BMVC99

Non-Linear Registration using a Finite Element Method

M. Valdivieso-Casique and S. Arridge
Department of Computer Science
University College London, London WC1 6BT
{mvaldivi, s.arridge}@cs.ucl.ac.uk

Abstract

An automated method to register Medical Images is presented in this work. This new technique employs an iterative warping scheme and is performed, either over fully pixel information or using geometrical features such as ridges. The warp at each iteration is given by a number of local Finite Element transformations which provide a non-linear global warp. Results show that this technique can be applied satisfactorily to images with a severe level of distortions.

1 Introduction

Image registration is an incipient image processing technique in continuous development and research. It is an important technique for a great variety of applications such as aerial image analysis, stereo vision, motion analysis and medical image diagnosis.

In medical imaging, image registration is used in order to enhance the information content of the data and thus, its clinical value. Images from the same or different modalities are registered, providing complementary information to the analyst. Due to its importance, there is a need to find a good registration algorithm which does not require human interaction.

There exists a variety of registration methods that have been proposed, a good classification being found in [14]. In our work we pay main attention to two aspects of the mentioned classification, the origin of the image properties and the elasticity of the transformation used to transform the coordinates of the source to the target image. Intrinsic image properties like pixel intensities and ridges are used here. Pixel intensity techniques are mainly used in monomodal applications, since the image intensity changes in different modalities. Ridges on the other hand, convey meaningful structural information across modalities, as well as providing data reduction and are proven to be viable for registration [1, 10, 13].

Normally the variations in images involved in the registration are very complex and cannot be modelled through a rigid body transformation, which contains zero elasticity. To overcome this complexity, our algorithm makes use of a finite element transformation (FET), which produces a warp deformation and is computationally less expensive than other warping techniques like the thin plate spline [3].

BMVC 1999 doi:10.5244/C.13.22

2 Implementation

The method here described is a modification of Collins et al's approach[5] in which images to be registered are divided in a number of non overlapping blocks or subimages and their centre is used as a control point for the transformation. We do the same process here for two approaches, using pixel intensity or ridges with a hierarchical iterative scheme. This scheme starts by splitting the images into a minimum number of blocks and, a similarity metric is minimised to give a displacement vector of the movement of pixels within the block. The FET is computed between correspondent blocks which gives a number of continuous local warps that overall provide a non-linear global one. At the next iteration, the deformed image in the previous iteration is partitioned by a larger number of blocks, and the FET is computed again. This iterative process finishes when a similarity metric (mean squared error, mutual information, correlation) of the deformed and target images at the current iteration is greater than at the previous one or, the size of the block is too small. A more detailed explanation is given below.

The first part of our technique is the identification of the control points to be involved in the registration. As has been mentioned, our iterative process starts splitting the image into four non overlapping blocks (2 partitions in each axes), so four points are selected. Next step is to have a measure of the movement of these points in the image to be registered (the target). Here is where a difference exists between the use of pixel intensity or ridges, the way in which the similarity metric is computed. Both are explained separately in the next subsections.

2.1 Computing the similarity metric for pixel intensity

2.1.1 Cross-correlation

For the pixel intensity case, the similarity metric is obtained by means of applying a normalised cross-correlation between each block and the target.

$$C(u,v) = \frac{\sum_{x,y} B_{(i,j)}(x,y)T(x-u,y-v)}{\sqrt{\sum_{x,y} T^2(x-u,y-v)}} \qquad (1)$$

In equation 1 (u,v) are the coordinates of the entire target image, (x,y) the coordinates of the block (i,j), where (i,j) stand for the position of the block in the source image. The block is first centred at the centre of an imaginary correspondent block in the target image and then translated over a search window surrounding the imaginary block. If the source block matches the target one exactly, except for an intensity scale factor at a translation (u_{max}, v_{max}), the cross correlation would have its peak at $C(u_{max}, v_{max})$.

In our scheme a search window around the block is used. The size of this window is 7 pixels greater than the blocksize. Therefore, 49 correlation values are computed for each block. This number can be reduced in each iteration, since the algorithm starts converging to the solution, and every source block is more similar to its counterpart in the target image.

It is worth noting the limitations of the cross correlation scheme, first of all because it can only take account of a block being translated, and secondly because

in a multimodal application pixels are not similar due to the physical realities of the images to be registered.

2.1.2 Local Mutual Information

The mutual information metric (MI) cannot be applied to local registration unless the local joint histogram has sufficient statistical power, for example, the size of the window over which the MI is computed is very large [11]. This does not apply here, since at every iteration the size of the block is smaller. To overcome this problem we use the approach by Maintz et al [11]. In their work, they extract the conditional probability densities $p(x \mid y)$, i.e. given a grey value x from the source image S we compute the probability that this value corresponds to any grey value y ocurrying in target image T.

