Matching 2D Gel Electrophoresis Images

K. Kaczmarek,[†] B. Walczak,*[,†] S. de Jong,[‡] and B. G. M. Vandeginste[‡]

Institute of Chemistry, Silesian University, 9 Szkolna Street, 40-006 Katowice, Poland, and
Unilever R&D Vlaardingen, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

Received November 7, 2002

Automatic alignment (matching) of two-dimensional gel electrophoresis images is of primary interest in the field of proteomics. The proposed method of 2D gel image matching is based on fuzzy alignment of features, extracted from gels' images, and it allows both global and local interpolation of image grid, followed by brightness interpolation. Method performance is tested on simulated images and gel images available via the Internet databases.

INTRODUCTION

2D gel electrophoresis is a powerful method of protein separation, based on their two independent physicochemical properties, namely isoelectric point and molecular weight.[1,2] Nevertheless, serious bottleneck of 2D gel electrophoresis is image analysis for a comparative study of biological materials. The possibility of an automated comparison of two or more gel images simultaneously is of great importance in many cases. For instance, to study protein expression in healthy and diseased tissues, many samples ought to be compared, to eliminate differences caused by individual diversity of the patients and to identify proteins related to the disease. By comparing healthy and diseased tissues on the protein level, it is possible to identify the potential protein targets as diagnostic markers also, to evaluate drug efficiency and toxicity, or to design right drugs for the disease treatment.[3,4]

Software for the analysis of 2D gel images has a long history,[5−13] but the majority of the available packages requires intensive user interactions, and in the high-throughput laboratories this remains one of the major bottlenecks.

Comparison of 2D gels images can be performed based on the extracted features[6−11,14−16] (spot coordinates, intensities, and/or areas) or directly on images.[12,13,17] Both approaches have their advantages and drawbacks.[18] Feature extraction requires detection and quantification of protein spots in each gel image. This approach is often associated with manual steps such as choice of threshold value for the background estimation and selection of landmark spots. In automated approaches, lists of unlabeled features require efficient algorithms able to deal with feature correspondence and with feature mapping simultaneously. The main argument against this approach is that the information about proteins pattern is first reduced, and then it is used by the matching algorithm.

In the second approach, all information in the gel images is used. Gel images are first aligned to each other according to their intensity distribution and then compared. Effectiveness of this approach strongly depends on the similarity measure, transformation function, and the optimization procedure used.

The aim of our study is to develop an automatic feature based method of a comparison of the 2D gel images. In the previous paper,[19] the fuzzy alignment of image features was introduced and tested, whereas in this part, fuzzy alignment followed by global or local geometric transform and brightness interpolation is presented. The proposed approach offers two possibilities; one can either perform alignment of features being at one's disposal or perform image warping, while the images are available.

THEORY

Notation. Image 1, presented in form of rectangular array containing brightness values of pixels, is denoted as **I1**. Vectors **x1** and **y1** denote coordinates of all pixels from image **I1**. Brightness value of single, ith pixel in image **I1** is denoted as $I1_i$ and its coordinates are denoted as (x1, y1).

Image Matching. Let us consider two images, image 1 and image 2, and let us assume that image 1 is a target image (standard image).

Image matching can be considered as consisting of two steps: (1) coordinates transformation and (2) interpolation of image intensity (brightness).

Coordinate transformation maps the coordinates of the pixels of one image to the corresponding coordinates in the other. The second step, interpolation, finds the points, which match the transformed points and determines their brightness values.

1. Transformation of Coordinates. To align image 2 to image 1, there are two possibilities: coordinates of image 1 can either be mapped on the corresponding coordinates of image 2 or the opposite. The first type of transform is called the inverse transform, whereas the second type is known as the forward transform (see Figure 1):

* Corresponding author phone: (+48-32)-35-91-246; fax: (+48-32)-25-99-978; e-mail: beata@tc3.ich.us.edu.pl.
[†] Silesian University.
[‡] Unilever R&D Vlaardingen.

10.1021/ci0256337 CCC: $25.00
Published on Web 02/25/2003

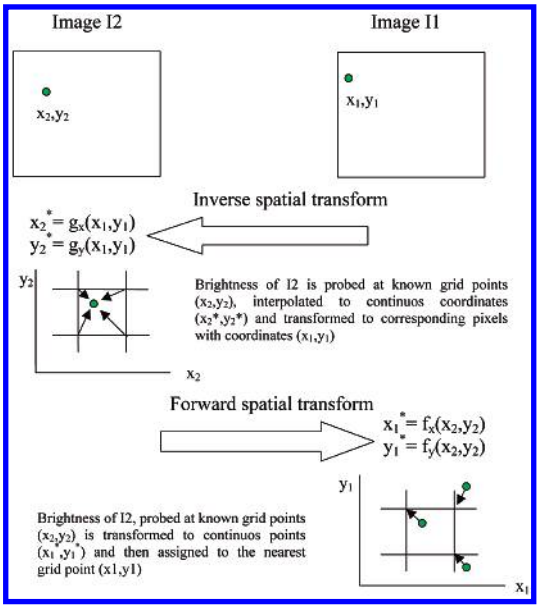

Figure 1. Scheme of relationships between inverse and forward spatial mapping and interpolation of pixel brightness.

inverse transform

$$x2^* = g_x(x1,y1) \qquad (1)$$
$$y2^* = g_y(x1,y1)$$

forward transform

$$x1^* = f_x(x2,y2) \qquad (2)$$
$$y1^* = f_y(x2,y2)$$

Forward transform maps a point with coordinates (x2, y2) to a noninteger (i.e., continuous or nongrid) point (x1*, y1*) between the four nearest pixels (see Figure 1). Because some pixels can map to points outside of the image and multiple pixels of image 2 can map to the same pixel location in output image, usually inverse transform is recommended. In inverse transform the output target image is constructed in a pixel by pixel manner, i.e., each pixel location (x1, y1) is taken in turn, and the inverse geometric mapping takes it back to a point (x2*, y2*), where (x2*, y2*) is often a nonpixel point with noninteger real number component. To this point the intensity value can be assigned based on the four corner points surrounding it.

