

# HYBRID REGISTRATION FOR TWO-DIMENSIONAL GEL PROTEIN IMAGES

XIUYING WANG

*Biomedical and Multimedia Information Technology (BMIT) Group  
School of Information technologies, The University of Sydney  
Sydney, NSW 2006, Australia*

DAVID DAGAN FENG

*Biomedical and Multimedia Information Technology (BMIT) Group  
School of Information Technologies, The University of Sydney  
Sydney, NSW 2006, Australia  
Center for Multimedia Signal Processing  
Department of Electronic and Information Engineering  
Hong Kong Polytechnic University, Hong Kong*

Image registration technique is fundamental and essential to the accurate and efficient analysis of protein sequence data. But due to the elastic deformations of two-dimensional gel protein electrophoresis images, their registration still remains a challenge. In this paper, a hybrid 2D gel protein image registration method is proposed. In the first stage of registration, the wavelet-based hierarchical registration which fully exploits the image intensity is used to correct global displacements and affine transformation parameters are achieved. In the second stage, the landmark-based elastic registration is introduced to correct local displacements and enhance registration performance and accuracy. In the proposed method, the hierarchical registration from low resolution to high resolution can accelerate the registration convergence and hence good computational efficiency can be achieved. Though wavelets have been widely used in image processing, its application in the area of gel protein image registration is not well investigated. In this paper, the wavelet-based hierarchical approach is introduced in 2D gel protein image registration. The algorithm makes use of the merits of the existing categories of registration techniques and achieves high performance registration results automatically.

## 1 Introduction

Proteomics is the study and analysis of the nature and function of DNA and protein sequence data. In the post-genome era, the analysis and study of proteomics are playing a more and more important role in the areas such as life science, therapeutics, and disease prevention and inhibition. Two-dimensional gel protein electrophoresis is one of the main technologies in the analysis and separation of complex protein mixtures [1]. This technology provides sufficient information about a variety of proteins simultaneously and hence can enhance and facilitate the quantitative and qualitative research of patterns of protein expression.

Image registration techniques are an important tool for the comparison and analysis of two-dimensional protein expression and structure. Due to its important role in image

analysis, registration has been studied intensively in different areas [2, 3]. Image registration aims to determine a transformation between two image spaces, which maps points of one image onto corresponding points of another image. Because of the current technology of 2D gel protein electrophoresis, the obtained images may be deformed randomly. Therefore, rigid registration which only corrects the rotation and translation deformations is not effective and the more complicated registration approaches are needed to solve these non-linear deformations. Principally, elastic registration approaches can be distinguished into intensity-based and feature-based methods [4].

Feature-based registration approaches are widely used in registering 2D gel protein images [5]. In these approaches, the features can be extracted manually or interactively. One main advantage of feature-based registration is that the transformation often can be stated in analytic form, which leads to efficient computational schemes. However, in feature-based registration methodologies, the preprocess step is needed and the registration results are highly dependent on the result of this preprocess. Because registration algorithms using landmarks often require users specify corresponding landmarks from the two images manually or semi-automatically, such methods cannot always provide very accurate registration [6].

Fully and directly exploiting the image intensities [7], the intensity-based gel image registration algorithms have the advantages of no segmentation required and most importantly, these methods have potentials to achieve automated registration. However, the computation of this category of schemes not efficient. To improve the computational efficiency and registration accuracy of intensity-based gel protein image registration, numerous multiresolution registration methods have been proposed [8].

In order to improve the registration performance, accuracy, and at same time avoiding the local minima, in this paper, a hybrid 2D gel protein image registration is proposed, which firstly corrects global displacements using hierarchical intensity-based method and then further improves the registration accuracy and performance by elastic registration of automatically localized landmark points.

## 2 Hybrid Registration Algorithm

### 2.1 Global Hierarchical Registration Based on Wavelet Decomposition

#### 2.1.1 Wavelet-based image decomposition

Because they are well localized in space and frequency domains, wavelets have been applied in the areas of image compression, denoising, enhancement, and deconstruction. The wavelet decomposition of a function  $f(x) \in L^2(\mathfrak{R})$  can be defined as:

$$(W_\psi f(x))(a, b) = \langle f(x), \psi_{ab}(x) \rangle = \frac{1}{\sqrt{|a|}} \int_{\mathfrak{R}} f(x) \psi\left(\frac{x-b}{a}\right) dx \quad (1)$$

