

局部放电脉冲群的实用快速分类技术及应用

司文荣, 李军浩, 杨景刚, 黎大健, 梁永春, 李彦明

(西安交通大学电气工程学院, 710049, 西安)

摘要: 对 10^8 次/s 高速采样率获取的超宽带局放脉冲波形-时间序列, 使用等效时频法提取了局部放电脉冲群的波形特征参数, 并研究了模糊 C 均值聚类算法的特性及其在局部放电脉冲波形特征向量参数处理中的应用. 基于 GIS 的局部放电试验数据处理表明: 由等效时频特征提取和模糊 C 均值聚类分析组成的局部放电脉冲群快速分类技术可以对多个局部放电源构成的脉冲群进行准确分类. 这为研制基于单个脉冲的局部放电超宽带检测与模式识别系统提供了实用的脉冲群快速分类技术.

关键词: 局部放电; 脉冲波形-时间序列; 等效时频; 模糊 C 均值聚类

中图分类号: TM835.4 **文献标志码:** A **文章编号:** 0253-987X(2008)08-1021-05

Fast and Practical Grouping Technique for Partial Discharge Pulse Sequence and Its Application

SI Wenrong, LI Junhao, YANG Jinggang, LI Dajian, LIANG Yongchun, LI Yanming

(School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A novel partial discharge (PD) detection and pattern recognition system are developed with ultra-wideband (UWB) based on a single pulse shape, where a fast grouping technique for PD pulse sequence is adopted. Equivalent time-frequency method (ETFM) is used to extract features for PD pulse shape-time sequence detected by UWB with a sampling rate of 10^8 per second. And the application of fuzzy C-means clustering (FCM) in fast grouping to PD pulse sequence is investigated based on feature parameters of pulse shapes. The analysis result of noised PD data of a gas insulated switch (GIS) shows that the fast grouping technique conducts a good separation for multi-PD pulses sequence which provides a feasible and practical fast grouping technique for PD pulse sequence to develop the multi-PDs detection and pattern recognition system.

Keywords: partial discharge; pulse shape-time sequence; equivalent time-frequency; fuzzy C-means clustering

对电力变压器、气体组合绝缘电器(GIS)等电力设备的绝缘进行在线监测时, 基于局部放电脉冲峰值-时间序列的检测系统对于存在多个局部放电源(包含两个)或异常干扰脉冲源的情况, 其获取的局部放电数据将是随机混叠的峰值-时间序列, 相应的用于局部放电模式识别的放电谱图也是随机混叠的^[1-4]. 在一般情况下, 局部放电检测与模式识别系统的放电指纹样本数据库都是根据单个人工缺陷模

型构造的. 因此, 基于脉冲峰值-时间序列的局部放电检测与模式识别系统对于实际工况中存在的多个局部放电源或异常干扰脉冲源问题时, 可能无法对放电模式做出准确判断.

在电工设备中, 虽然其内部存在的局部放电源是未知的, 但在局部放电源稳定的工况下, 该未知系统是确定的. 因此, 假若使用超宽带检测技术, 检测阻抗(或其他超宽频带传感器)获取的由同一局部放

电源或干扰源产生的脉冲群在时域波形上具有强相似性或高相似度. 如果将传统的脉冲峰值-时间序列检测改为脉冲波形-时间序列检测, 即记录单个脉冲波形及其捕获时间点, 使用某种方法将获取的脉冲群进行快速分类, 再将具有强相似性脉冲组成的各子类脉冲群转换成脉冲的峰值-时间序列, 就可以与由单个人工缺陷模型构造的放电模式进行对比, 与系统数据库中放电模式无关的子脉冲群可判断为噪声或是无效数据, 而对其他子脉冲群, 则做出相关的放电模式判断. 因此, 此系统不仅可以解决峰值-时间序列的混叠问题, 而且能够对存在干扰的多个局部放电源进行放电模式识别.