Although the process of transforming and interpolating a source image (image 2) into a target image (image 1) using the inverse transform is straightforward, there are interpolation methods based on Delaunay triangulation,[20] which allow efficient application of the forward transform as well.

1.1. Transformation Functions. Transform functions can have different forms and different degrees of freedom, i.e., the number of parameters to be calculated. Since the form and the amount of geometric distortion between two images is not known a priori, selection of the right mapping function is part of the image matching problem.

According to our experience, one of the most efficient and well-suited methods for the 2D gels matching is a **bilinear transform**:

inverse transform

$$x2^* = a_0 + a_1x1 + a_2y1 + a_3x1y1 \qquad (3)$$
$$y2^* = b_0 + b_1x1 + b_2y1 + b_3x1y1$$

forward transform

$$x1^* = c_0 + c_1x2 + c_2y2 + c_3x2y2 \qquad (4)$$
$$y1^* = d_0 + d_1x2 + d_2y2 + d_3x2y2$$

In the matrix-vector notation, the above equations can be presented as

inverse transform

$$\mathbf{x2^* = Wa} \qquad (5)$$
$$\mathbf{y2^* = Wb}$$

forward transform

$$\mathbf{x1^* = Wc} \qquad (6)$$
$$\mathbf{y1^* = Wd}$$

where vectors **a**, **b** and **c**, **d** contain transform parameters:

$$\mathbf{a} = [a_0\ a_1\ a_2\ a_3]^T \qquad (7)$$
$$\mathbf{b} = [b_0\ b_1\ b_2\ b_3]^T$$
$$\mathbf{c} = [c_0\ c_1\ c_2\ c_3]^T$$
$$\mathbf{d} = [d_0\ d_1\ d_2\ d_3]^T$$

and matrix **W** is constructed as

inverse transform

$$\mathbf{W = [1\ x1\ y1\ x1y1]} \qquad (8)$$

forward transform

$$\mathbf{W = [1\ x2\ y2\ x2y2]} \qquad (9)$$

The least squares solution is then as follows:

inverse transform

$$\mathbf{a} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{x2} \qquad (10)$$
$$\mathbf{b} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{y2}$$

forward transform

$$\mathbf{c} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{x1} \qquad (11)$$
$$\mathbf{d} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{y1}$$

Depending on the manner in which the transform parameters are calculated, one can distinguish the so-called area based matching (ABM) and the feature based matching (FBM).

1.2. Area Based Matching (ABM). In automated area based matching (ABM) of gel images, parameters of spatial transform are optimized according to the chosen similarity measure. The most popular measure of the similarity between

the gray value windows is defined as a function of the differences between the corresponding gray values.[18] This function can be the covariance between the intensities of image 1 and transformed image 2, i.e., t(**I2**)

$$S_1 = \sum_i ((I1_i - i1)(t(I2_i) - i2)) \quad (12)$$

where $I1_i$ denotes the brightness value of an ith pixel from the image 1, whereas i1 and i2 denote mean values of brightness of all pixels from images 1 and 2, respectively; or the cross-correlation coefficients

$$S_2 = \frac{\mathrm{cov}(\mathbf{I1}, t(\mathbf{I2}))}{\sigma(\mathbf{I1}) \cdot \sigma(t(\mathbf{I2}))} \quad (13)$$

where cov(**I1**,t(**I2**)) is the covariance between **I1** and t(**I2**), and σ denotes their standard deviation; or the negative sum of the squared pixel-by-pixel differences:

$$S_3 = -\sum_i (I1_i - t(I2_i))^2 \quad (14)$$

The above measures of similarity are simple and therefore widely used. However, due to the spatial variation of the covariance or cross-correlation coefficient, it can be a very extensive and difficult task to find its global maximum and quite often, ambiguous solutions arise.

1.3. Feature Based Matching. The mapping functions can also be approximated based on several corresponding pixels in both images, called control points or landmarks. In the majority of software, control points are selected by a user, because their automatic selection is very difficult. The problem is to extract from each image a set of *n* points and to match each point from the first image to its corresponding point in the second one. When feature extraction is performed automatically, the feature correspondence problem arises and has to be addressed. To explain the feature correspondence problem, let us assume that a set of feature points from each image was extracted. Each feature point has associated feature values (e.g. coordinates, intensity etc.) such that corresponding points in the two images would be close to each other in the feature space. Point matching could then be viewed as the pairing of points from the two images such that the sum of the distances between respective pairs of feature vectors attains its minimum. There is, of course, an assumption that noncorresponding points would have dissimilar feature values. When the feature points are distributed over the entire image, accidental similarities might occur between noncorresponding points. One way to eliminate such accidental similarities is to define a robust global transform. In our earlier study,[19] an automated approach to feature matching was proposed. According to this approach, one can establish feature correspondence based on fuzzy matching and use the corresponding spots for estimation of transform parameters.

