JOURNAL OF CHROMATOGRAPHY A

ELSEVIER

Journal of Chromatography A, 855 (1999) 305–315

www.elsevier.com/locate/chroma

Adaptive numerical morphological filter for identifying chromatographic signals

Feng Chen*, Shangxu Hu, Menghuai Yu

Intelligence Information Engineering Institution, Zhejiang University, HangZhou 310027, China

Received 15 March 1999; received in revised form 18 May 1999; accepted 19 May 1999

Abstract

A morphological algorithm has been proposed to filter away impulsive noises confounded in the chromatographic signal. Compared with the conventional median filtering method, the results showed that the proposed method has the advantages of a better filtering effect and less distortion. In particular, the morphological filter with adaptive scale gives very good results.

Keywords: Filters; Morphological signals; Signal processing

1. Introduction

Signals from analytical instruments are often confounded with noises, which may be imperfections in the experimental apparatus, or any of a number of causes which result in random, spurious fluctuations of the signal received at the detector [1–5]. Impulsive and continuous noises are usually found in the chromatographic data. The impulsive noise interferes with the identification of signal peaks. In order to filter the noise from the signal, the median filter was introduced in 1974 by Tukey [7,8] as a smoothing technique in time series analysis. In many practical cases, the median filter is employed for removing impulsive noise [2,3]. Moore and Jorgenson [2] applied the median filter to the extraction of peaks from chromatograms having strongly drifting baselines, and showed it provided a superior performance to the mathematically more complex Butterworth and Chebyshev digital filters. Later, Stone [3] applied median filtering to noise reduction in analytical chemistry. The main advantages of the median filter are its computational simplicity and fast speed. However, it has the disadvantage of having poor filtering results when the size of filter window is narrow, while the signal more easily distorted when the window is set wider.

Morphological filters are one of the most promising approaches to signal processing. They stem from the operations based on the theory of set for image analysis, called mathematical morphology, which was introduced by Matheron [10] and Serra [11]. Morphological filters come from a completely different mathematical background than other digital filters. Their field of application include image processing and analysis [11,12,19], biomedical image processing [11,18], shape recognition [13], nonlinear filtering [14], edge detection [15], etc.

*Corresponding author.

0021-9673/99/$ – see front matter
PII: S0021-9673(99)00663-9

However, they have been rarely used in the field of chemometrics [6] and chromatography [9].

2. Theoretical basis of morphological filters

The elementary morphological operations are restricted in the Euclidean space, which is important for image processing application. The two simplest basic morphological operations are erosion and dilation [16,17]. Other operations can be defined based on erosion and dilation.

Assume that the domain of an m-dimensional (mD) function is a subset of the domain space $D=R^m$ (R is a set of real numbers) or Z^m (Z is a set of integers), depending on the nature of the function $f(x)$. Assume also that the range of $f(x)$ is a subset of the range space $V=R$ or Z, depending on whether the amplitude of $f(x)$ varies continuously or discretely.

Since signals can be represented either by functions or by sets, and the set is the primary notion, the main issue is to represent functions by sets. This is done by following two different but equivalent approaches. That is, an mD function can be represented either by an ensemble of mD sets called its cross section or by a single $(m+1)$D set called its umbra.

For a one-dimensional (1D) function $f(x)$, the set

$$X_t(f)=\{x \in D\colon f(x)=t\},\ t \in V$$

is called the cross section of $f(x)$ at level t and is obtained by thresholding $f(x)$ at level t, see Fig. 1. By considering different levels of t we can associate $f(x)$ with a family of sets, which decrease monotonically as t increases. Since we work in a class of closed sets, all the cross section of $f(x)$ must be closed. The corresponding class of functions with which we will always deal is the class of upper semicontinuous (u.s.c.) functions on D. This correspondence is established because a real-valued function $f(x)$ defined on R^m is u.s.c. if and only if (iff) its cross sections $X_t(f)$ are closed sets in R^m for all $t \in R$. Qualitatively, we can take the u.s.c. function as results from continuous functions after the addition of some positive steps.

One of the most important links between sets and functions is the notion of the umbra, introduced by Sternberg as shown in Fig. 1. If a function $f(x)$ has a domain $D \subset R^n$ or $D \subset Z^n$ and takes values in R:

$$x \in D \rightarrow f(x) \in V$$

its umbra $U(f)$ is a subset of the Cartesian product $D \times V$ consisting of those points of $D \times V$ which occupy the space below the curve of $f(x)$ and down to $-\infty$:

$$U(f)=\{(x,y) \in D \times V\colon f(x) \geq y\}$$

There are morphological filters (MFs) where the input signal is an mD u.s.c. function $f(x)$ and the structure element is an nD u.s.c. function $g(x)$ with a compact region of support (where $n \leq m$). $g(x)$ is called the structure function. The simplest form of such a function is of the form:

$$g(x)=0,\ x \in G \quad (1)$$

where G is its domain, which is a subset of R^m. The umbras of functions with $G=[a,\ b] \in R$ are shown in Fig. 2. The symmetric function $g^s(x)$ with respect to the origin is given by:

$$g^s(x)=g(-x)$$

The definitions of the two basic operations erosion and dilation from function viewpoint are given as

