

自动跟踪中值滤波方法及应用

王 成 礼

(胜利油田地调指挥部研究所)

摘 要

王成礼: 自动跟踪中值滤波方法及应用, 《石油地球物理勘探》, 24 (3) 1989: 330~338

自动跟踪中值滤波方法, 不同于传统的水平方向中值滤波, 它首先要对地震道的每一点按一定的准则进行方向扫描, 再在所确定的方向上做中值滤波, 对边道要采用加权中值滤波。文中提出了两种加权因子, 对于大跨度中值滤波, 用近似最优快速中值算法计算中值。文中给出了三个应用实例: ①消除随机噪声, 提高记录的信噪比; ②分离 VSP 波场; ③用局部自动跟踪中值滤波来消除不同视速度的高能相干噪声。应用结果表明, 该方法在生产中具有很好的应用前景。

关键词: 中值滤波 自动跟踪中值滤波 波场分离

ABSTRACT

Wang Chengli: Automatically-adjusted median filtering and its application, OGP, 24(3)1989: 330~338

Automatically-adjusted median filtering is different from traditional median filtering for horizontal direction. It first makes directional scanning of each sample of a trace under certain criterion, then performs median filtering in a fixed direction, with weighted median filtering applied to boundary traces. Two kinds of weighting factors are recommended here. The median used in long span median filtering is calculated using an approximate optimum fast median algorithm. Presented here are the application examples in (1) suppressing random noises to raise S/N ratio, (2) separating VSP wave fields and (3) removing some high energy coherent noises with different apparent velocities by making use of partial automatically-adjusted median filtering. The results show that this method has brilliant application prospect.

Key words: median filtering, automatically-adjusted median filtering, wave field separation

问题的提出

地震资料处理的重要任务之一就是压制噪声,提高记录的信噪比。为此,人们研究出了许多方法,例如:频率域滤波,二维滤波和相干加强等。虽然这些方法各有所长,见到了一定的效果,但也有不足之处,如相干加强易产生假象,二维滤波存在严重的混波效应等。因此有必要在方法上作进一步的改进,使之既能提高记录的信噪比又能减少有效信息的损失。

中值滤波能成功地用于垂直地震剖面的波场分离。文献(1)对中值滤波方法进行了系统的理论研究,许多人又将中值滤波用于常规地震剖面处理,也见到了一定的效果。但这种中值滤波方法是针对水平方向同相轴做滤波的,对于陡度比较大的倾斜反射层,效果不够理想(图1b),这样就大大限制了中值滤波的作用。

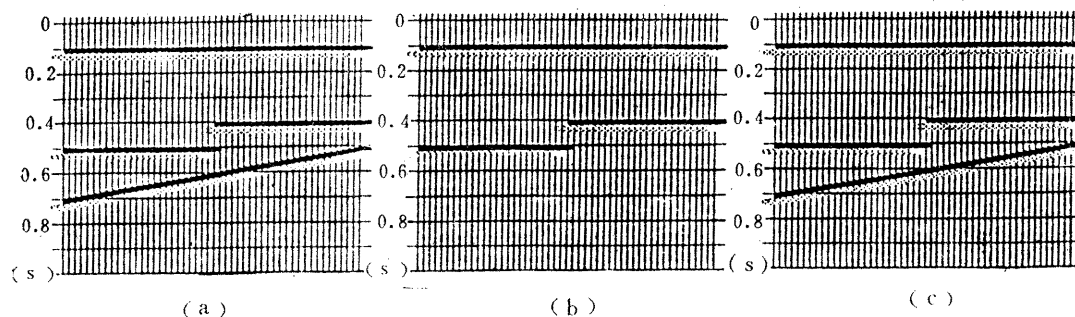


图1 滤波前后波场变化示意图

(a) 理论模型; (b) 经水平中值滤波; (c) 经跟踪中值滤波

本文提出的自动跟踪中值滤波方法,首先要利用有效信号的相干性,确定一个跟踪方向函数,在规定的方向上实现中值滤波。这种方法不受倾角限制,不损失有效信息(如图1c,滤波的结果没有压制倾斜层)。通过实际资料处理证明,该方法用于地面地震资料处理,波组的连续性增强,信噪比提高;用于垂直地震剖面处理,波场分离的效果也很好,此外还可用于剔除高能相干噪声。

