Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 26 (2008) 1000–1011
www.elsevier.com/locate/imavis

Optimised coupling of hierarchies in image registration

J. Wensch [a,*], A. Gerisch [b], S. Posch [c]

[a] Institut für wissenschaftliches Rechnen, Fachrichtung Mathematik, TU Dresden, 01062 Dresden, Germany
[b] Institut Mathematik, Martin-Luther-Universität Halle-Wittenberg, Postfach, 06099 Halle (Saale), Germany
[c] Institut für Informatik, Martin-Luther-Universität Halle-Wittenberg, Postfach, 06099 Halle (Saale), Germany

Received 27 February 2006; received in revised form 21 September 2007; accepted 2 November 2007

Abstract

Image registration algorithms rely on multilevel strategies in order to improve efficiency and robustness. Hierarchies in image resolution, the underlying grids for spline-based transformations, as well as the regularisation parameters are used. This paper deals with the optimisation of the coupling of these hierarchies.

An image registration procedure – suitable for 2D polyacrylamide gel electrophoresis images – using piecewise bilinear transformations and an intensity based objective function with a regularisation term based on the elastic deformation energy is described. The resulting nonlinear least squares problem is solved by the Gauss–Newton method.

Techniques reminiscent of dynamic programming are used to optimise the coupling of hierarchies in image and transformation resolution. Besides using these techniques to devise an advantageous fixed coupling of both hierarchies, we favour incorporating the dynamic programming ideas into the final registration algorithm. This leads to an adaptive and streamlined approach.

Numerical experiments on 2D-PAGE images show that the adaptive registration algorithm is much more reliable than the same algorithm with a fixed coupling of hierarchies. The proposed optimisation procedure for the coupling of hierarchies presents a valuable tool to optimise other registration algorithms.

Keywords: Elastic registration; Image hierarchies; Transformation hierarchies; Coupling of hierarchies; Dynamic programming; 2D-PAGE images

1. Introduction

Image registration is one of the most challenging tasks in image processing. The computation of a transformation that maps a given template image to a reference image requires the solution of a high-dimensional nonlinear minimisation problem. Efficiency and robustness of suitable optimisation algorithms are enhanced by multi-resolution techniques. Hierarchies in image and transformation resolution are used. We present here a general approach to couple multiple, in our case two, hierarchies based on techniques from dynamic programming. The efficiency of the approach is illustrated using the example of a registration algorithm based on piecewise bilinear transformations and elastic regularisation applied to 2D-PAGE (polyacrylamide gel electrophoresis) images.

The analysis of 2D-PAGE image pairs, or more general sets of image groups, aims at identifying corresponding spots (detection and assignment) and discovering differences in the spot pattern. A typical and widely used first step in performing this analysis is the registration of image pairs. The main objective of this first step is to remove those distortions in the gel images which are the result of uncontrolled experimental side conditions and of the handling of gels (*e.g.* the scanning procedure). A rough registration that leads to substantial overlap of corresponding spots is sufficient for this step. A particularity at 2D-PAGE

* Corresponding author. Tel.: +49 3319771365; fax: +49 3319771001.
E-mail addresses: joerg.wensch@tu-dresden.de (J. Wensch), gerisch@mathematik.uni-halle.de (A. Gerisch), posch@informatik.uni-halle.de (S. Posch).

0262-8856/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.imavis.2007.11.002

images is that spots are very local features and hence, in a registration of two images, spots may have no initial overlap at all.

A general overview on image registration is given in [25]. The authors of [5] review variational approaches with emphasis on regularisation and computational techniques, whereas [16] investigates variational approaches, viscous flow, and several regularisers in depth. The objective function of the registration problem may be either intensity based or landmark based. Landmark based registration requires either manual setting of landmarks [20] or feature detection [24,18]. A framework for the combination of landmark based and intensity based registration is given in [7]. Aiming at automatic registration procedures, intensity based registration is a natural approach. For images captured by the same machinery the L_2-distance of the images is a widely used choice as distance measure, see [14,11,13,15,12,16,23,5,24]. Otherwise, statistical approaches, based on mutual information, see [4], can be used.

Non-parametrised registration leads to ill-posed problems which have to be regularised, typically using a penalty term in the objective function. The elastic potential of the transformation is a commonly used penalty term [4,11,24], furthermore curvature [12], and also Tikhonov stabilisation [15]. In optical flow registration, viscosity serves as regularisation parameter [16,4]. For a comprehensive overview on regularisers, see [16]. In contrast, parametrised registration uses a finite-dimensional function space, where regularisation is not mandatory, see *e.g.* [13,18,20,23], however, with an increasing number of parameters, a regularisation is highly advisable.

Optimisation algorithms typically are based on derivative/gradient information of the objective function. The reader should note that features (spots) that do not overlap in the two images to register, do also not contribute to the gradient of the objective function and hence are not taken into account in the optimisation process. The use of image pyramids – a hierarchy of different resolutions of each image – solves this problem, see [13,18,15,23,10,22]. Furthermore, hierarchies in transformation classes are frequently used to improve computational efficiency, see [4,21,20,23,13]. For instance, piecewise bilinear transformations can be defined on increasingly finer grids. Also, in the case of non-parametric transformations, the smoothness controlled by a regularisation parameter can be incorporated in a hierarchy [11,12].

Hierarchies are coupled in several approaches. A hierarchy of piecewise bilinear transformations together with a hierarchy in the landmarks is used by Salmi et al. [20] for the registration of gel images. However, they give no comments on the relation between the hierarchies. A fixed coupling between image resolution and regularisation parameter is used by Gustafsson et al. and Rogers et al. [10,18]. Haber and Modersitzki [11] use a grid hierarchy to estimate suitable regularisation parameters by line search. The case of the coupling of a hierarchy of images at different resolutions with transformations defined on a hierarchy of underlying grids is used in [23,22]. Veeser et al. [23] use an image pyramid and piecewise bilinear transformations (linear B-splines) without a regularisation term. Sorzano et al. [22] use cubic B-splines to generate image hierarchies as well as transformation hierarchies. Their regulariser is based on gradients of divergence and vorticity of the displacements. Nevertheless, both approaches use a fixed coupling between image and transformation resolution.

To the best of our knowledge, there has not yet been devised a strategy to couple image and transformation hierarchies efficiently. This paper aims to close this gap by supplying a dynamic programming based approach. This approach is applied to the registration of gel images using the Euclidean distance measure, image resolution pyramids, a hierarchy of bilinear transformations and the elastic potential as regulariser.

The remainder of this paper unfolds as follows. In Section 2, we review the task of elastic image registration, without any hierarchies, using piecewise bilinear transformations. Then, in Section 3, we introduce transformation grid and image resolution hierarchies and devise a strategy for their coupling. Here, we use ideas from dynamic programming. This completes the definition of our elastic image registration algorithm. In Section 4, we apply the devised algorithm to the registration of pairs of sample gel images and discuss its performance. These experiments lead to an additional streamlining of the algorithm. Finally, Section 5 ends the paper with a summary and conclusions.

2. Elastic image registration

2.1. Basic terms of image registration

For the task of image registration we are given two equalised grey-scale images, a fixed reference image I_R defined on $\Omega_R \subset \mathbb{R}^2$ and a deformable template image I_T defined on $\Omega_T \subset \mathbb{R}^2$, and we aim to identify a continuous and invertible transformation M, which transforms approximately the template image into the reference image, *i.e.* $I_R \approx M(I_T)$. We assume that the images are defined by grey-scale values on an integer lattice (pixel values) and by bilinear interpolation in between. As usual we define $M(I_T)(x,y) := I_T(M(x,y))$ where M is a mapping from Ω_R in $\mathbb{R}^2$. Note, that the reference image and the transformed template image $M(I_T)$ overlap on $\Omega(M) := \Omega_R \cap M^{-1}(\Omega_T)$.