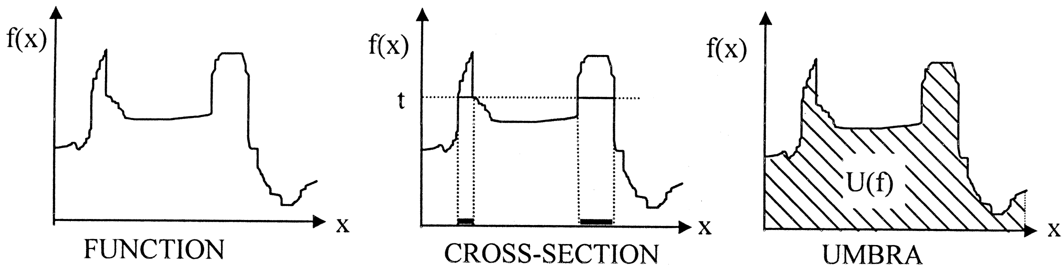

Fig. 1. Function $f(x)$, its cross section $X_t(f)$ at level t, and its umbra $U(f)$.

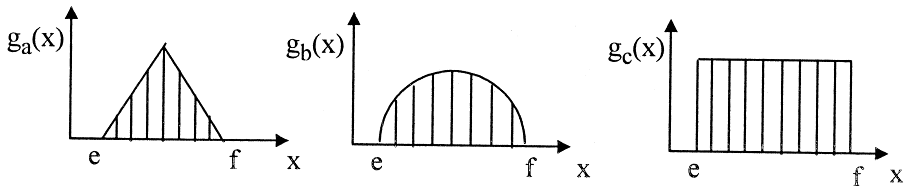

Fig. 2. Examples of the umbras of structured functions.

follows. More details are given by Serra [11] and Maragos [17,18].

Definition 1 (dilation): if D is the domain of $f(x)$, the dilation of function $f(x)$ by $g(x)$ is defined by:

$$[f \oplus g^s](x) = \max_{z \in D, z-x \in G} \{f(x) + g(z-x)\} \quad (2)$$

Definition 2 (erosion): the erosion of function $f(x)$ by $g(x)$ is defined by:

$$[f \Theta g^s](x) = \min_{z \in D, z-x \in G} \{f(x) - g(z-x)\} \quad (3)$$

Its direct geometric interpretation can be illustrated by Fig. 3, where X is an object to be processed and B is called a structure element. If B goes around the inner side of the boundary of X, the locus of B is called the erosion of X. After this operation, X keeps its basic shape while the two peaks disappear.

Definition 3 (opening): the opening of function $f(x)$ by $g(x)$ is defined by:

$$[f \circ g^s](x) = [f \Theta g^s \oplus g](x) \quad (4)$$

Definition 4 (closing): the closing of function $f(x)$ by $g(x)$ is defined by:

$$[f \cdot g^s](x) = [f \oplus g^s \Theta g](x) \quad (5)$$

The morphological operations defined by Eqs. (2) and (3) greatly resemble linear convolution. Definitions (2) and (3) are of great practical importance since they introduce a method of numerical computation of erosion and dilation. Other morphological operations can be defined through erosion and dilation. Since the structure element is important as a interface between the function $f(x)$ and the objective, structure elements with different shapes and sizes were designed according to different purposes. On the other hand, many papers proved that the morphological filter possesses a monotonic property [20]. When different filtering scales are used, different results will be obtained. When the size of structure element increases, impulsive noises wither away gradually while characteristic signals remain unchanged. For chromatographic signals, the opening operation and erosion operation can be employed to remove impulsive noise.

3. Standard data

In order to test the performance of the MF, a median filter is used for comparison, and three sets of standard data have been designed, in which three peaks and three impulsive spikes were generated to

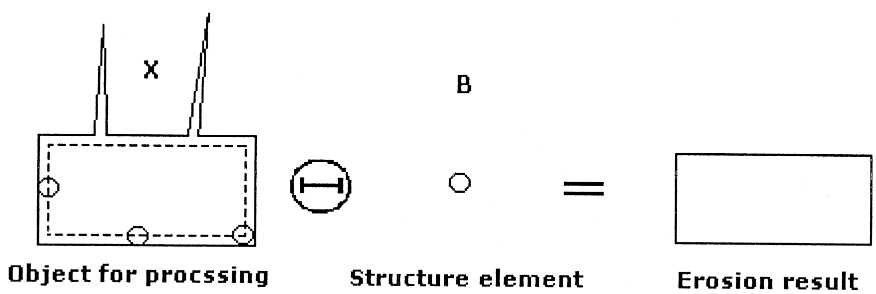

Fig. 3. Scheme of erosion of object X by structure element B.

imitate chromatographic signals and noises, respectively as shown in Fig. 4. In particular, Fig. 4a includes simple Gaussian peaks and overlapping peaks with three spikes which imitate the noises, Fig. 4b includes symmetric triangular peaks with three spikes, and Fig. 4c includes overlapping peaks with three spikes.

