2D-3D Rigid Registration of X-Ray Fluoroscopy and CT Images Using Mutual Information and Sparsely Sampled Histogram Estimators

L. Zöllei*, E. Grimson*, A. Norbash**, W. Wells* **
*Artificial Intelligence Laboratory, MIT
Cambridge, MA 02139
**Dept. of Radiology, Brigham and Women's Hospital
Boston, MA 02115
{lzollei,welg,sw}@ai.mit.edu, anorbash@partners.org

Abstract

The registration of pre-operative volumetric datasets to intra-operative two-dimensional images provides an improved way of verifying patient position and medical instrument location. In applications from orthopedics to neuro-surgery, it has a great value in maintaining up-to-date information about changes due to intervention. We propose a mutual information-based registration algorithm which establishes the proper alignment via a stochastic gradient ascent strategy. Our main contribution lies in estimating probability density measures of image intensities with a sparse histogramming method which could lead to potential speed-up over existing registration procedures and deriving the gradient estimates required by the maximization procedure. Experimental results are presented on fluoroscopy and CT datasets of a real skull, and on a CT-derived dataset of a real skull, a plastic skull and a plastic lumbar spine segment.

1. Introduction

1.1. 2D-3D Medical Image Registration

Registering pre-operative datasets to intra-operative image acquisitions can provide up-to-date information at the treatment site, aiding surgical guidance and other interventions. As three-dimensional intra-procedural image acquisition is uncommon in medical institutions, typically only two-dimensional datasets can be obtained for such purposes. Although these images lack the spatial detail of volumetric data, they have the advantages of faster acquisition time and reduced amount of radiation exposure to both patients and doctors. Ideally, one can recover the advantages of the volumetric data by aligning the intra-operative 2D images with pre-operative volumes.

For the specific case of fusing CT and X-ray images, the primary focus is on registering bony structures, since both modalities best visualize such information. Characterizing the rigid movement of bones implies six degrees of freedom. If we also wished to align finer details, such as soft tissues, we would define higher-dimensional transformations.

For medical image registration, a key challenge is to introduce an appropriate way to compare input images that are of different dimensionalities. The most common approach is to simulate one of the modalities given the other dataset and the current transformation estimate, so that the images can be compared in the same space. Then the transformation estimate is updated to maximize an alignment score according to some objective function. Reconstructing the 3D volume from 2D images is one alternative, but it requires numerous projection acquisitions and large computation time. It is more feasible to simulate 2D images from the 3D volume. Most existing applications follow this approach. Simulated projection images, whose construction is to model the production of X-ray acquisitions, are called Digitally Reconstructed Radiographs (DRRs). These images are formed by casting rays through the imaged 3D volume and summing up the attenuation coefficients of each voxel along the rays. Since during a registration procedure it might be necessary to evaluate the objective function a large number of times, several research studies have concentrated on defining practical methods for radiograph computation, such as voxel-projection [8] and shear-warp factorization [9].

In a registration system, the quality of registration is scored by similarity functions that aim to measure how well the input images align. Common registration methods can be grouped into two categories, feature-based and intensity-based. Feature-based approaches rely on the presence and identification of natural landmarks or fiducial markers to determine the best alignment. The most significant features need to be segmented in both of the input images and the alignment criterion is then optimized with respect to them. Contour- and point-based techniques [5, 6] are good examples of this approach. Although the reduced number of features to be registered provides computational speedup (after

the segmentation procedure is done), a major drawback of these methods lies in the inconvenience of planting artificial markers on the patient and the dependence on the segmentation procedure that can potentially introduce errors. Often they also require some level of user interaction.

Intensity-based measures operate on the pixel or voxel intensities directly. Though the number of points to be aligned is greater than in the above feature-based methods, no feature extraction step is required. An extensive study of intensity-based similarity measures used specifically in 2D-3D applications has evaluated the performance of six different target functions (normalized cross-correlation, entropy of the difference image, pattern intensity, mutual information, gradient correlation and gradient difference) in matching single X-ray fluoroscopy and corresponding CT images [4] . The imaged organ was a phantom spine, where only a user-defined region of interest (usually an individual vertebra) was registered. The study found that the best objective functions for such a multi-modal registration scenario are pattern intensity and gradient difference. These proved to be the most robust with respect to the presence of soft tissue and of a surgical instrument appearing only on one of the modalities. Both of these objective functions were implemented to use the whole input image in order to evaluate the current quality of alignment.

Usually, examining only a single 2D image is not sufficient to recover all six degrees of freedom robustly. In-plane rotation and displacement components can be approximated, but the out-of-plane transformations are difficult to accurately recover. In order to reliably determine all of the transformation components, it is often preferable to use two or more 2D acquisitions [2, 7, 8].