The image registration task is posed as a minimisation problem, where the objective function

$$f[M] := E_1(M;\, I_R, I_T) + E_2(M) \tag{1}$$

has to be minimised over the class $\mathcal{M}$ of admissible transformations. Here, E_1 is a measure of the distance between I_R and $M(I_T)$ and E_2 serves as a penalty term to regularise the problem. It is easily verified that the registration

problem without regularisation is ill-posed. To this end, note the fact that all transformations M that have the isolines of I_T as invariant sets yield the same $M(I_T)$.

2.2. Measure of distance $E_1(M; I_R, I_T)$

Our measure of distance between I_R and $M(I_T)$ is based on the L_2-norm of their difference on their domain of overlap $\Omega(M)$. We weight this difference with $1/|\Omega(M)|$ to give preference to transformations with large domain of overlap. Considering only the pixel positions $\Omega(M) \cap \mathbb{Z}^2$ on the domain of overlap, we then define

$$E_1(M; I_R, I_T) := \frac{1}{|\Omega(M) \cap \mathbb{Z}^2|} \sum_{(i,j) \in \Omega(M) \cap \mathbb{Z}^2} (M(I_T)(i,j) - I_R(i,j))^2, \quad (2)$$

where $|\Omega(M) \cap \mathbb{Z}^2|$ is the number of grid points (pixels) contained in $\Omega(M)$.

We remark, that there are several alternatives. Mutual agreement – the scalar product of I_R and $M(I_T)$ – is used as well, but it is strongly correlated to the L_2-norm. We believe, that in the case of equalised images, taking the L_2-norm of the difference will lead to appropriate results. For non-equalised images or images captured by different imaging modalities, more sophisticated concepts based on statistical models or topological analysis can be applied.

2.3. The class of transformations $\mathcal{M}$

We denote by $\mathcal{M}_{pbl}(L_x, L_y)$ the classes of continuous and piecewise bilinear transformations on the rectangular domain Ω_R which is partitioned into a grid of $L_x \times L_y$ congruent rectangles. In order for such a transformation to be invertible on Ω_R any rectangle of the partition of Ω_R has to be mapped to a convex, non-degenerate quadrilateral. The subclass of invertible transformations is denoted by $\mathcal{M}^*_{pbl}(L_x, L_y)$.

Each transformation $M \in \mathcal{M}_{pbl}(L_x, L_y)$ is parametrised by a vector $U \in \mathbb{R}^{2(L_x+1)(L_y+1)}$ of displacements in the grid points; $M = M[U]$. Furthermore, with each transformation $M[U]$ we associate a displacement field

$$u(x,y) := M[U](x,y) - (x,y)$$

2.4. The elastic penalty term $E_2(M)$

Besides the intensity difference E_1, our objective function $f[M]$ includes a penalty term E_2. The penalty term in elastic image registration is the elastic energy associated with the transformation M. This term acts as a regulariser for the otherwise ill-posed registration problem. As a desired side effect, transformations close to pure translations are given preference.

Employing linear elasticity theory for a plate under plane strain [2], we obtain the elastic energy of $M[U]$ from its displacement field u as a positive semidefinite bilinear form

$$\mathcal{E}(M[U]) = \int_{\Omega_R} \mu Du : Du + \lambda/2(\nabla \cdot u)^2 \, dx \, dy \geqslant 0, \quad (3)$$

where $Du = (\nabla u + \nabla u^T)/2$ and $\mu, \lambda > 0$ are the Lamé constants. These constants characterise the elastic material being used. For transformations $M[U] \in \mathcal{M}_{pbl}(L_x, L_y)$, the elastic energy can be written as $\mathcal{E}(M[U]) = 1/2 U^T K U$, where K is called the stiffness matrix.

We define the term $E_2(M) := \mathcal{E}(M[U])/|\Omega_R|$ as the elastic energy of $M = M[U]$ scaled by the size of the domain Ω_R. Scaling by the size makes the choice of regularisation parameters independent of the image size.

2.5. Numerical solution of the nonlinear least squares problem

With the choices made in the previous subsections, the minimisation problem for the image registration task is a constrained nonlinear least-squares problem. The constraint derives from the fact that we require invertible transformations $M[U]$. This implies that the set of admissible vectors U is an open set and hence the constraint should never become active. For this reason, we treat the problem as an unconstrained minimisation problem. We opted for the Gauss–Newton method for the solution of unconstrained nonlinear least-squares problem, see *e.g.* [8]. The line search strategy (Goldstein–Armijo) is adjusted to ensure that all computed transformations are invertible. Furthermore, the Jacobian is regularised by a threshold in the QR-decomposition. For alternative nonlinear minimisation schemes, see [11,15,5].

The selection of a suitable predictor (starting value) $U^{(0)}$ for the Gauss–Newton method is crucial for its efficiency and is the main issue discussed in Section 3.

3. Optimised coupling of hierarchies in image and transformation resolution

In Section 2, we considered an image registration algorithm applicable to an equalised image pair (I_R, I_T) of fixed resolution. Also the class of transformations $\mathcal{M}^*_{pbl}(L_x, L_y)$ was fixed with specified values L_x and L_y for the underlying grid. The algorithm is based on the iterative Gauss–Newton method for the minimisation of the objective function $f[M]$ – a highly nonlinear problem. For the efficiency and reliability of the procedure the choice of the predictor, *i.e.* starting value for the iteration, is critical. A well-chosen predictor reduces the number of iterations necessary for convergence and therefore the computational effort.

The objective of this section is to devise a strategy, such that suitable predictors can be provided for the Gauss–Newton method, as described in Section 2, in order to guide the overall algorithm to a global solution of our optimisation problem. The key to this strategy is to employ two hierarchies: a pair of image pyramids (one pyramid for

each of the images I_R and I_T) containing the same image at different resolutions and a hierarchy of transformation classes $\mathcal{M}^*_{\mathrm{pbl}}(L_x, L_y)$, where the underlying grid, and hence the number of free parameters, varies. These hierarchies are described in the following two subsections. A suitable coupling of both hierarchies is subsequently developed employing ideas from dynamic programming.

We remark that the coupling of the two hierarchies may be fixed or adaptive. For a fixed coupling the dynamic programming procedure is run on a subset of the image class, the best coupling is determined and used for subsequent registration tasks. Adaptive coupling uses the dynamic programming technique for each registration task.

3.1. The image pyramids

The choice of the search direction in the Gauss–Newton method is based on a linearisation of the functional $f[M]$. In order to globalise the information contained in the linearised functional the utilisation of image resolution pyramids is indispensable, see [13,18,23,10]. Multi-resolution approaches based on B-splines are used in [15,22]. We use image pyramids based on the algorithm of Burt–Adelson [3]. We note that when the image resolution is decreased by sub-sampling then it has to be low-pass filtered in advance to avoid aliasing, see [1].