Where the function  $\psi_{ab}(x)$  defines the family of the wavelet functions with  $a, b \in \mathfrak{R}$  and  $a \neq 0$  is the dilation parameter and  $b$  is the translation parameter. Wavelets  $\psi_{ab}(x)$  are obtained from mother wavelet  $\psi(x)$  by dilation and translation [9]:

$$\psi_{ab}(x) = \sqrt{2^{-a}} \psi(2^{-a}x - b) \quad (2)$$

For a given 2-dimensional image of size  $2^n * 2^m$ , the wavelet-based image decomposition can be achieved by convolving the wavelet low-pass filter  $L$  and the wavelet high-pass filter and down-sampling by a factor of 2 along rows and columns independently.

### 2.1.2 The Proposed Wavelet-based Hierarchical Registration

The application of wavelet in gel image registration has not been well studied. In this paper, we propose a hierarchical gel image registration approach based on wavelet decomposition (Figure 1).

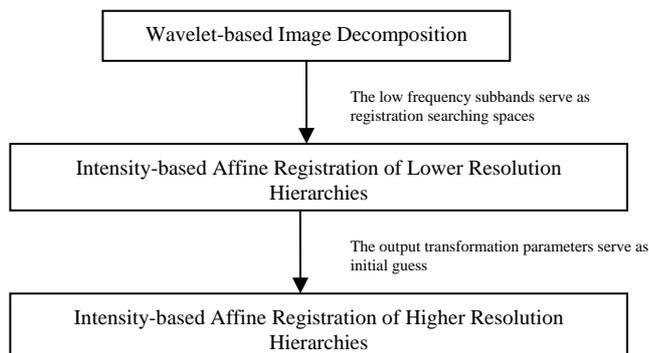


Figure 1. Wavelet-based Hierarchical Image Registration

The intensity-based image registration, especially for the mutual information based registration, has the property of high precision, but it is also time consuming. Therefore, to improve the computational efficiency, we register the images from low resolution (high level) of the registration pyramid to high resolution (low level) of the registration pyramid.

- Based on wavelets, the images are decomposed into subbands. Because the low-frequency subbands are the compact and smooth information of their original images, we use the low-frequency subbands as searching spaces in each hierarchy of the registration pyramid.

- In each hierarchy, based on mutual information criterion, the affine registration is performed to correct rotation, translation, and scaling displacements between the images. The results of the current registration hierarchy are used as the initial guess for the next hierarchy by keeping the same rotation and scaling parameters and double the transformation parameters.

Hierarchical registration approach has the property of high computing efficiency, which registers images from high level registration pyramid with coarse resolution to low level registration pyramid with fine resolution. In this paper, Daubechies' 4-coefficient and 8-coefficient are used to decompose the images because of their ability of keeping energy in the low frequencies.

### 2.1.3 Registration criterion-mutual information (MI)

Maximization of MI [10] is a general and powerful registration criterion, because no assumption are made regarding the nature of the dependence and no limiting constraints are imposed on the image content of the modalities involved.

Let  $R$  be the reference data presented by  $m$  samples  $\{r_0, r_1, \dots, r_{m-1}\}$  with a marginal probability distribution  $P_R(r)$ . Analogously, the study data  $S$  consists of  $n$  samples  $\{s_0, s_1, \dots, s_{n-1}\}$  with a marginal probability distribution  $P_S(s)$ . The mutual information  $I$  of the reference image  $R$  and study image  $S$  measures the degree of dependence of  $R$  and  $S$  by measuring distance between the joint distribution  $P_{RS}(r, s)$  and the distribution associated to  $P_R(r)$  and  $P_S(s)$ . MI can be defined as:

$$I_{R,S} = \sum_{(r,s)} P_{RS}(r, s) \log\left(\frac{P_{RS}(r, s)}{P_R(r)P_S(s)}\right) \quad (3)$$

Under the assumption that the mutual information of the two images is maximum when the images are in registration, registration can be performed by maximizing the mutual information as a function of a geometric transformation  $T$  of the study image  $S$ :

$$S_T = T(S) \quad (4)$$

$$I_{R,S_T} = \sum_{(r,T(s))} P_{R,S_T}(r, T(s)) \log\left(\frac{P_{R,S_T}(r, T(s))}{P_R(r)P_{S_T}(T(s))}\right) \quad (5)$$

$$T_{reg} = \arg \max_T (I_{R,S_T}) \quad (6)$$

where  $T_{reg}$  is registration transformation that will bring the images into registration.

