Direct Estimation of Non-Rigid Registrations

Adrien Bartoli, Andrew Zisserman

BMVC'04 — IN PROCEEDINGS OF THE FIFTEENTH BRITISH MACHINE VISION CONFERENCE, LONDON, UK, PP. 899-908, VOL. II, SEPTEMBER 2004

Direct Estimation of Non-Rigid Registrations

Adrien Bartoli and **Andrew Zisserman**
Department of Engineering Science, University of Oxford
{Bartoli,az}@robots.ox.ac.uk

Abstract

Registering images of a deforming surface is a well-studied problem. Solutions include computing optic flow or estimating a parameterized motion model. In the case of optic flow it is necessary to include some regularization.

We propose an approach based on representing the induced transformation between images using Radial Basis Functions (RBF). The approach can be viewed as a direct, i.e. intensity-based, method, or equivalently, as a way of using RBFs as non-linear regularizers on the optic flow field.

The approach is demonstrated on several image sequences of deforming surfaces. It is shown that the computed registrations are sufficiently accurate to allow convincing augmentations of the images.

1 Introduction

The objective of this paper is registration of images of a non-rigidly deforming surface, such as a flag being gently blown by the wind. Our goal is to compute dense image transformations, mapping pixels from one image to corresponding pixels in the other images. This task is important in domains such as augmented reality and medical imaging. We are particularly interested in images of surfaces whose behaviour is difficult to explain using specific physics-based or learnt models, such as the motion of cloth in a skirt as a person walks.

One solution to this problem is to compute a regularized optic flow field, by minimizing an energy functional based on a data term, from e.g. the brightness constancy assumption, and a regularizer, encouraging smoothness of the flow field. A survey can be found in e.g. [12]. For example, Irani uses subspace constraints to regularize the optic flow field [8].

Another solution is to compute a parameterized image transformation. Two main approaches are possible: feature-based and direct methods. Feature-based methods first compute a set of matched features extracted from the images, such as corners or contours, and then use them to estimate the image transformations [15]. On the other hand, direct methods usually minimize an error function similar to the one used for computing the optic flow field, based on the brightness constancy assumption [9]. Both feature-based and direct methods have been shown to give good results when computing rigid transformations, such as affinities or homographies. However, feature-based methods may fail

to capture the non-rigidities in the image areas where few features are present. In other words, in contrast to the rigid transformation case where e.g. an affinity is encapsulated in any three point correspondences, an arbitrary unknown number of correspondences might be needed to represent all the image deformations. For example, a low texture area might be subject to deformations, while no corner points might be found in this area. This is one reason in favour of methods using the intensity information, i.e. optic flow and direct methods, for computing non-rigid transformations.

The method in [6] draws on the strength of both optic flow and direct methods. The idea is to learn linear motion models that are used as bases for the optic flow field.

Our objective is to avoid the need for learning the motion model. We propose to represent it using *Radial Basis Mappings* (RBM). Such transformations have been shown to be very effective in representing various image distortions induced by different kinds of non-rigidities. Radial Basis Functions (RBF) are non-linear functions defined by centres and coefficients, see e.g. [13]. An example of such a function is the Thin-Plate Spline [2, 4].

The traditional approach to estimating RBMs is feature-based, for the main reason that when landmarks are used as centres, then the coefficients of the transformation can be computed by a linear algorithm. In the Thin-Plate Spline case, the linear algorithm minimizes the 'bending energy' [4]. Chui *et al*. [5] propose an integrated feature-based approach to match points while computing a RBM.

On the other hand, very few attemps have been made towards computing RBMs by directly considering the image intensity, i.e. using a direct method, or equivalently using a RBM to regularize the optic flow field. Some progress has been made in this direction in the medical imaging community, but mainly assuming that the centres of the transformation are given by user-defined landmarks, see e.g. [10].

We propose a scheme for intensity-based estimation of RBMs. The novelty of our approach is that the number of centres is estimated on-the-fly, and directly depends on the degree of non-rigidity between the images. Our algorithm is based on estimating an affine transformation and adding centres until a registration criterion is met, while minimizing the registration residual. At each iteration, both the position of centres and the coefficients of the transformation are estimated. Our algorithm has therefore two main characteristics: it minimizes an intensity-based registration error and fits a parameterized non-rigid motion model, whose intrinsic complexity is tuned depending on the amount of non-rigid deformations. Note that an alternative method of choosing centres is given in [11].

