Anal. Chem. **2008,** *80,* 2664−2671

Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC×GC−MS Data

Dabao Zhang,[†] Xiaodong Huang,[‡] Fred E. Regnier,[‡,§] and Min Zhang*[,†]

Department of Statistics, Department of Chemistry, and Bindley Bioscience Center, Purdue University, West Lafayette, Indiana 47907

A two-dimensional (2-D) correlation optimized warping (COW) algorithm has been developed to align 2-D gas chromatography coupled with time-of-flight mass spectrometry (GC×GC/TOF-MS) data. By partitioning raw chromatographic profiles and warping the grid points simultaneously along the first and second dimensions on the basis of applying a one-dimensional COW algorithm to characteristic vectors, nongrid points can be interpolatively warped. This 2-D algorithm was directly applied to total ion counts (TIC) chromatographic profiles of homogeneous chemical samples, i.e., samples including mostly identical compounds. For heterogeneous chemical samples, the 2-D algorithm is first applied to certain selected ion counts chromatographic profiles, and the resultant warping parameters are then used to warp the corresponding TIC chromatographic profiles. The developed 2-D COW algorithm can also be applied to align other 2-D separation images, e.g., LC×LC data, LC×GC data, GC×GC data, LC×CE data, and CE×CE data.

Current research in biological, environmental, and health sciences is demanding new techniques for the analysis of complex samples, which include thousands of different species and endogenous information. Two-dimensional (2-D) gas chromatography coupled with time-of-flight mass spectrometry (GC×GC/TOF-MS) is a powerful tool for analyzing complex samples and quantifying the underlying compounds with two different stationary phases performing 2-D separations. Small portions of the effluent from a first-dimension column (typically nonpolar) are continuously trapped and released via a modulator onto a second chromatographic dimension (typically more polar) for further separation on the basis of a different separation mechanism. Improved resolution and an order of magnitude increase in peak capacity can be achieved relative to one-dimensional (1-D) GC. Moreover, it provides higher sensitivity and wider dynamic range than conventional 1-D GC.[1]

GC×GC/TOF-MS has so far gained considerable attention in analyzing petrochemical products, environmental pollutants, and biological metabolites but produces very complicated data sets. Chemometric methods have been proposed to glean information from these profiles using multivariate tools such as principal component analysis,[2,3] hierarchical cluster analysis,[4] partial least-squares discriminant analysis,[5] and parallel factor analysis.[2] However, successful application of these multivariate approaches requires that the experimental data from GC×GC/TOF-MS should be preprocessed to be of good quality. In particular, the retention times in separate columns of the GC×GC should ensure high repeatability. One important goal of preprocessing GC×GC/TOF-MS data is to correct retention time shifts, which are usually caused by several uncontrollable factors such as the fluctuation of pressure and temperature, sample matrix effects, and stationary-phase degradation.

It is preferrable to use the entire chromatographic data in chemometric analyses, because reduced peak data usually result in loss of information and confound the problem of profiling the chromatograms.[6,7] Therefore, several 1-D alignment algorithms have been proposed to correct retention time shifts in the original chromatograms from 1-D GC. Wang and Isenhour[8] applied a dynamic time warping (DTW) algorithm, which was initially developed for aligning spectra in speech recognition,[9,10] to align 1-D chromatograms. Since DTW is sensitive to peak intensities, Nielsen et al.[6] proposed the correlation optimized warping (COW) algorithm to warp, i.e., piecewise stretch and compress, the time axis of targeted profiles and optimize the correlation between the warped and reference chromatographic profiles. Both DTW and COW algorithms are implemented with dynamic programming for the corresponding combinatorial optimization[11] (codes available at www.models.kvl.dk/source/DTW_COW/). Instead of maximizing the correlation by interpolatively stretching and compressing

* To whom correspondence should be addressed. Phone: (765) 496-7921, Fax: (765) 494-0558. E-mail: minzhang@stat.purdue.edu.
† Department of Statistics.
‡ Department of Chemistry.
§ Bindley Bioscience Center.

(1) Dalluge, J.; Beens, J.; Brinkman, U. A. *J. Chromatogr., A* **2003**, *1000*, 69−108.
(2) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 2700−2709.
(3) Pierce, K. M.; Hope, J. L.; Hoggard, J. C.; Synovec, R. E. *Talanta* **2006**, *70*, 797−804.
(4) Weckwerth, W.; Wenzel, K.; Fiehn, O. *Proteomics* **2004**, *4*, 78−83.
(5) Jonsson, P.; Gullberg, J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjoestroem, M.; Moritz, T. *Anal. Chem.* **2004**, *76*, 1738−1745.
(6) Nielsen, N.-P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr., A* **1998**, *805*, 17−35.
(7) Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 3289−3295.
(8) Wang, C. P.; Isenhour, T. L. *Anal. Chem.* **1987**, *59*, 649−654.
(9) Itakura, F. *IEEE Trans. ASSP* **1975**, *AS23*, 67−72.
(10) Sakoe, H.; Chiba, S. *IEEE Trans. ASSP* **1978**, *26*, 43−49.
(11) Tomasi, G.; van den Berg, F.; Andersson, C. *J. Chemom.* **2004**, *18*, 231−241.