当跨度较大时,可采用快速中值算法,所谓中值就是 M 个地震数据(M 为跨度)按递增或递减排序后中间位置上的数。对于排序算法,D.E.Knuth认为:“利用一个不坏的通用排序算法,对 M 个记录排序所需的时间,大约同 $M\log M$ 成正比……,这是至少要花的时间。”本文采用的快速中值算法,是ZHANG排序算法^[2]的改进,其运算量为 $O(M)$,是一种近似最优快速算法。

自动跟踪中值滤波的方法原理

1. 方向斜率的概念

如图2所示,假设 $x_1(i), x_2(i), \dots, x_L(i), \dots, x_M(i)$ 为地震记录道集中的一个相邻道组,其中 M 为奇数, $L =$

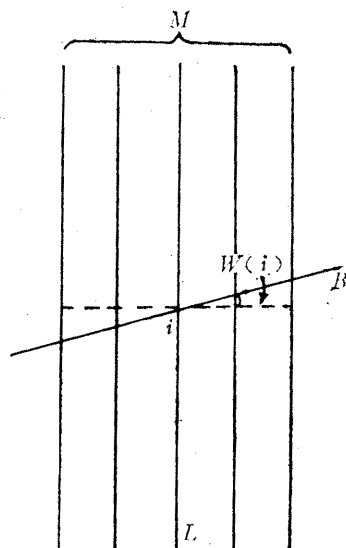


图2 方向斜率定义示意图

$(M+1)/2, 1 \leq i \leq N$ (N 为采样点数)。过第 L 道第 i 个采样点的直线方程可以用下式表示

$$W_m(i) = i + W(i) \frac{2m - M - 1}{2} \quad (m = 1, 2, \dots, M)$$

则我们称 $W(i)$ 为 L 道第 i 个采样的一个方向斜率, 其量纲为毫秒/道。称映射

$$W: \{1, 2, \dots, N\} \longrightarrow \{-N_a, \dots, N_a\}$$

为第 L 道的方向斜率族 (方向函数)。其中 N_a 为 $W(i)$ 的上界。

2. 自动跟踪中值滤波法

上面定义了某个采样点的方向斜率的概念, 因此我们可以对地震道的每一个采样点在某一特定的方向斜率所指的直线方向上做快速中值滤波处理, 以压制随机噪声。

如图 2 所示, 假定直线 A, B 为第 L 道第 i 个采样点方向斜率 $W(i)$ 所规定的直线, 它与道组相交处的数据为 $x_1, x_2, \dots, x_L, \dots, x_M$ 。并设 $\bar{x}_i$ 为数据组 $\{x_i\}$ 的中位数, 即中值。并假设 $\bar{x}$ 为第 L 道第 i 个采样点经滤波后的输出数据, 要求跨度的误差绝对值之和达到最小, 即令

$$Q = \sum_{j=1}^M |\bar{x} - x_j|$$

达最小。则有

$$\frac{\partial Q}{\partial x} = 0$$

由于

$$\sum_{j=1}^M |\bar{x} - x_j| = \sum_{j=1}^M \sqrt{(\bar{x} - x_j)^2}$$

所以

$$\begin{aligned} \frac{\partial Q}{\partial x} &= \frac{\partial}{\partial x} \left(\sum_{j=1}^M \sqrt{(\bar{x} - x_j)^2} \right) \\ &= \sum_{j=1}^M \frac{\bar{x} - x_j}{|\bar{x} - x_j|} \\ &= \sum_{j=1}^M \text{sign}(\bar{x} - x_j) = 0 \end{aligned}$$

式中, sign 表示取符号位。

由上式可知, $\bar{x}$ 必须为数据组 $\{x_i\}$ 经排序后中间位置上的数, 即取中值 $\bar{x} = \bar{x}_i$ 。

3. 边道处理

对于地震道集中的边道, 我们采用了加权中值滤波方法, 即令

$$Q = \sum_{j=0}^M |c_j| \cdot |\bar{x} - x_j|$$

最小。这里, c_i 为加权系数; $\{c_i\} = \{c_1, c_2, \dots, c_M\}$ 为加权因子。

在实际处理中, 针对不同资料我们采用了两种加权因子: ①中间加权因子, 例如 5 道中值滤波, 加权因子为 $\{1, 1, 3, 1, 1\}$, 7 道中值滤波, 加权因子为 $\{1, 1, 1, 3, 1, 1, 1\}$; ②滑动加权因子, 以 7 道中值滤波为例, 当第 1 道时可采用中间加权因子, 第 2 道时加权因子为 $\{0, 0, 1, 1, 1, 0, 0\}$, 第 3 道时加权因子为 $\{0, 1, 1, 1, 1, 1, 0\}$, 当道号大于或等于 4 时, 加权因子为 $\{1, 1, 1, 1, 1, 1, 1\}$ 。

