Multi-modal Medical Image Registration by Local Affine Transformations

Liliana Lo Presti and Marco La Cascia
DIID - University of Palermo, Italy

Keywords: Image Registration, Mutual Information, Medical Images.

Abstract: Image registration is the process of finding the geometric transformation that, applied to the *floating* image, gives the *registered* image with the highest similarity to the *reference* image. Registering a pair of images involves the definition of a similarity function in terms of the parameters of the geometric transformation that allows the registration. This paper proposes to register a pair of images by iteratively maximizing the empirical mutual information through coordinate gradient descent. Hence, the registered image is obtained by applying a sequence of local affine transformations. Rather than adopting a uniformly spaced grid to select image blocks to locally register, as done by state-of-the-art techniques, this paper proposes a method which is similar in spirit to boosting strategies used in classification. In this work, a probability distribution over the pixels of the registered image is maintained. At each pixel, this distribution represents the probability that a local affine transformation of a block centered on this pixel should be computed to improve the similarity between the registered and the reference images. The distribution is updated iteratively during the registration process to move probability mass towards pixels unaffected by the estimated local transformation. The paper presents preliminary results by a qualitative evaluation on several pairs of medical images acquired by different sources.

1 INTRODUCTION

Image registration, also known as image alignment, is the process of finding the geometric transformation T that, applied to the *floating* image f, gives the *registered* image r with the highest similarity to the *reference* image g. Image registration is used in medical imaging for allowing comparative and/or diachronic studies of the patients. Physicians can take advantage of image registration to monitor the course of diseases such as Alzheimer or multiple sclerosis cases, to compare the anatomical structures of different patients or to study anomalies in groups of individuals (Rueckert and Schnabel, 2010).

In the general image registration pipeline, an objective function is defined in terms of the parameters of the geometric transformation that transforms the floating image into the registered image. A numerical optimization algorithm is then applied to optimize the objective function with respect to the geometric transformation parameters.

State-of-the-art algorithms for image registration mainly differs in the kind of geometric transformation that is estimated for aligning the images and in the objective function. In particular, the work in (Ardizzone et al., 2007) applies a non-rigid transformation obtained by piecewise affine transformations estimated by local mutual information maximization. An image pyramidal approach allows a coarse-to-fine details registration. At each level of the pyramid, blocks of the floating image are aligned to the reference image, and changes in pixel coordinates are accumulated in a deformation field, which stores the displacement along the x and y directions (columns and rows in the image respectively) of each pixel of the floating image. To avoid a checkerboard effect at the edges of the aligned blocks on the registered images, the pixel displacements originated by the local affine transformation are smoothed by a filter before accumulating them into the deformation field.

The main limitation of the work in (Ardizzone et al., 2007) is the use of a uniform grid to estimate local affine transformations. Image blocks extracted based on the uniform grid are registered several times and, eventually, the step of the grid is decreased in order to deal with finer details of the images.

This approach has two main drawbacks: first, regions of the floating image that are already well aligned

Presti, L. and Cascia, M.
Multi-modal Medical Image Registration by Local Affine Transformations.
DOI: 10.5220/0006656405340540
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 534-540
ISBN: 978-989-758-276-9

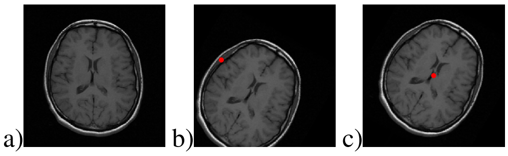


Figure 1: Effect of changing the origin of the coordinate system when applying a transformation. a) Original image; b) and c) images obtained by applying a rotation of 30 degrees while changing the origin of the coordinate system, which is represented by a red dot on the resulting images.

to the reference image may be processed several times by the method without gaining any benefits in terms of registration accuracy but, rather, growing the computational complexity of the registration process; second, geometrical transformations are applied in a local coordinate system centered in the selected block. Local transformations are affected by the choice of the origin of such local coordinate system. A fixed uniform grid strongly limits the flexibility of the piecewise affine transformations. To make clear the problem, we show in Fig. 1 the effects of the same rotation transformation applied in different coordinate system whose origin is depicted by a red dot. The use of a uniform grid limits the possibility of changing the origin of the coordinate system, thus limiting the strength of the deformation that can be achieved by adopting a piecewise affine transformation.