After having done this, we proceed to evaluate the local gray value correspondence probability cp between source and target windows, as has been done for the cross-correlation.

$$cp(u,v) = \sum_{x,y} p(T(x-u, y-v) \mid B_{(i,j)}(x,y)) \tag{2}$$

The notation is the same as for equation 1. The coordinates (u,v) that maximise cp represent the best matching between blocks. This measure can be applied to the multimodal case, however there are some conditions that have to be met. Firstly, the images must present variations due to small local elastic deformations. Secondly, the conditional probabilities from the joint histogram must be adequate approximations of the "real probabilities" after elastic registration.

It is important to point out that both measures (cross-correlation and local mutual information) are only computed over the overlapping part of the images and thus, the pixels falling outside the borders are discarded.

2.2 Computing the similarity metric for ridges

By obtaining the ridges of the images we are reducing the amount of information and thus, the computational time invested to perform the registration.

The way we obtain the displacement vectors for each of the blocks is through the following procedure. Let $R(S)$ be the ridge image of the source. We compute the chamfer distance transform [4] of the target ridge-like image $D[R(T)]$. According to the warping scheme, we divide $R(S)$ and $D[R(T)]$ into blocks, with the block size in the latter being 10 pixels larger so that they are overlapping. This is done in order to take account of the movement of ridges from one block to another.

The goal of this process is to find a transformation that maps the ridges in the source blocks to those in the target. The model of the transformation we employ is an affine transformation, which for the 2D case can be expressed as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{3}$$

This transformation will map any ridge point (x,y) in the source image to a coordinate (x',y') in the target image. For a given affine transformation and for

every ridge point in the source, we compute the sum of the values of the distanced image at (x', y'), to define a metric $M = \sum_{x'y'} D[R(T)]$, where T is the target image, R the ridge finding operator and D the distance operator. The coordinates (x', y') are given by equation 3.where T is the target image, R the ridge finding operator and D the chamfer distance operator. The coordinates (x', y') are given by equation 3.

Next point is optimize the measure M and find the minimum value of it. A value of $M = 0$ would mean that the matching of the ridges is perfect, however this is not possible since the number of ridges in the source differs that from the target. Choosing the transformation that minimises metric M represents the most of the ridges in the source block fall in a small distance from the ridges in the target. The optimisation procedure is implemented by using the simplex method [7].

The optimisation problem then is a multidimensional minimisation over 6 variables (the parameters of the affine). The simplex method requires $n + 1$ estimates in order to solve an n dimensional problem. Our initial estimate of the affine transformation comes from a Principal Component match between correspondent source and target blocks, while the other n come from varying each of the parameters of the aforementioned affine by a small offset.

After the optimisation is done, we end up having an affine transform for each of the blocks. This affine transformation cannot be applied directly to the original block source image since the deformation would not be continuous. To overcome this problem we have developed the next deformation process.

2.3 Deformation

After the similarity metric is computed, the displacement given by the crosscorrelation or the affine transformation is applied only to the centre of the block, so that we end up having four points in the source image which are displaced from the original one, as is shown in Figure 1.

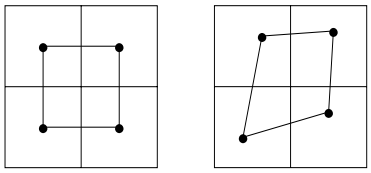

Figure 1: Centres of each block before and after applying the displacement.

In Figure 1 the initial square has been deformed to a quadrilateral. To make a continuous warp we employ a Finite Element technique which provides a bilinear mapping from a point within the square into a point within the quadrilateral, as illustrated in Figure 2.

The formula for the finite element transformation from a unit square to a quadrilateral is given by:

$$\begin{aligned} x &= (1-\epsilon)(1-\eta)P_{1x} + \epsilon(1-\eta)P_{2x} + (1-\epsilon)\eta P_{3x} + \epsilon\eta P_{4x} \\ y &= (1-\epsilon)(1-\eta)P_{1y} + \epsilon(1-\eta)P_{2y} + (1-\epsilon)\eta P_{3y} + \epsilon\eta P_{4y} \\ & 0 \leq \epsilon \leq 1 \qquad 0 \leq \eta \leq 1 \end{aligned} \tag{4}$$

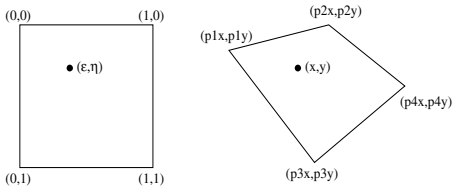

Figure 2: FET mapping.