由上述分析可知, 研制基于单个脉冲的局部放电超宽带检测与模式识别系统的关键是研究出一种实用的脉冲群快速分类技术. 由于局部放电脉冲波形的不确定性, 该分类技术必须为非监督的或无教师的. 本文作者针对超宽带检测技术获取的局部放电脉冲波形特性, 研究了用于局部放电波形的等效时频特征提取和模糊 C 均值聚类分析组成的脉冲群快速分类技术, 并以 GIS 的实测局部放电数据为例, 对该方法进行了试验应用.

1 脉冲波形-时间序列采集及其特征提取

1.1 局部放电波形-时间序列检测系统

作者使用 NI5112 双通道数字示波采集卡(每通道 100 MHz 模拟带宽, 10^8 次/s 实时采样率)结合其他信号处理模块, 研制了基于超宽带检测技术的局部放电脉冲波形-时间序列检测系统^[4]. 系统的局部放电脉冲群检测与快速分类数据处理子界面, 可对实时获取的脉冲群进行特征提取及聚类分析, 从而完成脉冲群的快速分类, 为后续的局部放电模式识别做数据预处理.

1.2 波形特征提取

考虑在实际局部放电数据在线监测中的应用, 脉冲波形的特征提取需要满足如下特性: ①计算的实时快速性; ②与时间记录点无关; ③与脉冲的幅值、极性无关. 据此, 系统对超宽带检测获取的单个脉冲波形的幅值-时间数据的处理如下

$$\left. \begin{aligned} p_j(t_i) &= \left\{ m_0, m_1, \dots, m_i, \dots, m_{n-1} \right. \\ &\quad \left. 0, \Delta t, \dots, \Delta t(i-1), \dots, \Delta t(n-1) \right\} \\ P_j(f_i) &= \left\{ M_0, M_1, \dots, M_i, \dots, M_{n/2-1} \right. \\ &\quad \left. 0, \Delta f, \dots, \Delta f(i-1), \dots, \Delta f(n/2-1) \right\} \end{aligned} \right\} \quad (1)$$

式中: p_j 表示第 j 个脉冲; P_j 为脉冲 p_j 的快速傅里叶变换(FFT); n 表示脉冲波形由 n 个点组成, 脉冲波形的 FFT 频谱则由 $n/2$ 个点组成; m_i 为第 i 个点对应的时域波形幅值; M_i 为第 i 个点的频谱幅值; $\Delta t(i-1)$ 为第 i 个点对应的时间($\Delta t=1/f_s$ 为采样时间间隔, 采样率 $f_s=n\Delta f$); $\Delta f(i-1)$ 为第 i 个点的频率值.

针对脉冲波形特征提取需要满足的特性, 作者对文献[1-3]提出的等效时频法进行改进, 对脉冲 p_j 做如下处理

$$\left. \begin{aligned} T_0^j &= \sum_{i=0}^n t_i p_j(t_i)^2 / \sum_{i=0}^n p_j(t_i)^2 \\ F_0^j &= \sum_{i=0}^{n/2} f_i P_j(f_i)^2 / \sum_{i=0}^{n/2} P_j(f_i)^2 \end{aligned} \right\} \quad (2)$$

$$\left. \begin{aligned} (T_j)^2 &= \sum_{i=0}^n (t_i - T_0^j)^2 p_j(t_i)^2 / \sum_{i=0}^n p_j(t_i)^2 \\ (F_j)^2 &= \sum_{i=0}^{n/2} (f_i - F_0^j)^2 P_j(f_i)^2 / \sum_{i=0}^{n/2} P_j(f_i)^2 \end{aligned} \right\} \quad (3)$$

由式(2)、式(3)对脉冲群中所有的脉冲波形进行特征提取后, 可得到脉冲群在等效时频面内的特征向量分布 (T_j, F_j) , $j=1, 2, \dots, N$, N 为脉冲个数.

1.3 聚类分析前的预处理

脉冲波形-时间序列由上述等效时频法处理后, 脉冲群中的第 j 个脉冲 p_j 对应一个向量 (T_j, F_j) , T_j 和 F_j 分别是脉冲 p_j 在等效时域和等效频域上的特征值. 根据聚类分析的数学模型, 分析脉冲群中 N 个脉冲波形所对应的模式矢量间的相似性, 可以将脉冲群划分为多个不相交的子脉冲群.

