Multi-modal Image Registration Using Polynomial Expansion and Mutual Information

Daniel Forsberg[1,2,3], Gunnar Farnebäck[1],
Hans Knutsson[1,2], and Carl-Fredrik Westin[4]

[1] Department of Biomedical Engineering, Linköping University, Sweden
[2] Center for Medical Image Science and Visualization, Linköping University, Sweden
[3] Sectra Imtec, Linköping, Sweden
[4] Laboratory of Mathematics in Imaging, Brigham and Women's Hospital,
Harvard Medical School, Boston, MA, USA

Abstract. The use of polynomial expansion in image registration has previously been shown to be beneficial due to fast convergence and high accuracy. However, earlier work has only been for mono-modal image registration. In this work, it is shown how polynomial expansion and mutual information can be linked to achieve multi-modal image registration. The proposed method is evaluated using MRI data and shown to have a satisfactory accuracy while not increasing the computation time significantly.

1 Introduction

The use of image registration within the medical image domain is vast and includes various tasks, such as; surgical planning, radiotherapy planning, image-guided surgery, disease progression monitoring and image fusion. Especially image fusion is becoming more and more important in several clinical workflows, since patients tend to have multiple examinations from different imaging modalities that need to be registered before image fusion is possible. Intensity-based similarity metrics, such as; sum-of-squared-difference or normalized cross-correlation, are often inadequate for handling multi-modal image registration. A frequently applied measure that can handle multi-modal registration is mutual information. Since the introduction of mutual information in image registration by Collignon *et al.* [1] and by Viola and Wells [10], it has become a well-established similarity metric for multi-modal image registration [8].

Polynomial expansion was introduced by Farnebäck [3] as a method to locally approximate a signal with a polynomial. It was later shown by Farnebäck and Westin [4] how polynomial expansion could be used to perform both linear (e.g. translation and affine) and non-rigid image registration. The idea of image registration using polynomial expansion was further developed by Wang *et al.* [11]. Both Farnebäck and Westin [4] and Wang *et al.* [11] showed that image registration using polynomial expansions has some valuable qualities. Firstly, since it is

based on an analytical solution, the convergence rate is fast, typically only need-ing a few iterations per scale. Secondly, also the accuracy of the registration has been demonstrated to be on a similar or even better level than some of the algo-rithms included in ITK. However, thus far, image registration using polynomial expansion has only been applicable for mono-modal image registration.

The contribution of this paper, is to present how mutual information can be integrated into polynomial expansion in order to achieve multi-modal image registration, thus, making image registration based on polynomial expansion feasible for not only mono-modal registration but also for multi-modal image registration.

2   Background

2.1   Polynomial Expansion

The basic idea of polynomial expansion is to locally approximate each signal value with a polynomial. In case of a linear polynomial, this approximation can be expressed as $f(\mathbf{x}) \sim \mathbf{b}^T \mathbf{x} + c$, where the coefficients are determined by a weighted least squares fit to the local signal. The weighting depends on two factors, certainty and applicability. These terms are the same as in normalized convolution [3,6], which forms the basis for polynomial expansion.

2.2   Image Registration Using Polynomial Expansion

Global Translation. Let both the fixed and the moving images be locally approximated with a linear polynomial expansion and assume that the moving image is a globally translated version of the fixed image, thus,

$$I_{\mathrm{f}}(\mathbf{x}) = \mathbf{b}_f^T \mathbf{x} + c_f, \tag{1}$$

$$I_{\mathrm{m}}(\mathbf{x}) = \mathbf{b}_m^T \mathbf{x} + c_m = I_{\mathrm{f}}(\mathbf{x} - \mathbf{d}) = \mathbf{b}_f^T(\mathbf{x} - \mathbf{d}) + c_f, \tag{2}$$

which gives $c_m = c_f - \mathbf{b}_f^T \mathbf{d}$ and hence, the translation $\mathbf{d}$ can be is estimated as

$$\mathbf{d} = (\mathbf{b}_f \mathbf{b}_f^T)^{-1} \mathbf{b}_f (c_f - c_m). \tag{3}$$

In practice, a point-wise polynomial expansion is estimated, i.e. $\mathbf{b}_f(\mathbf{x})$, $c_f(\mathbf{x})$, $\mathbf{b}_m(\mathbf{x})$, and $c_m(\mathbf{x})$. Since it cannot be expected that $\mathbf{b}_f(\mathbf{x}) = \mathbf{b}_m(\mathbf{x})$ holds, they are replaced with the average

$$\mathbf{b}(\mathbf{x}) = \frac{\mathbf{b}_f(\mathbf{x}) + \mathbf{b}_m(\mathbf{x})}{2}. \tag{4}$$

