31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

A New Multi-Modal Similarity Measure for Fast Gradient-Based 2D-3D Image Registration

Mark R. Pickering, *Member, IEEE,* Abdullah A. Muhit, *Student Member, IEEE,* Jennie M. Scarvell, and Paul N. Smith

***Abstract*—2D-3D image registration has been adopted in many clinical applications such as image-guided surgery and the kinematic analysis of bones in knee and ankle joints. In this paper we propose a new single-plane 2D-3D registration algorithm which requires far less iteration than previous techniques. The new algorithm includes a new multi-modal similarity measure and a novel technique for the analytic calculation of the required gradients. Our experimental results show that, when compared to existing gradient and non-gradient based techniques, the proposed algorithm has a wider range of initial poses for which registration can be achieved and requires significantly fewer iterations to converge to the true 3D position of the anatomical structure.**

I. Introduction

Image registration is the process of spatially aligning one image to another. Registration algorithms consist of two main components: a similarity measure and an optimization technique [1].

If the images to be registered are captured using different sensors, multi-modal similarity measures such as Mutual Information (MI), Cross-Correlation, Correlation Ratio [2, 3] and more recently Cross-Correlation Residual Entropy (CCRE) [4] can be adopted. All these similarity measures quantify the relationship between two images using probability distributions rather than intensity values.

Optimization techniques can be categorized as either gradient or non-gradient-based approaches. The non-gradient-based approaches do not require a numerical or analytic calculation of the gradient of the similarity measure at each iteration of the optimization search. The most commonly used non-gradient approaches are Powell's conjugate direction search and the downhill simplex method. The gradient-based approaches include: steepest descent, quasi-Newton, conjugate gradient and non-linear least squares methods (which include the classic Gauss-Newton and Levenberg-Marquardt algorithms) [1]. These approaches can be further divided into those that calculate the required gradients numerically and those that use an analytic approach. Numerically calculating the gradients involves perturbing the position of the 3D volume in a small positive and negative direction from the current position for every 3D transform parameter. Analytically calculating the required gradients at each iteration does not require these small perturbations and is the fastest of all techniques.

2D-3D image registration has been adopted in many clinical applications such as image-guided surgery and the kinematic analysis of bones in knee and ankle joints. The 2D image is typically a single X-ray or Fluoroscopy video frame and the 3D image is typically 3D CT or MRI data of the same anatomical structure. A distinction should be made between those 2D-3D registration algorithms which use two or more 2D views as compared to those that use a single-plane view. Techniques which use multiple 2D views have the advantage of more 2D information to correctly determine the true 3D pose of the anatomical structure. The disadvantage of these techniques is the requirement for two fluoroscopy units to be specially setup for this purpose which is costly and not normally required for standard medical imaging procedures. It also limits the field of view of the images captured to the relatively small area where the two X-ray sources intersect and requires a double dose of radiation for each patient.

The optimization procedures used in previously reported single-plane 2D-3D algorithms include many non-gradient approaches (see for example [5, 6]) and gradient-based approaches using numerically calculated gradients (see for example [7, 8]).

In this paper we propose a new single-plane 2D-3D registration algorithm which requires far less iteration than previous techniques. The new algorithm includes a new multi-modal similarity measure and a novel technique for the analytic calculation of the required gradients.

II. 2D-3D Registration Algorithm

In this paper, the new registration algorithm is described for the application of registering 2D fluoroscopy data to 3D CT data for the kinematic analysis of knee joints. However the new techniques described in the algorithm can be utilized in any 2D-3D registration applications.

There are 6 parameters that describe the 3D rigid-body motion of the object depicted by the 3D data. These 6 parameters control translation in the x, y and z directions (denoted by T_x, T_y and T_z respectively) and rotation about the x, y and z axes (denoted by R_x, R_y and R_z respectively) as shown in Figure 1.

M. R. Pickering and A. A. Muhit are with the School of Information Technology and Electrical Engineering, The University of New South Wales at the Australian Defence Force Academy, Canberra, Australia (e-mail: m.pickering@adfa.edu.au).

J. M. Scarvell and P. N. Smith are with the Trauma and Orthopaedic Research Unit in the Department of Surgery at The Canberra Hospital, Canberra, Australia.