10.1021/ac7024317 CCC: $40.75
Published on Web 03/20/2008

Table 1. Chemical Standards Used in the Experiment[a]

AA standards		FA standards			OA standards	
name	(t_1, t_2)	name	(t_1, t_2)	m/z	name	(t_1, t_2)
glutamate	(1046, 1.16)	decanoic acid	(710, 1.12)	229	acetic acid	(162, 0.92)
glycine	(622, 1.10)	docosanoic acid	(1326, 1.20)	397	adipic acid	(878, 1.18)
L-alanine	(606, 1.06)	dodecanoic acid	(834, 1.18)	257	benzoic acid	(578, 1.58)
L-aspartic acid	(986, 1.14)	heneicosanoic acid	(1286, 1.18)	383	butanoic acid	(282, 1.06)
L-cystine	(1438, 1.22)	heptadecanoic acid	(1106, 1.14)	327	citric acid	(1210, 1.22)
L-histidine	(1198, 1.32)	heptanoic acid	(498, 1.14)	187	formic acid	(126, 0.78)
L-isoleucine	(734, 1.08)	hexadecanoic acid	(1054, 1.18)	313	fumaric acid	(774, 1.14)
L-leucine	(714, 1.06)	hexanoic acid	(422, 1.10)	173	isobutyric acid	(246, 1.02)
L-lysine	(1098, 1.10)	icosanoic acid	(1242, 1.20)	369	lactic acid	(578, 1.06)
L-methionine	(882, 1.20)	nonadecanoic acid	(1198, 1.16)	355	maleic acid	(746, 1.20)
L-phenylalanine	(950, 1.28)	nonanoic acid	(642, 1.14)	215	malic acid	(962, 1.18)
L-proline	(758, 1.14)	octadecanoic acid	(1154, 1.18)	341	malonic acid	(678, 1.14)
L-serine	(894, 1.08)	octanoic acid	(570, 1.12)	201	oxalic acid	(614, 1.16)
L-threonine	(910, 1.04)	pentadecanoic acid	(1002, 1.64)	299	succinic acid	(754, 1.18)
L-tyrosine	(1218, 1.22)	tricosanoic acid	(1366, 1.22)	411	tartaric acid	(1118, 1.16)
L-valine	(686, 1.08)	tridecanoic acid	(890, 1.16)	271		
		tetracosanoic acid	(1406, 1.16)	425		
		Tetradecanoic acid	(950, 1.70)	285		
		Undecanoic acid	(774, 1.10)	243		

[a] The retention times along the first and second columns are designated as t_1 and t_2, respectively. For the FA standards, the mass-to-charge Ratios of the $[M - 57]^+$ ions are listed under the column m/z.

local regions, Pierce et al.[12] proposed a piecewise alignment algorithm to maximize the correlation for simple scalar shifts of local regions.

Lack of alignment algorithms to correct 2-D retention time shifts in the entire chromatographic data is a substantial limitation in the chemometric analyses of GC×GC/TOF-MS data. Fraga et al.[13] extended the work by Prazen et al.[14] and developed a rank-based algorithm for 2-D retention time alignment. This algorithm aligns small regions of interest by estimating the rank of a subregion of the entire chromatographic data. Mispelaar et al.[15] also developed a correlation-optimized shifting algorithm to align local regions of GC×GC chromatograms, using optimal correction for each subregion of interest on the basis of shifting the sample subregion around on a predefined grid and maximizing its correlation to a standard target subregion. Aiming to correct the entire chromatogram and preserve the separation information in both dimensions, Pierce et al.[16] proposed a new indexing scheme to extend their earlier 1-D piecewise alignment algorithm[12] for 2-D GC data.

While the original piecewise alignment algorithm of Pierce et al.[12] only allowed simple scalar shifts of local regions, the COW algorithm interpolatively stretches and compresses local regions to maximize the correlation between the warped and reference chromatographic profiles. Hence, the COW algorithm is conceptually more powerful and flexible in correcting retention time shifts. In this paper, a general framework to develop a 2-D COW algorithm for warping GC×GC data is presented. Raw chromatographic profiles are first partitioned, and then time warping of the grid points is applied simultaneously along the first and second columns using 1-D COW algorithms. With the shifted grid points, it is possible to interpolatively warp nongrid points.

(12) Pierce, K. M.; Hope, J. L.; Johnson, K. J.; Wright, B. W.; Synovec, R. E. *J. Chromatogr., A* **2005**, *1096*, 101−110.
(13) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **2001**, *73*, 5833−5840.
(14) Prazen, B. J.; Synovec, R. E.; Kowalski, B. R. *Anal. Chem.* **1998**, *70*, 218−225.
(15) van Mispelaar, V. G.; Tas, A. C.; Smilde, A. K.; Schoenmakers, P. J.; van Asten, A. C. *J. Chromatogr., A* **2003**, *1019*, 15−29.
(16) Pierce, K. M.; Wood, L. F.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2005**, *77*, 7735−7743.

EXPERIMENTAL SECTION

Materials and Reagents. Amino acid (AA; Catalog No. AA518), fatty acid (FA; Catalog No. 0C9-1KT and EC10-1KT), and organic acid (OA; Catalog No. 47264) standards along with anhydrous pyridine were obtained from Sigma-Aldrich (St. Louis, MO). All AA standards were at a concentration of 2.5 μmol/mL in 0.1 N HCl except L-cystine, which was used at 1.25 μmol/mL. The FA and OA mixtures were prepared by mixing the 0.5 mg/mL FA standards and 0.5 mg/mL OA standards in pyridine, respectively. The derivatization reagent (*N*-methyl-*N*-*tert*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) was obtained from Regis Technologies (Morton Grove, IL). The different standards used are listed in Table 1.

Derivatization of Standard Samples. A 200-μL AA mixture was dried with nitrogen flow and redissolved in 100 μL of pyridine before derivatization. Two types of test mixtures were prepared. A FA + AA mixture was generated by mixing a 10-μL AA mixture and a 10-μL FA mixture. A FA + OA mixture was generated by mixing a 10-μL FA mixture and a 10-μL OA mixture. For each of the FA + AA and FA + OA mixtures, 11 test samples were derivatized with 20 μL of MTBSTFA for 30 min at 60 °C.