Suppose S denotes the intensity data array of chromatographic signals alone, $\hat{S}$ the intensity data array of signals confounded with noise, and L the signal array length. The signal-to-noise ratio (S/N) is defined as:

$$S/N = 10 \log_{10} \frac{\frac{1}{L}\sum_{i=1}^{L} S_i^2}{\frac{1}{L}\sum_{i=1}^{L} (\hat{S}_i - S_i)^2}$$

The S/N of data in Fig. 4a–c is 32.61, 37.30 and 51.42, respectively. On the other hand, if not specially indicated, the data of Fig. 4a are used to study the properties of median filter and how to select the structure element.

4. Performance of the median filter

The median filter has been widely used for removing impulsive noise [2,3], and the performance of the median filter varies with window width. The window width should be set wider in order to remove impulsive noise with wider bottom width. At the same time, the filtering performance is affected by window width. This is shown in Fig. 5, from which we can see the S/N changes with the window width and reaches a maximum when the window width m is about 43. Standard data were filtered by a window of 43 unit width, the results are shown in

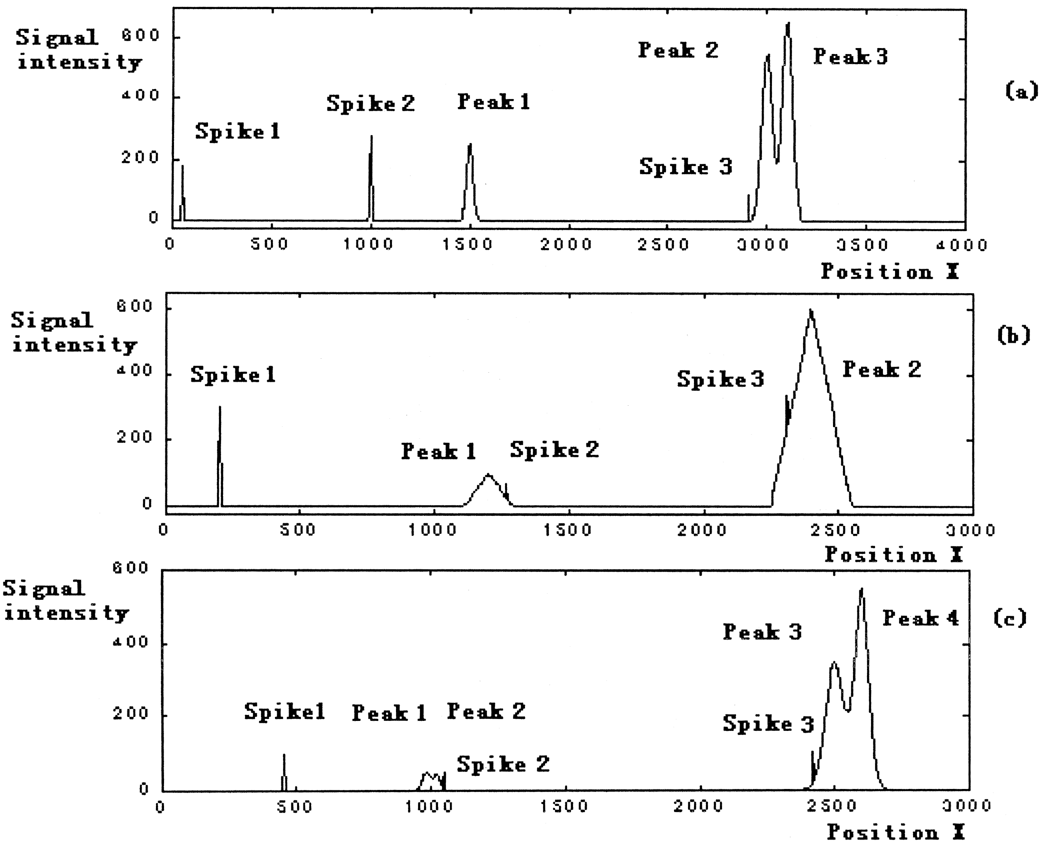

Fig. 4. Three sets of standard data used to compare filtering methods.

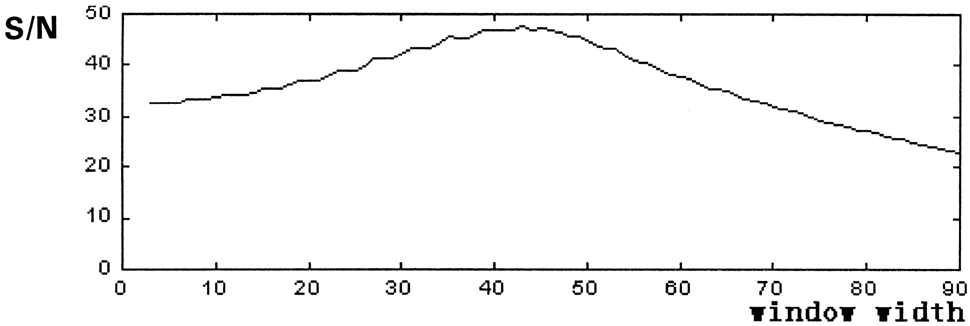

Fig. 5. Performance of median filter affected by window width.