Also, set

$$\boldsymbol{\Delta} c(\mathbf{x}) = c_f(\mathbf{x}) - c_m(\mathbf{x}) \tag{5}$$

and thus, the primary constraint is given by:

$$\mathbf{b}(\mathbf{x})^T \mathbf{d} = \boldsymbol{\Delta} c(\mathbf{x}) \tag{6}$$

To solve (6), compute $\mathbf{d}$ by minimizing the squared error in the constraint over the whole image,

$$\epsilon^2 = \sum_{\mathbf{x}} \|\mathbf{b}(\mathbf{x})\mathbf{d} - \boldsymbol{\Delta}c(\mathbf{x})\|^2. \tag{7}$$

Affine Registration. If a space-variant displacement is assumed, then the previous solution can be extended to handle affine registration. Let $\mathbf{d}(\mathbf{x}) = \mathbf{S}(\mathbf{x})\mathbf{p}$, where

$$\mathbf{S}(\mathbf{x}) = \begin{pmatrix} x\ y\ 0\ 0\ 1\ 0 \\ 0\ 0\ x\ y\ 0\ 1 \end{pmatrix}, \tag{8}$$

$$\mathbf{p} = \begin{pmatrix} a_1\ a_2\ a_3\ a_4\ a_5\ a_6 \end{pmatrix}^T. \tag{9}$$

To estimate the parameters of the affine displacement field, replace $\mathbf{d}$ with $\mathbf{d}(\mathbf{x}) = \mathbf{S}(\mathbf{x})\mathbf{p}$ in (7),

$$\epsilon^2 = \sum_{\mathbf{x}} \|\mathbf{b}(\mathbf{x})^T \mathbf{S}(\mathbf{x})\mathbf{p} - \boldsymbol{\Delta}c(\mathbf{x})\|^2 \tag{10}$$

and the parameters $\mathbf{p}$ are given by the least squares solution of (10).

Non-rigid Registration. A non-rigid registration algorithm can be achieved if the assumption about a global translation is relaxed and we instead sum over a neighborhood around each pixel in (7), thereby obtaining an estimate for each pixel. In this case, a local translation is assumed but it could easily be changed to a local affine transformation as is done by Wang *et al.* [11]. More precisely (7) is changed to

$$\epsilon^2(\mathbf{x}) = \sum_{\mathbf{y}} w(\mathbf{y}) \left(\|\mathbf{b}(\mathbf{x} - \mathbf{y})^T \mathbf{d}(\mathbf{x}) - \boldsymbol{\Delta}c(\mathbf{x} - \mathbf{y})\|^2 \right), \tag{11}$$

where w weights the points in the neighborhood around each pixel. This weight can be any low-pass function, but here it is assumed to be Gaussian. This equation can be interpreted as a convolution of the point-wise contributions to the squared error in (7) with the low-pass filter w. The solution is given as

$$\mathbf{G}(\mathbf{x}) = \mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T, \tag{12}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{b}(\mathbf{x})\boldsymbol{\Delta}c(\mathbf{x}), \tag{13}$$

$$\mathbf{G}_{\mathrm{avg}}(\mathbf{x}) = (\mathbf{G} * w)(\mathbf{x}), \tag{14}$$

$$\mathbf{h}_{\mathrm{avg}}(\mathbf{x}) = (\mathbf{h} * w)(\mathbf{x}), \tag{15}$$

$$\mathbf{d}(\mathbf{x}) = \mathbf{G}_{\mathrm{avg}}(\mathbf{x})^{-1}\mathbf{h}_{\mathrm{avg}}(\mathbf{x}). \tag{16}$$

Note that in this work only a linear polynomial expansion have been used, but as shown in [4,11] a quadratic polynomial expansion along with similar derivations

can also be used for image registration. In fact, it is possible to combine both in order to obtain a more robust solution. Further, the described method for non-rigid registration can be extended to handle diffeomorphic deformations by using diffeomorphic field accumulation as described by Forsberg *et al.* [5].