Derivatization of Serum Samples. A 100-μL human serum sample was mixed with 400 μL of solvent (water/methanol/chloroform, 2:5:2, v/v) for removal of proteins. After using a sample sonicated for 10 min and sitting at room temperature for 1 h, mixtures were centrifuged at 16 000 rpm to form a pellet of proteins. The liquid phase was collected and evaporated to dryness with a SpeedVac and then redissolved in 100 μL of pyridine. FA standards were spiked in 20 μL of the serum extract. The mixture was reacted with 10 μL of ethoxyamine hydrochloride solution at 50 mg/mL for 30 min at 60 °C and, subsequently, derivatized with

20 μL of MTBSTFA for 1 h at 60 °C. The same process was repeated 5 months later to derivatize another mixture. FA standards were again spiked in human serum extracts but at a slightly different concentration.

GC×GC/TOF-MS Analysis. Analyses of the derivatized standard mixtures and the extracted serum samples were performed using a Leco Pegasus 4D GC×GC/TOF-MS instrument (Leco Corp., St. Joseph, MI) equipped with a cryogenic modulator. The GC portion of this instrument is an Agilent 6890 gas chromatograph (Agilent Technologies, Palo Alto, CA) and the injector is a CTC Combi PAL autosampler (CTC Analytics, Zwingen, Switzerland). The first-dimension chromatographic column was a 10-m DB-5 capillary column with an internal diameter of 180 μm and a stationary-phase film thickness of 0.18 μm. The second-dimension chromatographic column was a 1-m DB-17 capillary column with an internal diameter of 100 μm and a film thickness of 0.1 μm. High-purity helium was used as the carrier gas at a flow rate of 1.0 mL/min. After applying a 110-s solvent delay for test mixtures and 250-s solvent delay for serum samples, the mass spectrometer started to detect signals. The first-dimension column oven temperature began at 50 °C with a hold time of 0.2 min, was then programmed to 300 °C at a rate of 10 °C/min, and held at this temperature for 2 min. The second-dimension column oven temperature was 20 °C higher than the corresponding first-dimension column oven, but was temperately programmed at the same rate and hold time. The second-dimension separation time was set for 4 s. The 2-μL derivatization solutions were injected in the split mode with a split ratio of 50:1. The inlet and transfer line temperatures were set at 280 °C. The ion source was held at 200 °C. The detector voltage was set at 1600 V and the filament bias at −70 V. Mass spectra were collected from 50 to 800 m/z at 50 spectra/s. The Leco ChromaTOF (v2.32) software was equipped with the NIST MS database (NIST MS Search 2.0, NIST/EPA/NIH Mass Spectral Library; NIST 2002). This software was used for data processing and peak matching. The proposed 2-D COW algorithm is implemented in Matlab, with the WarpingTB package by Tomasi et al.[11] for its 1-D COW algorithm. Under this implementation, each of the alignments using an Intel Pentium 4 with CPU 3.80 GHz, took less than 4 s.

Methods. With GC×GC-MS data, it may be of interest to align chromatographic profiles calculated as either the total ion counts (TIC) or the selected ion counts (SIC). Using one profile as a reference, sample profiles were piecewise stretched and compressed on the basis of maximizing the correlations between 1-D characteristic vectors of the reference profiles and warped sample profiles. Specifically, raw chromatographic profiles were first partitioned into pieces, then 1-D characteristic vectors were defined along the first and second dimensions as in (1) and (2), respectively, and finally the retention times of the grid points were simultaneously shifted along the first and second dimensions by applying the 1-D COW algorithm to those characteristic vectors. The nongrid points were interpolatively warped on the basis of the shifted grid points. The 1-D COW algorithm is briefly reviewed below and the 2-D COW algorithm is then presented.

1. 1-D COW Algorithm. Let $X = (X_1, X_2, \cdots, X_{L_x})$ and $Y = (Y_1, Y_2, \cdots, Y_{L_y})$ be a pair of 1-D chromatographic profiles sampled at regularly spaced time points; i.e., each of X_k and Y_k is sampled at the kth time point. It is intended to warp X to be aligned well with Y. The original X is first partitioned into n sections, i.e.

$$X = (X_{(l_0+1):l_1}, X_{(l_1+1):l_2}, \cdots, X_{(l_j+1):l_{j+1}}, \cdots, X_{(l_{n-1}+1):l_n})$$

where $l_0 = 0$, $l_n = L_x$, and each section $X_{(l_j+1):l_{j+1}} = (X_{l_j+1}, X_{l_j+2}, \cdots, X_{l_{j+1}})$ usually has m data points, $j = 0, 1, \cdots, n-1$. Here $l_0, l_1, l_2, \cdots, l_n$ are called nodes, which define the sections. Similarly, Y can also be partitioned into n sections. Each section $X_{(l_j+1):l_{j+1}} = (X_{l_j+1}, X_{l_j+2}, \cdots, X_{l_{j+1}})$ is stretched or compressed, namely, warped, to $\tilde{X}_{(l_j+u_j+1):(l_{j+1}+u_{j+1})} = (\tilde{X}_{l_j+u_j+1}, \tilde{X}_{l_j+u_j+2}, \cdots, \tilde{X}_{l_{j+1}+u_{j+1}})$ by linear interpolation. Each u_j is a shift of l_j and the difference $u_{j+1} - u_j$ defines warping of the $(j+1)$-st section. The resultant vector, $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_{L_y})$, has the same length as Y and is better aligned with Y.