Fig. 6. These results are also be listed in Table 2 for further comparison. It is obvious that spike 2 has not been removed completely because the bottom width is too wide. Furthermore, the shapes of peaks were changed. If window width is increased, the impulsive noises can be removed more easily, while S/N will be declining and the shapes of peaks will be distorted much more seriously, especially at the peak tops.

5. Performance of morphological filters

The performance of the MF has a relationship with structure function. The main procedure of MF operation is as follows:

(1) Select a suitable structure function and suitable size of structure function.

(2) Apply an erosion operation to remove noise from chromatographic data.

(3) Apply a dilation operation to the result of above procedure.

Three structure functions $g_a(x)$, $g_b(x)$, $g_c(x)$ were employed as shown in Fig. 2. In order to remove the impulsive noise of Fig. 4a, define the domain G of $g(x)$, $G \in [1, 27]$. The value of G is defined as the size of the structure function in our study. The structure function $g_a(x)$ of Fig. 2 is a triangle, which is defined as:

$$g_a(x) = \begin{cases} 0.08(x-1), & x \in [1, 14] \\ 0.08(26-x+1), & x \in [14, 27] \end{cases}$$

Structure function $g_b(x)$ is a circle, which is defined as:

$$g_b(x) = \left[13^2 - (x-14)^2\right]/28^2, x \in G$$

Structure function $g_c(x)$ is a straight line, which is defined as:

$$g_c(x) = 0, x \in G$$

The standard data were filtered by using the MF of above three structure functions. The results for quantitative evaluation are given in Table 1. It can be seen that the performance of MF of the structure

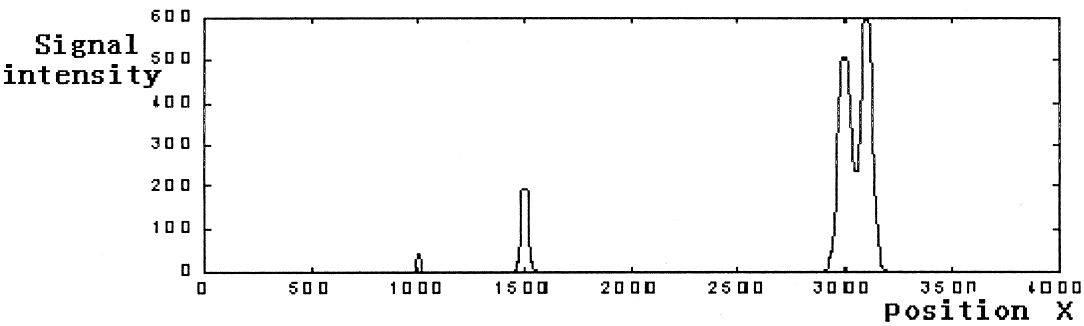

Fig. 6. Results of standard data (see Fig. 4a) processing with median filter ($n=43$).

Table 1
Performance of filtering standard data (see Fig. 4a) with MF of different structure function

Structure function	Criteria								
	S/N	Area error (%)		Deviation in height (%)			Deviation in position		
		Peak 1	Peak 2 + peak 3	Peak 1	Peak 2	Peak 3	Peak 1	Peak 2	Peak 3
Triangle	14.66	69.62	31.77	28.68	11.85	11.42	13	13	13
Circle	14.67	59.02	31.65	28.68	11.85	11.42	13	13	13
Straight line	14.80	55.25	30.61	27.88	11.48	11.11	13	13	13

function $g_c(x)$ is the best. As shown in Fig. 7, three impulsive spikes had been removed completely, however the distortion of peaks occurs. For improving the performance, an adaptive MF was proposed in the following based upon the structure function $g_c(x)$ since it is the best of three structure functions tested.

Suppose that the size of the original structure function is L, when a peak is formed, the size of the structure function decreases gradually from the bottom of peak to the top, and then increases gradually from the top of peak to the bottom. In general, the initial size of the structure function should be greater than the maximum bottom width of the impulsive spikes. The changes of size are taken such that the size of the structure function decreases to a single unit as it goes to the top of peak and returns to the initial size as it reaches the baseline again. The results of filtering the standard data by an adaptive MF with initial size 27 is shown in Fig. 8. The quantitative evaluation of the adaptive morphological filter (AMF) is given in Table 2. After erosion operation, impulsive noises were removed and signal peaks were preserved properly. Generally speaking,