2.3   Mutual Information

Mutual information is typically based on the *Shannon Entropy* and defined as

$$H = \sum_i p_i \log \frac{1}{p_i} = -\sum_i p_i \log p_i, \tag{17}$$

where p_i is the probability that event e_i occurs. Based on the Shannon entropy we can define the mutual information of the random variables A and B as

$$MI(A, B) = H(A) + H(B) - H(A, B), \tag{18}$$

where $H(A, B)$ is the joint entropy of A and B. The Shannon entropy for the joint distribution is given by

$$-\sum_{i,j} p_{i,j} \log p_{i,j}. \tag{19}$$

3   Polynomial Expansion and Mutual Information

3.1   Algorithm Overview

The basic idea of combining polynomial expansion and mutual information lies in replacing $\boldsymbol{\Delta} c(\mathbf{x})$ in (5). The term $\boldsymbol{\Delta} c(\mathbf{x})$ can be interpreted as how much $I_f(\mathbf{x})$ should change in order to match $I_m(\mathbf{x})$ based on intensity-difference. Instead, estimate $\boldsymbol{\Delta} c(\mathbf{x})$ as how much $I_f(\mathbf{x})$ should change in order to match $I_m(\mathbf{x})$ but based on a pixel-wise minimization of the conditional entropy, i.e. a pixel-wise maximization of the mutual information. This is estimated according to the following scheme:

- Estimate the joint distribution $p(I_f, I_m)$ given I_f and I_m.
- For each $\mathbf{x}$:
    - Retrieve the corresponding pixel values $i_f = I_f(\mathbf{x})$ and $i_m = I_m(\mathbf{x} + \mathbf{d})$.
    - Estimate the conditional distribution $p(I_f | I_m = i_m)$.
    - Estimate how the conditional entropy $H(I_f | I_m = i_m)$ is affected by i_f.
    - Find $i^\star$ for which the conditional entropy has a local minimum and estimate $\boldsymbol{\Delta} c(\mathbf{x})$ as $i_f - i^\star$.

Knowing that mutual information can also be defined as

$$MI(A, B) = H(A) - H(A|B), \tag{20}$$

ensures that the last step is reasonable for finding a $\boldsymbol{\Delta} c(\mathbf{x})$ that matches I_f and I_m based on mutual information, since minimizing $H(I_f | I_m)$ will maximize $MI(I_f, I_m)$ in a point-wise manner.

3.2   Channel Coding and Probability Distribution Estimation

A well-known issue with mutual information is how to estimate the needed probability distributions, since the straightforward solution of using a standard histogram is insufficient. A common approach is to use some sort of *soft histograms* where the intensity of each pixel is distributed fractionally over multiple bins. This will give a more stable and often close to continuous result. In this work, channel coding and B-splines have been used to achieve this, which in practice is similar to *kernel density estimation* introduced by Rosenblatt and Parzen to estimate probability density functions [7,9].

The basic concept of channel coding consists of representing a value x by passing it through a set of localized basis functions $\{B(x-k)\}_1^K$. The output signal from each basis function is called a channel, thus, the channel representation of x is given by a vector of a set of channel values

$$\varphi(x) = [B(x-1) \ B(x-2) \ \ldots \ B(x-K)]^T . \tag{21}$$

A suitable basis function is the quadratic B-spline function, defined as the unit rectangle convolved with itself two times. A B-spline of degree two, $B_2(x)$, is chosen because it is continuously differentiable and has compact support.

The probability distribution of an image, using K channels, is, thus, estimated as:

1. Map each intensity value $I(\mathbf{x})$ to $[1, K-1]$, $\tilde{I}(\mathbf{x})$.
2. Compute the channel representation of each mapped value $\varphi(\tilde{I}(\mathbf{x}))$.
3. Compute the bins h_k of the soft histogram as the element-wise sum of the channel representations.
4. Estimate a probability distribution p_k by normalizing the soft histogram.

The fractional bins of the joint histogram can then be estimated as the sum of the element-wise products of the channel representations over the respective images

$$h_{k,l} = \sum_{\mathbf{x}} \varphi_k\left(\tilde{I}_f(\mathbf{x})\right) \varphi_l\left(\tilde{I}_m(\mathbf{x} + \mathbf{d})\right), \tag{22}$$

which is then normalized to obtain the joint probability distribution, $p_{k,l}$.