Besides the length m of each section, the 1-D COW algorithm also takes another integer parameter δ, which defines the maximum warping for each section. As the lengths of sections in X and Y differ by $d = L_y/n - m$, each warping $u_{j+1} - u_j$, given the maximum warping parameter δ, varies in the range $(d - \delta, d + \delta)$. An optimal warping is pursued by finding $(u_1, u_2, \cdots, u_n)$, under the above constraints, to maximize the following summation of a series of empirical correlation coefficients

$$\sum_{j=0}^{n-1} \widehat{\mathrm{corr}}(\tilde{X}_{(l_j+u_j+1):(l_{j+1}+u_{j+1})}, Y_{(l_j+u_j+1):(l_{j+1}+u_{j+1})})$$

where $u_0 = 0$, and $u_n = L_y - l_n$. As each u_j is only allowed to take a finite number of values, a backward dynamic programming strategy is developed to solve the above optimization problem by examining all possible combinations of the variables $u_1, u_2, \cdots, u_n$.[6]

2. 2-D COW Algorithm. Here it is assumed that X and Y are 2-D chromatographic profiles, both sampled at regularly spaced time points on the first and second columns. The goal is to warp X and align it with Y by a 2-D COW algorithm. For simplicity, both X and Y are assumed to be $R \times C$ matrices with X_{ij} and Y_{ij} denoting the intensity values at pixel (i, j) of X and Y, respectively. The 2-D COW algorithm, as will be described below, requires two pairs of prior parameters, the section length m_r and maximum warping δ_r along the first column, and the section length m_c and maximum warping δ_c along the second column. Note that the first column time and the second column time are arbitrarily arranged along the row and column of $R \times C$ chromatograms, respectively. This arrangement can also be done the other way.

Given the section lengths m_r and m_c, the chromatograms X and Y are first partitioned into pieces, each represented by a $m_r \times m_c$ matrix; see Figure 1a. Suppose both chromatograms are partitioned into $n_r \times n_c$ pieces with grid nodes $\{i_0, i_1, \cdots, i_{n_r}\}$ along the first column, and $\{j_0, j_1, \cdots, j_{n_c}\}$ along the second column, where $i_0 = j_0 = 1$, $i_{n_r} = R$, and $j_{n_c} = C$. As $\{i_0, i_1, \cdots, i_{n_r}\}$ and $\{j_0, j_1, \cdots, j_{n_c}\}$ are subsets of $\{1, 2, \cdots, R\}$ and $\{1, 2, \cdots, C\}$, respectively, the grid nodes are $\{(i_k, j_l) : k = 0, 1, \cdots, n_r; l = 0, 1, \cdots, n_c\}$.

At each row i_k of X, a new row vector $\tilde{\mathbf{X}}_{i_k} = (\tilde{X}_{i_k1}, \tilde{X}_{i_k2}, \cdots, \tilde{X}_{i_kC})$ is calculated, with the jth component $\tilde{X}_{i_kj}$ calculated through

$$\tilde{X}_{i_kj} = \sum_{i=1}^{m} X_{ij} W\left(\frac{i - i_k}{h}\right) \Big/ \sum_{i=1}^{m} W\left(\frac{i - i_k}{h}\right) \qquad (1)$$

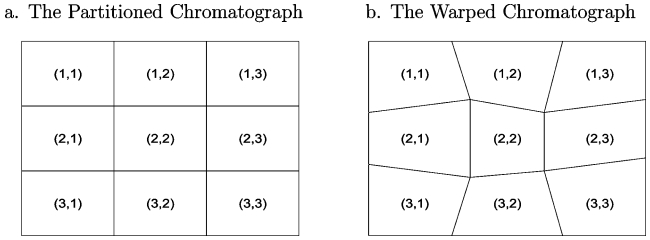

Figure 1. Hypothetical example. The chromatograph is partitioned into patches first (a), then the 2-D COW algorithm is applied to warp the intersection points, i.e., nodes, which essentially help stretching and compressing each patch (b).

Here $W(\cdot)$ is a symmetric kernel density function and $h > 0$ is the bandwidth. In practice, it is preferred to use the Epanechnikov kernel $W(t) = 3/4(1 - t^2)I_{\{|t|<1\}}$, which, as a function of distance to the center, has decreasing weights on a bounded support.[17] Similarly, a new vector $\tilde{\mathbf{Y}}_{i_k} = (\tilde{Y}_{i_k1}, \tilde{Y}_{i_k2}, \cdots, \tilde{Y}_{i_kC})$ is calculated with the j-th component $\tilde{Y}_{i_kj}$ as follows

$$\tilde{Y}_{i_kj} = \sum_{i=1}^{m} Y_{ij} W\left(\frac{i - i_k}{h}\right) \Big/ \sum_{i=1}^{m} W\left(\frac{i - i_k}{h}\right) \qquad (2)$$

By applying the 1-D COW algorithm to align $\tilde{\mathbf{X}}_{i_k}$ to $\tilde{\mathbf{Y}}_{i_k}$, the indices $(j_0, j_1, \cdots, j_{n_c})$ are essentially warped to $(\tilde{j}_{i_k0}, \tilde{j}_{i_k1}, \cdots, \tilde{j}_{i_k,n_c})$. Similarly, for each column j_l, two new column vectors are calculated from the original chromatograms X and Y, respectively. Employing the 1-D COW algorithm to align these two column vectors will warp the indices $(i_0, i_1, \cdots, i_{n_r})$ to $(\tilde{\imath}_{0j_l}, \tilde{\imath}_{1j_l}, \cdots, \tilde{\imath}_{n_rj_l})$.

On the basis of applying the 1-D COW algorithm, each grid node (i_k, j_l) of X is warped to a new location at $(\tilde{\imath}_{i_kj_l}, \tilde{\jmath}_{i_kj_l})$; see Figure 1b. For any pixel (i, j) of X, a pair of k and l can be found such that $i_k \le i \le i_{k+1}$ and $j_l \le j \le j_{l+1}$. The index i can be warped along the first column (see Figure 2a) and the index j can be warped along the second column (see Figure 2b) simultaneously. Specifically, i and j will be warped, respectively, to