In our registration approach, affine transformation, which has six degrees of freedom, is used in each registration hierarchy to correct the rotation, translation, and the scaling displacements between images.

## 2.2 Local Elastic Registration Based on Automatically Localized Landmark Points

After the affine hierarchical registration, the global displacements have been corrected. However, there still exist local elastic differences between the images. In this paper, the automatic landmark point selection method is proposed (Figure 2) and the elastic registration is carried out to further improve the registration precision.

### 2.2.1 Automatic Landmark Points Localization

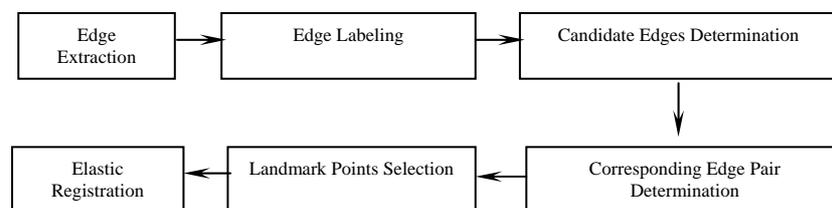


Figure 2. Automatic Landmark Point Selection Procedure

The detailed algorithm explanation is described as follows:

- **Edge extraction:** Due to its optimal property of noise suppression, Canny edge detection algorithm is selected to extract the edges from the output images of the first global intensity-based registration. In Canny method, the image is convolved with a Gaussian filter; then the local maxima of the image gradient is achieved; using two thresholds, the edge candidates are examined and the connectivity is maximized.
- **Edge labeling:** The region-growing technique is used to label the 8-connected objects in the binary edge images and then the edges are ranked according to their perimeters.
- **Candidate edge determination:** The edges which perimeters are greater than a predetermined threshold are chosen as candidate edges for the following automatic landmark point selection procedure. Through this step, those edges with small perimeter are eliminated and the influence of noise is avoided.
- **Corresponding edge pair searching:** For each labeled edge of study image, in order to find its corresponding edge in reference image, we iteratively move the edge of study image on the reference image within a searching range and select the one with the most similar perimeter and least distance from it as its corresponding edge.
- **Landmark points selection:** For each pair of corresponding edge, the centroid points, the points with maximum and minimum distances to their centroid points are selected as landmark points for the elastic registration.

### 2.2.2 Thin-Plate Splines (TPS)

The use of thin-plate spline interpolation for registration purpose was first proposed by Bookstein in 1989 [11]. The main reason for choosing the thin-plate splines in our algorithm is because it can produce a smoothly interpolated spatial mapping.

Given two sets of  $n$  landmarks (2D data are used in this section)  $p_i = (x_i, y_i)$  and  $q_i = (x_i', y_i')$ ,  $i = 1, 2, \dots, n$  in the study and reference image respectively. Finding a transformation  $f : f(p_i) = q_i, i = 1, 2, \dots, n$  to minimize the following energy function can give us a smooth spline interpolation which is capable of warping points  $p_i$  arbitrarily close to points  $q_i$ . A regulation parameter  $\lambda > 0$  determines the closeness of the fit.

$$E(f) = \sum_{i=1}^n \|q_i - f(p_i)\|^2 + \lambda \iint_{\infty} \left( \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 \right) dx dy \quad (7)$$

The thin-plate spline interpolation function can be written as:

$$f(x, y) = \sum_{k=1}^3 a_k \phi_k(t) + \sum_{i=1}^n w_i U(\|(x, y) - p_i\|) \quad (8)$$

where  $[\phi_1(t), \phi_2(t), \phi_3(t)] = [1, x, y]$ ,  $U(\|(x, y) - p_i\|) = \|(x, y) - p_i\|^2 \log(\|(x, y) - p_i\|)$ .

In order to keep  $f(x, y)$  having square integrable second derivatives, we require that

$$\begin{aligned} \sum_{i=1}^n w_i &= 0 \\ \sum_{i=1}^n w_i x_i &= \sum_{i=1}^n w_i y_i = 0 \end{aligned} \quad (9)$$

The coefficient vector  $a = (a_1, a_2, a_3)^T$  and  $w = (w_1, w_2, \dots, w_n)^T$  can be computed through the following linear equations:

$$\begin{aligned} Kw + Pa &= v \\ P^T w &= 0 \end{aligned} \quad (10)$$

where  $v$  represents column vectors of landmarks;  $k_{ij} = U_i(p_j) = U(\|(x_i, y_i) - (x_j, y_j)\|)$ ;  $(1, x_i, y_i)$  is the  $i$ th row in the  $P$ .