The most commonly used 2D medical image modalities for the 2D-3D alignment task are X-ray fluoroscopy and portal images. The creation of portal images employs high-energy treatment radiation beams instead of low-energy imaging radiation, hence they could be considered byproducts of a procedure and their quality is extremely poor. Fluoroscopic images are taken by an X-ray machine and they capture much more detail about the examined anatomy (Figure 1). Their major disadvantage stems from the fact that without correction, their geometric accuracy degrades due to pincushion and radial distortion effects.

1.2. Our Registration Approach

We are interested in registering biplanar 2D fluoroscopic images to a 3D CT volume. As the region of interest is the whole imaged organ, the full input volume is considered in the registration task. The geometry of the imaging environment is assumed to be known, so the location of the two imaging sources is fixed. By updating the transformation components, we adjust the imaging coordinate system to best approximate the observed fluoro images with DRR
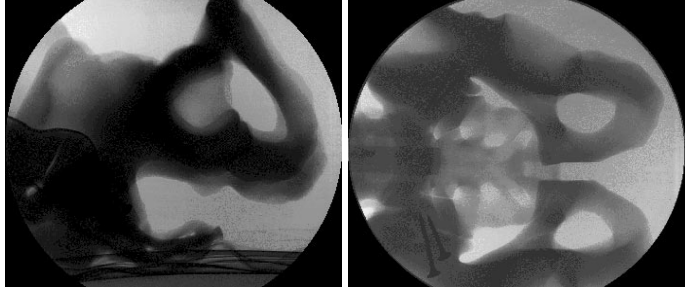

Figure 1: Lateral and AP Acquisitions of X-Ray Fluoroscopic Images of the Pelvis Phantom.

projections from the 3D volume. An information theoretic measure called mutual information (MI) [10, 14] is used as our objective function. It has been used in many applications, as it does not assume a linear relationship between the random variables being compared. It has been successfully applied to 3D-3D multi-modal and 2D-3D video-frame to model surface registration applications [11, 12, 13]. We decided to use that particular objective function because its useage allows for generalizing the code for image modalities other than the ones currently presented and because it is possible to conveniently approximate its value even without using all available information provided by the inputs.

To evaluate the robustness of MI with respect to all of the transformation variables, we ran extensive probing experiments, verifying that the objective function peaked when the correct alignment was evaluated. To automatically locate the transformation that corresponds to the best alignment, we use a gradient ascent procedure. Our choice of an optimization method was driven by efficiency constraints. Although good registration results could be achieved by the non-gradient-based Powell-method, the average execution time ranged between 3-3.5 hrs. We preferred a gradient-guided search because of its computational efficiency. Using all corresponding pixel pairs to obtain intensity information, however, is still a redundant and time intensive procedure: hence we apply a stochastic gradient estimation strategy which uses noisy approximations of the required measures instead of relying on the true gradients. We also estimate probability distributions of image intensities by a sparse ray-casting method (using only 100-150 rays) as opposed to by full DRR construction.

1.3. Paper Summary

We first define the transformation which establishes the relationship between the imaging and data coordinate systems. We then introduce our objective function and the stochastic gradient optimization theme, and we derive the update calculations used to iteratively approximate the optimal transformation components. Finally, we present our

registration results on fluoro and CT images of a real skull and more thoroughly evaluate the performance of the algorithm on CT-derived datasets of a plastic and real skull, a phantom lumbar spine and a plastic pelvis.

2. Background

2.1. Transformation Description

The transformation parameter T which relates the coordinate frames of the imaging environment and the data volume is constructed from a rotational and a translational component. However, in order to distinguish constant and variable elements, we decompose T into three individual sub-transforms.

$$T = D_c \circ R \circ D_d$$

D_c is a constant displacement term that is responsible for positioning the data volume into the center of its own coordinate system (so that rotation may be performed around the origin). R encodes the rotational component required to perform the match, and translation D_d positions the object in the imaging coordinate system. As we specify T to be the transformation that expresses imaging coordinates in terms of data coordinates, the appropriate order of the subtransforms is D_d, R and D_c (Figure 2). Decoupling the components of the transformation in such a way is useful because it makes the parameter space more directly searchable.
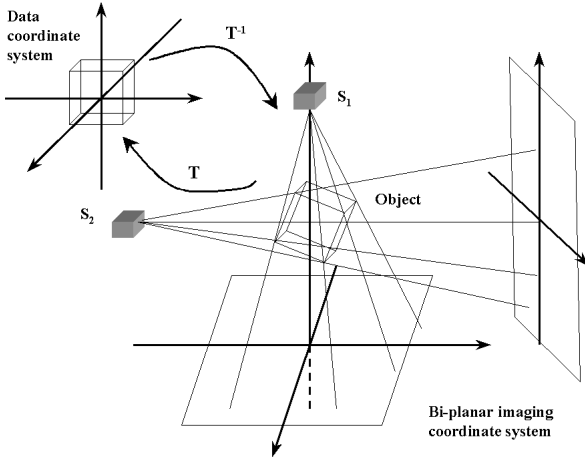

Figure 2: The transformation parameter T which relates the coordinate frames of the imaging environment and the data volume; $T = D_c \circ R \circ D_d$.