Given an image $I : \Omega \to \mathbb{R}$ ($I \in \{I_R, I_T\}$), the image at level l has a resolution decreased by a factor of 2^l in each direction. The image at level $l = 0$ is defined to be the original image, $I^0 := I$. The image at level $l + 1$ is computed from the image at level l via Gaussian filtering (G_σ) and sub-sampling (S_2)

$$I^{l+1} := S_2(G_\sigma(I^l)). \tag{4}$$

The sampling operator S_2 simply selects every second pixel in each direction. The filter operator G_σ is a finite impulse response filter working with a 5×5 window. It is constructed as a tensor product of Gaussian distributions in one dimension with mean zero and variance σ^2. The filter is therefore separable. Using indices $-2 \leqslant i, j \leqslant 2$ in the 5×5 window, the convolution kernel is given by

$$w_{ij} = d \exp(-(i^2 + j^2)/(2\sigma^2)), \tag{5}$$

where the normalisation factor d is chosen such that $\sum_{i,j} w_{ij} = 1$.

A good choice for the parameter σ in G_σ is obtained by the sampling theorem. Note that signals $\exp(\mathrm{i}\omega x)$ of frequency ω will be damped by a factor $\exp(-(\omega^2\sigma^2/4))$. With $\sigma = 1$, the critical frequency $\omega = \pi$ is damped by a factor $\exp(-\pi^2/4) \approx 0.08$.

We will start the global minimisation procedure to be devised on a suitable image level $l_{\max}$ (*i.e.* a suitably low resolution) selected such that at this resolution important global information is captured only. Then the resulting transformation of the Gauss–Newton method applied to the image pair $(I_R^{l_{\max}}, I_T^{l_{\max}})$ and started with the identity transformation as predictor results in a transformation which is a good candidate for a predictor on the next lower image level $l_{\max} - 1$ (next finer resolution) and so on.

3.2. The transformation hierarchy

The piecewise bilinear transformations incorporate a natural hierarchy defined below. For our application we choose the class parameters L_x and L_y equal and as powers of two: $L_x = L_y := 2^k$. The value $k = 0, 1, 2, \ldots$ is called the grid level of the transformation. Then the classes of transformation $\mathcal{M}^*_{\mathrm{pbl}}(2^k, 2^k)$ form the hierarchy

$$\mathcal{M}^*_{\mathrm{pbl}}(1,1) \subset \mathcal{M}^*_{\mathrm{pbl}}(2,2) \subset \mathcal{M}^*_{\mathrm{pbl}}(4,4) \subset \ldots \tag{6}$$

where coarse transformations are extended to finer grids by bilinear interpolation.

Similar as for the image levels, the resulting transformation of grid level l of the Gauss–Newton method provides a good candidate for a predictor on the next higher grid level $l + 1$ where computations are more expansive.

3.3. Optimised coupling by dynamic programming techniques

Dynamic programming [6] is a technique applied in combinatorial optimisation. Roughly speaking, the solution of complex optimisation problems is computed from the solution of smaller subproblems. The main ideas of dynamic programming are the decomposition of the problem into overlapping subproblems and the memoisation of the solution of previously solved subproblems in a dynamic programming array. The data stored in the array are then used to solve other subproblems, eventually the problem one is interested in. So the dynamic programming array is built up gradually. If the subproblems possess an optimal substructure, then using the dynamic programming approach will yield optimal results. A typical example of the dynamic programming approach is for instance the Needleman–Wunsch algorithm for global string alignment, see [9]. In contrast to that, in our application we have to assume that a better registration on lower resolutions will lead to a better (or faster) registration on finer resolutions – there is no proof. Although it lacks a strict optimal substructure property the approach seems to be a good heuristic.

We next describe how the ideas from dynamic programming can help to solve the image registration problem efficiently by connecting the image and transformation hierarchies. For each image level l and each grid level k, we denote by $M^{k,l}$ an optimal transformation from class $\mathcal{M}^*_{\mathrm{pbl}}(2^k, 2^k)$ for the image pair (I_R^l, I_T^l). The transformations $M^{k,l}$ will form the entries of the dynamic programming array. The final goal of the algorithm is the computation of an optimal transformation $M^{k_{\max}, l_{\min}}$ for a pair of images (I_R, I_T). In subsequent steps we use minimisers from coarser grids and/or coarser images as predictors for Gauss–Newton iterations on finer grids and/or finer images (starting with the identity transformation on the coarsest levels). The natural choice for predictors for the

computation of entry $M^{k,l}$ of the dynamic programming array among all previously computed entries/minimisers are $M^{k-1,l}$, $M^{k-1,l+1}$, and $M^{k,l+1}$ (if they are entries of the array). Because we have experienced that for a too large maximum image level $l_{\max}$, the transformations computed on the coarsest image levels are sometimes of no use as predictors for finer image levels, the identity transformation I is used as an additional predictor on all levels. By that, the choice of $l_{\max}$ becomes uncritical provided it is large enough.

The ideas from dynamic programming may now be used in two different ways

(a) run the algorithm on a subset of the image class[1] to identify a promising pathway through the dynamic programming array that is fixed afterwards for the registration process of the whole image class or
(b) use the algorithm above (with some modifications) to determine the path adaptively for each image pair.

We have followed both, the preprocessing (a) and the adaptive (b), approaches and they are described in further detail in the following subsections.

3.4. The preprocessing approach

At each level (k,l) we have a set $S^{k,l}$ of up to 4 predictors available. Besides the identity we use transformations from the coarser levels $(k,l+1),(k-1,l),(k-1,l+1)$ whenever these levels exist and the transformations have been computed successfully. We execute Gauss–Newton iterations for each of the up to four predictors $P \in \{I, M^{k,l+1}, M^{k-1,l}, M^{k-1,l+1}\}$ and compare the resulting transformations $M(P)$ with respect to a predictor selection criterion $H = H(M(P))$. The predictor leading to the smallest value of that criterion defines the new transformation $M^{k,l}$.

3.4.1. Predictor selection criterion

The idea behind the predictor selection criterion $H(M(P)) := T_\tau(M(P))$ with parameter $\tau \geqslant 0$ is to combine a high quality of the registration with reasonably fast computation times. We choose the predictor that leads to minimal total computation time among those predictors that lead to almost optimal quality (small values of $f[M]$).

To this end, define $T(M(P))$ as the total time spent to compute the transformation $M(P)$. Thus $T(M(P))$ is given recursively by

$$T(I) = 0,$$
$$T(M(P)) = T(P) + T_{P \to M(P)},$$

where $T_{P \to M(P)}$ is the computing time for $M(P)$ with the Gauss–Newton method when P is used as predictor.

[1] In our application, an *image class* is a large group of gel images from similar experiments which share certain qualitative properties.

For $\tau \in \mathbb{R}, \tau \geqslant 0$, we define $T_\tau(\cdot)$ by

$$T_\tau(M(P)) := \begin{cases} T(M(P)) & \text{for } f[M(P)] \leqslant (1+\tau) f_{\min} \\ \infty & \text{otherwise} \end{cases} \text{ with}$$

$$f_{\min} := \min_{\widetilde{P} \in S^{k,l}} f[M(\widetilde{P})].$$

For decreasing values of τ, criterion $T_\tau(\cdot)$ should lead to better registration results at, in general, increasing computational effort. Note that T_0 corresponds to the objective function, whereas T_∞ corresponds to the computing time as predictor selection criterion.

3.4.2. Building the dynamic programming array

Starting with the identity as the only predictor available on level $(k_{\min}, l_{\max})$, the array is filled row by row. When an optimal predictor on the finest levels is obtained, the optimal path through the array is obtained by backtracking.