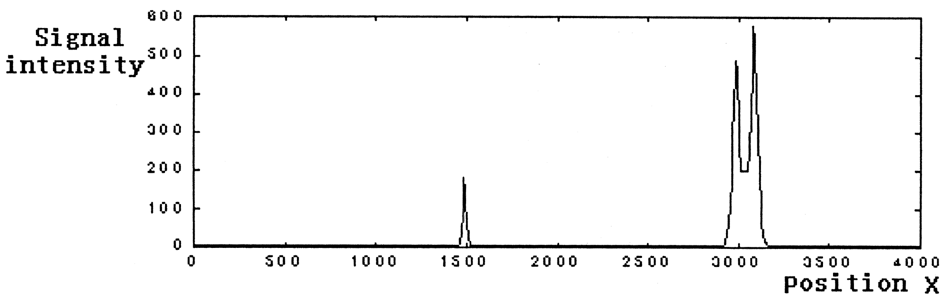

Fig. 7. Results of standard data (see Fig. 4a) processing by MF of structure function $g_c(x)$.

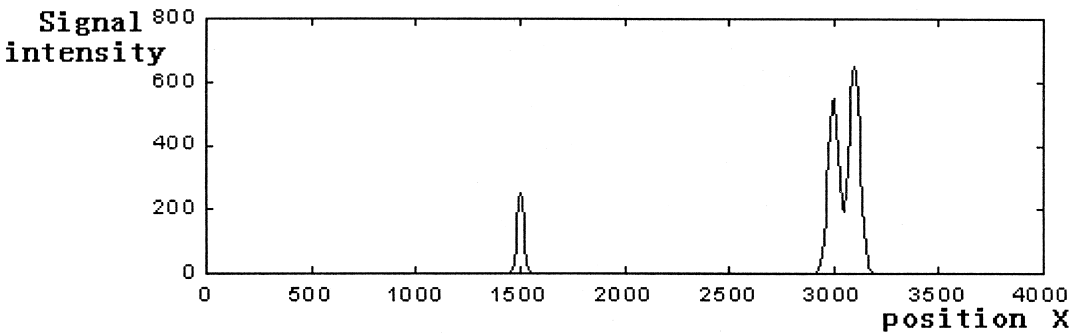

Fig. 8. Results of standard data (see Fig. 4a) processing with adaptive morphological filter ($n \leq 27$).

Table 2
Performance of processing standard data Fig. 4a with median and morphological filters

Filter type	Indication								
	S/N after filter	Area error (%)		Deviation in height (%)			Deviation in position		
		Peak 1	Peak 2+peak 3	Peak 1	Peak 2	Peak 3	Peak 1	Peak 2	Peak 3
Median filter ($m=43$)	47.63	7.75	0.80	21.50	8.25	8.03	−12~+12	−11~+12	−12~+11
AMF filter ($L=36$)	143.99	0	0.004	0	0	0	0	0	0
IAMF filter ($L=18$)	69.15	0	0.0005	0	0	0	0	0	0
IAMF filter ($L=87$)	120.98	0	0.0068	0	0	0	0	0	0
IAMF filter ($L=36$)	181.17	0	0.0005	0	0	0	0	0	0

as the size of structure function becomes large enough, impulsive noise should disappear. On the other hand, a series of tests proved that if the initial size of structure function is smaller than half-peak width of signal peaks, information of signal peaks can be held very well, including height, position, etc.

After the standard data were filtered with the AMF, peaks 1 and 2 had been removed completely, while a residual spike 3 on the overlapping peak still can be seen. The filtering process of spike 3 is shown in Fig. 9 with an enlarged scale. It can be seen from Fig. 9B that the slope of the straight line *cb* is 0. In order to increase the S/N further, an improved adaptive morphological filter (IAMF) was proposed. The IAMF method is to make linear interpolation between points *a* and *b* after processing with an AMF. The quantitative analysis of IAMF is summarized in Table 2. The results of the IAMF are superior in comparison with the other methods.

In order to complete our study, we also used the standard data of Fig. 4b and c to compare the performance of the MF with that of median filter. The quantitative analysis of the results are shown in Tables 3 and 4, respectively. It can be seen that the smaller the peaks and the overlapping peaks are, the more distorted these peaks will be. At the same time, the size of the structure function should be selected carefully according to the size of impulsive noise and the size of peaks. If the size selected is too small, the impulsive noise cannot be removed to meet our demands. If the size selected is too large, the distortion of peaks becomes serious, especially when the chromatographic peaks are relatively small. If the size of peaks is small, the size of the structure function selected should be relatively smaller. According to our experiments, if the size of the structure function is greater than the bottom of impulsive noise, the noise can be removed. On the other hand, if the size of the structure function is smaller than the half-width of the signal peaks, the signal peaks can be retained totally. Altogether, the impulsive noise can be removed effectively, no matter the noise is on the baseline, on single chromatographic peaks or on overlapping chromatographic peaks. The critical elements in designing the MF is that the size of the structure function selected