For mapping general quadrilaterals we first transform to a unit square, and from this to the target quadrilateral, as shown in Figure 3.

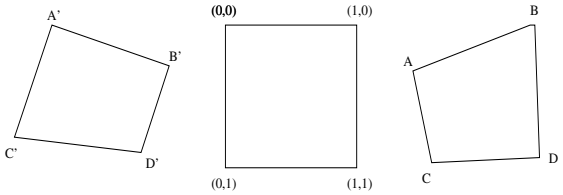

Figure 3: Transformation from quadrilateral to quadrilateral.

For this step we need the inverse of equation 4. A Newton-Raphson process was implemented to find this inverse. Having done this we are capable of transforming any quadrilateral into another one and it is possible to warp all the regions in our source image.

This kind of warp is simpler than the Thin Plate Spline [3] and computationally less expensive since less operations are needed. Another intention of using this transformation instead of the TPS is that the latter technique leads to a global unique warp whereas with the Finite Element method, the global warp is a composition of local warps. Moreover, points in the border of the image must remain fixed, and the TPS stretches all points as a consequence of its global nature.

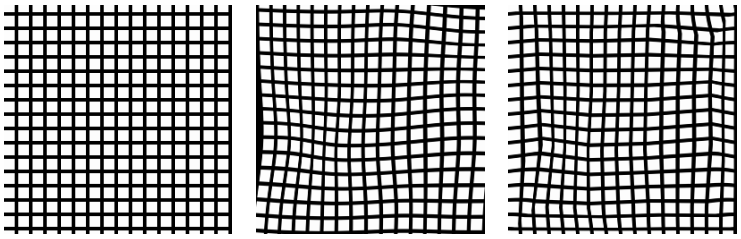

Figure 4: From left to right. Original, resulting deformed image using the TPS, FET.

Figure 4a) shows a synthetic image of a grid (64*64 pixels). The same set of control points was used to define both a TPS and FET warping procedures. In 4b) the movement of all the pixels within the image (including borders) which produces a dark gap in the left side, representing points outside the support of the original image.

We ensure border preservation by defining particular quadrilaterals around the border, where two vertices are fixed and two are movable, Figure 5. These warps preserve the continuity in the image, since pixels within the border of adjacent blocks are mapped into the same position in the global coordinate system, in other words, final conditions in the first adjacent warp are initial conditions in the second.

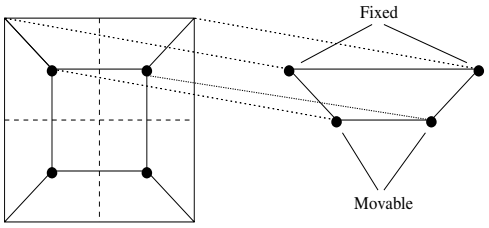

Figure 5: Different warps are performed outside the original quadrilateral.

The finite element mapping produces a transformation in which the coordinates are no longer integer points, therefore it is necessary to resample the image. The resample is accomplished by means of cubic spline interpolation, in order to ensure a continuous deformation.

After all five warps for the first iteration are obtained and the image is resampled, we perform an iterative process in which the previous result serves as the basis for the next iteration but with the number of divisions along each axes doubled.

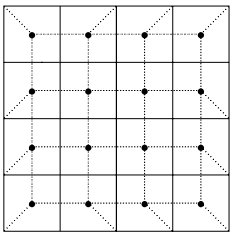

Figure 6: Points involved in the second iteration and different warping areas.

In Figure 6 we can see the number of local warps that need to be performed in the second iteration. The iterative scheme terminates when the similarity measure at the current iteration is greater than the previous one.

3 Results

Some results are shown below for synthetic and real images:

3.1 Test on synthetic images

In Figure 7(left) we show a synthetic image of 64*64 pixels consisting of a dark rectangle over a light background as the source image, while the target 7(right),

consists of a deformed version using the Thin Plate Spline(TPS) and 32 landmarks.

Figure 7: Synthetic images. (Left) Source, (Right) Target

The test of the algorithm on these synthetic images, using the cross-correlation approach, provided the results in the column entitled "Synthetic" of Table 1. This data contains the mean squared error (MSE) of the resultant image at each iteration in comparison with the target. Figure 8 on the other hand, depicts the deformation process of the source image in the iterative scheme.

Figure 8: First 5 iterations of the algorithm applied to the synthetic image pair

3.2 Test on real data

Next pair ($pairB$) are 256*256 MRI images (Figure 9). The source was deformed using the TPS and 16 landmarks to produce the target image. The severe local distortions in the target are quite noticeable.