跟踪方向的确定

1. 倾角扫描跟踪方向

假定理想地震剖面不含随机噪声, 不存在静校问题, 且波组可以在某一道组内近似用直线方向来描述。那么, 我们同样可以假设 $x_1(i), x_2(i), \dots, x_L(i), \dots, x_M(i)$ 为一相邻道组, 以 $W(i)$ 表示过第 L 道第 i 个采样点地震层位倾斜变化的方向斜率。这样, 过该点的直线方程为

$$W_m(i) = i + W(i) \frac{2m - M - 1}{2} \quad m = 1, 2, \dots, M$$

并可用下列准则确定 $W(i)$:

①振幅和的绝对值极大准则 即要求使下式的值达最大的 $W(i)$

$$Q = \left| \sum_{m=1}^M x_m(W_m(i)) \right|$$

②方差最小准则 直线在道组内的 M 个地震数据的方差为

$$Q = \sum_{m=1}^M x_m^2(W_m(i)) - \frac{1}{M} \left[\sum_{m=1}^M x_m(W_m(i)) \right]^2$$

在 $W(i)$ 允许变化的范围内, 搜索使 Q 为最小的 $W(i)$ 。

2. 利用纵向渐变特征平滑方向斜率函数

假设 $W_i (1 \leq i \leq N)$ 为第 L 道的方向斜率函数, 则可以用平滑滤波来修正方向斜率函数。滤波器的频率响应为

$$\begin{aligned} H(f) &= \sum_{n=-k}^k \frac{1}{2k+1} \exp(-2\pi i f n) \\ &= \frac{1}{2k+1} \cdot \frac{\sin(2k+1)\pi f}{\sin \pi f} \end{aligned}$$

式中, k 为平滑宽度。

另外, 也可以用纵向约束特性来限制方向斜率函数 $W(i)$, 即规定 $W_1 \leq W(i) \leq W_2$ 。这里, W_1 为最小方向斜率, W_2 为最大方向斜率。

快速计算中值方法

设序列

$$x_1, x_2, \dots, x_L, \dots, x_M \quad (1)$$

为第 L 道第 i 个采样点由某一方向斜率所决定的直线与相邻道组相交点的值, M 为跨度。所谓中值滤波就是将序列(1)的中值作为该点的输出, 而中值则是将(1)排成递增或递减序列后中间位置上的数, 如排成递增的序列, 则为

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(M)} \quad (2)$$

其中, $x_{(i)}$ 为序列(1)中的一个数据。

快速算法就是将序列(1)快速排为序列(2), 并取中值。具体步骤如下:

1. 计算极值

对序列(1)计算极大值和极小值

$$x_{\max} = \max(x_1, x_2, \dots, x_L, \dots, x_M) = x_{(M)}$$

$$x_{\min} = \min(x_1, x_2, \dots, x_L, \dots, x_M) = x_{(1)}$$

若极差 $L = x_{\max} - x_{\min} = 0$, 则(1)不需重新排序, 否则, $L > 0$, 则可分组排序。

2. 数据分组

将数据 x_i 分在第 k_i 组, 这里整型数 k_i 为

$$k_i = \left[(x_i - x_{\min}) \frac{M}{L} + 1 \right] \quad (3)$$

可取值 $1, 2, 3, \dots, M, M+1$, 其中 $[x]$ 表示取不大于 x 的整数值。这样, 按照 k_i 取值的不同, (1)被分成 $M+1$ 组。

实现分组计算过程为: ①计算 $M+1$ 组中各组数记录数据 $\{x_i\}$ 所占的频数 $n_1, n_2, \dots, n_M, n_{M+1}$ 。显然, $n_i \geq 0$, 且 $\sum_{i=1}^{M+1} n_i = M$ 。根据组频数 $n_i (i=1, 2, \dots, M, M+1)$, 确定各组数据占用存储单元的起始地址 $1, n_1+1, n_1+n_2+1, \dots, n_1+n_2+\dots+n_M+1$; ②把(1)中数据, 按(3)中 k_i 的取值和所得的起始地址单元, 分成 $M+1$ 组, 同一组的数据连续存放。