Remember that the registration procedure is supposed to be applied to a large number of images all stemming from the same application. In this case it can be expected that there exist a few paths through the image and grid hierarchies which lead to good results for most image pairs. The algorithm described above can be used to identify these paths when it is executed on a characteristic subset of the images. This leads to the identification of a suitable coarsest image level $l_{\max}$ for the application and, more importantly, the sets of suitable predictors $S_H^{k,l}$ can be streamlined and some transformations $M_H^{k,l}$ need not be computed at all because they will never be used as predictors. In the above notation the subscript H indicates that this streamlining depends on the choice of the predictor selection criterion H. With this *once* precomputed knowledge for an image class, the adjusted algorithm will be much more efficient when applied to the remaining image pairs of the class. This optimisation procedure is illustrated on sample images in Section 4.

3.5. The adaptive and the streamlined adaptive approach

In a pure adaptive approach the complete dynamic programming array is computed for each image pair using all available predictors (up to four) in order to obtain the transformation on the finest grid and image levels. Since in the adaptive approach we are only interested in the registration quality (the computation time has been spent anyway), we use $H = T_0$ as predictor selection criterion in that case.

Especially on the finer image and grid levels, the adaptive approach leads to very time consuming computations. Therefore it appears natural to identify promising predictors for the grid and image level combinations on these finer levels by making use of results from the preprocessing approach, ultimately leading to a streamlined adaptive technique. We stress however, that investing small additional amounts of computing time on coarser image and grid levels, by using the (pure) adaptive approach there,

will improve the quality of the overall registration and reduce the total computational effort as well.

4. Results

4.1. Sample images

Image set A is the set of four images shown in Fig. 1. They serve as samples for the preprocessing approach. These 2D-PAGE images result from experiments on transgenic plants which are transformed to produce a bacterial avirulence gene. The resistance reaction leads to the apoptosis of cells and finally dead leaves. The control group (bottom row in Fig. 1) lacks the avirulence gene and thus the corresponding resistance reaction. All images have roughly the same size of 1300×1000 pixels. Note that for a typical image height of 200 mm, this corresponds approximately to a resolution of 5 pixels per mm.

Image set B consists of five images taken from the Bielefeld 2D-PAGE *Escherichia coli* database [17]. The image size is roughly 700×1000.

Image pair C is a pair of 2D-PAGE images from experiments with the plant *Arabidopsis thaliana* which has been exposed to cadmium ions. Under these conditions the synthesis of metal binding peptides is triggered, see [19] for details. Image size is as in set A. We consider all these images to belong to the same image class.

4.2. The preprocessing approach

We apply the preprocessing approach developed in Section 3 to the image registration algorithm described in Section 2 using image set A.

The goal of the numerical experiment is to determine an optimised coupling of the transformation and image hierarchies for the given image class. Furthermore, we will compare the registration results obtained using predictor selection criterion T_τ for three values of τ and also for several strategies with respect to the maximum number of iterations in the Gauss–Newton process. More precisely, we consider

- The predictor selection criteria $H = T_\tau$ for $\tau \in \{0, 0.05, 0.1\}$, see Section 3.4.
- Four strategies for the maximum number of iterations in the Gauss–Newton process: when applying the Gauss–Newton iterations to the set of predictors $S_H^{k,l}$, we fix a maximum number of iterations N_P. When the decision on the best predictor has been made, we continue the Gauss–Newton process with the selected predictor and a possibly higher number of maximal iterations N_F for the final best transformation on level (k, l). We consider fixed numbers for N_P, N_F as well as numbers $N_{GN}(k, l)$ depending on the grid/image level. These numbers $N_{GN}(k, l)$ vary from 20 for coarse grids/images to 4 for fine grids/images. Here, we consider the following strategies:

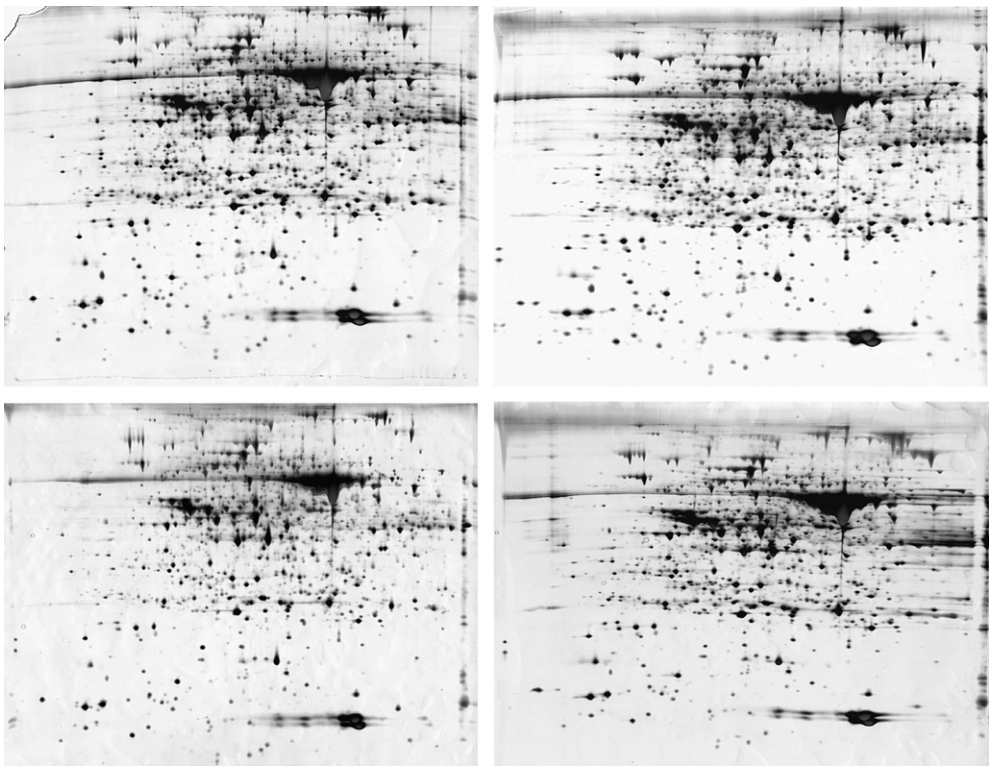

Fig. 1. The four sample gel electrophoresis images from set A used in the preprocessing approach. Upper row: images I_0 (left) and I_1 (right), lower row: images I_2 (left) and I_3 (right).

1. $N_P = 1$, $N_F = N_{GN}(k,l)$.
2. $N_P = N_{GN}(k,l)$, $N_F = N_{GN}(k,l)$.
3. $N_P = 1$, $N_F = 1$.
4. $N_P = 3$, $N_F = 3$.

The ratio of the Lamé constants $\mu = 10$ and $\lambda = 20$ for the elastic material has been chosen such that the resulting Poisson number is $\nu = 1/3$.

For each of the different image pairs, criteria, and strategies we compute the optimal transformations $M^{k,l}$ for $0 \leqslant k \leqslant k_{max}$ and $l_{min} \leqslant l \leqslant l_{max}$ based on the full set of predictors as defined in Section 3. This means that $12 \cdot 3 \cdot 4 = 144$ dynamic programming tables $(M^{k,l})$, $0 \leqslant k \leqslant k_{max}$ and $l_{min} \leqslant l \leqslant l_{max}$, are computed. We have used $k_{max} = 4$ as maximum grid level and $l_{max} = 6$ as coarsest and $l_{min} = 2$ as finest image level. From each table we can extract the path taken from the identity transformation to the final transformation $M^{k_{max},l_{min}}$ by backtracking. We display in Fig. 2 the 36 paths for the first strategy to limit the number of iterations in the Gauss–Newton process, *i.e.* $N_P = 1$ and $N_F = N_{GN}(k,l)$. The rows in each diagram correspond to different image levels l, whereas the columns correspond to the grid levels k (denoted by the corresponding number of grid cells 2^k in each dimension, for convenience).