$$\begin{aligned}\tilde{\imath} = {} & \tilde{\imath}_{i_kj_l} + (\tilde{\imath}_{i_kj_{l+1}} - \tilde{\imath}_{i_kj_l}) \times \frac{j - j_l}{j_{l+1} - j_l} + (\tilde{\imath}_{i_{k+1}j_l} - \tilde{\imath}_{i_kj_l}) \times \frac{i - i_k}{i_{k+1} - i_k} \\ & + (\tilde{\imath}_{i_{k+1}j_{l+1}} - \tilde{\imath}_{i_{k+1}j_l} - \tilde{\imath}_{i_kj_{l+1}} + \tilde{\imath}_{i_kj_l}) \times \frac{i - i_k}{i_{k+1} - i_k} \times \frac{j - j_l}{j_{l+1} - j_l} \\ \tilde{\jmath} = {} & \tilde{\jmath}_{i_kj_l} + (\tilde{\jmath}_{i_{k+1}j_l} - \tilde{\jmath}_{i_kj_l}) \times \frac{i - i_k}{i_{k+1} - i_k} + (\tilde{\jmath}_{i_kj_{l+1}} - \tilde{\jmath}_{i_kj_l}) \times \frac{j - j_l}{j_{l+1} - j_l} \\ & + (\tilde{\jmath}_{i_{k+1}j_{l+1}} - \tilde{\jmath}_{i_{k+1}j_l} - \tilde{\jmath}_{i_kj_{l+1}} + \tilde{\jmath}_{i_kj_l}) \times \frac{i - i_k}{i_{k+1} - i_k} \times \frac{j - j_l}{j_{l+1} - j_l} \end{aligned} \qquad (3)$$

On the other hand, for each pixel $(\tilde{\imath}, \tilde{\jmath})$ of the warped X, the pixel (i, j) of X can be identified accordingly.

RESULTS AND DISCUSSION

The COW algorithm is powerful in correcting retention time shifts between two similar chromatograms. For similar chemical samples, the generated TIC chromatograms will share common features and differ by very few features, which we call homogeneous chromatograms. However, the TIC chromatograms from different chemical samples may share few common features, which we call heterogeneous chromatograms. Although the COW algorithm can be applied directly to homogeneous TIC chromatograms, it may fail in aligning heterogeneous TIC chromatograms, since the peak corresponding to a chemical component unique in one sample may be matched to a neighboring peak that corresponds to a different chemical component unique in another sample. In the extreme case where chromatograms share no common features, any algorithm would fail in aligning either 1-D or 2-D chromatograms without utilizing chemical component information. In practice, it is feasible to generate SIC chromatograms for certain specific components with known mass-to-charge ratios (m/z values). These SIC chromatographic profiles can be aligned using the COW algorithm and therefore provide warping parameters for aligning TIC chromatographic profiles.

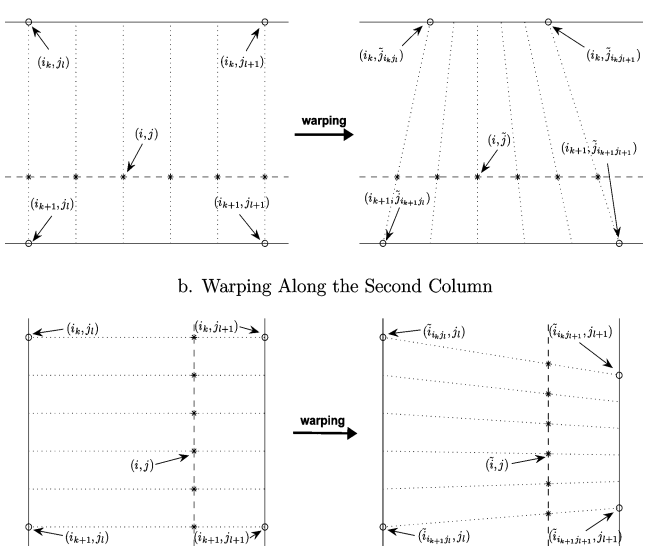

Figure 2. 2-D COW algorithm. Using the 1-D COW algorithm, the circled points are vertically warped along the first column (a) and horizontally warped along the second column (b) simultaneously. The starred points in the warped chromatograph are interpolated based on those circled points.

Aligning Homogeneous Chromatograms. When GC×GC/TOF-MS is used to analyze similar chemical samples, or even repeated samples from the same subject, the corresponding TIC chromatograms have similar features although the retention times may shift. As shown in Figure 3a,b, the TIC chromatograms for a pair of FA + AA samples are homogeneous and therefore can be directly aligned using the COW algorithm. These two samples are respectively from the first and last GC×GC/TOF-MS analyses. Shown in Figure 3a are the chromatographic contours with peaks for two FA + AA samples before alignment, and shown in Figure 3b are the corresponding contours with peaks after alignment with the prior parameters $m_r = 10$, $m_c = 40$, $\delta_r = 1$, and $\delta_c = 8$. In the original TIC chromatograms, the peaks for tetracosanoic and tricosanoic acids shift two units along the second column, while the other three peaks shift one unit along the second column, respectively. These shifts are perfectly corrected after the COW alignment. Aligning with prior parameters $m_r = 8$, $m_c = 30$, $\delta_r = 1$, and $\delta_c = 8$ presents similar results (not shown).

(17) Epanechnikov, V. A. *Theory Probability Its Appl.* **1969**, *14*, 153−158.