2.2. Objective Function: Mutual Information

Our preferred choice of objective function is a sum of mutual information-based measures which evaluates the matches with respect to the two input fluoroscopic images. If the fluoro images are denoted as $U_1(X)$, $U_2(X)$ and the simulated DRRs (created from the CT volume given the current transformation estimate) are $V_1(T(X))$ and $V_2(T(X))$, we write:

$$g = I(U_1(X), V_1(T(X))) + I(U_2(X), V_2(T(X))). \quad (1)$$

The goal is to find T that maximizes g.

The mutual information of two discrete random variables expresses how much the knowledge about one of these variables increases the information about the other. More informally, instead of assuming a linear relationship between the values of the random variables that are compared, it proposes that, in the registration problem, with perfect alignment the intensity values from the corresponding images maximally explain each other. In order to write its definition, we introduce another closely related notion, the Shannon entropy. The entropy of a random variable describes the uncertainty associated to that variable and is defined as the expected value of the negative log probability [3].

$$H(A) = E_A[-\log p(A)]$$

Mutual information can be expressed as the sum of individual entropy terms of the random variables less their joint entropy.

$$I(A, B) = H(A) + H(B) - H(A, B) = \quad (2)$$
$$E_{(A,B)}[\log p(A, B)] - E_A[\log p(A)] - E_B[\log p(B)] \quad (3)$$

3. Optimization

In order to find the local maximum of the mutual information-based objective function, we use a stochastic version of the gradient ascent procedure. Gradient ascent techniques are maximization methods whose local search for the optimal parameter settings is guided by calculations of the objective function gradient. The stochastic approximation [15, 16, 17] approach uses noisy estimates of the derivatives instead of the true ones (in order to increase computational efficiency). The stochastic nature of our algorithm originates from two sources: we approximate the probability distributions of our variables by a sparse histogramming approach and we use various simplifications to compute the derivatives.

At each iteration of the algorithm, we use the current estimate of transformation T to simulate the creation of two X-ray images and we compute the goodness of their alignment with the corresponding observed fluoro images. To find the local extremum of the objective function, we compute the partial derivative of it with respect to the transformation parameters. Then to improve our estimate of T, we write:

$$T_{\text{updated}} = T_{\text{current}} + \lambda * \frac{\partial g}{\partial T_{\text{current}}}. \quad (4)$$

In Eq.(4), λ, represents the learning rate (or step-size) of the algorithm, and determines the magnitude of an individual update operation. Finding the appropriate range for λ forms a crucial part of the experiments. If the magnitude is too small, convergence takes a long time, however, if it is set to be too large, convergence to the searched value might not even occur. One way to avoid fixing an ideal value for the learning rate is to vary it over time. This practice is called *annealing*, and it assigns decreasing values to λ as the iteration number increases. Another approach, which we take advantage of, is the application of a hierarchical structure. Step size settings are less sensitive when registration takes place on successively increasing levels of resolution. We differentiate between the learning rates of rotational and displacement components. Further distinction could be made between components corresponding to in-plane and out-of-plane operations.

The second part of the update term in Eq.(4) is the derivative of our objective function g. Although g is the sum of mutual information terms, for simplicity, we will describe the update calculations with respect to only one of the mutual information terms. Using Eq.(3) to express MI, we get

$$
\begin{aligned}
&\frac{\partial I(U(X), V(T(X)))}{\partial T} \\
&= \frac{\partial H(V(T(X)))}{\partial T} - \frac{\partial H(U(X), V(T(X)))}{\partial T} \\
&= \frac{\partial \{ E_{(u,v)}[\log(p(u,v))] - E_v[\log(p(v))] \}}{\partial T}, \quad (5)
\end{aligned}
$$

where we denote $U(x)$ as u and $V(T(x))$ as v.