We see, that there is no clear preference for a single path. Nevertheless, several conclusions can be drawn.

1. For $H = T_0$ there is a unique best predictor at the finest levels $(k,l) = (4,2)$ of both hierarchies, namely $P^{4,2} = M^{3,2}$. For the criteria $H = T_\tau$, $\tau = 0.1$ or 0.05, this predictor is preferred, too.
2. For the initial grid level $k = 0$ is a promising choice. The initial image level l_{max} (coarsest resolution) should be four, five or six. Since the computation on coarser resolutions are rather fast, we advocate $l_{max} = 6$.
3. The approaches with $H = T_{0.1}$ and $H = T_{0.05}$ give very similar results.
4. For $H = T_0$ the diagonal predictor $P^{k,l} = M^{k-1,l+1}$ is preferred in only a few cases, whereas for $H = T_{0.1}$ and $H = T_{0.05}$ the diagonal predictor is preferred frequently.

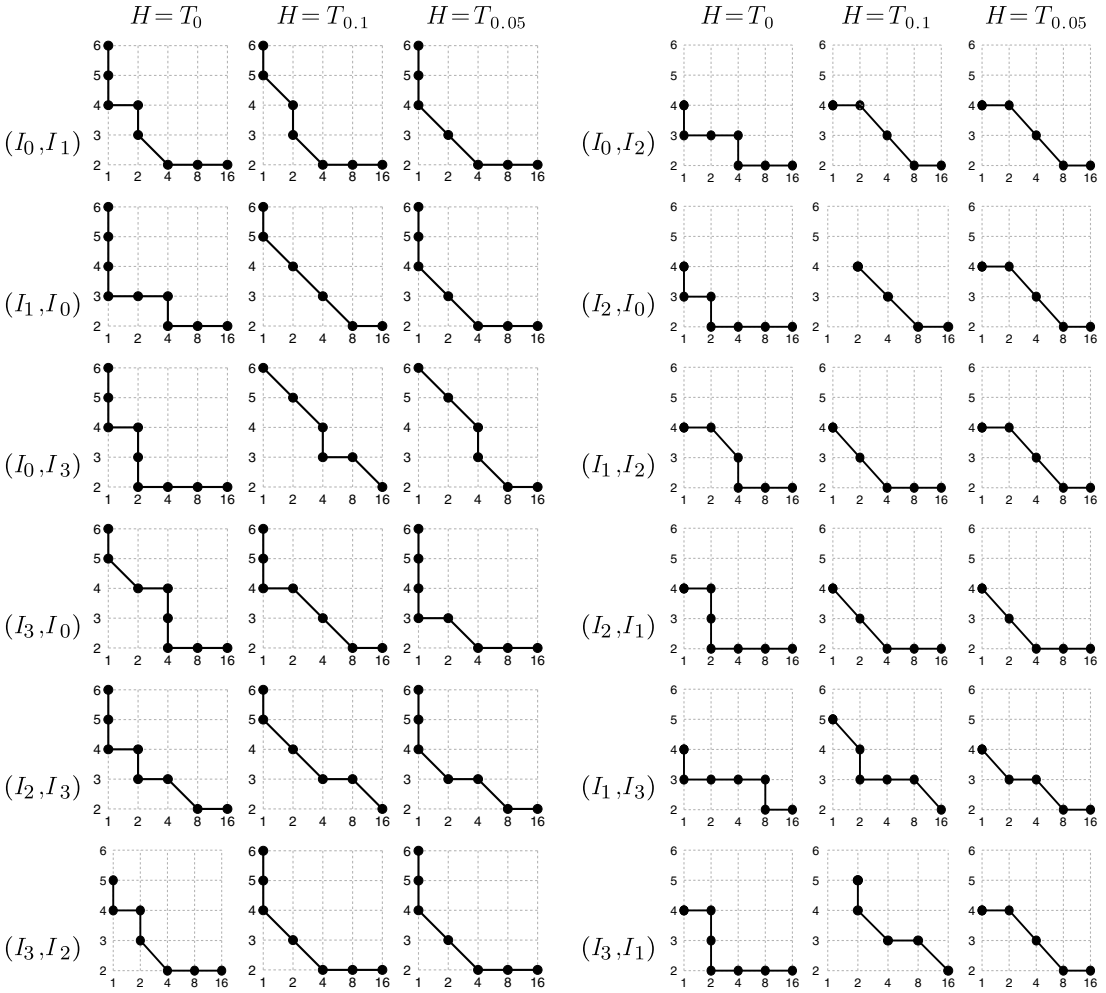

Fig. 2. Each of the 36 subplots shows the path selected between the final transformation $M^{k_{max},l_{min}}$ in the bottom right corner through its optimal predictor and so on to a transformation for which the identity transformation is the optimal predictor. Axes are grid size 2^k (horizontal) and image level l (vertical). The predictor selection criterion is given in the column heads and the image pair to the left of the plots. See text for more details.

5. Some levels (k,l) are never or rarely used in optimal paths: level $(k,l)=(0,2)$ is never used; levels (k,l) with $k+l=6$ are rarely used (except for the final level $(4,2)$ and the initial level (0,6)), and levels with $k+l>6$ are never used.

We have displayed all the results in condensed form in Fig. 3. In each picture we see an overlay of 12 paths corresponding to the 12 image pairs. The thickness of the edges corresponds to the frequency a predictor is used. The grey discs denote vertices where the identity is the selected predictor. The size of the disc corresponds to the frequency the identity has been chosen. The first column in Fig. 3 corresponds to the choice of maximum number of iterations in the Gauss–Newton process as used in Fig. 2, *i.e.* $N_{\mathrm{P}}=1,\ N_{\mathrm{F}}=N_{\mathrm{GN}}$. The other three columns correspond to the other choices of the maximum number of iterations in the Gauss–Newton process. The results are similar and the conclusions above remain true, although the original choice ($N_{\mathrm{P}}=1,\ N_{\mathrm{F}}=N_{\mathrm{GN}}$) seems to have the least variations in the optimal pathways.

The preprocessing process identifies $l_{\max}=6$ as a good choice for our gel images. It does not indicate a clear preference for a fixed path through the dynamic programming array. Nevertheless, the results obtained can be used to identify promising predictors for the adaptive approach.

4.3. The adaptive approach

The results from the preprocessing approach suggest not to fix a unique predictor for the Gauss–Newton process for coarse image/grid levels. For each level (k,l), we instead restrict the set of predictors to those that have been chosen at least once in the preprocessing step when predictor selection criterion $H=T_0$ has been used. The resulting coupling is visualised in Fig. 4. The selected predictors are represented by bold lines and, for the identity predictor, by grey discs in Fig. 4.

We note that only a subset of all transformations $M^{k,l}$ ($0\leqslant k\leqslant k_{\max}$ and $l_{\min}\leqslant l\leqslant l_{\max}$) is computed. On the finest image and grid level $(k,l)=(4,2)$, only one or two Gauss–Newton iterations are necessary to compute the final transformation. By using a unique predictor there we avoid computations for the other predictors – so the computational effort on the finest level (which is approximately 80% of overall runtime) is reduced by a factor of two at least.

Therefore our final optimised coupling procedure between the grid and image resolution hierarchies is streamlined (leading to computational efficiency) and at the same time adaptive (leading to a flexible algorithm).