These two vector equations can be solved by:

$$\begin{aligned} w &= K^{-1}(v - Pa) \\ a &= (P^T K^{-1} P)^{-1} P^T K^{-1} v \end{aligned} \quad (11)$$

The proposed registration process is optimized by Powell optimization algorithm [12]. The Powell algorithm performs a succession of one-dimensional optimizations, finding in turn the best solution along each freedom degree, and then returning to the first degree of freedom. The algorithm stops when it is unable to find a new solution with a significant improvement to the current solution.

### 3 Experimental Validation and Discussion

To access the performance of our proposed hybrid registration algorithm, experiments are carried out using 2D gel images of human blood plasma, human hela cell, and cerebrospinal fluid downloaded from Swiss-2DPAGE [13]. To validate our proposed algorithm, firstly, we deform the reference images with predetermined polynomial warping, translations and rotations to create the study images. Then the proposed registration procedure is carried out to transform the study images to their corresponding reference images. Because there is no gold standard available for measuring the performance of registration algorithms, in our paper, the registration results are tested by subtracting the registered images and visually investigating the difference image. The more the corresponding parts overlap, the more successful the registration is.

#### 3.1 Intensity-based Hierarchical Registration of Human Blood Plasma Gel Images

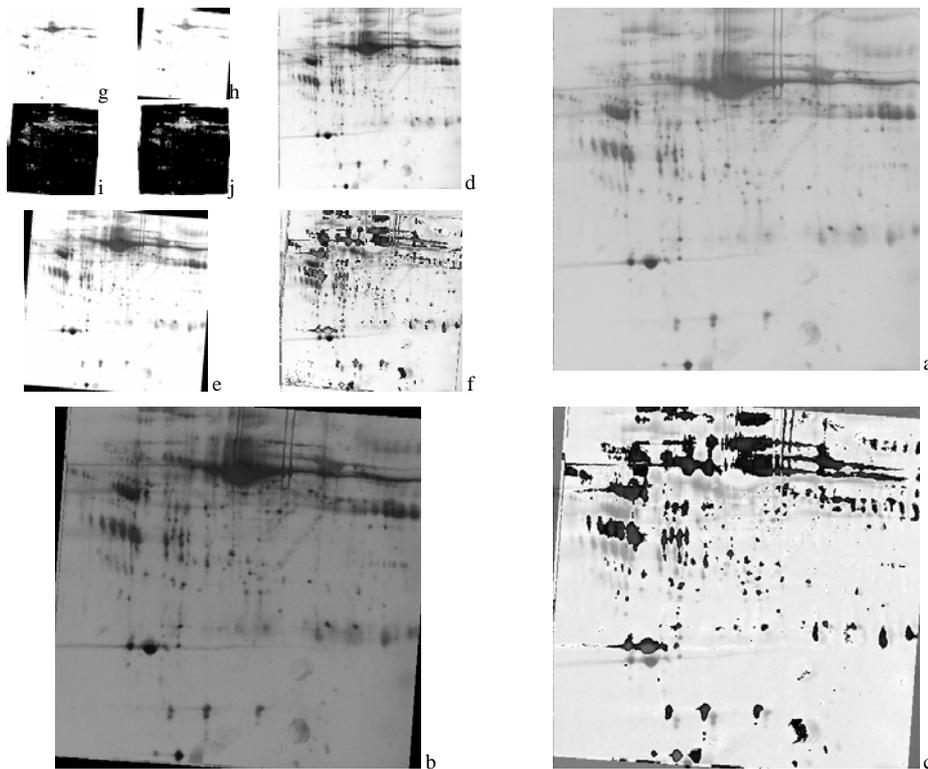


Figure 3. Intensity-based Hierarchical Gel Image Registration: (a), (b), and (c) are the 1<sup>st</sup> hierarchy reference image, study image, and difference image before registration; (d), (e), and (f) are the 2<sup>nd</sup> hierarchy reference image, study image, and difference image before registration; (g), (h), and (i) are the 3<sup>rd</sup> hierarchy reference image, study image, and difference image before registration; (j) is the difference image after affine registration of (g) and (h).