For every iteration of the registration algorithm a 3D rigid-body geometric transform is applied to the CT volume to produce a change in the 3D position of the bone. The 3D volume is then reduced to a 2D digitally reconstructed radiograph (DRR) by summing the voxel values of the transformed CT volume in the *z* direction. Both the DRR and fluoroscopy frames are then filtered using a Laplacian-of-Gaussian (LoG) filter to highlight the edges of the objects. This LoG filtering approach to emphasize the edges of bony structures was also adopted by Ma in [9].

A. Similarity Measure: Sum of Conditional Variances

Assume pixel values of the filtered DRR are denoted by I_i and pixel values of the filtered fluoroscopy frame are denoted by R_i. Here the subscript *i* indicates the pixel value at coordinates (x_i', y_i') and (x_i, y_i) in *I* and *R* respectively for $i = 1 \ldots N$ pixels in the images.

The new multi-modal similarity measure calculates the sum of the conditional variances (SCV) for images *I* and *R* and is given by

$$S(\mathbf{m}) = \sum_j E\left(\left[I_i - E(I_i \,|\, R_i \in \Delta_j) \right]^2 \,\middle|\, R_i \in \Delta_j \right) \qquad (1)$$

where $E(\ )$ denotes the expectation operator, Δ_j denotes a set of histogram bins that span the range of values in *R* and **m** is the vector of motion parameters $[T_x\ T_y\ T_z\ R_x\ R_y\ R_z]^T$ where $[\ ]^T$ denotes the matrix transform operation. The conditional mean required in (1) is calculated using the joint probability distribution of *I* and *R*. Equation (1) can be rewritten in the form

$$S(\mathbf{m}) = \sum_{i=1}^{N} \left(I_i - \hat{R}_i \right)^2 \qquad (2)$$

where

$$\hat{R}_i = E(I_i \,|\, R_i \in \Delta_k) \qquad (3)$$

and Δ_k is the histogram bin which includes R_i.

The expression for *S* in (2) is now in the same form as the sum-of-squared difference (SSD) similarity measure. This form is convenient as it can be used directly in standard non-linear least squares optimization approaches. However, unlike SSD, since $\hat{R}$ is calculated using the joint probability distribution of *I* and *R,* it is possible to use this measure for multi-modal registration problems.

B. Optimization Algorithm: Gauss-Newton

The optimization procedure is required to find the values of the 3D rigid-body transform parameters which minimize *S*. The first step in describing the minimization process is to estimate the values of *S* in a small neighbourhood around the current value of **m** using a second order Taylor series approximation as follows:

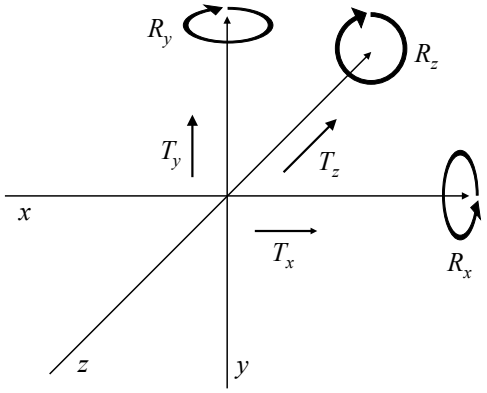


Fig. 1. The 6 parameters that are used to describe 3D rigid-body motion.

$$S(\mathbf{m}+\mathbf{p}) = S(\mathbf{m}) + \mathbf{p}^T \nabla S(\mathbf{m}) + \tfrac{1}{2}\mathbf{p}^T \nabla^2 S(\mathbf{m})\mathbf{p} \qquad (4)$$

The value of the vector **p** that minimizes $S(\mathbf{m}+\mathbf{p})$ in the local neighbourhood of **m** is required at each iteration of the optimization procedure. This vector can be found by setting to zero the derivative of $S(\mathbf{m}+\mathbf{p})$ with respect to **p** and rearranging as follows:

$$\mathbf{p} = -\left(\nabla^2 S(\mathbf{m})\right)^{-1} \nabla S(\mathbf{m}) \qquad (5)$$