In particular, the adaptivity in our coupling is in contrast to earlier work of others, see Section 1, where only

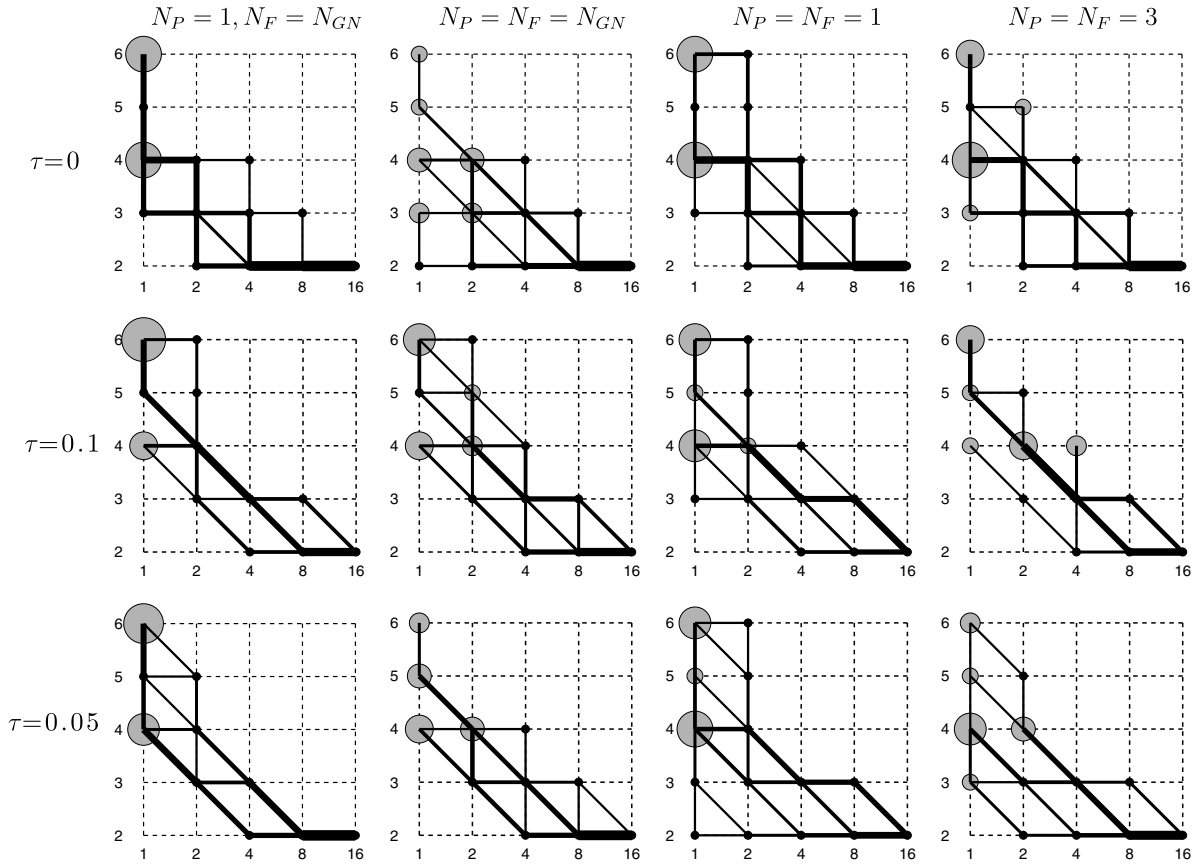

Fig. 3. Summary of selected paths for criterion $H=T_\tau$. Axes are grid size 2^k (horizontal) and image level l (vertical). Thickness of lines indicates how often a predictor is part of an optimal path. Thickness of circles indicates how often the identity has been chosen as predictor.

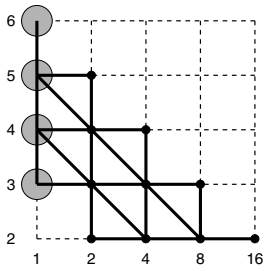

Fig. 4. The final optimised coupling between the grid and image resolution hierarchies. Axes are grid size 2^k (horizontal) and image level l (vertical).

a single fixed path is used to go from the identity predictor through various levels (k, l) to the desired final transformation at level $(k_{\max}, l_{\min})$.

We mention, that the algorithm performs well for all the different strategies for the selection of the maximum number of Gauss–Newton iterations. Nevertheless, when emphasis is put on the quality of the registration, the following choice turns out to be slightly favourable:

- We have opted for the optimisation criterion $H = T_0$ because we see it as advantageous to spend more time on (less expensive) coarse image and grid resolutions with the likely outcome of a better or faster registration on fine image and grid levels.
- For the maximum number of iterations in the Gauss–Newton procedure we have chosen $N_P = 1$ and $N_F = N_{GN}$ which leads to less variations in the pathways chosen.

We next demonstrate that with the streamlined adaptive approach, based on the streamlined coupling as visualised in Fig. 4, a considerable improvement over a static coupling of hierarchies, *i.e.* using a fixed path, can be achieved in registration quality.

We consider two fixed paths for the coupling. Using image levels $l = 2, \ldots, 6$ and grid sizes $2^k = 1, 2, 4, 8, 16$, we use a diagonal path through the dynamic programming array, *i.e.* we start with $l = 6$, $k = 0$ and increase image and grid resolution simultaneously up to $l = 2$, $k = 4$. This self-evident choice for the coupling of image and grid resolution by simultaneous refinement is also suggested by Sorzano et al. [22]. Our preprocessing approach suggests that the staircase-type path using $(l, k) = (6, 0), (5, 0), (5, 1), (4, 1), \ldots, (2, 3), (2, 4)$ is better. We compare therefore

FD the self-evident fixed coupling by a diagonal path,
FS the fixed coupling by a staircase path that is suggested from the preprocessing,
A the streamlined adaptive coupling with predictors from Fig. 4.

Comparison of the achieved average square differences E_1 (for convenience we display $\sqrt{E_1}$ where grey values range from 0 to 255) and the CPU times for the diagonal path (FD), the staircase path (FS) and the adaptive approach (A) are given in Tables 1 and 2. Particular results for each image pair from set A are given in Table 1. Results for pair C and statistics for set A and set B are given in Table 2. In Fig. 5 we visualise the results for image pair $(3, 1)$ from set A.

Comparing solely the achieved quality, the initial fixed coupling using a diagonal path can be improved considerably by using the staircase coupling that has been suggested by the preprocessing approach. The streamlined adaptive approach gives even better results than the improved fixed path. In the cases of the image set A and the image pair C, the streamlined adaptive approach is even faster than the fixed approaches. Nevertheless, in the case of image set B more computational work is necessary with approach A in order to achieve the results of higher quality.

Among the images in set A the differences between the three registration results seem to be smallest for image pair $(3, 1)$. Even for this pair, the difference between the three approaches is enormous. We have displayed difference images for the three couplings in Fig. 5. To visualise differences in a grey-scale image, we use the intensity function

Table 1
For each image pair of set A we display the quadratic mean of the image differences (with values ranging from 0 to 255) for the unregistered images, $\sqrt{E_1(Id)}$, and the registered images, $\sqrt{E_1}$, using the coupling with the fixed diagonal path (FD), the fixed staircase path (FS) and the streamlined adaptive approach (A)

Images	$\sqrt{E_1(Id)}$	$\sqrt{E_1}$			CPU time (s)		
		FD	FS	A	FD	FS	A
(0,1)	63.6	27.6	25.6	23.7	36	44	58
(0,2)	53.2	34.0	30.9	24.4	28	34	125
(0,3)	61.3	35.4	24.7	26.0	171	80	84
(1,0)	63.6	34.3	27.1	24.7	33	51	58
(1,2)	54.7	41.2	28.4	27.7	29	48	61
(1,3)	53.6	31.4	24.3	23.3	181	218	65
(2,0)	53.2	35.6	26.3	26.6	75	121	69
(2,1)	54.8	40.3	33.6	28.1	70	135	56
(2,3)	56.8	46.7	31.5	28.2	147	212	63
(3,0)	61.3	29.1	26.2	25.9	69	114	123
(3,1)	53.6	28.7	25.2	22.5	66	47	91
(3,2)	56.8	43.3	34.2	29.1	57	62	72

The last three columns give the corresponding CPU times.