The first approximation step in our algorithm results from the fact that we estimate statistical expectation terms with sample averages. If the size of the sample collection is M and x_i represents the ith sample point in the collection, $u_i = U(x_i)$ and $v_i = V(T(x_i))$,

$$
\begin{aligned}
&\frac{\partial I(U(X), V(T(X)))}{\partial T} \\
&= \frac{1}{M} \sum_{i=1}^{M} \left[\frac{1}{p(u_i, v_i)} \frac{\partial p(u_i, v_i)}{\partial T} - \frac{1}{p(v_i)} \frac{\partial p(v_i)}{\partial T} \right] (6)
\end{aligned}
$$

3.1. Density Estimators and their Partials

The probability density functions of image intensities in Eq.(6) are estimated from the data using 1D and 2D histograms. Using all available information about corresponding image pixels is undesirable as it requires the generation of the full DRR which is computationally expensive. We chose to experiment with a sparse sampling approach where we draw only a small number of samples from the overlapping image regions and base our probability estimates on them. The smaller the sample size we use, the faster the

estimations become. However, with each intensity value ignored we trade off accuracy (and possibly convergence of the algorithm). Experimentally, building histograms from 100-150 random samples and using 32 intensity bins on intensity range 0-255 proved to be adequate. Decreasing the bin sizes did not improve the registration results.

In order to complete the optimization task, the density estimators need to be differentiated (Eq.(6)). To carry out that task on histogram-estimated probability distributions, we adopted ideas introduced for dense histograms in [12]. Given a histogramming function f approximating the probability density function of random variable A based upon a collection of sample points B, the probability of $a \in A$ is given by

$$
p(a) \approx f(a, B).
$$

and the derivative of f with respect to variable s is estimated according to

$$
\frac{d}{ds} f(a, B) = \frac{\partial}{\partial a} f(a, B) \frac{da}{ds} + \frac{\partial}{\partial B} f(a, B) \frac{dB}{ds}. \quad (7)
$$

The application of the chain rule in Eq.(7) makes an implicit assumption. It holds only for cases when the histogram estimator function f is not explicitly dependent on the variable s with respect to which the derivative is taken. Although this assumption is not quite valid in our scenario (the histograms depend on transformation T with respect to which we take derivatives), empirically, it was established that infinitesimal changes in the parameters of T are unlikely to (greatly) alter the nature of the density estimator. Hence, we apply the above simplification. Furthermore, the last term on the right of Eq.(7) can be ignored if differential changes in the sample intensities in B result in vanishingly small changes in the density values estimated by f. Based on our experiments, that condition also held for sparse sampling. Utilizing these assumptions and after some algebraic manipulations, the terms in Eq.(6) can be expressed as:

$$
\frac{\partial p(u_i, v_i)}{\partial T} \approx \frac{\partial p(u_i, v_i)}{\partial v_i} * \frac{\partial v_i}{\partial T}. \quad (8)
$$

and

$$
\frac{\partial p(v_i)}{\partial T} \approx \frac{\partial p(v_i)}{\partial v_i} * \frac{\partial v_i}{\partial T}. \quad (9)
$$

The terms in Eq.(8) and (9) correspond to changes in the DRR image intensity values resulting from modifications in the transformation parameters and to changes in the probability densities as a result of changes in sample intensities. We approximate the derivatives of the probability densities by the use of finite differences calculated from their corresponding histogram estimates. Deriving the $\frac{\partial v_i}{\partial T}$ term is more complex, and the details are explained below. We emphasize that at each iteration of the algorithm we perform an update which corresponds to a *small* angle rotation and

small displacement applied to the current transform estimate. As among the components of transformation T, only rotation R and displacement D_d need to be recovered, we only take partial derivatives with respect to these terms.

For calculations of the update elements, we introduce a new term, the *update* rotation R_u. This operator represents a *small* angle rotation r which adjusts the value of the current rotation estimate at the end of each iteration cycle. (Note that at the beginning of each iteration of the registration algorithm r is reset to be a zero angle rotation.) Hence we write the new rotation component of the transform as $(R_u \circ R)$ and the transformation itself as

$$T = D_c \circ R_u \circ R \circ D_d.$$

A transformed point is

$$T(x) = T(r, d, x) = D_c(R_u(r, R(D_d(d, x)))). \quad (10)$$

From Eq.(8) and (9), we need to compute

$$\frac{\partial V(T(x_i))}{\partial T} = \left\{ \frac{\partial V(T(x_i))}{\partial r}; \frac{\partial V(T(x_i))}{\partial d} \right\}. \quad (11)$$