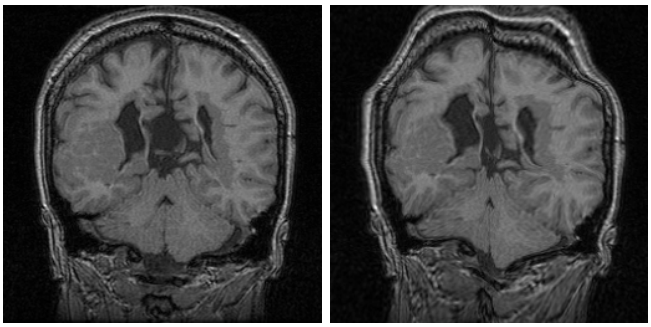

Figure 9: Image pair B, (Left) Source and (Right) Target

This pair of images served as input to our algorithm using the local mutual information approach. The reduction in the similarity metric (MSE) for this pair is also shown in Table 1, while the series of warped images from the first five iterations are depicted in Figure 10.

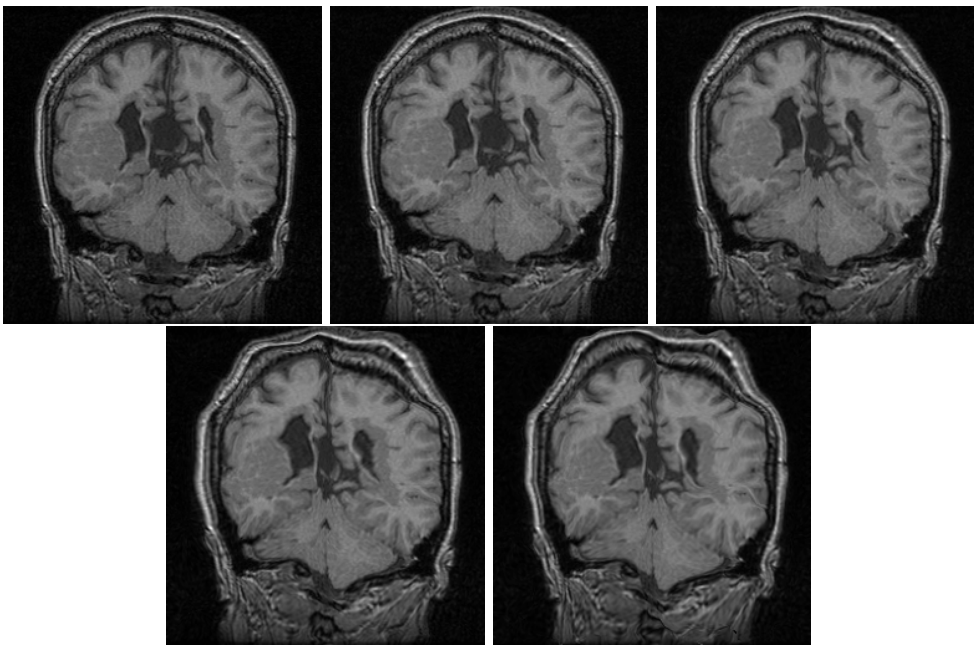

Figure 10: Warped images at the first five iterations using the local mutual information approach

Figure 11, corresponds to *pairC* images. These images contain two properties that make them ideal for testing the ridge-based approach. Firstly, they present relevant local deformations, like those found in the ventricles. Secondly, they contain different graylevel scale, as can be seen in the composite images in Figure 12. This makes them inappropriate for the cross-correlation approach. In figure 12 (left) the result at the last iteration of the algorithm is shown. Figures 12 (centre and right) depict the composite between the original source and target C, and the registered image with the target, respectively.