在实际应用中, 获得脉冲波形的等效时频特征参数比较复杂, 往往不是 $[0, 1]$ 区间中的数, 因此需要把各个原始特征参数标准化. 以 T_j 处理为例

$$T = \frac{1}{N} \sum_{j=1}^N T_j; \sigma_T^2 = \frac{1}{N-1} \sum_{j=1}^N (T_j - T)^2 \quad (4)$$

然后求各数据的标准化值

$$T'_j = (T_j - T) / \sigma_T \quad (5)$$

这样得到的标准化数据 T'_j 还不一定在 $[0, 1]$ 闭区间内. 为了把标准化数据压缩到 $[0, 1]$ 闭区间, 再采用极值标准化公式处理

$$T_j^{\text{new}} = \frac{T'_j - \min(T'_j)}{\max(T'_j) - \min(T'_j)} \quad (6)$$

2 模糊 C 均值聚类

在基于目标函数的聚类算法中, 模糊 C 均值聚

类(FCM)的理论最为完善,其算法具有高效性和优越性.本文鉴于其是一种非监督聚类算法,将其应用于局部放电脉冲群的快速分类中.

2.1 FCM 算法^[5-9]

FCM算法通过对目标函数的迭代优化来获取数据集的模糊分类,目标函数为样本到聚类中心距离的平方和,然后采用目标函数进行聚类,算法如下.

将样本空间 $P=(p_1, p_2, \dots, p_N)$ 中的 N 个样本分为 b 类, $R=(r_{ij}), i=1, 2, \dots, N, j=1, 2, \dots, b$, 其中 R 为分类结果, r_{ij} 为样本 p_i 对第 j 类的隶属度.

FCM算法的聚类准则是各样本到各聚类中心加权距离的平方和最小,即求 $O_{\text{FCM}}(R, V)$ 的最小值.定义目标函数

$$O_{\text{FCM}}(R, V) = \sum_{i=1}^N \sum_{j=1}^b r_{ij}^m d_{ij}^2 \quad (7)$$

式中: $V=(v_j)$ 表示聚类中心的位置; d_{ij} 为距离测度,表示样本 p_i 和聚类中心 v_j 之间的距离; $m \in (1, \infty)$, 是影响聚类模糊度的加权指数.通过对目标函数的迭代来获取对数据集的模糊分类,迭代函数为

$$\left. \begin{aligned} v_j &= \sum_{i=1}^N r_{ij}^m p_i / \sum_{i=1}^N r_{ij}^m \\ u_{ij} &= \left[\sum_{k=1}^b \left(\frac{p_i - v_k}{p_i - v_j} \right)^{2/(m-1)} \right]^{-1} \quad k=1, 2, \dots, b \end{aligned} \right\} \quad (8)$$

在确定聚类数 b 和加权指数 m 的基础上,初始化隶属度矩阵 $U=(u_{ij})$.由式(7)求目标函数,确定分类结果 $R^{(n)}$, n 为迭代次数,计算 $\|R^{(n)} - R^{(n+1)}\| \leq \varepsilon$, 迭代计算到满足要求为止.

2.2 加权指数 m 的确定

加权指数 m 又称为平滑因子^[5],控制着模式在模糊类间的分享程度.因此,要实现模糊聚类就必须选定一个合适的 m ,然而最佳 m 的选取目前尚缺乏理论指导.文献[6]从聚类有效性的试验研究中得到 m 的最佳选取区间为 $[1.5, 2.5]$,在不做特殊要求时,可取区间中值 $m=2$.文献[7]给出一个经验范围 $1 \leq m \leq 5$,后又从物理解释上得出 $m=2$ 时最有意义.文献[8]从算法收敛性角度着手,得出 m 的最佳取值与样本数 n 有关的结论,建议 m 的取值要大于 $n/(n-2)$.依据上述研究结果,本文取 FCM 算法中的加权指数 $m=2$.