In the following calculations, the vector r encodes a rotation transform such that its magnitude determines the angle of rotation and its direction stands for the axis of rotation. In order to express these partial derivative terms, we use the ray-casting algorithm to model the formation of the fluoro image intensities. (The ray-casting algorithm is used instead of a more efficient procedure, as we only sample a small fraction of the image intensities and the whole image is not constructed.) In particular, a sample of the simulated fluoroscopic image at location x_i on the image plane (or at $T(x_i)$ in data coordinates) is approximated as

$$V(T(x_i)) = \sum_{z \in \text{ray}(T(x_i), S)} \text{Vol}(z),$$

where ray refers to the line segment which connects the imaging source S with $T(x_i)$ on the imaging plane and z indicates uniformly distributed steps along that ray within the volume. As the steps are located in the transformed coordinate space, we could write $z = T(y) = T(r, d, y)$. Therefore,

$$\frac{\partial V(T(x_i))}{\partial T} = \sum_{z \in \text{ray}(T(x_i), S)} \frac{\partial \text{Vol}(T(r, d, y))}{\partial T}. \quad (12)$$

3.2. Update wrt Displacement

We first calculate the partial derivative of the volume intensity with respect to the ith component of the displacement component d, denoted as d_i. In Eq.(13) and (18), e_i stands

for a unit vector whose components are all zero except for the ith one which equals 1.

$$\left[\frac{\partial}{\partial d_i} \text{Vol}(T(y)) \right] =$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(D_c(R_u(r, R(D_d(d,y)))))}{\partial d_i} =$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(R_u(r, R(D_d(d,y))))}{\partial d_i} =$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(R_u(r, R(y+d)))}{\partial d_i} =$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(R_u(r, R(y)) + R_u(r, R(d)))}{\partial d_i} =$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(R_u(r, R(d)))}{\partial d_i} =$$
$$\nabla \text{Vol}(T(y)) \bullet (R_u(r, R(e_i))) \quad (13)$$

The full expression is

$$\frac{\partial}{\partial d} \text{Vol}(T(y)) = \begin{bmatrix} \nabla \text{Vol}(T(y)) \bullet (R_u(r, R(e_1))) \\ \nabla \text{Vol}(T(y)) \bullet (R_u(r, R(e_2))) \\ \nabla \text{Vol}(T(y)) \bullet (R_u(r, R(e_3))) \end{bmatrix}. \quad (14)$$

3.3. Update wrt Small Angle Rotation

We derive the partial derivative terms of the volume intensities with respect to the rotation component of transformation T similarly to the above. First we only consider the ith element of r, which we denote as r_i.

$$\left[\frac{\partial}{\partial r_i} \text{Vol}(T(y)) \right] =$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(D_c(R_u(r, R(D_d(d,y)))))}{\partial r_i} = \quad (15)$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(R_u(r, R(D_d(d,y))))}{\partial r_i} = \quad (16)$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(R(D_d(d,y)) + r \times R(D_d(d,y)))}{\partial r_i} = \quad (17)$$
$$\nabla \text{Vol}(T(y)) \bullet \frac{\partial(r \times R(D_d(d,y)))}{\partial r_i} =$$
$$\nabla \text{Vol}(T(y)) \bullet (e_i \times R(D_d(d,y))) = \quad (18)$$
$$e_i \bullet (R(D_d(d,y)) \times \nabla \text{Vol}(T(y))) =$$
$$(R(D_d(d,y)) \times \nabla \text{Vol}(T(y)))_i$$

Hence, with respect to the full vector r,

$$\frac{\partial}{\partial r} \text{Vol}(T(y)) = R(D_d(d,y)) \times \nabla \text{Vol}(T(y)). \quad (19)$$

We note two of the steps in the above derivation. First, Eq.(16) is a result of a simplification to the formula in Eq.(15). As the constant displacement operation D_c only happens after the rotation, this has no effect on the partial derivatives that are being calculated. That term disappears from the numerator. Secondly, to arrive at Eq.(17), we use the fact that R_u is strictly defined to stand for small angle rotations. In that case can we make the assumption that a coordinate point p, after such a rotation can be expressed in the form:

$$p' = R_u(r, p) = r(p) = p + r \times p. \quad (20)$$

4. Algorithm Outline & Results

The top-level outline of the registration code is summarized as follows:

1. *Preprocessing the input images and input volume*: During the preprocessing step, we smooth the fluoro images to better match the resolution of the CT-derived DRR images and eliminate all artificial labels from them that were placed there for patient identification. We also define the desirable window and level settings for the CT volume. These determine the range and average value of intensities which are usually set by radiologists after image acquisition.

2. *Initializing the imaging environment*: The initialization step involves reading in the parameters that are known about the imaging environment.