Fuzzy Alignment of Images Features—Finding Control Points. Features automatically derived for the two images are unlabeled features, i.e., their correspondence is unknown and their number can differ to a certain extent. To solve the problem of feature correspondence, we propose to transform features of image 1 to features of image 2 using an iterative algorithm, alternating between correspondence and spatial global mapping:[19] 1. for a given correspondence matrix calculate parameters of transform and 2. for transformed images update the correspondence matrix and go to 1, till the convergence is attained, where correspondence between the features of image 1 and image 2 is represented in form of a matrix **M** ($n_1 \times n_2$), where n_1 and n_2 denote numbers of spots extracted from the two considered images.

To calculate the elements of matrix **M**, we use fuzzy matching between all pairs of features with the varying degrees of fuzziness, i.e., starting with a high degree of fuzziness and decreasing it in the consecutive steps of the algorithm. The starting value of fuzziness, i.e., the value of σ of the Gaussian function, is calculated as the radius of a circle, the area of which is equal to the one-quarter of the area occupied by all spots from a more numerous set. In each iteration the degree of fuzziness is decreased by the factor equal to 0.833. At each iteration, the results of fuzzy matching are collected in a matrix, to which a process of alternating row and column normalization is applied.[21] This procedure leads to the matrix with the rows and columns summing to unit, i.e., the one-to-one correspondence can be established. These elements are then used as weights in the feature transform.

The main steps of an iterative automated matching are as follows:

0. Initialize the degree of fuzziness, i.e., the value of σ of the Gaussian function

1. Construct matrix **G** ($n_1 \times n_2$), the elements thereof representing outputs of the Gaussian functions, centered at the features of image 1, for all the features of image 2

2. Use Sinkhorn's theorem[21] to obtain a "double stochastic matrix" **M**

3. For a given correspondence (**M**) calculate parameters of weighted transform, i.e., perform matching

4. If convergence is not achieved, change a degree of fuzziness and return to step 1.

In our earlier study,[19] spot coordinates were used as the only spot features, but fuzzy matching can be performed for any type and any number of features. In the present study, fuzzy matching was performed for spots coordinates (x, y), area (A) and intensity at the maximum (I), i.e., each spot was described by four parameters [x y A I], but spot coordinates were considered as two times more important than the area and intensity.

To enhance efficiency of the fuzzy matching procedure, the extracted sets of features were preliminarily examined and the features from one image, for which there were no features in the second image within a predefined neighborhood, were eliminated from the feature set. The size of the neighborhood was determined based on our a priori knowledge about the possible degree of image distortion.

Additionally all spots with very irregular shapes and spots with the area higher than 300 pixels were removed before fuzzy matching.

1.4. Global versus Local Transform. In a global approach one mapping function is used to characterize the overall geometric distortion between the images. However, to deal with local distortions of 2D gel images, piecewise transformation should be used.[22,23]

Taking into the account that approximation of transform function is equivalent with calculation of two smooth surfaces passing through the coordinates of control points, (**x1**^{c}, **y1**^{c},

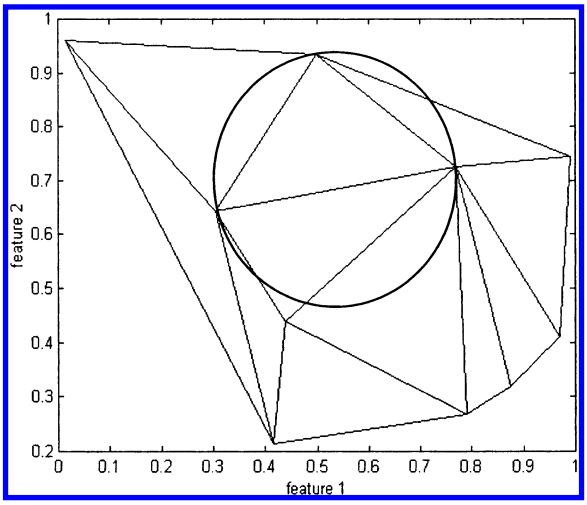

Figure 2. The principle of pseudocolor image construction.

x2^c) and (**x1**^{c}, **y1**^{c}, **y2**^{c}) in the case of inverse transform, or through (**x2**^{c}, **y2**^{c}, **x1**^{c}) and (**x2**^{c}, **y2**^{c}, **y1**^{c}) in the case of forward transform, local approximation of the transform function can be based on interpolation of these surfaces over a whole data grid. To the set of control points, the edges of both images are added, but edges of image 2 are already transformed, using the global transformation.

In this case, i.e., for the data available for a random set of points and the interpolation generated on a rectangular grid, an interpolation method based on Delaunay triangulation can be applied.

Delaunay triangulation[24,25] converts a given set of points into a network of nonoverlapping triangles such that no point in the network is enclosed by the circumscribed circle that passes through all the three vertices of the triangle (see Figure 2).

Delaunay function is used to tessellate the data, and then interpolation is based on the Delaunay tessellation. For instance, in the case of linear interpolation, the three vertices ($\mathbf{x} = [x_1,x_2,x_3]$, $\mathbf{y} = [y_1,y_2,y_3]$, $\mathbf{z} = [z_1,z_2,z_3]$) of a triangle can be used to calculate coefficients A, B, C, and D of the plane:

$$A\mathbf{x} + B\mathbf{y} + C\mathbf{z} + D = \mathbf{0}$$

Namely, coefficients are computed as

$$A = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2)$$

$$B = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2)$$

$$C = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)$$

$$D = -Ax_1 - By_1 - Cz_1$$

Another way to present plane equation is

$$\mathbf{z} = f(\mathbf{x},\mathbf{y}) = -(A/C)\mathbf{x} - (B/C)\mathbf{y} - (D/C)$$

In this form plane equation is used to compute transform at any point on the grid.