2.3 聚类数 b 对数据集分类的影响及确定

FCM中的聚类数 b 一般是不可预先知道的.图1所示为对一组相同数据在分类数 $b=3$ 和 $b=4$ 时

的 FCM 聚类分析结果.如何得到最佳聚类数 b 是众多学者所关心的问题,目前许多学者提出了各种分类准则^[9],他们根据不同的分类准则,采用自组织迭代技术或是遗传算法得到最优分类,但在计算时,必须事先设定最大分类数后再利用穷举法选定最佳分类数.显然,穷举法在数据点比较少时是可行的,但当数据点很多时穷举几乎是不可能的,况其如何确定最大分类数至今也没有理论依据和实施方法.

针对上述情况以及所开发系统的自身特性(系统处理界面可以查看所有记录脉冲波形以及实时显示脉冲群波形的特征向量分布),本文采用人工干预法对脉冲群聚类数 b 进行确定,方案如下.

(1)对所记录的脉冲波形进行查看.根据同一局部放电源或是干扰脉冲源产生脉冲波形的强相似性,由脉冲波形类数确定聚类数 b .此方法适用于各类脉冲数相差不多的情况,如果其他类型脉冲数或干扰脉冲数非常少,在脉冲数很多的脉冲群中做波形查看是较困难和费时的.

(2)观察等效时频面上的特征向量分布情况.文献[1-6]提出的等效时频法经过大量的试验数据验证表明,该方法对脉冲波形特征提取能够使得脉冲群在等效时频面上的分布为类间距离大、类内距离小.本文改进的等效时频法同样具有此特性,因此在一般情况下,不同类的脉冲群特征向量分布之间会有明显的分界线,据此可确定聚类数 b .

在某些特殊场合下,可结合上述2种方法做出综合评判,从而确定聚类数 b 的值.

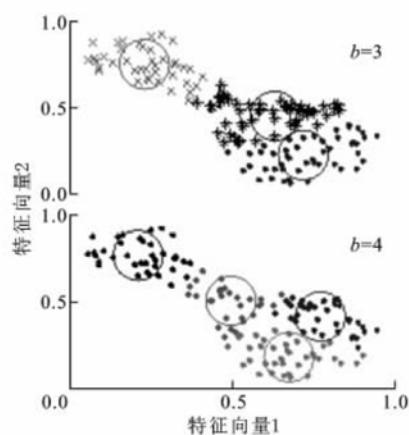


图1 分类数 b 对聚类分析结果的影响

3 试验应用与分析

为了验证由局部放电快速分类技术实现的系统在线处理复杂数据的能力,作者对 220 kV GIS 装置内部设置人工缺陷,并进行局部放电试验,产生具有

相位特征分布的局部放电序列,以区分于其他脉冲源信号。

3.1 悬浮电极

电极布置如图 2a 所示,在 GIS 母线上使用绝缘带悬挂一金属块作为悬浮电极。该缺陷在 0.35 MPa 的 SF_6 气体气压下,产生的单个局部放电脉冲波形如图 2b 所示,脉冲波形记录长度为 $2\ \mu\text{s}$ (系统采样率为 10^8 次/s,采集 200 个点)。由于电极自身结构,产生的单个放电脉冲幅值很大。据此,检测系统脉冲域值设置为 200 mV,从而可以避开其他小幅值干扰脉冲。