3. *Registration loop* : for fixed number of iterations – for all fluoro-DRR image pairs

   (a) Randomly sample image points from the fluoro image and extract their intensities ($U(X)$ where X is the sample collection)

   (b) Calculate corresponding DRR values ($V(T(X))$) by ray-casting

   (c) (*) Evaluate the objective function (see Eq.(1))

   (d) Compute transformation updates and assign new transformation estimate according to Eq.(19) and (14)-(23)

$$T_{1\,updated} = T_1 + \frac{\lambda}{M} \sum_{j=1}^{2} \sum_{i=1}^{M} \left[K_{i,j} * L_{i,j} \right] \quad (21)$$

$$K_{i,j} = \sum_{z \in \text{ray}(T_j(x_i), S)} \frac{\partial \text{Vol}(z)}{\partial T} \quad (22)$$

$$L_{i,j} = \left(\frac{1}{p(u_i, v_{ji})} \frac{\partial p(u_i, v_{ji})}{\partial v_{ji}} - \frac{1}{p(v_{ji})} \frac{\partial p(v_{ji})}{\partial v_{ji}} \right) \quad (23)$$

The step indicated with a (*) is not an integral part of the algorithm. We evaluate the MI estimate at each iteration just to analyze the convergence behavior as a function of iterations, but it is not a required step in the maximization procedure. The algorithm currently runs for a fixed number of iterations. We set the iteration number to be 5000. However, with an appropriate set of learning rates, it usually takes less then half that many rounds to arrive at the optimal set of parameters. We could also halt the registration procedure if convergence were detected.

4.1. Registration results

We first present registration results obtained from controlled experiments. High-quality, CT-derived DRR datasets of a plastic pelvis, plastic skull, real skull, real head and plastic lumbar spine were registered to their volumetric counterparts. The data specifications for the CT datasets are listed in Table 1. The first column contains row, column, and slice information in that order and the second column specifies voxel dimensions (dx, dy, dz). Some of the *observed* DRR images corresponding to these settings are displayed in Figure 3.

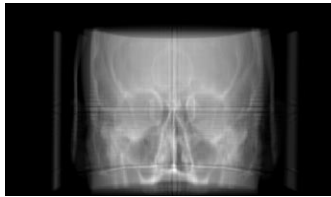| CT Data Specifications:		
plastic pelvis	265 x 455 x 107	[0.662109; 0.662109; 2.0]
plastic skull	188 x 128 x 105	[1.015620; 1.015620; 2.0]
plastic spine	512 x 512 x 103	[0.234375; 0.234375; 1.5]
real head	512 x 512 x 127	[0.488281; 0.488281; 1.0]
real skull	512 x 512 x 424	[0.447300; 0.447300; 0.5]

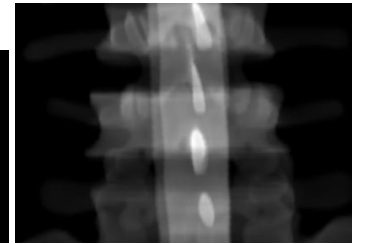Table 1: CT Dataset Specifications.
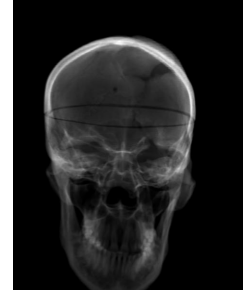
(a) Real skull   (b) Pelvis phantom

(c) Real head   (d) Lumbar spine segment

Figure 3: Example DRRs from the Controlled Experiments.

Given the ground-truth pose parameter T_{GT}, the capture range of the algorithm with respect to a particular dataset is established by finding the greatest perturbation of individual components that could be reset by our application. In the controlled experiments, we found that a displacement offset of 35-40 mm and rotation angle offset of 20-25 degrees could generally be registered. The quality of the registration results is determined by calculating an *error* transformation,

T_{error}, which is defined in

$$T_{GT} = T_{error} \circ T_{output}.$$

REG	DATA	LEV	d_e (mm)	r_e (deg)	TIME
Hi	Pelvis	sm0	0.5368	0.1071	3620
		sm1	0.3170	0.1031	2422
		sm2	1.4763	0.1576	1388
		sm3	1.9451	0.2131	699
	R. Skull	sm0	0.8145	0.0869	4666
		sm1	1.0348	0.1331	3553
		sm2	11.5438	1.1923	2820
Pow	Pelvis	sm0	1.1164	0.1472	25358
		sm1	2.9557	0.1495	4839
		sm2	5.3905	0.3676	1478
		sm3	13.6580	1.298	240
	R. Skull	sm1	2.7144	0.1432	13744
		sm2	5.2431	0.4744	2583
		sm3	10.9750	1.1474	561