3.3   Conditional Entropy

To estimate $\mathbf{\Delta}c(\mathbf{x})$ for the corresponding pixel values $i_f = I_f(\mathbf{x})$ and $i_m = I_m(\mathbf{x} + \mathbf{d}(\mathbf{x}))$, we start by estimating the conditional probability distribution $p_k(I_f|I_m = i_m)$ as

$$p_k(I_f|I_m = i_m) = \frac{h_k(I_f|I_m = i_m)}{\sum_{k=1}^{K} h_k(I_f|I_m = i_m)} \tag{23}$$

where

$$h_k(I_f|I_m = i_m) = \sum_{l=1}^{L} \varphi_l(\tilde{i}_m) \, p_{k,l}(I_f, I_m) \quad \text{for} \quad k = 1 \ldots K, \tag{24}$$

and where $\tilde{i}_m$ is the mapped i_m value. Here $\varphi_l\left(\tilde{i}_m\right) p_{k,l}\left(I_f, I_m\right)$ acts as a sub-channel interpolation term.

Given the conditional probability distribution $p_k\left(I_f | I_m = i_m\right)$, the conditional entropy is estimated as

$$H\left(I_f | I_m = i_m\right) = -\sum_k p_k\left(I_f | I_m = i_m\right) \log p_k\left(I_f | I_m = i_m\right). \qquad (25)$$

However, in our case we are interested in understanding how i_f affects the conditional entropy. This can be investigated by an infinite small addition of the intensity value i_f to create a new distribution

$$p_k\left(i_f\right) = \frac{p_k\left(I_f | I_m = i_m\right) + \epsilon \times \varphi_k\left(\tilde{i}_f\right)}{1+\epsilon}, \qquad (26)$$

where ϵ is small. Thus, the modified entropy is given by

$$H\left(i_f\right) = -\sum_k p_k\left(i_f\right) \log p_k\left(i_f\right) = \frac{1}{1+\epsilon}\left(H\left(I_f | I_m = i_m\right) - \right.$$
$$\left. \epsilon \sum_k \varphi_k\left(\tilde{i}_f\right)\left(1 + \log p_k\left(I_f | I_m = i_m\right)\right)\right) + \mathrm{O}\left(\epsilon^2\right). \qquad (27)$$

Differentiating (27) with respect to i_f and omitting $\mathrm{O}\left(\epsilon^2\right)$ gives

$$\frac{\partial H\left(i_f\right)}{\partial i_f} = \frac{\epsilon}{1+\epsilon}\left(\sum_k \varphi_k\left(\tilde{i}_f\right)\left(1 + \log p_k\left(I_f | I_m = i_m\right)\right)\right). \qquad (28)$$

Using the expressions in (27) and in (28), it is straightforward to find a $i^\star$, which minimizes $H\left(i^\star\right)$. This in turn, can be viewed as a pixel-wise minimization of the conditional entropy. The found $i^\star$ is used to estimate $\boldsymbol{\Delta}c(\mathbf{x})$ as

$$\boldsymbol{\Delta}c(\mathbf{x}) = i_f - i^\star. \qquad (29)$$

The term $\boldsymbol{\Delta}c(\mathbf{x})$ can, as previously explained, be interpreted as how much i_f should change in order to match i_m in a mutual information sense, instead of in a intensity-difference sense. The estimated $\boldsymbol{\Delta}c(\mathbf{x})$ can now be used for image registration as previously described in Sect. 2.2, but with the difference that $\mathbf{b}(\mathbf{x}) = \mathbf{b}_f(\mathbf{x})$ instead of as defined in (4).

4   Results

The proposed method was evaluated using both affine and non-rigid image registration as described in Sect. 2.2 and 2.2. For the non-rigid registration, diffeomorphic field accumulation was used for achieving diffeomorphic deformations. For some of the performed experiments, the proposed method was compared with its mono-modal counterpart on mono-modal data. The image data used in the experiments were obtained from BrainWeb [2].

In the first experiment, the proposed multi-modal method was compared with the previously presented mono-modal method using a linear polynomial expansion [4]. A single T1-weighted brain image was used as moving image. From the moving image, 20 fixed images were created using known affine transforms. The affine transforms were obtained according the following schema: the scaling factors within $[0.80, 1.30]$, the rotation factor within $[-\pi/4, \pi/4]$, and the translation factors within $[-15, 15]$. The moving image was then registered to each of the 20 fixed images using affine registration. The obtained affine transforms where then compared with the known transforms, yielding the results in Table 1, comparing the average error for each parameter type. The experiment was run twice, one time where noise was added (Gaussian noise with $\sigma = 200$) and one time without any added noise. Timing the experiment showed that the computation time increased with a mere 40% for adding the computational steps for multi-modal image registration, i.e. from approximately 1.0 seconds to 1.4 seconds.