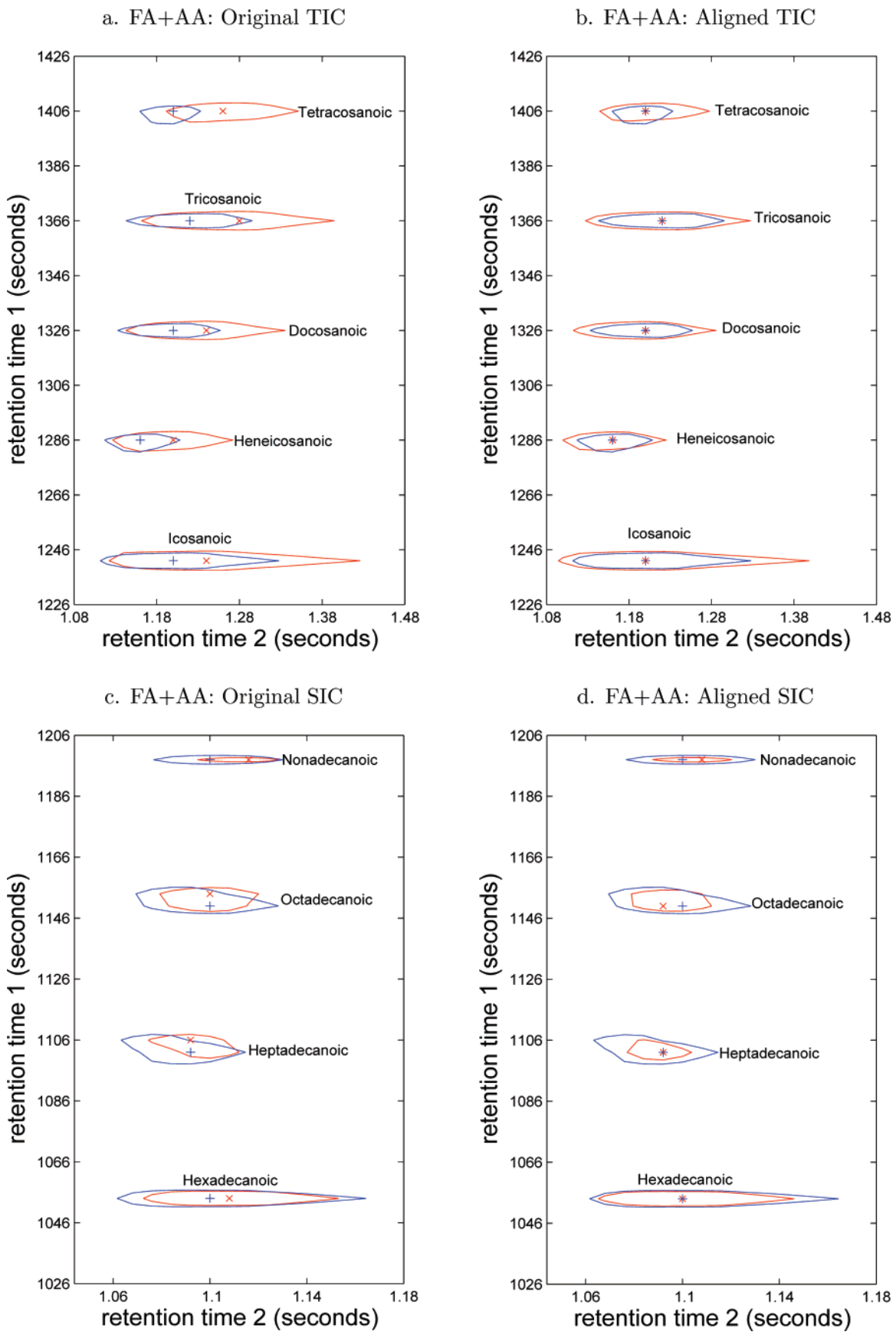

Figure 3. Aligning homogeneous TIC/SIC images. (a) Shown are the contours and peaks of TIC chromatograms for a pair of FA + AA samples, which are respectively from the first and last GC×GC/TOF-MS analyses. (b) The TIC chromatograms were aligned using the COW algorithm. (c) Shown are the contours and peaks of SIC chromatograms for the same two samples. (d) The SIC chromatograms were aligned using the COW algorithm.