2. Brightness Interpolation. After grid transformation, image intensity interpolation is required. Interpolation step is more complicated, because transform of coordinates maps a set of integers to a set of continuous values (real numbers),

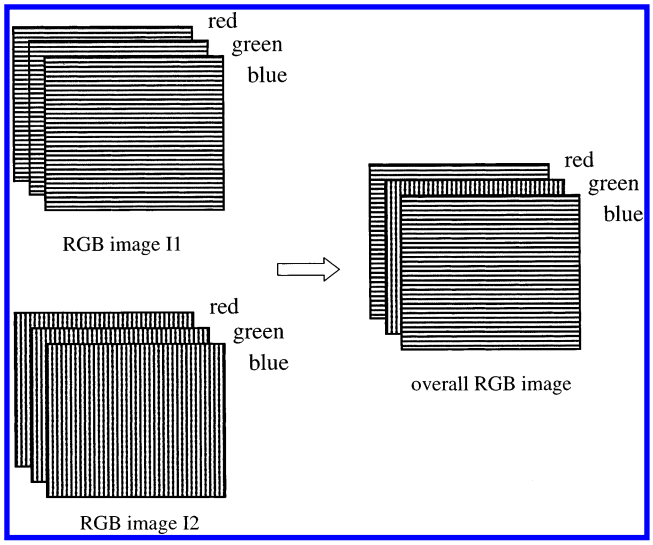

Figure 3. The Delaunay triangulation and an example of the circumcircle of the triangles.

and the real-valued positions assigned by mapping functions present complications at the discrete output.

In fact there are two distinct types of 2D interpolation. In the first approach, data are available for a rectangular grid of points, and interpolation is performed for the points off the grid. In the second case, data are available for a random set of points and the interpolation is generated on a rectangular grid. (These two types of interpolation can be performed in matlab using the interp2 and griddata functions, respectively[20]).

2.1. Interpolation Methods. There are many methods of interpolations, such as linear, polynomial, nearest neighbor, etc. Their choice determines the results of final matching and of course depends on the image types. The most popular interpolation methods are nearest neighbor, linear, and cubic.

The nearest neighbor interpolation method fits a piecewise constant surface through the data values. The value of an interpolated point is the value of the nearest point.

Linear interpolation method fits a linear surface through the data points. The value of an interpolated point is a combination of the values of the four closest points. This method is piecewise bilinear.

Cubic interpolation method fits a cubic surface through the existing data points. The value of an interpolated point is a combination of the values of the 16 closest points. This method is piecewise cubic and produces a much smoother surface than bilinear interpolation. This can be a key advantage for applications to, e.g., image processing.

3. Visualization of Matched Images. Matched images may be visualized by superimposing them and displaying in a "pseudocolor" mode. Pseudocolor image is constructed by a special combination of three color components (red, green, and blue) from both images, and namely, red and blue components are taken from image 1, whereas the green component is taken from image 2 (see Figure 3).

Spots from image 1, which have no corresponding spots in image 2, are displayed in overall image as green spots, and spots observed in image 2 only appear as magenta spots. When two images with the similar spot patterns are well matched, i.e., spots from both images overlap, the entire image has gray scale appearance. Weak spots are light gray, and the strongest spots are black.

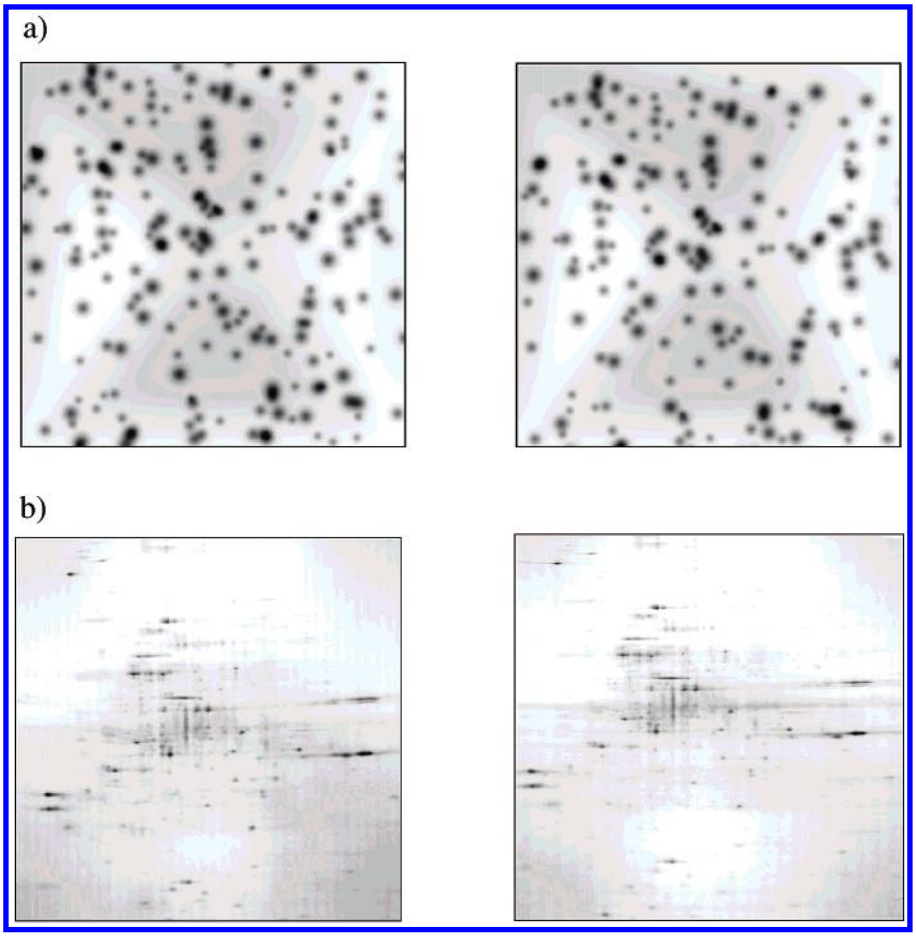

Figure 4. Simulated (a) and real (b) gel image pairs.