3. 按组排序

对已分好的 $M+1$ 组数据, 按组分别独立地进行排序: ① $n_i \leq 1$ 时, 该组无数据 ($n_i = 0$) 或只有一个数据 ($n_i = 1$), 无需排序; ②当 $n_i > 1$ 时, 采用同时取极大、极小法排序。

在一般情况下, 上述快速分组排序法的运算量为 $O(M)$ 。下面给出均匀分布和任意连续分布两种情况下的证明。

若 η 为 (a, b) 区间上均匀分布的随机变量, 则有

$$x_{\min} = \min(x_1, x_2, \dots, x_M) \geq a$$

$$x_{\max} = \max(x_1, x_2, \dots, x_M) \leq b$$

把 $(x_{\min}, x_{\max})$ 分为 M 个相等的子区间 $[x_{\min} + (i-1)(x_{\max} - x_{\min})/M, x_{\min} + i(x_{\max} - x_{\min})/M] (i=1, 2, 3, \dots, M)$, 由附录2可知, 序列(1)落入每个子区间内的样本平均个数不超过1, 因此利用上述算法进行排序的总运算量为 $O(M)$ 。

若连续分布随机变量 η 具有分布函数 $F(x)$, 由附录3可知, $R = F(\eta)$ 为 $(0, 1)$ 上均匀分布的随机变量。由 x_i 变换到 $R_i = F(x_i)$, 对 R_i 进行排序后再变换回到 $x_{(i)} = F^{-1}$

$(R_{(i)}) (i = 1, 2, \dots, M)$ ，故上述算法的总运算量为 $O(M)$ 。

在实际程序设计中，我们对上述方法进行了一些改进，即在计算过程中，只记住地址，不进行数据存储移动。当计算至中值后，后边的数据不再进行排序。为了进一步提高速度，同组数据也可不再进行排序。对于向量计算机来说，则通过式(3)可快速将组号求出。

程序设计与理论数据试验

1. 时效分析

本程序是在 TIMAP-1 机上实现的。

由于机器内存很小 (64K)，这样的程序运行时间大部分花在与外围设备进行交流方面。为了试验快速算法的效果，在不同的跨度内对 N35 井资料进行了处理，不同算法所用的计算时间情况如图 3 所示。从图中可以看出，当跨度越来越大时，快速算法所用时间基本呈线性变化，而大小比较算法随着跨度的增大，所用时间增加很快。因而，做大跨度中值滤波时快速算法可以节约大量的时间，提高了工效。

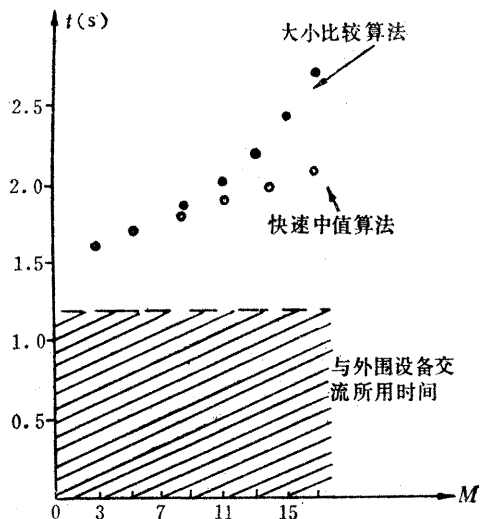


图 3 不同方法时效比较图

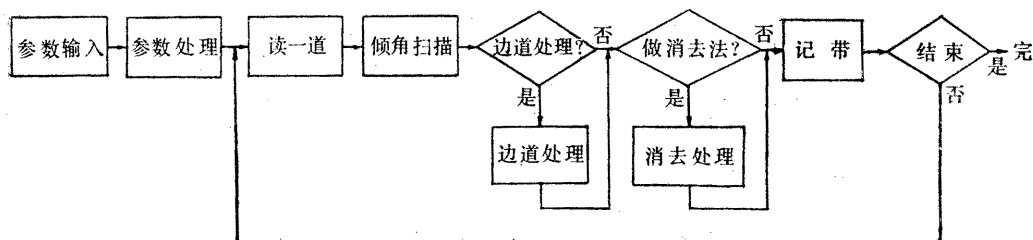


图 4 程序流程图

2. 程序流程图

自动跟踪中值滤波程序流程如图 4 所示。

3. 理论数据试验

为了检验本方法的效果，对理论数据做了分离波场的试验。

从图 5d 的结果可以看出，原来的交叉噪声(图 5b)已基本消除，与原始波场记录(图 5c)比较，断点清楚，斜层没有被削弱，只是在几处有波组变粗的现象，这是由于其中一组噪声与斜层波组视速度差距太小所致。