We give some background in §2, and describe our method for the direct estimation of RBMs in §3. We propose an extension of the method to deal with image sequences in §4 and report experimental results in §5. Finally, we give our conclusions and discuss further work in §6.

Notation. Vectors and matrices are respectively typeset using bold and sans-serif fonts, e.g. $\mathbf{x}$ and A. We do not use homogeneous coordinates, i.e. image point coordinates are 2-vectors: $\mathbf{x}^\mathsf{T} = (x\ y)$, where $^\mathsf{T}$ is transposition. We denote as $\mathscr{I}$ the images, and $\mathscr{I}(\mathbf{x})$ the colour or gray-level at pixel $\mathbf{x}$. Index i is used for the frames ($i = 1,\ldots,n$), k for the centres ($k = 1,\ldots,l$) and j for point correspondences. The evaluation of an expression at some value is denoted as in $\mathsf{S}|_{\mathbf{x}}$. Matrix vectorization is written as in $\mathbf{a} = \mathrm{vect}(\mathsf{A})$. The identity matrix is denoted I and the zero matrix and vector as $\mathsf{0}$ and $\mathbf{0}$.

2 Background

In the following two sections, we describe direct methods and RBMs.

2.1 Direct Methods

We describe the principle of the direct, i.e. intensity-based, alignment of two images $\mathscr{I}$ and $\mathscr{I}'$, related by a point-to-point transformation. The main idea, common to most algorithms, is to minimize the sum of squared intensity differences between the aligned images, over the parameters of the transformation [9]. Let us consider the case of an affine transformation, and derive one of the possible algorithms based on Gauss-Newton. A more detailed formulation of the image alignment problem is described in [1]. In particular, a robust, coarse-to-fine framework is used to speed up convergence and prevent the algorithm getting trapped into local minima, see [3].

An affine transformation has 6 parameters and is modelled by a 2×3 matrix $\mathsf{A} = (\,\bar{\mathsf{A}}\ \ \mathbf{t})$, where $\bar{\mathsf{A}}$ is the leading 2×2 submatrix and $\mathbf{t}$ the last column. A point $\mathbf{x}$ is mapped from the first image to a point $\mathbf{x}'$ in the second image by $\mathbf{x}' = \bar{\mathsf{A}}\mathbf{x} + \mathbf{t}$. The transformation is parameterized by the 6-vector $\mathbf{a} = \mathrm{vect}(\mathsf{A})$. The direct estimation of A consists in solving $\min_{\mathbf{a}} E(\mathbf{a})^2$, where E is an error function defined by the mean intensity difference induced by A between $\mathscr{I}$ and $\mathscr{I}'$ over a region of interest $\mathscr{X}$ with N pixels $E(\mathbf{a})^2 = \frac{1}{N}\sum_{\mathbf{x}\in\mathscr{X}} e(\mathbf{x},\mathbf{a})^2$. Each error term is given by $e(\mathbf{x},\mathbf{a})^2 = \left(\mathscr{I}(\mathbf{x}) - \mathscr{I}'(\bar{\mathsf{A}}\mathbf{x}+\mathbf{t})\right)^2$. We employ the Gauss-Newton algorithm [14] to solve this minimization problem, using the identity transformation as an initial solution.

2.2 Radial Basis Mappings

In their basic form, RBFs define a mapping from $\mathbb{R}^d$ to $\mathbb{R}$, where d is the dimension. A general description can be found in [13]. A 2D RBF f is defined by a $\mathbb{R}^2 \to \mathbb{R}$ *basis function* $\phi(\eta)$, *coefficients* represented by an $l+3$-vector $\mathbf{h}^\mathsf{T} = (w_1\ \cdots\ w_l\ \lambda\ \mu\ \nu)$ and a set of l *centres* $\mathbf{q}_k$ as:

$$f(\mathbf{x}) = \lambda x + \mu y + \nu + \sum_{k=1}^{l} w_k \phi(\|\mathbf{x} - \mathbf{q}_k\|). \tag{1}$$

It consists of a linear part, with parameters $(\lambda\ \mu\ \nu)$, and a non-linear part, a sum of l weighted terms with coefficients w_k of the basis function applied to the distance between $\mathbf{x}$ and the centre $\mathbf{q}_k$. Amongst others, the basis function can be chosen as a Gaussian $\phi(\eta) = \exp(-\eta^2/(2\sigma^2))/(2\pi\sigma^2)$ or as a Thin-Plate Spline $\phi(\eta) = \eta^2\log(\eta)$.