This way of visualization allows fast identification of the spots with no-correspondence and the spots with different expression.

4. Algorithm of Image Matching. The main steps of the proposed approach are as follows:

1. select automatically a set of features
2. calculate correspondence between selected features and select the best set of features
3. choose the method of coordinate transform
4. calculate parameters of the global or local transformation
5. transform coordinates of image pixels
6. choose method of brightness interpolation
7. perform brightness interpolation
8. visualize matched images.

DATA

Performance of the proposed approach was thoroughly tested on simulated and real 2D gel images (real gel images were downloaded from the available net databases[26−28]).

Gel images were simulated in the following steps:

1. randomly generate coordinates of 100 features, i.e., those representing features of image 1
2. construct features of image 2 by locally distorting features of image 1 and performing their global transform
3. randomly remove certain fraction of features (up to 5%)
4. generate images image 1 and image 2 (100 × 100 pixels each) with the intensities of spots described by the Gaussian functions centered at the coordinates of features.

For an illustrative purpose two pairs of images (simulated and real) are presented in Figure 4.

Each of the studied pairs of images was inverse transformed with use of global linear transform, and different methods of brightness interpolation were tested, such as nearest neighbor, linear, and cubic. In the case of local transformation, the surface fitting was performed, using linear and cubic interpolation.

RESULTS

To illustrate the performance of the proposed approach, let us consider two 2D gel images presented in Figure 4b. Applying the thresholding procedure, we can identify the most intense spots (see Figure 5a) and calculate their coordinates, area, and maximum intensity. At this stage of feature identification, spots with very irregular shapes, or those with no correspondence, etc. are removed (see Figure 5b). Spots remaining after preselection are used as input to the fuzzy matching procedure, aiming to establish the spots' correspondence (see Figure 5c). Corresponding spots are then used to calculate the parameters of global transform of the image 2 grid, and this transform is followed by interpolation of brightness of image 2. In this example, parameters of the inverse linear transform (calculated with aid of the selected "landmarks") were