In this paper, we propose a method which is similar in spirit to boosting strategies used in classification. In boosting, a set of weak classifiers is trained to build stronger classifiers. The weak classifiers are trained sequentially on samples of the training set that are more difficult to classify. This is in general achieved by keeping a distribution on the training examples; this distribution is used to draw examples to train a weak classifier (Zhu et al., 2009). The distribution is update based on the weak classifier performance by moving probability mass towards those samples that are hard to classify.

In a similar way, in this paper we propose to keep a distribution over the pixels of the registered image. At each pixel, this distribution represents the probability that a local affine transformation of a block centered on this pixel should be computed to improve the similarity between the registered and the reference images. The distribution is updated iteratively during the registration process to move probability mass towards pixels unaffected by the estimated local transformation.

Local affine transformations are estimated by empirical mutual information maximization based on the seminal work in (Viola and Wells, 1997).

Fig. 2 shows the pipeline of the proposed method. The advantages of the proposed registration technique are mainly two:

- a local transformation is computed based on the probability that it might improve the similarity of registered and reference images;
- image registration is not constrained by a uniform grid, which adds more flexibility to the registration process.

We present preliminary results of our technique on a set of pairs of images acquired with different modalities. The plan of the work is as follows. In section 2 we review work on image registration; in section 3 we define the empirical mutual information that we adopt as objective function; in section 4, we present the adopted geometric transformation, while in section 5 we describe the optimization procedure. In section 6 we present experimental results and, finally, in section 7 we present conclusions and future work.

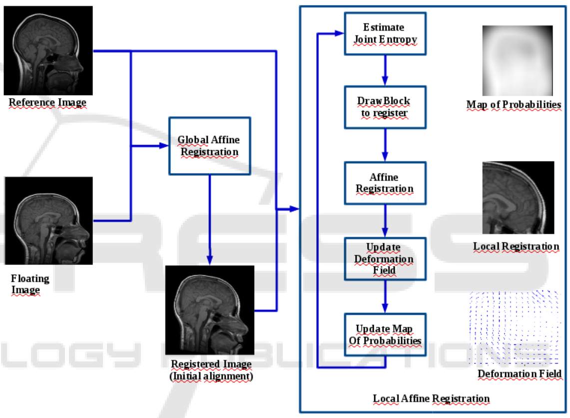


Figure 2: After pre-processing the images (noise reduction and estimation of a global affine transformation), the method estimates a sequence of local affine transformations. Based on the values of local joint entropies of the current solution and the reference image, a map of probabilities is computed. The map is used to draw the center of the image blocks to align. Hence, a local affine transformation is estimated by maximizing the local empirical mutual information. The local deformation is then used to update the deformation field. The process is repeated till convergence.

2 RELATED WORK

Image registration is a fascinating field that has attracted much attention in past years. Several methods have been proposed in literature, and they often differ in the initial hypotheses or prior knowledge available to solve the problem, in the adopted geometrical transformation or in the objective function to optimize.

Landmark and surface based registration (Crum et al., 2004; Fornefett et al., 2001) define a set of points or surfaces correspondences between the two images, and use this information to extend the correspondences to all other pixels in the floating image by interpolation. Identification of correspondences in images acquired by different modalities (such as PETs) is not always straightforward. Another approach is that of considering force fields (Pekar et al., 2006) or some physical model capable of representing the registration problem in terms of partial differential equations (i.e. in viscous fluid approach, image is modeled like a fluid and deformation comes from the solution of the Navier-Stokes equation (Fookes and Maeder, 2003)).

In works such as (Ardizzone et al., 2007; Viola and Wells, 1997), there are no assumptions about geometric shapes and spatial positions of the structures in the two images but, instead, it is assumed some relation among their intensity distribution functions.

This is especially convenient when dealing with medical images. Indeed, in mono-modal registration, images are acquired with the same type of exam (for example MR-PD); In this case, anatomical structures are represented with similar intensity distribution and comparison of the images to align can be performed by estimating the mean squared error of the intensities at the aligned pixels. Instead, in multi-modal registration, images are acquired by different modalities (for example MR-PD versus MR-T1). In this case, since the same anatomical structure is characterized by different intensity distributions in the two images, criteria based on statistical properties of the gray levels distribution function can be adopted.