The gradient of *S* is defined by

$$\nabla S(\mathbf{m}) = 2\sum_{i=1}^{N} \hat{E}_i(\mathbf{m}) \nabla \hat{E}_i(\mathbf{m}) \qquad (6)$$

where $\hat{E}_i(\mathbf{m}) = I_i - \hat{R}_i$ and the gradient of $E_i(\mathbf{m})$ is given by

$$\nabla \hat{E}_i(\mathbf{m}) = \left[\frac{\partial \hat{E}_i}{\partial T_x} \frac{\partial \hat{E}_i}{\partial T_y} \frac{\partial \hat{E}_i}{\partial T_z} \frac{\partial \hat{E}_i}{\partial R_x} \frac{\partial \hat{E}_i}{\partial R_y} \frac{\partial \hat{E}_i}{\partial R_z} \right]^T \qquad (7)$$

For the Gauss-Newton optimization algorithm, the Hessian matrix $\nabla^2 S(\mathbf{m})$ is approximated as follows:

$$\nabla^2 S(\mathbf{m}) = \mathbf{J}(\mathbf{m})\mathbf{J}(\mathbf{m})^T \qquad (8)$$

where the *N* columns of **J** are given by

$$\mathbf{J}_i = \nabla \hat{E}_i(\mathbf{m}). \qquad (9)$$

Using equations (5) to (9) it is now possible to iteratively update the 3D rigid-body motion parameters until a minimum for *S* is found. For each iteration, an estimate of the parameter update vector **p** is calculated and these updates are then added to the transform parameters. A new version of the filtered DRR *I* is used in the next iteration and the process continues until some threshold is reached, e.g. a minimum change in **m** or *S*.

The proposed 2D-3D registration algorithm also includes a new technique to calculate the gradients of the registration error with respect to *out-of-plane rotations* R_x and R_y for use in (7). The required formulae for these gradients are not easily defined since these parameters describe 3D motion

and must be estimated using only images from a single 2D view of the object.

In the proposed algorithm these gradients are given by

$$\frac{\partial \hat{E}_i}{\partial R_x} = \hat{z}_i \frac{dI_i}{dy_i'} \quad \text{and} \quad \frac{\partial \hat{E}_i}{\partial R_y} = \hat{z}_i \frac{dI_i}{dx_i'} \tag{10}$$

where $\partial I_i / \partial x'$ and $\partial I_i / \partial y'$ are the ith pixels in the horizontal and vertical spatial gradient images of I respectively. The pixel values of $\hat{z}_i$ are an estimate of the z coordinates of the edges of the 3D object and are given by

$$\hat{z}_i = \frac{\sum_z z|V_e|}{\sum_z |V_e|}. \tag{11}$$

where V_e is a Laplacian-of-Gaussian filtered version of the segmented 3D CT data.

To increase the range of initial displacements for which the algorithm will converge to the global minimum, a coarse-to-fine approach is adopted. This is achieved by applying the algorithm using progressively narrower LoG filter kernels. The wider filter kernels produce a much smoother edge image which allows the algorithm to avoid being trapped in a local minimum. The narrower filter kernels have a more limited range for which the algorithm will converge but provide greater accuracy. The parameters produced by the smoother edge image are used as the initial parameters in the following stage which uses a narrower filter kernel. This approach enables the algorithm to successively produce a more accurate registration result.

III. Experimental Results

To evaluate the performance of the new registration algorithm we compared the algorithm with two other techniques. The first was a non-gradient based approach using Powell's conjugate direction optimization algorithm and SCV as the similarity measure. Powell's algorithm has been shown to perform well in many non-gradient based approaches [1]. The second approach was a gradient-based approach using Gauss-Newton optimization and CCRE as the similarity measure. When compared to mutual information, CCRE has been shown to provide faster convergence to the global maximum and a greater range of initial disparities for which successful convergence can be obtained [4, 10]. For this approach the values for the gradient vector and the Hessian matrix were calculated using the approach defined in [11]. All the approaches were performed in three stages with progressively narrower LoG filter kernels used in each stage.