$$\mathbf{a} = [-10.2948\ 0.9993\ -0.0240\ 0.0001]$$

$$\mathbf{b} = [-21.9775\ 0.0309\ 1.0025\ -0.0001]$$

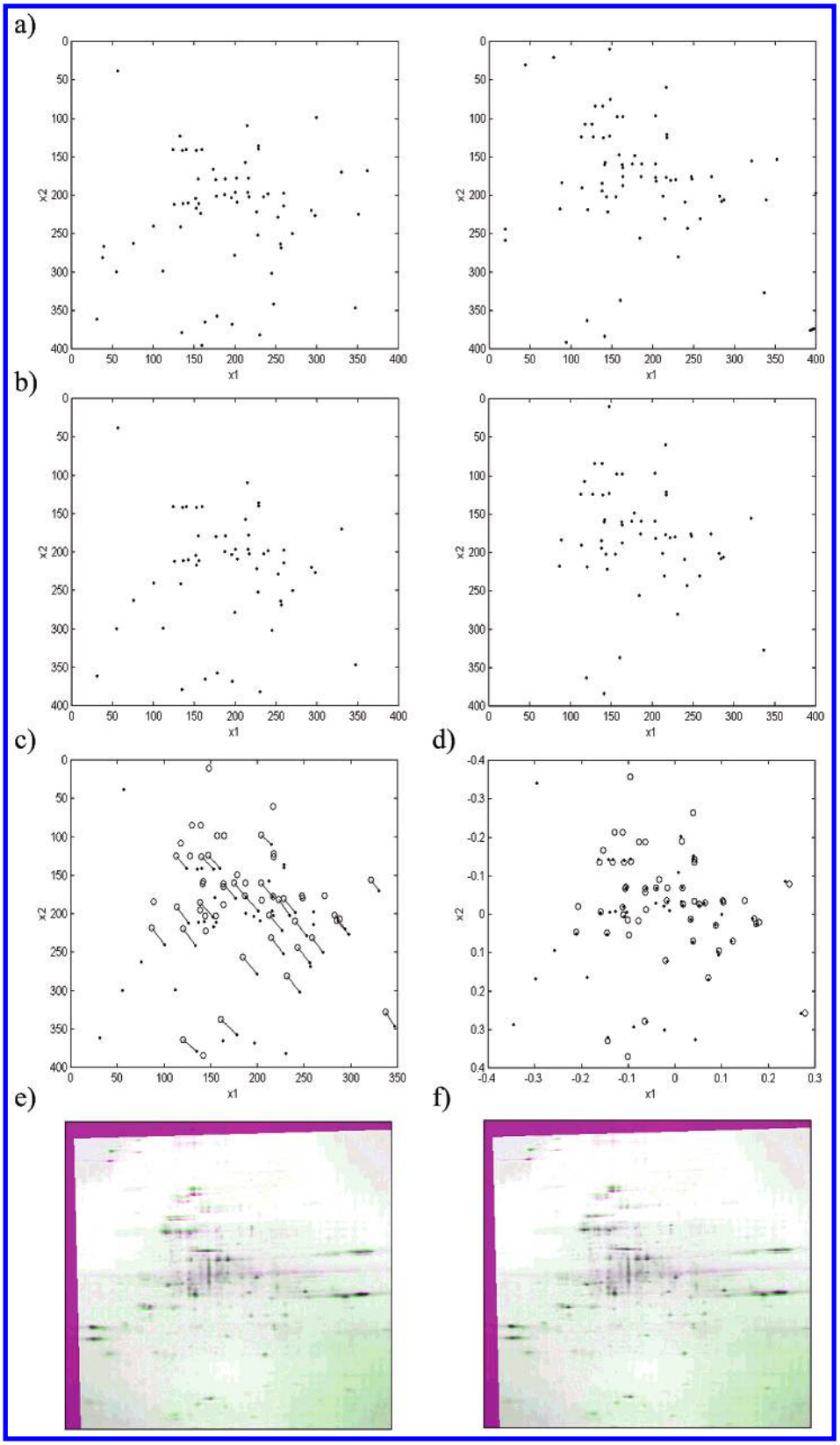

Figure 5. The main steps of image matching: (a) features selected from the real gel images presented in Figure 4a, (b) features after preselection, i.e., after discarding the irregular spots, etc., (c) features with the corresponding spots connected with lines, (d) matched features, and (e) superimposed images after (e) global and (f) local transformation.

Features aligned using the aforementioned transform parameters are presented in Figure 5d. Interpolation of brightness of image 2 was performed, using the cubic interpolation method. Final results of the described procedure, performed for the discussed pair of images, are presented in Figure 5e.

As one can easily notice, in the pseudocolor image (presented in Figure 5e) there are some spots not too well

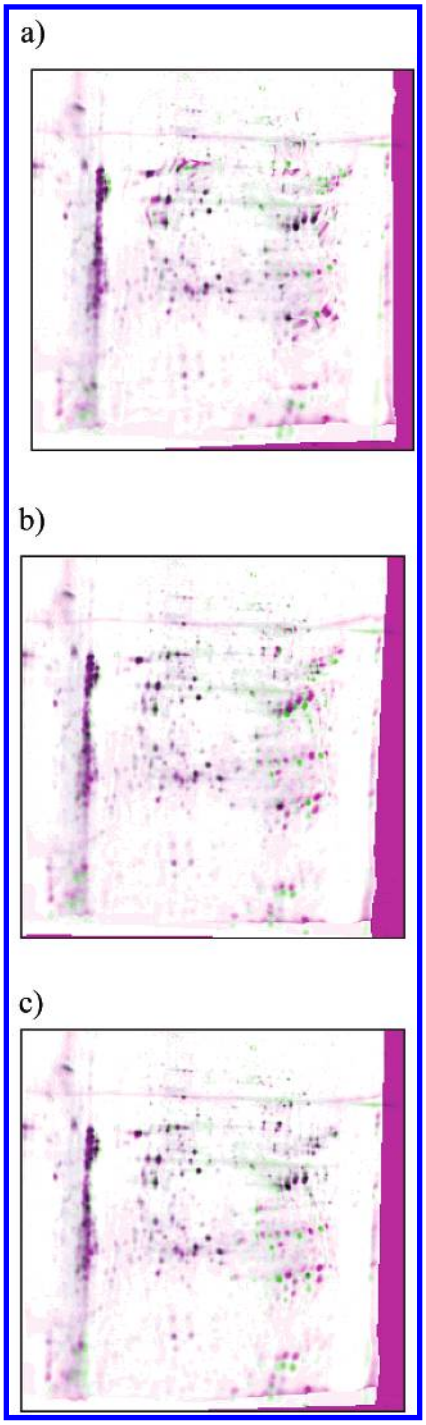

Figure 6. Matching of images performed, using the different numbers of spots: (a) 100; (b) 50; and (c) 20.

aligned, based on the described approach. However, further improvement can be achieved using local transform of the image 2 grid (Figure 5f).

Final alignment of images obviously depends on the performance of every step of the described procedure and on the applied transform and interpolation functions.

Performance Evaluation. The proposed approach is fast, with the calculation time mainly depending on the number of the extracted features. According to our study, this number can strongly be limited, without affecting the performance of image matching. Examples of the aligned images using different numbers of features are presented in Figure 6. Differences between these images are quite negligible, and hence it is possible to use a smaller amount of features without any considerable deterioration of matching quality. For evaluation of performance, all spots with the correspondence higher than 0.95 were taken into consideration.

Table 1. Results of Performance Evaluation for the Proposed Approach on Simulated Gel Images Presented as Means ($\bar{R}_c$ and $\bar{R}_c^p$) and Medians of Correlation Coefficients Ratios (R_c^* and R_c^{p*}) between the Standard, I1, and the Transformed, t(I2), Images

grid transform		brightness interpolation	$\bar{R}_c$	$\bar{R}_c^p$	R_c^*	R_c^{p*}
global		NN	4.04	4.57	3.76	4.08
		linear	4.11	4.76	3.82	4.24
		cubic	4.10	4.78	3.80	4.26
local	linear grid interpolation	NN	4.20	4.81	3.95	4.34
		linear	4.27	4.99	3.99	4.46
		cubic	4.27	5.03	3.98	4.52
	cubic grid interpolation	NN	4.20	4.78	3.93	4.31
		linear	4.28	4.98	3.98	4.44
		cubic	4.27	5.01	3.99	4.48

Performance of the proposed approach was tested for the simulated pairs of images, because of the simulated data only the correspondence spots are known precisely and the different degrees of local image distortions can be studied in detail. The results of the performed study are summarized in Table 1.