几个应用特例分析

1. 压制随机干扰，提高记录信噪比

图 6a 是 7371 测线原始地震记录剖面。图 6b 是经自动跟踪中值滤波处理后的剖面。

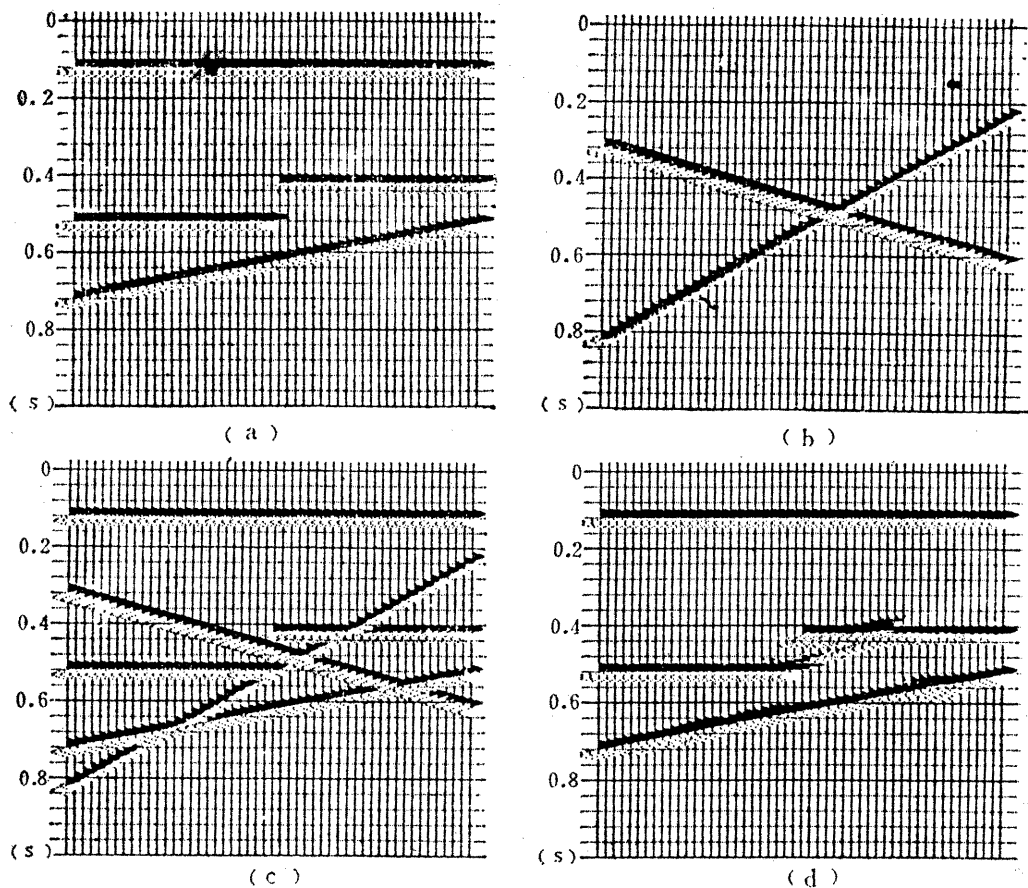


图 5 理论数据试验

(a) 为原始地震波场模型; (b) 为两组不同视速度的交叉噪声; (c) 为 (a) (b) 的合成波场记录, 信噪比为1; (d) 为经11道跟踪中值滤波后的记录

从整体看, 图 6b 的浅、中、深层波组的连续性得到了不同程度的加强, 断层更加清楚, 信噪比有了提高。从细节上分析, 600ms 以上几套波组比较突出, 1 250ms 处强波的断点清晰, 4.0s 以下几套倾斜组得到加强, 波组特征更加明显。