Radial Basis Mappings as $\mathbb{R}^2 \to \mathbb{R}^2$ Radial Basis Functions. The usual way to construct a $\mathbb{R}^2 \to \mathbb{R}^2$ mapping m, i.e. a RBM, is to stack two $\mathbb{R}^2 \to \mathbb{R}$ RBFs f^x and f^y sharing their centres [4]:

$$m(\mathbf{x}) = \begin{pmatrix} f^x(\mathbf{x}) \\ f^y(\mathbf{x}) \end{pmatrix} = \bar{\mathsf{A}}\mathbf{x} + \mathbf{t} + \sum_{k=1}^{l} \begin{pmatrix} w_k^x \\ w_k^y \end{pmatrix} \phi(\|\mathbf{x} - \mathbf{q}_k\|), \tag{2}$$

where $\bar{\mathsf{A}}$ and $\mathbf{t}$ form an affine transformation given by:

$$\mathsf{A} = \begin{pmatrix} \lambda^x & \mu^x & \nu^x \\ \lambda^y & \mu^y & \nu^y \end{pmatrix}.$$

The coefficients are encapsulated in an $(l+3)\times 2$ matrix $\mathsf{h} = (\mathbf{h}^x\ \mathbf{h}^y)$, partitioned in a non-rigid and a rigid part as $\mathsf{h}^\mathsf{T} = (\mathsf{W}^\mathsf{T}\ \mathsf{A})$.

Computation from point correspondences. RBMs are often used to create dense smooth transformations interpolating point correspondences $\mathbf{n}_j \leftrightarrow \mathbf{n}'_j$ between two images. Points $\mathbf{n}_j$ are used as the centres of the transformation. By writing the interpolating conditions $\mathbf{n}'_j = m(\mathbf{n}_j)$ using equation (2), one obtains:

$$\mathbf{n}'_j = \bar{\mathsf{A}}\mathbf{n}_j + \mathbf{t} + \sum_{k=1}^{l} \begin{pmatrix} w_k^x \\ w_k^y \end{pmatrix} \phi(\|\mathbf{n}_j - \mathbf{n}_k\|),$$

which can be rewritten as a linear system $\mathsf{Ch} = \mathsf{D}$, see [2, 4].

The side conditions. The last three equations of the linear system $\mathsf{Ch} = \mathsf{D}$ are the 'side conditions'. They ensure that the computed transformation has square integrable second derivatives, i.e. that it smoothly behaves outside the region of interest $\mathscr{X}$. The side conditions are expressed between the coefficients $\mathbf{w}^x$ and $\mathbf{w}^y$ and the centres $\mathbf{n}_j$ as $\sum_{k=1}^{l} w_k = \sum_{k=1}^{l} w_k n_{k,1} = \sum_{k=1}^{l} w_k n_{k,2} = 0$, or, defining the r-th row of matrix P as $(\mathbf{n}_r^\mathsf{T}\ 1)$, in matrix form:

$$\mathsf{P}^\mathsf{T}\mathsf{W} = \mathbf{0}_{(3\times 1)}. \tag{3}$$

3 Direct Estimation of Radial Basis Mappings

We describe our approach for estimating a RBM by using a direct method.

3.1 Outline of the Approach

Constructing a direct method for estimating a RBM raises specific concerns since the number of centres l, as well as the centres $\mathbf{q}_k$ themselves and the coefficients h of the transformation have to be estimated. More formally, we formulate the problem as:

$$\min_{l,\alpha} E(\alpha)^2 \quad \text{such that } \mathsf{P}^\mathsf{T}\mathsf{W} = \mathbf{0}_{(3\times 1)},$$

where α encapsulates the set of parameters of the transformation as:

$$\alpha^\mathsf{T} = (q_1^x\ q_1^y\ \ldots\ q_l^x\ q_l^y\ w_1^x\ w_1^y\ \ldots\ w_l^x\ w_l^y\ a_1\ \ldots\ a_6).$$

A possible approach is to use a pre-defined set of centres, e.g. on a regular grid or corner points in the first image (as in a feature-based approach). This approach is not satisfactory. If too few centres are used, the mapping may fail to capture all the deformations, and if too many centres are used, then the computational cost might be extremely high.