Table 1. Results (average error and its standard deviation) for comparing accuracy per parameter type (scale, rotation and translation) of affine mono-modal registration with the proposed multi-modal registration on mono-modal data

Transformation factor	Scale	Rotation	Translation
Mono-modal (noise added)	1.0e-5 ± 1.1e-3	-5.8e-4 ± 2.4e-3	2.0e-4 ± 3.2e-2
Mono-modal	-4.1e-6 ± 4.9e-5	1.8e-5 ± 1.4e-4	6.9e-4 ± 3.2e-3
Multi-modal (noise added)	7.4e-4 ± 1.7e-3	-1.0e-3 ± 4.1e-3	1.6e-4 ± 5.3e-2
Multi-modal	-2.9e-5 ± 1.3e-4	-2.1e-6 ± 5.7e-5	6.3e-4 ± 3.3e-3

In the second experiment, the proposed multi-modal method was evaluated in the same manner as in the first experiment but on multi-modal data, i.e. on a T1-weighted image and a T2-weighted image, where the T2-weighted image was transformed according to a set of known affine transforms (in this experiment the rotation factor was limited to $[-\pi/6, \pi/6]$), and without running the mono-modal registration algorithm. The experiment was also executed twice, one time where noise was added and one time without any added noise (Gaussian noise with $\sigma = 200$). The results for this experiment are given in Table 2. Example images from the second experiment are depicted in Fig. 1. In both the experiments for affine registration, the number of scales and iterations per scale were kept constant, five scales and ten iterations per scale.

To evaluate the accuracy for non-rigid registration, a similar setup was used as for affine registration. A single image was deformed using 20 known displacement fields (randomly created and with an average displacement of 5 pixels and maximum displacement of 15 pixels). The original image was then registered to the deformed images and the obtained displacements fields were compared with the known displacement fields using the target registration error (TRE). As before, the experiment was first run on mono-modal data to compare the accuracy of multi-modal registration with mono-modal registration (with and without added

Table 2. Results (average error and its standard deviation) for comparing accuracy per parameter type (scale, rotation and translation) of affine multi-modal registration on multi-modal 2D data

Transformation factor	Scale	Rotation	Translation
With added noise	-4.1e-3 $\pm$ 6.8e-3	-5.5e-4 $\pm$ 4.3e-3	1.3e-2 $\pm$ 4.9e-2
Without added noise	-9.82e-4 $\pm$ 2.0e-4	-2.8e-4 $\pm$ 4.9e-4	-8.7e-3 $\pm$ 9.0e-3
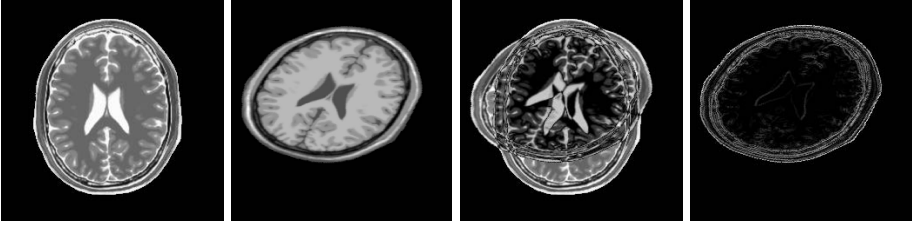
Fig. 1. An example of affine multi-modal registration, showing from left the moving image (T2-weighted), the fixed image (T1-weighted), the absolute difference between fixed and moving, and fixed and deformed for T2-weighted images

noise) and then it was run on multi-modal data (also with and without added noise). Also here Gaussian noise with $\sigma = 200$ was used. The results for this experiment are given in Table 3 and some example images are shown in Fig. 2. Also in this experiment, the number of scales and iterations per scale were kept constant, four scales and ten iterations per scale. Timing the experiment further showed that for non-rigid registration the computation time increased with only 30% for adding the computational steps for multi-modal image registration, i.e. from approximately 3.0 seconds to 3.8 seconds per registration.

Table 3. Average target registration error (TRE) of 20 non-rigid 2D registrations, comparing accuracy of multi-modal registration on mono-modal and multi-modal data with mono-modal registration on mono-modal data, with and without added noise. TRE is given in pixels.