The quality of image matching is described by two measures, R_c and R_i, defined as the ratio of the correlation coefficients (R_c) and the ratio of the differences in brightness (R_i) between the original (nonmatched) and the transformed (matched) images. These measures allow overcoming of the problem with imperfectly determined brightness of individual pixels and with the lack of an ideal geometrical transform of pixel coordinates. Because of the presence of the noncorresponding spots, R_c and R_i calculated for the whole images are a "rough" quality measures only. For this particular reason the goodness of matching was also calculated in terms of the ratios of correlation coefficients and brightness differences determined for the parts of images with the corresponding spots, R_c^p and R_i^p, only.

In general, better results were obtained for local transformations, both linear and cubic, than for the global ones. The best results are obtained for the local linear transformation of a grid, followed by cubic interpolation of brightness (see Table 1). However, each combination of the grid and brightness interpolations yields acceptable results that differ to a small extent only.

The bar plots of the both ratios, R_c^p and R_i^p, calculated for the locally transformed pairs of gel images, are presented in Figure 7a,b. Calculations were performed for 750 pairs of the simulated gel images. For each pair of images, all combinations of interpolation methods (i.e., linear and cubic local interpolation of the grid followed by the nearest neighbor, linear and cubic brightness interpolation) were evaluated. Due to the presence of image pairs with initial ratios almost equal to zero, some ratios attain very high values, so that 10% of the highest values were discarded. This yields the totality of 4050 results, i.e., 675 results for each of the six possible combinations of interpolation methods. According to these results, there are no considerable differences between any of the methods of grid and brightness interpolation used. However, in agreement with the results presented in Table 1, the best results are obtained

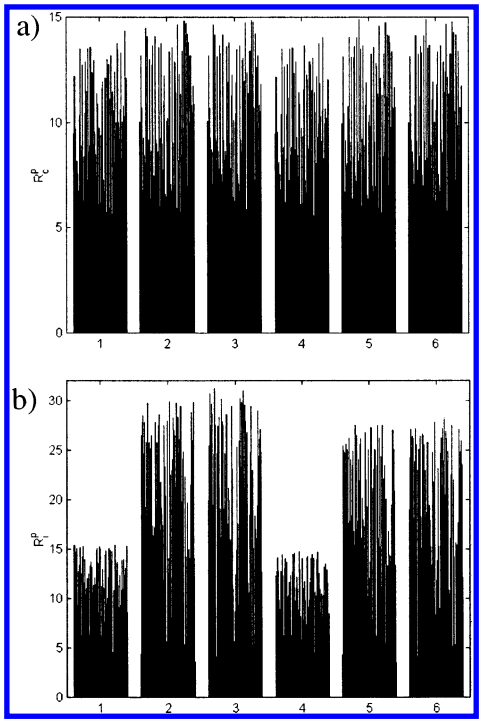

Figure 7. The results of performance evaluation of the proposed approach on simulated gel images, calculated for the six different combinations of grid and brightness interpolation (linear interpolation of grid, followed by nearest neighbor (1), linear (2) or cubic (3) brightness interpolation and cubic interpolation of grid, followed by nearest neighbor (4), linear (5) or cubic (6) brightness interpolation), presented as (a) ratio of correlation coefficients, R_c^p, and (b) ratio of differences in pixel intensities, R_i^p.

with the third combination, i.e., in the case of local transformation performed with linear interpolation and cubic brightness interpolation using both quality measures. Additionally, for the ratio of brightness differences there is a noticeable difference between the nearest neighbor brightness interpolation and the other ones, i.e., the linear and cubic interpolations. It means that in the case of this quality measure the greatest influence on the obtained results exerts a method used for brightness interpolation.

The proposed approach was tested on 100 pairs of real gel images also, downloaded from the Internet databases.[26−28] In this case matching was performed with the local linear grid transformation, followed by cubic brightness interpolation, which yielded the best results for the simulated gel image pairs.

R_c^p and R_i^p calculated for real images are presented in Figure 8(parts a and b, respectively). As it can be seen from this figure, all correlation ratios are higher than unity, so that the matching has proved successful and the transformed image t(**I2**) has a higher correlation with the standard image **I1** than with the original image **I2**. Ratios of differences in pixel intensities also indicate that for the majority of the 100 image pairs tested, matching of images was successful. For only a few image pairs does this ratio have the value close to unity, i.e., matching does not significantly improve the similarity between images.

CONCLUSIONS

The presented method of gel image matching allows efficient automated matching of two-dimensional gel electrophoresis images. Its efficiency is limited by the performance of fuzzy alignment used to align the sets of the extracted features. Corresponding features can be used to calculate the parameters of global transform or for local interpolation of image grid to improve alignment of images in these cases, where local distortions are present. Final goodness of a fit depends on a type of transform also and on the interpolation method used. Best results were obtained for linear interpolation of the grid and for cubic interpolation of the values of pixel brightness.

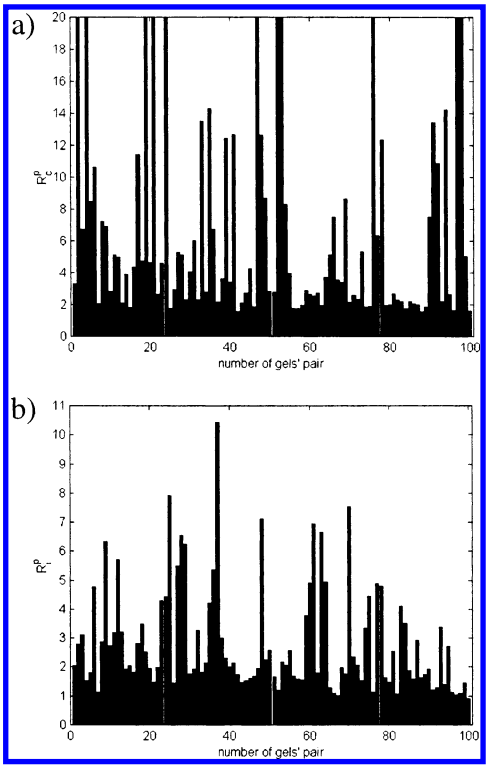

Figure 8. The results of image matching for 100 pairs of real gels images: (a) ratio of correlation coefficients, R_c^p, and (b) ratio of differences in pixel intensities, R_i^p.