Table 2
Values (image pair C) and mean and standard deviation σ (image sets A and B) of the quadratic mean of the image differences and CPU times

Images		$\sqrt{E_1(Id)}$	$\sqrt{E_1}$			CPU time		
			FD	FS	A	FD	FS	A
Set A	Mean	57.2	35.6	28.2	25.9	80	97	77
Set A	σ	4.1	6.1	3.5	2.1	55.1	64.4	24.2
Set B	Mean	22.0	20.7	18.0	15.1	17	28	37
Set B	σ	2.4	4.1	4.5	3.5	13.7	18.9	12.5
Pair C		47.6	33.5	25.0	23.2	54	62	46

See Fig. 1 for labelling.

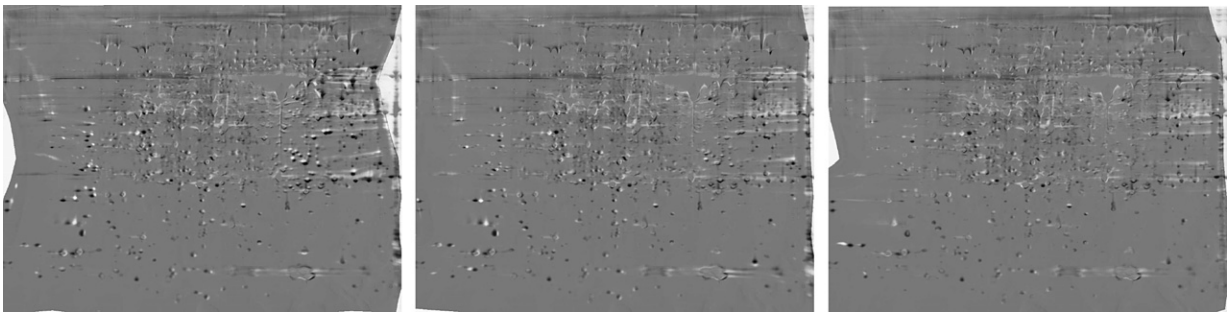

Fig. 5. From left to right: difference image for image pair (3,1) with the diagonal approach, the staircase approach and the adaptive approach.

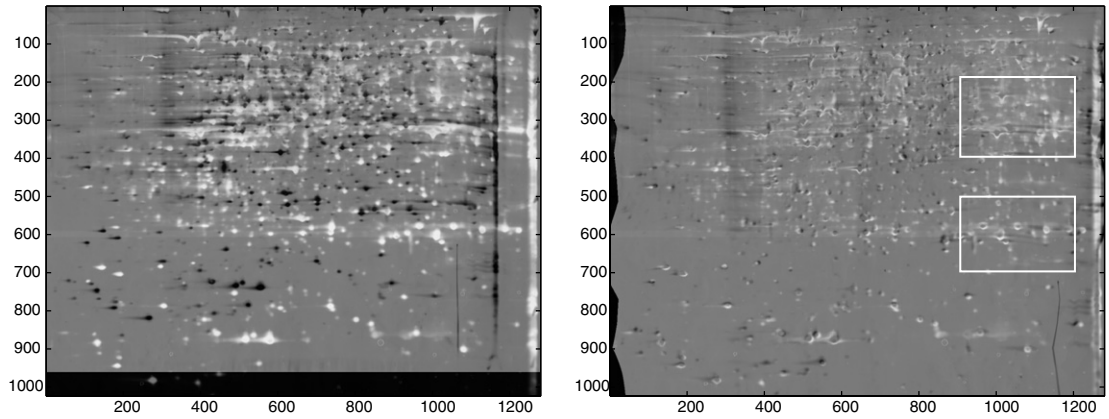

Fig. 6. The difference between the template and the reference image (before (left) and after registration) for image pair C.

$$I_{\text{diff}} := ((I_{\text{T}} - I_{\text{R}}) + 255)/2. \tag{7}$$

Hence, a medium grey in the difference image indicates no difference between reference and template image, a white area corresponds to a dark area (protein spot) in the reference image and a black area corresponds to a dark area (protein spot) in the template image. Although most of the spots have been registered, there is a considerable amount of unregistered spots to be seen in the left part of the difference images for the fixed coupling approaches, among which the improved staircase coupling gives better results than the original diagonal coupling.

We remark, that the adaptive algorithm spends approximately 20% of the total computing time on the computation of predictors and transformations $M^{k,l}$ that are not used for the computation of the final transformation $M^{k_{\max},l_{\min}}$. We consider this to be quite a mild overhead compared to the flexibility achieved in the coupling of both hierarchies.

In Fig. 6 we see the registration result for image pair C using the registration algorithm with the streamlined coupling, see Fig. 4, between the image resolution and grid resolution hierarchies.

In Fig. 6 (left), the difference between the template and the reference image before registration is displayed. The differences are clearly visible. The right image of Fig. 6 shows the difference after registration, *i.e.* between the transformed template image and the reference image. The final transformation $M^{k_{\max},l_{\min}}$ is visualised in Fig. 7. The predominant medium grey intensity in the right difference image in Fig. 6 indicates a very successful registration. Almost all spot configurations present in both, the reference and the template image, are mapped onto each other by the computed registration map. The remaining differences serve to detect interesting spots (spots that have changed size or spots that vanish in the template image)

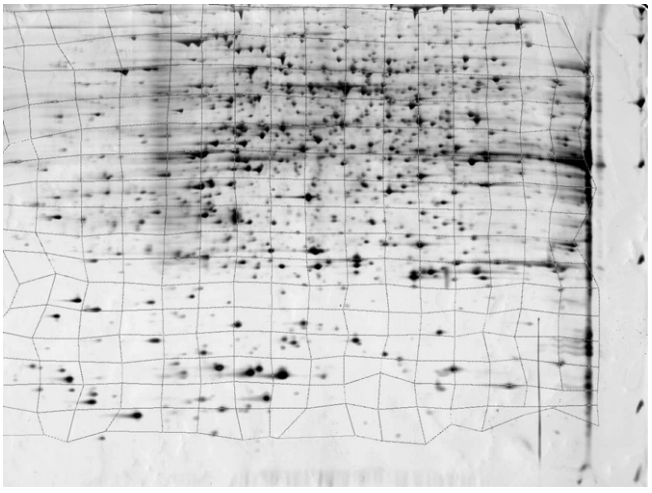

Fig. 7. The transformed grid of the bilinear transformation $M^{k_{\max},l_{\min}}$ in the untransformed template image of image pair C.