2. 分离 VSP 波场

图 7a 是 N41 井原始 VSP 记录, 图 7b 是经自动跟踪中值滤波处理后所得到的上行波记录剖面。从分离的结果可以看到, 下行波已完全被消除, 上行波被清晰地分离出来, 几套弱波也能明显地被突出出来, 分离的效果十分明显。

3. 利用消去法, 消除高能相干噪声

消去法的基本思路就是利用噪声在能量上的优势, 以及与有效波在视速度上有差异的特点, 首先把噪声突出出来, 然后再将其从原始剖面中消去。图 8a ZHUANG 29 井原始 VSP 剖面, 记录中一套声波噪声严重影响了剖面的质量 (噪声能量强, 视速度为 940m/s~1 000m/s), 图 8b 是经消去法自动跟踪中值滤波消除噪声后得到的剖面。处理后的效果十分明显, 声波噪声已基本被消除, 下行波的波组特征比较清楚, 箭头所示处的上行波已可连续追踪, 原始剖面的质量有了很大改善。

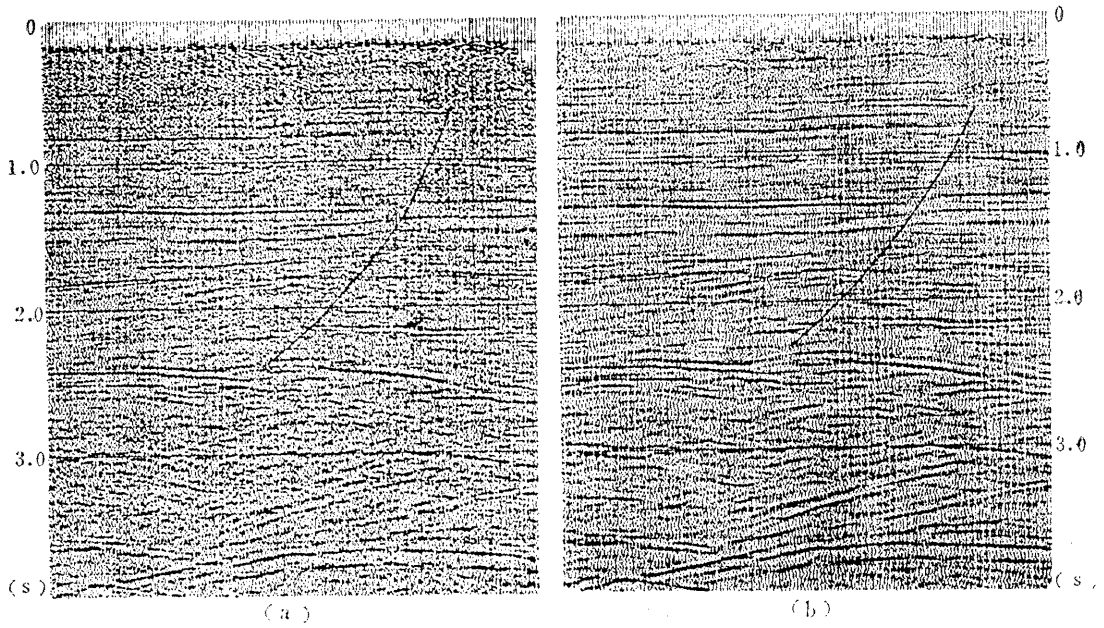


图 6 7371测线原始剖面 (a) 经自动跟踪中值滤波处理后的剖面 (b)

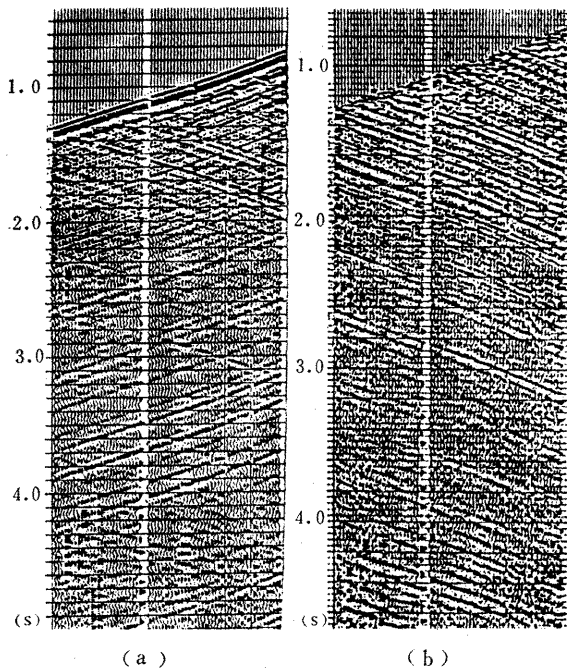


图 7 N41井原始 VSP 记录 (a) 和波场分离后的上行波剖面 (b)