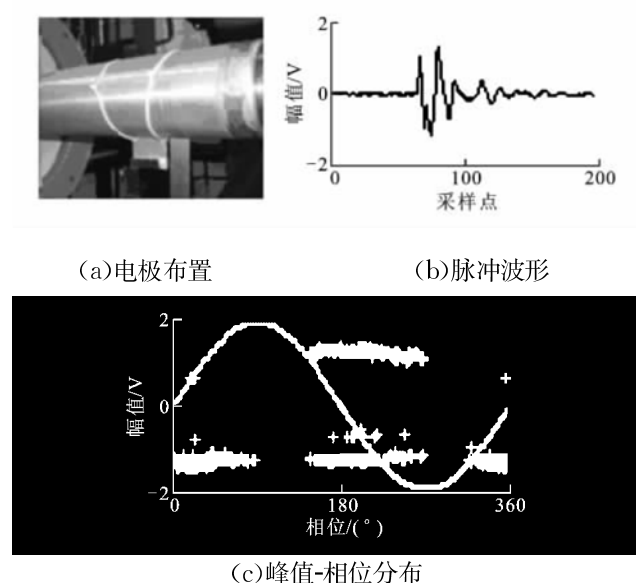


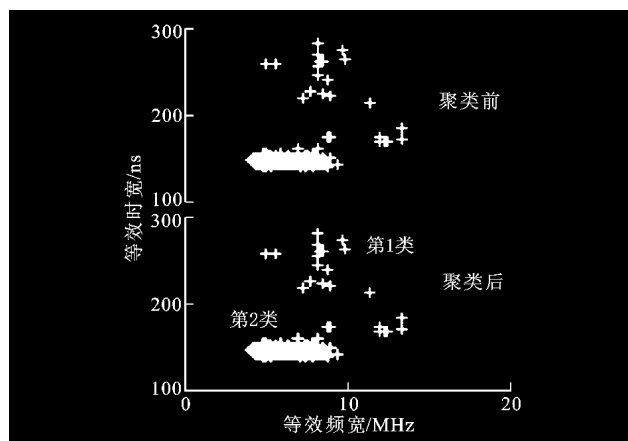
图2 GIS 高压母线的悬浮电极及其检测结果

在某个试验电压下,利用研制的脉冲波形-时间序列,检测系统采集 100 个工频周期获取的脉冲峰值-时间序列,如图 2c 所示。该悬浮电极放电具有明显的相位特征,但可以发现有一些随机分布的低幅值脉冲,其个数较少(在确定聚类数时选用方案 2)。

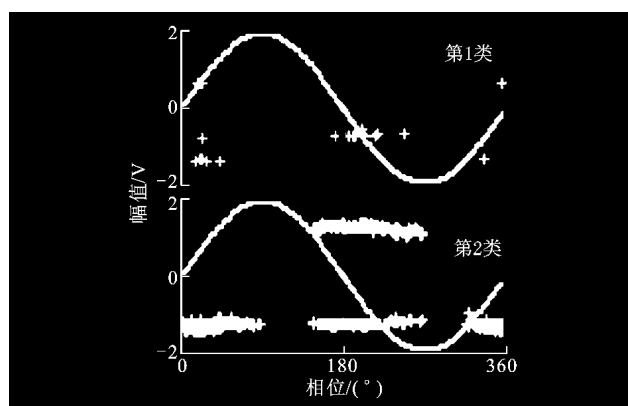
图 2c 所示峰值-相位分布对应脉冲群的脉冲波形-时间序列由等效时频特征提取及 FCM 分析后的分类结果,如图 3 所示。脉冲群对应的波形特征向量在等效时频面上的分布明显为 2 类,由人工干预法方案 2 确定的聚类数为 2。第 2 类子脉冲对应的脉冲波形与试验时单个脉冲触发获取的相同(如图 2b 所示),第 1 类子脉冲对应的脉冲波形如图 4 所示。比较图 4 和图 2b 所示的单个脉冲时域波形,可以判断其并非为干扰脉冲,而是由于检测系统脉冲域值设置较低引起的。

3.2 尖端缺陷

如图 5a 所示,在 GIS 母线上使用螺纹连接一金



(a) 等效时频特征向量分布



(b) 各子类对应的峰值相位分布

图3 悬浮电极放电脉冲群快速分类结果

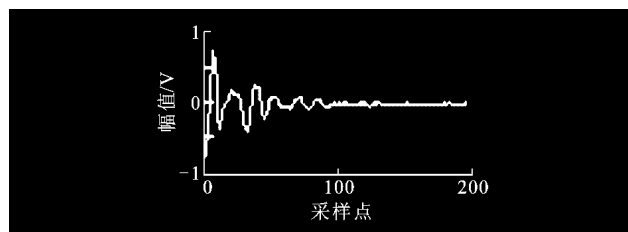


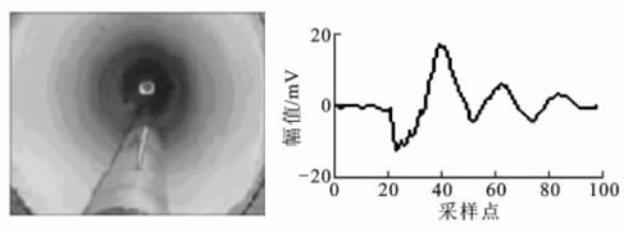
图4 第1类子脉冲群对应的脉冲波形

属突出物以模拟尖端放电。同样,该缺陷在 0.35 MPa 的 SF_6 气体气压下,产生的单个局部放电脉冲波形如图 5b 所示,脉冲波形记录长度设置为 $1\ \mu\text{s}$,脉冲域值设置为 10 mV。