The seminal work in (Viola and Wells, 1997) presents an image alignment technique derived from information theory. The technique adopts a framework for estimating the empirical entropy from a set of data samples. As we will detail in Section 3, entropy is not easy to differentiate. However, by adopting a Parzen window density estimation technique to approximate the gray level distribution of an image, empirical entropy can be made differentiable. Mutual information, which is a function of entropies, becomes differentiable as well. Nonetheless, optimization of the empirical mutual information is computationally intensive especially when huge sample sets are used to obtain a reliable estimation of the mutual information. Viola and Wells proposed then to use stochastic gradient descent to optimize empirical mutual information. At each step of their algorithm, a small sample set is used to align the images. Of course, the gradient they use to optimize their objective function is only an inaccurate estimate of the true gradient, but the expected value of the gradients will tend to approach the true gradient. Nowadays, we know that stochastic gradient is very effective in numerical optimization of non-linear functions of a large number of parameters.

Our work is inspired to both the work in (Ardizzone et al., 2007; Viola and Wells, 1997). We have adopted the geometrical transformation proposed in (Ardizzone et al., 2007), which is a piecewise affine transformation. Effects of the estimated local transformations are accumulated in a deformation field that stores, for each pixel of the registered image, the position of the corresponding pixel in the floating image. In (Ardizzone et al., 2007), local transformations are estimated by maximizing the mutual information computed in image blocks. However, mutual information is highly non-linear and is not differentiable without a parametric model of the gray level distribution. Hence, in (Ardizzone et al., 2007), a gradient-free optimization algorithm is used to optimize the mutual information. Such kind of approach is heavily affected by the initial solution to the optimization problem. A more reliable initial solution is found by applying a coarse-to-fine registration strategy, which is implemented in (Ardizzone et al., 2007) by using a pyramid of images.

In (Yang et al., 2015), optimization of the normalized mutual information is carried out by combining the limited memory Broyden-Fletcher-Goldfarb-Shanno with boundaries (L-BFGS-B) with cat swarm optimization (CSO).

In our work, rather than optimizing the mutual information, we follow the work in (Viola and Wells, 1997) and optimize the empirical mutual information. In contrast to the work in (Viola and Wells, 1997), we estimate a sequence of local transformations that are used to compute the deformation field of the registered image. In contrast to (Ardizzone et al., 2007), we do not adopt a uniform grid to select the image blocks to register. Instead, we use an approach that is similar in spirit to boosting, and sample the image block to register during the alignment process.

3 EMPIRICAL MUTUAL INFORMATION

Entropy $H(X)$ is a measure derived by information theory that allows us to quantify the randomness of a random variable X, and is defined as follows:

$$H(X) = -\sum_{\forall x} p(X = x) \cdot \log p(X = x) \tag{1}$$

In image registration, the joint entropy $H(X,Y)$ is used to measure the extent to which two random variables – Y, the gray levels in the registered image, and X, the gray levels in the reference image – are dependent. Low values of the entropy indicate a strong dependency of Y and X, while higher values indicate a higher level of randomness.

Mutual information is a measure of the reduction in the entropy of Y given X (Viola and Wells, 1997), and is defined as follows:

$$I(X,Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y). \tag{2}$$

Mutual information has already been adopted in image registration (Yang et al., 2015; Ardizzone et al., 2007; Viola and Wells, 1997) and it showed to be highly non-linear and difficult to differentiate. Such difficulty arises from the lack in our problem of a parametric representation of the probability distribution of the gray levels of the image. This affects the optimization process, and techniques that do not require an explicit gradient of the objective function are often adopted. As an example, the work in (Ardizzone et al., 2007) adopts the Nelder-Mead simplex-based optimization algorithm (Nelder and Mead, 1965). However, without gradient information, the found solution strongly depends on the initial guess, which can be hard to find in image registration.

To deal with such difficulty, Viola proposed in (Viola and Wells, 1997) to adopt an approximated but differentiable expression of the mutual information that relies on empirical entropies. Given a sample set A, the empirical entropy of a variable X is defined as the expected value of the log-likelihood:

$$H_A(X) = -E_A[\log p(X)] = -\frac{1}{|A|}\sum_{x\in A}\log p(x). \tag{3}$$

However, this representation is still difficult to differentiate and, in (Viola and Wells, 1997) it is proposed to adopt Parzen window density estimation(Bishop, 2006) to approximate the probability distribution of X.

Given another sample set B, Parzen window density estimation allows to approximate a distribution as follows:

$$P^*(X = x) = \frac{1}{|B|}\sum_{x_b\in B} R(x - x_b) \tag{4}$$

where $R(\cdot)$ is a density function, such as a Gaussian density function ($G(\cdot)$).