ACKNOWLEDGMENT

K. Kaczmarek thanks Unilever R&D Vlaardingen (The Netherlands) for financial support of his Ph.D. study.

REFERENCES AND NOTES

(1) O'Farrell, P. H. High-resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007−4021.
(2) Westermeier, R. *Electrophoresis in practice: a guide to theory and practice*; VCH: Weisenheim, 1993.
(3) Young, D. S.; Tracy, R. P. Clinical applications of two-dimensional electrophoresis. *J. Chromatogr. A* **1995**, *698*, 163−179.
(4) Bandara, L. R.; Kennedy, S. Toxicoproteomics−a new preclinical tool. *Drug Discov. Today* **2002**, *7*, 411−418.
(5) Harry, J. L.; Wilkins, M. R.; Herbert, B. R.; Packer, N. H.; Gooley, A. A.; Williams, K. L. Proteomics: capacity versus utility. *Electrophoresis* **2000**, *21*, 1071−1081
(6) Lemkin, P. F.; Merril, C.; Lipkin, L.; Van Keuren, M.; Oertel, W.; Shapiro, B.; Wade, M.; Schultz, M.; Smith, E. Software aids for the analysis of 2D gel electrophoresis images. *Comput. Biomed. Res.* **1979**, *12*, 517−544.
(7) Lemkin, P. F.; Lipkin, L. E.; Lester, E. P. Some extensions to the GELLAB 2D electrophoresis gel analysis system. *Clin. Chem.* **1982**, *28*, 840−849.
(8) Garrels, J. I. The QUEST system for quantitative analysis of two-dimensional gels. *J. Biol. Chem* **1989**, *264*, 5269−5282.
(9) Lemkin, P. F. Comparing Two-Dimensional gels across the Internet. *Electrophoresis* **1997**, *18*, 461−470.

(10) Appel, R. D.; Vargas, J. R.; Palagi, P. M.; Walther, D.; Hochstrasser, D. F. Melanie II−a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. *Electrophoresis* **1997**, *18*, 735−748.
(11) Pleissner, K. P.; Hoffmann, F.; Kriegel, K.; Wenk, C.; Wegner, S.; Sahlström, A.; Oswald, H.; Alt, H.; Fleck, E. New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases. *Electrophoresis* **1997**, *20*, 755−765.
(12) Veeser, S.; Dunn, M. J.; Yang, G. Z. Multiresolution image registration for two-dimensional gel electrophoresis. *Proteomics* **2001**, *1*, 856−870.
(13) Smilansky, Z. Automatic Registration for Images of Two-dimensional Protein Gels. *Electrophoresis* **2001**, *22*, 1616−1626.
(14) Hoffmann, F.; Kriegel, K.; Wenk, C. An applied point pattern matching problem: comparing 2D patterns of protein spots. *Discrete Appl. Math.* **1999**, *93*, 75−88.
(15) Takahashi, K.; Nakazawa, M.; Watanabe, Y.; Konagaya, A. Fully automated spot recognition and matching algorithm for 2-D gel electrophoretogram of genomic DNA. *Genome Informatics* **1998**, *9*, 173−182.
(16) Pánek, J.; Vohradsk×c6, J. Point pattern matching in the analysis of two-dimensional gel electrophoretograms. *Electrophoresis* **1999**, *20*, 3484−3491.
(17) Conradsen, K.; Pedersen, J. Analysis of 2-dimensional electrophoretic gels. *Biometrics* **1992**, *48*, 1273−1287.
(18) Heipke, C. Overview of Image Matching Techniques. *OEEPE-Workshop "Applications of Digital Photogrammetric Workstations", Proceedings*; Lausanne, Switzerland, 1996; pp 173−191.
(19) Kaczmarek, K.; Walczak, B.; de Jong, S.; Vandeginste, B. G. M. Feature Based Fuzzy Matching of 2D Gel Electrophoresis Images. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1431−1442.
(20) *Matlab Function Reference*; The MathWorks, Inc.: 1999.
(21) Sinkhorn, R. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Stat.* **1996**, *35*, 876−879.
(22) Goshtasby, A. Piecewise linear mapping functions for image registration. *Pattern Recogn*. **1986**, *19*, 459−466.
(23) Goshtasby, A. Piecewise cubic mapping functions for image registration. *Pattern Recogn.* **1987**, *20*, 525−533.
(24) de Berg, M.; van Kreveld, M.; Overmars, M.; Schwarzkopf, O. *Computational Geometry−algorithms and applications*; Springer-Verlag: Berlin, 1997.
(25) Bern, M.; Eppstein, D. Mesh Generation and Optimal Triangulation. In *Computing in Euclidean Geometry*; Du, D.-Z., Hwang, F. K., Eds.; World Scientific: Singapore, 1992; Vol. 1, pp 23−90.
(26) WORLD-2DPAGE−2D PAGE databases and services at http://www.expasy.ch/ch2d/2d-index.html.
(27) NCI Flicker Database at http://www-lecb.ncifcrf.gov/2dwgDB/.
(28) GelBank at http://gelbank.anl.gov.

CI0256337