Our approach is built on a *dynamic centre insertion* procedure. The idea is to iteratively insert new centres, i.e. add non-rigidity to the transformation, until it becomes satisfactory. The centres are inserted based on examining the error image, to detect where the mapping fails to provide a proper registration. At each iteration, a centre is inserted, and the error is minimized. The number of centres grows until the algorithm converges.

The initial number of centres is set to 4, since if less than 4 centres are present, the corresponding coefficients are constrained to be zero by the side-conditions (3). The algorithm is summarized in table 1 and illustrated in figure 1.

1. *Initialization:* use algorithm of §2.1 to obtain the parameters A of an initial affine transformation. Insert 4 centres as indicated in §3.3, set $l \leftarrow 4$ and setup α accordingly.
2. *Transformation refinement:* (§3.2) compute the parameters α' which minimize the error in intensity, starting from α.
3. *Convergence test:* (§3.4) if $E(\alpha) - E(\alpha') < \varepsilon$ then remove the last inserted centre(s) and stop.
4. *Parameters updating:* $\alpha \leftarrow \alpha'$.
5. *Centre insertion:* (§3.3) $l \leftarrow l+1$. The new centre is $\mathbf{q}_l$, with corresponding coefficients $w_l^x \leftarrow 0$ and $w_l^y \leftarrow 0$.
6. *Main loop:* return to step 2.

Table 1: Direct alignment of two images $\mathscr{I}$ and $\mathscr{I}'$ based on estimating a RBM α with l centres $\mathbf{q}_k$ and coefficients h, using the Gauss-Newton algorithm and dynamic centre insertion. The iterations terminate when the decrease in the error becomes insignificant.

Figure 1: Registration of two images of a deforming Tshirt, shown top and bottom. (from left to right) Original images, the centres of the transformation and a grid mesh illustrating the mapping.

3.2 Refining the Transformation

We describe the alignment algorithm given the number l of centres. The algorithm draws on the previously described algorithm for the direct estimation of an affine transformation, see §2.1. The problem is to solve:

$$\min_{\alpha} E(\alpha)^2 \quad \text{such that } \mathsf{P}^{\mathsf{T}}\mathsf{W} = \mathbf{0}_{(3\times 1)}.$$

We use the Gauss-Newton algorithm. The differences with the affine transformation refinement algorithm are two-fold: the Jacobian matrix of the mapping is more complicated, and a special parameterization is used to enforce the side-conditions[1].

3.3 Dynamic Centre Insertion

A centre accounts for non-rigidity in the transformation around its position. Our strategy is to insert centres until the transformation gives a satisfactory registration of the two images, by looking at the error image $\mathscr{E}$, i.e. the difference between the first image, and the warped second image.

We proceed as follows. First, we compute a blurred version $\hat{\mathscr{I}}$ of the first image $\mathscr{I}$ using a Gaussian kernel. Second, using the current transformation, we warp the second image $\mathscr{I}'$ as $\hat{\mathscr{I}}'$, and blur using the same Gaussian kernel. Blurring the images before computing their difference is important to get rid of effects such as the partial pixel effect. The error image $\mathscr{E}$ (the absolute difference between $\hat{\mathscr{I}}$ and $\hat{\mathscr{I}}'$) indicates the regions where the registration is not satisfactory. One may simply insert the new centre where $\mathscr{E}$ attains a maximum. We perform an integration step by convolving with a Gaussian kernel, before looking for the maximum, to emphasize the regions where the registration is not satisfactory.

3.4 A Stopping Criterion

Inserting a centre increases the number of degrees of freedom of the transformation and reduce the registration error. One strategy to stop the iterations would be to penalize the registration error by the number of degrees of freedom, and stop the algorithm when the minimum of this function is reached, similar to a nested model selection algorithm, e.g. [16]. Another strategy is based on the fact that when the 'best' number of centres is reached, inserting a new centre will only produce a slight decrease in the error, proportional to the noise level and quantization / warping error. Thresholding the difference between two consecutive errors is consequently used as a stopping criterion.