The algorithms were used to register segmented 3D CT data of a female tibia and femur to synthetic fluoroscopy images that were generated using the CT data. We have used low-resolution versions of the original data for this experiment where 1 pixel = 1 mm. The synthetic fluoroscopy images were generated using a logarithmic attenuation function on the sum of the voxel values of the un-segmented CT data along rays that simulate the path of a point source of X-rays. The use of the logarithmic attenuation function and the inclusion of soft tissue and other bones produces a synthetic fluoroscopy image that has almost identical properties to a real fluoroscopy frame. Figure 2 (a) shows the synthetic Fluoroscopy frame of the tibia and Figure 2 (b) shows the DRR produced from the CT data at the same position that was used to generate the fluoroscopy frame. Figure 2 (c) and (d) show the LoG filtered versions of the fluoroscopy frame and the DRR respectively for the LoG filter kernel used in the second stage of the registration algorithm. Note that the bones used in these experiments contained implanted tantalum beads but these were not used in the evaluation process. The exact 3D position of the CT data that was used to produce the synthetic fluoroscopy image is known and is used as the gold standard for measuring registration errors.

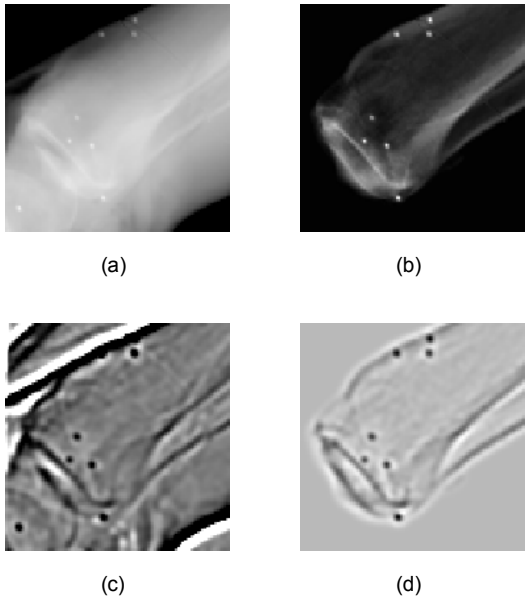
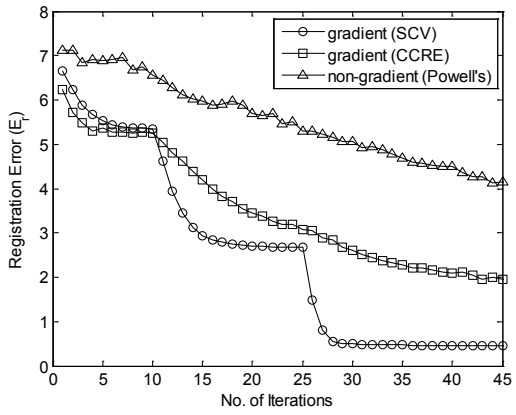


Fig. 2 (a) Synthetic fluoroscopy frame, (b) DRR produced from CT data, (c) LoG filtered version of (a), (d) LoG filter version of (b)

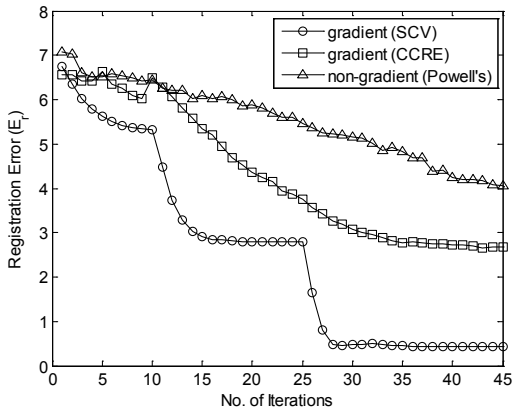
To test the algorithm, a known 3D rigid body transform was applied to the segmented CT data of the bones. Each of the algorithms was then used to register this data to the synthetic fluoroscopy frame. The true 3D rigid body transform parameters that align the CT data with the fluoroscopy frame are all exactly zero, so the goal of the registration algorithms was to produce a final set of transform parameters that are as close to zero as possible. The registration error is therefore defined as the RMS error between the transform parameters and an all zero vector and is given by

$$E_r = \sqrt{T_x^2 + T_y^2 + T_z^2 + R_x^2 + R_y^2 + R_z^2} \tag{12}$$

The values for the translational and rotational parameters were measured in pixels and degrees respectively. It is worth noting that registration error due to translational and



(a)



(b)

Fig. 3 Average registration error at each iteration for the (a) tibia and (b) femur of the knee.

rotational parameters should ideally be expressed separately. However, for conciseness we have combined them to show the cumulative effect. Moreover, our experiments show that this combined error quite accurately reflects the registration error due to individual parameters over iterations.