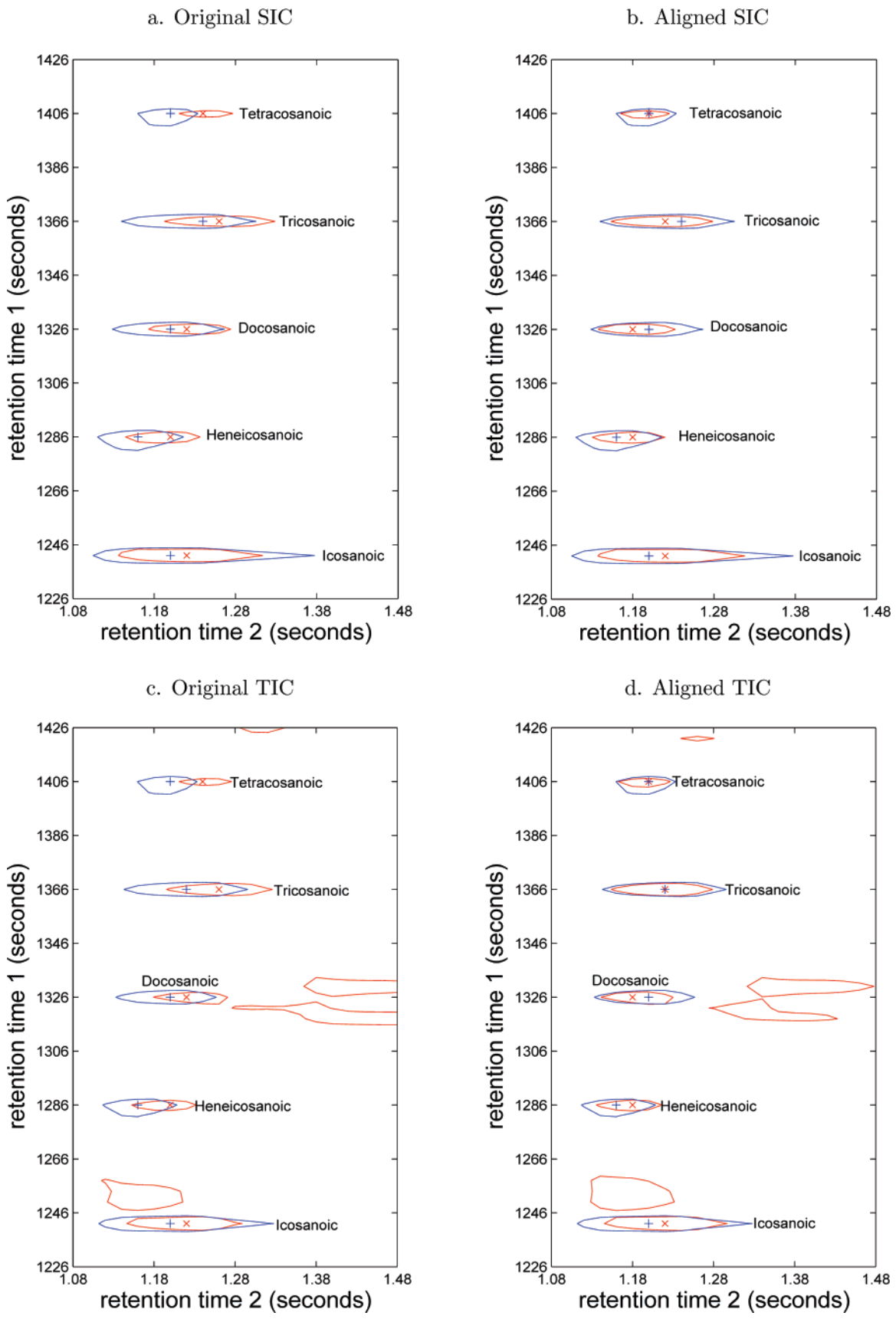

Figure 4. Aligning heterogeneous SIC/TIC images. (a) Shown are the contours and peaks of SIC chromatograms from the first FA + OA sample and the last FA + AA sample analyzed by the GC×GC/TOF-MS. (b) The SIC chromatograms were aligned using the COW algorithm. (c) Shown are the contours and peaks of TIC chromatograms for the same two samples. (d) The warping parameters obtained by aligning the SIC chromatograms were applied to the TIC chromatograms.

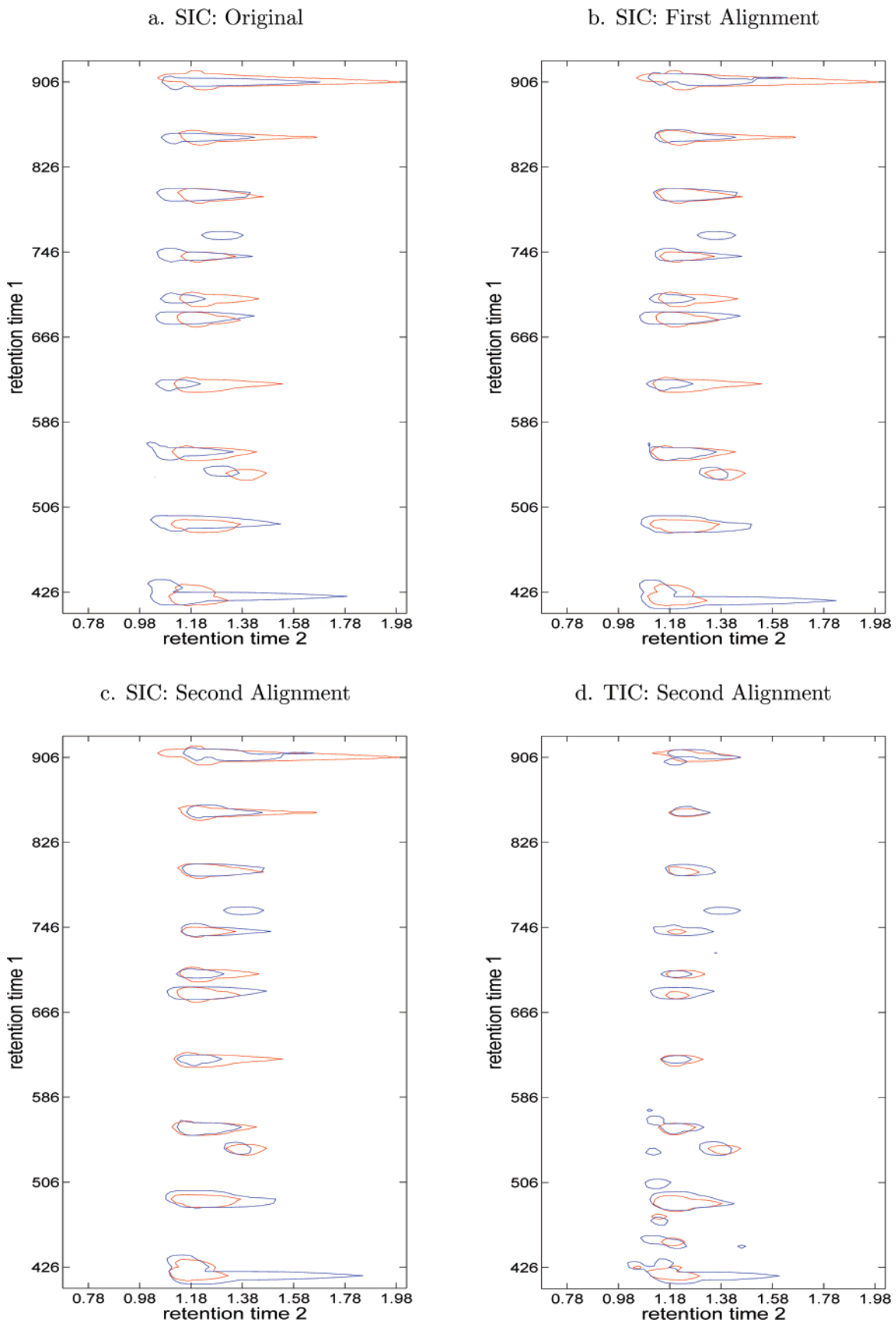

Figure 5. Aligning chromatograms from serum samples. (a) Shown are the contours of SIC chromatograms from two serum samples which were recorded 5 months apart. (b) The SIC chromatograms were aligned using the COW algorithm. (c) The resultant SIC chromatograms from the first alignment were aligned again using the same prior parameters for the COW algorithm. (d) The two sets of warping parameters were applied to the TIC chromatograms.