The empirical entropy estimated by Parzen window density estimation takes the following form:

$$H^*(X) = -\frac{1}{|A|}\sum_{x_a\in A}\log\frac{1}{|B|}\sum_{x_b\in B} G(x_a - x_b). \tag{5}$$

Such form of the empirical entropy is differentiable and is adopted in this paper to estimate the empirical mutual information.

4 NON-RIGID IMAGE REGISTRATION

Figure 2 shows the approach we followed to compute a sequence of local affine transformations. First, a probability map is computed by normalizing the values of the joint entropy of the current registered image and the reference image within a block.

The map is used to sample a pixel. An image block centered on the sampled pixel is then considered. From this block, the sets A and B are sampled and used to estimate the empirical mutual information. A local affine registration is then performed by applying coordinate gradient descent. Once the optimal local transformation has been estimated, it is used to update the global deformation field by smoothing the effects of the transformation outside the registered image block.

The probability map is then updated by decreasing the values of the probabilities of the pixels within the registered image block. The whole process is repeated by sampling a new pixel point and registering a new image block.

In the following we provide more details about each of the above mentioned steps.

4.1 Affine Transformation

In image registration, the goal is finding the linear transformation T that allows to map a pixel x^{reg} of the registered image to a pixel x^{float} in the floating image, that is

$$x^{float} = T \cdot x^{reg}. \tag{6}$$

Since coordinate pixel should be integer, the gray level to be assigned to the pixel x^{reg} is estimated by interpolating the gray levels of the pixels in a neighborhood of the point x^{float}. Nearest neighbor or bilinear interpolation are often used to limit computational complexity of the registered image estimation.

Affine transformations are linear transformations that originate from a combination of scaling, rotation, shearing and translation transformations. An affine transformation can be represented by a 3×3 matrix and is applied to the homogeneous coordinates of the pixels.

In our implementation, we have considered affine transformations that can be represented as a composition of simpler geometric transformations. For example, a geometric transformation T might be estimated as follows:

$$T = R(\theta) \cdot S_h(k) \cdot S(s_x, s_y) \cdot T(t_x, t_y), \tag{7}$$

where $R(\theta)$ is a rotation matrix that rotates the image by an angle θ counter-clockwise; $S_h(k)$ performs a shearing transformation; $S(s_x, s_y)$ performs a scaling in the x and y directions; $T(t_x, t_y)$ does perform a translation transformation of the image.

Differently than (Ardizzone et al., 2007), where an affine transformation is represented by 6 parameters representing the first two rows of matrix T, in this paper we explicitly estimate parameters $\theta, k, s_x, s_y, t_x, t_y$ and later assemble the affine transformations. Our approach is more robust since it guarantees that the estimated transformation is affine and, hence, is invertible. This cannot be easily guaranteed by estimating directly the first two rows of matrix T.

4.2 Deformation Field

The deformation field we use to store the displacements of each pixel is similar to that adopted in (Ardizzone et al., 2007; Pekar et al., 2006). When deformation fields are adopted, it is possible to compute the pixel coordinates of the floating image that correspond to the pixel of the registered images as follows:

$$x^{float} = x^{reg} + \Delta x \tag{8}$$

$$y^{float} = y^{reg} + \Delta y. \tag{9}$$

The displacements $\Delta x, \Delta y$ are estimated iteratively by considering the deformations yielded by each local affine transformation.

To avoid checkerboard artifacts, a smoothing mask is adopted to limit the effects of the affine transformation locally to the selected image block. The mask is obtained by setting to 1 all pixels within the image block, and setting to 0 the remaining ones. The mask is then filtered by a Gaussian smoothing filter whose effect is that of decreasing the influence of the affine transformation near the border of the image block.

4.3 Probability Map

Rather than adopting a uniform grid for processing images, we sample the center of the image block to register. Whilst we might sample image block from a uniform distribution, in order to put more effort in registering those areas of the images where strong misalignments are observable, we build a probability map that represents the probability that aligning an image block centered in a point would improve the registration accuracy.