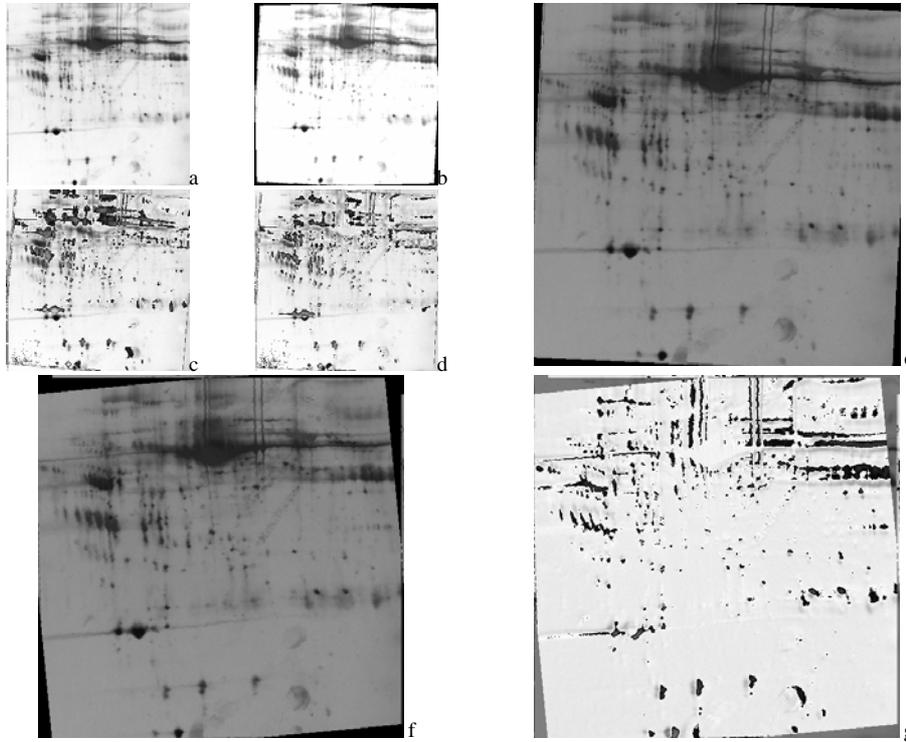


Figure 4. Intensity-based Hierarchical Gel Image Registration: (a) is the 2<sup>nd</sup> hierarchy reference image; (b) is the 2<sup>nd</sup> hierarchy study image after being initialized using the results of the 3<sup>rd</sup> hierarchy registration results (Figure 3 (j)); (c) is the difference image before registration being carried out in the 2<sup>nd</sup> hierarchy; (d) is the difference image after the registration in the same hierarchy; (e) is the study image of the 1<sup>st</sup> hierarchy after being initialized using the results of hierarchy 2; (f) is the final study image after the registration being carried out in the 1<sup>st</sup> hierarchy and (g) is the difference image after the intensity-based registration.

From the first stage registration experiments, we find out that the registration of very low resolution hierarchy images is not very helpful for the final registration results. It is mainly because we carry out our registration procedure in the low frequency domains, and excessive decomposition will lead to the loss of too much information. Therefore, in our experiments, we decompose the images into three hierarchies.