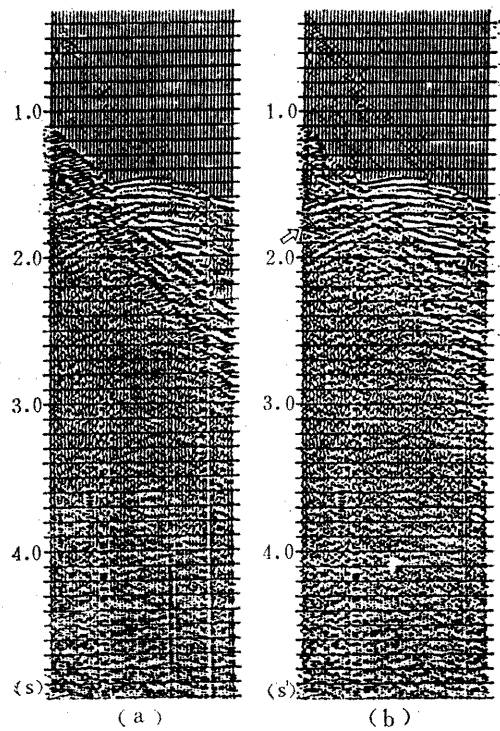


图 8 ZHUANG 29井原始VSP剖面 (a) 经消除噪声处理后的VSP剖面 (b)

结束语

通过理论数据试验与实际地震资料的处理结果说明, 本文所提出的自动跟踪中值滤波方法, 弥补了传统中值滤波方法只适用于水平层的缺陷, 扩大了应用范围。对于边道处理, 采用加权中值滤波可弥补方法本身的缺陷, 当需要做大跨度中值滤波时, 可采用快速算法, 提高工作效率。

实际资料处理结果证明, 该方法可用于低信噪比地震资料的处理及VSP资料处理等。

李梦庚审阅了全文, 并提出了修改意见, 田召才同志帮助分析了效果, 吕功勋、兰永庆同志也给予了帮助, 在此表示感谢。

参考文献

- [1] 李承楚: 关于中值滤波的理论基础, 《石油地球物理勘探》, 21(4)1986: 372~379
- [2] 张建中: 快速分组排序, 《数值计算与计算机应用》, 9(3)1988
- [3] 王伟重等: 含噪地震剖面中反射信号的增强, 《石油物探》, 27(2)1988: 28~39
- [4] 徐常练: 地震数据的中值处理方法, 《物探科技通报》, 5(1)1987

附录 1

对数据记录 (1) 进行线性变换

$$y_i = ax_i + b \quad (i = 1, 2, \dots, M)$$

其中, a, b 为任意常数, 且 $a > 0$, 则有

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}$$

这里

$$x_{\min} = \min(x_1, x_2, \dots, x_M)$$

$$x_{\max} = \max(x_1, x_2, \dots, x_M)$$

$$y_{\min} = \min(y_1, y_2, \dots, y_M)$$

$$y_{\max} = \max(y_1, y_2, \dots, y_M)$$

附录 2

设 R 为 $(0, 1)$ 上均匀分布的随机变量, 数据记录 (1) 为 R 的 M 个相互独立的观测值, 将 $(0, 1)$ 区间分为 M 个等区间 $((i-1)/M, i/M)$ ($i = 1, 2, \dots, M$), 则在任一子区间内, 平均只有一个观测值。

证明: 由附录 1 可知, 对任意区间 (a, b) 上均匀分布的随机变量 η , 因有

$$R = (\eta - a) / (b - a)$$

均匀分布在 $(0, 1)$ 区间上, 故上述结论成立。

附录 3

设连续分布的随机变量 η 具有密度函数 $f(x)$ 和分布函数

$$F(x) = \int_{-\infty}^x f(t) dt$$

则 $R = F(\eta)$ 为 $(0, 1)$ 上均匀分布的随机变量。