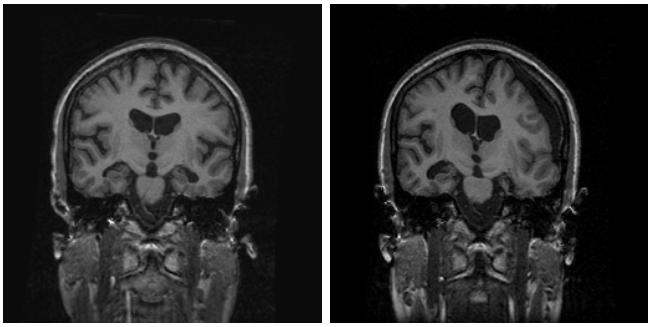

Figure 11: pair C, (Left)Source and (Right)Target

In Table 1 the decrease in the MSE for the first two pair of images is evi-

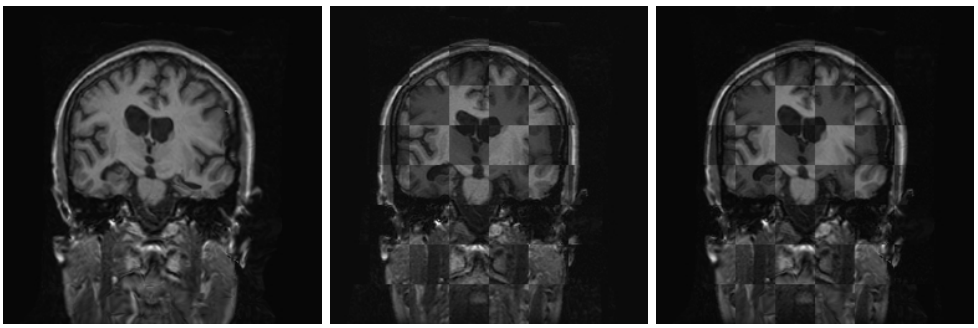

Figure 12: Registered Image (Left), composite images of sourceC and TargetC (Centre), composite of registered and TargetC (Right)

	MSE		
Iteration	Synthetic	Pair B	Pair C
0	40.715	23.6419	54.3557
1	34.9199	22.679	52.7988
2	33.4544	21.6005	52.208
3	24.6298	18.4217	49.9876
4	20.6473	13.8778	48.9237
5	16.1964	10.0338	47.1226

Table 1: Computed MSE for the three approaches

dent, from 40.71 to 16.19, and from 23.64 to 10.03, which gives a reduction of 61% and 58%, respectively. These results make this algorithm very promising for monomodal registration in images with severe level of distortions. The ridge-based approach did not perform as well in reducing the MSE, however the composite image 12(right) shows that the ventricles were perfectly matched. Moreover, the ridge-based approach can be used for multimodal registration such as between MRI and CT, whereas intensity based transformations cannot as the intensity varies from one modality to another.

4 Conclusions

We have presented a simple hierarchical non-linear registration algorithm that is faster than other techniques like the TPS. Use of geometrical features and pixel intensity means that it will be applicable to both, mono and multimodal problems. The implementation for 3D images does not differ a great deal from the actual one.

We are currently investigating the addition of edges in conjunction with ridges, to compute the affine transform in subsection 2.2. We are also studying a scale space hierarchy where the final warp at coarse scale is used to initialise a warp at a lower scale.

References

[1] N Ayache. Medical Computer Vision, Virtual Reality and Robotics. Image and Vision Computing. 13(4), pp 295-313. 1995.

[2] R Bajcsy et al. Multiresolution Elastic Matching. Computer Vision, Graphics and Image Processing. 46(1). pp 1-21. 1989.

[3] F Bookstein. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Anlysis and Machine Intelligence. 2(6).pp 567-585. 1989.

[4] G Borgefors. Distance Transformations in Arbitrary Dimensions. Computer Vision, Graphics and Image Processing. 27(1), pp 321-345. 1984.

[5] D Collins et al. Evans. An Automated 3D Non-Linear image deformation procedure for Determination of Gross Morphometric Variability in Human Brain. SPIE 2359, pp 180-190. 1994.

[6] L Brown. A Survey of Image Registration Techniques. ACM Computing Surveys, 24(4). 1992.

[7] W Ledermann. Handbook of applicable Mathematics. John Wiley and Sons, 1981.

[8] T Lindeberg. Scale-Space theory in Computers. Kluwer Academic Publishers, 1994.

[9] B ter Haar Romeny et al. Scale Space: its Natural Operators and Differential Invariants. Proceedings IPMI 1991.

[10] J Maintz et al. Evaluation of Ridge Seeking Operators for Multimodality Medical Image Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 4, April 1996.

[11] J Maintz, et al. General Multimodal Elastic Registration Based on Mutual Information. SPIE 3338, pp 144-154. 1998.

[12] M Valdivieso-Casique et al. Hierarchical Non-Linear Deformation with the Use of Finite Element Transformations. Proceedings of Computer Assisted Radiology and Surgery, CARS'99. 1999.

[13] P van den Elsen et al. Automatic Registration of CT and MR Brain Images Using Correlation of Geometrical Features. IEEE Transactions on Medical Imaging, Vol. 14, No. 2, June 1995. Pages 384-396.

[14] P van den Elsen et al. Medical Image Matching- A review with Classification. 1(1), pp 26-39. 1993.

[15] W Wells et al. Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1(1) pp 35-51. 1996.