Table 2: Registration results of the histogramming (*Hi*) and the Powell (*Pow*) methods on controlled experiments of a phantom pelvis and a real skull. Hierarchy levels (with decreasing resolution): sm0, sm1, sm2, sm3

Table 2 displays the average accuracy and running time of a subset of our experiments after 12 runs. The same experiments were run using the Powell and the newly introduced histogramming optimization methods in order to compare their performance. At the beginning of the experiments, the rotation angle estimate was randomly perturbed relative to the ground truth value in the range of $(0°, 20°)$ and the displacement estimates were offset by $0 - 15$ mm in each direction. In the hierarchical framework, we downsampled and smoothed the input image volume by a factor of 2 at each level. The table entries display the average magnitude of the displacement component and of the rotation angle of the registration T_{errors}. As the resolution level increases, the error measures decrease resulting in sub-volume accuracy at the resolution of the original input. The table entries also show that the new method converges quicker to the optimal solution as the running time for the Powell method largely increases with each level of resolution. (In the case of the non-gradient method, the experiments were sometimes not even completed on the top level of the pyramid as the running time became so high.)

We monitored the convergence behavior of the alignment procedure by displaying the MI-estimates and the T_{error} components at each iteration. One such convergence plot is demonstrated by Figure 4. This presents a real skull experiment, where one of the translational components was perturbed by 35 mm. Although the MI curve nicely converges,

it is also quite noisy, which is explained by the fact that it is only evaluated on a small-sized, random set of samples.
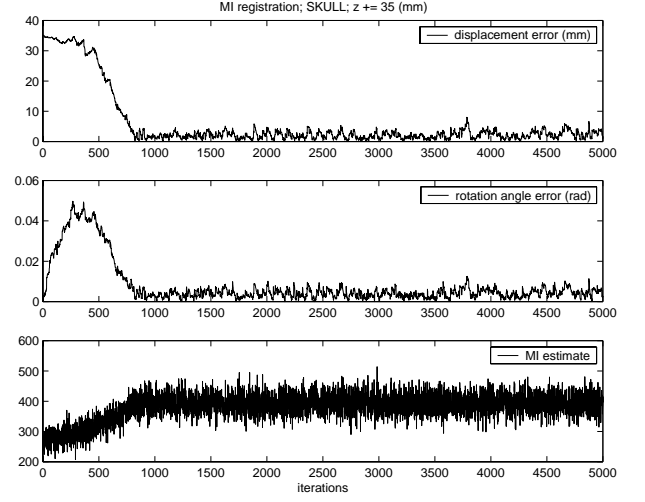

Figure 4: Sample output from a controlled run of the algorithm. Dataset: real skull; Initial displacement offset: 35 mm. Plots display the magnitude of displacement error, rotation angle and the MI estimate at each iteration.

We carried out registration experiments with real fluoro images as well. The same validation/accuracy measures as the ones discussed with respect to the controlled datasets however are not yet available. This is due to the fact that ground-truth transformation measure could only be estimated by manual registration. Therefore, in Figure 5, we display one of the promising real fluoro experiments. The input fluoro image pair is overlaid with the main edge contours of the projection images created by both the initial and final transformation estimates. We can see that while at the outset of the algorithm the DRR outlines do not really fit the edges in the fluoro acquisitions, the edges at the final stage nicely match the fluoro boundaries. Qualitative evaluation of the algorithm will be available as soon as data with sufficient ground-truth information is obtained.

5. Conclusion

We have introduced a novel approach to MI-based 2D-3D rigid-body registration. We improved the execution time of the algorithm by estimating probability density functions with sparse histogramming and by optimizing our similarity function via a stochastic gradient ascent strategy. We presented promising experimental results on accurately registering controlled datasets of real and plastic organs and demonstrated fast convergence on fluoroscopy-CT images of a real skull.
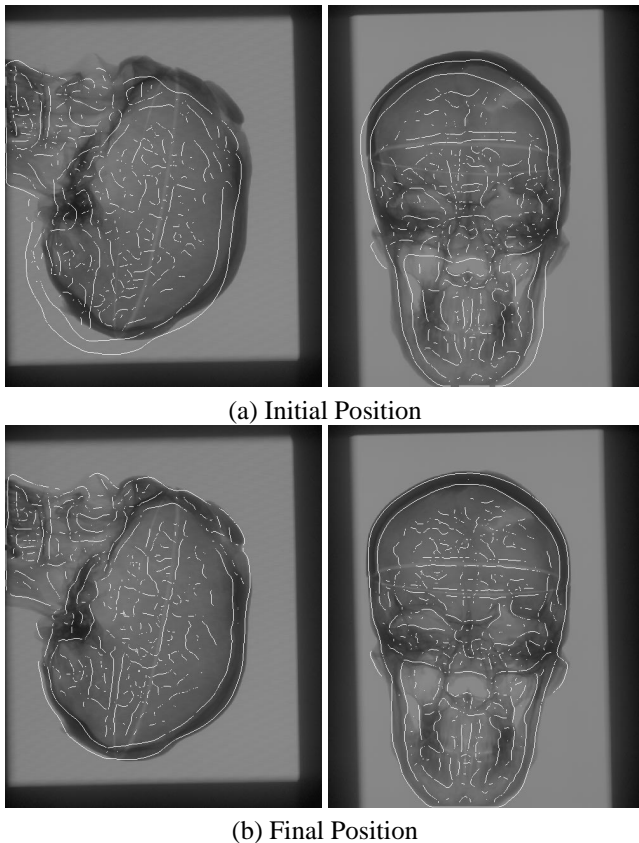
(a) Initial Position