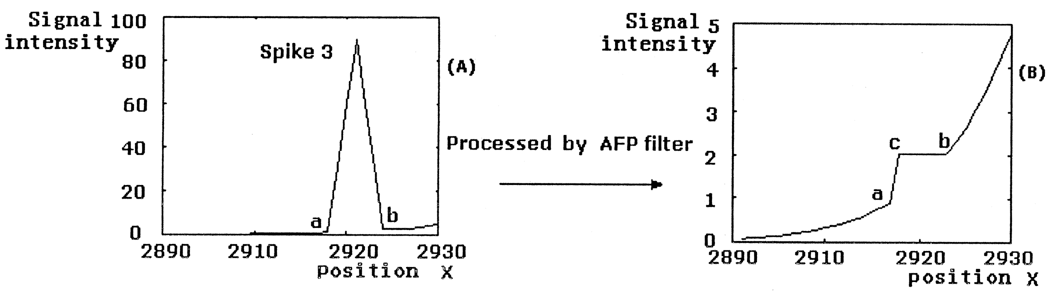

Fig. 9. Filtering result of spike 3 of standard data (see Fig. 4a) processed by AMF filter.

Table 3
Performance of processing standard data Fig. 4b with median and morphological filters

Filter type	Indication						
	S/N after filter	Area error (%)		Deviation in height (%)		Deviation in position	
		Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2
Median filter ($m=43$)	65.42	0.316	0.223	0.12	7.33	$-11\sim+11$	$-11\sim+11$
AMF filter ($L=26$)	91.89	0.185	0.067	0	0	0	0
IAMF filter ($L=15$)	88.33	0	0	0	0	0	0
IAMF filter ($L=57$)	114.47	1.43	0	0	0	0	0
IAMF filter ($L=26$)	741.40	0	0	0	0	0	0

Table 4
Performance of processing standard data Fig. 4c with median and morphological filters

Filter type	Indication										
	S/N after filter	Area error (%)		Deviation in height (%)				Deviation in position			
		Peak 1+peak 2	Peak 3+peak 4	Peak 1	Peak 2	Peak 3	Peak 4	Peak 1	Peak 2	Peak 3	Peak 4
Median filter ($m=43$)	56.33	0.971	0.235	15.46	25.13	4.59	7.43	$-8\sim+9$	$-3\sim+9$	$-10\sim+11$	$-11\sim+10$
AMF filter ($L=22$)	94.46	0.413	0.072	0	0	0	0	0	0	0	0
IAMF filter ($L=14$)	77.709	0.601	0.180	0	0	0	0	0	0	0	0
IAMF filter ($L=42$)	112.42	0.199	0.005	0.117	0.132	0	0	1	1	0	0
IAMF filter ($L=22$)	137.81	0.068	0.005	0	0	0	0	0	0	0	0

should be less than the half-width of signal peaks which we want to retain, and greater than the bottom width of impulsive noise which needs to be removed.

6. Practical data

Fig. 10 includes three sets of practical chromatographic data obtained by the Shimadzu GC-14B chromatography system. Two impulsive noises are included in each set of data, one of the noises (spike 2) is on the chromatographic peak. Fig. 11 shows a chromatographic spectrogram processed by the median filter with a window width of 43. The shape of peak is obviously distorted. The top parts of the peaks became somewhat flatter and the heights of peaks were reduced. Table 5 illustrates the quantitative results of processing the practical data with the median filter and IAMF. The results obtained using the IAMF are satisfactory. Fig. 12 shows the spectrogram of Fig. 10 after filtering by the IAMF. Two impulsive spikes were removed and the information of peaks retained perfectly. For practical chromatographic data, it can be seen that the impulsive noises either on the baseline or on the shoulder peaks can be removed by IAMF. It can be deduced from the filtered spectrogram in Fig. 12 that continuous noises can be reduced.

7. Conclusions

In this paper, the basic definitions and properties of the four simplest morphological filters are given. Standard data confounded with impulsive noises were constructed to examine the properties of digital filters. Several morphological filters with good characteristics were proposed to process practical chromatographic data. The results are satisfactory for removing impulsive noises confounded in chromato-