A set of 100 initial 3D positions for each bone was used with transform parameters uniformly randomly distributed between -5 and +5. The registration algorithm was assumed to have failed if any of the final parameters had a magnitude greater than 3. The proposed approach was successful for 96% of the initial positions while the gradient-based approach using CCRE and the non-gradient approach were successful for only 62% and 49% of the initial positions respectively.

All four initial positions for which the proposed algorithm failed contained a parameter which had a value close to 5. This indicates that the limit to the capture range of the algorithm is approximately ±4 pixels of translation and ±4 degrees of rotation. This range could possibly be extended by using more coarse stages with larger LoG filter kernels.

Figure 3 shows the average registration error at each iteration for the successful registration attempts performed by the three algorithms. For clarity only the first 45 iterations are shown, however the CCRE gradient-based algorithm required between 75 and 100 iterations to converge to a maximum and the non-gradient algorithm required between 150 and 200 iterations to converge. The proposed approach required only 30 to 35 iterations to estimate the true 3D pose of the tibia.

IV. Conclusions

In this paper we have presented a new 2D-3D registration algorithm for applications when only a single 2D view is available. The proposed approach includes a new multi-modal similarity measure that is based on calculating the sum of conditional variances from the joint probability distribution of the two images to be registered. Unlike existing multi-modal similarity measures, this new measure allows the gradient vector and Hessian matrix required for standard non-linear least squares approaches to be calculated using the same procedure as that used for SSD. The algorithm also includes a novel technique for calculating the gradient of the registration error for out-of-plane rotations.

Our experimental results show that, when compared to existing gradient and non-gradient based techniques, the proposed algorithm has a wider range of initial poses for which registration can be achieved and requires significantly fewer iterations to determine the true 3D position of the anatomical structure.

V. References

[1] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Medical Image Analysis,* vol. 3, pp. 373-386, 1999.

[2] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging,* vol. 22, pp. 986-1004, 2003.

[3] N. D. Cahill, et al., "Revisiting overlap invariance in medical image alignment," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008. CVPRW '08, 2008, pp. 1-8.

[4] F. Wang and B. Vemuri, "Non-Rigid Multi-Modal Image Registration Using Cross-Cumulative Residual Entropy," *International Journal of Computer Vision,* vol. 74, pp. 201-215, 2007.

[5] R. Komistek, D. Dennis, and M. Mahfouz, "In vivo fluoroscopic analysis of the normal human knee," *Clinical Orthopaedics & Related Research,* vol. 410, pp. 69-81, 2003.

[6] S. Aouadi and L. Sarry, "Accurate and precise 2D-3D registration based on X-ray intensity," *Computer Vision and Image Understanding,* vol. 110, pp. 134-151, 2008.

[7] R. Munbodh, et al., "Automated 2D--3D registration of portal images and CT data using line-segment enhancement," *Medical Physics,* vol. 35, pp. 4352-4361, 2008.

[8] S. A. Merritt, L. Rai, and W. E. Higgins, "Real-time CT-video registration for continuous endoscopic guidance," in *SPIE Virtual Endoscopy*, 2006, p. 614313.

[9] B. Ma, et al., "2D/3D Registration of Multiple Bones," in 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007, 2007, pp. 860-863.

[10] M. Pickering, Y. Xiao, and X. Jia, "Registration of multi-sensor remote sensing imagery by gradient-based optimization of cross-cumulative residual entropy," *SPIE Image Registration and Change Detection,* vol. 6966, 2008.

[11] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Transactions on Image Processing,* vol. 9, pp. 2083-2099, 2000.