在某个试验电压下,脉冲波形-时间序列检测系统采集 100 个工频周期获取的脉冲峰值-时间序列如图 5c 所示。可以明显看出,此峰值相位分布受到了其他脉冲源的干扰,如果直接生成用于局部放电模式识别(PRPD)放电谱图,并从中提取放电指纹,系统则无法对其放电模式做出准确判断。

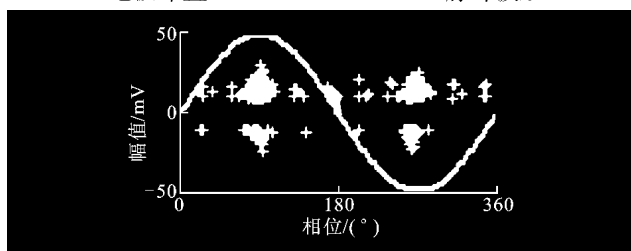
图 5c 所示的脉冲波形-时间序列由等效时频特征提取及 FCM 分析后的分类结果如图 6 所示。脉冲群对应的波形特征向量在等效时频面上的分布明显为 3 类,由人工干预法方案 2 确定聚类数为 3(此

时由于3类脉冲波形数均较多,利用方案1同样可确定聚类数为3)。第2类子脉冲群对应的脉冲波形与试验时单个脉冲触发获取的相同(如图5b所示),



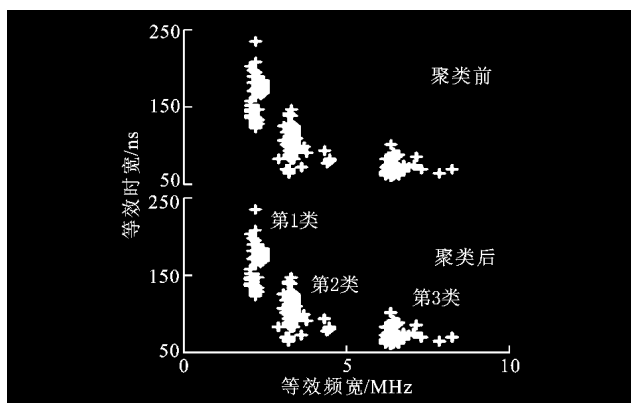
(a)电极布置

(b)脉冲波形

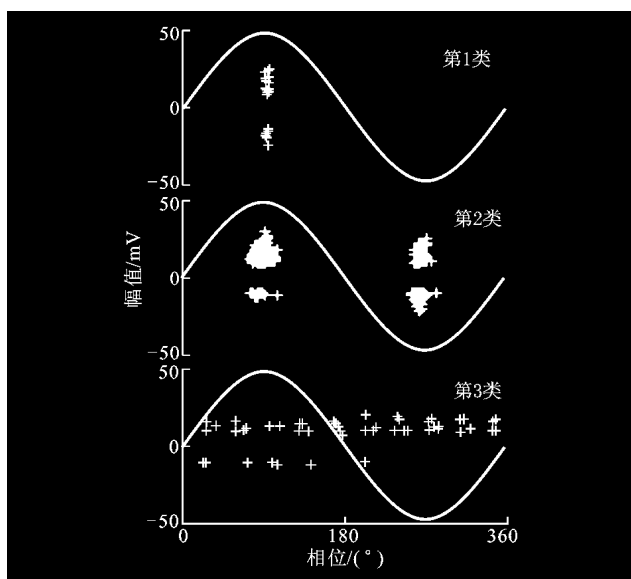


(c)峰值-相位分布

图5 GIS高压母线的尖端缺陷及其检测结果



(a)等效时频特征向量分布



(b)各子类对应的峰值相位分布

图6 尖端缺陷放电脉冲群快速分类结果

其峰值-相位分布具有明显的尖端放电相位分布特征。第1类脉冲波形由于记录时间设置为 $1\mu\text{s}$,图7所示并非其完整时域波形。一般情况下,局部放电脉冲波形持续时间小于 $1\mu\text{s}$,结合其相位分布,可判断为随机干扰信号,后证实为试验大厅内的大灯产生的长达 $10\mu\text{s}$ 的振荡脉冲信号。第3类脉冲波形类似图2a缺陷产生的局部放电脉冲波形,但无明显的相位特征分布,并且在以后的试验中没有发现,同样判断其为外部随机干扰脉冲(快速分类后,可将各子脉冲群的峰值-相位序列生成 PRPD 谱图,分别提取放电指纹后再进行放电模式识别,这里仅是简单的人为经验判断)。

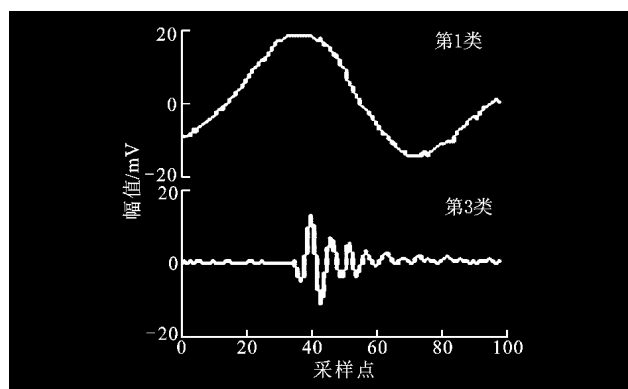


图7 其他子类脉冲群对应的脉冲波形

4 结论

提出利用超宽带检测技术,研制基于单个脉冲的多个局部放电源检测与模式识别系统。利用等效时频法提取脉冲波形特征参数结合模糊C均值聚类,分析实现了研制该系统需要解决的关键技术,即局部放电脉冲群快速分类技术。自行研制的超宽带局部放电脉冲波形-时间序列检测系统在GIS局部放电试验数据分析中的应用,表明了该技术的可行性和实用性。

参考文献:

- [1] CONTIN A, CAVALINI A, MONTANARI G C, et al. Digital detection and fuzzy classification of partial discharge signals[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2002, 9(2): 335-348.