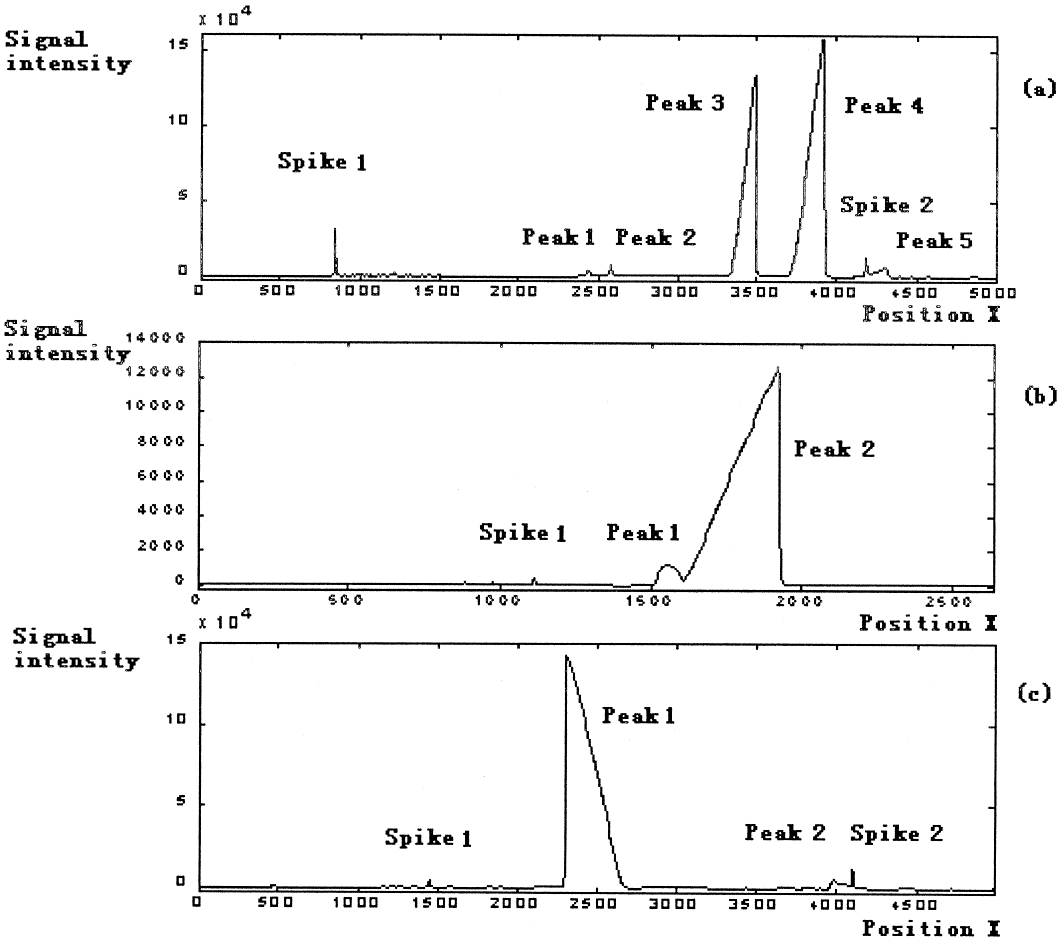

Fig. 10. Practical chromatographic spectrogram.

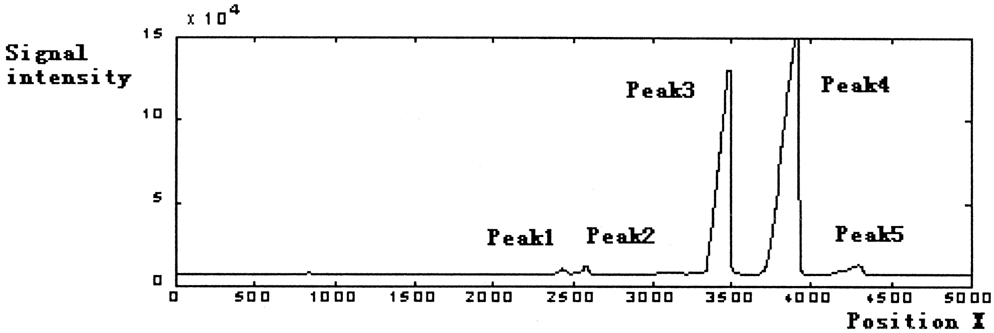

Fig. 11. Practical chromatographic spectrogram filtered with median filter ($m=43$).

Table 5
Performance of filtering practical data with the median filter and IAMF

Filter type		Indication									
		Deviation in height (%)					Deviation in position				
		Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Fig. 10a	Median filter ($m=43$)	12.46	14.42	7.34	9.31	4.24	−10~+10	−12~+9	−18~+4	−17~+6	−17~−5
	IAMF ($L=36$)	0	0	0	0	0	0	0	0	0	0
Fig. 10b	Median filter ($m=43$)	3.88	4.6	–	–	–	−10~+11	−17~+2	–	–	–
	IAMF ($L=30$)	0	0	–	–	–	0	0	–	–	–
Fig. 10c	Median filter ($m=43$)	3.77	4.34	–	–	–	−1~+20	+5~−17	–	–	–
	IAMF ($L=20$)	0	0	–	–	–	0	0	–	–	–