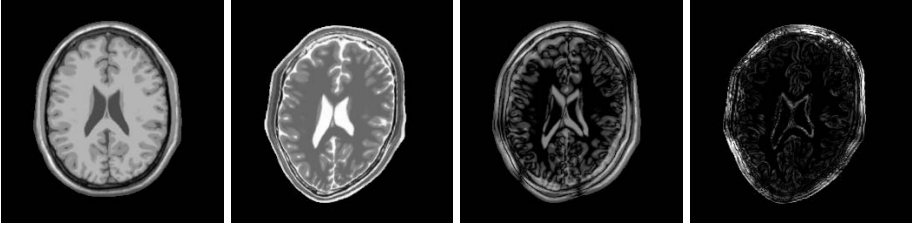| Registration	Mono	Multi	Multi
Data	Mono	Mono	Multi
With added noise	0.59	0.68	1.16
Without added noise	0.35	0.42	0.91

Fig. 2. An example of non-rigid multi-modal registration on multi-modal data, showing from left the moving image (T1-weighted), the fixed image (T2-weighted), the absolute difference image between fixed and moving, and fixed and deformed for T1-weighted images

5 Discussion

In this work, we have presented how polynomial expansion can be utilized for multi-modal image registration (both linear and non-rigid). This is achieved by incorporating a pixel-wise minimization of the conditional entropy, in order to estimate a pixel-wise intensity change that maximizes the mutual information.

The results in Tables 1 and 2 demonstrate the accuracy of the proposed method for multi-modal image registration but also when compared to its mono-modal counterpart in mono-modal image registration. In the case of affine registration of mono-modal image data, both algorithms achieves a descent registration accuracy. The accuracy of the registration decreases with a factor of approximately ten when adding Gaussian noise to the images but is still sufficient. For affine registration of multi-modal data, the accuracy is on a similar level as for registration of mono-modal data, hence sufficient for achieving a good registration. One thing that differed between the multi-modal registration of mono-modal data and multi-modal data was that the capture range for rotations decreased from $[-\pi/4, \pi/4]$ to $[-\pi/6, \pi/6]$ when running the registration on multi-modal data.

In the case of non-rigid registration, see Table 3, the accuracy of mono-modal registration and multi-modal registration of mono-modal data is also on similar levels and achieving sub-pixel accuracy. For non-rigid registration of multi-modal data the accuracy decreased by a factor of two and reached an average TRE of approximately one pixel.

A difference that could be noted between the mono-modal registration and the multi-modal registration was that the multi-modal was more iterative in its nature, i.e. the mono-modal algorithm converged within a few iterations, whereas the multi-modal often required at least twice as many iterations. Hence, the large number of iterations per scale used in the experiments.

Future work includes a more thorough investigation of how the number of channels affect the end result of the registration. In our experiments we only empirically decided how to set the number of channels, typically eight channels

for coarse scales and step-wise increasing up to 24 or 32 for the finest scales. For instance, a better use of the number of channels might decrease the TRE further for non-rigid registration and making sub-pixel accuracy also possible for multi-modal registration. Further evaluation and implementation in 3D will also be of interest to better compare with other existing algorithms for multi-modal image registration and on various types of multi-modal data.

References

1. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multimodality image registration based on information theory. In: XIVth International Conference on Information Processing in Medical Imaging, IPMI 1995, vol. 3, pp. 263–274. Kluwer Academic Publishers (1995)
2. Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A.: Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging 17(3), 463–468 (1998)
3. Farnebäck, G.: Polynomial expansion for orientation and motion estimation. Ph.D. thesis, Linköping University, Sweden (2002)
4. Farnebäck, G., Westin, C.-F.: Affine and Deformable Registration Based on Polynomial Expansion. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, Part I. LNCS, vol. 4190, pp. 857–864. Springer, Heidelberg (2006)
5. Forsberg, D., Andersson, M., Knutsson, H.: Non-rigid diffeomorphic image registration of medical images using polynomial expansion (2012), Accepted at the International Conference on Image Analysis and Recognition
6. Knutsson, H., Westin, C.F.: Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data. In: Proceedings of Computer Vision and Pattern Recognition 1993, New York City, USA, pp. 515–523 (1993)
7. Parzen, E.: On Estimation of a Probability Density Function and Mode. The Annals of Mathematical Statistics 33(3), 1065–1076 (1962)
8. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: A survey. IEEE Transactions on Medical Imaging 22(8), 986–1004 (2003)
9. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 27, 832–837 (1956)
10. Viola, P., Wells III., W.M.: Alignment by maximization of mutual information. In: Proceedings of Fifth International Conference on Computer Vision, pp. 16–23 (1995)
11. Wang, Y.J., Farnebck, G., Westin, C.F.: Multi-affine registration using local polynomial expansion. Journal of Zhejiang University - Science C 11, 495–503 (2010)