### 3.2 Landmark-based Elastic Registration of Human Blood Plasma Gel Images

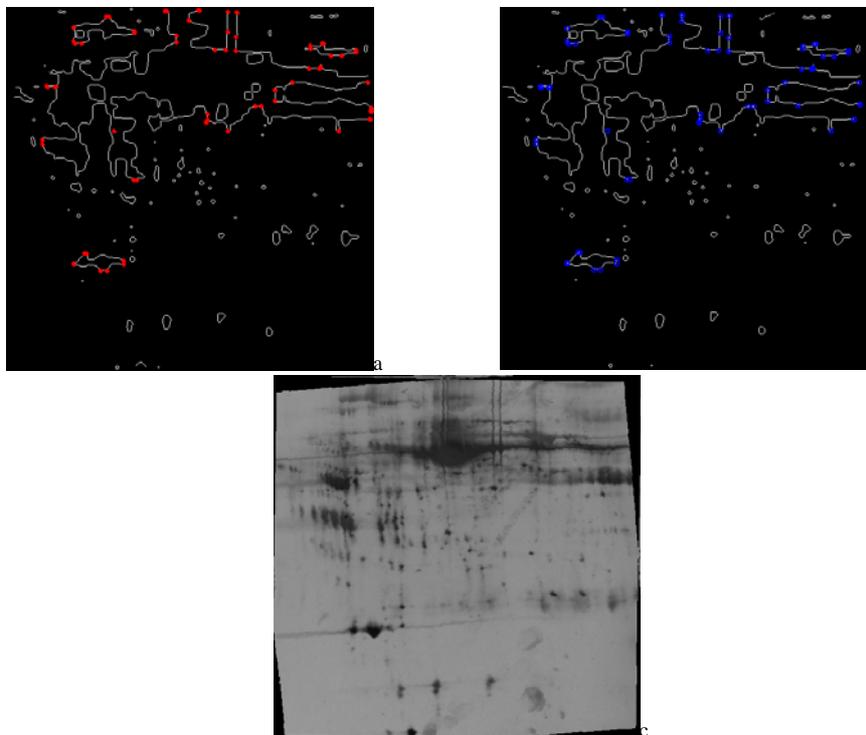


Figure 5. Landmark-based Elastic Gel Image Registration: (a) and (b) are the corresponding landmark points selected automatically in the reference edge image and the study edge image; (c) is the study image after elastic registration using TPS.

In our experiments of automatic landmark selection, to eliminate redundant and small-sized edges, we carry out morphological operations, closing and opening, before the edge extraction procedure is carried out.

#### 4 Conclusion

Image registration technique is fundamental and essential to the accurate and efficient analysis and comparison of protein sequence data. In this paper, a hybrid 2D gel protein image registration method is proposed to facilitate the investigation of protein expression. The first stage registration corrects global displacements using mutual information method and provides initial results for the second registration stage. Registration of the images from low resolution to high resolution enables fast registration convergence and therefore improves the computational efficiency. By introducing the wavelet-based hierarchical algorithm to the 2D gel protein image registration area, the algorithm uses the low-frequency subbands as searching spaces in each registration pyramid hierarchy and provides high registration accuracy. In the second registration stage, the landmark points are selected automatically and then thin-plate splines are used

to further improve the registration accuracy and performance by elastic registration. Because the proposed algorithm makes use of the merits of the intensity-based techniques and the feature-based registration methods, it can achieve high performance registration results automatically. The experiments demonstrate that the proposed algorithm has high registration accuracy and performance.

### Acknowledgments

This work is supported by ARC, UGC grants, HLJNSF and HLJE10541169.

### References

1. M. J. Dunn, A. Gorg. *From protein sequence to function*. In S.R. Pennington, M. J. Dunn (Eds.) *Proteomics*, BIOS Scientific Publishers, Oxford, 43-63, 2001.
2. Brown, L.G. A survey of image registration techniques. *ACM Computing Surveys* 24(4): 325-376, 1992.
3. T. Mäkelä, P. Clarysse, O. Sipilä, N. Pauna, Q. C.,Pham, T. Katila, and I. E. Magnin. A review of cardiac image registration methods. *IEEE Transaction on Medical Imaging*, 21(9): 1011-1021, 2002.
4. K. Rohr. Elastic registration of multimodal medical images: A Survey. *Auszug aus: Kunstliche Intelligenz, Heft3*: 11-17, 2000.
5. P. B. Hojte and X. X. Wang. Registering electrophoresis images for bioinformatics study of protein. *ICME'03*, 3:465-468, 2003.
6. X. Y. Wang, D. D. Feng and H. Hong. Novel elastic registration for 2-D medical and gel protein images. *APBC'03*, 19: 223-226, 2003.
7. Z. Smilansky. Automatic registration for images of two-dimensional protein gels. *Electrophoresis*, 22:1616-1626, 2001.
8. S. Veeseer, M. J. Dunn and G-Z. Yang. Multiresolution image registration for two-dimensional gel electrophoresis. *Proteomics* 2001, 1: 865-870, 2001.
9. S. Mallat, and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Inform Theory*. 38: 617-643, 1992.
10. A. Collignon, F. Maes, D.Delaere, D.Vandermeulen, P.Suetens and G.Marchal. Automated multi-modality image registration based on information theory. In: *Information Processing in Medical Imaging*, Y. J. C. Bizais (Eds.) 263-274,1995.
11. F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(6):567-585, June 1989.
12. M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.*, 7: 155-163, 1964.
13. <http://kr.expasy.org>.