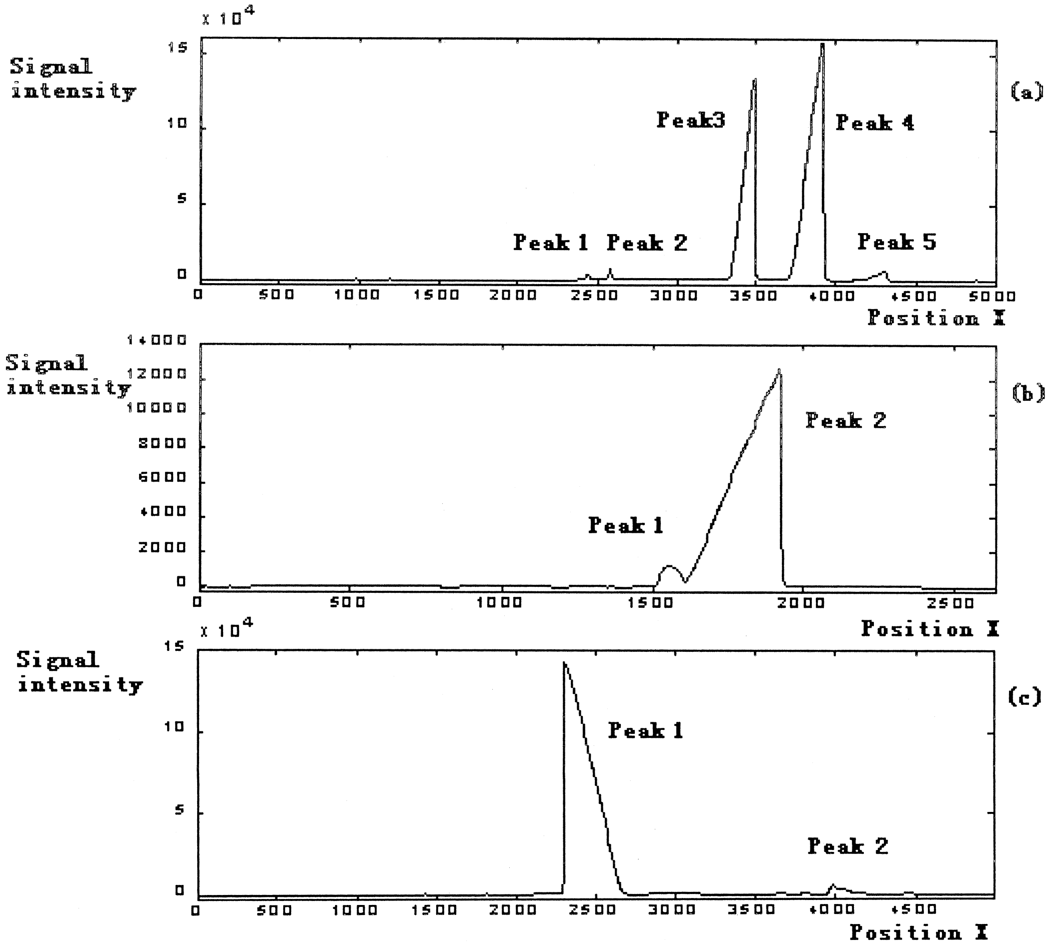

Fig. 12. Practical chromatographic spectrogram filtered with IAMF.

graphic spectrograms. The original information of chromatographic peaks can be reserved completely, and the heights and positions of chromatographic peaks are unchanged. The critical point is that the size of the structure function selected should be less than the half-width of the signal peaks and at the same time greater than the bottom width of impulsive noises.

References

[1] V.J. Barclay, R.F. Bonney, Anal. Chem. 69 (1997) 78–90.
[2] A.W. Moore, J.W. Jorgenson, Anal. Chem. 65 (1993) 168.
[3] D.C. Stone, Can. J. Chem. 73 (1995) 1573–1581.
[4] P.P. Camus, D.J. Larson, Appl. Surf. Sci. 76/77 (1994) 416.
[5] T.A. Nodes, IEEE Trans. Acoustic, Speech, Signal Processing APPS-30 (1982) 739–745.
[6] K.L. Barry, Anal. Chem. 70 (1998) 209–228.
[7] J.W. Tukey, in: Conf. Rec., 1974 EASCON, 1974, p. 673.
[8] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977.
[9] A.E. Gray, H.H. Herbert, G.T. Forge, Anal. Chem. 70 (1998) 321–339.
[10] G. Matheron, Random Sets and Integral Geometry, Wiley, New York, 1975.
[11] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, New York, 1982.
[12] P. Maragos, in: Proceedings of SPIE – The International Society for Optical Engineering, Visual Communication Image Processing, Vol. 707, 1986, pp. 64–74.
[13] T.R. Crimmins, W.M. Brown, IEEE Trans. Aerosp. Electron. Syst. AES-21 (1985) 60–69.
[14] P. Maragos, R.W. Schafer, in: Proc. 1985 IEEE Int. Conf. Acoust., Speech, Signal Process, Tampa, FL, March 1985, pp. 1329–1332.
[15] V. Goetcherian, Pattern Recog. 12 (1980) 7–15.
[16] P. Maragos, IEEE Trans. ASSP 35 (1987) 1153–1169.
[17] P. Maragos, IEEE Trans. ASSP 35 (1987) 1170–1184.
[18] S.R. Sternberg, Computer (Jan. 1983) 22–34.
[19] S.R. Sternberg, Comput. Vision, Graphics Image Process. 35 (1986) 333–355.
[20] D.Y. Wu, M. Songde, Chin. J. Comput. 20 (1997) 989–995.