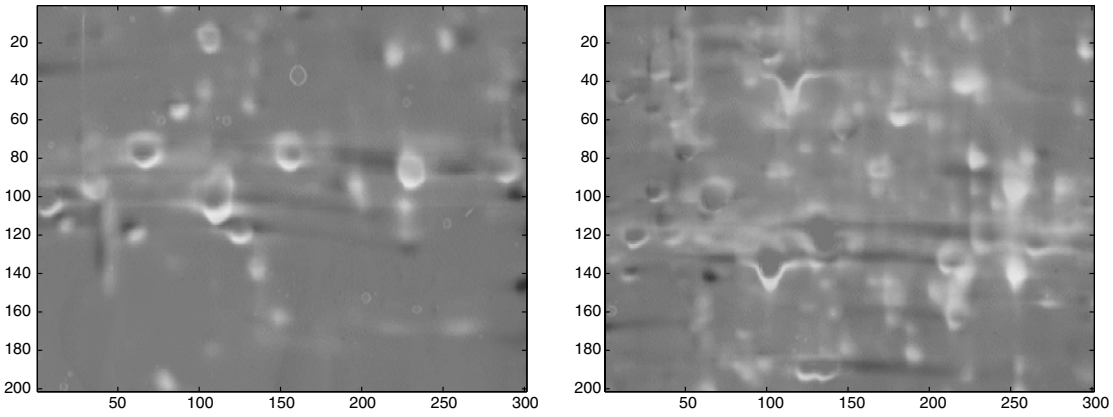

Fig. 8. Interesting spot configurations: spots that change size (left) and vanishing spots in the template image (right). Magnifications of the areas marked in Fig. 6 (right).

by visual inspection. Two such regions are highlighted in Fig. 6 with white boxes. In Fig. 8 (left), we see the configuration located in the box $[900, 1200] \times [500, 700]$ of Fig. 6. The white rings in the centre of the image indicate a higher production of a certain protein under the exposition of the plant to cadmium ions. In the right picture in Fig. 8 (which shows the section located in the box $[900, 1200] \times [200, 400]$ of Fig. 6) spots can be recognised that do not occur in the template image, *i.e.* the corresponding protein is produced when the plant is exposed to cadmium ions.

5. Summary

In this paper, we have presented a procedure to optimise the coupling between different hierarchies used in image registration algorithms. We have proved the efficiency of the optimisation procedure for a registration algorithm based on piecewise bilinear transformations, Euclidean distance measure and an elastic registration term. Although our grid spacing on the finest grid level is much larger than the pixel width (so the algorithm has to be classified as parametric) we have decided to add an elastic regularisation term to increase the robustness of the registration. For our choice of Lamé parameters and images we have not experienced that the elastic penalty term prevents registration.

Experiments on 2D-PAGE images show very promising results. In our tests we devised an improved strategy for determining fixed couplings as well as a streamlined adaptive strategy that lead to a very reliable and efficient registration algorithm. For the images processed, the adaptive approach worked always properly, whereas the less sophisticated couplings failed several times to overlap all spots, the improved coupling (FS) being superior to the original coupling (FD).

The devised optimisation procedure offers a valuable tool to be used by specialists in the field to improve state of the art registration algorithms.

The approach presented can be a base for future state of the art image registration algorithms. Our vision is to use continuous hierarchies in image resolution and transformation resolution (or regularisation parameters) and to couple them via the continuous pendant of dynamic programming – the Hamilton–Jacobi–Bellmann equation. Such a future generation registration algorithm could be able to identify bifurcations in the pathway of optimal transformations for fixed resolutions of images and transformations and therefore to identify all local minima of the registration problem on the finest levels.

Acknowledgements

We thank Edda v. Röpenack-Lahaye, Udo Roth and Stephan Clemens from the Leibniz Institute of Plant Biochemistry for providing gel images and background information. This work was in part supported by the Bundesministerium für Bildung und Forschung under Grant 0312706D.

References

[1] A. Bovik (Ed.), Image & Video Processing, Academic Press, San Diego, 2000.
[2] D. Braess, Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics, Cambridge University Press, 2001.
[3] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Transactions on Communication 31 (1983) 532–540.
[4] G.E. Christensen, Deformable Shape Models for Anatomy, PhD Thesis, 1994.
[5] U. Clarenz, M. Droske, S. Henn, M. Rumpf, K. Witsch, Computational methods for nonlinear image registration, in: O. Scherzer (Ed.), Mathematical Method for Registration and Applications to Medical Imaging, Mathematics in Industry, Vol. 10, Springer, Berlin, Heidelberg, 2006, pp. 81–102.
[6] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, MIT Press and McGraw-Hill, 1990.
[7] B. Fischer, J. Modersitzki, Combining landmark and intensity driven registrations, PAMM 3 (2003) 32–35.
[8] R. Fletcher, Practical Methods of Optimization, second ed., Wiley Chichester, 1990.
[9] D. Gusfield, Algorithms on Strings, Trees and Sequences, Cambridge University Press, 1997.
[10] J.S. Gustafsson, A. Blomberg, M. Rudemo, Warping two-dimensional electrophoresis gel images to correct for geometric distortions of the spot pattern, Electrophoresis 23 (2002) 1731–1744.

[11] E. Haber, J. Modersitzki, A multilevel method for image registration, SIAM Journal on Scientific Computing 27 (2006) 1594–1607.
[12] S. Henn, K. Witsch, Image registration based on multiscale energy information, SIAM Journal on Multiscale Modeling and Simulation 4 (2005) 584–609.
[13] C.-L. Huang, P.-Y. Chang, A multi-resolution image registration method for multimedia application, in: Proceedings of 1998 IEEE ISCAS, 1998.
[14] B. Josso, E. Zindy, H. Aldemir, Automatic 2-D gel registration using distance minimization of image morphing, in: International Conference on Information Visualisation, vol. IV, 2000, pp. 357–361.
[15] J. Kybic, M. Unser, Multidimensional elastic spline registration of images using splines, in: Proceedings of ICIP2000, 2000.
[16] J. Modersitzki, Numerical Methods for Image Registration, Oxford University Press, 2004.
[17] University of Bielefeld, Fermentation Engineering Group. 2DBase:2D-PAGE Database of *Escherichia coli*. <http://2dbase.techfak.uni-bielefeld.de>.
[18] M.D. Rogers, J. Graham, R.P. Tonge, R.M. Leahy, C. Roux, 2D electrophoresis gel registration using feature matching, in: IEEE International Symposium on Biomedical Imaging, 2004.
[19] U. Roth, E. von Roepenack-Lahaye, S. Clemens, Proteome changes in *Arabidopsis thaliana* roots upon exposure to cd^{2+}, Journal of Experimental Botany 57 (15) (2006) 4003–4013.
[20] J. Salmi, T. Aittokallio, J. Westerholm, M. Griese, A. Rosengren, T.A. Nyman, R. Lahesmaa, O. Nevalainen, Hierarchical grid transformation for image warping in the analysis of two-dimensional electrophoresis gels, Proteomics 2 (2002) 1504–1515.
[21] Z. Smilansky, Automatic registration for images of two-dimensional protein gels, Electrophoresis 22 (2001) 1616–1626.
[22] C.Ó.S. Sorzano, P. Thévenaz, M. Unser, Elastic registration of biological images using vector-spline regularization, IEE Transactions on Biomedical Engineering 52 (4) (2005) 652–663.
[23] S. Veeser, M.J. Dunn, G.-Z. Yang, Multiresolution image registration for two-dimensional gel electrophoresis, Proteomics 1 (2001) 856–870.
[24] X.Y. Wang, D.D. Feng, H. Hong, Novel elastic registration for 2-D medical and gel protein images, in: Conferences in Research and Practice in Information Technology, vol. 19, 2003.
[25] B. Zitová, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (2003) 977–1000.