This probability map M is computed by estimating the joint entropy of the reference and the floating images in small blocks (by a sliding window approach). Joint entropy measures the randomness of the gray levels in the two selected image blocks (one from the reference image and the other one from the floating image). Hence, joint entropy may provide useful hints to locate areas that "need more deformation". To initialize the map M, at each pixel (i, j) we consider a block $W_{i,j}$ centered at (i, j), then we estimate the joint entropy:

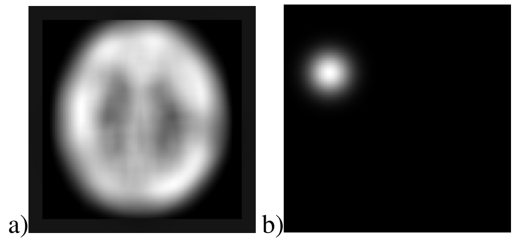


Figure 3: a) Probability map M used for sampling image blocks to register. Higher values in the map indicate a high value of the joint entropy and, hence, might indicate the need to align the image blocks. b) The map G: weights used to update the map are distributed as a Gaussian centered in the processed image block.

$$M(i, j) = H(f_{|W_{i,j}}, g_{|W_{i,j}}). \tag{10}$$

To get a probability distribution, we normalize M in such a way that all values in M sum to 1:

$$M(i, j) = \frac{M(i, j)}{\sum_{i=1}^{r} \sum_{j=1}^{c} M(i, j)} \tag{11}$$

where r and c represent the number of rows and columns of the registered image respectively.

Estimating the probability map at each iteration is computationally complex. We estimate the map once at the beginning of the registration process. Every time we draw an image block from the map, we perform a registration process to estimate a local transformation and update the map in order to move probability mass from the aligned area. This is achieved by decreasing the map values by means of a local weight map G. The map G is obtained by normalizing the probability values of a Gaussian distribution centered on the point drawn from the map. The updating of the map follows the following rule:

$$M(i, j) = M(i, j) \cdot (1 - \alpha \cdot G(i, j)). \tag{12}$$

After the map is updated, it is renormalized in order to get a probability distribution. Fig. 3, shows a probability map initialized by the joint entropy values in small image blocks.

5 OPTIMIZATION

In our framework, we used a coordinate descent approach to optimize the empirical mutual information.

At each iteration, we maximize the objective function with respect to a single parameter $(\theta, k, s_x, s_y, t_x, t_y)$. Among the 6 transformations, we accept the one that yields the highest value of (non empirical) mutual information.

Maximization of the empirical mutual information is achieved by using the limited memory Broyden-FletcherGoldfarbShanno. In contrast to (Viola and Wells, 1997), which adopts stochastic gradient descent by sampling different sets A and B to estimate the empirical mutual information, we sample the two sets and use them to estimate a local optimal solution. Then, we repeat the transformation estimation while varying sample sets. In our implementation, small sample sets (of about 10 samples each one) proved to be sufficient for gradually maximizing the mutual information. The small size of the sample sets reduces the computational complexity of the objective function evaluation.

6 EXPERIMENTAL RESULTS

Preliminary evaluation of the proposed method has been carried out on both synthetic and real data for a total of 20 pairs of images. All images were 8-bit images with 256×256 spatial resolution and acquired by different sources. Following (Ardizzone et al., 2007), in some experiments we used an atlas from the Simulated Brain Database (SBD), default parameters were used to generate these volumes. Other images have been extracted from Alzheimer and multiple sclerosis diseased volumes available on the whole brain atlas at http://www.med.harvard.edu/AANLIB/home.html.

When running experiments, we set α to 0.1. The size of the image block $W_{i,j}$ is set to $\frac{1}{4}$ of the size of the floating image. We show a qualitative evaluation by reporting in Table 1 some of our results. In the table, the first column shows the floating image, the second column shows the reference image, the third column shows the registered image.

7 CONCLUSIONS AND FUTURE WORK

This paper proposes a new registration algorithm in which image deformation is iteratively computed by maximizing the local empirical mutual information. In contrast to other work at the-state-of-the-art, which use uniformly sampled image blocks, we use a strategy inspired to the boosting approach popular in classification. We update and maintain a probability distribution over the pixels of the registered image. At each point, such distribution models the probability that, by registering image blocks centered on that point, the similarity between the registered and the reference images would improve.

This approach increases the flexibility and the power of the piecewise affine transformation by avoiding that the transformation is constrained to a uniformly spaced grid.

Testing of our method has been conducted on multi-modal images of different patients, and qualitatively accurate results have also been obtained on inter-patient volumes. To deal with occasional failures of the local affine transformation estimation method, local transformations are accepted and globally applied only whether global mutual information (not the empirical one) increases.

Our iterative approach is repeated till convergence, that is until the maximal number of iterations is reached or only negligible changes in the mutual information can be measured.