4 Registering Multiple Images

The goal in this section is to exploit the two-image registration algorithm proposed above, to register a sequence of images. More precisely, we aim at computing the transformation between a reference image of the sequence and all other images. Without loss of generality, we choose the first image as the reference one. Denoting $f_{i_1 i_2}$ the transformation between image i_1 and image i_2, our goal is to compute f_{1i}, $i = 2, \ldots, n$.

We perform sequential processing: we first compute f_{12} starting from an identity transformation. Then, we compute f_{13} starting from f_{12} and so on.

One problem with this approach is that throughout the sequence, shadows might appear or disappear, meaning that the appearance of the surface might change. To overcome

[1] We use a QR decomposition-based subspace projection, as described in e.g. [7, §12.1.4].

this problem, we investigated two approaches. The first approach consists in updating the appearance of the reference frame while registering the frames. After having computed f_{12}, we use it to align $\mathscr{I}_2$ with the reference frame. This aligned image is called $\mathscr{I}_{21}$, and is used as an updated reference frame when computing f_{13} as the transformation between $\mathscr{I}_{21}$ and $\mathscr{I}_3$. Of course, this approach gets rid of appearance variations, but might drift since registration errors are accumulated through the process.

The second approach we propose is to apply a shadow mask. This mask is computed as the residual error image between the reference image $\mathscr{I}_1$ and $\mathscr{I}_{21}$. It is then applied to the reference image before registering it with $\mathscr{I}_3$. As in the previous approach, this gets rid of appearance variations, but might drift through the sequence. We have found however, that this approach is less likely to drift. This is due to the fact that only the shadow mask is updated, and not the reference frame itself.

5 Experimental Results

This section reports experimental results on simulated and real data.

5.1 Simulated Data

The goal of these experiments is to validate the algorithm and determinate conditions under which it converges. Starting from a first image, we generate a second image, that will be used in the algorithm. The second image is generated as illustrated on figure 2. First, a rigid transformation is applied to the first image: three points are selected and

Figure 2: The simulation setup: the original Tshirt image, after affine transformation, after full non-rigid transformation and after having applied a global illumination change and added random noise.

randomly perturbed to define an affine transformation. The direction of the perturbation is chosen at random, while its norm δ_R is user-defined. Second, a non-rigid transformation is applied: points on a regular grid are offset by a random perturbation as above, with norm δ_{NR}. Third, a global illumination change is applied, as well as a random Gaussian noise with variance σ, on the intensity of all pixels. We perform the experiments using the Tshirt image of figure 2 and an image from the Newspaper sequence, figure 5.

The default values of these three parameters are $\delta_R = 3$ pixels, $\delta_{NR} = 2$ pixels and $\sigma = 1$ (over 256 gray levels). We vary independently these parameters while measuring the residual error E at convergence. A residual error close to the noise level σ means that the algorithm successfully converged. The results are averages over 50 trials.

Figure 3 shows the results we obtained. Based on these, we can say that the convergence is independent of the rigid part δ_R, on the 0 to 10 pixels range. However, convergence is strongly affected by the non-rigid part δ_{NR}. It gracefully degrades until a break-point is reached, at roughly $\delta_{NR} = 6$ pixels for the Tshirt image and $\delta_{NR} = 4$ pixels

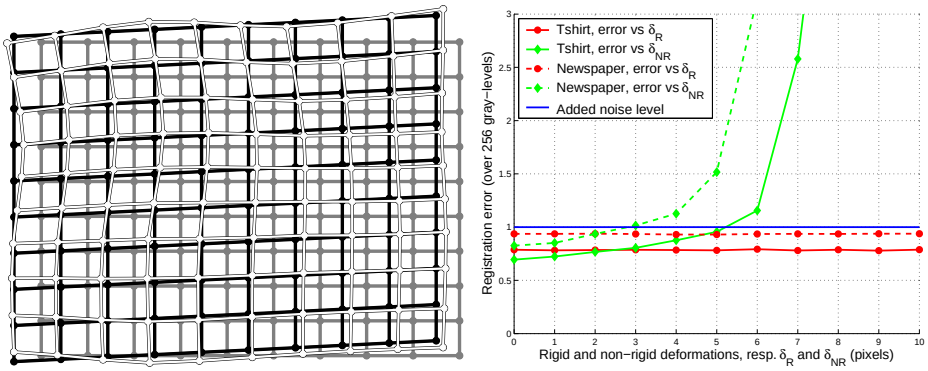

Figure 3: (left) A grid mesh illustrating the kind of simulated deformations. (right) Simulation results.

for the Newspaper image. Beyond these break-points, the algorithm does not converge to the right solution.