(b) Final Position

Figure 5: Biplanar Fluoro Images Overlaid with DRR Outlines in Initial (a) and Registered (b) Positions.

Acknowledgments

References

[1] R. Bansal, L.H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath and J. Duncan "A Novel Approach for the Registration of 2D Portal and 3D CT Images for Treatment Setup Verification in Radiotherapy" *Proceedings of MICCAI 1998*, p 1075-1086

[2] R. Bansal, L.H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath and J. Duncan "Entropy-Based, Multiple-Portal-to-3D CT Registration for Prostate Radiotherapy Using Iteratively Estimated Segmentation" *Proceedings of MICCAI 1999*, p 567-578

[3] T.M. Cover, J.A. Thomas "Elements of Information Theory" *John Wiley & Sons, Inc* 1991

[4] G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill, D.J. Hawkes "A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration" *IEEE Transactions on Medical Imaging*, 17(4), 1998, pp. 586-595

[5] T.S.Y. Tang, R.E. Ellis, G. Fichtinger "Fiducial Registration from a Single X-ray Image; A New Technique for Fluoroscopic Guidance and Radiotherapy" *Proceedings of MICCAI 2000*, p 502-511

[6] Z.R. Yaniv "Fluoroscopic X-ray Image Processing and Registration for Computer-Aided Orthopedic Surgery" *Master's Thesis*, The Hebrew University of Jerusalem; 1998

[7] D. LaRose, L. Cassenti, B. Jaramaz, J. Moody, T. Kanade, A. DiGioia "Post-operative Measurement of Acetabular Cup Position Using X-ray/CT Registration" *Proceedings of MICCAI 2000*, p 1104-1113

[8] E. Cosman,Jr. "Rigid Registration of MR and Biplanar Fluoroscopy" *Master's Thesis*, Massachusetts Institute of Technology; 2000

[9] J. Weese, R. Gocke, G.P. Penney, P. Desmedt, T.M. Buzug, H. Schumann "Fast voxel-based 2D/3D registration algorithm using a volume rendering method based on the shear-warp factorization" *Medical Imaging 1999: Image Processing; Proceedings of SPIE* Vol 3661, 1999, p 802-810.

[10] P. Viola and W. Wells. "Alignment by Maximization of Mutual Information." *Proceedings of the 5th International Conference on Computer Vision*, 1995

[11] W. Wells, P. Viola, H. Atsumi, S. Nakajima and R. Kikinis. "Multi-modal Volume Registration by Maximization of Mutual Information." *Medical Image Analysis* 1996; 1:35-52.

[12] W. Wells, M. Halle, R. Kikinis, P. Viola "Alignment and Tracking using Graphics Hardware" *Image Understanding Workshop, DARPA* 1996; 837-842.

[13] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer,Jr.,R.M. Kessler, R.J. Maciunas, et al. "Comparison and evaluation of retrospective intermodality image registration techniques" *Journal of Computer Assisted Tomography* Jul/Aug 1997; vol. 21 , p. 554-566

[14] A. Colignon et al. "Automated multi-modality image registration based on information theory" *Proceeding of Information Processing in Medical Imaging Conference*, p.263-174; Kluwer Academic Publisher, 1995

[15] B. Widrow, M. Hoff "Adaptive Switching Circuits" *1960 IRE WESCON Convention Record* vol 4, p. 96-104; IRE, New York

[16] L. Ljung, T. Soderstrom "Theory and Practice of Recursive Identification" *MIT Press* 1983

[17] S. Haykin "Neural Networks: A comprehensive Foundation" *Macmillan College Publishing*

[18] J.J. Craig *Introduction to Robotics; Mechanics and Control* Addison-Wesley Publishing Company, 1986