Figure 3c shows the contours and peaks of original SIC chromatograms for the same two samples. The peaks for both octadecanoic and heptadecanoic acids mismatch by one unit along the first column, and the peaks for hexadecanoic and nonadecanoic acids mismatch by one and two units along the second column, respectively. With the prior parameters $m_r = 5$, $m_c = 40$, $\delta_r = 1$, and $\delta_c = 8$, the COW algorithm warps along both columns to optimally match the shapes of the corresponding contours; see Figure 3d. While the alignment matches the peaks for heptadecanoic and hexadecanoic acids perfectly, the mismatch between peaks for nonadecanoic acid is shrunken to one unit. For octadecanoic acid, the one unit mismatch along the first column is switched into one unit mismatch along the second column. In general, TIC chromatograms are preferred to SIC chromatograms for homogeneous samples as TIC chromatograms usually have stronger common peaks.

Aligning Heterogeneous Chromatograms. When chromatographic profiles are diverse, a chemical component unique to one sample may be close in retention times to a neighboring chemical component unique to another sample. Direct alignment between these chromatograms may mismatch features. In such case, it is suggested that SIC chromatograms be obtained for certain specified chemical components with known m/z values and aligned using the COW algorithm. The corresponding TIC chromatograms are then warped using the warping parameters derived from aligning SIC chromatograms.

Results from aligning one pair of heterogeneous chromatograms are shown in Figure 4, with chromatograms from the first FA + OA sample and the last FA + AA sample analyzed by the GC×GC/TOF-MS. SIC chromatograms are first created on the basis of summing up intensities across m/z values specified by the 19 chemical standards in the FA mixture (see Table 1). Figure 4a presents the contours with peaks of the two original SIC chromatograms, and Figure 4b presents the contours with peaks after aligning the FA + OA SIC chromatogram to the FA + AA SIC chromatogram with the prior parameters $m_r = 5$, $m_c = 40$, $\delta_r = 1$, and $\delta_c = 8$. While the two-unit shift between peaks for tetracosanoic acid is corrected by the COW algorithm, the shift of heneicosanoic acid is improved by one unit, and the other one-unit shifts are either untouched or overcorrected. Nonetheless, the COW algorithm successfully matches all the contours, which help correct the misalignment in TIC chromatograms as shown in Figure 4c,d. The contours and peaks of the original TIC chromatograms are shown in Figure 4c, while Figure 4d shows the contours and peaks after the TIC chromatogram of the FA + OA sample being warped using the parameters obtained in aligning the corresponding SIC chromatograms. The well-matched contours of TIC chromatograms in Figure 4d demonstrate a successful transition of warping parameters from the SIC chromatogram to the corresponding TIC chromatogram.

Aligning Chromatograms of Serum Samples. Shown in Figure 5 are the results of aligning the chromatograms from two serum samples, which were recorded 5 months apart. Treating the samples as heterogeneous, SIC chromatograms (see Figure 5a) were created for the 19 FA standards listed in Table 1 and then were aligned to generate warping parameters for the corresponding TIC chromatograms. The contours of the aligned SIC chromatogram are shown in Figure 5b. With conservative prior parameters $m_r = 24$, $m_c = 12$, $\delta_r = 2$, and $\delta_c = 1$, the mismatches between the pairs of chromatograms were reduced. The warping procedure on the SIC and TIC chromatograms derived from the previous alignment was repeated without opting for better parameters. Figure 5c and Figure 5d show the contours of the resultant SIC and TIC chromatograms, respectively, from the second alignment, which demonstrate slightly better correction of mismatches than those from the first alignment.

CONCLUSIONS

A general framework of the 2-D COW algorithm is presented to align GC×GC/TOF-MS data. By partitioning raw chromatographic profiles and warping the grid points along the first and second columns simultaneously on the basis of applying the 1-D COW algorithm to characteristic vectors, nongrid points can be interpolatively warped. With homogeneous chemical samples, this 2-D algorithm can be directly applied to TIC chromatographic profiles. When the chemical samples are heterogeneous such that each sample has its unique chemical components, it is suggested to collect SIC chromatograms for chemical components specified in all samples. The 2-D COW algorithm is applied directly to align these SIC chromatograms first, and the resultant warping parameters are then used to warp the TIC chromatograms. In principle, the developed 2-D COW algorithm can be applied to align any 2-D separation images, e.g., LC×LC data, LC×GC data, GC×GC data, LC×CE data, and CE×CE data.

Unlike the piecewise alignment algorithm proposed by Pierce et al.[12] that only allows simple scalar shifts of local regions, the COW algorithm interpolatively warps local regions to maximize the correlation between the warped and reference chromatographic profiles and therefore is conceptually more powerful and flexible in correcting retention time shifts. Practically, values of the two pairs of prior parameters, i.e., the section lengths m_r and m_c, and the maximum warpings δ_r and δ_c, can be explored and set on the basis of a priori knowledge of the instrumentation and environment. Otherwise, with conservatively preset values of the prior parameters, the COW algorithm can be repeatedly applied to chromatograms until they converge under certain criteria (e.g., the cross-correlation). The resultant chromatograms usually stabilize after several repeats of the alignment and are insensitive to the choice of conservative values of prior parameters.

ACKNOWLEDGMENT

D. Zhang is grateful to the funding from Purdue University Research Foundation. F. E. Regnier acknowledges the support from NIH (AG13319). M. Zhang acknowledges the Research Support Grant from the Center on Aging and the Life Course at Purdue University.

Received for review May 24, 2007. Accepted February 1, 2008.

AC7024317