5.2 Real Data

Comparing grid-based and dynamic centre insertion approaches. We compare the traditional approach based on placing the centres at the nodes of a fixed, regular grid, and our dynamic centre insertion procedure. The results for the Tshirt images shown in figure 1 are displayed on figure 4. Our approach converged after 5 centres were introduced, with an error of 13.80 over 256 gray-levels.

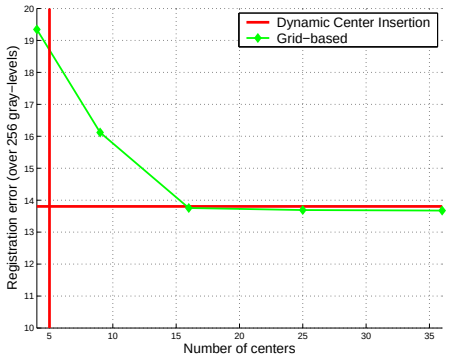

Figure 4: Comparison of our dynamic centre insertion and a fixed grid approach.

We observe that the fixed grid approach needs many more centres than ours to minimize the error function to a similar order of magnitude. More precisely, 16 centres are needed for the fixed grid approach to reach the same alignment.

The Newspaper sequence. We apply our algorithm to the 225-frame Newspaper sequence, shown in figure 5. This sequence was acquired by waving the newspaper, which non-rigidly deformed, in front of a hand-held digital camera. The movie shows that the surface is undergoing highly non-rigid deformations. We select a reference frame, shown on figure 5, to which all other frames are registered, using pair-wise non-rigid motion estimation, as described in §4.

Visually it is evident that the registration is good. The average intensity registration error over the sequence is 5.24 over 256 gray-levels, which is acceptable. The final column of figure 5 shows an augmented sequence: the original cartoon has been replaced by a new one. The video presentation associated with this paper clearly shows that visual deformations have been eliminated.

Figure 5: Registration results on the Newspaper sequence. (left) Original images. (middle) A mesh grid illustrating the computed mapping. (right) The cartoon on the reference image (indicated by a black frame) is replaced and mapped onto the other frames.

6 Conclusions and Further Work

We have proposed an intensity-based algorithm for the computation of Radial Basis Mappings, that can equivalently be viewed as a way to compute a regulatized optic flow field. The basic algorithm is intended to register pairs of views, and an extension for the registration of multiple views is proposed. Amongst the possible avenues for future research, experimenting with different, robust cost functions would be important, as well as computing super-resolution.

References

[1] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, February 2004.

[2] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, pages 454–461, 2001.

[3] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 237–252, 1992.

[4] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[5] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2):114–141, February 2003.

[6] D. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson. Design and use of linear models for image motion analysis. *International Journal of Computer Vision*, 36(3):171–193, 2000.

[7] G.H. Golub and C.F. van Loan. *Matrix Computation*. The Johns Hopkins University Press, Baltimore, 1989.

[8] M. Irani. Multi-frame optical flow estimation using subspace constraints. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, volume I, pages 626–633, September 1999.

[9] M. Irani and P. Anandan. About direct methods. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, number 1883 in LNCS, pages 267–277, Corfu, Greece, July 1999. Springer-Verlag.

[10] H. J. Johnson and G. E. Christensen. Consistent landmark and intensity-based image registration. IEEE *Transactions on Medical Imaging*, 21(5):450–461, May 2002.

[11] S. Marsland and C. J. Twining. Constructing data-driven optimal representations for iterative pairwise non-rigid registration. In *Second International Workshop on Biometric Image Registration*, pages 50–60, 2003.

[12] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *International Journal of Computer Vision*, 19(1):29–55, July 1996.

[13] M. J. L. Orr. Introduction to radial basis function networks. Technical report, University of Edinburgh, 1996.

[14] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[15] P. H. S. Torr and A. Zisserman. Feature based methods for structure and motion estimation. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, number 1883 in LNCS, pages 278–295, Corfu, Greece, July 1999. Springer-Verlag.

[16] P.H.S. Torr. Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society of London A*, 356:1321–1340, 1998.