The optical fiber sensing system is widely used in many fields, such as long-distance oil and gas pipelines, tunnel safety detection, large structural safety detection, perimeter security detection and so on. Optical fiber sensing signal identification plays a key role in real-time monitoring, abnormal alarm and other aspects. Its working performance directly determines the performance of the real-time, accuracy and stability of the optical fiber sensing detection system. Therefore, fast and accurate identification and classification is of great significance to ensure the safety of various fields and reduce the cost loss caused by equipment damage. To improve the real-time and accuracy of optical fiber vibration signal pattern recognition, a feature extraction algorithm based on compensation distance estimation is proposed. The algorithm draws on the human auditory perception mechanism and extract the MEL frequency cepstrum coefficients from the optical fiber sensing vibration signals. The algorithm uses the compensation distance estimation technology to formulate the feature selection strategy, and finally realizes feature evaluation and optimization. The MFCC feature extraction algorithm can extract the feature of the vibration signal acquired by the optical fiber sensing system, and then identifies the interference signal according to the modal prediction. However, the extracted feature vector by the MFCC feature extraction algorithm has the problems of high dimension and vector redundancy. When these feature vectors are trained and recognized by the classifier, it will increase the time cost and reduce the recognition accuracy. Therefore, effectively reduce the dimension of the MFCC eigenvector is the key to improving the real-time performance and accuracy of optical fiber sensing vibration signals. This paper proposes a feature extraction method based on compensated distance estimation. The CDET algorithm jointly evaluates the intra-class and inter-class discreteness of eigenvectors. The feature evaluation is performed on different dimensions of the feature matrix, and the redundant vector of low score is deleted from the original feature vector matrix, thereby realizing feature dimension reduction. Solve the influence of redundant vectors on classification, avoid the problem of complex operation caused by too many extracted feature dimensions, and improve real-time performance. The steps of the algorithm are to calculate the average distance between samples of the same condition and different conditions, then average the within-class and between-class distances, and then

calculate the within-class variance and between-class variance factors. Then calculate the compensation coefficient between the two variance factors, calculate the ratio of the inter-class distance to the intra-class distance, multiply the compensation coefficient to obtain the distance evaluation standard, and select a better feature dimension according to the distance evaluation standard. The experimental results show that the feature extraction algorithm of vibration signal based on compensation distance estimation technology can effectively reduce the redundant vectors that affect the classification accuracy in optical fiber sensing system. It solves the problems of feature representation and operation complexity for the vibration signal, and further improves the effectiveness and real-time performance of vibration signal pattern recognition of the optical fiber sensor system. Compared with Principal Component Analysis (PCA), our algorithm has the same performance in the low-dimensional case. With the increase of dimension, the performance of the algorithm in this paper is better in recognition accuracy. In terms of anti-noise performance, in the presence of noise, in order to improve the feature identification, the dimension of the feature vector extracted from the MFCC feature increases. The PCA method is difficult to distinguish the features caused by noise, and the algorithm in this paper can reduce the influence of superimposed noise by pruning redundant vectors, and can extract feature vectors with high feature recognition. Therefore, the proposed algorithm also has certain anti-noise performance.

Key words: Optical fiber sensing; Feature extraction; Compensation distance estimation; Feature optimization; Pattern recognition

OCIS Codes: 070.4340; 070.5010; 060.2380