- [2] CAVALINI A, MONTANARI G C, CONTIN A, et al. A new approach to the diagnosis of solid insulation systems based on PD signal inference[J]. IEEE Electrical Insulation Magazine, 2003, 19(2): 23-30.
- [3] CAVALINI A, MONTANARI G C, PULETTI F. A

- fuzzy logic algorithm to detect electrical trees in polymeric insulation systems[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2005, 12(6): 1134-1144.
- [4] 司文荣,李军浩,李彦明,等. 基于宽带检测的局放脉冲波形快速特征提取技术[J]. 电工电能新技术, 2008, 27(2): 11-15.
- SI Wenrong, LI Junhao, LI Yanming, et al. Fast feature extraction technique for PD pulse shape based on wideband detection [J]. Adv Tech of Elec Eng & Energy, 2008, 27(2): 11-15.
- [5] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms[M]. New York, USA: Plenum Press, 1981.
- [6] PAL N R, BEZDEK J C. On clustering for the fuzzy c-means model[J]. IEEE Trans FS, 1995, 3(3): 370-379.
- [7] BEZDEK J C. A physical interpretation of fuzzy ISO-DATA[J]. IEEE Trans SMC, 1976, 6(3): 387-390.
- [8] BEZDEK J C, HATHAWAY R J. Convergence and theory for fuzzy c-means clustering: counterexamples and repairs[J]. IEEE Trans PAMI, 1987, 17(5): 873-877.
- [9] KIM D W, LEE K H, LEE D. On cluster validity index for estimation of optimal number of fuzzy clusters [J]. Pattern Recognition, 2004, 37: 2009-2024.

(编辑 杜秀杰)