In our approach, the size (or spatial scale) of the image block is fixed. We have experimented with different values of the spatial scales and found out that too small size of the image blocks may lead to degenerate solutions. In future work we will investigate the possibility of introducing within the framework a method for automatically selecting the scale of the local transformation and, hence, the size of the image blocks. This might be achieved, for example, by considering an energy function upon the probability map that we used to sample the image block positions.

ACKNOWLEDGEMENT

This work was partially supported by the grant DM. 46965 LATO CIPE2.

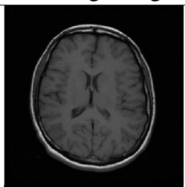
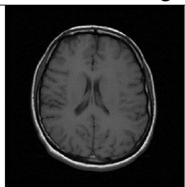
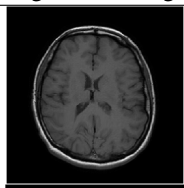
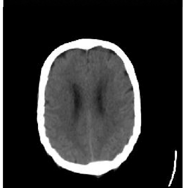
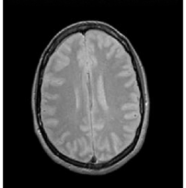
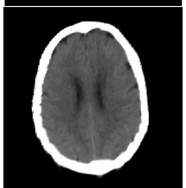
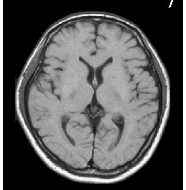
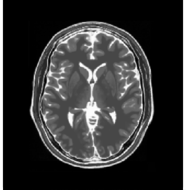
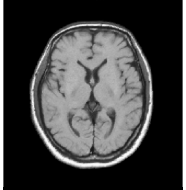
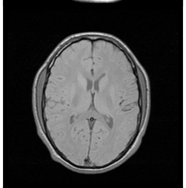
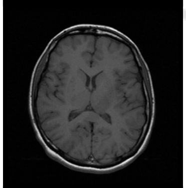
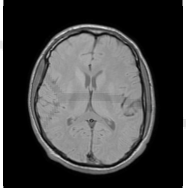
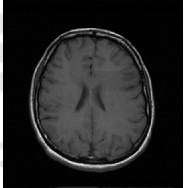
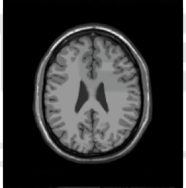
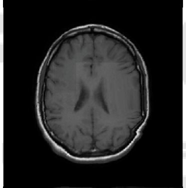
REFERENCES

Ardizzone, E., Gambino, O., La Cascia, M., Lo Presti, L., and Pirrone, R. (2007). Multi-modal non-rigid registration of medical images based on mutual information maximization. In *Proc. of International Conference on Image Analysis and Processing (ICIAP 2007)*, pages 743–750. IEEE.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Crum, W. R., Hartkens, T., and Hill, D. (2004). Non-rigid image registration: theory and practice. *The British journal of radiology*, 77(suppl_2):S140–S153.

Table 1: Qualitative evaluation of the proposed registration technique.

Floating Image	Reference Image	Registered Image	Notes
			Intra-patient MR-T1 images
			Inter-patient CT vs MR-PD images
			inter-patient MR-T1 vs MR-PD images
			Inter-patient MR-PD vs MR-T1 images
			MR-T1 vs MR-T1 (atlas) images
			intra-patient MR-T1 images of knee

Fookes, C. B. and Maeder, A. J. (2003). Comparison of popular non-rigid image registration techniques and a new hybrid mutual information-based fluid algorithm.

Fornefett, M., Rohr, K., and Stiehl, H. S. (2001). Radial basis functions with compact support for elastic registration of medical images. *Image and vision computing*, 19(1):87–96.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

Pekar, V., Gladilin, E., and Rohr, K. (2006). An adaptive irregular grid approach for 3d deformable image registration. *Physics in Medicine and Biology*, 51(2):361.

Rueckert, D. and Schnabel, J. A. (2010). Medical image registration. In *Biomedical Image Processing*, pages 131–154. Springer.

Viola, P. and Wells, W. (1997). Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154.

Yang, F., Ding, M., Zhang, X., Hou, W., and Zhong, C. (2015). Non-rigid multi-modal medical image registration by combining l-bfgs-b with cat swarm optimization. *Information sciences*, 316:440–456.

Zhu, J., Zou, H., Rosset, S., Hastie, T., et al. (2009). Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360.